

Artificial Intelligence in Ophthalmology

Andrzej Grzybowski
Editor

Second Edition

Artificial Intelligence in Ophthalmology

Andrzej Grzybowski
Editor

Artificial Intelligence in Ophthalmology

Second Edition

 Springer

Editor

Andrzej Grzybowski
Department of Ophthalmology
University of Warmia and Mazury
Olsztyn, Poland

Institute for Research in Ophthalmology
Foundation for Ophthalmology Development
Poznan, Poland

ISBN 978-3-031-83755-5 ISBN 978-3-031-83756-2 (eBook)
<https://doi.org/10.1007/978-3-031-83756-2>

1st edition: © Springer Nature Switzerland AG 2021

2nd edition: © The Editor(s) (if applicable) and The Author(s), under exclusive license to
Springer Nature Switzerland AG 2025

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

If disposing of this product, please recycle the paper.

Contents

1	Artificial Intelligence in Ophthalmology: Promises, Hazards and Challenges	1
	Andrzej Grzybowski	
2	Basics of Artificial Intelligence for Ophthalmologists	19
	Ikram Issarti and Jos J. Rozema	
3	Overview of Artificial Intelligence Systems in Ophthalmology	37
	Natsuda Kaothanthong, Niracha Arjkongharn, Nattaporn Vongsa, Varis Ruamviboonsuk, and Paisan Ruamviboonsuk	
4	Autonomous Artificial Intelligence Safety and Trust	69
	Michael D. Abramoff	
5	Technical Aspects of Deep Learning in Ophthalmology	83
	Ashkan Abbasi, Sowjanya Gowrisankaran, Wei-Chun Lin, Gadi Wollstein, Joel S. Schuman, and Hiroshi Ishikawa	
6	Threats to the Validity of the Predictive System Performance for Ophthalmology	93
	Michał Woźniak and Andrzej Grzybowski	
7	Experimental Artificial Intelligence Systems in Ophthalmology: An Overview	103
	Luis de Sisternes, Joelle A. Hallak, Kathleen Romond, and Dimitri T. Azar	
8	Artificial Intelligence in Age-Related Macular Degeneration (AMD)	121
	Souvik Mukherjee, Yifan Peng, Qingyu Chen, Tiarnan D. L. Keenan, Emily Y. Chew, and Zhiyong Lu	
9	AI and Glaucoma	137
	Sowjanya Gowrisankaran, Ashkan Abbasi, Wei-Chun Lin, Gadi Wollstein, Joel S. Schuman, and Hiroshi Ishikawa	

10	Artificial Intelligence in Retinopathy of Prematurity	153
	Brittni A. Scruggs, Adam M. Hanif, Michael F. Chiang, and J. Peter Campbell	
11	Artificial Intelligence in Diabetic Retinopathy	169
	Andrzej Grzybowski and Piotr Brona	
12	Application of AI in Angle Closure Diagnosis and Management	195
	Tin Aung and Xiulan Zhang	
13	Automatic Retinal Imaging and Analysis: Age-Related Macular Degeneration (AMD) Within Age-Related Eye Disease Studies (AREDS)	217
	T. Y. Alvin Liu and Neil M. Bressler	
14	Artificial Intelligence for Keratoconus Detection and Refractive Surgery Screening	223
	José Luis Reyes-Luis, Ana Silvia Serrano-Ahumada, and Roberto Pineda	
15	Artificial Intelligence in the Treatment of Keratoconus: The Use of Intracorneal Ring Segments	235
	C. Fariselli, F. Versace and J. L. Alio	
16	Artificial Intelligence for Cataract Management	243
	Haotian Lin, Xiaohang Wu, Lixue Liu, and Zizheng Cao	
17	Artificial Intelligence in Refractive Surgery	251
	Yan Wang, Haohan Zou, Pinghui Wei, Mohammad Alzogool, and Xuan Chen	
18	Artificial Intelligence in Cataract Surgery Training.	261
	Nouf Alnafisee, Waverly Rose Brim, Sidra Zafar, Bassell Hammoud, Kristen Park, S. Swaroop Vedula, and Shameema Sikder	
19	Deep Learning Applications in Ocular Oncology	273
	T. Y. Alvin Liu, Neslihan Dilruba Koseoglu, and Zelia M. Correa	
20	Artificial Intelligence Applied to Neuro-Ophthalmic Conditions.	281
	Samy Zaher, Raymond P. Najjar, and Dan Milea	
21	Artificial Intelligence Using the Eye as a Biomarker of Systemic Risk	287
	Jinyuan Wang, Rachel Marjorie Wei Wen Tseng, Tyler Hyungtaek Rim, Carol Y. Cheung, and Tien Yin Wong	
22	Overview of Intraocular Lens Power Calculation Formulas Based on Artificial Intelligence	305
	Wiktor Stopyra and Andrzej Grzybowski	

23	Relevance and Impact of AI in the Pharmaceutical Industry—Focus on Clinical Development	319
	Andreas Maunz, Yun Yvonna Li, Carl Glittenberg, and Sascha Fauser	
24	Artificial Intelligence in the Diagnosis of Dry Eye	333
	Mohammad Hassan Emamian, Roqayeh Aliyari, Carla Lanca, and Andrzej Grzybowski	
25	Artificial Intelligence in Refractive Errors	349
	Carla Lanca, Mohammad Hassan Emamian, and Andrzej Grzybowski	
26	Artificial Intelligence and Prediction of Eye Diseases	373
	Tahereh NaseriBooriabadi, Mohammad Hassan Emamian, and Andrzej Grzybowski	
27	Leveraging Artificial Intelligence (AI) to Enhance Nursing Care for Patients Undergoing Eye Surgery	417
	Morteza Shamsizadeh	
28	Exploring the Potential of ChatGPT in Ophthalmology: A Vision for Future Healthcare.	433
	Anfisa Ayalon, Lauren M. Wasser, Andrew M. Williams, and José-Alain Sahel	
29	Analysis of International Publication Trends in Artificial Intelligence in Ophthalmology (2019–2023)	443
	Kai Jin, Wenyue Shen, Lu Yuan, Andrzej Grzybowski, and Juan Ye	
30	Machine Learning Studies in Ocular Oncology	461
	David E. Pelayes and Arun D. Singh	
31	Application of Artificial Intelligence in Oculoplastics	469
	Yilu Cai, Xuan Zhang, Jing Cao, Andrzej Grzybowski, Juan Ye, and Lixia Lou	
	Index.	483

Artificial Intelligence in Ophthalmology: Promises, Hazards and Challenges

1

Andrzej Grzybowski

*“If you do not get feedback, your confidence grows much faster than your accuracy”
Tetlock P., Gardner D. Superforecasting:
The Art and Science of Prediction,
Crown Publishing, 2016.*

protection, testing AI algorithms on data sets that did not correspond real-world conditions, and its vulnerability to cybersecurity attacks. Finally, some aspects of cost-effectiveness of AI-based devices are presented.

Abstract

The chapter presents the overview of the present achievements and applications of artificial intelligence in medicine, including the possible benefits for future medicine. It also presents the increase in published scientific studies using artificial intelligence in medicine in last decade.

The regulation frameworks for medical devices, including AI medical devices, in the USA and in the European Union is discussed. Moreover, the problem of access to reliable data is given with the explanation of transfer learning, generative adversarial networks, and continual machine learning. Some of the hazards, and challenges of the AI in ophthalmology are described, including privacy

Keywords

Medicine · Ophthalmology · Artificial intelligence · Deep learning · Machine learning · Artificial intelligence in ophthalmology · AI devices regulations · AI devices safety

The Promise of Artificial Intelligence

The term “artificial intelligence” (AI) was coined on August 31, 1955, when John McCarthy, Marvin L. Minsky, Nathaniel Rochester, and Claude E. Shannon submitted “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence.” [1, 2]. It was, however, Alan Turing who during a public lecture in London in 1947 mentioned computer intelligence, and in 1948 he introduced many of the central concepts of AI in a report entitled “Intelligent Machinery.” [3]. Moreover, Turing proposed in 1950 the test, originally called the imitation game, and later known as the Turing

A. Grzybowski (✉)
Department of Ophthalmology, University of
Warmia and Mazury, Olsztyn, Poland
e-mail: ae.grzybowski@gmail.com

A. Grzybowski
Institute for Research in Ophthalmology, Foundation for
Ophthalmology Development, Poznan, Poland

test, as a way to confirm that the intelligent behavior of a machine was equivalent to that of a human. A human evaluator is asked to determine the nature of a partner (human or machine) based on a text-only conversation [1–3]. After decades of slow progress since the Turing test was proposed, AI has finally blossomed. Many new technologies and applications are available, and there is great enthusiasm about the promise of AI in health care. It holds the potential to improve patient and practitioner outcomes, reduce costs by preventing errors and unnecessary procedures, and provide population-wide health improvements. We have entered the fourth stage of the Industrial Revolution that began in the eighteenth century, and its defining feature may well be the use of AI technologies (Fig. 1.1). The results of an annual competition known as the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) provide interesting insights into recent developments in AI technology (Fig. 1.2). Over the years 2010–2016 there was a steady decrease in the error rates of the algorithms presented, and in 2017, 29 of the 38 competing teams had error rates lower than 5% (considered to be the human threshold). Thus in 10 years AI algorithms exceeded human performance in image recognition.

There are many promising applications for AI in health care, addressing a variety of aims and taking many different approaches (Table 1.1). For example, misdiagnoses constitute a huge, although poorly recognized, medical problem. A

study published in 2014 estimated that diagnostic errors affect at least 5% of US adults (12 million people) per year [4]. More recently a systematic review and meta-analysis reported that the rate of diagnostic errors causing adverse events among hospitalized patients was 0.7% [5]. Furthermore, diagnostic error is the most important reason for malpractice litigation in the United States, accounting for 31% of malpractice lawsuits in 2017 [2]. The creation of AI programs to identify and analyze diagnostic errors could be an important step in addressing this problem [6].

Eric Topol has proposed that AI could help shift into “deep medicine,” by allowing the physicians to devote more time to crucial relationships with their patients—an aspect of medicine that cannot be replaced by any AI technology [2]. It is also interesting to consider whether AI might enrich the doctor-patient relationship, enabling a shift from the present “shallow medicine” into “deep medicine,” based on deep empathy and connection [2]. Success in building such relationships is very much related to the amount of time doctors can spare for patients and the extent of the personal contact they have with their patients. The average time of a clinic visit in the United States for an established patient is 7 min and for a new patient 12 min. In many Asian countries, clinic visits last as little as 2 min per patient [2]. Making this situation even worse, part of this time must be devoted to completing electronic health records, further limiting personal contact. A study published in

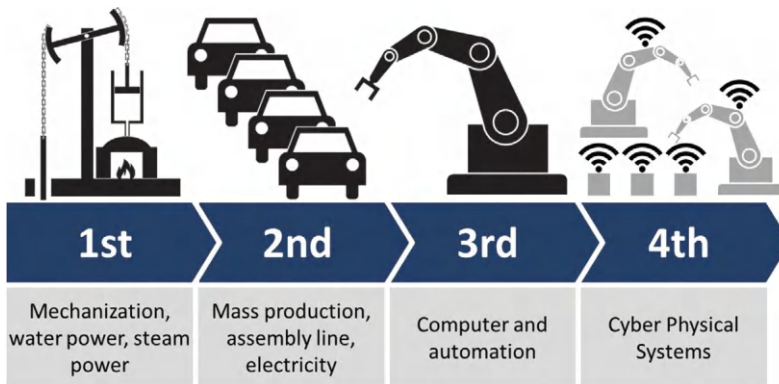


Fig. 1.1 The four main stages of the Industrial Revolution that began in the eighteenth century

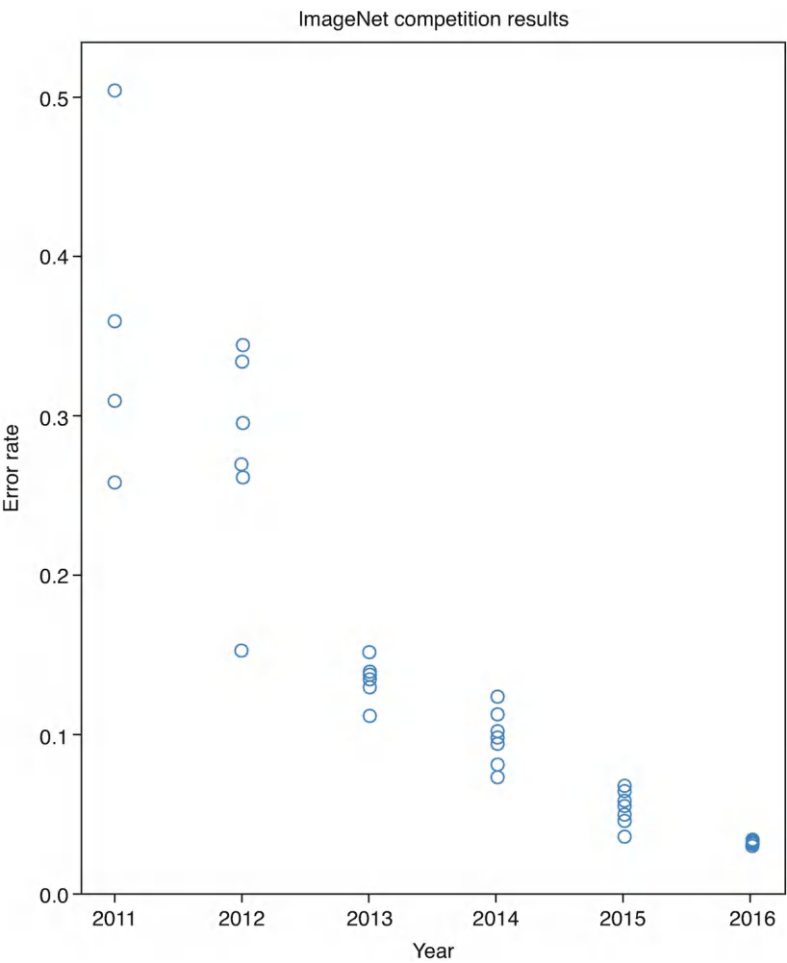


Fig. 1.2 Error-rate history on ImageNet

Table 1.1 Some ambitious expectations for AI in health care.

- outperform doctors,
- help to diagnose what is presently undiagnosable,
- help to treat what is presently untreatable,
- to recognize on images what is presently unrecognizable,
- predict the unpredictable,
- classify the unclassifiable,
- decrease the workflow inefficiencies,
- decrease hospital admissions and readmissions,
- increase medication adherence
- decrease patient harm
- decrease or eliminate misdiagnosis

Adapted from Topol E. Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again. Basic Books, New York 2019

2017 that asked patients to describe how they perceive their physician found that the most common negative responses were “rushed,” “busy,” and “hurried.” [7]. These reactions are manifestations of “shallow medicine.”

One of the arguments supporting the use of AI in medicine is that human cognitive capacity to effectively manage information is often exceeded by the quantity of data generated. Each year the world produces zettabytes of data (roughly, enough to fill a trillion smartphones) [2]. Moreover, unlike humans, who have bad days and emotions, and who get tired, with subsequent decreases in performance and accuracy, AI works 24/7 without vacations or complaints [2].

AI-based technologies employing deep-learning (DL) approaches have proven effective in supporting decisions in many medical specialties, including radiology, cardiology, oncology, dermatology, ophthalmology, and others. For example, AI/DL algorithms (also referred to as AI/DL models in the following text) have been shown to reduce waiting times, improve medication adherence, customize insulin dosages, and help interpret magnetic resonance images. The number of AI life-science papers listed in PubMed increased from 596 in 2010 to 12,422 in 2019 [8]. The number of papers on the use of AI in the field of ophthalmology also has increased dramatically (Figs. 1.3 and 1.4).

AI/DL algorithms have been used to detect diseases based on image analysis, with fundus photos and optical coherence tomography (OCT) scans analyzed for retinal diseases, chest radiographs assessed for lung diseases, and skin photos analyzed for skin disorders. Retinal photos have also been used to identify risk factors related to cardiovascular disorders, including blood pressure, smoking, and body mass index [9]. Using DL models trained on data from over 280,000 patients and validated on two independent data sets, Poplin et al. predicted cardiovascular risk factors not previously thought to be

present or quantifiable in retinal images, such as age (mean absolute error within 3.26 years), gender (area under the receiver operating characteristic curve=0.97), smoking status (AUC=0.71), systolic blood pressure (mean absolute error within 11.23 mmHg) and major adverse cardiac events (AUC=0.70) (Fig. 1.5) [9].

The COVID-19 pandemic has raised expectations for the use of AI in data analysis. So far it has been used in epidemic modeling, detection of misinformation, diagnostics, vaccine and drug development, triage and patient outcomes, and identification of regions of greatest need [10].

Regulating AI-Based Medical Devices: Demonstrating Benefit and Safety

One of many challenges in the field of AI is determining what constitutes evidence of impact and benefit for AI medical devices and who should assess the evidence [2]. The majority of AI studies are conducted in experimental conditions and based on preselected data. They might provide inadequate insight into the use of AI applications in heterogeneous, real-world care settings.

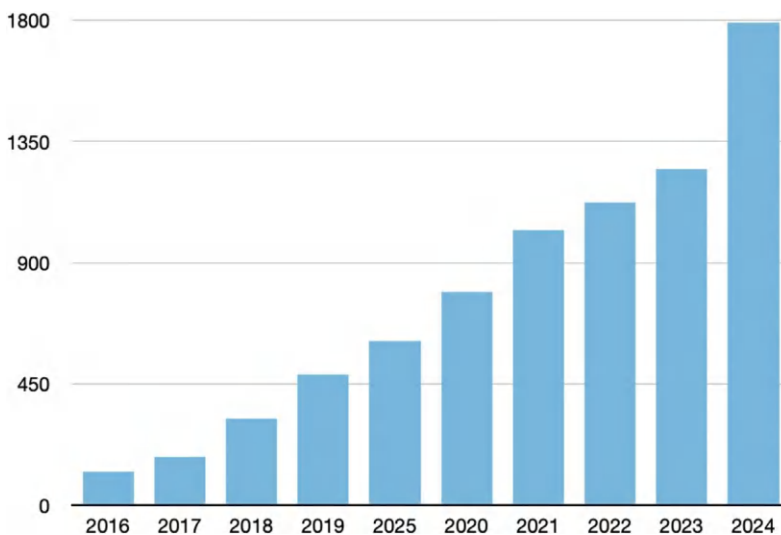


Fig. 1.3 The number of PubMed articles on Artificial Intelligence (AI) and the eye that were published between 2016 and 2024

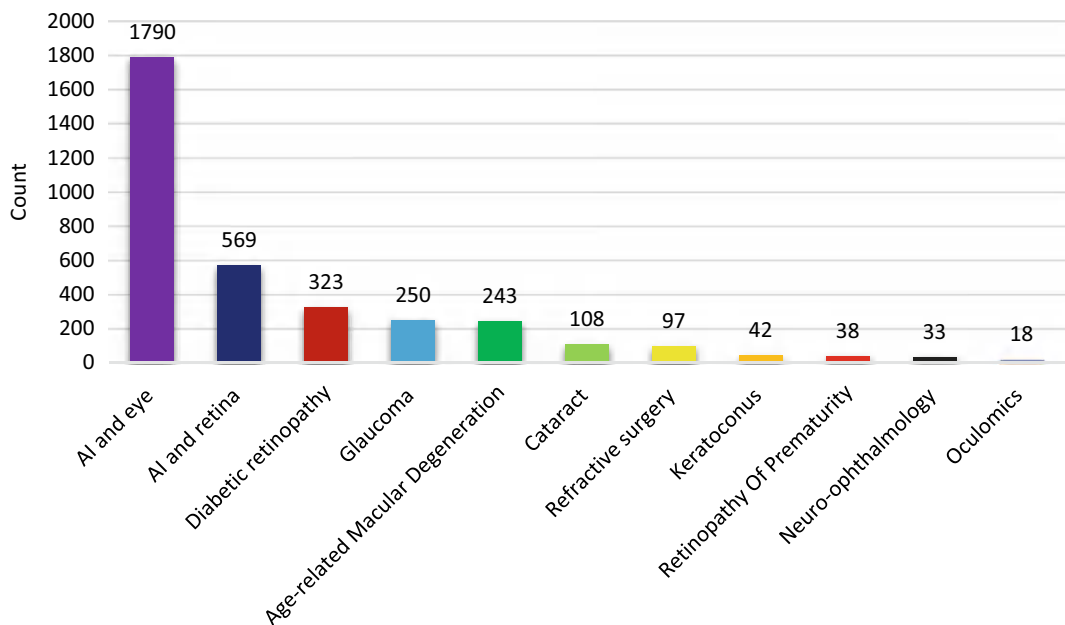


Fig. 1.4 The number of Pubmed articles relating to AI and eye diseases published in 2024

Lee et al. tested seven algorithms being used clinically around the world, including one with US Food and Drug Administration (FDA) approval and four whose developers have submitted applications for FDA approval. They found that most of these algorithms performed worse in real-world, compared with experimental, situations, with only three of seven and one of seven having comparable sensitivity and specificity to the human graders, respectively. Only one algorithm performed as well as human graders [11]. Another of the algorithms tested performed significantly worse than human graders at all levels of DR severity—it missed 25.58% of cases of advanced retinopathy, which could have serious consequences. One of the potential hazards of the clinical use of algorithms identified in this study was the risk of applying an algorithm trained on a particular demographic group to a population that differs in factors such as ethnicity, age, and sex. Moreover, many studies of algorithms developed with AI have excluded low-quality images, treating them as ungradable images, and patients with comorbid eye diseases, making them less reflective of real-world conditions.

The study by Lee et al. shows the importance and limitations of the registration process of AI-based medical devices. FDA registration is based on a centralized system, which does not have a specific, easily accessible regulatory pathway for AI-based medical devices. FDA clears the medical devices through three pathways: the premarket approval pathway, the de-novo premarket review, and the 510(k) pathway [12, 13]. The leading AI disciplines in medicine are radiology, cardiology, internal medicine/endocrinology, neurology, ophthalmology, emergency medicine, and oncology. FDA approvals of AI-based medical devices have increased steadily in recent years; there were 9 in 2015, 13 in 2016, 32 in 2017, 67 in 2018, and 77 in 2019, with the majority of devices designed for use in radiology, cardiology, and neurology [12]. Interestingly, 85% of FDA-approved medical devices in the years 2015–2019 were intended for use by health-care professionals, and only 15% for use by patients. The best-known, FDA-approved, AI-based medical devices in the field of ophthalmology are IDx-DR (2018), the first software to provide screening decisions that do not have to be interpreted by a clinician, and

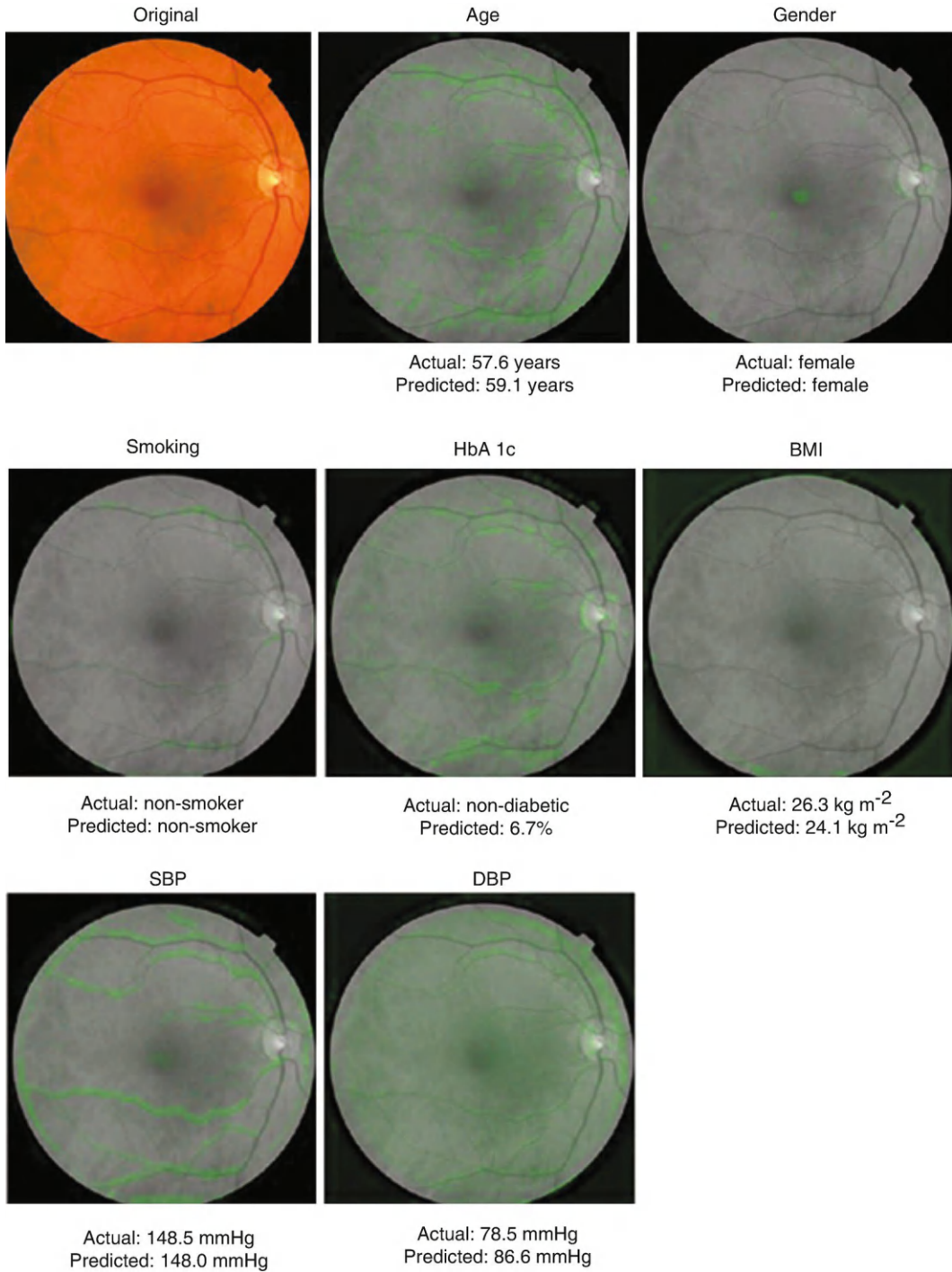


Fig. 1.5 Attention maps for a single retinal fundus image. The top left image is a sample retinal image in color from the UK Biobank data set. The remaining images show the same retinal image in black-and-white. The soft attention heat map for each prediction is overlaid in green, indicating the areas of the heat map that

the neural-network model is using to make the prediction for the image. Source: Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, Peng L, Webster DR. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng.* 2018 Mar;2(3):158–164

Eyenuk (2020), which, like IDx-DR, screens for diabetic retinopathy.

In European Economic Area, which includes the European Union (EU) countries and the European Free Trade Association (EFTA) members (Iceland, Lichtenstein, Norway, and Switzerland), medical devices are approved in a decentralized manner.

Conformité Européenne (CE) marking indicates conformity with EU health, safety, and environmental-protection standards. For the low-risk medical devices (CE class I), the manufacturer ensures that the product complies with regulations and an approval procedure is not required. The registration procedure for higher-risk medical devices (CE class IIa, IIb, and III) is handled by private entities, called notified bodies, that have been accredited to assess the devices and issue a CE mark.

Thirteen CE-marked AI-based medical devices were approved in 2015, 27 in 2016, 26 in 2017, 55 in 2018, and 100 in 2019. The majority were designed for use in radiology, general hospital care, cardiology, neurology, ophthalmology (12 devices), and pathology, and most were class IIa (40%), class I (35%), or class IIb (12%) devices [12]. Of the AI-based devices that were CE-marked between 2015 and 2019, 124 (52%) were also FDA approved, making up 56% of the AI-based tools that the FDA approved in those. Bigger companies were more likely to get both approvals, whereas smaller companies were more likely to obtain only a CE mark. The authors of this study suggested that the European approval system was less rigorous than the US one. This conclusion is supported by an FDA report on 12 devices that received CE approval only and later were found to be unsafe or ineffective [13, 14]. A major problem in studying CE-marked devices in the European Economic Area is the lack of a publicly available register of approved devices comparable to the FDA register. Moreover, the information submitted to the notified bodies is confidential. In 2022, a new European database on medical devices (Eudamed), providing a live picture of the lifecycle of medical devices, will become operational. It will be composed of six modules,

including actor registration, unique device identification (UDI), device registration, notified bodies and certificates, clinical investigations and performance studies, and vigilance and market surveillance [15].

Access to Reliable Data

DL algorithm training requires large data sets with thousands or even hundreds of thousands of diverse, well-balanced, and accurately labeled images [16]. The resources required for an AI study are presented in Fig. 1.6. The enormous numbers of required images rarely can be obtained from individual centers; thus they are secured from data repositories or centers that agree to share data. There is a growing need for consensus on standardized definitions of medical entities; conventions for data formatting; identification of units of measure; protocols for data cleaning, harmonization, and validation; standards for sharing and reusing data and sharing of code implementing AI models; and the adoption of open application program interfaces to AI models [17]. This is required for data sharing and open communication in AI, which is critical for conducting the reproducible research that is necessary before AI technology can be adopted in health care.

Keremany et al. used a DL analysis of a data set of optical coherence tomography images for triage and diagnosis of choroidal neovascularization, diabetic macular edema, and drusen. They demonstrated performance comparable to that of human experts and provided a more transparent and interpretable diagnosis by highlighting the regions recognized by the neural network. Further, they showed that a transfer-learning approach produced only modestly worse results (twofold increase of error, compared with the full data set) while using approximately 20 times fewer images. They also demonstrated the wider utility of this approach by applying it to the identification of pediatric pneumonia using chest X-ray images. They provided their data and code in a publicly available database to facilitate their use by other

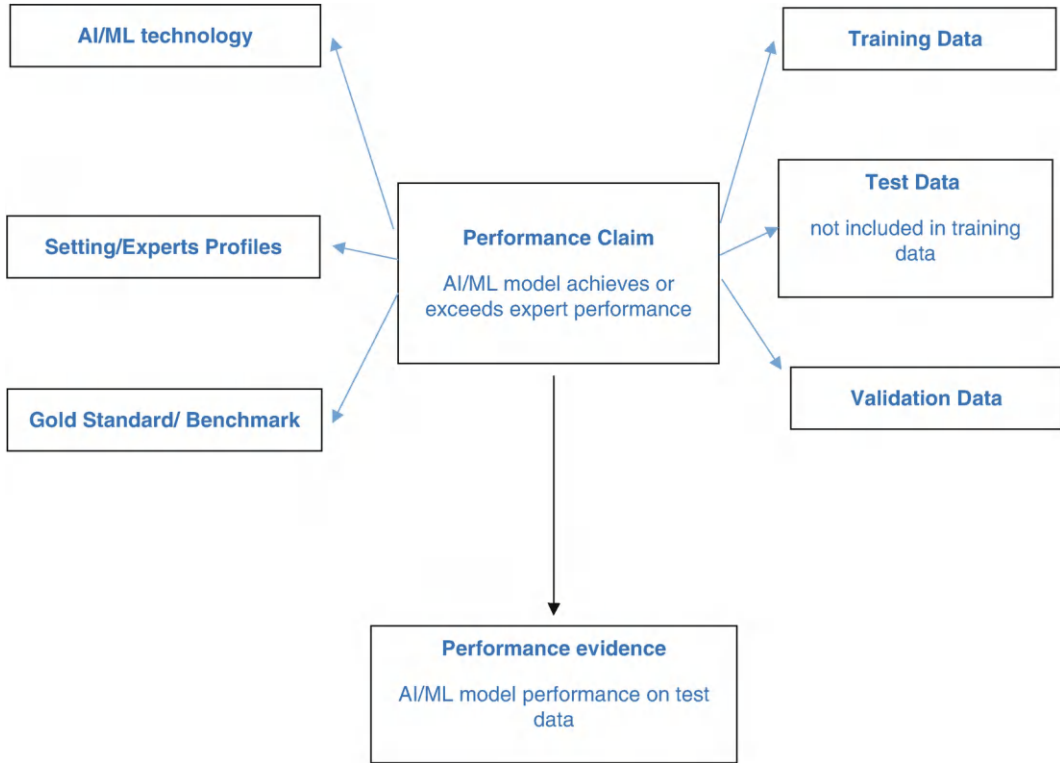


Fig. 1.6 The schematic presentation of resources required for an AI study

biomedical researchers in order to improve the performance of future models [18].

Transfer learning (Figs. 1.7 and 1.8) has been used in recent years to build classification models for medical images because the number of images that can be used for training is relatively small compared to the number of images available to train general models [19] (Fig. 1.9). Another approach to meet the need for large, annotated training data sets might be the use

of low-shot DL algorithms. Low-shot learning (LSL), also known as few-shot learning, is a type of machine learning (ML) problems where the training dataset contains limited information. It is well known that many real-life situations, including rare diseases (e.g., serpiginous choroidopathy or angiod streaks in pseudoxanthoma elasticum) and non-typical presentations or subtypes of common disorders, are prone to AI bias due to the paucity or imbalance of data.

Transfer learning: idea

Instead of training a new model from scratch a specific task:

- Take a model trained on a different domain for a different source task
- Adapt the knowledge accumulated in the model for your domain and your target task

Variants

- Same domain, various tasks
- Various domains, same task

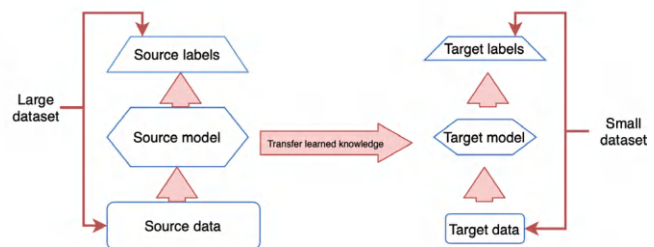


Fig. 1.7 The idea of transfer learning

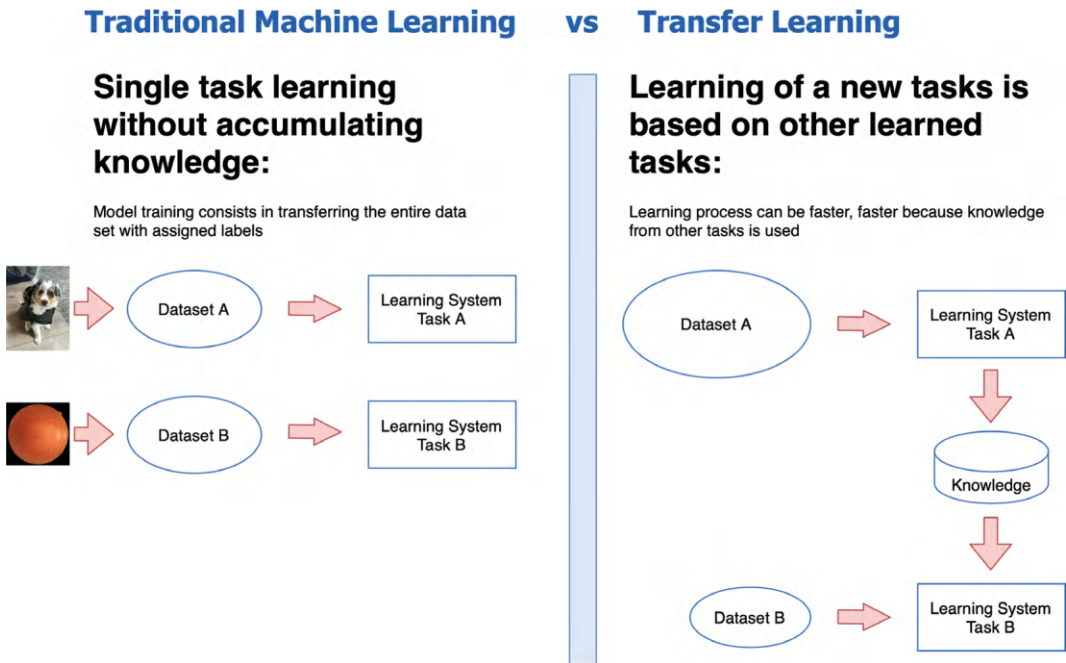


Fig. 1.8 Schematic of a convolutional neural network and transfer learning

These deficiencies may also result in less accurate future models. When addressing this sort of bias, dividing data according to some patient features (e.g., age, sex, and race/ethnicity) may result in smaller data sets that may be insufficient for training models for these particular groups. The study by Burlina et al. showed that the performance of widely used DL methods degraded substantially when used with limited data sets, but LSL methods performed better and might be applied in retinal diagnostics when a limited number of retina images are available for training [20].

Another approach that has been suggested by several authors to address the problem of limited data sets is the use of generative adversarial networks (GANs) to synthesize new images from a training data set of real images. GANs are ML models that can generate new data with the same statistics as the training set (Fig. 1.10). For example, a GAN trained on photographs can generate photographs of non-existing persons that look as authentic as real humans (Fig. 1.11). Artificial photos can be found at <https://this-person-does-not-exist.com>. Many applications of

GAN have been proposed, including, art, fashion, advertising, science, video games, however, concerns about malicious uses were also raised, e.g., to produce fake, possibly incriminating, photographs and videos.

Burlina et al. used the Age-Related Eye Disease Study data set of over 130,000 fundus images to generate a similar number of synthetic images to train DL models. The performance of DL models trained with the synthetic images was nearly as good as the performance of models trained on real images [21]. Liu et al. have shown that 92% of synthetic OCT images had sufficient quality for further clinical interpretation. Only about 26–30% of synthetic post-therapeutic images could be accurately identified as synthetic images (Fig. 1.8) [22]. The accuracy of models trained on synthetic images to predict wet or dry macular status was 0.85 (95% CI 0.74–0.95) [22]. In a study by Zheng et al., the image quality of real versus synthetic OCT images was similar as assessed by two retinal specialists. The accuracy of discrimination of real versus synthetic OCT images was 59.50% for retinal specialist 1 and 53.67% for retinal

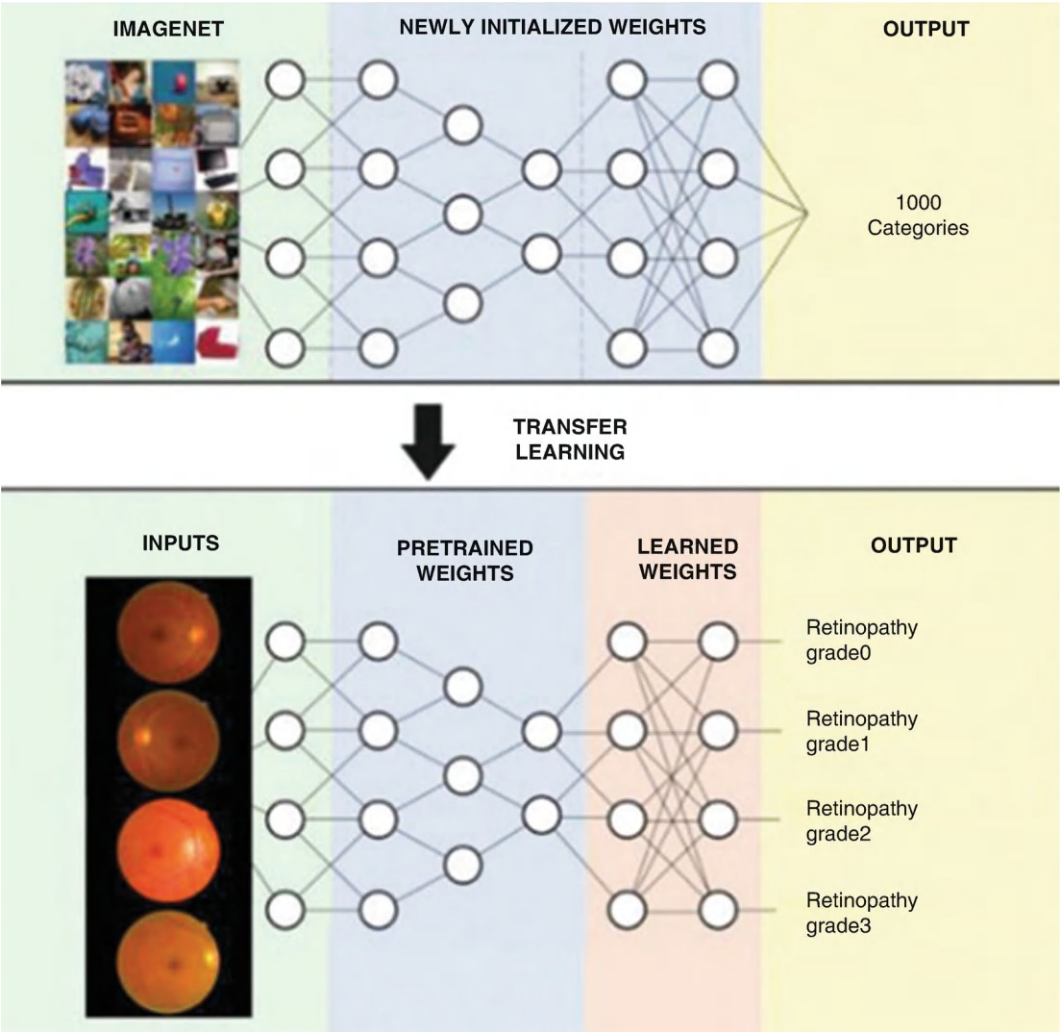


Fig. 1.9 The schematic diagram of transfer learning. Source: Lingling Li et al. Diabetic retinopathy identification system based on transfer learning. 2020. *J. Phys.:* *Conf. Ser.* 1544 012,133. <https://doi.org/>. Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence

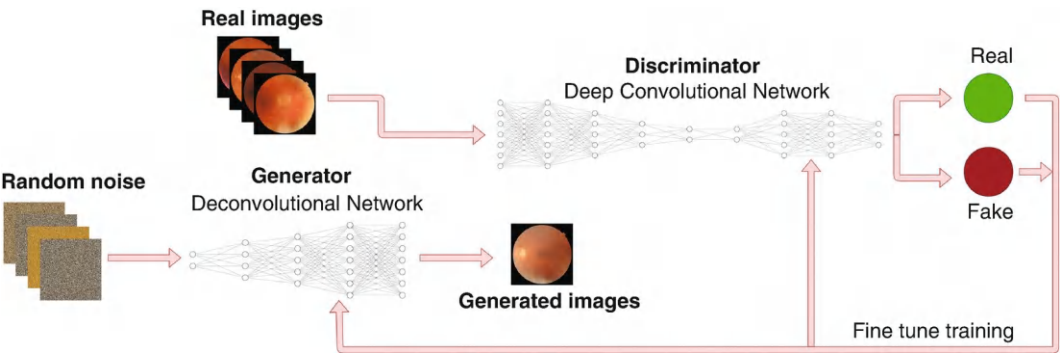


Fig. 1.10 The schematic presentation of generative adversarial network (GAN)



Fig. 1.11 The image of a young woman generated by StyleGAN, an generative adversarial network (GAN). The person in this photo does not exist, but is generated by an artificial intelligence based on an analysis of portraits. Source: https://commons.wikimedia.org/wiki/File:Woman_1.jpg. This file is in the public domain because, as the work of a computer algorithm or artificial intelligence, it has no human author in whom copyright is vested

specialist 2. For the local data set, the DL model trained on real and synthetic OCT images had an area under the curve of 0.99 and 0.98, respectively. For the clinical data set, the area under the curve was 0.94 for the real model and 0.90 for the synthetic one [23]. These studies suggest the GAN synthetic images can be used by clinicians for educational purposes and developing DL algorithms [24].

An important and interesting issue is the clinical application of continual ML, i.e., continuous learning and development from new data while retaining previously learned knowledge [25]. However, there are technical challenges to the implementation of this promising concept, including the need to prevent interference between new and old data, and old and new knowledge. In the catastrophic interference phenomenon, the acquisition of new data can lead to an abrupt decrease in the performance of an algorithm. Practical applications of AI tools in health care must be cautiously introduced because of the existence of such risks. FDA

regulations require that FDA-approved autonomous algorithms be locked for safety to prevent unpredictable future changes. This requirement, however, is designed to ensure the safety of the model rather than improving its performance. Continual learning could refine the performance of machine-learning algorithms by the gradual correction and elimination of mistakes. It will be necessary to consider how this technology can be introduced safely to health care.

Hazards and Challenges of AI in Ophthalmology

The future development of ophthalmology depends on better and possibly unlimited access to the medical data stored within electronic health records. However, this access cannot be allowed to compromise of privacy of this very sensitive data. There is a need for effective regulations that will set a balance between individual protection and the common good. One approach to protecting privacy and increasing sample size is to share DL algorithms with local institutions for retraining purposes, but without sharing the private data used to build the algorithms. This model-to-data approach, also known as federated learning, was tested in ophthalmology and was shown to work effectively [26].

According to the US National Institute of Standards and Technology, biometric data, including retina images, are personally identifiable information and should be protected from inappropriate access. Although AI models have been shown to diagnose and stage some ocular diseases from fundus photographs, OCT, and visual-field images, most AI algorithms were tested on data sets that did not correspond well to real-world conditions. Patient populations were usually homogenous, and poor-quality images and patients with multiple pathologies were excluded. Future studies are needed to validate algorithms on ocular images from heterogeneous populations, including both good- and poor-quality images. Otherwise, we may face the situation of “good AI gone bad.” The tendency to cherry-pick the best results might

make the situation even worse. AI algorithms can behave unpredictably when applied in real life. Algorithm performance can degrade after deployment due to the changes between the training and testing conditions (dataset shift), caused, for example, by using to images generated by a different device than this in the training set or collected in a different clinical environment [27–30]. Moreover, algorithms may return different outputs at different times when presented with similar inputs [31, 32]—they can be affected by minor changes in image quality or extraneous data on an image [32–35]. All these problems might lead to misdiagnosis and erroneous treatment suggestions, breaching trust in AI technologies. An error in an AI system could harm hundreds or even thousands of patients.

A recent report from the National Academy of Medicine [36] highlights some important challenges in the further development of AI applications in health care (Table 1.2). Its

authors advocate the use of openly accessible, standardized, population-representative data; addressing explicit and implicit biases related to AI; developing and deploying appropriate training and educational programs for health workers to support health-care AI; and balancing innovation and safety through the use of regulation and legislation to promote trust.

To understand the limitations of AI-based models in health care and the responsibilities of manufacturers and users of AI software as a medical device (SaMD), an MI-CLAIM checklist was proposed for use in AI software development [37]. Its purpose is to enable a direct assessment of clinical impact, including considerations of fairness and bias, and to allow rapid replication of the technical design by any legitimate clinical AI study. The MI-CLAIM checklist has six parts (Table 1.3), including (1) Study design; (2) Separation of data into partitions for model training and model testing; (3) Optimization and final model selection;

Table 1.2 Practical challenges to the advancement and application of AI tools in clinical settings

Workflow integration	Understand the technical, cognitive, social, and political factors in play and incentives impacting integration of AI into health care workflows.
Enhanced explainability and interpretability	To promote integration of AI into health care workflows, consider what needs to be explained and approaches for ensuring understanding by all members of the health care team.
Workforce education	Promote educational programs to inform clinicians about AI/machine learning approaches and to develop an adequate workforce.
Oversight and regulation	Consider the appropriate regulatory mechanism for AI/machine learning and approaches for evaluating algorithms and their impact.
Problem identification and prioritization	Catalog the different areas of health care and public health where AI/machine learning could make a difference, focusing on intervention-driven AI
Clinician and patient engagement	Understand the appropriate approaches for involving consumers and clinicians in AI/machine learning prioritization, development, and integration, and the potential impact of AI/machine learning algorithms on the patient-provider relationship.
Data quality and access	Promoting data quality, access, and sharing, as well as the use of both structured and unstructured data and the integration of non-clinical data is critical to developing effective AI tools.

Source Matheny ME, Thadaney Israni S, Ahmed M, Whicher D. AI in Health Care: The Hope, the Hype, the Promise, the Peril. Washington, DC: National Academy of Medicine; 2019. <https://nam.edu/artificial-intelligence-special-publication>

Table 1.3 The MI-CLAIM checklist [Source: Norgeot B, Quer G, Beaulieu-Jones BK, Torkamani A, Dias R, Gianfrancesco M, Arnaout R, Kohane IS, Saria S, Topol E, Obermeyer Z, Yu B, Butte AJ. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. Nat Med. 2020 Sep;26(9):1320–1324]

Before paper submission		
<i>Study design (Part 1)</i>	<i>Completed: page number</i>	<i>Notes if not completed</i>
The clinical problem in which the model will be employed is clearly detailed in the paper	☐	
The research question is clearly stated	☐	
The characteristics of the cohorts (training and test sets) are detailed in the text	☐	
The cohorts (training and test sets) are shown to be representative of real-world clinical settings	☐	
The state-of-the-art solution used as a baseline for comparison has been identified and detailed	☐	
<i>Data and optimization (Parts 2, 3)</i>	<i>Completed: page number</i>	<i>Notes if not completed</i>
The origin of the data is described and the original format is detailed in the paper	☐	
Transformations of the data before it is applied to the proposed model are described	☐	
The independence between training and test sets has been proven in the paper	☐	
Details on the models that were evaluated and the code developed to select the best model are provided	☐	
Is the input data type structured or unstructured?	☐	
<i>Model performance (Part 4)</i>	<i>Completed: page number</i>	<i>Notes if not completed</i>
The primary metric selected to evaluate algorithm performance (e.g., AUC, F-score, etc.), including the justification for selection, has been clearly stated	☐	
The primary metric selected to evaluate the clinical utility of the model (e.g., PPV, NNT, etc.), including the justification for selection, has been clearly stated	☐	
The performance comparison between baseline and proposed model is presented with the appropriate statistical significance	☐	
<i>Model examination (Part 5)</i>	<i>Completed: page number</i>	<i>Notes if not completed</i>
Examination technique 1 ^a	☐	
Examination technique 2 ^a	☐	
A discussion of the relevance of the examination results with respect to model/algorithm performance is presented	☐	
A discussion of the feasibility and significance of model interpretability at the case level if examination methods are uninterpretable is presented	☐	
A discussion of the reliability and robustness of the model as the underlying data distribution shifts is included	☐	
<i>Reproducibility (Part 6): choose appropriate tier of transparency</i>		<i>Notes</i>
Tier 1: complete sharing of the code	☐	
Tier 2: allow a third party to evaluate the code for accuracy/fairness; share the results of this evaluation	☐	
Tier 3: release of a virtual machine (binary) for running the code on new data without sharing its details	☐	

(continued)

Table 1.3 (continued)

Before paper submission		
Tier 4: no sharing	☐	

PPV positive predictive value, *NNT* numbers needed to treat

^aCommon examination approaches based on study type: for studies involving exclusively structured data, coefficients and sensitivity analysis are often appropriate; for studies involving unstructured data in the domains of image analysis or natural language processing, saliency maps (or equivalents) and sensitivity analyses are often appropriate

Table 1.4 Major topics of CONSORT-AI extension

1. State the inclusion and exclusion criteria at the level of participants
2. State the inclusion and exclusion criteria at the level of the input data
3. Describe how the AI intervention was integrated into the trial setting, including any onsite or offsite requirements
4. State which version of the AI algorithm was used
5. Describe how the input data were acquired and selected for the AI intervention
6. Describe how poor-quality or unavailable input data were assessed and handled
7. Specify whether there was human—AI interaction in the handling of the input data, and what level of expertise was required for users
8. Specify the output of the AI intervention
9. Explain how the AI intervention’s outputs contributed to decision-making or other elements of clinical practice
10. Describe results of any analysis of performance errors and how errors were identified, where available. If no such analysis was planned or done, explain why not.

Source Adapted from Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK; SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Lancet Digit Health*. 2020 Oct;2(10):e537–e548

(4) Performance evaluation; (5) Model examination; and (6) Reproducible pipeline. The CONSORT-AI and SPIRIT-AI working groups have proposed reporting guidelines for clinical trials of interventions involving AI. A summary of these guidelines is presented in Table 1.4.

Inherent conflicts of interest should be acknowledged. Manufacturers who develop and market SaMD have a strong financial interest in presenting their products positively. Thus, conflicts of interest exist if they fund, conduct, and publish results of studies, including those that might report deficiencies in their products. Many of the published papers in the field of AI-based diabetic retinopathy screening, particularly those using CE-marked and FDA-approved algorithms, were conducted by manufacturers or patent owners.

It should be also remembered that AI algorithms can be designed to perform in unethical ways. For example, Uber’s software Greyball allowed the company to identify and circumvent local regulations, and Volkswagen’s algorithm allowed vehicles to pass emission tests by reducing their emissions of nitrogen oxide during testing. AI algorithms could be tuned to generate

increased profits for their owners by recommending particular drugs, tests, or the like without clinical users’ awareness. AI systems are vulnerable to cybersecurity attacks that could cause their algorithms to misclassify medical information [31].

Seven essential factors to design AI for social good were proposed by Floridi et al. (Table 1.5) [38]. The authors propose falsifiability as an essential factor to improve the trustworthiness of technological application, i.e., for an SaMD to be trustworthy, its safety should be falsifiable. Critical requirements for a device to be fully functional must be specified and must be testable. If falsifiability is not possible, then the critical requirements cannot be checked, and the system should not be deemed trustworthy [38].

Cost-Effectiveness of AI-Based Devices

One of the arguments for AI-based medical devices is that they can reduce medical costs and eliminate unnecessary procedures. A study from Singapore found that a semiautomated model that

Table 1.5 Essential Factors to design AI for social good

Factors	Corresponding best practices	Corresponding ethical principle
Falsifiability and incremental deployment	Identify falsifiable requirements and test them in incremental steps from the lab to the “outside world”	Nonmaleficence
Safeguards against the manipulation of predictors	Adopt safeguards which (i) ensure that non-causal indicators do not inappropriately skew interventions, and (ii) limit, when appropriate, knowledge of how inputs affect outputs from AI4SG systems, to prevent manipulation	Nonmaleficence
Receiver-contextualised intervention	Build decision-making systems in consultation with users interacting with and impacted by these systems; with understanding of users’ characteristics, the methods of coordination, the purposes and effects of an intervention; and with respect for users’ right to ignore or modify interventions	Autonomy
Receiver-contextualised explanation and transparent purposes	Choose a Level of Abstraction for AI explanation that fulfils the desired explanatory purpose and is appropriate to the system and the receivers; then deploy arguments that are rationally and suitably persuasive for the receiver to deliver the explanation; and ensure that the goal (the system’s purpose) for which an AI4SG system is developed and deployed is knowable to receivers of its outputs by default	Explicability
Privacy protection and data subject consent	Respect the threshold of consent established for the processing of datasets of personal data	Nonmaleficence; autonomy
Situational fairness	Remove from relevant datasets variables and proxies that are irrelevant to an outcome, except when their inclusion supports inclusivity, safety, or other ethical imperatives	Justice
Human-friendly semanticisation	Do not hinder the ability for people to semanticise (that is, to give meaning to, and make sense of) something	Autonomy

Source Floridi L, Cowls J, King TC, Taddeo M. How to Design AI for Social Good: Seven Essential Factors. *Sci Eng Ethics*. 2020 Jun;26(3):1771–1796. Springer

combined a DL system with human assessment achieved the best economic returns, leading to savings of 19.5% in screening for diabetic retinopathy. An earlier study from the UK reported cost-savings of 12.8–21.0%; however, a simple comparison between them is not possible due to the different models of DR screening in the two countries (two-stage screening in Singapore and three-stage screening in the UK), and their

different DR classification systems. The authors of both studies argued that a semiautomated system produces more savings than a fully automated system due to the lower rate of false positives and unnecessary specialist visits [39, 40].

This book aims to provide ophthalmologists and other visual professionals and researchers with an overview of current research into the use of AI in ophthalmology. Together with a team

of international experts from Europe, North America, and Asia, we present an overview of the most important documentary research in ophthalmology on ML and AI technologies and their benefits. We discussed the use of AI in the diagnosis of some retinal and corneal disorders, the diagnosis of congenital cataract, neuro-ophthalmology, glaucoma, intraocular lens calculation methods, ocular oncology, ophthalmology triaging, cataract-surgery training, refractive surgery, and the assessment and prediction of systemic diseases through the use of the eye. Chapters on digital-image analysis, AI basics, and technical aspects of AI provide the reader with knowledge not commonly possessed by ophthalmologists, but required to understand the topic in both its field-specific and broader contexts. The very important chapter on AI safety and efficacy outlines the challenges ophthalmology will face with the introduction and widespread dissemination of this technology. Although we have covered all of the major areas of AI/ML technology in ophthalmology, research in this field is progressing so quickly that some new concepts that emerged at the end of 2020 and in early 2021 do not appear on these pages. However, evidence-based medicine often demands that we await for more evidence to verify early reports and assess the real value of new medical technologies or applications. I would like to thank all the contributors for sharing their knowledge in this new and fascinating discipline, which has great potential to change ophthalmology.

Acknowledgements I would like to thank Aleksandra Lemanik, Foundation for Ophthalmology Development, Poznan, Poland and Tomasz Krzywicki, Faculty of Mathematics and Computer Science, University of Warmia and Mazury, Olsztyn, Poland for their help in preparing illustrations, and Szymon Wilk, Faculty of Computing and Telecommunications, Poznan University of Technology, Poznan, Poland, for his valuable discussion on this chapter.

References

1. Mitchell M. Artificial intelligence: a guide for thinking humans. Penguin UK; 2019.
2. Topol E. Deep medicine: how artificial intelligence can make healthcare human again. New York: Basic Books; 2019.
3. Copeland BJ. Artificial intelligence. *Encyclopedia Britannica*, 11 August 2020. <https://www.britannica.com/technology/artificial-intelligence>. Accessed 18 Mar 2021.
4. Singh H, Meyer AN, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. *BMJ Qual Saf*. 2014;23(9):727–31.
5. Gunderson CG, Bilan VP, Holleck JL, et al. Prevalence of harmful diagnostic errors in hospitalised adults: a systematic review and meta-analysis. *BMJ Qual Saf*. 2020;29:1008–18.
6. Zwaan L, Singh H. Diagnostic error in hospitals: finding forests not just the big trees. *BMJ Qual Saf*. 2020;29(12):961–4.
7. Singletary B, Patel N, Heslin M. Patient perceptions about their physician in 2 words: the good, the bad, and the ugly. *JAMA Surg*. 2017;152(12):1169–70.
8. Benjamins S, Dhunoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *NPJ Digit Med*. 2020;3:118.
9. Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, Peng L, Webster DR. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng*. 2018;2(3):158–64.
10. Chen J, See KC. Artificial intelligence for COVID-19: rapid review. *J Med Internet Res*. 2020;22(10):e21476.
11. Lee AY, Yanagihara RT, Lee CS, Blazes M, Jung HC, Chee YE, Gencarella MD, Gee H, Maa AY, Cockerham GC, Lynch M, Boyko EJ. Multicenter, head-to-head, real-world validation study of seven automated artificial intelligence diabetic retinopathy screening systems. *Diabetes Care*. 2021;dc201877. <https://doi.org/10.2337/dc20-1877>.
12. Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis. *Lancet Digit Health*. 2021;3(3):e195–203.
13. Hwang TJ, Kesselheim AS, Vokinger KN. Lifecycle regulation of artificial intelligence- and machine learning-based software devices in medicine. *JAMA*. 2019;322(23):2285–6.

14. Hwang TJ, Sokolov E, Franklin JM, Kesselheim AS. Comparison of rates of safety issues and reporting of trial outcomes for medical devices approved in the European Union and United States: cohort study. *BMJ*. 2016;353: i3323.
15. European Commission. Medical devices—EUDAMED. 17 June 2020. https://ec.europa.eu/growth/sectors/medical-devices/new-regulations/eudamed_en. Accessed 15 Jan 2021.
16. Ting DSW, Liu Y, Burlina P, Xu X, Bressler NM, Wong TY. AI for medical imaging goes deep. *Nat Med*. 2018;24(5):539–40.
17. Wang SY, Pershing S, Lee AY, AAO Taskforce on AI and AAO Medical Information Technology Committee. Big data requirements for artificial intelligence. *Curr Opin Ophthalmol*. 2020;31(5):318–23.
18. Kermanny DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, McKeown A, Yang G, Wu X, Yan F, Dong J, Prasadha MK, Pei J, Ting MYL, Zhu J, Li C, Hewett S, Dong J, Ziyar I, Shi A, Zhang R, Zheng L, Hou R, Shi W, Fu X, Duan Y, Huu VAN, Wen C, Zhang ED, Zhang CL, Li O, Wang X, Singer MA, Sun X, Xu J, Tafreshi A, Lewis MA, Xia H, Zhang K. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*. 2018;172(5):1122–1131.e9.
19. Rampasek L, Goldenberg A. Learning from everyday images enables expert-like diagnosis of retinal diseases. *Cell*. 2018;172(5):893–5.
20. Burlina P, Paul W, Mathew P, Joshi N, Pacheco KD, Bressler NM. Low-shot deep learning of diabetic retinopathy with potential applications to address artificial intelligence bias in retinal diagnostics and rare ophthalmic diseases. *JAMA Ophthalmol*. 2020;138(10):1070–7.
21. Burlina PM, Joshi N, Pacheco KD, Liu TYA, Bressler NM. Assessment of deep generative models for high-resolution synthetic retinal image generation of age-related macular degeneration. *JAMA Ophthalmol*. 2019;137:258–64.
22. Liu Y, Yang J, Zhou Y, Wang W, Zhao J, Yu W, Zhang D, Ding D, Li X, Chen Y. Prediction of OCT images of short-term response to anti-VEGF treatment for neovascular age-related macular degeneration using generative adversarial network. *Br J Ophthalmol*. 2020;104(12):1735–40.
23. Zheng C, Xie X, Zhou K, Chen B, Chen J, Ye H, Li W, Qiao T, Gao S, Yang J, Liu J. Assessment of generative adversarial networks model for synthetic optical coherence tomography images of retinal disorders. *Transl Vis Sci Technol*. 2020;9(2):29.
24. Liu TYA, Farsiu S, Ting DS. Generative adversarial networks to predict treatment response for neovascular age-related macular degeneration: interesting, but is it useful? *Br J Ophthalmol*. 2020;104(12):1629–30.
25. Lee CS, Lee AY. Clinical applications of continual learning machine learning. *Lancet Digit Health*. 2020;2(6):e279–81.
26. Mehta N, Lee CS, Mendonça LSM, Raza K, Braun PX, Duker JS, Waheed NK, Lee AY. Model-to-data approach for deep learning in optical coherence tomography intraretinal fluid segmentation. *JAMA Ophthalmol*. 2020;138(10):1017–24.
27. Larson DB, Harvey H, Rubin DL, Irani N, Tse JR, Langlotz CP. Regulatory frameworks for development and evaluation of artificial intelligence-based diagnostic imaging algorithms: summary and recommendations. *J Am Coll Radiol*. 2021;18(3 Pt A):413–24.
28. Wang X, Liang G, Zhang Y, Blanton H, Bessinger Z, Jacobs N. Inconsistent performance of deep learning models on mammogram classification. *J Am Coll Radiol*. 2020;17:796–803.
29. Subbaswamy A, Schulam P, Saria S. Preventing failures due to dataset shift: learning predictive models that transport. *Proc Mach Learn Res*. 2019;89:3118–27.
30. Subbaswamy A, Saria S. From development to deployment: dataset shift, causality, and shift-stable models in health AI. *Biostatistics*. 2020;21:345–52.
31. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med*. 2019;17:195.
32. Winkler JK, Fink C, Toberer F. Association between surgical skin markings in dermoscopic images and diagnostic performance of a deep learning convolutional neural network for melanoma recognition. *JAMA Dermatol*. 2019;155:1135–41.
33. Finlayson SG, Bowers JD, Ito J. Adversarial attacks on medical machine learning. *Science*. 2019;363:1287–9.
34. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. *PLoS Med*. 2018;15: e1002683.
35. Antun V, Renna F, Poon C, Adcock B, Hansen AC. On instabilities of deep learning in image reconstruction and the potential costs of AI. *Proc Natl Acad Sci U S A*. 2020; pii:201907377. <https://doi.org/10.1073/pnas.1907377117>.
36. Matheny ME, Thadaneey Israni S, Ahmed M, Whicher D. AI in health care: the hope, the hype, the promise, the peril. Washington, DC: National Academy of Medicine; 2019. <https://nam.edu/artificial-intelligence-special-publication>
37. Norgeot B, Quer G, Beaulieu-Jones BK, Torkamani A, Dias R, Gianfrancesco M, Arnaut R, Kohane IS, Saria S, Topol E, Obermeyer Z, Yu B, Butte AJ. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. *Nat Med*. 2020;26(9):1320–4.
38. Floridi L, Cowls J, King TC, Taddeo M. How to design AI for social good: seven essential factors. *Sci Eng Ethics*. 2020;26(3):1771–96.

39. Xie Y, Nguyen QD, Hamzah H, Lim G, Bellemo V, Gunasekaran DV, Yip MYT, Qi Lee X, Hsu W, Li Lee M, Tan CS, Tym Wong H, Lamoureux EL, Tan GSW, Wong TY, Finkelstein EA, Ting DSW. Artificial intelligence for teleophthalmology-based diabetic retinopathy screening in a national programme: an economic analysis modelling study. *Lancet Digit Health*. 2020;2(5):e240–9.
40. Tufail A, Rudisill C, Egan C, Kapetanakis VV, Salas-Vega S, Owen CG, Lee A, Louw V, Anderson J, Liew G, Bolter L, Srinivas S, Nittala M, Sadda S, Taylor P, Rudnicka AR. Automated diabetic retinopathy image assessment software: diagnostic accuracy and cost-effectiveness compared with human graders. *Ophthalmology*. 2017;124(3):343–51.

Basics of Artificial Intelligence for Ophthalmologists

2

Ikram Issarti and Jos J. Rozema

Abstract

The applications of Artificial Intelligence are slowly beginning to take over tasks into the field of ophthalmology that until recently required a trained clinician to perform. Given their exceedingly level of high performance, many are inclined to trust the results provided by such programs, without much insight into how these results were accomplished. The exponential advancement of AI, especially in topics such as Natural Language Processing (NLP) and Generative AI expanded the use cases and the capabilities of AI, triggering boarder concerns and capabilities of the impact and usability of such technologies in the area of Ophtalmology. This chapter comprehensively elucidates the basic principles and latest updates of Artificial Intelligence to help ophthalmologists get a feel for the strengths and limits of the techniques, as well as how they may use them in clinical practice.

Keywords

Artificial intelligence · Machine learning · Deep learning · Ophthalmology · Introduction · Basic principles · Generative AI · Large language models

Introduction

The past decade has seen a steep rise in the number of applications of Artificial Intelligence (AI), especially for repetitive or complex tasks where humans may quickly suffer from either a drifting attention span or subtle inconsistencies. Such systems are often more cost efficient, thus accelerating their adoption and acceptance, consequently increasing people's reliance on AI. But an understanding of its inner workings is often lacking, many tend to approach it as a 'black box' at the risk of uncritically accepting whatever output it produces. Although by its very nature, AI is opaque about how it reaches a result, there are statistical methods to objectively assess the quality of its output. As AI becomes a popular subject within the scientific community and health-care practice, this chapter explains the basic principles of AI in a comprehensive step-by-step manner, along with examples of ophthalmological applications. Special

I. Issarti · J. J. Rozema (✉)
Visual Optics Lab Antwerp (VOLANTIS), Faculty
of Medicine and Health Sciences, University of
Antwerp, Wilrijk, Belgium
e-mail: jos.rozema@uantwerpen.be

Department of Ophthalmology, Antwerp University
Hospital, Edegem, Belgium

attention will be paid to the differences between AI, Machine Learning (ML), and Deep Learning (DL), highly interconnected techniques that are often confused for one another. This work's scope will extend to cover recent trends in AI including Generative AI and Large Language Models, together with certain recent applications in Ophthalmology.

Short History

The concept of artificial intelligence dates back to ancient times, with legends describing artificial beings created by master craftsmen or brought into life by magic. For many centuries, and without the required scientific and technological basis, the aspiration of creating intelligent beings remained fictitious and only present in art, myths, and novels.

It was not until the eighteenth century that the mathematical and statistical foundations of AI were established, such as Bayes' Theorem, which would later become the groundwork for the current AI developments. Simultaneously, the first executable algorithms were developed by Ada Lovelace, who argued that computers could be used to execute instructions beyond basic calculations.

In the nineteenth century, Neuroscience has inspired the first establishment of the theoretical foundations of AI. Warren McCulloch and Walter Pitts introduced a computational Artificial Neuron [1], with the intent of explaining how the biological neuron wires. The computational Neuron functioning was described by simple logical operators (AND, OR, and NOT). A few years later, Donald Hebb published his famous paper "The Organization of Behavior" [2], in which he suggested that the biological neuronal pathways are reinforced with repeated use. Hebb's work inspired the earliest AI algorithms, Artificial Neural Networks, which are assumed to mimic certain intelligent characteristics, such as learning, adaptability, and generalization.

From the 1950s onward, specific AI concepts have developed, yet the term Artificial Intelligence isn't yet known to the scientific

public. At the time, philosophers, and scientists were exploring the idea of achieving machines that could 'Think' and perform tasks restricted to human intelligence. This was the scope of the summer workshop organized by John MacCarthy at Dartmouth College [3]. In the same vein, the famous English mathematician Alan Turing published his paper '*Computing Machinery and Intelligence*' [4], in which he argued that the focus may need to be shifted toward developing machines that could perform human-like tasks instead of focusing on achieving '*Thinking*' machines. Turing also developed his famous Turing test, which is an open-ended interview with a human evaluator that must determine whether the respondent is human or machine. Although AI has yet to pass the test even today [5], it helped create the concept of 'Artificial Intelligence' as a result of MacCarthy's conference.

Between 1956 and 1974, advancements in computing power enabled the testing of various AI concepts. Among them, the 'Logic Theorist' [6], represented the initial AI program that emulated human problem-solving abilities. Followed by Samuel's self-learning checkers-playing program. Thus, introducing the term 'Machine Learning' [7]. These critical milestones, however, marked a benchmark in the evolution of AI and machines, as learning algorithms emerged. Subsequently, a successful first development of the Natural Language Processing program ELIZA [8].

In spite of the initial successes, the 1970s witnessed the first AI Winter, due to limited funding, interest, and unmet expectations. It was concluded that if strong AI was not achievable, then other sciences could solve typical AI problems. Yet, notable AI breakthroughs continued. The first Optical Character Recognition (OCR) and text-to-speech synthesizer were developed, allowing computers to read text out loud for blind patients [9]. The following years saw a resurgence in AI activities, particularly 'Expert Systems', which were designed to capture human knowledge in specific domains and mimic the human decision-making process. Meanwhile, Artificial Neural Networks

(a subfield of AI), and backpropagation algorithms, witnessed significant progress which is considered the fundamental basis of Machine and Deep Learning today. This progress was joined with the development of parallel distributed processing, robotics, AI-medical diagnosis applications, etc. This continued up to the 90s, when AI progress recessed again in the second AI winter due to limitations in the computing power required for implementing robust expert systems. Even so, during this time the first ‘Intelligent Agents’ emerged, a broad term for an AI that can perceive its environment and respond accordingly, such as IBM’s Deep Blue that defeated chess world champion Garry Kasparov or LeNet-5, a Convolutional Neural Network for digit recognition. Industrialized and marketed AI progress appeared from the early 2000s onwards due to a significant increase in the availability of datasets and computing power, leading to large breakthroughs in Machine Learning algorithms and applications. These include for instance the iRobot Roomba autonomous vacuum cleaner released in 2002 and the self-driving vehicles introduced by Google, Tesla and Ford.

In 2006, Deep Belief Networks were introduced, sparking more interest in the potential of Deep Learning. Various successful Deep Learning applications were developed, such as AlexNet, a Convolutional Neural Network that successfully competed in the ImageNet competition for image recognition. Similar breakthroughs included Google’s Word2Vec, a Shallow Neural Network that represents words, as well as DeepMind’s AlphaGo, which defeated the Go World Champion after only 30 hours of unsupervised learning.

In the following years the Machine Learning and Deep Learning algorithms expanded in scope and capabilities as dozens of new algorithms were developed and commercialized while the scientific community explored other types of AI, such as the Generative Adversarial Networks (GANs, 2014) [10] and dozens of subsequent variants. In the famous 2017 paper ‘*Attention is All you Need*’ by Vaswani et al. [11], GANs were combined with a Transformer

architecture, the key components behind Large Language Models like ChatGPT and BERT, or image generators like DALL-E and CLIP. Such algorithms have revolutionized the AI landscape by providing proven versatility without the need for fine-tuning. The development of Large Language Models is still ongoing, however, as OpenAI recently released GPT-4o, that will bridge into GPT-5 later in 2024.

Overview

The following gives an overview of key terminologies and algorithms employed in AI.

Artificial Intelligence is a very broad field of study encompassing a wide range of techniques that allow machines to display ever more intelligent behaviour (Fig. 2.1). **Machine learning** is one of the most important subfields of AI. Although ML and AI are often confused, AI also includes other approaches not included in machine learning, such as Expert Systems, knowledge - or rule-based systems that emulate human cognitive and reasoning abilities by following certain guidelines to perform a decision-making process [12]. Meanwhile, ML refers to a group of mathematical algorithms that learn from experience (data) by mimicking human learning behaviour to perform new tasks. ML is able to fit complex data sets, to extract new knowledge, imitate complex behaviour, predict and classify based on prior data. Another well-known group of algorithms is **Deep learning** (DL), which is a subset of machine learning based on artificial neural networks. DL is able to simultaneously analyse multiple layers of data. These layers consist of data processing units, called neurons, that allow them to analyse large amounts of data at once while preserving the data’s spatial distribution. DL systems have seen significant successes in applications such as pattern recognition, image processing, and speech recognition. The application of DL in linguistics and semantics is referred to as **Natural Language Processing** (NLP), a subset of DL that focuses on the interaction between computers and humans through natural language by interpreting

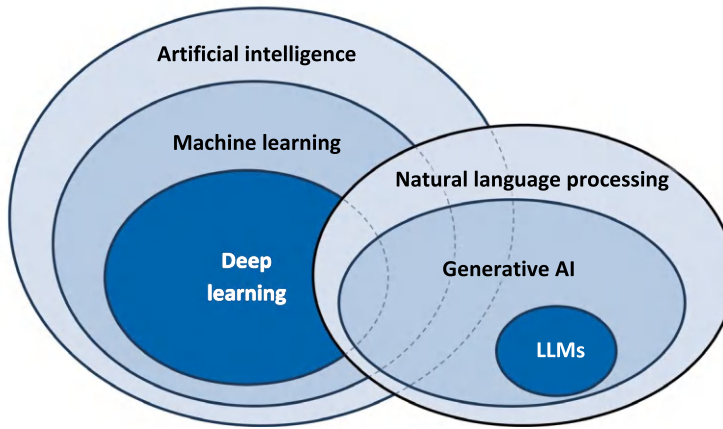


Fig. 2.1 Artificial intelligence techniques

and manipulating sequential semantics. It has been around since the 1960s, but due to recent advances in ML and DL, NLP has become the foundation for the disruptive digital applications being developed, ranging from automating clinical documentation to enhancing patient experience with chatbots. Building on the capabilities of NLP, **Generative AI** is able to learn to model the underlying distribution of a given dataset and subsequently generate new content, such as text, images, and music. Applications, may include the generation of new survey questions or synthetic medical data for research.

While the above-mentioned learning types (ML, DL, NLPs, and Generative AI) focus on specific tasks, they are still categorized as narrow AIs as they are solely for single-purpose applications. Strong AI or **Artificial General Intelligence (AGI)**, on the other hand, aspires to achieve human levels of intelligence, with intelligence features such as understanding, reasoning, and learning across various domains without re-training. Although strong AI is currently still a theoretical concept, the ongoing effort in developing large multimodal (combined) AI shows signs of approaching AGI as for instance **Large Language Models (LLMs)** have proven themselves to be a powerful general purpose platform capable of performing a wide range of tasks without additional training. These LLMs (e.g., GPT-4, BERT, etc.) are NLPs with an underlying Transformer architecture,

trained on trillions of data by parallel supercomputers to interpret human language, generate text, and engage in human-like conversations. Unlike standard, data-driven AI (e.g., ML, DL), LLMs are trained on vast amounts of data and billions of parameters, which enables them to be general-purpose and perform a variety of tasks without fine-tuning. One example is ChatGPT, a Generative AI chatbot that generates human-like responses in a conversation through prompts and without the need for re-training or expert coding. But an LLM can also be fine-tuned on specific data to make it more relevant and avoid incorrect or inappropriate responses.

The **training process** of most AI algorithms (e.g., ML, DL, etc.) is very similar to that found in schools, with a professor teaching his students. From a large amount of given data, the algorithm learns how to describe a specific topic into a model (knowledge acquisition), which will subsequently be validated using unseen data to evaluate its generalizability. Finally, the performance of the algorithm is evaluated based on several guidelines given in the “Performance Evaluation” section.

Data Basis

Data is the fuel of AI, which can come from different sources such as, webs, videos, audios, text, etc. It is comprised of massive amounts of

bits, binary values of zeros and ones, that can be reorganized to form structured data that is usually easier to process by AI algorithms, such as a relational database or a spreadsheet. It is also possible to work with unstructured data without predefined formatting (e.g. audio, video, text, etc.), or with a hybrid form of a structured and unstructured data called semi-structured data. Finally, one can consider time series data, consisting of structured or unstructured data in sequential time steps [13]. A good understanding of data structures allows a proper AI implementation. Some highlights are given in the section “Conducting a Machine Learning Analysis”, but more details are available in the data mining literature and data pre-processing text books [14].

Common Tasks

In medicine, Machine Learning is mostly used to assist physicians with diagnosis, monitoring, and decision making by providing insight into the structure and patterns within large datasets. The most typical tasks for ML are classification, clustering and prediction.

- **Classification** involves sorting new cases into two or more groups (Fig. 2.2a). In healthcare, classification could be used for diagnosis (healthy or abnormal) or the identification of biological markers.
- **Clustering** the algorithm divides a dataset into several, previously unknown clusters

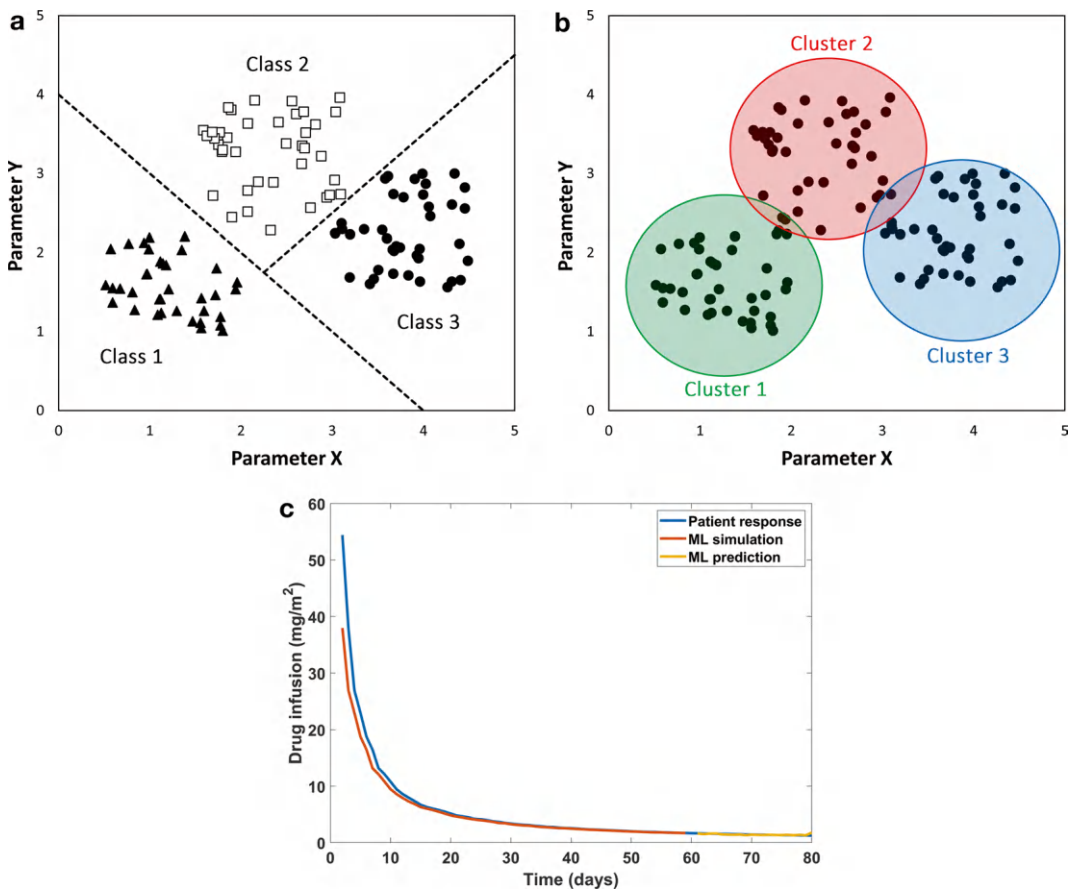


Fig. 2.2 Examples of (a) classification, (b) clustering and (c) prediction

(groups) with certain properties in common (Fig. 2.2b). Clustering can be used to e.g. distinguish the different stages of a disease.

- **Prediction** consists of building a model based on historical data to forecast unknown parameter values in the future e.g. to predict the outcome of a surgical procedure or treatment (Fig. 2.2c).
- **Regression** while classification problems classify data into different sets of classes or categories, regression problems predict the values of a continuous variable rather than categorical real variables. This problem is also referred to as prediction task.
- **Content Generation** involves modelling the probabilistic relationships and patterns in a given dataset, out of which the algorithm can, e.g., predict what should come next in a sentence.

Learning Models

AI algorithms can be trained in any one of four methods:

- **Supervised Learning** ('with professor') teaches a ML algorithm the desired output (answer) given an input with labelled categories. Based on this, the algorithm learns the characteristics of each category, so when it is presented with an unseen input, it will be able to assign it to the right output class (category). Supervised algorithms are mostly used for classification problems (Fig. 2.2a) where points can be assigned to three pre-defined classes (e.g. healthy, pathological, and suspect) or for prediction problems (Fig. 2.2c), such as predicting the future evolution of a tumour.
- **Unsupervised Learning** ('without professor') algorithms assign data to multiple subgroups (clusters) with similar properties within the input data without being given the desired answers or outputs. Unsupervised learning can be applied for classification problems with unknown outputs, as illustrated in Fig. 2.2b, where the algorithm

identified three clusters based on the available data.

- **Semi-supervised learning** combines supervised and unsupervised learning by giving the desired output for only a small number of inputs. After training based on the labelled data, the algorithm uses unsupervised learning for the unlabelled data to create new clusters. Ultimately these clusters are themselves labelled and added to the previous outputs. This method is used when not all outputs are available.
- **Reinforcement learning** is a training method in which an algorithm must define its own response based on trial and error, much as in human learning. This can be applied when there is a continuous change in the situation to which the machine needs to adapt and respond to. Although quite advanced, its use remains limited within the field of medicine to e.g. systems that learn from the successes and failures of clinical trials in the literature to suggest new approaches for testing.
- Certain AIs use **hybrid forms of learning**, combining the previous learning types. For example, generative AI and LLMs primarily rely on unsupervised learning techniques, but also on supervised and reinforcement learning.

Machine Learning Algorithms

There are dozens of machine learning algorithms described in the literature. For reasons of conciseness, only the most common ones will be listed below.

(Non)-linear Regression

Regression analysis is a well-known statistical method that builds a mathematical **model** from prior observations to make a prediction, which constitutes the basis of machine learning. If the relation between the input and output is linear, the model is called **linear regression** (Fig. 2.3). For example, one can score the progress of a

disease based on several observed variables ($x_1, x_2, \dots, x_n$) by assigning a weight ($w_1, w_2, \dots, w_n$) to each variable indicating their relative importance. The overall score is then defined as a function of the weighted variables as follows: with weights that are estimated through an over-

$$\text{Score} = w_1.x_1 + w_2.x_2 + \dots + w_n.x_n$$

all fit of the observed dataset. Similarly, for more complex relationships non-linear regression may be used, using higher orders of the variables ($x_1, x_2, \dots, x_n$). A second order non-linear regression for n parameters would look like:

$$\begin{aligned} \text{Score} = & w_1.x_1 + \dots + w_n.x_n + w_{11}.x_1^2 \\ & + w_{12}.x_1.x_2 + \dots + w_{nn}.x_n^2 \end{aligned}$$

Logistic Regression

This is an easy to implement but powerful classification algorithm that gives binary outputs (e.g. diseased or healthy) [13, 15].

Naïve Bayes Algorithm

Naïve Bayes is one of the simplest supervised classification algorithms available, based only on Bayes' theorem. This theorem relates the probability $P(c|x)$ that a certain event c occurs

under circumstances x with the probability $P(x|c)$ that circumstances x are present once event c has already happened and the ratio of the probabilities $P(c)$ and $P(x)$ of event c and circumstances x individually. Formally, this is written as:

$$P(c|x) = P(x|c).P(c)/P(x)$$

Although this may seem complicated, think for example of the situation where a patient must be classified as normal (NL) or keratoconus (KC) using only minimum pachymetry $P_{\min}$ given as 'Thin' where $P_{\min} < 500 \mu\text{m}$ and 'Thick' with $P_{\min} \geq 500 \mu\text{m}$. Suppose a new patient appears with $P_{\min} < 500 \mu\text{m}$, then the probability that this patient really has keratoconus is given by:

$$P(KC|Thin) = P(Thin|KC) \cdot \frac{P(KC)}{P(KC).P(Thin|KC) + P(NL).P(Thin|NL)}$$

where all the terms on the right-hand side can easily be estimated on forehand from a large data set after weighting $P(KC|Thin, W)$ with a weight W . In practice the classification is based on many variables, however, increasing the chance of interdependence between parameters. Naïve Bayes chooses to ignore this interdependence, which, despite being a severe oversimplification, demonstrated good results in classification, especially in text recognition, spam detection, and medical diagnosis.

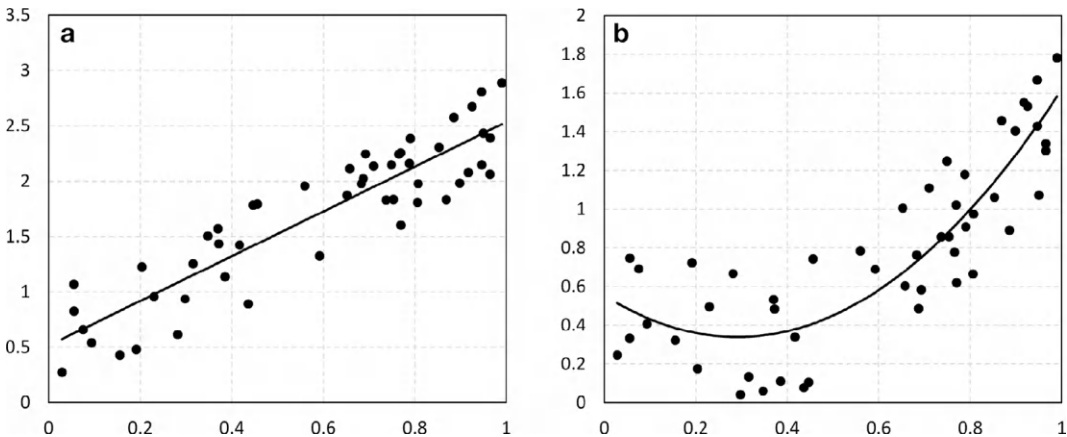


Fig. 2.3 Examples of (a) linear and (b) non-linear fitting

Support Vector Machine (SVM)

SVM is a recently developed form of supervised machine learning used for both regression and classification. These algorithms are preferred by experts for their accurate and reproducible results with less computation power, while being robust enough to handle small samples. In short, SVM is a binary classifier that searches for a dividing plane to separate distinct data. The orientation and the position of this dividing plane are determined by the closest points, called support vectors, in an attempt to maximize the margins between distinct groups [13, 15].

K-Means

As a form of unsupervised clustering, this method groups unlabelled input data into a predefined number of K clusters. As such, it can operate on a large dataset. For example, if a dataset of normal and diseased subjects, but without clear classification of the disease's features, then K-means can be used to define K distinct clusters with statistical characteristics that may be used clinically. It does this by first randomly choosing centroid points (Fig. 2.4) and seeing whether the neighbouring points can efficiently be divided into the requested number of clusters by assessing their distance to the

centroids. If this is not the case, the centroids points are iteratively shifted until a certain minimum distance is achieved.

K-Nearest Neighbor (KNN)

KNN is a supervised machine learning method to address regression and classification problems. It is considered a 'lazy' algorithm as it does not require training, relying instead on the assumption that similar inputs remain close to one another. The algorithm computes a distance between a test data set and its K nearest neighbours to form a large cluster. Whenever a new data point is presented for classification, KNN will look in the database at the K points nearest to the new point to determine to what group this point should belong.

Decision Trees

A **decision tree** is a directed data structure that uses a series of yes/no questions for classification. The start of the decision tree is the root node with a yes/no input question. From this starts a series of decision paths (branch) where the algorithm makes a decision through a computed probability that ultimately leads to the leaf of the tree, corresponding with the outcome [15].

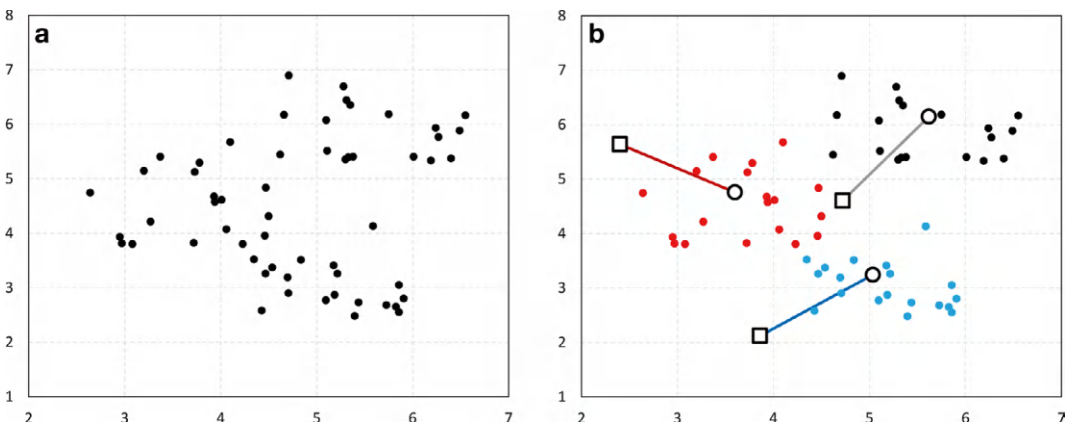


Fig. 2.4 (a) Input data; (b) K-means clustering, starting randomly placed centroids (open squares) that are gradually adjusted until an equilibrium has been reached (open circles)

Decision trees can also be combined into a structure, called a **Random Forest**, as a way to average multiple Decision Trees built from different parts of the training set to reduce the risk of overfitting often experienced by a single Decision Trees. The increased performance comes at the cost of some loss of interpretable and accurate fit to training data, however.

Artificial Neural Networks (ANN)

Artificial neural networks are a family of algorithms that, inspired by the human brain, form interconnected structures of **artificial neurons**. These structures interact with each other to mimic the complex behaviour found in real neurons, such as self-adapting, self-organizing, and real-time learning from examples. The human brain consists of many neurons that are interconnected through a large number of axons to exchange signals. When a neuron receives a specific input signal through its connections, its cellular body will generate a new signal through the axons and transmit it to other dendritic cells (Fig. 2.5) [16]. This biological architecture is emulated in artificial neurons, where dendritic

signals represent the neural inputs $X=(x_j)_j$ that are assigned synaptic weights θ_{ij} . The cellular body is represented by a nonlinear **activation function** that operates on the input signal to create an output signal y that is passed on to the next layer of neurons. The strength of ANNs lies in their several interconnected artificial neurons, exchanging signals in a forward and backward direction. Dozens of ANNs algorithms have been described in the literature, some supervised, such as multilayer perceptron (MLP) and feedforward neural network (FFN), and some unsupervised, such as self-organization maps (SOM). Artificial neural networks can be used for regression, classification, clustering problems and have been applied to highly complicated tasks like forecasting and system control.

Deep Learning (DL)

Similarly to machine learning there are dozens of deep learning algorithms, some of which are listed below.

Deep Neural Networks (DNN)

These are neural networks with many layers and a large number of neurons. Although this design comes at a high computational cost, DNN show very accurate results in medical imaging, vision recognition, etc. Some classic examples of deep learning algorithms are the Conventional Neural Network, Recurrent Neural Network, Self-Organization Maps, etc. Deep neural network is an equivalent term for deep learning.

Conventional Neural Network (CNN)

This is an advanced neural network algorithm able to simultaneously analyse multiple input layers while preserving their spatial connection. In image analysis, for example, neural networks, would translate each pixel to a vector, thus losing the spatial connection and correlations between them [12]. Meanwhile, the multilayer architecture of CNN, overcomes this limitation and preserves the relationship between the pixels.

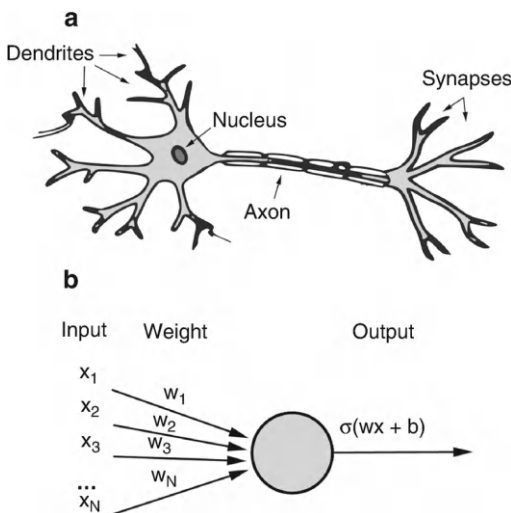


Fig. 2.5 Comparison of (a) biological and (b) artificial neurons

Recurrent Neural Network (RNN)

RNNs are neural network models designed for processing sequential data, such as a time series or a sequence of data with repetitive properties (e.g. DNA, which has long sequences of base pairs).

Self-organization Maps (SOM)

These unsupervised neural network algorithms are inspired by the brain's visual cortex. Their neurons are organised in close spatial proximity to one another to process the input with the shortest possible transmission of signals during processing. SOM is mostly used for pattern recognition, to extract features, and to compress highly dimensional data [17, 18].

Reinforcement Learning (RL)

This method is an idealized, goal-directed computational approach that closely imitates human leaning behaviour, consisting of interacting with an environment and assessing its response as a source of knowledge. Reinforcement learning algorithms are not told what to do, but instead figure out their best course of action through trials, mapping situations and maximizing a reward function [19]. In theory, this method could be applied to e.g. minimize the postoperative complications of cataract surgery by choosing the best surgical strategy that minimizes postoperative refractive errors.

Generative AI Algorithms

Variational Autoencoder (VAEs)

A Variational Autoencoder is a stochastic algorithm that learns the distribution of a given dataset and subsequently output a latent representation of the data within a reduced dimensional space. This algorithm uses unsupervised learning consisting of an encoder and a decoder that work together to determine the underlying relationship of the input data and compress it into a latent space to capture its essential features [20]. These models are used for anomaly detection, image generation, and dimensionality reduction.

Generative Adversarial Network (GAN)

Generative Adversarial Networks were introduced in 2014 by Goodfellow et al. [10] and contain two competing neural networks: one called the generator to create new data, and the second called discriminator to identify whether the generated data is fake or real. The competition between both neural networks allows it to generate high-quality synthetic data, which has allowed it to currently become one of the most used methods for this purpose.

Autoregressive Generative AI (AR)

Autoregressive AI relies on the principles of autoregression to predict the next component in a sequence by taking measurements from previous inputs in a sequence. The ability to capture sequential dependencies makes this algorithm efficient for text generation. For instance, OpenAI's ChatGPT uses an autoregressive language Transformer as part of its building blocks.

Large Language Models (LLM)

LLM are typically implemented as web-based platforms that allow for interaction through text- or speech-based prompting, making it ideal for non-experts. They can be used to interpret and generate texts, make translations, write various types of creative content, and answer questions. The most widely used LLMs are:

- **Generative Pre-trained transformer (GPT-4)** was launched in March 2023 by OpenAI, this fourth generation GPT Foundation model was trained on 1.8 trillion parameters. This Transformer can be accessed with an API for development purposes, and it has been adopted by Microsoft, ChatGPT, Dropbox, etc.
- **Bidirectional Encoder Representations from Transformers (BERT)** is an NLP developed by Google to understand natural language based on a Deep Neural Network

Table 2.1 Performance metrics

True positive	TP	Abnormal cases identified correctly
True negative	TN	Normal cases identified correctly
False positive	FP	Normal cases classified as abnormal (Type I error)
False negative	FN	Abnormal cases classified as normal (Type II error)
Accuracy	$(TP + TN)/(TP + FP + FN + TN)$	Percentage of times the algorithm is correct
Sensitivity	$TP/(TP + FN)$	Percentage of true positives correctly identified
Specificity	$TN/(TN + FP)$	Percentage of true negatives correctly identified
Precision	$TP/(TP + FP)$	Ratio of correctly classified positives; high precision relates to low false positive
Cut-off		Value or point designed as a limit of a group

- Transformer architecture. The Tranformer engine was trained on 3.3 billion words of text.
- **Gemini:** Gemini is a LLM similar to GPT-4 introduced in December 2023 by Google DeepMind, capable of responding to questions and perform multipurpose tasks. Interaction with GPT-4 is done through an API for development purposes or through an online interface for public users.
 - **Llama:** next generation open-source LLM trained on massive text data developed by Meta with abilities similar those of to GPT and Gemini.

Performance Evaluation

Metrics

Performance evaluation assesses how an algorithm handles unseen data that is representative of a general population. The common practice, called **holdout validation**, is to randomly split the available data into a training set (60%), a test set (15%) and a validation set (25%), each with a different purpose. The training set is fed into the machine learning algorithm, the test set is used to perform internal validation at each iteration of the training, while the validation set helps to assess the algorithm’s performance after it stabilized.

ML models use **metrics** that compare the real, measured values with the models’ prediction to assess performance in each iteration using the training and test. This procedure is

done using not only the training set to evaluate the learning performance, but also the validation set to evaluate the models’ generalisability. The most common metrics include accuracy, sensitivity, specificity, and precision, as defined in Table 2.1.

During the training process it is especially important to be mindful of **overfitting and underfitting** of the model. Overfitting occurs when the model has become overly detailed to the point where it begins to fit random statistical variations (noise). This can be noticed by a continued improvement of the metrics of the training set, but a stabilization or worsening of the test set’s metrics (Fig. 2.6). Underfitting, on the other hand, is the opposite where the model cannot account for various relations due to a lack of well-discriminating parameters. Consequently, it is a good practice to avoid overfits and underfits cases before computing the performance metrics

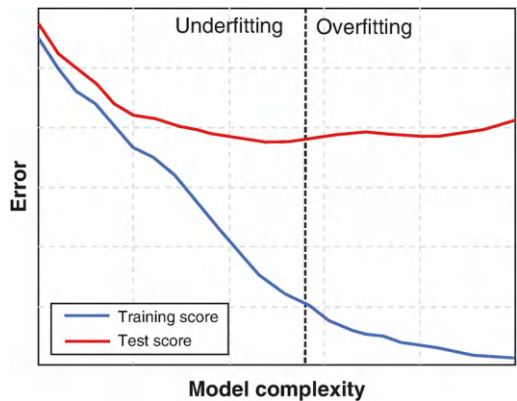


Fig. 2.6 Under- and overfitting

Table 2.2 Confusion matrix

	Actually positive	Actually negative
Predicted positive	True positive	False positive
Predicted negative	False negative	True negative

on the validation set. Meanwhile, in a good fit the performance of the training and test sets should be very similar.

Confusion Matrix

One popular way to represent performance is by using a **confusion matrix** that compares the algorithm’s classifications to the actual classification using the metrics in Table 2.2. Ideally, the non-diagonal values should remain 0. Usually confusion matrices are binary, but they may be expanded to include more than two classes.

Receiver Operating Characteristic Curve (ROC)

The ROC curve is a plot of the true positive rate as a function of the false positive rate (i.e. 100–Specificity) for different cutoffs. This aims to find the optimal cut-off that maximizes the balance between specificity and sensitivity. Typically curves close to the top-left corner of the plot represent “ideal” models with a false positive rate of zero, and a true positive rate of one, while curves close to the diagonal approximate random noise (Fig. 2.7). ROC curves can also be represented by the Area Under the Curve (AUC), which ranges between 0.5 for random noise and 1 for perfect models.

K-Fold Cross Validation Testing

This validation technique splits the training data into K folds, selects one for validation and builds the model based on the remaining (K – 1) folds. It can be considered a K-repetitive hold-out validation, where the test set and the

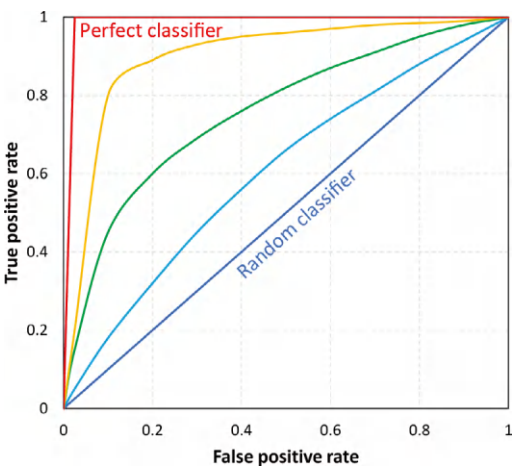


Fig. 2.7 Examples of ROC curves

validation set are independent for each iteration or run. There are several variations of the technique, such as stratified K-Fold Cross Validation, Leave-P-Out Cross Validation, etc.

Conducting a Machine Learning Analysis

To start a machine learning or deep learning analysis, you first need to have a clear understanding of the problem you are trying to address. First, look at the available parameters and what kind of outcome can be expected. Based on this assessment, you can decide what parameters may be the most appropriate for the analysis, or whether parameters must be combined or transformed before proceeding. Next, perform the pre-processing to remove missing values and outliers, and restructure the data to an [input, output] format for supervised learning, or an [input] format for unsupervised learning. Once the data is ready for training, select an AI algorithm based on the task to perform (e.g.

classification, clustering, time series prediction, etc.). If multiple AI algorithms are available, select the one that is easiest to implement and adjust to avoid a loss of time and computational cost. The actual training of the AI involves a non-linear optimisation based on your inputs and outputs. This step is mostly done in a black box format that is incorporated in your chosen software environment (e.g. Python, Weka, Matlab, etc.), yielding a trained model, along with its performance metrics. During this phase you should keep an eye out for overfitting, underfitting or large errors. If any of these occur, despite retraining several times, the parameters selected in data pre-processing step are not representative enough and other parameters should be chosen before retraining the model. Once the model performs well based on the metrics, ROC curves and confusion matrices based on the test sets, it is time to assess the model performance using the unseen validation data. Ideally, this should be completely independent data from a different centre, if possible. If this validation is also satisfactory, the model development is complete. A full overview of these steps is given in Fig. 2.8.

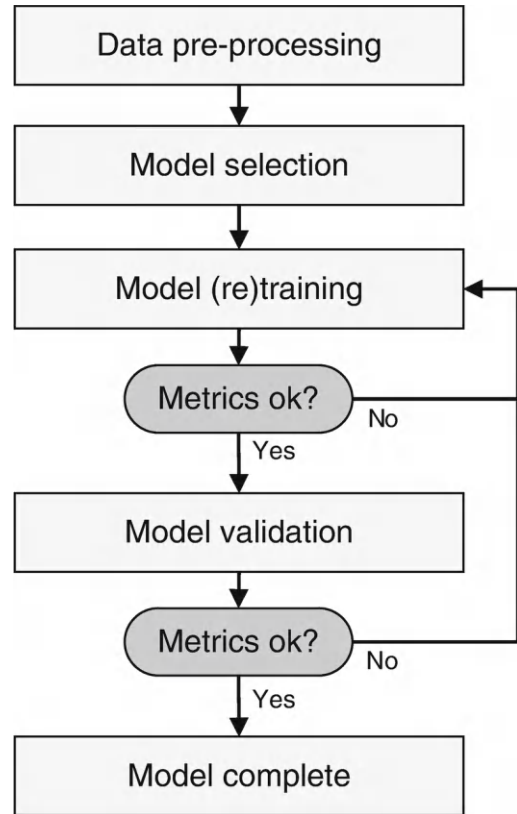


Fig. 2.8 Building a machine learning model

Software for Machine Learning Implementation

There are many software packages available that allow quick implementation of machine learning. One of the most popular is **Python**. This open source environment has a user-friendly syntax, is easy to learn and allows for rapid prototyping, which is often used for e.g. web development, mathematics or system scripting. Although implementing an AI algorithm can be time consuming, Python offers many handy libraries dedicated to AI, such as Keras, TensorFlow, and SciKit-learn. Other than a Python library, **TensorFlow** is also a leading stand-alone open source software package for machine learning and deep learning, used by ML beginners and practitioners alike. It has an open source toolkit for building machine learning pipelines, as well as support for applications

such as computer vision, natural language processing, speech recognition, and general predictive analysis [21]. Meanwhile, **Keras** is an open-source neural network library that is popular for its ease of use for fast experimentation and prototyping. It supports implementing deep learning algorithms such as convolutional neural networks and recurrent neural networks.

Weka is a Java-based open source package with a graphical user interface that can be operated with just a few mouse clicks or some minimal programming. This makes it ideal for beginners in machine learning. It implements several machine learning algorithms for classification, clustering and data pre-processing [22], as well as data mining.

Another widely used programming language is called **R**, which is optimized for statistical data analysis. It uses mathematical data structures such as vectors, lists and data frames

that are easy to manipulate by ML. R supports various machine learning algorithms in the form of community-authored packages such as K-nearest neighbours, Naive Bayes, decision trees, regression methods, neural networks and support vector machine [23].

Finally, **MATLAB** is a high-end, high-performance commercial package for scientific computing that integrates programming, computation and visualisation into a matrix-based interactive system. It has many extensions, called toolboxes, that include machine learning and deep learning. These allow implementing, training and validating various algorithms with applications in classification, clustering, and prediction. One major advantage of MATLAB is its user-friendly visualization and data pre-processing facilities, along with its ability to interface with other ML tools such as TensorFlow and Keras.

Applications in Ophthalmology

Current medical diagnostic technologies generate **massive amounts of data that are difficult to analyse** and hamper practitioners by their lack of interpretability. This prompts clinical practitioners to focus on a limited number of diagnostic criteria alongside their subjective expertise, creating the risk that certain indicators of a pathology may escape their attention. With its ability to extract meaningful knowledge in an objective and automatic way, AI has become an important tool for medical diagnosis and for decision making. The following, along with the following chapters, highlight some examples of AI applications in ophthalmology, with a focus on the technique and the achieved results.

Glaucoma

The earliest application of ML in ophthalmology [24] demonstrated that neural networks could be used to screen visual field data for signs of glaucoma with a performance similar to that of expert observers (ML: sensitivity=65%,

specificity=71%; Experts: sensitivity=59%, specificity=74%). Later studies improved upon these results and confirmed that ML can outperform clinical practitioners in the detection of glaucoma [25]. Current advanced ML algorithms, such as self-organisation maps with decision trees [26], are now able to detect glaucoma with a very high accuracy of 0.98 [27].

Retinal Disease

Deep learning is often used to accurately diagnose retinal conditions such as Diabetic Retinopathy (DR), Age-related Macular Degeneration (AMD), and Retinopathy of Prematurity (ROP). For example, DL was used to assess the quality of fundus images for ROP, achieving an AUC of 0.95 [28], while a CNN for the diagnosis of ROP achieved an AUC of (0.94, 0.99) for normal and diseased cases, respectively [29]. These values are similar as those of a human expert. The latter DL algorithm [30] was also applied to retinal images for the detection of referable DR, resulting in a sensitivity of 97.5% and a specificity of 93.4%. Meanwhile, predictive models identified the retinal biomarkers for drusen regression in intermediate AMD with an AUC of 0.75 [31]. Similarly, a DL system by Arcadu et al. [32] predicted a two-step worsening of diabetic retinopathy for a period of 12 months with an AUC of 0.79. These examples demonstrate the potential for using ML to assess the progression risk of retinal diseases.

Keratoconus

Although, the earliest attempts with AI targeted advanced stages of the disease, it showed potential for early detection and improved interpretability. Nowadays, AI is used to identify the earliest stages of the disease and score its severity. This early detection is especially important since it allows therapeutic action to halt progression, or screen these patients out from refractive surgery, thus avoiding postoperative

complications. For example, Sousa et al. [33] evaluated three classifiers (SVM, MLP, Radial Basis Functions), reporting a sensitivity of 98–99%, Arbelaez et al. [34] applied SVM to tomographic and pachymetry data, yielding an accuracy of 98.2% for keratoconus and 97.3% for early keratoconus, and other studies reported even better results for the earliest cases [34, 35]. Lopes et al. [36] examined tomographic data using Random Forests, with a sensitivity of 85.2%, specificity of 96.6% for the earliest cases. Finally, a hybrid ML algorithm combining supervised and unsupervised learning based on FFNN reported a sensitivity of 97.8% and specificity of 95.6% for suspect keratoconus detection [37].

Refractive Surgery

Recently, a Decision Forest classifier was developed for the individual, postoperative risk assessment after LASIK refractive surgery for follow-up periods of 12 years [38]. Meanwhile, Yoo et al. [39] developed a decision support system to screen out unsuitable candidates for refractive surgery. This study combined both medical devices and experts to achieve an AUC of 0.983, highlighting the benefit of human-AI collaboration. Finally, an artificial multilayer perceptron was used to predict the improvement in visual quality for KC patients implanted with intracorneal ring segments based on corneal curvature and astigmatism, yielding a best error of 0.97D of corneal power and 0.93D of astigmatism [40].

Other Applications

One of the striking demonstrations of the abilities of deep learning is that it was able to extract new knowledge from retinal fundus images to predict cardiovascular risk factors [41] or assess the retinal findings and Alzheimer disease [42]. Although, ophthalmologists have been looking at fundus images for many years, this knowledge was impossible to extract before because of

the large variability in retinal features, patterns, colours, etc.

LLMs and Generative AI were also adopted for various applications in Ophthalmology, primarily medical image synthesis and Teleophthalmology. VisionFM [43] is a Multi-Modal Multi-Task Vision Foundation Model trained on 3.4 million images to serve as a generalist ophthalmic AI that is supposed to diagnose, classify disease phenotype, and screen for various eye conditions at expert level. The model has been reported to outperform ophthalmologists in diagnosing 12 diseases including cataract, myopia, glaucoma, etc. GANs have also been used on OCT images to predict the short-term response to anti-VEGF treatment [44], detect glaucoma [45] and synthesize OCT images for training purposes [46]. Meanwhile, LLMs were employed for medical education, answering ophthalmic board questions, and writing operation reports [47].

Common Misconceptions

Since there are many misconceptions about AI, it is important to address these to inspire ophthalmologist users to have realistic expectations.

Technical Aspects

First of all, AI is **not a magical tool** that can solve any problem. Instead, AI are knowledge-driven algorithms that analyse massive amounts of data and extract patterns much faster than human biological synapses. But current AI systems require clearly defined problems to perform their tasks and are unable to surpass focused specific tasks. Therefore, if the user does not provide such a well-delineated problem or high-quality data to work with, the outcome will likely be underwhelming.

Contrary to what many users think, **larger datasets will not always yield better results**. The training data can be increased when it is not descriptive enough, or if there are patterns that

cannot be detected by the algorithm. But when a trained AI algorithm has an acceptable validation error, adding more data may lead to over-training the algorithm, divergence issues and an increased computational cost. Similarly, switching between classic ML algorithms or between classic DL algorithms should not make much difference. For instance, Multilayer Perceptron, Random Forest, SVM would all provide similar outcomes in most instances as they are blackbox combinations of similar (non-)linear functions. Instead, to get a remarkable improvement, you can revise the design strategy, add more descriptive features, remove correlations, increase the data set size, implement a hybrid machine learning system (e.g. supervised combined with unsupervised), or use advanced generative AI and LLMs that combine hybrid learning types etc.

Note, that for advanced applications the appropriate AI algorithm choice can be important. This however depends on the study objectives, and the dataset. If the dataset is multimodal exhibiting diversity in structure and the source, a multi-modal hybrid AI would be the most suitable choice. This also applies in cases where the goal is to implement a multi-purpose AI, such as screening multiple ocular diseases at the same time.

Social Aspects

There is a widespread fear that in many professions AI will eventually make human workers obsolete. While this may be true for boring, repetitive tasks that can be well described, new job opportunities will arise in the training, supervision and maintenance of AI systems. For now, the list of jobs that can be replaced by AI is limited, given the technical limitations listed above, along with the fact that computers are still **unable to emulate the important human characteristics**, such as teamwork, interaction, creativity, adaptability, empathy, etc., that are essential in a medical environment [48]. Although one could argue that AI could emulate some of these important human characteristics, it has yet to show true

creativity in open-ended problems, awareness or empathy. As such, AI lacks the all-important human aspect of the doctor-patient relationship that enhances the patient's understanding of his situation, his treatment compliance, the therapeutic effectiveness [49], as well as his physical, emotional, and social needs. Some aspects, such as e.g. initial anamnesis or questionnaires, could perhaps be handled by a virtual assistant similar to Siri or Alexa, provided a physician will later go over the responses with the patient to ensure nothing important was missed. This same is true for the screening and forecasting examples given in section "Applications in Ophthalmology", which should serve as a Decision Support System to help the physician reach a conclusion, rather than a Diagnostic System to replace the physician's expertise altogether. Finally, AI may help with screening, patient follow-up and scheduling, filling out patient files, letters and administration, provided it is supervised and corrected afterwards by a physician. **Current AI systems would therefore have to be embedded in a human context.** Provided such AI systems are developed with respect for human interaction, empathy and privacy, this could optimize time use and reduce waiting times in hospitals.

The exponential increase of AI in the field of ophthalmology raises concerns about whether these applications ensure quality and reliability guidelines, especially AIs incorporated in ophthalmic clinics. To this end, guidelines on AI ophthalmic clinical research evaluation were recently proposed [50] that emphasize the quality and the diversity of the data used to train the model, the importance of selecting appropriate models, as well as conducting robust validation. If an algorithm is to be integrated into practice, it is crucial to ensure user friendliness, along with continuous monitoring and close consideration for ethical use and regulatory compliance.

Conclusion

For very well-delineated tasks in ophthalmology, AI can reach exceptional levels of performance that supersede human ophthalmologists.

The technology suffers from a number of limitations, however, that make it unwise to rely solely on its output. Instead, AI systems are ideal to work in partnership with ophthalmologists, for example, for disease detection or as a decision support system.

References

- McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys.* 1943;5:115–33.
- Hebb DO. *The organization of behavior; a neuropsychological theory.* Oxford, UK: Wiley; 1949. p. xix, 335.
- McCarthy J, Minsky ML, Rochester N, Shannon CE. A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. *AI Mag.* 2006;27:12.
- Turing AM. Computing machinery and intelligence. *Creat Comput.* 1980;6:44–53.
- Mei Q, Xie Y, Yuan W, Jackson MO. A Turing test of whether AI chatbots are behaviorally similar to humans. *Proc Natl Acad Sci.* 2024;121: e2313925121.
- Newell A, Simon H. The logic theory machine—a complex information processing system. *IEEE Trans Inform Theory.* 1956;2:61–79.
- Samuel AL. Some studies in machine learning using the game of checkers. *IBM J Res Dev.* 1959;3:210–29.
- Weizenbaum J. ELIZA—a computer program for the study of natural language communication between man and machine. *Commun ACM.* 1966;9:36–45.
- Flanagan JL. *Speech analysis synthesis and perception.* Springer Science & Business Media; 2013.
- Goodfellow I et al. Generative adversarial nets. In: *Advances in neural information processing systems*, vol. 27. Curran Associates, Inc.; 2014.
- Vaswani A et al. Attention is all you need. In: *Advances in neural information processing systems*, vol. 30. Curran Associates, Inc.; 2017.
- Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Campbell JP. *Introduction to machine learning, neural networks, and deep learning.* Transl Vis Sci Technol. 2020;9:14.
- Taulli T. *Artificial intelligence basics: a non-technical introduction.* Berkeley, CA: Apress; 2019. <https://doi.org/10.1007/978-1-4842-5028-0>.
- Aggarwal CC. Outlier analysis: advanced concepts. In: Aggarwal CC, editor. *Data mining: the textbook.* Cham: Springer International Publishing; 2015. p. 265–83. https://doi.org/10.1007/978-3-319-14142-8_9.
- Rebala, G., Ravi, A. & Churiwala, S. *An Introduction to Machine Learning.* (Springer, 2019).
- Lo JT-H. Functional model of biological neural networks. *Cogn Neurodyn.* 2010;4:295–313.
- Gupta N, Trindade BL, Hooshmand J, Chan E. Variation in the best fit sphere radius of curvature as a test to detect keratoconus progression on a scheimpflug-based corneal tomographer. *J Refract Surg.* 2018;34:260–3.
- Kohonen T. Self-organization of very large document collections: state of the art. In: Niklasson L, Bodén M, Ziemke T, editors. *ICANN 98.* London: Springer; 1998. p. 65–74. https://doi.org/10.1007/978-1-4471-1599-1_6.
- Sutton RS, Barto AG. *Reinforcement learning, second edition: an introduction.* MIT Press; 2018.
- Kingma DP, Welling M. Auto-encoding variational Bayes; 2022. <https://doi.org/10.48550/arXiv.1312.6114>.
- Hope T, Resheff YS, Lieder I. *Learning TensorFlow: a guide to building deep learning systems.* O'Reilly Media, Inc.; 2017.
- Witten I et al. *Weka: practical machine learning tools and techniques with Java implementations;* 1999.
- Lantz B. *Machine learning with R: learn how to use R to apply powerful machine learning methods and gain an insight into real-world applications;* 2013.
- Goldbaum MH, et al. Interpretation of automated perimetry for glaucoma by neural network. *Invest Ophthalmol Vis Sci.* 1994;35:3362–73.
- Goldbaum MH, et al. Comparing machine learning classifiers for diagnosing glaucoma from standard automated perimetry. *Invest Ophthalmol Vis Sci.* 2002;43:162–9.
- Huang M-L, Chen H-Y, Lin J-C. Rule extraction for glaucoma detection with summary data from StratusOCT. *Invest Ophthalmol Vis Sci.* 2007;48:244–50.
- Kim SJ, Cho KJ, Oh S. Development of machine learning models for diagnosis of glaucoma. *PLoS One.* 2017;12: e0177726.
- Coyner AS, et al. Automated fundus image quality assessment in retinopathy of prematurity using deep convolutional neural networks. *Ophthalmol Retina.* 2019;3:444–50.
- Brown JM, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. *JAMA Ophthalmol.* 2018;136:803–10.
- Gulshan V, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA.* 2016;316:2402–10.
- Bogunovic H et al. Machine learning of the progression of intermediate age-related macular degeneration based on OCT imaging. *Invest Ophthalmol Vis Sci.* 2017;58: BIO141–BIO150.
- Arcadu F et al. Deep learning algorithm predicts diabetic retinopathy progression in individual patients. *NPJ Digit Med.* 2019;2:1–9.

33. Souza MB, Medeiros FW, Souza DB, Garcia R, Alves MR. Evaluation of machine learning classifiers in keratoconus detection from orbscan II examinations. *Clinics*. 2010;65:1223–8.
34. Arbelaez MC, Versaci F, Vestri G, Barboni P, Savini G. Use of a support vector machine for keratoconus and subclinical keratoconus detection by topographic and tomographic data. *Ophthalmology*. 2012;119:2231–8.
35. Smadja D, et al. Detection of subclinical keratoconus using an automated decision tree classification. *Am J Ophthalmol*. 2013;156:237–246.e1.
36. Lopes BT, et al. Enhanced tomographic assessment to detect corneal ectasia based on artificial intelligence. *Am J Ophthalmol*. 2018;195:223–32.
37. Issarti I, et al. Computer aided diagnosis for suspect keratoconus detection. *Comput Biol Med*. 2019;109:33–42.
38. Achiron A, et al. Predicting refractive surgery outcome: machine learning approach with big data. *J Refract Surg*. 2017;33:592–7.
39. Yoo TK et al. Adopting machine learning to automatically identify candidate patients for corneal refractive surgery. *NPJ Digit Med*. 2019;2:1–9.
40. Valdés-Mas MA, et al. A new approach based on Machine Learning for predicting corneal curvature (K1) and astigmatism in patients with keratoconus after intracorneal ring implantation. *Comput Methods Programs Biomed*. 2014;116:39–47.
41. Poplin R, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng*. 2018;2:158–64.
42. Schrijvers EMC, et al. Retinopathy and risk of dementia. *Neurology*. 2012;79:365–70.
43. Qiu J et al. VisionFM: a multi-modal multi-task vision foundation model for generalist ophthalmic artificial intelligence. <https://doi.org/10.48550/arXiv.2310.04992>.
44. Liu Y, et al. Prediction of OCT images of short-term response to anti-VEGF treatment for neovascular age-related macular degeneration using generative adversarial network. *Br J Ophthalmol*. 2020;104:1735–40.
45. Zheng C, et al. Semi-supervised generative adversarial networks for closed-angle detection on anterior segment optical coherence tomography images: an empirical study with a small training dataset. *Annals of Translational Medicine*. 2021;9:1073–1073.
46. Zheng C, et al. Assessment of generative adversarial networks model for synthetic optical coherence tomography images of retinal disorders. *Transl Vis Sci Technol*. 2020;9:29.
47. Waisberg E, et al. Generative artificial intelligence in ophthalmology. *Surv Ophthalmol*. 2024. <https://doi.org/10.1016/j.survophthal.2024.04.009>.
48. Korot E, et al. Will AI replace ophthalmologists? *Transl Vis Sci Technol*. 2020;9:2.
49. Di Blasi Z, Harkness E, Ernst E, Georgiou A, Kleijnen J. Influence of context effects on health outcomes: a systematic review. *Lancet*. 2001;357:757–62.
50. Yang W-H, Shao Y, Xu Y-W, Expert Workgroup of Guidelines on Clinical Research Evaluation of Artificial Intelligence in Ophthalmology. Ophthalmic Imaging and Intelligent Medicine Branch of Chinese Medicine Education Association, Intelligent Medicine Committee of Chinese Medicine Education Association. Guidelines on clinical research evaluation of artificial intelligence in ophthalmology. *Int J Ophthalmol*. 2023;16:1361–72.

Overview of Artificial Intelligence Systems in Ophthalmology

3

Natsuda Kaothanthong, Niracha Arjkongharn,
Nattaporn Vongsa, Varis Ruamviboonsuk,
and Paisan Ruamviboonsuk

Abstract

Many recent artificial intelligence (AI) systems in ophthalmology, starting with deep learning (DL) for screening of diabetic retinopathy (DR), has stirred excitement due to their robust performances. Abundance of DL models not only in common retina diseases, such as age-related macular degeneration (AMD) and retinopathy of prematurity (ROP), but in many other fields of ophthalmology, such as glaucoma and cataract, has proliferated since the first publications of DL for DR screening. Most of the published models are claimed to have performances at least on par with, if not better than, expert

ophthalmologists based on internal validation. Published articles on external validation of AI in ophthalmology are far fewer. Only handful articles on deploying AI for prospective, real-world validation are published. The general tasks for AI in ophthalmology, mostly performed on images, can be categorized into (1) classification or screening, (2) segmentation, and (3) prediction of disease progression or treatment outcome. Various architectures of deep convolutional neural networks, such as AlexNet, Inception, VGG, ResNet, etc., are used for classification of DR, AMD, glaucoma, ROP, cataract, etc. Conventional machine learning (ML) methods, such as Bayers' theorem, support vector machine, decision tree, etc., are still useful for automated segmentation of retinal layers, the optic nerve head area, retinal fluid, etc. Both conventional ML and DL may be applied to predict disease progression or outcome of treatment, such as progression of DR severity or vision after intravitreal injections. Although the performance of AI for this task can be improved with the combination of risk scores, the accuracy of approximately 70–75% is still lower than the other two tasks. There is an increase in articles applying generative adversarial networks for creating synthetic images for research where there is limitation of datasets. Large language models: generative pretrained transformer, or chatGPT-4.0, has

N. Kaothanthong
Sirindhorn International Institute of Technology,
Thammasat University, Pathumtani, Thailand
e-mail: natsuda@siit.tu.ac.th

N. Arjkongharn · N. Vongsa · P. Ruamviboonsuk (✉)
Department of Ophthalmology, Rajavithi Hospital,
College of Medicine, Rangsit University,
Bangkok, Thailand
e-mail: paisan.trs@gmail.com

N. Arjkongharn
e-mail: nirachaarj@gmail.com

N. Vongsa
e-mail: nattaporn.von@gmail.com

V. Ruamviboonsuk
Department of Ophthalmology, Faculty of Medicine,
Chulalongkorn University, Bangkok, Thailand
e-mail: varis.ruam@gmail.com

recently been shown to interactively generate texts with very high performance. The roles of these AI systems in ophthalmology are remained to be investigated.

Keywords

Artificial intelligence · Deep learning · Machine learning · Diabetic retinopathy · Age-related macular degeneration · Glaucoma · Retinopathy of prematurity · Cataract

One of the first successful systems of artificial intelligence (AI) in health care can be traced back to a study in the late 1970s. In this study by Yu et al. [1], a computer was able to recommend choices of antibiotics for treatment of meningitis with acceptability rate of 65%. This rate may not be very high but the corresponding acceptability rates of faculty specialists who performed the same task were only from 42.5% to 62.5%. It is obvious that this early AI system had a better performance than the specialists.

About two decades later, in the late 1990s, a system of AI in ophthalmology by Sinthanayothin et al. [2] was able to recognize the structures on retinal images, such as the optic disc, fovea, with sensitivity and specificity, both as high as 80.4% to 99.1%. It took another two decades later when another system of AI in ophthalmology, the iDx-DR [3], became the first system of AI in health care approved by the United States Food and Drug Administration (U.S. FDA) for automated detection of diabetic retinopathy (DR) for referrals to ophthalmologists in primary care settings. The iDx-DR has a sensitivity of 87.2%, and a specificity of 90.7% with the approval.

The approval of the iDx-DR has placed AI in ophthalmology at the forefront of AI in health care even though the studies of AI in other fields, such as pathology and radiology, may outnumber the studies of AI in ophthalmology. While majority of studies of AI in health care

focus on screening or early detection of diseases, such as screening chest diseases using chest X-ray images, AI in ophthalmology is also useful for other tasks [4], such as automated segmentation of retinal layers in macular edema due to DR, age-related macular degeneration (AMD), and retinal vein occlusion (RVO); automated segmentation of optic nerve head (ONH) in glaucoma; automated extraction of features, such as nucleus and capsule in cataract. In addition, AI has been studied for therapeutic and prognostic predictions, such as prediction of requirement for anti-vascular endothelial growth factors (anti-VEGF) injections and prediction of visual outcome after treatment of AMD.

This chapter provides an overview of the commonly used systems or models of AI in ophthalmology in a basic bioengineering perspective and how the models can be applied to common diseases in ophthalmology.

AI, Machine Learning, and Deep Learning

In overview, deep learning (DL), the widely used system of AI today, is a subset of machine learning (ML) whereas ML is a subset of AI. ML was commonly used before the current era of DL. The mention of ML in today's perspective usually means this conventional ML, not including DL. Though the current use of AI has generally been shifted from conventional ML into DL, it is misleading to understand that ML is rarely useful today. It is depended on the tasks. ML can still achieve robust performances in many applications in health care, particularly when combined with DL.

ML is an application in AI that enables a computer system to learn from input data. The learning process can also be called training. Knowledge obtained in this training phase is used by the system to infer the output of new data; this process is also called inferencing, see Fig. 3.1. There are two common learning approaches, supervised and unsupervised

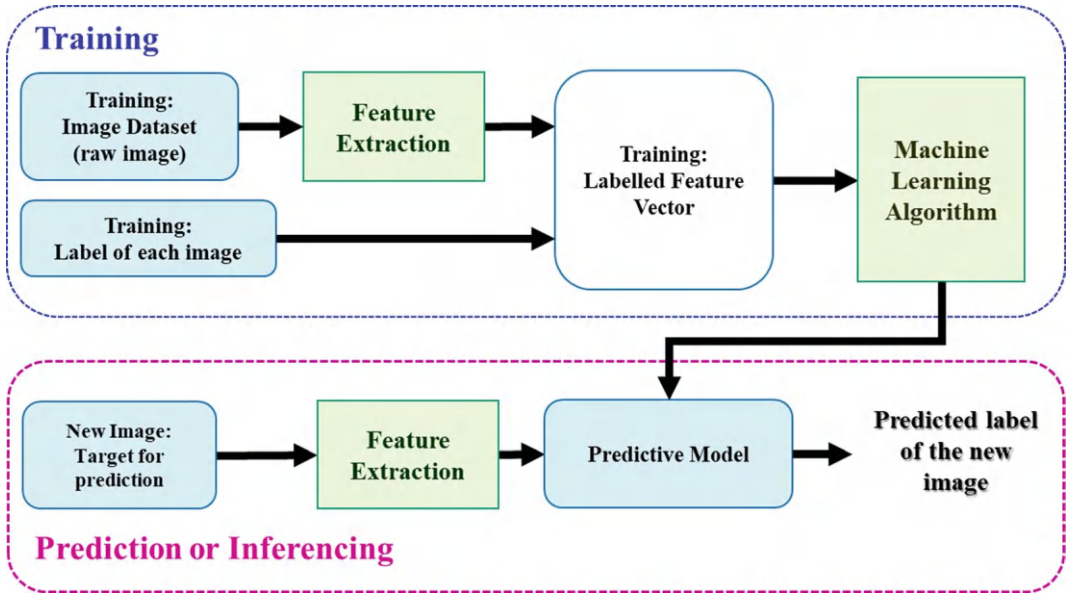


Fig. 3.1 Manual feature extraction in machine learning

learning. The difference between the two is the input knowledge which is provided by humans. The supervised learning requires the labelled inputs, such as images labelled with the outcomes: disease or no disease, whereas the unsupervised learning does not. It learns from finding homogenous groups of the input data using similarity of their features [5, 6].

On technical perspectives, as shown in Fig. 3.1, most conventional ML algorithms, either supervised or unsupervised approaches, require the process of feature extraction, which usually needs a domain expert. The application of the conventional ML requires an expert to guide a feature extraction process on how to transform the raw data, such as the pixel values of an image into a feature vector, for a model to learn for pattern recognition. To obtain a good model or algorithm, therefore, we need a suitable set of good features for training.

In the past, a natural solution towards a good model is to apply several feature extraction methods on the same image to obtain a set of different feature representations. However, as a trade-off, this plentiful number of the generated features could in turn lead to the sparseness of

the features. This is called a curse of dimensionality [7]. As a result, similar patterns of the same class of features could be hardly detected [8].

DL, on the other hand, allows the system to automatically extract features without assisting from the domain expert. This is why it is known as “black box”. The architecture of DL is a multilayered stack of simple modules known as a neural network, see Fig. 3.2. It is capable of discovering the feature representations from a set of raw input data for classification [9]. Each module transforms the input from the previous levels into a representation at a higher, slightly more abstract level. With multiple levels of transformation, very complex features, which are much too high dimensional to be accessible for human interpretation, are extracted and inferences can be performed [10].

Overview of Conventional ML Models

Supervised Learning Approach

Given input-output pairs assigned by humans, supervised learning finds a pattern of input

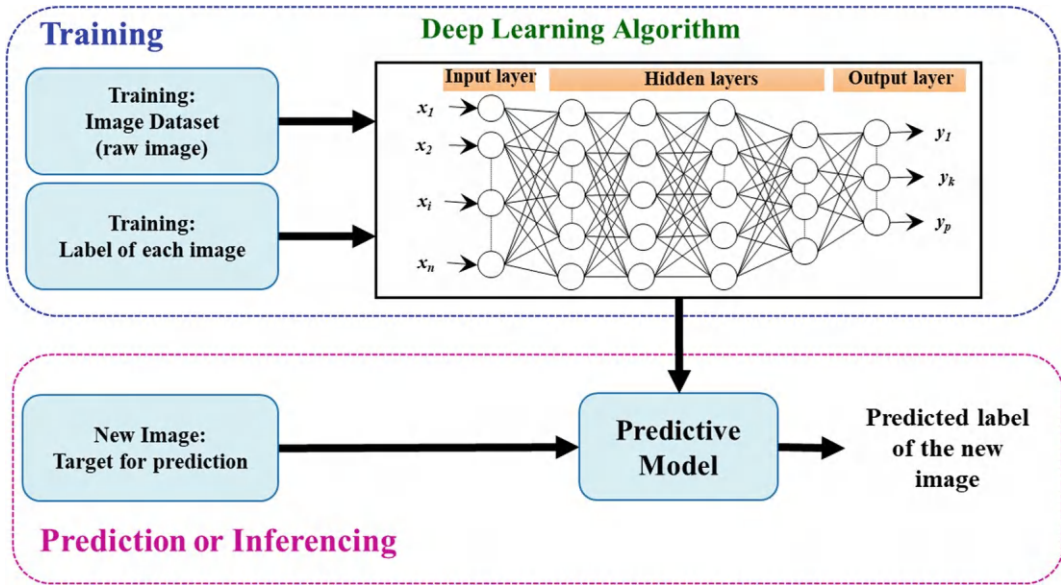


Fig. 3.2 Automatic feature extraction using the connectionist approach: deep learning

features that discriminates the different desired output. This pattern is considered as knowledge that is used to predict the output of a new input feature. There are many methods for supervised learning.

Naïve Bayes is a probabilistic method based on Bayes' theorem. It finds the probability that the desired output, denoted by B , will occur when the input feature, denoted by A , is presented, see the equation below:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

For example, in an image segmentation application, A is defined as a class, such as disease or no disease, and B is a feature extracted from a pixel in an image. The pixel B is classified as either disease or no disease using a ratio of the likelihood of the feature B occurring in the area of class A .

Naïve Bayes can also be applied for an image classification. For example, the probability that the features of an input image are DME or Non-DME. There are many Naïve Bayes theorem-based methods such as multinomial Naïve Bayes for predicting classes and Gaussian Naïve Bayes for predicting continuous value [11].

Support Vector Machine (SVM) finds a line segment or a hyperplane that optimally discriminates the features of two classes. The data on one side of the hyperplane should contain the data of the same class as much as possible. Figure 3.3 shows an example of an image classification application using SVM, where each point represents an image feature. New data, when given to SVM, are classified according to their positions on the plane [12]. SVM can be applied as both image classification and image segmentation. The latter is by assessing features of each pixel and classify them as either

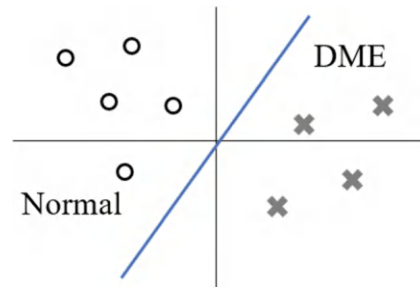


Fig. 3.3 An illustration of a classification using a support vector machine (SVM)

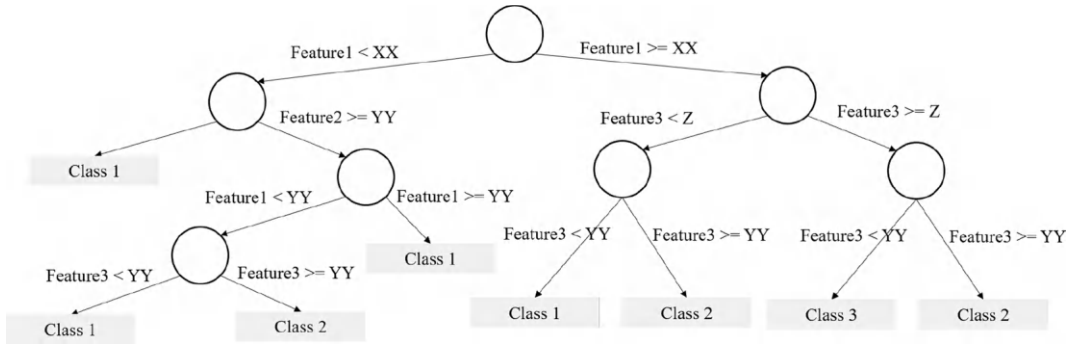


Fig. 3.4 Example of a decision tree for classifying the diseases

background (no disease) or foreground (disease) for segmentation (Fig. 3.11).

Decision Tree is a binary tree-based method that recursively divides the features of input in the training dataset into two parts until the optimal split between each class of the output has been reached. For each separation, the value of a feature is used to optimally divide the features into the two subgroups, such as disease and no disease. Each subgroup is divided further using the same or different value of feature. Figure 3.4 shows an example of a Decision Tree for classifying the disease. The features in the training set is separated into two subgroups using “Feature1”. Each subgroup is divided further using “Feature2”, where the training data of “Feature2 < YY” are mostly in Class 1. The subdivision is continued until the optimal class separation is found, as shown in the gray-shaded labels.

Instead of relying on only one tree, **Random Forest** applies multiple trees to learn the input features [12]. Criteria for separating the input features into homogeneous groups are defined differently for each tree. To predict the output of a new image, features are extracted and used by decision Trees in the forest. The classification outputs that are decided by each tree are voted. The output that achieved the highest number of votes becomes the prediction result.

Artificial Neural Network (ANN) extracts relevant features from the input data by learning from examples without explicitly stating the rules to perform classification tasks [13]. It

applies the concept of connected neural network where each neuron adjusts the weights (the optimal parameters) from the precedent neurons for the learning process. ANN has been applied in many tasks and is also a fundamental of deep learning.

Unsupervised Learning Approach

Whereas the supervised learning approach relies on the input-output pairs assigned by humans, Unsupervised learning approach requires only the input features to separate them homogeneously. The aim of the unsupervised learning approach is to *discover* a structure or distribution in the input data in order to *learn* more about each separated group.

The unsupervised learning is used when the input-output pair is not provided. It has widely been used in an image segmentation task to separate the set of pixels into a group of background and foreground or a region of an interest object (Fig. 3.11). In addition, it is also applied for studying objects in each homogeneous group.

K-Nearest Neighbor (KNN) finds sets of objects whose features are similar to the input features. Distance among the input features is used as a similarity measure [14]. Given a feature of new data, the classification result is achieved by voting the number of the k closest objects to the new data.

Boosting Algorithms

Boosting is a generic algorithm that aims to improve the accuracy of the prediction result. Instead of relying on the prediction outcome of a single model, boosting algorithms apply multiple weak classifiers trained with new data to achieve a good classifier [15]. Outcome of the precedent weak model is connected to a new model together with the new data to train and improve the prediction outcome. There are many boosting algorithms, where each applies different measures to improve the prediction accuracy. Adaboost [16] and Gradient Boosting are examples of these algorithms that are used in ophthalmology, particularly for the prediction task.

Overview of DL Model

Methods of DL that are commonly used in ophthalmology may be classified into Convolutional Neural Networks (CNNs), Pre-trained Unsupervised Networks (PUNs) [17], and Recurrent/Recursive Neural Networks (RNNs) [18]. Among these three different categories of DL networks, CNN has been used more extensively in medical image recognition [19] including in ophthalmology.

Convolutional Neural Network (CNN)

The CNN is designed to automatically extract features in two-dimensional data, such as images, while merging semantically similar features into

one to reduce sparseness. The features extracted using CNN can preserve important information for obtaining a good prediction. In addition, one or multiple images can be used as input, and a single diagnostic feature is designed as the output, such as disease presence or absence.

The architecture of CNN comprises three layers: Input Layer, Convolution Layer, and Pooling Layer (see Fig. 3.5). An input image is placed in the first Image layer, as shown in Fig. 3.5a. The image is a two-dimensional array, where each cell in the array is in a three-color-channel, red, green, and blue. Each channel is considered as a matrix and applied for a feature extraction. To obtain a rich representation, the input image is divided into smaller sub-images. Each sub-image is used in the subsequent layers.

The Convolution Layer in Fig. 3.5b then applies a filter to extract a feature from the input matrix. The objective of this layer is for feature extraction using different filters. The output of the Convolution Layer is varied according to the filters, such as edge or texture [20]. Figure 3.6 illustrates two examples of the convolution result using two different gradient filters [21]. Blood vessels can be clearly represented using Filter A while the optic disc can be visualized using Filter B. Since the performance of the prediction model is also depended on the weights (optimal parameters) [22], many filters, therefore, are applied to extract features from an image. The problem of using too many filters to obtain a rich representation of an image is the sparseness of the feature which results in a low accuracy. To cope with this limitation, CNN utilizes the Pooling Layer for a dimensionality reduction.

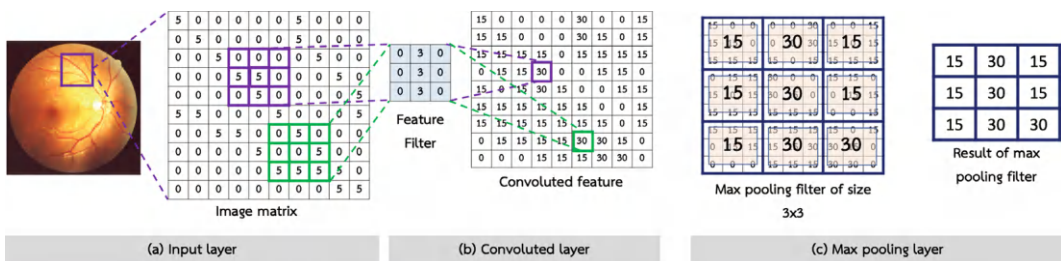


Fig. 3.5 Architecture of convolutional neural network

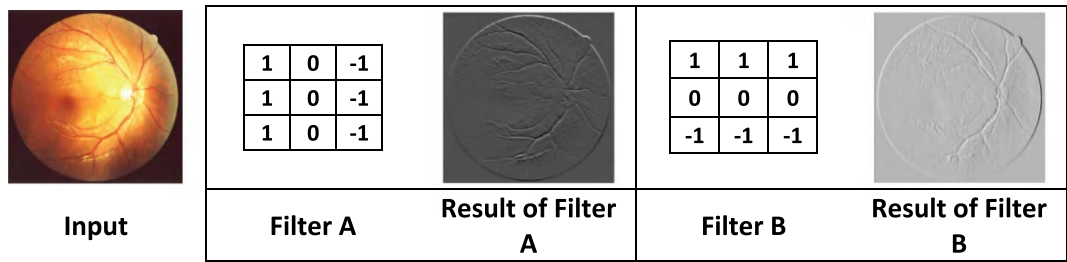


Fig. 3.6 Output of an input image using a filter

The Pooling Layer in Fig. 3.5c applies a filter to preserve important information of the extracted feature in the previous layer and down-sampling it into a smaller size. The filter can be in any size, such as 3×3 filter as shown in Fig. 3.5c. The value of the extracted features is summarized using one of the three mappings: Max Pooling, Average Pooling, and Sum Pooling. In addition to the dimension reduction, the Pooling Layer is useful for extracting dominant features to achieve rotational and positional invariant. In other words, it can distinguish between non-disease and disease location in an image.

Applying the CNN in DNN is typically structured as a series of stages where each stage consists of a CNN unit as shown in Fig. 3.7. The number of the units depends on the number of filter size, also called a network *width*. The stages are connected as shown in Fig. 3.8a. The number of stages refers to a network *depth*. The first few stages focus on mapping features from input images. The later stages apply the features from the previous stages as input to merge semantically similar features into one [9, 19]. The last stage is fully-connected layer for

predicting the classification result. This fully-connected layer can be viewed as a conventional ML that applies an input feature from the Pooling Layer for a classification.

The architectures of DNN can be varied in the arrangement of the number and size of filters for feature extraction, in the connections between these features, and in the network depth as depicted in Fig. 3.8. The number of filters defines the width of the network while the depth refines the learning capability of the network. Many DNN architectures have been created, such as AlexNet [17], VGG [23], Inception [24], ResNet [25], and EfficientNet [26]. They employ the same structure of CNN for feature extraction, however, the number of layers and feature mappings (filters) are varied. Their efficiency, such as accuracy, and time for training are also varied.

AlexNet

The difference of the original CNN and AlexNet is the number of convolution layers. AlexNet comprises eight layers where five layers are convolution and the rest are fully-connected layer [17]. It is the first architecture that applies the

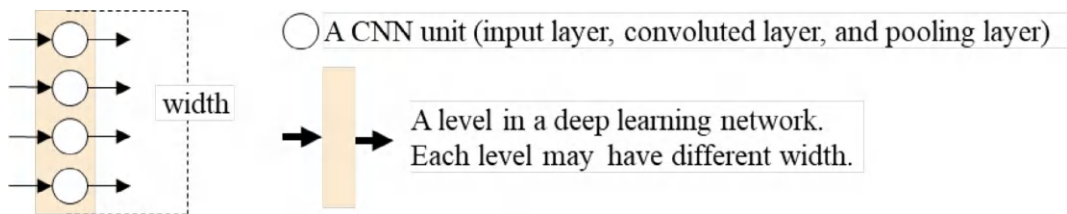


Fig. 3.7 An overview of the deep learning neural network (DNN) using CNN

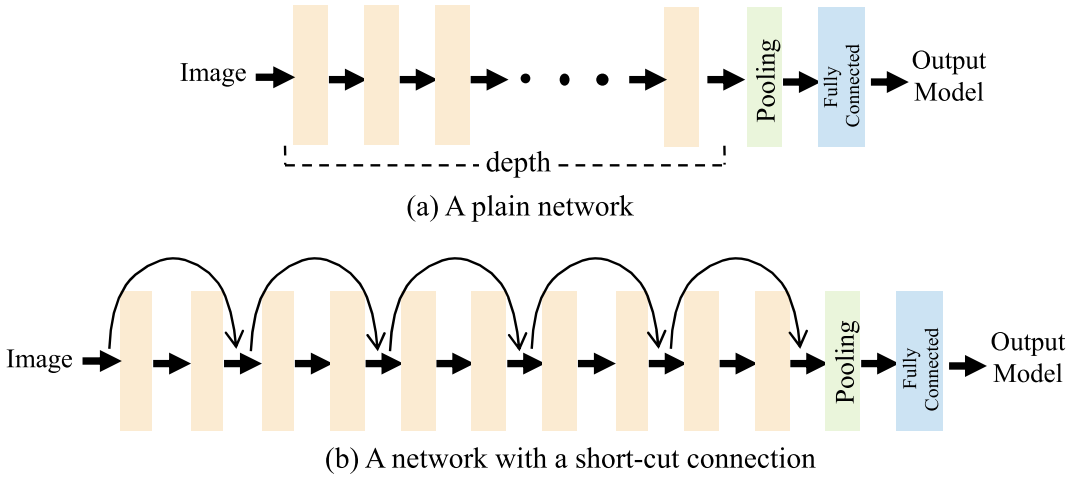


Fig. 3.8 Examples of the deep neural network using CNN. (a) A plain network (b) A network with a short-cut connection to skip some layers

multiple-layered CNN with graphic processing unit (GPU) to accelerate computational time of the DL.

VGG

Since increasing the depth of the CNN showed more accurate results of its classification [23], the architecture of VGG applies a small-sized filter of 3×3 and a deeper weight layers to the CNN. The architecture of VGG16 and VGG19 are similar, except only in the depth of the weight layers. With a smaller filter, compared to the original CNN, VGG shows a significant improvement of the result.

Inception

The drawbacks of increasing depth and width of deep neural networks are an overfitting of the prediction model when the training size was limited. Overfitting means the model performs much poorer in a new dataset than in the training dataset. Also, a uniformly increased network size means the dramatically increased use of computational resources. The architecture of Inception is to increase the depth and the width of the network to achieve a higher accuracy while keeping the computational budget constant [24].

ResNet

Although increasing the depth of a network improves its performance, it can lead to a degradation. The problem of degradation causes saturation of the network's performance. Instead of passing every layer in the network, ResNet applied a "shortcut" connection to skip some CNN layers, see Fig. 3.8b. The shortcut is used when the output feature of the layer is the same as the one before, then this particular layer can be skipped, and the degradation problem can be resolved [25]. There are many configurations of ResNet architecture: ResNet35, ResNet50, etc. The number refers to the depth of the network.

EfficientNet

Other than scaling up the depth of deep neural networks, such as in VGG [23], Inception [24], ResNet [25], scaling up the width [27] and increasing image resolution [28] are other means to improve network performance. EfficientNet [26] presents a compound scaling method which achieves the maximal accuracy by uniformly scaling network width, depth, and resolution of an input. However, it requires a specialized pipeline parallelism library to train.

Pre-Trained Unsupervised Network (PUN)

PUN is a family of DL models that uses unsupervised learning to train each of the hidden layers in a neural network to achieve a more accurate fitting of the dataset. Types of PUNs include Autoencoders, Deep Belief Networks (DBN), and Generative Adversarial Networks (GAN). In image processing tasks, PUN is trained using the very large number of images without considering labels of the data [29]. The weights obtained from the “pre-training” could be used as an initialized value for refining weights for a different target domain such as detecting glaucomatous optic neuropathy (GON) in fundus images [30]. Utilizing pre-training data in another different task is called “transfer learning”. Since the sample and dataset sizes of medical images are usually small, compared to non-medical images, transfer learning has become an important popular technique.

Recurrent Neural Network (RNN)

The Recurrent Neural Network (RNN) is a neural network designed to process an input sequence one element at a time [18]. The characteristic of RNN is the prediction of the next occurrence of the element based on the previously learned data. In addition, it accepts a series of an input with no pre-determined limit on the size. It has been found to be very good at predicting the next character in the text. The Long Short-Term Memory (LSTM) is a kind of an artificial recurrent neural network (RNN) architecture. Davidson et al. used RNN for localization of cone photoreceptor cells in healthy and patients with Stargadt’s disease [31].

Training a DL Model

On engineering perspectives, the objective of training a DL model is to find the optimal parameters, also called weights, that yield the

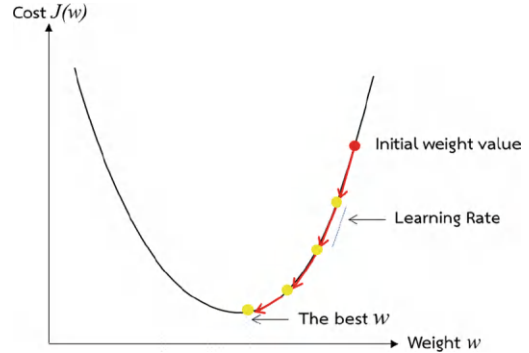


Fig. 3.9 The relationship of the weight values, the learning rate, and the cost value

best model performance. In DL, a cost, denoted by $J(w)$, is the penalty on any error that the model makes with a specific set of weights, denoted by w . In the beginning of the training process, the initial weight is a set of small random values. For each training iteration, a hyperparameter called “learning rate” is used to scale the magnitude of the weight, which is updated for the next iteration, see Fig. 3.9. To achieve the best performance, the model is trained several times to obtain the weights that minimize the cost.

The duration of the training process and the optimal weight values depend on the learning rate. A smaller learning rate results in a more reliable model, but the duration of training will be longer than using a larger learning rate. However, using a large learning rate may result in skipping the optimal weight values. Figure 3.10 shows the comparison between using the small learning rate and the large learning rate. It can be seen in Fig. 3.10b that the update of the weight using the large learning rate may skip the optimal weight value. On the other hand, using the small learning rate as in Fig. 3.10a requires many iterations to achieve the optimal point. Assigning an appropriate learning rate, therefore, is essential in training the model.

Epochs (the number of rounds that the entire dataset is passed through in a DL network [32]) is another hyperparameter that limits the number

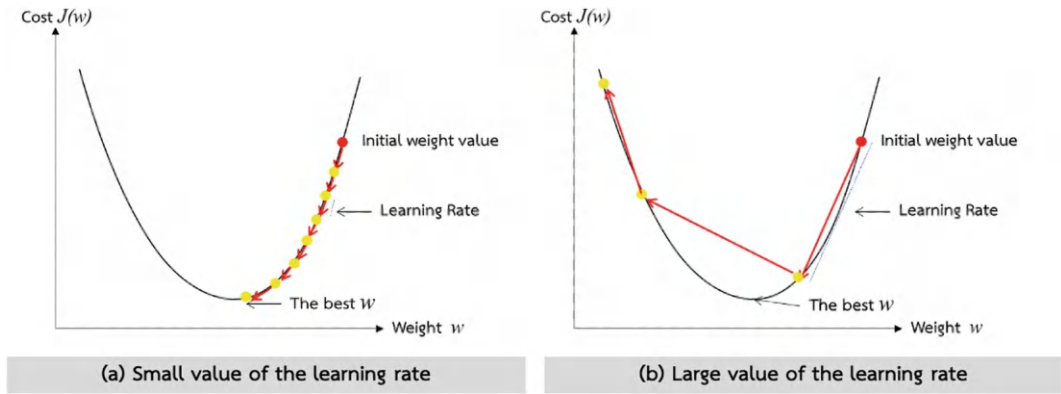


Fig. 3.10 A comparison of a small and a large value of the learning rate

of training iterations. Although a smaller learning rate provides a more reliable result, the training iteration may stop before obtaining the optimal weight value due to the epochs hyperparameter. Assigning appropriate initial weight values can reduce the number of training iterations without the limitation due to Epochs. With transfer learning, these initial weight values can be assigned from the already available model, also called “pre-trained” model. There is a number of pre-trained models for CNN-based DL [33].

Testing a DL Model

There are many terms in evaluation of an AI model which are not yet settled [34]. “Validation” means evaluation of performance of an AI model. When “Internal Validation” is used, it means evaluation of the model’s performance in the same dataset as training. Some studies call this “testing”. On the other hand, if the evaluation is conducted in the new dataset which is different from the training dataset, this is called “External Validation”. Generally, data in the original dataset for developing an AI model will be divided into training set and testing (internal validation) set with 80:20 proportion [34]. There are many public datasets of retinal fundus images available for development of AI systems. The top three most commonly use are Kaggle, Messidor-2, and EyePACS [35].

A confusion matrix is used to show the performance of a supervised learning model. It visualizes the predicted results versus the actual labels assigned by humans. Shows a confusion matrix of ‘Disease’ and ‘No Disease’ outcomes. The ‘True Positive’ and the ‘True Negative’ show the numbers of data that are correctly predicted. The ‘False Negative’ and the ‘False Positive’ show the numbers of wrong predictions. To measure the performance of unsupervised learning, log likelihood or distance measure is preferable [36].

The confusion matrix can also be used to compute sensitivity and specificity values. The sensitive value shows the model performance on predicting the ‘disease’ (positive) class, while the specificity refers to prediction of the ‘no disease’ (negative) class. Either Area Under the Curve of the Receiver Operating Characteristics (AUC) which is the area under the plot between sensitivity and 1-specificity or accuracy which is the proportion of the True Positive and True Negative and the total can also be used for judging performance of an AI model (Table 3.1).

Types of Learning Methods

As mentioned earlier, there are two learning methods of machine learning which are supervised learning and unsupervised learning. Due to a large dataset for DL, a continuous

Table 3.1 A confusion matrix for predicting two outcomes

Predicted results	Actual labels	
	Disease	No disease
Disease	True positive	False positive
No disease	False negative	True negative

generation of new data, and a limit labelled data, employing a traditional learning method is not feasible. For example, re-training the prediction model for every new data generated is inefficient and may take time for a large training dataset. In this section, techniques for learning such as reinforcement learning, transfer learning, incremental learning, and few-shot learning are explained.

Traditional Learning Method

Traditional learning method takes a training dataset to train a prediction model. It employs a static set of data to perform feature and knowledge extraction. In Fig. 3.11, the labelled feature vector of training set is used by a prediction model to extract knowledge to separate the classes of samples. The prediction model obtained from the training is used to predict the label of the new data. With a static dataset, the predicted label of the new data is limited to the known label in the training set. For example, the training dataset contains OCT images with labelled ‘AMD’ and ‘Normal’, therefore, the prediction model can either give the result of

‘AMD’ or ‘Normal’ when the new OCT image is used for testing.

To enhance the knowledge of the prediction model, samples of new labels are combined with the training dataset to train a new prediction model. For example, the new labels are OCT of ‘DME’ and ‘PCV’. In this way, the images of labelled ‘AMD’ and ‘Normal’ are combined with ‘DME’ and ‘PCV’ to train a new prediction model. When an image is input in the testing, this model can predict four labels, which are ‘AMD’, ‘Normal’, ‘DME’, and ‘PCV’.

A limitation of the tradition learning is the whole training process must be restarted when the new label is introduced to upgrade the model’s capability. As the number of samples increased, the training time also increased.

Reinforcement Learning

Reinforcement learning (RL) is a machine learning method that an agent learns to decide an action to respond to the environment as shown in Fig. 3.12. It does not rely on historical data solely to make decision, but any new data. RL enables the agent to make a series of decisions that maximize a reward for the task without human intervention or program explicitly to achieve the task. The agent knows the outcome of the action towards the obtained reward then adjusting the knowledge state of the agent to improve decision next time [37]. If the reward was not obtained, the agent improves the models for the decision making.

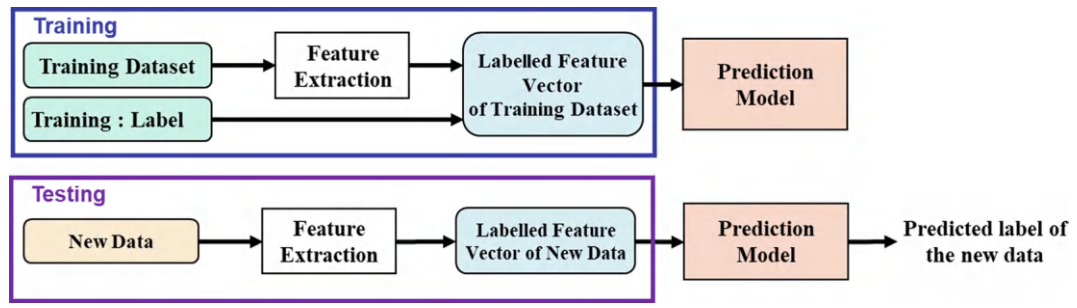


Fig. 3.11 Overview of tradition learning

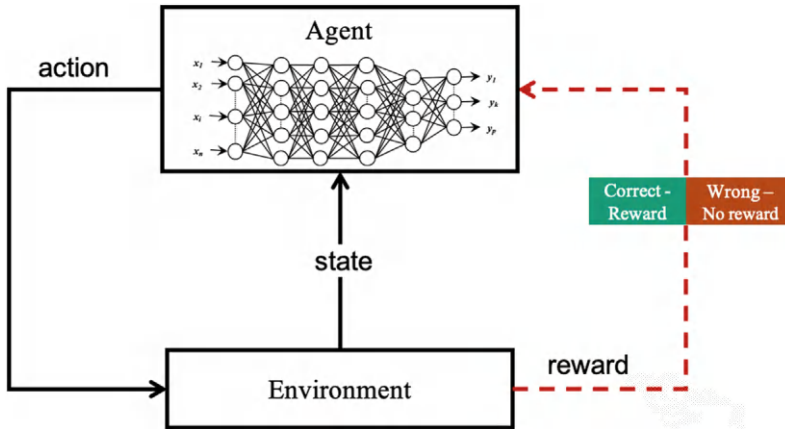


Fig. 3.12 Overview of reinforcement learning

For a classification task of ‘AMD’ or ‘Normal’ on fundus images using reinforcement learning, the goal is to teach the agent to separate the images of the two classes. The environment is the fundus images in the clinic. The reward could be considered from a signal given by an ophthalmologist. When the environment and the reward are defined, the agent is created by selecting machine learning algorithms and training multiple models to accomplish the task. Given a new data (state) into the environment, the agent selects the action from multiple models and a reward is given if the task is successful. Training an agent using reinforcement learning is an iterative process since the training process could be retrained to improve the agent’s performance.

RL has been used in many applications such as robot controllers in factory and self-driving cars [38]. ChatGPT uses human feedback as a reward to improve its reliability and conversation to human [39]. It is also integrated to health data to enable the capability of the agent to answer health-related conversation.

Transfer Learning Method

When developing a model using deep neural network, a vast amount of training data is needed to allow the model to extract knowledge.

With a large dataset, the parameters weights are adjusted to accomplish a high performance of the assigned task such as a classification. However, not every task has a large training dataset.

Transfer learning (TL) is a learning method that applies the knowledge of a solved task to a different but related task [40]. To train the model with a limited training data, the weight parameters of the other task are used to as an initial weight. TL enable the deep neural network to shorten the training time of computing initial weights during the parameters fine-tuning of the network. Example in Fig. 3.13, Task 1 was trained with a dataset to classify image fundus of ‘AMD’ and ‘Normal’ classes. Task 2 is aimed to classify ‘DME’ and ‘PCV’ fundus images. Since the training data set applies the same data of fundus images, the weight parameters of Task 1 were used as initial weight during training of the classification model for Task 2 to accomplish the objective that is to classify ‘DME’ and ‘PCV’ fundus images.

Incremental Learning Method

Traditional learning method trained a model on an entire data set. In a real-world application like a face recognition system of a building, images of new faces enter the system weekly or

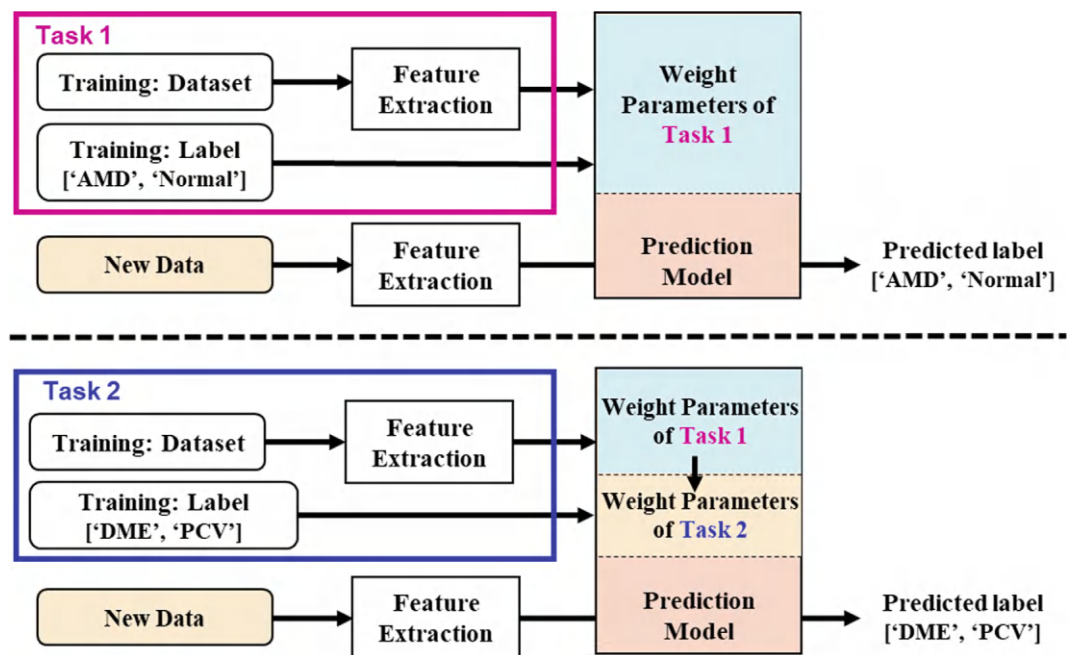


Fig. 3.13 Overview of transfer learning

monthly. To enable the model to recognize the new faces as well as the previously known faces, the new data must be included into the previously trained dataset to allow and re-train the entire dataset. A problem of training the model entirely is time and performance of recognizing the new faces since the number of images is limited. To avoid a re-train of the entire dataset, an incremental learning was introduced.

Incremental learning (IL) is a learning method that a model learns and enhances its knowledge progressively; without re-training the entire dataset and forgetting previously acquired information as well as achieving an acceptable performance. Figure 3.14 shows training classification model of the based classes from the training dataset, which is called based classes. When the data of the new classes that have never been in the based classes were introduced, they are trained incrementally on the existed model to enhance the capability to separate the new class. Reducing re-training time is an advantage of IL, however, it suffers from catastrophic forgetting that leads to a dramatic decrease in

overall performance when training with new classes added incrementally [41]. Catastrophic forgetting is a phenomenon that the prediction performance of knowledge that was learned previously is significantly lower while the new classes are better. With deep neural network in IL, the parameter weights are adjusted in accordance with the current data.

IL is different from transfer learning. The transfer learning applied the knowledge of a solved problem to a different but related problem. For example, employing the trained deep neural network of 'AMD' and 'Normal' fundus image classification to train another model for classifying 'DME' and 'PCV' images. On the other hand, IL continuously and efficiently learn new knowledge (classes) while maintaining the performance on old knowledge [42]. For example, the classification model was trained with 'AMD' and 'Normal' fundus image, later 'DME' and 'PCV' images are additionally trained. In this way, the capability of model is enhanced from binary- to multiple-class classification, i.e., from two to four classes.

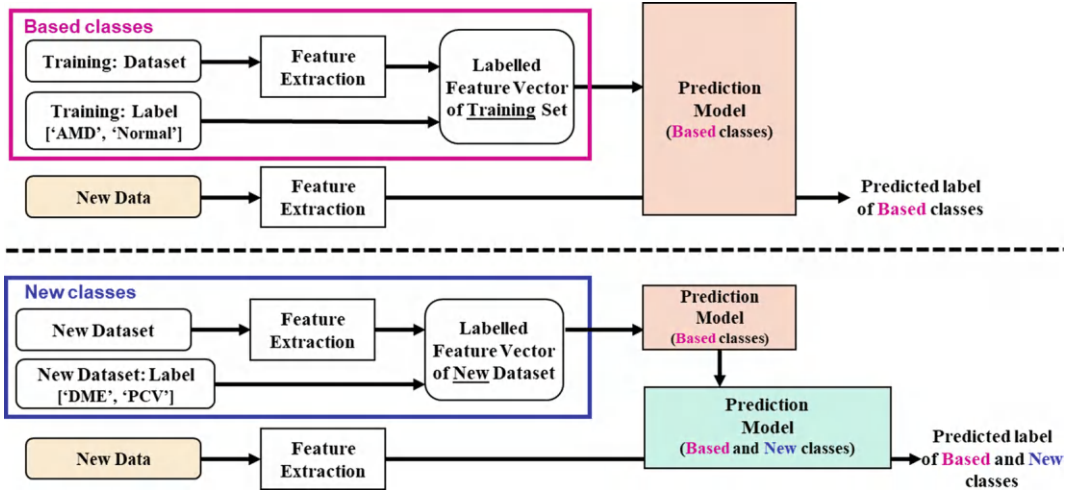


Fig. 3.14 Overview of incremental learning in classification task

Few-Shot Learning Method

Machine learning is successful in data-intensive applications, but it often restricted when the dataset is small. The term “Few-Shot” refers to classes with a very small number of samples. Few-Shot learning (FSL) is a method of learning of a limited data with supervised information for the task such as classification [43]. The goal of FSL is not to let the model recognize the sample of novel classes in the small training set and then generalize to the test set [44]. However, in traditional machine learning, the novel classes may not be identified due to a relatively small sample. Scenario of FSL has been employed to develop a model for the rare cases such as drug discovery and rare diseases [45].

One intuitive method with deep neural network is similar to the transfer learning that first trained on numerous available base class data as a prior knowledge and then fine-tuned using few-shot novel class data, see Fig. 3.15. In the survey of Wang et al. [43], the prior knowledge obtained by transforming samples with label from the training dataset, samples from a weakly label or unlabeled dataset, and samples from similar datasets were studied. The result shows that FSL performs better on the transformed knowledge from weakly supervised or unlabeled

samples. However, if the few-shot class has some similar classes, the prior knowledge from samples of similar classes is a good choice.

FSL is different from transfer learning. FSL method focuses on training a dataset with a very small number of samples, while TL focuses on transforming the knowledge of a solved problem to a different one. With deep neural network, FSL employed the prior knowledge from the network previously trained with based classes then trained with novel classes of a very small dataset. In addition, the goal of FSL is to develop a model that can identify the classes of few-shot. While the transfer learning only utilized the prior knowledge to shorten the network fine-tuning.

Few-Shot Incremental Learning Method

As mentioned earlier, the term “Few-Shot” refer to a very small samples while Incremental learning (IL) is a learning method that enhances the knowledge of the model progressively; without re-training the entire dataset and forgetting previously acquired information. Few-Shot Incremental Learning (FSIL) is a method of learning that intensify the knowledge of the model gradually with few-shot samples. The

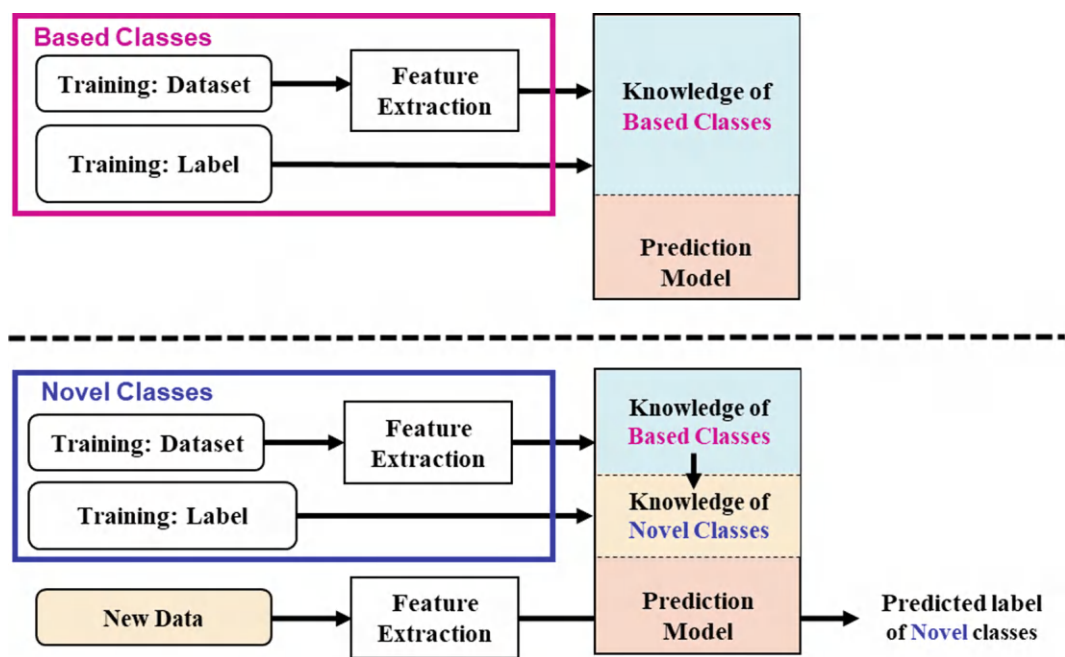


Fig. 3.15 Overview of few-shot learning

incrementally trained model must be able to prediction the classes that was previously learned with based classes as shown Fig. 3.16 as well as the novel classes.

Similar to IL, FSIL also suffers from catastrophic forgetting problems and also

“overfitting new” which the model performs well on predicting the new class [46]. However, compared with IL that learns new classes with unlimited, usually large-scale training samples, FSCIL is more challenging, since the number of new training samples is very small.

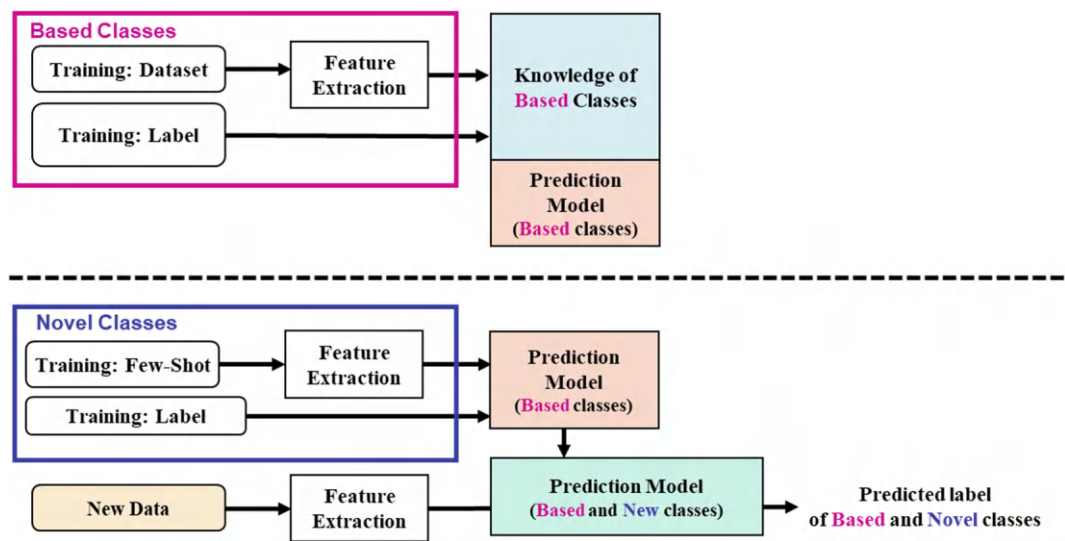


Fig. 3.16 Overview of few-shot incremental learning

Systems for Screening and Classification

Wilson and Jungner, on behalf of the World Health Organization (WHO), proposed principles and practices for diseases screening in 1968 [47]. Dobrow et al. then conducted a systematic review and modified Delphi consensus process on screening diseases [48]. The newly consolidated principles by Dobrow et al. was focused on screening programs and system principles rather than disease, test, and treatment as in the original principles by Wilson and Jungner.

Eye diseases can be screened easily using conventional visual acuity charts. Independent to visual acuity, many eye diseases can be screened using ocular imaging modalities. Of all these images, color retinal photographs (CRP) may be the most ubiquitous, practical, and least expensive. CRP can be applied for screening or classification of eye diseases in both posterior and anterior segments of an eye, with the highest number of studies in AI for this task, compared to other images, such as optical coherence tomography, scanning laser ophthalmoscopy, and angiography.

Screening or Classifying DR

Screening Based on CRP

DR is an eye disease which is found to be cost-effective for screening which has been conducted worldwide. Among the many available systems of AI for DR screening, the first autonomous system (no requirement of additional intervention by clinicians) approved by the U.S. FDA to date is the iDX-DR system [3]. This AI model evolved from integrating only conventional ML algorithms to combining with DL algorithms (AlexNet and VGG) and showed improvement of the AUC for detecting referable DR from 0.937 to 0.987 [49]. This hybrid system requires two fields of CRP, one with macula-centered and another with disc-centered from Topcon NW-400 non-mydratic retinal cameras, as inputs [50].

The U.S. FDA cleared the performance of iDX-DR with a clinical trial [50] that used this system to prospectively screen 900 patients with diabetes in 10 primary care units in the U.S. The standardization of this trial was four wide-field stereoscopic retinal images, representing the 7-field Early Treatment Diabetic Retinopathy Study (ETDRS) stereoscopic standard photographs, interpreted by certified graders from Wisconsin Reading Center. This trial found iDX-DR provided 96.1% interpretability, sensitivity of 87.2%, and specificity of 90.7% to detect more than mild DR (mtmDR).

The second AI system for DR screening approved by the U.S. FDA was EyeArt. It was the first AI approved for using with multiple cameras (Cannon CR-2 AF, Cannon CR-2 Plus AF, and Topcon NW-400). The approval was based on the sensitivity for detecting mtmDR of 94.9%, specificity of 86.7%, and imageability of 98.6% [51]. Both EyeArt and RetmarkerDR, another system of AI for DR screening, have been available since before the current era of DL. Both AI systems adopt feature-based extraction of conventional ML that can also detect turnover of microaneurysms. For DR classification, in a study of both systems in the UK national DR screening programs [52], using arbitrated results of human grading as comparators, RetmarkerDR provided a sensitivity of 73% for any retinopathy, 85% for referable DR, and 97.9% for proliferative DR (PDR) whereas those of EyeArt were 94.7%, 93.8%, and 99.6% respectively. However, the false positive rate of both systems was relatively high at 50%.

The third AI system for DR screening approved by the U.S. FDA was AEYE-DS. Like iDX-DR, it was approved to be used with a specific retinal camera, Topcon NW-400, which is the same as iDX-DR. However, the approval was for screening using only one image per eye which could achieve a sensitivity of 93%, a specificity of 91.4%, and an imageability of 99.1% [53].

One of the first AI systems that used CNN (Inception V3) for DR screening is from Google Research by Gulshan et al. [54]. The algorithm

in this study was developed from more than 100,000 retinal images and was validated in the other two independent datasets of more than 10,000 images. This study was the first that showed the achievement of both sensitivity and specificity at 95% and AUC at 99% for detecting referable DR when validated in independent datasets.

This system was further validated in another retrospective dataset of more than 20,000 images of 7000 patients in a nationwide screening program of DR to detect moderate NPDR or worse [55]. This AI model achieved 97% sensitivity compared to 74% of human graders in this screening program whereas slightly lower specificity of 96%, compared to 98% of human graders, was found. When validated prospectively in two private hospitals in India with more than 3000 patients to detect moderate NPDR or worse, this model still achieved about 90% sensitivity with slightly higher than 90% of specificity, which were better than manual grading [56].

Another large-scale study of DL system for DR screening is from Singapore in which more than 70,000 images were used for development of the algorithm. The highlight of this study by Ting et al. [57] was the largest population with more than 100,000 images of independent datasets of various races for validation to date. This AI-based software, now called SELENA (VGG-19), was able to detect STDR with 100% sensitivity, 91% specificity, and 0.96 AUC.

A recent study on systematic review and meta-analysis on AI for DR screening from CRP found 2748 articles published between January 2017 and December 2022 [58]. The authors identified 21 studies for the analysis and found the pooled sensitivity of 0.880, pooled specificity of 0.912, pooled AUC of 0.9798, Cochrane Q index of 0.9388, and pooled diagnostic odds ratio of 206.80. The authors concluded that AI has a clear value for DR screening. Although they attempted to analyze on prospective studies, many retrospective studies were included.

Another recent review identified only 8 prospective studies on 7 AI systems for DR screening in the literature. The authors found, despite

high performance in diagnostic accuracy, inconsistency in setting standardization, varied study population, and higher rate of ungradable images due to lower threshold in determination of image ungradability, compared to human graders [59].

Optical Coherence Tomography (OCT)

Images from OCT are primarily applied for detecting and monitoring diabetic macular edema (DME) in retinal clinics, rather than screening for DR in primary care settings like CRP.

A recently published systematic review and meta-analysis identified 19 studies, with 41,005 subjects, on AI for detecting DME from OCT images. Although the pooled sensitivity and specificity were relatively high at 96.0% and 99.3% respectively, data set selection, sample size of training set, and the choice of OCT devices contributed to the heterogeneity of this analysis [60]. Most of DL models applied for OCT classification for DME in these studied were common CNN, such as VGG16, Inception v3, AlexNet, etc.

The identification of DME from CRP, however, could be problematic since in the real clinical practice DME was identified using images from OCT which is 3-dimensional. To overcome the limitation of 2-dimensional images of CRP, the presence of hard exudates in the macular area was used as proxy for detecting DME when grading CRP. There are occurrences for which the identification of DME from CRP and from OCT are not in concordance [61]. In addition, the prevalence of DME based on CRP and OCT is significantly different [62]. An interesting study by Varadarajan et al. used paired data of CRP and OCT to train a DL model to detect OCT-derived DME from grading on only CRP [63].

Developed from more than 6000 of the paired images, this model (CNN: Inception V3) was able to detect centered-involved DME (CI-DME) from CRP in the testing set of 1000 CRP images with 85% sensitivity and 80% specificity whereas three retinal specialists who graded the same CRP images using hard

exudates as proxy for CI-DME had, in average, similar sensitivity but about half of the specificity at 45%. In validation on another independent dataset of 990 CRP images, sensitivity and specificity of this model was lowered to be at 57% and 91% respectively, whereas sensitivity and specificity of graders who graded the same images were even lowered than the model at 55% and 79% respectively.

A recent study validated this model in multi-national datasets and found high performance of this model [64]. The analysis in a combined data set from Australia, India, and Thailand showed the model had 80% specificity and 81% sensitivity, compared with expert graders, who had 59% specificity and 70% sensitivity. Relative to human experts, the model had significantly higher specificity ($P=0.008$) and noninferior sensitivity ($P<0.001$).

This AI system showed a potential of AI to make a detection across multiple imaging modalities or across multiple kinds of labelled data (other examples: predict gender from CRP, blood pressure from CRP, etc. [65]), when trained with pairs of both imaging modalities or trained with pairs of both data labeling (label gender data with CRP, label blood pressure data with CRP, etc.). This concept is sometimes called label transfer.

Advanced Imaging

Some studies applied DL for detecting DR on more advanced retinal imaging, such as OCT angiography (OCTA) and ultra-widefield scanning laser ophthalmoscopy (UWF) retinal images. Ryu et al. applied ResNet 101 model for automated detection of DR and referable DR on OCTA images, using images from ultra-wide-field fluorescein angiography as the gold standard, achieving an overall accuracy of 91–98%, sensitivity of 86–97%, specificity of 94–99%, and an overall AUC of 0.919–0.976 on validation on external dataset [66]. The high performance of these diagnostic parameters was found in detecting from superficial capillary plexus layer, deep capillary plexus layer, and full-thickness retinal layer on OCTA.

Tang F et al. also achieved excellent performance of a CNN model for detecting referable DR (RDR) and vision-threatening DR (VTDR) on UWF from 4 external validation datasets [67]. The AUCs and accuracies for detecting both RDR and VTDR were >0.9 and $>80\%$, respectively, for the geographical external validation datasets.

The roles of the AI systems for detecting DR from these sophisticated imaging technologies may be limited due to their availability and practicality in primary care and limited resourced settings.

Screening or Classifying AMD

Screening Based on CRP

Most of the AI systems for screening and classification of AMD are developed using CNN and use CRP as inputs. SELENA, one of the first systems for screening AMD, was initially applied in patients with diabetes. Although use for screening AMD, the algorithm in SELENA was developed from a training dataset of more than 72,000 images of patients with diabetes in Singapore and Malaysia and a testing dataset of almost 36,000 images of patients in the same population. The output in this study was defined as referable AMD [57].

There are other studies on AI for screening AMD developed from CRP of Age-Related Eye Disease Study (AREDS) [68] which is a large randomized controlled trial compared between vitamin supplements and placebo for AMD development and progression. Since the AREDS collected CRP as films, they are digitized for applying AI. A study by Burlina et al. [69] (CNN: AlexNet) used a training and testing dataset of almost 54,000 and 14,000 images from AREDS while a study by Grassman et al. [70] (using various CNNs) used approximately 87,000 and 34,000 images. The former study used existing grades from AREDS for training while the latter required a trained ophthalmologist to label data for the training, both studies provided outputs as grades of AMD according to AREDS Classification.

Burlina et al. classified the outputs into two classes, referable and non-referable, whereas Grassman et al. classified the outputs into 9 steps of AREDS and 3 late AMD stages. Both studies achieved sensitivity and specificity of around 90% for the testing dataset. The study by Grassman et al. conducted validation in the external dataset of more than 5000 images and achieved a sensitivity and specificity of 82.2% and 97.1% for detecting intermediate AMD, and a sensitivity and specificity of 100% and 96.5% for late AMD. The system by Burlina et al. could later classify the 9-step AREDS and predict 5-year risk of progression to advanced AMD with acceptable error [71].

Another system of AI, DeepSeeNet, was also developed from AREDS datasets for automated assignment of individual patients to AMD risk categories based on the AREDS Simplified Severity Scale. A total of 59,302 color fundus images from 4,549 participants were extracted from the AREDS dataset for this study. The DeepSeeNet was a CNN model with Inception-v3 architecture and consisted of three CNN models for detecting drusen, retinal pigmentary abnormalities, and late AMD. DeepSeeNet performed better on patient-based, multi-class classification (accuracy=0.671; kappa=0.558) than retinal specialists (accuracy=0.599; kappa=0.467) with high AUCs in the detection of large drusen (0.94), pigmentary abnormalities (0.93) and late AMD (0.97), respectively.

The DeepSeeNet was later developed for detecting geographic atrophy (GA) [72], another manifestation of late AMD, with high accuracy for detecting GA (0.965), and central GA (0.966). This was on par with grading by retinal specialists (accuracy of 0.975 for GA and 0.990 for central GA).

In another system from the same team of investigators, the CNN DenseNet (version 152) was used to develop two AI models for detecting reticular pseudodrusen (RPD), a biomarker in early AMD which carries the risk of developing late AMD, from CRF and fundus autofluorescence (FAF) images. The model for CRF was

trained from labels transferred from FAF, which detect RPD better than CRF, and outperformed all other conventional models of CNN trained with data from CRF without the transfer of labels from FAF (AUC 0.832 whereas the AUC of rest was less than 0.80) [73].

Screening Based on OCT

There are other AI systems for classification of AMD from spectral-domain OCT (SD-OCT) images. Some systems use DL for classifying AMD directly from OCT images whereas some systems apply conventional ML for automated segmentation of fluid or biomarkers on OCT images as the first step then use DL classifiers for classification later. Studies by Kermany et al. [74] and Treder et al. [75] are examples of the former. Both used transfer learning from existing, open-sourced, pre-trained, ImageNet deep neural network (DNN) with 1000 output categories to train on OCT images for AMD.

Kermany et al. trained the ImageNet on four categories: choroidal neovascularization (CNV), DME, drusen, and normal. With the training dataset of more than 100,000 images (37,000 CNV, 11,000 DME, 8600 drusen, and 51,000 normal) and 1000 images for validation with equal distribution of the four categories, the system achieved AUC of 98% with accuracy, sensitivity, and specificity around 95%.

Lee et al. [76] linked data from electronic medical record (EMR) with OCT images to develop CNN system (VGG16) to classify AMD. With approximately 100,000 OCT images with linked data points of EMR were used, half normal and another half AMD, the system achieved AUC and accuracy around 90%.

Studies by Prahs et al. [77] and Hwang et al. [78] are examples of applying AI for classification of OCT images for decision-making. Prahs et al. trained CNN (GoogLeNet or Inception) with the inputs of AMD, DME, RVO, CSC, and the outputs of requiring or not requiring anti-VEGF treatment, labeled by treating clinicians. This study conducted validation on external dataset of more than 5500 images and achieved

sensitivity of 90%, specificity of 96%. Hwang et al. used three different cloud-based DL systems (VGG16, InceptionV3, and ResNet50) to train on OCT images of normal, dry AMD, active wet AMD, and inactive wet AMD. The authors found that the three CNN systems performed similarly for classification of the four categories of AMD with slightly lower performance on dry AMD. They also found a potential on prediction of longitudinal changes after treatment of wet AMD with 90% accuracy.

Another major AI system designed to perform automated segmentation of OCT images, and classify the segmented images for AMD and other retinal diseases is by De Fauw et al. [79]. For the segmentation, the authors applied a 3-dimensional U-net architecture for deep segmentation network to delineate OCT scanned images using more than 1000 manually segmented training images to form tissue segmentation maps of the OCT scans. Another classification network, a customized 29 CNN layers with 5 pooling layers, developed from 14,884 training tissue maps with confirmed diagnosis and referral decision, was applied to the segmented OCT maps.

This system then makes classification in different retinal diagnoses, for example, normal, CNV, macular hole, central serous chorioretinopathy, vitreomacular traction, etc., and referral suggestions as urgent, semi-urgent, routine, and observation. The authors found, on test performance of the model on an independent test set of 997 patients (252 urgent, 230 semi-urgent, 266 routine, 249 observation), that an AUC for urgent referral was 99.9 whereas the error rate of 3.4% was on par with those of retinal specialists. This performance was better than optometrists.

A research question on whether OCT images should be segmented before classification was of interest in another study. Kaothanthong et al. found that applying auto-segmentation to OCT images prior to classification of common macular diseases by DL models might improve their performance [80].

Screening or Classifying Glaucoma

The diagnosis of glaucoma may require identification of many co-existing parameters, such as family history, increased optic nerve head cupping or glaucomatous optic neuropathy (GON), characteristic loss of retinal nerve fiber layer or characteristic defects of visual field (VF). Some AI systems are developed to detect GON which appears on CRP.

The SELENA system by Ting et al. [57], apart from detecting DR and AMD, can also detect GON, this part of the algorithms was developed from CRP of more than 120,000 patients with diabetes. Li et al. [81] also developed another DL system (VGG) for detecting GON from more than 50,000 CRP graded by more than 20 ophthalmologists. The detection of referable glaucoma in these studies by Ting et al. and Li et al. may have a limitation from relying on only GON because even ophthalmologists might not have high agreement among them on grading GON [82].

Before this era of DL, many models of conventional ML were applied for detecting glaucoma from both time- and spectral-domain OCT images of optic nerve head (ONH) with acceptable performance [83]. Muhammad et al. showed that a hybrid DL model (AlexNet and Random Forest Classifier) was able to analyze single-scan of SS-OCT images to classify between normal and glaucoma suspects with 93% accuracy [84].

Christopher et al. applied Principal Component Analysis (PCA) approach of unsupervised ML to analyze retinal nerve fiber layer (RNFL) thickness maps from SS-OCT and showed that this approach could achieve the highest AUC of 0.95, compared to SD-OCT-based circumferential RNFL thickness measurements and visual field global indices for detection of glaucoma. Using stereoscopic CRP as standard for defining glaucoma, this model could also detect glaucoma progression with the highest AUC [85] compared to the other means.

Automated detection of VF progression, another important indicator of worsening glaucoma, had been studied in the Advanced Glaucoma Intervention Study (AGIS) using back-propagation neural network achieving an AUC of 0.92 [86]. In another study, Yousefi et al. introduced a new glaucoma VF index calculated by an unsupervised ML, Gaussian Mixture Model and Expectation (GEM), to detect VF progression. This model was trained on more than 2000 VFs and tested on a longitudinal cohort of 270 eyes followed every 6 months. This new AI-based index outperformed existing indices, such as Global or Region-wise, by finding the time to progression of 25% of the eyes in the longitudinal cohort at 3.5 years, compared with 4.5 years for the Region-wise and 5.2 years for the Global index [87].

Li et al. [88] compared (1) DL-CNN (VGG architecture) (2) conventional ML models (SVM, RF, k-NN) (3) rule-based algorithms (AGIS and enhanced glaucoma staging system [GSS2]) and (4) human experts for grading 300 VFs to differentiate between glaucoma and non-glaucoma patients. The CNN, developed from the same data set of 4000 VF images, achieved the accuracy of 0.876, while the specificity and sensitivity were 0.826 and 0.932 respectively. The accuracy of the three ML models was around 0.65, the human experts was around 0.6, and AGIS and GSS2 was around 0.5.

A study by Bowd et al. [89] combined structural data (OCT) and functional data (VF) to train conventional ML models (Bayesian Classifiers) and found their performances improved with the combined data for classification between glaucoma and non-glaucoma patients.

Medeiros et al. [90], on the other hand, applied the concept of label transfer in glaucoma. They trained a CNN (ResNet34) with more than 30,000 paired data of CRP images and RNFL thickness to predict the RNFL thickness for analyses on only CRP. In the test set of around 6200 CRP images, the model could predict the RNFL thickness with a strong correlation between predicted and observed RNFL thickness values (Pearson $r = 0.832$; $R^2 = 69.3\%$;

$P < 0.001$), with mean absolute error of the predictions of $7.39 \mu\text{m}$. The AUC of the classification of glaucoma and normal made by the CNN on grading only CRP images was 0.94.

In the following study by Jammal et al., the same team of authors validated their label transfer AI model with another set of 490 CRP images of 490 eyes of 370 subjects graded by two glaucoma specialists for the probability of glaucomatous optical neuropathy (GON) and estimates of cup-to-disc ratios (C/D). The AUC for classifying GON from CRP from their AI model, 0.529, was higher than the AUC of the glaucoma specialists, 0.411 [91].

Screening or Classifying Cataract

Anterior Segment Imaging

One of the first AI systems for grading cataract was by Gao et al. [92]. This system (Recursive Neural Networks combined with Support Vector Regression) was trained to grade severity of cataract from slit lamp photography in a decimal score, from 0.1–5.0, based on Wisconsin Grading System. The testing was performed in 5378 images, this system achieved a 70.7% exact agreement ratio (R0), a 88.4% decimal grading error of less than 0.5 (Re0.5), a 99.0% decimal grading error of less than 1.0 (Re1.0) when compared against clinical integral grading.

A much larger scale of an AI system for classification of cataract was by Wu et al. [93] in China. The authors trained a DL-CNN (ResNet) to classify photographs of anterior segment in three steps. The first was to classify according to modes of illumination: slit or diffuse, and methods of capture: mydriatic or non-mydriatic; the second was according to diagnosis: normal or cataract or postoperative cataract; the third was according to severity: mild or severe nuclear sclerosis, and visual axis involvement or not. The development dataset was from 37638 slit lamp photographs of 16,611 patients with 80% for training and 20% for testing. Validation of the models in the testing set found relatively good performances with confusion metrics more than 90% across the board. The performances on

images from slit or diffuse illumination, mydriatic or non-mydriatic were similar.

However, when the models were validated in four separate community hospitals in real-world situations, some of the confusion metrics of the AI dropped. Sensitivity for classifying normal at 71.3%, specificity for classifying cataract at 83.9%, sensitivity/specificity for classifying severe nuclear sclerosis at 73%/86% and sensitivity/specificity for classifying mild nuclear sclerosis at 86%/73% were found in the real world.

AI, in theory, would also be ideal as a tool for intraocular lens (IOL) power calculation. Sramka et al. [94] demonstrated that a conventional ML model, Support Vector Machine Regression, and multilayered neural network ensemble achieved significantly better performance for IOL power calculation compared to conventional methods of calculation without AI. Another study by Koprowski et al. [95] deployed ANN to evaluate corneal power after refractive surgery.

Interestingly, other than patient care, AI has also been used for ophthalmic training in cataract surgery. The aim is to use DL to identify manually pre-segmented phases of cataract surgery procedures in assisting to develop efficient and effective skill training tools. In a study by Yu et al. [96], from a dataset of videos of 100 cataract surgery procedures, the authors applied different AI algorithms, SVM, CNN, CNN-RNN, to identify phases in the cataract surgery videos. Modelling time series of labels of instruments in use provided the highest accuracy for the identification. In another study by Morita et al. [97], a DL-CNN model, Inception V3, was also trained to extract important phases, capsulorhexis and nuclear extraction, of cataract surgery videos with high correct response rate and low errors.

Screening Cataract Using CRP

Screening cataract, particularly visually significant cataract, in primary care settings using CRF may be more useful than anterior segment imaging, such as slit lamp. This is due to general availability of retinal cameras for DR screening.

Applying AI for automated detecting visually significant cataract can also be an adjunct to DR screening to differentiate cataract in the ungradable retinal images.

Tham YC et al. developed a ResNet-50 pretrained from ImageNet dataset to perform feature extractions on CRF, then an XGBoost classifier was used to classify visually significant cataract. Using more than 25,000 CRF from population-based studies, the model was able to achieve AUCs of 91.6%-96.5% on external validation. In a separate test set of 186 eyes, the system performed comparably, if not being slightly more superior than 4 ophthalmologists' evaluation (sensitivity of 93.3% vs 51.7-96.6% by ophthalmologists and specificity of 99.0% vs 90.7-97.9% by ophthalmologists) [98].

In another study, Elsayy A et al. developed a DL system to detect cataract from the AREDS dataset comprised 8,681 CRP from eyes with cataract and 8,833 CRP from eyes without cataract. The CNN achieved performance scores of 0.6683, 0.6691, 0.6686, 0.6686, which were higher than those of the combined ophthalmologists of 0.6025, 0.6119, 0.5988, and 0.5988 for accuracy, precision, recall, and AUC, respectively [99].

Screening for Retinopathy of Prematurity and Pediatric Cataract

Similar to DR screening, telemedicine has been deployed successfully for screening of retinopathy of prematurity (ROP) [100], the leading cause of childhood blindness. Telemedicine allows addressing two important barriers for ROP screening: interexaminer variability due to subjectivity and inadequate number of qualified trained examiners [101]. Further alleviation of these barriers is expected with application of AI.

There was a number of studies on conventional ML for determination of tortuosity and width of retinal vessels in Plus ROP in the era prior to DL [102-106]. All these models required manual annotations in implementation. The fully automated software, without additional manual annotations, for detection of ROP

came with application of CNN. DeepROP by Wang et al. [107] is one of the first systems of CNNs for ROP detection. The system (modified inception-BN nets pre-trained on ImageNet) was developed in China with the largest dataset of ROP to date (more than 20,000 images).

Another system, i-ROP-DL (the earlier version of i-ROP had not yet adopted DL [106]) developed in the U.S. with more than 5000 images, was demonstrated in a study by Redd et al. [108] to successfully classify severity of ROP into type 1 and type 2. Another study of i-ROP-DL by Brown et al. [109] compared the DL model (Inception V1 and U-Net) with eight international ROP experts on the output of Plus disease. Using the consensus of image-based diagnosis combined with ophthalmoscopy as the gold standard, the authors found the DL system agreed with this standard more than six of the eight experts. This i-ROP-DL system, in addition, was able to compute the probability of its prediction via a linear formula to be ROP severity scores.

In the subsequent study of i-ROP-DL, Taylor et al. [110] showed that these ROP severity scores could potentially be applied to monitor ROP disease progression. In monitoring more than 870 infants with ROP over time, the median scores of those who progressed to require treatment were significantly higher than those who did not progress at each of the postmenstrual age time points in the study.

There was another CNN system for ROP developed from a smaller dataset of 1500 images from Canada and England. This system could also detect Plus ROP with the accuracy which is comparable with other systems [111].

Apart from ROP, another major AI system was developed for detection of congenital cataract in China [112–114]. This system called Congenital Cataract-Cruiser, or CC-Cruiser, is a cloud-based AI applied to slit-lamp images and comprised three CNNs (AlexNet) for screening, assessment of severity, and treatment recommendation. It was compared with experienced pediatric ophthalmologists in five clinics in China in a prospective randomized controlled trial evaluating 350 patients younger than 14

years old. Applying grading of images by a panel of three experienced pediatric ophthalmologists as the gold standard, the accuracy of this AI system was significantly lower than the experts for detection of cataract (87% vs 99%) and recommendation of treatment (71% vs 97%). The duration for reaching results, however, was significantly shorter for the AI system (2.8 vs 8.5 min).

Other pediatric ophthalmology conditions that AI is studied for include detection of strabismus, refractive error, prediction of future high myopia and reading disability [115].

Screening Multiple Diseases on CRP

A recent study attempted to apply DL models to detect 8 common retinal diseases, DR, dry AMD, neovascular AMD, retinal vein occlusion, epiretinal membrane, retinal detachment, retinitis pigmentosa, and macular hole, based on CRP [116]. The authors found that VGG 19 model was able to classify between normal and abnormal CRP with 99.12% accuracy. For nine-class diagnosis, ResNet50 model with 128 nodes showed the highest accuracy of 87.42%. However, in a separate set of 180 CRP evaluated by extracting 20 images for each retinal disease, the CNN performed poorer than each of the 4 residents in ophthalmology (accuracy of 83.9% compared to 85–91.7%). However, the residents had reduced time spent for image interpretation by approximately 34 to 70% when they were notified of AI diagnosis results in advance.

Systems for Automated Segmentation

The difference between AI-based classification and AI-based automated segmentation may be clarified from the outputs. The classification systems generally provide outputs as stages of a disease classification, or binary outputs as referable or non-referable. Most of the automated segmentation systems, on the other hand, provide outputs as biomarkers or pathological characteristics, such as subretinal fluid, retinal

pigment epithelial detachment, or optic nerve head. The objectives of automated segmentation are assisting specialists in busy clinics or assisting researchers in research for time and cost saving. The objectives of the classification systems are generally for the benefits of patients, such as early detection or screening disease.

On bioengineering perspective, the objective of image segmentation task is to find the interesting area, such as an area of characteristics of disease in images, then separate image pixels in that area into background and foreground. The pixels of the fluid area on OCT images are separated as foreground from the rest which are background. Each of the image pixels is then classified into either a class of background or foreground. To train the model, a set of training images with labelling of the fluid area by experts, also called as ground truth, is provided. The ML model then learns to find the patterns of the pixels that separate them into foreground and background classes. To validate the model, the ML model then uses these patterns to categorize image pixels in the new images as either background or foreground.

The conventional ML method, such as Principle Component Analysis (PCA), is generally applied to reduce the sparseness of the extracted features on OCT images. Another method, such as SVM, is applied to classify each pixel into the two classes [117]. ANN, DNN, and CNN can also be applied for the classification. A common statistic that is used for assessment of accuracy of automated segmentation is Dice coefficient [118]. The hand-drawn of the segmented area by human experts is used as ground truth.

Retinal OCT images are excellent inputs for automated segmentation. Fang et al. [119] was able to segment the nine retinal layers in OCT images of patients with dry AMD. While macular fluid was generally the most common outputs for segmentation of retinal OCT images, photoreceptor cells loss was output for segmentation of retinal OCT in patients with retinitis pigmentosa and choroideremia [120]. Segmentation of choroidal thickness in patients with AMD from enhanced-depth OCT (EDI-OCT) images was also

performed successfully in another study [121]. Adaptive Optics Scanning Light Ophthalmoscopy images from patients with Stargadt's disease was also segmented by a Multi-Dimensional RNN for localization of cone cells [31].

While the automated segmentation of OCT images of macular region is important in retinal disease, the region of interest (ROI) for automated segmentation in glaucoma is at the ONH. Many studies used different segmentation algorithms for ONH segmentation from CRP. Almost all of the studies reported higher than 90% of accuracy for this task [83]. Hagiwara et al. [122] proposed a common workflow for a conventional computer-aided diagnosis of glaucoma using CRP images as the followings: (1) Image Input, (2) Pre-Processing, (3) Segmentation, (4) Feature extraction, (5) Feature Selection and Ranking, and (6) Classification. Some studies performed segmentation of peripapillary area which included RNFL, neural retina, retinal pigment epithelium, choroid, peripapillary sclera, and lamina cribrosa from OCT images [123, 124].

Some recent studies performed automated segmentation of retinal vasculatures on CRP [125–127], retinal angiographic images, such as invasive techniques of fluorescein angiography in DME [128] or indocyanine green angiography in PCV [129], and non-invasive technique of OCTA [130]. These papers focused on engineering techniques, rather than clinical validation.

Systems for Prediction of Natural History or Treatment Outcomes

Prediction aims to foresee the outcome of the disease, either with or without treatment, when a set of features, such as best corrected visual acuity (BCVA), or recurrence of macular edema at certain period, which are related to but not necessarily within the images themselves, is input. Longitudinal data or images are usually required for development of AI systems for prediction.

There has been an attempt to predict progression of DR at 6 months, 12 months, and 2

years for individual patients who did not have any ocular intervention during that period [131]. Using CRP of patients with DME who were randomized to be in the sham arm in the clinical trials of RISE and RIDE, a system of deep CNN (Inception V3) was developed to predict progression of DR [132]. The CRP in this AI model was based on the standard 7-field ETDRS photography, each field had its own CNN for prediction of 2-step or more worsening of DR according to Diabetic Retinopathy Severity Scale (DRSS), the results of each CNN was then combined using Random Forest. The prediction at 12 months of the system of the combined models achieved the best performance with AUC of 0.75 whereas the prediction of this system based on some peripheral fields of ETDRS surprisingly outperformed the prediction based on the posterior fields of ETDRS. However, the generalizability of this system might still be doubt due to (1) the limited availability of the ETDRS 7-field photography (2) the lack of validation in independent dataset and (3) the applicability of the models in patients without DME since this study included only CRP of patients with DME.

A recent study attempted to validate the application of the DR progression prediction model in the real-world screening workflow [133, 134]. The model was developed from a dataset of CRP of >575,000 eyes, almost 29,000 eyes of these had known DR progression outcomes. External validation of this model for predicting the risk of progression from no DR to DR provided an AUC of 0.71 when the prediction was combined with risk score whereas prediction from risk score alone provided the AUC of 0.62. This model was then applied to prioritize patients with diabetes for screening in next visits in the real world based on their risks. The model could predict the first half of patients who came for the next screening visits, based on the model prioritization order, as having moderate DR or worse correctly at 90.38% whereas the first half of patients who came in random order had moderate NPDR at 50%.

A study by Rohm et al. [135] reflected an attempt to develop AI systems to improve patient care and facilitate research from database of their own institute. Based on electronic health record collected in a form of Data Warehouse [136], the authors developed and compared different conventional ML models (AdaBoost.R2, Gradient Boosting, Random Forests, Extremely Randomized Trees, and Lasso) that predicted VA at 3 months and 12 months after 3 monthly injections of anti-VEGF agents for neovascular AMD. The prediction by the models at 3 months had mean absolute error between 5.5 to 9 letters, and root mean square between 7 to 10 letters. The prediction at 12 months was slightly worse.

Other studies on AI-based prediction of treatment outcomes were also centered on data from major clinical trials. For example, data from the Protocol T [137] of drcr.net trials were used for prediction of BCVA at 12 months after treatment with anti-VEGF injections for patients with DME [138]. Data from HARBOUR Study [139] were used for prediction of (1) BCVA at 12 months after treatment [140] (2) requirement of low or high injections of anti-VEGF [141] and (3) advanced AMD conversion for patients with AMD [142]. For retinal vein occlusion, data from CRYSTRAL Study [143] were used for prediction of BCVA and recurrence macular edema at 12 months [144].

Large Language Models and Data Synthesis

There is a recent interest in applying GANs to generate data, particularly images, in ophthalmology [145]. The main objective of GANs applications is to overcome limited access to large datasets for developing AI models. Ocular images synthesis by GANs include common modalities, such as CRP and OCT, or translating images between modalities, such as from anterior segment ultrasonography into anterior segment OCT [146], or CRP into fundus autofluorescence [147], with or without pathology

depended on purposes. However, there are challenges and limitations in validation of GANs outputs in both qualitative and quantitative perspectives.

The real recent interest in AI comes from the announcement of the large language models (LLMs) Generative Pretrained Transformer (GPT) [148] version 4, which is cooperated with chatbot and called ChatGPT-4 [149]. LLMs have a capability to leverage DL neural networks to learn complex associations between unstructured texts and use these learned patterns to produce outputs in response to custom queries in text format [150]. ChatGPT-4 has generated great excitement owing to its performance in cognitive tasks including medical problem-solving [151]. In ophthalmology, LLMs offer unique opportunities to revolutionize digital eye care, address clinical workflow inefficiencies, and enhance patient experiences across diverse global eye care landscapes [152]. Yet there are still tangible and ethical challenges related to data privacy, security, and the intricacies of applying LLMs into clinical eye care.

The Future of AI in Ophthalmology

We are in a transition phase, which changes rapidly, of AI in ophthalmology. There will be abundance of AI systems coming for more specific tasks in ophthalmology in the future. No more doubt will be cast on AI's performance in research, particularly in validation in the internal datasets, but many questions remain for clinical deployment. The robust performance of many systems may not be carried over into the real world. Since comparison between different AI systems is difficult, the choosing of an AI system to be used in clinical eye care should neither be easy. While researchers will be unravelling "black box", other issues, such as patient acceptability, data privacy, data protection, regulations, including medico-legal aspects [153] will still be challenged that AI systems, their developers, and users will be faced in the future.

References

1. Yu VL, Fagan LM, Wraith SM, et al. Antimicrobial selection by a computer. A blinded evaluation by infectious diseases experts. *JAMA*. 1979;242(12):1279–82.
2. Sinthanayothin C, Boyce JF, Cook HL, et al. Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images. *Br J Ophthalmol*. 1999;83(8):902–10. <https://doi.org/10.1136/bjo.83.8.902>.
3. He J, Baxter SL, Xu J, et al. The practical implementation of artificial intelligence technologies in medicine. *Nat Med*. 2019;25(1):30–6. <https://doi.org/10.1038/s41591-018-0307-0>.
4. Schmidt-Erfurth U, Sadeghipour A, Gerendas BS, et al. Artificial intelligence in retina. *Prog Retin Eye Res*. 2018;67:1–29. <https://doi.org/10.1016/j.preteyeres.2018.07.004>.
5. Han J, Kamber M, Pei J. Data mining: concepts and techniques. *Data Min Concepts Tech*; 2012.
6. Alpaydin E. Introduction to machine learning. MIT Press; 2020.
7. Indyk P, Motwani R. Approximate nearest neighbors: towards removing the curse of dimensionality. In: *Proceedings of the 30th annual ACM symposium on theory of computing*; 1998.
8. Bengio Y, LeCun Y. Scaling learning algorithms towards AI. In: *Proceedings of 2007*; 2007.
9. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436–44. <https://doi.org/10.1038/nature14539>.
10. Hosseini M-P, Lu S, Kamaraj K, et al. Deep learning architectures. In: *Pedrycz W, Chen S-M, editors. Deep learning: concepts and architectures*. Cham: Springer International Publishing; 2020. p. 1–24. https://doi.org/10.1007/978-3-030-31756-0_1.
11. Lewis DD. Naïve (Bayes) at forty: the independence assumption in information retrieval. In: *Nédellec C, Rouveirol C, editors. Proceedings of machine learning: ECML-98*. Berlin: Heidelberg; 1998.
12. Baştanlar Y, Özuysal M. Introduction to machine learning. In: *Yousef M, Allmer J, editors. miRNomics: MicroRNA biology and computational analysis*. Totowa, NJ: Humana Press; 2014. p. 105–128. https://doi.org/10.1007/978-1-62703-748-8_7.
13. Graupe D. Principles of artificial neural networks. World Scientific; 2013.
14. Weinberger KQ, Saul LK. Distance metric learning for large margin nearest neighbor classification. *J Mach Learn Res*. 2009;10(2):207–44.
15. Hastie T, Rosset S, Zhu J, et al. Multi-class adaboost. *Statistics and its Interface*. 2009;2(3):349–60.
16. Schapire RE. Explaining AdaBoost. In: *Schölkopf B, Luo Z, Vovk V, editors. Empirical inference: festschrift in honor of Vladimir N. Vapnik*. Berlin

- Heidelberg: Springer; 2013. p. 37–52. https://doi.org/10.1007/978-3-642-41136-6_5.
17. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM*. 2017;60(6):84–90. <https://doi.org/10.1145/3065386>.
 18. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997;9(8):1735–80. <https://doi.org/10.1162/neco.1997.9.8.1735>.
 19. Lu W, Tong Y, Yu Y, et al. Applications of artificial intelligence in ophthalmology: general overview. *J Ophthalmol*. 2018;2018:5278196. <https://doi.org/10.1155/2018/5278196>.
 20. Andrearczyk V, Whelan PF. Using filter banks in convolutional neural networks for texture classification. *Pattern Recogn Lett*. 2016;84:63–9. <https://doi.org/10.1016/j.patrec.2016.08.016>.
 21. Robinson R. Convolutional neural networks—basics; 2017. <https://mlnotebook.github.io/post/CNN1/>. Accessed 20 Mar 2020.
 22. Saxe AM, Koh PW, Chen Z, et al. On random weights and unsupervised feature learning. In: *Proceedings of the 28th international conference on machine learning*. Bellevue, Washington, USA; 2011.
 23. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: *3rd international conference on learning representations ICLR 2015—conference track proceedings*; 2015.
 24. Szegedy C, Wei L, Yangqing J, et al. Going deeper with convolutions. In: *Proceedings of IEEE conference on computer vision and pattern recognition*; 2015.
 25. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. . In: *Proceedings of IEEE conference on computer vision and pattern recognition*; 2016.
 26. Tan M, Le Q. EfficientNet: rethinking model scaling for convolutional neural networks. In: *Proceedings of 36th international conference on machine learning*; 2019.
 27. Zagoruyko S, Komodakis N. Wide residual networks 2016. *Br Mach Vis Conf*. 2016;87(12):1–87.
 28. Huang Y, Cheng Y, Bapna A, et al. Gpipe: efficient training of giant neural networks using pipeline parallelism. *Adv Neural Inf Process Syst*. 2019;32.
 29. Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. *IEEE Trans Pattern Anal Mach Intell*. 2013;35(8):1798–828. <https://doi.org/10.1109/tpami.2013.50>.
 30. Christopher M, Belghith A, Bowd C, et al. Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs. *Sci Rep*. 2018;8(1):16685. <https://doi.org/10.1038/s41598-018-35044-9>.
 31. Davidson B, Kalitzeos A, Carroll J, et al. Automatic cone photoreceptor localisation in healthy and stargardt afflicted retinas using deep learning. *Sci Rep*. 2018;8(1):7911. <https://doi.org/10.1038/s41598-018-26350-3>.
 32. Diaz GI, Fokoue-Nkoutche A, Nannicini G, et al. An effective algorithm for hyperparameter optimization of neural networks. *IBM J Res Dev*. 2017;61(4/5):9:1–9:11. <https://doi.org/10.1147/JRD.2017.2709578>.
 33. Models for image classification with weights trained on ImageNet. <https://keras.io/applications/#models-for-image-classification-with-weights-trained-on-imagenet>. Accessed 27 Mar 2020.
 34. Ting DSW, Lee AY, Wong TY. An ophthalmologist's guide to deciphering studies in artificial intelligence. *Ophthalmology*. 2019;126(11):1475–9. <https://doi.org/10.1016/j.ophtha.2019.09.014>.
 35. Raman R, Srinivasan S, Virmani S, et al. Fundus photograph-based deep learning algorithms in detecting diabetic retinopathy. *Eye (Lond)*. 2019;33(1):97–109. <https://doi.org/10.1038/s41433-018-0269-y>.
 36. Dy JG, Brodley CE. Feature selection for supervised learning. *J Mach Learn Res*. 2004;5:845–89.
 37. Sutton RS, Barto AG. Reinforcement learning: an introduction. Cambridge, MA: MIT Press; 2018.
 38. Matsuo Y, LeCun Y, Sahani M, et al. Deep learning, reinforcement learning, and world models. *Neural Netw*. 2022;152:267–75. <https://doi.org/10.1016/j.neunet.2022.03.037>.
 39. Ray PP. ChatGPT: a comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet Things Cyber Phys Syst*. 2023;3:121–54. <https://doi.org/10.1016/j.iotcps.2023.04.003>.
 40. Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng*. 2010;22(10):1345–59. <https://doi.org/10.1109/TKDE.2009.191>.
 41. Castro FM, Marín-Jiménez MJ, Guil N, et al. End-to-end incremental learning. In: *Proceedings of the European conference on computer vision (ECCV)*; 2018.
 42. Luo Y, Yin L, Bai W, et al. An appraisal of incremental learning methods. *Entropy (Basel)*. 2020;22(11):1190. <https://doi.org/10.3390/e22111190>.
 43. Wang Y, Yao Q, Kwok JT, et al. Generalizing from a few examples: a survey on few-shot learning. *ACM Comput Surv*. 2020;53(3):1–34. <https://doi.org/10.1145/3386252>.
 44. Tao X, Hong X, Chang X, et al. Few-shot class-incremental learning. In: *Proceedings of IEEE conference on computer vision and pattern recognition*; 2020.
 45. Chen Y, Guo X, Pan Y, et al. Dynamic feature splicing for few-shot rare disease diagnosis. *Med Image Anal*. 2023;90: 102959. <https://doi.org/10.1016/j.media.2023.102959>.
 46. French RM. Catastrophic forgetting in connectionist networks. *Trends Cogn Sci*. 1999;3(4):128–35. [https://doi.org/10.1016/s1364-6613\(99\)01294-2](https://doi.org/10.1016/s1364-6613(99)01294-2).

47. Wilson JMG, Jungner G. Principles and practice of screening for disease. Geneva: World Health Organization; 1968.
48. Dobrow MJ, Hagens V, Chafe R, et al. Consolidated principles for screening based on a systematic review and consensus process. *Can Med Assoc J.* 2018;190(14):e422–9. <https://doi.org/10.1503/cmaj.171154>.
49. Abràmoff MD, Lou Y, Erginay A, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Invest Ophthalmol Vis Sci.* 2016;57(13):5200–6. <https://doi.org/10.1167/iovs.16-19964>.
50. Abràmoff MD, Lavin PT, Birch M, et al. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit Med.* 2018;1:39. <https://doi.org/10.1038/s41746-018-0040-6>.
51. EyeArt 510(k) Summary (K200667). https://www.accessdata.fda.gov/cdrh_docs/pdf20/K200667.pdf. Accessed 31 Dec 2023.
52. Tufail A, Rudisill C, Egan C, et al. Automated diabetic retinopathy image assessment software: diagnostic accuracy and cost-effectiveness compared with human graders. *Ophthalmology.* 2017;124(3):343–51. <https://doi.org/10.1016/j.ophtha.2016.11.014>.
53. AEYE-DS Device 510(K) Summary (K221183). https://www.accessdata.fda.gov/cdrh_docs/pdf22/K221183.pdf. Accessed 31 Dec 2023.
54. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA.* 2016;316(22):2402–10. <https://doi.org/10.1001/jama.2016.17216>.
55. Raumviboonsuk P, Krause J, Chotcomwongse P, et al. Deep learning versus human graders for classifying diabetic retinopathy severity in a nationwide screening program. *NPJ Digit Med.* 2019;2:25. <https://doi.org/10.1038/s41746-019-0099-8>.
56. Gulshan V, Rajan RP, Widner K, et al. Performance of a deep-learning algorithm vs manual grading for detecting diabetic retinopathy in india. *JAMA Ophthalmol.* 2019;137(9):987–93. <https://doi.org/10.1001/jamaophthalmol.2019.2004>.
57. Ting DSW, Cheung CY, Lim G, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA.* 2017;318(22):2211–23. <https://doi.org/10.1001/jama.2017.18152>.
58. Wang Z, Li Z, Li K, et al. Performance of artificial intelligence in diabetic retinopathy screening: a systematic review and meta-analysis of prospective studies (systematic review). *Front Endocrinol (Lausanne).* 2023;14:1197783. <https://doi.org/10.3389/fendo.2023.1197783>.
59. Rajesh AE, Davidson OQ, Lee CS, et al. Artificial intelligence and diabetic retinopathy: AI framework, prospective studies, head-to-head validation, and cost-effectiveness. *Diabetes Care.* 2023;46(10):1728–39. <https://doi.org/10.2337/dci23-0032>.
60. Li HY, Wang DX, Dong L, et al. Deep learning algorithms for detection of diabetic macular edema in OCT images: a systematic review and meta-analysis. *Eur J Ophthalmol.* 2023;33(1):278–90. <https://doi.org/10.1177/11206721221094786>.
61. Mackenzie S, Schmermer C, Charnley A, et al. SDOCT imaging to identify macular pathology in patients diagnosed with diabetic maculopathy by a digital photographic retinal screening programme. *PLoS One.* 2011;6(5): e14811. <https://doi.org/10.1371/journal.pone.0014811>.
62. Wang YT, Tadarati M, Wolfson Y, et al. Comparison of prevalence of diabetic macular edema based on monocular fundus photography vs optical coherence tomography. *JAMA Ophthalmol.* 2016;134(2):222–8. <https://doi.org/10.1001/jamaophthalmol.2015.5332>.
63. Varadarajan AV, Bavishi P, Ruamviboonsuk P, et al. Predicting optical coherence tomography-derived diabetic macular edema grades from fundus photographs using deep learning. *Nat Commun.* 2020;11(1):130. <https://doi.org/10.1038/s41467-019-13922-8>.
64. Liu X, Ali TK, Singh P, et al. Deep learning to detect OCT-derived diabetic macular edema from color retinal photographs: a multicenter validation study. *Ophthalmol Retina.* 2022;6(5):398–410. <https://doi.org/10.1016/j.oret.2021.12.021>.
65. Poplin R, Varadarajan AV, Blumer K, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng.* 2018;2(3):158–64. <https://doi.org/10.1038/s41551-018-0195-0>.
66. Ryu G, Lee K, Park D, et al. A deep learning model for identifying diabetic retinopathy using optical coherence tomography angiography. *Sci Rep.* 2021;11(1):23024. <https://doi.org/10.1038/s41598-021-02479-6>.
67. Tang F, Luenam P, Ran AR, et al. Detection of diabetic retinopathy from ultra-widefield scanning laser ophthalmoscope images: a multicenter deep learning analysis. *Ophthalmol Retina.* 2021;5(11):1097–106. <https://doi.org/10.1016/j.oret.2021.01.013>.
68. Chew EY, Clemons TE, SanGiovanni JP, et al. Lutein + zeaxanthin and omega-3 fatty acids for age-related macular degeneration: the Age-Related Eye Disease Study 2 (AREDS2) randomized clinical trial. *JAMA.* 2013;309(19):2005–15. <https://doi.org/10.1001/jama.2013.4997>.
69. Burlina PM, Joshi N, Pekala M, et al. Automated grading of age-related macular degeneration

- from color fundus images using deep convolutional neural networks. *JAMA Ophthalmol.* 2017;135(11):1170–6. <https://doi.org/10.1001/jamaophthalmol.2017.3782>.
70. Grassmann F, Mengelkamp J, Brandl C, et al. A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography. *Ophthalmology.* 2018;125(9):1410–20. <https://doi.org/10.1016/j.ophtha.2018.02.037>.
 71. Burlina PM, Joshi N, Pacheco KD, et al. Use of deep learning for detailed severity characterization and estimation of 5-year risk among patients with age-related macular degeneration. *JAMA Ophthalmol.* 2018;136(12):1359–66. <https://doi.org/10.1001/jamaophthalmol.2018.4118>.
 72. Keenan TD, Dharssi S, Peng Y, et al. A deep learning approach for automated detection of geographic atrophy from color fundus photographs. *Ophthalmology.* 2019;126(11):1533–40. <https://doi.org/10.1016/j.ophtha.2019.06.005>.
 73. Keenan TDL, Chen Q, Peng Y, et al. Deep learning automated detection of reticular pseudodrusen from fundus autofluorescence images or color fundus photographs in AREDS2. *Ophthalmology.* 2020;127(12):1674–87. <https://doi.org/10.1016/j.ophtha.2020.05.036>.
 74. Kermany DS, Goldbaum M, Cai W, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell.* 2018;172(5):1122–1131.e9. <https://doi.org/10.1016/j.cell.2018.02.010>.
 75. Treder M, Lauer mann JL, Eter N. Automated detection of exudative age-related macular degeneration in spectral domain optical coherence tomography using deep learning. *Graefes Arch Clin Exp Ophthalmol.* 2018;256(2):259–65. <https://doi.org/10.1007/s00417-017-3850-3>.
 76. Lee CS, Baughman DM, Lee AY. Deep learning is effective for the classification of OCT images of normal versus age-related macular degeneration. *Ophthalmol Retina.* 2017;1(4):322–7. <https://doi.org/10.1016/j.oret.2016.12.009>.
 77. Prahs P, Radeck V, Mayer C, et al. OCT-based deep learning algorithm for the evaluation of treatment indication with anti-vascular endothelial growth factor medications. *Graefes Arch Clin Exp Ophthalmol.* 2018;256(1):91–8. <https://doi.org/10.1007/s00417-017-3839-y>.
 78. Hwang DK, Hsu CC, Chang KJ, et al. Artificial intelligence-based decision-making for age-related macular degeneration. *Theranostics.* 2019;9(1):232–45. <https://doi.org/10.7150/thno.28447>.
 79. De Fauw J, Ledsam JR, Romera-Paredes B, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med.* 2018;24(9):1342–50. <https://doi.org/10.1038/s41591-018-0107-6>.
 80. Kaothanthong N, Limwattanayingyong J, Silpa-Archa S, et al. The classification of common macular diseases using deep learning on optical coherence tomography images with and without prior automated segmentation. *Diagnostics (Basel).* 2023;13(2):189. <https://doi.org/10.3390/diagnostics13020189>.
 81. Li Z, He Y, Keel S, et al. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. *Ophthalmology.* 2018;125(8):1199–206. <https://doi.org/10.1016/j.ophtha.2018.01.023>.
 82. Abrams LS, Scott IU, Spaeth GL, et al. Agreement among optometrists, ophthalmologists, and residents in evaluating the optic disc for glaucoma. *Ophthalmology.* 1994;101(10):1662–7. [https://doi.org/10.1016/s0161-6420\(94\)31118-3](https://doi.org/10.1016/s0161-6420(94)31118-3).
 83. Zheng C, Johnson TV, Garg A, et al. Artificial intelligence in glaucoma. *Curr Opin Ophthalmol.* 2019;30(2):97–103. <https://doi.org/10.1097/icu.0000000000000552>.
 84. Muhammad H, Fuchs TJ, De Cuir N, et al. Hybrid deep learning on single wide-field optical coherence tomography scans accurately classifies glaucoma suspects. *J Glaucoma.* 2017;26(12):1086–94. <https://doi.org/10.1097/jgg.0000000000000765>.
 85. Christopher M, Belghith A, Weinreb RN, et al. Retinal nerve fiber layer features identified by unsupervised machine learning on optical coherence tomography scans predict glaucoma progression. *Invest Ophthalmol Vis Sci.* 2018;59(7):2748–56. <https://doi.org/10.1167/iovs.17-23387>.
 86. Lin A, Hoffman D, Gaasterland DE, et al. Neural networks to identify glaucomatous visual field progression. *Am J Ophthalmol.* 2003;135(1):49–54. [https://doi.org/10.1016/s0002-9394\(02\)01836-6](https://doi.org/10.1016/s0002-9394(02)01836-6).
 87. Yousefi S, Kiwaki T, Zheng Y, et al. Detection of longitudinal visual field progression in glaucoma using machine learning. *Am J Ophthalmol.* 2018;193:71–9. <https://doi.org/10.1016/j.ajo.2018.06.007>.
 88. Li F, Wang Z, Qu G, et al. Automatic differentiation of Glaucoma visual field from non-glaucoma visual field using deep convolutional neural network. *BMC Med Imaging.* 2018;18(1):35. <https://doi.org/10.1186/s12880-018-0273-5>.
 89. Bowd C, Hao J, Tavares IM, et al. Bayesian machine learning classifiers for combining structural and functional measurements to classify healthy and glaucomatous eyes. *Invest Ophthalmol Vis Sci.* 2008;49(3):945–53. <https://doi.org/10.1167/iovs.07-1083>.
 90. Medeiros FA, Jammal AA, Thompson AC. From machine to machine: an OCT-trained deep learning algorithm for objective quantification of glaucomatous damage in fundus photographs. *Ophthalmology.* 2019;126(4):513–21. <https://doi.org/10.1016/j.ophtha.2018.12.033>.

91. Jammal AA, Thompson AC, Mariottoni EB, et al. Human versus machine: comparing a deep learning algorithm to human gradings for detecting glaucoma on fundus photographs. *Am J Ophthalmol*. 2020;211:123–31. <https://doi.org/10.1016/j.ajo.2019.11.006>.
92. Gao X, Lin S, Wong TY. Automatic feature learning to grade nuclear cataracts based on deep learning. *IEEE Trans Biomed Eng*. 2015;62(11):2693–701. <https://doi.org/10.1109/tbme.2015.2444389>.
93. Wu X, Huang Y, Liu Z, et al. Universal artificial intelligence platform for collaborative management of cataracts. *Br J Ophthalmol*. 2019;103(11):1553–60. <https://doi.org/10.1136/bjophthalmol-2019-314729>.
94. Sramka M, Slovak M, Tuckova J, et al. Improving clinical refractive results of cataract surgery by machine learning. *PeerJ*. 2019;7: e7202. <https://doi.org/10.7717/peerj.7202>.
95. Koprowski R, Lanza M, Irregolare C. Corneal power evaluation after myopic corneal refractive surgery using artificial neural networks. *Biomed Eng Online*. 2016;15(1):121. <https://doi.org/10.1186/s12938-016-0243-5>.
96. Yu F, Silva Croso G, Kim TS, et al. Assessment of automated identification of phases in videos of cataract surgery using machine learning and deep learning techniques. *JAMA Netw Open*. 2019;2(4): e191860. <https://doi.org/10.1001/jamanetworkopen.2019.1860>.
97. Morita S, Tabuchi H, Masumoto H, et al. Real-time extraction of important surgical phases in cataract surgery videos. *Sci Rep*. 2019;9(1):16590. <https://doi.org/10.1038/s41598-019-53091-8>.
98. Tham Y-C, Goh JHL, Anees A, et al. Detecting visually significant cataract using retinal photograph-based deep learning. *Nature Aging*. 2022;2(3):264–71. <https://doi.org/10.1038/s43587-022-00171-6>.
99. Elsayy A, Keenan TDL, Chen Q, et al. Detecting cataract from color fundus photographs using explainable deep learning. *Invest Ophthalmol Vis Sci*. 2022;63(7):217-F0064.
100. Daniel E, Quinn GE, Hildebrand PL, et al. Validated system for centralized grading of retinopathy of prematurity: telemedicine approaches to evaluating acute-phase retinopathy of prematurity (e-ROP) STUDY. *JAMA Ophthalmol*. 2015;133(6):675–82. <https://doi.org/10.1001/jamaophthalmol.2015.0460>.
101. Ting DSW, Pasquale LR, Peng L, et al. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol*. 2019;103(2):167–75. <https://doi.org/10.1136/bjophthalmol-2018-313173>.
102. Capowski JJ, Kylstra JA, Freedman SF. A numeric index based on spatial frequency for the tortuosity of retinal vessels and its application to plus disease in retinopathy of prematurity. *Retina*. 1995;15(6):490–500. <https://doi.org/10.1097/00006982-199515060-00006>.
103. Heneghan C, Flynn J, O’Keefe M, et al. Characterization of changes in blood vessel width and tortuosity in retinopathy of prematurity using image analysis. *Med Image Anal*. 2002;6(4):407–29. [https://doi.org/10.1016/s1361-8415\(02\)00058-0](https://doi.org/10.1016/s1361-8415(02)00058-0).
104. Swanson C, Cocker KD, Parker KH, et al. Semiautomated computer analysis of vessel growth in preterm infants without and with ROP. *Br J Ophthalmol*. 2003;87(12):1474–7. <https://doi.org/10.1136/bjo.87.12.1474>.
105. Gelman R, Martinez-Perez ME, Vanderveen DK, et al. Diagnosis of plus disease in retinopathy of prematurity using retinal Image multiScale analysis. *Invest Ophthalmol Vis Sci*. 2005;46(12):4734–8. <https://doi.org/10.1167/iovs.05-0646>.
106. Ataer-Cansizoglu E, Bolon-Canedo V, Campbell JP, et al. Computer-based image analysis for plus disease diagnosis in retinopathy of prematurity: performance of the “i-ROP” system and image features associated with expert diagnosis. *Transl Vis Sci Technol*. 2015;4(6):5. <https://doi.org/10.1167/tvst.4.6.5>.
107. Wang J, Ju R, Chen Y, et al. Automated retinopathy of prematurity screening using deep neural networks. *EBioMedicine*. 2018;35:361–8. <https://doi.org/10.1016/j.ebiom.2018.08.033>.
108. Redd TK, Campbell JP, Brown JM, et al. Evaluation of a deep learning image assessment system for detecting severe retinopathy of prematurity. *Br J Ophthalmol*. 2018. <https://doi.org/10.1136/bjophthalmol-2018-313156>.
109. Brown JM, Campbell JP, Beers A, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. *JAMA Ophthalmol*. 2018;136(7):803–10. <https://doi.org/10.1001/jamaophthalmol.2018.1934>.
110. Taylor S, Brown JM, Gupta K, et al. Monitoring disease progression with a quantitative severity scale for retinopathy of prematurity using deep learning. *JAMA Ophthalmol*. 2019;137(9):1022–8. <https://doi.org/10.1001/jamaophthalmol.2019.2433>.
111. Worrall DE, Wilson CM, Brostow GJ. Automated retinopathy of prematurity case detection with convolutional neural networks. In: Carneiro G, Mateus D, Peter L, Bradley A, Tavares JMRS, Belagiannis V, et al, editors. *Proceedings of deep learning and data labeling for medical applications*. Cham; 2016.
112. Long E, Lin H, Liu Z, et al. An artificial intelligence platform for the multihospital collaborative management of congenital cataracts. *Nat Biomed Eng*. 2017;1:0024. <https://doi.org/10.1038/s41551-016-0024>.
113. Lin H, Li R, Liu Z, et al. Diagnostic efficacy and therapeutic decision-making capacity of an artificial intelligence platform for childhood cataracts in eye clinics: a multicentre randomized controlled trial. *EClinicalMedicine*. 2019;9:52–9. <https://doi.org/10.1016/j.eclinm.2019.03.001>.

114. Liu X, Jiang J, Zhang K, et al. Localization and diagnosis framework for pediatric cataracts based on slit-lamp images using deep features of a convolutional neural network. *PLoS One*. 2017;12(3): e0168606. <https://doi.org/10.1371/journal.pone.0168606>.
115. Reid JE, Eaton E. Artificial intelligence for pediatric ophthalmology. *Curr Opin Ophthalmol*. 2019;30(5):337–46. <https://doi.org/10.1097/icu.0000000000000593>.
116. Kim KM, Heo TY, Kim A, et al. Development of a fundus image-based deep learning diagnostic tool for various retinal diseases. *J Pers Med*. 2021;11(5):321. <https://doi.org/10.3390/jpm11050321>.
117. Alsaih K, Lemaitre G, Rastgoo M, et al. Machine learning techniques for diabetic macular edema (DME) classification on SD-OCT images. *Biomed Eng Online*. 2017;16(1):68. <https://doi.org/10.1186/s12938-017-0352-9>.
118. Zou KH, Warfield SK, Bharatha A, et al. Statistical validation of image segmentation quality based on a spatial overlap index. *Acad Radiol*. 2004;11(2):178–89. [https://doi.org/10.1016/s1076-6332\(03\)00671-8](https://doi.org/10.1016/s1076-6332(03)00671-8).
119. Fang L, Cunefare D, Wang C, et al. Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search. *Biomed Opt Express*. 2017;8(5):2732–44. <https://doi.org/10.1364/boe.8.002732>.
120. Camino A, Wang Z, Wang J, et al. Deep learning for the segmentation of preserved photoreceptors on en face optical coherence tomography in two inherited retinal diseases. *Biomed Opt Express*. 2018;9(7):3092–105. <https://doi.org/10.1364/boe.9.003092>.
121. Chen M, Wang J, Oguz I, et al. (2017) Automated segmentation of the choroid in EDI-OCT images with retinal pathology using convolution neural networks. *Fetal Infant Ophthalmic Med Image Anal*. 2017;10554:177–84. https://doi.org/10.1007/978-3-319-67561-9_20.
122. Hagiwara Y, Koh JEW, Tan JH, et al. Computer-aided diagnosis of glaucoma using fundus images: a review. *Comput Methods Programs Biomed*. 2018;165:1–12. <https://doi.org/10.1016/j.cmpb.2018.07.012>.
123. Devalla SK, Chin KS, Mari JM, et al. A deep learning approach to digitally stain optical coherence tomography images of the optic nerve head. *Invest Ophthalmol Vis Sci*. 2018;59(1):63–74. <https://doi.org/10.1167/iovs.17-22617>.
124. Devalla SK, Renukanand PK, Sreedhar BK, et al. DRUNET: a dilated-residual U-Net deep learning network to segment optic nerve head tissues in optical coherence tomography images. *Biomed Opt Express*. 2018;9(7):3244–65. <https://doi.org/10.1364/boe.9.003244>.
125. Lahiri A, Roy AG, Sheet D, et al. Deep neural ensemble for retinal vessel segmentation in fundus images towards achieving label-free angiography. *Annu Int Conf IEEE Eng Med Biol Soc*. 2016;2016:1340–3. <https://doi.org/10.1109/embc.2016.7590955>.
126. Wang D, Haytham A, Pottenburgh J, et al. Hard attention net for automatic retinal vessel segmentation. *IEEE J Biomed Health Inform*. 2020;24(12):3384–96. <https://doi.org/10.1109/jbhi.2020.3002985>.
127. Rodrigues EO, Conci A, Liatsis P. ELEMENT: multi-modal retinal vessel segmentation based on a coupled region growing and machine learning approach. *IEEE J Biomed Health Inform*. 2020;24(12):3507–19. <https://doi.org/10.1109/jbhi.2020.2999257>.
128. Jin K, Pan X, You K, et al. Automatic detection of non-perfusion areas in diabetic macular edema from fundus fluorescein angiography for decision making using deep learning. *Sci Rep*. 2020;10(1):15138. <https://doi.org/10.1038/s41598-020-71622-6>.
129. Loo J, Teo KYC, Vyas CH, et al. Joint multimodal deep learning-based automatic segmentation of indocyanine green angiography and OCT images for assessment of polypoidal choroidal vasculopathy biomarkers. *Ophthalmol Sci*. 2023;3(3): 100292. <https://doi.org/10.1016/j.xops.2023.100292>.
130. Ma F, Li S, Wang S, et al. Deep-learning segmentation method for optical coherence tomography angiography in ophthalmology. *J Biophotonics*. 2023;e202300321. <https://doi.org/10.1002/jbio.202300321>.
131. Arcadu F, Benmansour F, Maunz A, et al. Deep learning algorithm predicts diabetic retinopathy progression in individual patients. *NPJ Digit Med*. 2019;2:92. <https://doi.org/10.1038/s41746-019-0172-3>.
132. Nguyen QD, Brown DM, Marcus DM, et al. Ranibizumab for diabetic macular edema: results from 2 phase III randomized trials: RISE and RIDE. *Ophthalmology*. 2012;119(4):789–801. <https://doi.org/10.1016/j.ophtha.2011.12.039>.
133. Bora A, Balasubramanian S, Babenko B, et al. Predicting the risk of developing diabetic retinopathy using deep learning. *Lancet Digit Health*. 2021;3(1):e10–9. [https://doi.org/10.1016/s2589-7500\(20\)30250-8](https://doi.org/10.1016/s2589-7500(20)30250-8).
134. Bora A, Tiwari R, Bavishi P, et al. Risk stratification for diabetic retinopathy screening order using deep learning: a multicenter prospective study. *Transl Vis Sci Technol*. 2023;12(12):11. <https://doi.org/10.1167/tvst.12.12.11>.
135. Rohm M, Tresp V, Müller M, et al. Predicting visual acuity by using machine learning in patients treated for neovascular age-related macular degeneration. *Ophthalmology*. 2018;125(7):1028–36. <https://doi.org/10.1016/j.ophtha.2017.12.034>.

136. Kortüm KU, Müller M, Kern C, et al. Using electronic health records to build an ophthalmologic data warehouse and visualize patients' data. *Am J Ophthalmol*. 2017;178:84–93. <https://doi.org/10.1016/j.ajo.2017.03.026>.
137. Wells JA, Glassman AR, Ayala AR, et al. Aflibercept, bevacizumab, or ranibizumab for diabetic macular edema. *N Engl J Med*. 2015;372(13):1193–203. <https://doi.org/10.1056/NEJMoA1414264>.
138. Gerendas BS, Bogunovic H, Sadeghipour A, et al. Computational image analysis for prognosis determination in DME. *Vision Res*. 2017;139:204–10. <https://doi.org/10.1016/j.visres.2017.03.008>.
139. Suner IJ, Yau L, Lai P. HARBOR study: one-year results of efficacy and safety of 2.0 mg versus 0.5 mg ranibizumab in patients with subfoveal choroidal neovascularization secondary to age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 2012;53(14):3677–3677.
140. Schmidt-Erfurth U, Bogunovic H, Sadeghipour A, et al. Machine learning to analyze the prognostic value of current imaging biomarkers in neovascular age-related macular degeneration. *Ophthalmol Retina*. 2018;2(1):24–30. <https://doi.org/10.1016/j.oret.2017.03.015>.
141. Bogunovic H, Waldstein SM, Schlegl T, et al. Prediction of anti-VEGF treatment requirements in neovascular AMD using a machine learning approach. *Invest Ophthalmol Vis Sci*. 2017;58(7):3240–8. <https://doi.org/10.1167/iovs.16-21053>.
142. Schmidt-Erfurth U, Waldstein SM, Klimscha S, et al. Prediction of individual disease conversion in early AMD using artificial intelligence. *Invest Ophthalmol Vis Sci*. 2018;59(8):3199–208. <https://doi.org/10.1167/iovs.18-24106>.
143. Larsen M, Waldstein SM, Boscia F, et al. Individualized ranibizumab regimen driven by stabilization criteria for central retinal vein occlusion: twelve-month results of the CRYSTAL study. *Ophthalmology*. 2016;123(5):1101–11. <https://doi.org/10.1016/j.ophtha.2016.01.011>.
144. Vogl WD, Waldstein SM, Gerendas BS, et al. Analyzing and predicting visual acuity outcomes of anti-VEGF therapy by a longitudinal mixed effects model of imaging and clinical data. *Invest Ophthalmol Vis Sci*. 2017;58(10):4173–81. <https://doi.org/10.1167/iovs.17-21878>.
145. Wang Z, Lim G, Ng WY, et al. Generative adversarial networks in ophthalmology: what are these and how can they be used? *Curr Opin Ophthalmol*. 2021;32(5):459–67. <https://doi.org/10.1097/icu.0000000000000794>.
146. Wanichwecharungruang B, Kaothanthong N, Pattanapongpaiboon W, et al. Deep learning for anterior segment optical coherence tomography to predict the presence of plateau iris. *Transl Vis Sci Technol*. 2021;10(1):7. <https://doi.org/10.1167/tvst.10.1.7>.
147. Song F, Zhang W, Zheng Y, et al. A deep learning model for generating fundus autofluorescence images from color fundus photography. *Adv Ophthalmol Pract Res*. 2023;3(4):192–8. <https://doi.org/10.1016/j.aopr.2023.11.001>.
148. Radford A, Narasimhan K, Salimans T, et al. Improving language understanding by generative pre-training; 2018.
149. GPT-4 Technical Report. 2023. <https://cdn.openai.com/papers/gpt-4.pdf>. Accessed 31 December 2023.
150. Tan TF, Thirunavukarasu AJ, Campbell JP, et al. Generative artificial intelligence through ChatGPT and other large language models in ophthalmology: clinical applications and challenges. *Ophthalmol Sci*. 2023;3(4): 100394. <https://doi.org/10.1016/j.xops.2023.100394>.
151. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLOS Digit Health*. 2023;2(2): e0000198. <https://doi.org/10.1371/journal.pdig.0000198>.
152. Betzler BK, Chen H, Cheng CY, et al. Large language models and their impact in ophthalmology. *Lancet Digit Health*. 2023;5(12):e917–24. [https://doi.org/10.1016/s2589-7500\(23\)00201-7](https://doi.org/10.1016/s2589-7500(23)00201-7).
153. Grzybowski A, Brona P, Lim G, et al. Artificial intelligence for diabetic retinopathy screening: a review. *Eye (Lond)*. 2020;34(3):451–60. <https://doi.org/10.1038/s41433-019-0566-0>.

Autonomous Artificial Intelligence Safety and Trust

4

Michael D. Abramoff

Abstract

Artificial Intelligence, or Augmented Intelligence (AI), is the term for systems capable of making decisions of high cognitive complexity. In health-care, Autonomous AI systems are those AI systems that make clinical decisions without human oversight, and where the autonomous AI creator assumes medical liability (American Medical Association (AMA) Board of Trustees Policy Summary in Augmented intelligence in healthcare, 2019). Autonomous AI systems are thus different from Assistive AI systems, which help a clinician make better diagnostic or management decisions, and where the liability for the medical decision remains with the clinician (Horton et al. in *Telemed J E Health* 26(4):495–543, 2020).

Keywords

Autonomous artificial intelligence · AI safety · Augmented intelligence · Clinical decision-making · Healthcare AI · Autonomous systems · Assistive AI

Introduction

Artificial Intelligence, or Augmented Intelligence (AI), is the term for systems capable of making decisions of high cognitive complexity. In health-care, Autonomous AI systems are those AI systems that make clinical decisions without human oversight, and where the autonomous AI creator assumes medical liability [1]. Autonomous AI systems are thus different from Assistive AI systems, which help a clinician make better diagnostic or management decisions, and where the liability for the medical decision remains with the clinician [2].

As an example, diagnostic autonomous AI systems for the point of care diagnosis of diabetic retinopathy and diabetic macular edema, provide a direct diagnostic recommendation. In doing so, they perform a cognitive highly complex task that was previously only performed by ophthalmologists and optometrists—representing 0.02% of all Americans—after extensive, specialized, training.

M. D. Abramoff (✉)

The Robert C Watzke Professor of Ophthalmology and Visual Sciences, University of Iowa, Iowa City, IA, USA

e-mail: michael-abramoff@uiowa.edu

Retina Service, University of Iowa, Iowa City, IA, USA

Digital Diagnostics, Coralville, IA, USA

Rigorously validated medical diagnostic autonomous AI systems hold great promise for improving patient access, as well as increasing accuracy, and lowering cost, while enabling specialist physicians to provide the greatest value by managing and treating those patients whose outcomes can be improved [3, 4]. There are significant ethical and legal concerns around the introduction of autonomous AI into healthcare [5, 6]. Ensuring the public's and healthcare systems' trust that autonomous AI provides these benefits, requires creators to negotiate multiple ethical and practical challenges.

In 2018, after extensive work on the ethics and accountability implications of a computer making a medical diagnosis, culminating in a clinical trial for autonomous AI comparing it to clinical outcome, the first autonomous point of care diabetic retinopathy exam was de novo authorized by the US FDA [7]. This milestone marked that a pathway for other autonomous AI systems to be introduced safely exists. In 2020, this same autonomous AI system became part of the American Diabetes Association's Standard of Medical Care in Diabetes [8]. In the same year, the US National Committee for Quality Assurance updated its quality measurements to support the use of autonomous AI to perform the diabetic eye exam, at an equivalent level as an ophthalmologist or optometrist [9]. Also in 2020, US Medicare national insurance formally adopted autonomous AI for the diabetic eye exam in its reimbursement system, thus offering a pathway for other autonomous AI to be reimbursed.

No prior processes existed for determining the safety, efficacy and equity of such autonomous AI systems that might serve as guidance. This chapter goes through the why and how of ethical and accountability principles for autonomous AI, explains how they were practically addressed and how they form the foundation for a series of requirements for autonomous AI, that were used in design, in the de novo FDA clearance process, and in ongoing implementation. It leans heavily on the concepts addressed in "Lessons Learned About Autonomous AI:

Finding a Safe, Efficacious, and Ethical Path Through the Development Process" [10].

Ultimately, successful introduction of autonomous AI into the healthcare system is dependent on trust in this technology, and it is thus vital that everyone involved with autonomous AI helps build trust [11].

Autonomous AI for the Diabetic Retinopathy Exam

The patient and societal benefits of early detection of diabetic retinopathy are well established [12–15]. This diabetic eye exam is typically performed as an indirect dilated retinal examination, using slitlamp biomicroscopy and indirect binocular ophthalmoscopy, as well as macular optical coherence tomography, by an ophthalmologist, retina specialist, or optometrist. Thus, many practice guidelines referred to this method as the standard of care [16]. In the 1990s, evidence showed that telemedicine for diabetic retinopathy, where retinal images are taken and then evaluated remotely, has at least a similar patient safety profile as the dilated retinal examination [17, 18]. While improving access, traditional telemedicine does not allow instantaneous, point of care diagnosis, nor has the safety, efficacy and equity of its process been established in a scientifically valid, hypothesis testing, manner. Thus, trust has generally been limited. There is a delay of typically days between when the patient is imaged, and the availability of an ophthalmologist to read images and diagnose the patient. In many cases, low image quality requires calling the patient back for another examination. Thus, safer, more trustworthy solutions were desirable.

Patient adherence to any form of regular diabetic eye examination is generally low around the world, a state of affairs primarily caused by lack of access and convenience. Recent research shows that in the US, adherence to a regular documented diabetic eye exam remains as low as 15.3%, even though all practice guidelines and standards recommend regular diabetic

eye exams [16, 19, 20]. Compared to testing at a remote laboratory, point of care diagnosis increases access, which has been shown for point of care A1C testing [21, 22], and has even been shown to improve clinical outcome [23].

Historically, clinicians are trained during medical school, residency and possible fellowship, and then their medical competency continues to be overseen by Medical Boards. However, in practice, clinicians are rarely validated on their safety, efficacy and equity for a specific diagnostic process, against a valid standard. Furthermore, consistency of clinicians performing the diabetic eye exam is limited, and the only scientifically valid studies comparing clinicians to the most rigorous patient outcome-based prognostic reference standard, show that clinician accuracy, expressed as sensitivity, does not exceed 50% [24, 25]. Additionally, documentation of a diabetic eye exam in the chart, once performed, is a human driven process with many points of potential failure, so that too often, the fact that a diabetic eye exam was performed cannot be verified from the patient's chart. Operational coding systems for physician activities are typically not fine grained enough to establish that a diabetic eye exam was performed [26–28]. Instead, typically only the fact that the physician spent time interacting with the patient is documented, if it is documented at all.

Together, these are major challenges with process integrity for the diabetic eye exam (both traditional, in the chair, as well as through telemedicine). In other words, the traceability of what exactly happens to the patient when the need for a diabetic eye exam is established, is limited, because the accuracy of the diagnostic process is unknown, and because the documentary evidence that a completely determined by fallible clinician documentation [29]. To increase trust, program integrity for the diabetic eye exam needs to be as high as possible.

The use of artificial intelligence for medical diagnostic purposes, started in the 1960s with Mycin, an AI that helped physicians prescribe antibiotics [30]. This continued through the 1980s with algorithms such as the perceptron [31] and multilayer neural network learning

using backpropagation [32]. Safe performance of these AI systems was limited, mainly because of a lack of high quality maximally objective input data. Input consisted of physicians interpreting the patient's symptoms and signs and then typing them in, a process with an inherently low signal to noise ratio. Instead much of the foundational methodology for modern autonomous AI has been developed over the past decades. The recent introduction of affordable digital retinal cameras with high-quality complementary metal-oxide semiconductor (CMOS) image sensors was a pivotal moment. These CMOS image sensors make it possible to acquire images of high fidelity and consistency, such as retinal images of people with diabetes, and thus provide highly objective input data for AI algorithms. The resulting higher performance and potentially higher safety, is necessary for increasing trust in autonomous AI.

In summary, diagnostic autonomous AI systems, such as IDx-DR, provide a direct diagnostic recommendation for the point of care diagnosis of diabetic retinopathy and diabetic macular edema. Autonomous AI allows the diagnosis for diabetic retinopathy and diabetic macular edema to be performed in real time, with the goals of higher adherence, improved access, higher cost-effectiveness [33], increased accuracy and diagnosability [34], and increased program integrity.

Can We Trust It? Concerns About Autonomous AI

The idea of 'a computer making a diagnosis', generates concerns in both physicians and patients, as is to be anticipated with any new technology. We will address the most common ones: undesired racial, sex or ethnic bias, validation, what type of trial and what to compare the AI against, what safety threshold to use, data usage, and liability in this section [10].

A recent study of an AI showed that using medical cost as a proxy for patients' overall health needs led to inappropriate racial bias in allocating healthcare resources, as black

patients were incorrectly deemed to have lower risk compared to white patients because their incurred costs were lower for a given health risk status [35]. Another study showed a consistent decrease in performance for underrepresented sex categories in the AI's machine learning training data when a minimum balance was not fulfilled [36].

To demonstrate AI systems' safety, scientifically valid studies that are replicable are essential. The design of these studies is a concern. Some, like healthcare pundit Eric Topol, have proposed that Randomized Clinical Trials (RCT) are the gold standard for diagnostic AI [37], though there is clear evidence that other study designs can work as well or better [38]. For example, RCTs for diagnostic AI require an arm where patient outcome, including need for intervention, is only decided based on the AI output—without possible override by a clinician. If the AI is not perfect, this requires potentially withholding effective treatment for a treatable condition where we know how to improve outcome. Most Institutional Review Boards see this as unethical, as the benefit to other patients and society may not outweigh the harm to the patient who has been diagnosed by AI [39, 40]. As diagnostic AIs are typically designed for conditions where effective interventions are available, RCTs are ethically unsound. In other words, while a null hypothesis of “no effect” works well in most interventional trials, a null hypothesis of “not informative” is not appropriate for validation of diagnostic AI [41].

Replicability is another big problem affecting trust in safety, and without preregistration, AI performance tends to be overestimated, and successful study replication becomes less likely. In fact when comparing trials with and without pre-registration, trial effect sizes are larger when they lack preregistration [34, 42].

Often, to demonstrate AI systems' safety, their output is compared against the diagnosis by clinicians or groups of clinicians, called a “reference standard” [43]. This approach assumes that the AI system safety is highest when it most closely matches clinicians' diagnosis. There are several problems with this

approach: (a) showing that an AI system is safer than clinicians is impossible, as a discrepancy between clinician and AI system will by definition be attributed to an error by the AI rather than the clinician; (b) clinicians differ greatly on typical diagnostic tasks, in many cases in 30% of cases or more [44], and it is thus impossible to determine which clinician is right and which is wrong; (c) if compared to clinical outcome, such as with a prognostic standard, the ultimate determinant of clinical relevance, clinicians perform poorly, for example achieving only 33 and 34% in the only two studies comparing clinicians diagnosing diabetic retinopathy as determined by the Wisconsin Reading Center ETDRS standard [24, 25]. Thus, estimates of safety are greatly affected by the choice of reference standard.

In 2007, Fenton and colleagues first demonstrated the importance of rigorous validation of AI in the actual workflow setting, rather than in a modeled laboratory setting [45]. In this pivotal study, the outcomes of women undergoing breast cancer screening by a radiologist assisted by a previously FDA approved assistive AI system, were compared to women who underwent breast cancer screening by a radiologist without such an assistive AI. The assistive AI had, in 2000, been approved by FDA, on the basis of a study that showed that, when used in isolation, the assistive AI had high diagnostic accuracy compared to radiologists. When this assistive AI system was tested in a study design that reflected actual usage, assisting a radiologist who makes the final clinical decision, the study showed worse outcomes for the women who underwent breast cancer screening with AI assistance.

For continued trust and acceptance, the autonomous AI's impact on clinic workflow needs to be optimized, and optimal clinic workflow of the patient around the autonomous AI is an important aspect. One metric is so-called *population diagnosability*. Population diagnosability is defined as

$$PS = \frac{n_n + n_p}{n_n + n_p + n_x + n_i}$$

where

n_p = number of subjects that received
a positive diagnostic result

n_n = number of subjects that received
a negative diagnostic result

n_x = number of subjects that were excluded
for any reason from study completion

n_i = number of subject that received an
insufficient input quality result

For example, if in a completed validation study, the total number of subjects recruited for the study, n , is 1000, n_x , 200 subjects were excluded from analysis for various reasons, and the autonomous AI gave an invalid input quality, n_i , for 100 subjects, $PS=0.7$. Obviously, a higher population diagnosability improves efficiency and especially workflow, as there is less risk of an individual patient not being diagnosed by the autonomous AI, and needing to fall back on the in person exam.

Safety, for a diagnostic process, including for an autonomous AI system, is typically measured as sensitivity, as appropriate for binary outcomes, more so than metrics such as Receiver Operator Characteristics (ROC) analysis [46]. In many cases, the purpose of the diagnostic process is to find (true) cases in a population. While a high sensitivity maximizes the efficiency of such a process, a population level analysis will show that adherence with the diagnostic process plays as important a role. For example, if a diagnostic process has high sensitivity of 90%, but only 10% of a population undergoes the diagnostic process, the cases in the remaining 90% of the population will not be found, lowering the “population achieved sensitivity”. In order to account for this diagnostic access bias [47], the population achieved sensitivity (PAS) or “compliance corrected sensitivity” can be calculated as follows:

$$PAS = \frac{s_c c p_c}{c p_c + (1 - c) \widehat{p}_{nc}}$$

where

- s_c =sensitivity (as determined in compliant population)
- c =compliance
- p_c =measured prevalence in the compliant population
- $\widehat{p}_{nc}$ = estimated prevalence, in the noncompliant population

If we assume that $p_c \cong \widehat{p}_{nc}$ i.e. the prevalence of the disease is the same in the non-compliant as in the compliant part of the population, we can use the simplified estimate $s_c c$.

$$PAS \cong s_c c$$

This estimated PAS would then form an upper bound, as in most cases, prevalence in the non-compliant subpopulation is higher than in the compliant part.

For example, if compliance c , with the diabetic eye exam, is 15% [19], and the minimal acceptable sensitivity is 85% [34], the population achieved sensitivity (PAS)=0.13. In other words, only 13% of all cases in the population will be identified correctly with this diagnostic system.

These metrics show the importance of including workflow and patient-experience in general in the study design, as focus on sensitivity in the compliant population may not give the optimal population benefit, and increasing compliance with the autonomous AI may have more effect on its safety than increasing sensitivity.

The development of any AI requires vast amounts of clinical data. There are many statutes and regulations covering patient derived data, such as HIPAA and HITECH [48]. Ultimately, whether patient derived data belongs to the patient, the physician, the hospital system, or even who paid for acquisition, has not been fully litigated, and as such can easily lead to concerns and controversy. For example, in one case patient data for training AI through was obtained through an agreement with a health system [49]. While agreements were in

place, patients and physicians were not aware of this data usage, leading to confusion, so that the Department of Health and Human services became involved. In another example, a class action lawsuit alleging failure to adequately de-identify patient data for AI training was initiated against an academic health system.

Autonomous AI is typically designed and validated for a narrow diagnostic task, and so-called incidental findings, that potentially could have been discovered by a general exam, are not flagged. For our example, which diagnoses diabetic retinopathy and diabetic macular edema, other diagnoses such as glaucoma, or macular degeneration will not be made, as the autonomous AI is neither designed nor validated for anything but DR. While there is widespread evidence for the effectiveness and cost-effectiveness of early detection of diabetic retinopathy [15], this is currently not yet the case for glaucoma [50], macular degeneration [51] and many other eye diseases. Therefore, clinical trials for a specific AI will typically not be designed or powered to analyze diagnostic accuracy on other retinal abnormalities or other eye abnormalities in people with diabetes. However, it is important to note that little evidence exists on how accurately clinicians can diagnose these incidental findings: their performance has not been evaluated in formal studies. In fact, such studies may be logistically impossible and challenging to power and therefore are unlikely to be performed because of the enormous number of subjects required. For example, at a prevalence of 5 per million, evaluating the diagnostic performance of either ophthalmologists or autonomous AI for choroidal melanoma diagnosis would require a clinical trial with approximately 40 million subjects [52].

In the past, liability for medical errors contributed to by AI has generally been attributed to the physician using it [53]. While this is acceptable practice for assistive AI, as ultimately the medical decision is made the physician using the AI, this may not be the case for autonomous AI, where after all, the medical decisions is made by the AI, without input by that physician. Nevertheless many AI creators have publicly

refused to take liability for their AI products, as shown by the ongoing liability debate concerning autonomous driving [54].

Trust in Autonomous AI is essential, and definitely, lack of trust has negatively affected other medical innovations. In the early 2000s, gene therapy effectively went through a moratorium on research funding, including closure of research institutions, after several young people died in poorly planned and executed gene therapy studies [55]. Only in 2017, almost two decades later, did FDA approve the first ever gene therapy, for the RPE65 variant of Leber's Congenital Amaurosis [56]. In sort, trust in autonomous AI needs to be earned.

Building Trust: An Ethical Foundation for Autonomous AI Requirements

Successful introduction of autonomous AI into the healthcare system is dependent on trust, and it is essential that everyone involved with autonomous AI helps build such trust [6, 11]. This is best done through an ethical foundation. Previously, Char, Abramoff et al., derived a foundation for autonomous AI in healthcare from bioethical principles as well as accountability principles, implemented in Abramoff et al. [10]. We ensured alignment with classical bioethics principles, such as Beauchamp and Childress [57]. The following table is modified from our paper [10]. Copyright retained by the Authors (Table 4.1).

The following sections will illustrate these various requirements for autonomous AI as they are relevant. Addressing such requirements during the design, validation and implementation is expected to increase trust from all healthcare stakeholders, including patients, physicians, regulators, and payers.

Autonomous AI System Design Requirements

Considerations for Autonomous AI design can have unexpected and profound ethical implications.

Table 4.1 Deriving requirements from bioethical principles

Autonomous AI requirement	Relevant bioethical principles [57]
Improve patient outcome as shown either by direct evidence or linked clinical literature, and aligned with evidence based clinical standards of care/practice patterns from quality of care organizations, professional medical societies and patient organizations, while accounting for safety, efficacy and equity	Nonmaleficence, Beneficence, and JEquity
Design so its operations are maximally reducible to characteristics aligned with scientific knowledge of human clinician cognition	Beneficence Non-maleficence
Maximize traceability of patient derived data, and commensurate data stewardship, accountability, and authorization; including by adherence to accepted standards	Accountability and Respect for Autonomy
Validate rigorously for safety, efficacy and equity, using preregistered clinical studies, by comparing against clinical outcome, or outcome surrogates in the case of chronic diseases, in the intended clinical workflow and usage, as shown by either direct or linked evidence	Nonmaleficence, and Equity
Assume liability commensurate with indications for use and autonomy	Accountability, and Equity

When autonomous AI is designed so its operations are maximally reducible to characteristics aligned with scientific knowledge of human clinician cognition, it ethically aligns with non-maleficence. The use of patho-physiologically sound priors, such as biomarkers, and exploiting higher order coherence in the input data, not only helps gain trust from regulators, physicians and patients, but also improves safety and equity, thus aligning with non-maleficence and justice. Machine learning algorithms that mimic closely how clinicians diagnose have been shown to be more robust to small perturbations in the input, to show less catastrophic failure, and to be less likely to exhibit inappropriate racial and other bias [58, 59]. Black box or gray box algorithm designs makes such bias harder to mitigate and detect, while the speed and scalability can multiply the effect of inappropriate bias faster than traditional enforcement efforts can react. Take as an example, the design of autonomous AI for diabetic retinopathy: for over 150 years, clinicians have evaluated a patient’s retina for the different indicators of diabetic retinopathy such as hemorrhages, microaneurysms, and neovascularization [60, 61]. These are all indicators or biomarkers that are invariant with regards to race, ethnicity, sex, and age. Using multiple, statistically dependent detectors for such biomarkers [62, 63], each optimized using machine learning algorithms mitigates the problems sketched.

Designing the autonomous AI so that it improves patient outcome as shown either by direct evidence or linked clinical literature, ensures patient benefit, and ethically aligns with non-maleficence and justice. Similarly, alignment with evidence based clinical standards of care/practice patterns from quality of care organizations, professional medical societies and patient organizations, while accounting for safety, efficacy and equity, again aligns ethically with non-maleficence and justice. ‘Glamour AI’ or AI which is technologically of high interest, but does not improve outcome, is thus avoided by this requirement.

For our example, when the goal is to create a real time point of care autonomous retinal exam for diabetic retinopathy, non-maleficence was maximized, by ensuring that diabetic macular was diagnosed accurately by the autonomous AI system—as is standard of care [8, 16]. Somewhat surprisingly, this design requirement is different from that of most AI designs, which focus only on the classic ischemic variant of diabetic retinopathy, and ignore macular edema. At most, they only look for exudates, which in many cases are not proxies for the presence or absence of diabetic macular edema [34, 64].

Evidence of the relationship to outcome is essential to confirm that the autonomous AI improves outcome. The level of disease diagnosed by an autonomous AI can align directly

with the risk of poor outcome: an IDx-DR negative output (less than level ETDRS 35 and no macular edema) implies a risk of 1.7% or less of proliferative retinopathy in 3 years, and a risk of 2.4% or less of DME in 1 year, while an IDx-DR positive output (ETDRS level 35 or higher or center-involved or clinically significant macular edema) confers a risk of at least 18% of proliferative retinopathy in 3 years, as well as a risk of 17.7% of developing DME in 1 year—if left untreated [34, 65]. Tying the autonomous AI output to patient relevant clinical outcome.

Autonomous AI System Validation Requirements

Just as is the case for its design, the validation of an autonomous AI can also have unexpected and profound ethical implications. To maximize trust, autonomous AI should be validated rigorously for safety, efficacy and equity, ideally using preregistered clinical studies, by comparing the AI against clinical outcome, or prognostic standards, in the case of chronic diseases, within the intended clinical workflow and usage, as shown by either direct or linked evidence, non-maleficence and equity are maximized. The meaning of each these terms is explained here.

In accordance with the bioethical principle of non-maleficence [6, 10], AI validation studies should test hypotheses of safety, efficacy and equity. Scientific validity of such studies is higher the more replicable they are, and common reporting standards [66], CONSORT-AI [67], preregistration of study and analysis protocols [42, 68], and validated relationship to patient outcome [10] are important factors to enhance replicability, and consistent with US Federal regulations. While standards have been set for preregistration, and especially Good Clinical Practice (GCP) [7], depending on the risk of harm to the patient, these can be burdensome and can require substantial resources from AI creators, and thus may not always be achievable for AI validation studies, in accordance with risk of patient harm. Broadly, preregistration includes public registration of the

in- and exclusion criteria, the entire protocol, and statistical on a site such as clinicaltrials.gov, a hypothesis-testing design with predefined endpoints, a predefined method for statistical analysis, predefined inclusion and exclusion criteria, a predefined sampling protocol, a plan for handling of the trial data by an independent Contract Research Organization or third party, and prohibition of access by the researchers to the subject level results before finalizing the statistical analysis.

Also in accordance with non-maleficence, is to validate the autonomous AI against what matters to the patient. This is most likely clinical outcome: the clinical event relevant to the patient, or, an event of which the patient is aware and wants to avoid, including death, loss of vision, visual field loss, the need for ventilatory support, or other events causing a reduction in quality of the patient's life [69].

While for acute diseases or interventions, clinical outcome may be immediate and easy to measure—such as visual acuity in the case of myopia or central retinal artery occlusion. However, for the many chronic diseases that autonomous AI has particular potential for, such as diabetic retinopathy, glaucoma or macular degeneration, clinical outcome may take years to manifest. As a consequence, *surrogate* endpoints have been developed [70], to reduce the cost and shorten the duration of trials, especially in the drug approval process. One type of surrogate endpoint is a phenotype: a laboratory measurement, or a physical sign, used as a substitute for a clinical outcome that measures directly how a patient feels, functions or survives. Another type is a prognostic standard, which are (combinations of) biomarkers that are associated with prognosis. Changes induced by a therapy on a prognostic standard, or other surrogate endpoints should reflect changes in clinical outcome. Examples of surrogate endpoints are suppression of ventricular arrhythmias, or reduction in cholesterol level, in cardiovascular trials, Prognostic standards include positive pathology in a biopsy, or progression on mammograms, in breast cancer, or evidence of mitosis in skin cancer.

Within diabetic retinopathy the Early scale and the [DRCR.net](#) macular edema scale are prognostic standards [65, 71].

Obviously, while requiring less time and fewer resources than true outcome in the case of chronic disease, prognostic standard based reference standards still requires considerable effort. This is an important reason why clinician derived reference standards, rather than outcome or surrogate based reference standards, are so widely used in AI. This practice is so widespread that it is almost a standard. A widely cited meta-analysis of the quality of evidence of AI accuracy, while mentioning the potential of AI to improve outcome, takes as a given the comparison to clinician derived ground truth. Validation against (surrogate) clinical outcome is not even considered [72].

In addition to unknown validity against outcome, reproducibility—different clinicians evaluating the same patient differently in 30–50% of cases—; repeatability—the same clinician evaluating the same patient differently in 20–30% of cases—; and temporal drift—clinicians systematically evaluating the same hypothetical patient differently over generations of clinicians—; are other major issues to be addressed [24, 25, 73]. As the evidence for a given treatment based on a given evaluation may have been derived decades ago, the latter, temporal drift, is a form of bias that is especially pernicious, and difficult to correct for. When prognostic standards, or outcome is unfeasible, optimal correction for

reproducibility and repeatability requires strict evaluation protocols and independent verification where possible.

Summarizing the above, clinical outcome or prognostic standards should be used for primary endpoints, are preferred, provided their validity against true clinical outcome has already been rigorously established. Only if these are not available, can non outcome associated endpoints be used, but that should then be clear in the autonomous AI labelling [69] (Table 4.2).

Validation in the envisioned context, environment and workflow, in “locked” form, so that its performance is known and persists in real world clinical practice, is desirable for aligning with non-maleficence. For example, the negative effect of AI on outcome, that was shown in the Fenton study, could thus have been prevented [45].

Updates to the autonomous AI should be evaluated on their potential effect on patient risk of harm. Changing the font size of the autonomous AI user interface will has substantially lower risk than adding training data for potential performance improvement. A standardized process that links the level of evidence to potential risk of patient harm for each type of system update, allows analysis of such “continuous learning”. At the highest risk of patient harm, a fully locked autonomous AI, once validated, cannot have its training data automatically updated based on new inputs, as then the safety, efficacy and equity are not known. In a narrow

Table 4.2 Reference standards © 2020 Abramoff

• Level I Reference Standard: A reference standard that either is a clinical outcome, a prognostic standard or other surrogate outcome. If the surrogate outcome is derived from an independent reading center, validation against outcome is required, as is published evidence of temporal drift, reproducibility, and repeatability metrics
• Level II Reference Standard: A reference standard established by an independent reading center with published temporal drift, reproducibility, and repeatability metrics. A B level reference standard has not been validated to correlate to a clinical outcome
• Level III Reference Standard: A reference standard created from the same modality as used by the AI, by adjudicating or voting of multiple independent expert readers, documented to be masked, with published reproducibility and repeatability metrics. A C level reference standard has not been established an independent reading center, and has not been validated to correlate with a clinical outcome
• Level IV Reference Standard: All other reference standards, created by single readers or non-expert readers, without an established protocol. A D level reference standard has not been derived from an independent reading center, has not been validated to correlate with a clinical outcome, and there are no published reproducibility and repeatability metrics

sense, “Continuous learning” AI systems, where learning is used to describe incorporating new training data as it processes new inputs during deployment, will require new validation at the same level as the original, in chronic disease.

Ensuring that the validation is applicable for the clinical use case, requires workflow analysis, and where possible, mimicking workflow during the trial. In our example, this required the trial to be performed in primary care clinics, in the standard diabetes management workflow, without modifications to the clinic environment, and with operators recruited from existing staff without prior experience or training.

Autonomous AI System Implementation Requirements

For enhancing trust, ethical alignment with patient autonomy from a patient-derived perspective is important, as we have derived previously [10]. Focusing on maximizing traceability of patient derived data, including commensurate data stewardship, accountability, and authorization, as well as adherence to accepted standards. Obviously, this applies to data usage during the design phase, but is of primary importance during deployment. Alignment with patient autonomy may make so-called “data-plays”, where the purpose of the autonomous AI is maximizing the value of resellable patient derived data, rather than providing a patient diagnosis.

Operationally, autonomous AI creators have an obligation to lawfully collect data, in the US requiring compliance with HIPAA/HITECH, as well as other applicable statutory and regulatory rules, in a manner that is transparent about the purpose and scope for which the data will be used [48]. Data used by the autonomous AI creator should be traceable to an authorization to use such data. Transparency on the part of autonomous AI creators, through written agreements, is essential to assess whether patients have adequately authorized use of data. Physicians and AI creators together are accountable directly to patients and each must take full responsibility for protecting patient rights as

stewards of patient derived data. Additionally, the rules require auditable processes and security controls to ensure that data is being used in accordance with the scope for which it was authorized and to protect the data from unauthorized use or access.

It is critical that autonomous AI system validation requirements also include ongoing monitoring of real-world performance after deployment. Typically, this is achieved by instituting a comprehensive Quality Management System (QMS), such as that under 21 CFR 820, that accommodates user feedback, complaints, reportable events, and ongoing product monitoring. Performance data monitored under the QMS should include a predefined protocol for determining whether the autonomous AI system results remain within the specified performance range that aligns with safe, effective, and equitable use of the AI system. In addition, ongoing monitoring of real-world performance includes all other quality responsibilities that remain within developers’ control such as usability, user experience, product performance (which include uptime, bugs, and issues), and necessary safety controls (which include a comprehensive framework for cybersecurity, data protection, and data privacy).

Program integrity is maximized through the inherent validation of the autonomous AI safety and equity, as well through design controls for the integration of the AI with the wider clinical EHR and charting system, ensuring automated population of EHR with the diagnostic outputs and management.

Creators of healthcare autonomous AI should assume liability for harm related to the inadequate performance of the device when used properly and on-label. This is essential for adoption: it is inappropriate for a clinician, using an autonomous AI to make a diagnosis they are not comfortable making themselves, to have full medical liability for harm caused by that AI. This position has been confirmed by the American Medical Association in its 2019 AI Policy [1]. Just like a physician grading an examination would be held responsible for their diagnosis, creators of autonomous AI products

have obtained medical malpractice insurance. This paradigm for responsibility shifts medical liability for a medical diagnostic from the provider managing the patient's diabetes, who orders the autonomous point of care retinal examination, to the autonomous AI creator.

However, medical decisions by autonomous AI on *individual patients* typically cannot be unequivocally labeled as correct or incorrect, especially in chronic diseases where outcomes may emerge years later. On *populations of patients* however, the medical decisions can be compared statistically to the desired decisions, for example to claimed correctness, and it is thus where the liability will be focused. Another issue is that, while autonomous AI is preferably compared to patient outcome, or surrogate outcome, this requires enormous resources that will be not available for the individual patient where liability is at stake. Then, the autonomous AI decision will be compared to an individual physician or group of physicians, lacking validation and thus, with unknown correspondence to outcome or surrogate outcome. This is obviously an issue for so-called continuous learning AI systems.

These distinctions will need to be resolved as various AI applications move forward.

Summary

Successful introduction of autonomous AI into the healthcare system can be achieved. For example, at the University of Iowa, introduction of autonomous AI for the diabetic eye exam has led to greatly increased compliance of people with diabetes with their annual eye exams. This was especially important during the recent COVID-19 pandemic, where patients, already at higher risk of morbidity and mortality from the virus, were reluctant to make the extra visit to an ophthalmologist that would otherwise have been required. Even after all clinics were shut down for a few weeks, the care gaps resulting from patients not getting their diabetic eye exam, were closed within weeks, so that there is almost complete compliance in this diabetes population.

Using the autonomous AI system, patients were able to get in and get the exam within minutes who would have otherwise not been able to get in for an exam for at least 6 months. Because of the program integrity enabled by autonomous AI, patients and payers were ensured that the diabetic eye exam was fully documented, billed and coded.

Trust in autonomous AI, is essential, and we need to do this “the right way” [11]. It is thus vital that everyone involved with autonomous AI helps build trust. Scientifically valid considerations, and best practices such as being currently developed by the “Foundational Principles of Ophthalmic imaging and Algorithmic Interpretation” workgroup allow any autonomous AI creator to contribute rather than detract from building trust in autonomous AI. Ultimately, the benefits of autonomous AI to patients, physicians and the public in terms of better quality and consistency of care, better access and lower cost will thus be realized [11].

References

1. American Medical Association (AMA) Board of Trustees Policy Summary. Augmented intelligence in healthcare. 2019. <https://www.ama-assn.org/system/files/2019-08/ai-2018-board-policy-summary.pdf>.
2. Horton MB, Brady CJ, Cavallerano J, Abramoff M, Barker G, Chiang MF, et al. Practice guidelines for ocular telehealth-diabetic retinopathy, 3rd edition. *Telemed J E Health*. 2020;26(4):495–543. <https://www.ncbi.nlm.nih.gov/pubmed/32209018>.
3. Helmchen LA, Lehmann HP, Abramoff MD. Automated detection of retinal disease. *Am J Manag Care*. 2014;11(17).
4. Centers for Medicare and Medicaid Services. Artificial Intelligence (AI) health outcomes challenge. 2019.
5. Char DS, Shah NH, Magnus D. Implementing machine learning in health care—addressing ethical challenges. *N Engl J Med*. 2018;378(11):981–3. <https://www.ncbi.nlm.nih.gov/pubmed/29539284>.
6. Char DS, Abramoff MD, Feudtner C. Identifying potential ethical concerns in the conceptualization, development, implementation, and evaluation of machine learning healthcare applications. *Am J Bioethics*. 2020;20(11):7–17.
7. US Food and Drug Administration (FDA). E6(R2) Good clinical practice: integrated addendum to ICH E6(R1). 2018.

8. American Diabetes A. 11. Microvascular complications and foot care: standards of medical care in diabetes-2020. *Diabetes Care*. 2020;43(Suppl 1):S135–S151. <https://www.ncbi.nlm.nih.gov/pubmed/31862754>.
9. National Committee for Quality Assurance (NCQA). HEDIS measurement year 2020 and measurement year 2021. Volume 2L technical specifications for health plans. Washington, DC: National Committee for Quality Assurance (NCQA); 2020.
10. Abramoff MD, Tobey D, Char DS. Lessons learnt about autonomous AI: finding a safe, efficacious and ethical path through the development process. *Am J Ophthalmol*. 2020; <https://www.ncbi.nlm.nih.gov/pubmed/32171769>.
11. Robeznieks A. (American Medical Association). This ophthalmologist is doing health care AI the right way AMA website. 2019. <https://www.ama-assn.org/practice-management/digital/ophthalmologist-ng-health-care-ai-right-way>.
12. Bragge P, Gruen RL, Chau M, Forbes A, Taylor HR. Screening for presence or absence of diabetic retinopathy: a meta-analysis. *Arch Ophthalmol*. 2011;129(4):435–44.
13. Rein DB, Zhang P, Wirth KE, Lee PP, Hoerger TJ, McCall N, et al. The economic burden of major adult visual disorders in the United States. *Arch Ophthalmol*. 2006;124(12):1754–60. <https://www.ncbi.nlm.nih.gov/pubmed/17159036>.
14. Fong DS, Aiello L, Gardner TW, King GL, Blankenship G, Cavallerano JD, et al. Retinopathy in diabetes. *Diabetes Care*. 2004;27(Suppl 1):S84–7.
15. Klonoff DC, Schwartz DM. An economic analysis of interventions for diabetes. *Diabetes Care*. 2000;23(3):390–404.
16. American Academy of Ophthalmology Retina/Vitreous Panel, Hoskins Center for Quality Eye Care. Preferred practice patterns: diabetic retinopathy. In: American Academy of Ophthalmology Retina Panel, editor. Updated 2016 ed. San Francisco, CA: American Academy of Ophthalmology; 2016.
17. Ahmed J, Ward TP, Bursell SE, Aiello LM, Cavallerano JD, Vigersky RA. The sensitivity and specificity of nonmydriatic digital stereoscopic retinal imaging in detecting diabetic retinopathy. *Diabetes Care*. 2006;29(10):2205–9. <http://www.ncbi.nlm.nih.gov/pubmed/17003294>.
18. Aiello LM, Bursell SE, Cavallerano J, Gardner WK, Strong J. Joslin vision network validation study: pilot image stabilization phase. *J Am Optom Assoc*. 1998;69(11):699–710.
19. Benoit SR, Swenor B, Geiss LS, Gregg EW, Saaddine JB. Eye care utilization among insured people with diabetes in the U.S., 2010–2014. *Diabetes Care*. 2019;42(3):427–33. <https://www.ncbi.nlm.nih.gov/pubmed/30679304>.
20. Solomon SD, Chew E, Duh EJ, Sobrin L, Sun JK, VanderBeek BL, et al. Diabetic retinopathy: a position statement by the American Diabetes Association. *Diabetes Care*. 2017;40(3):412–8. <https://www.ncbi.nlm.nih.gov/pubmed/28223445>.
21. Cagliero E, Levina EV, Nathan DM. Immediate feedback of HbA1c levels improves glycemic control in type 1 and insulin-treated type 2 diabetic patients. *Diabetes Care*. 1999;22(11):1785–9. <https://www.ncbi.nlm.nih.gov/pubmed/10546008>.
22. Lian J, Liang Y. Diabetes management in the real world and the impact of adherence to guideline recommendations. *Curr Med Res Opin*. 2014;30(11):2233–40. <https://www.ncbi.nlm.nih.gov/pubmed/25105305>.
23. Egbunike V, Gerard S. The impact of point-of-care A1C testing on provider compliance and A1C levels in a primary setting. *Diabetes Educ*. 2013;39(1):66–73.
24. Pugh JA, Jacobson JM, Van Heuven WA, Watters JA, Tuley MR, Lairson DR, et al. Screening for diabetic retinopathy. The wide-angle retinal camera. *Diabetes Care*. 1993;16(6):889–95. <http://www.ncbi.nlm.nih.gov/pubmed/8100761>.
25. Lin DY, Blumenkranz MS, Brothers RJ, Grosvenor DM. The sensitivity and specificity of single-field nonmydriatic monochromatic digital fundus photography with remote image interpretation for diabetic retinopathy screening: a comparison with ophthalmoscopy and standardized mydriatic color photography. *Am J Ophthalmol*. 2002;134(2):204–13.
26. Thorwarth WT Jr. From concept to CPT code to compensation: how the payment system works. *J Am Coll Radiol*. 2004;1(1):48–53. <https://www.ncbi.nlm.nih.gov/pubmed/17411519>.
27. Chiang MF, Casper DS, Cimino JJ, Starren J. Representation of ophthalmology concepts by electronic systems: adequacy of controlled medical terminologies. *Ophthalmology*. 2005;112(2):175–83. <https://www.ncbi.nlm.nih.gov/pubmed/15691548>.
28. Steindel SJ. A comparison between a SNOMED CT problem list and the ICD-10-CM/PCS HIPAA code sets. *Perspect Health Inf Manag*. 2012;9:1b. <https://www.ncbi.nlm.nih.gov/pubmed/22548020>.
29. Linder JA, Kaleba EO, Kmetik KS. Using electronic health records to measure physician performance for acute conditions in primary care: empirical evaluation of the community-acquired pneumonia clinical quality measure set. *Med Care*. 2009;47(2):208–16. <https://www.ncbi.nlm.nih.gov/pubmed/19169122>.
30. Shortliffe EH, Davis R, Axline SG, Buchanan BG, Green CC, Cohen SN. Computer-based consultations in clinical therapeutics: explanation and rule acquisition capabilities of the MYCIN system. *Comput Biomed Res*. 1975;8(4):303–20. <http://www.ncbi.nlm.nih.gov/pubmed/1157471>.
31. Fukushima K. Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol Cybern*. 1980;36(4):193–202. <http://www.ncbi.nlm.nih.gov/pubmed/7370364>.
32. Rumelhart DE, McClelland JL, University of California San Diego. PDP Research Group. Parallel distributed processing: explorations in the microstructure of cognition. Cambridge, MA: MIT Press; 1986.

33. Wolf RM, Channa R, Abramoff MD, Lehmann HP. Cost-effectiveness of autonomous point-of-care diabetic retinopathy screening for pediatric patients with diabetes. *JAMA Ophthalmol*. 2020. <https://www.ncbi.nlm.nih.gov/pubmed/32880616>.
34. Abramoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *Nat Digit Med*. 2018;1(1):39. <https://doi.org/10.1038/s41746-018-0040-6>.
35. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366(6464):447–53. <https://www.ncbi.nlm.nih.gov/pubmed/31649194>.
36. Larrazabal AJ, Nieto N, Peterson V, Milone DH, Ferrante E. Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. *Proc Natl Acad Sci USA*. 2020;117(23):12592–4. <https://www.ncbi.nlm.nih.gov/pubmed/32457147>.
37. Angus DC. Randomized clinical trials of artificial intelligence. *JAMA*. 2020; <https://www.ncbi.nlm.nih.gov/pubmed/32065828>.
38. Pearl J, Mackenzie D. *The book of why: the new science of cause and effect*. New York: Basic Books; 2018.
39. Bossuyt PM, Lijmer JG, Mol BW. Randomised comparisons of medical tests: sometimes invalid, not always efficient. *Lancet*. 2000;356(9244):1844–7. <https://www.ncbi.nlm.nih.gov/pubmed/11117930>.
40. Korevaar DA, Gopalakrishna G, Cohen JF, Bossuyt PM. Targeted test evaluation: a framework for designing diagnostic accuracy studies with clear study hypotheses. *Diagn Progn Res*. 2019;3:22. <https://www.ncbi.nlm.nih.gov/pubmed/31890896>.
41. Lu B, Gatsonis C. Efficiency of study designs in diagnostic randomized clinical trials. *Stat Med*. 2013;32(9):1451–66. <https://www.ncbi.nlm.nih.gov/pubmed/23071073>.
42. Kaplan RM, Irvin VL. Likelihood of null effects of large NHLBI clinical trials has increased over time. *PLoS One*. 2015;10(8):e0132382. <https://www.ncbi.nlm.nih.gov/pubmed/26244868>.
43. Ting DSW, Peng L, Varadarajan AV, Keane PA, Burlina PM, Chiang MF, et al. Deep learning in ophthalmology: the technical and clinical considerations. *Prog Retin Eye Res*. 2019;72:100759. <https://www.ncbi.nlm.nih.gov/pubmed/31048019>.
44. Van Dijk HW, Verbraak FD, Kok PHB, Oberstein SYL, Schlingemann RO, Russell SR, et al. Variability in photocoagulation treatment of diabetic macular oedema. *Acta Ophthalmol*. 2013;91(8):722–7. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84888203653&doi=10.1111%2fj.1755-3768.2012.02524.x&partnerID=40&md5=48a44cbc77f3b8682f5c428b10c88683>.
45. Fenton JJ, Taplin SH, Carney PA, Abraham L, Sickles EA, D’Orsi C, et al. Influence of computer-aided detection on performance of screening mammography. *N Engl J Med*. 2007;356(14):1399–409.
46. Sonka M, Fitzpatrick JM. *Handbook of medical imaging—volume 2, medical image processing and analysis*. Wellingham, WA: The International Society for Optical Engineering Press; 2000.
47. Sackett DL. Bias in analytic research. *J Chronic Dis*. 1979;32(1–2):51–63. <https://www.ncbi.nlm.nih.gov/pubmed/447779>.
48. Blumenthal D. Launching HITECH. *N Engl J Med*. 2010;362(5):382–5. <http://www.ncbi.nlm.nih.gov/pubmed/20042745>.
49. Copeland R, Needleman S. Google’s ‘Project Nightingale’ triggers federal inquiry. *WSJ*. 2019. <https://www.wsj.com/articles/behind-googles-project-nightingale-a-health-data-gold-mine-of-50-million-patients-11573571867>.
50. Moyer VA, Force USPST. Screening for glaucoma: U.S. preventive services task force recommendation statement. *Ann Intern Med*. 2013;159(7):484–9. <https://www.ncbi.nlm.nih.gov/pubmed/24325017>.
51. Chou R, Dana T, Bougatsos C, Grusing S, Blazina I. Screening for impaired visual acuity in older adults: updated evidence report and systematic review for the US preventive services task force. *JAMA*. 2016;315(9):915–33. <https://www.ncbi.nlm.nih.gov/pubmed/26934261>.
52. McLaughlin CC, Wu XC, Jemal A, Martin HJ, Roche LM, Chen VW. Incidence of noncutaneous melanomas in the U.S. *Cancer*. 2005;103(5):1000–7. <https://www.ncbi.nlm.nih.gov/pubmed/15651058>.
53. Sullivan HR, Schweikart SJ. Are current tort liability doctrines adequate for addressing injury caused by AI? *AMA J Ethics*. 2019;21(2):E160–6. <https://www.ncbi.nlm.nih.gov/pubmed/30794126>.
54. Maier S. Elon take the wheel. *Minnesota Law Rev*. 2017. <https://minnesotalawreview.org/2017/01/24/elon-take-the-wheel/>.
55. Chandler RJ, Venditti CP. Gene therapy for metabolic diseases. *Transl Sci Rare Dis*. 2016;1(1):73–89. <https://www.ncbi.nlm.nih.gov/pubmed/27853673>.
56. Russell S, Bennett J, Wellman JA, Chung DC, Yu ZF, Tillman A, et al. Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial. *Lancet*. 2017;390(10097):849–60. <https://www.ncbi.nlm.nih.gov/pubmed/28712537>.
57. Beauchamp TL, Childress JF. *Principles of biomedical ethics*. 8th ed. New York: Oxford University Press; 2019.
58. Shah A, Lynch S, Niemeijer M, Amelon R, Clarida W, Folk J, et al., editors. Susceptibility to misdiagnosis of adversarial images by deep learning based retinal image analysis algorithms. In: *Proceedings—International symposium on biomedical imaging*; 2018.
59. Finlayson SG, Bowers JD, Ito J, Zittrain JL, Beam AL, Kohane IS. Adversarial attacks on medical

- machine learning. *Science*. 2019;363(6433):1287–9. <https://www.ncbi.nlm.nih.gov/pubmed/30898923>.
60. Friedenwald J, Day R. The vascular lesions of diabetic retinopathy. *Bull Johns Hopkins Hosp*. 1950;86(4):253–4. <http://www.ncbi.nlm.nih.gov/pubmed/15411556>.
 61. MacKenzie S. A case of glycosuric retinitis, with comments (Microscopical Examination of the Eyes by Mr. Nettleship). *Roy London Ophthal Hosp Rep*. 1879;9(134).
 62. Hubel DH, Wiesel TN. Receptive fields of single neurones in the cat's striate cortex. *J Physiol*. 1959;148:574–91.
 63. Ts'o DY, Frostig RD, Lieke EE, Grinvald A. Functional organization of primate visual cortex revealed by high resolution optical imaging. *Science*. 1990;249(4967):417–20.
 64. Wang YT, Tadarati M, Wolfson Y, Bressler SB, Bressler NM. Comparison of prevalence of diabetic macular edema based on monocular fundus photography vs optical coherence tomography. *JAMA Ophthalmol*. 2016;134(2):222–8. <http://www.ncbi.nlm.nih.gov/pubmed/26719967>.
 65. Fundus photographic risk factors for progression of diabetic retinopathy. ETDRS report number 12. Early treatment diabetic retinopathy study research group. *Ophthalmology*. 1991;98(5 Suppl):823–33.
 66. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open*. 2016;6(11):e012799. <https://www.ncbi.nlm.nih.gov/pubmed/28137831>.
 67. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, Chan A-W, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med*. 2020;26(9):1364–74. <https://doi.org/10.1038/s41591-020-1034-x>.
 68. US Food and Drug Agency (FDA). FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems. Washington, DC; 2018. <https://www.fda.gov/news-events/newsroom/pressannouncements/ucm604357.htm>.
 69. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med*. 1996;125(7):605–13. <https://www.ncbi.nlm.nih.gov/pubmed/8815760>.
 70. Temple R. A regulatory authority's opinion about surrogate endpoints. In: Nimmo W, Tucker G, editors. *Clinical measurement in drug evaluation*. New York: Wiley; 1995.
 71. Browning DJ, Glassman AR, Aiello LP, Bressler NM, Bressler SB, Danis RP, et al. Optical coherence tomography measurements and analysis methods in optical coherence tomography studies of diabetic macular edema. *Ophthalmology*. 2008;115(8):1366–71, 71 e1. <http://www.ncbi.nlm.nih.gov/pubmed/18675696>.
 72. Nagendran M, Chen Y, Lovejoy CA, Gordon AC, Komorowski M, Harvey H, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *BMJ*. 2020;368:m689. <https://www.ncbi.nlm.nih.gov/pubmed/32213531>.
 73. Lin AP, Katz LJ, Spaeth GL, Moster MR, Henderer JD, Schmidt CM Jr, et al. Agreement of visual field interpretation among glaucoma specialists and comprehensive ophthalmologists: comparison of time and methods. *Br J Ophthalmol*. 2011;95(6):828–31. <http://www.ncbi.nlm.nih.gov/pubmed/20956271>.

Technical Aspects of Deep Learning in Ophthalmology

5

Ashkan Abbasi, Sowjanya Gowrisankaran,
Wei-Chun Lin, Gadi Wollstein, Joel S. Schuman,
and Hiroshi Ishikawa

Abstract

Accurate glaucoma diagnosis and management require understanding structural and functional measurements, and their complex relationship. In recent years, artificial intelligence (AI), particularly deep learning (DL), has shown great promise in analyzing fundus and optical coherence tomography images, as well as visual field data, for tasks such as diagnosis, progression analysis, knowledge

discovery, data quality improvement, and even synthetic data generation. Advances in natural language processing (NLP) have also enabled the use of electronic health records, creating new opportunities for personalized treatment and outcome prediction. This chapter concisely reviews the literature on these AI applications, covering diagnosis (using structural and functional data), structure-function correlation analysis, longitudinal progression prediction, and NLP-enabled EHR data utilization. It also highlights challenges, including the need for larger, more diverse datasets, improved model explainability, and the diversity of glaucoma definitions.

A. Abbasi · S. Gowrisankaran · W.-C. Lin
Department of Ophthalmology, Oregon Health &
Science University, Portland, OR, USA
e-mail: abbasi@ohsu.edu

S. Gowrisankaran
e-mail: gowrisan@ohsu.edu

W.-C. Lin
e-mail: linw@ohsu.edu

G. Wollstein · J. S. Schuman
Wills Eye Hospital, Philadelphia, PA, USA
e-mail: gwillstein@willseye.org

J. S. Schuman
e-mail: jschuman@willseye.org

H. Ishikawa (✉)
Department of Ophthalmology, Casey Reading
Center, Casey Eye Institute, Oregon Health &
Science University, Portland, OR, USA
e-mail: ishikawh@ohsu.edu

Department of Medical Informatics and Clinical
Epidemiology (DMICE), Oregon Health & Science
University, Portland, OR, USA

Keywords

Machine learning · Artificial intelligence ·
AI · Deep learning · Medical application ·
Ophthalmology

Deep learning (DL) is a specific category of machine learning that uses neural networks (NNs) with many layers to automatically learn and extract complex patterns and representations from large datasets. The first NN dates back to 1940s when Walter Pitts and Warren McCulloch designed a computation model to implement

logical operations while simulating of biological neurons in the human brain [1]. Early neural networks consisted of only one layer of neurons, significantly limiting their modeling capabilities. Although it was known that a NN with two layers is a universal function approximator, there was no effective method training its parameters (or weights) until the development of backpropagation in 1985. For many years, practical NNs typically had only two to three layers because training NNs with many layers, aka deep NNs, was still difficult due to optimization and gradient instability problems [2]. Another significant advancement was the advent of Convolutional Neural Networks (CNNs) in 1989 to address several key challenges of traditional NNs for image processing [3].

Later developments in optimization methods, the availability of large datasets like ImageNet (with over 14 million labeled images in 1000 categories) [4], and the accelerating the training process using graphics processing units (GPUs) enabled the effective training of deep NNs. In 2012, the success of AlexNet [5], a CNN with eight layers, that reduced the top-5 error rate from 26.1% to 15.3% in the ImageNet challenge, sparked widespread interest in DL across both academia and industry. This breakthrough highlighted the potential of DL to tackle complex tasks, leading to a surge in research and development that continues to drive innovation in the field. Since then, AI has outperformed humans on numerous benchmarks, particularly in areas such as image classification, visual reasoning, and English comprehension [6]. The high dependence of the field of Ophthalmology on imaging makes DL a perfect fit for this field. In recent years, research involving DL in ophthalmology has risen exponentially [7–9].

In this chapter, we briefly review concepts related to DL, main components of a deep NN, and common deep NN architectures. We talk about different forms of machine learning, a simple feed forward NN, and gradually present more sophisticated methods and architectures.

Learning from Data: Main Training Approaches

Machine learning models, including DL, rely on parameters learned from data. To achieve this, various approaches can be employed based on the objectives and available data. Below, we'll provide a concise overview of the various learning approaches.

Supervised Learning

Supervised learning is the most common training approach where a model's parameters are learned from labeled data, where the input and the corresponding output are provided. The goal is to enable the model to learn the mapping and make predictions on unseen data. Typical supervised learning tasks include classification, assigning discrete categories to inputs (e.g., diagnosis conditions: retinopathy, glaucoma, etc.), and regression, predicting continuous outcomes based on inputs (e.g., vision outcomes after surgeries).

Unsupervised Learning

Unsupervised learning involves training models on data without labeled outputs, allowing the discovery of hidden patterns or structures within the data. Key applications of unsupervised learning include clustering, where the method groups similar data points (e.g., clustering patients with similar symptoms), and dimensionality reduction, where the method projects data from a higher to lower dimensions while maintaining important information for data visualization, noise reduction, feature extraction, and data compression.

Semi-supervised Learning

Semi-supervised learning is a hybrid approach that combines the strengths of both supervised and unsupervised learning. It utilizes a small

amount of labeled data alongside a larger pool of unlabeled data to train models. This method is particularly useful when acquiring labeled data is expensive or time-consuming, but there is an abundance of unlabeled data available. By leveraging the structure and patterns within the unlabeled data, this approach can potentially improve model performance and generalization.

Self-supervised Learning

Self-supervised learning (SSL) aims to solve an auxiliary or pretext task that benefits a main downstream task. For example, instead of directly training a model for disease detection from an input image, SSL first exploits a pretext task, where labels can be automatically generated for all data. A model can be trained to predict missing parts of data or reconstruct its input itself. Once the model is trained on the pretext task, the pretrained model is reused for the downstream task (e.g., by training on a smaller subset of data that has labels). The main benefit of SSL is that it can reduce the need for human-annotated data by leveraging vast amounts of unlabeled data to create robust feature representations, which can be fine-tuned for specific tasks with minimal labeled data while it often leads to more robust and generalizable models.

Reinforcement Learning

Reinforcement learning (RL) is a type of machine learning where an agent learns to make decisions using a policy, a dynamic strategy aimed at optimizing the reward outcome. This process involves trial-and-error and numerous policy iterations over time until an optimal policy is identified. The rewards can be defined sparsely, and thus, it is suitable for training on incomplete data over time. In ophthalmology, the potential applications of RL remain largely unexplored, mostly due to major limitations of RL such as the difficulty in exploring potential policies in real-world scenarios, susceptibility to system delays (that complicates timely evaluation of an action's effects), and susceptibility

to prioritize short-term rewards and neglecting long-term outcomes [10].

Feedforward Neural Network

A feedforward neural network (FNN), or a multilayer perceptron (MLP), is a core building block in DL. It is composed of multiple layers of interconnected computation units, or neurons. The flow of information in an FNN is unidirectional, moving from the input layer through hidden layers, and finally to the output layer.

The objective of an FNN is to approximate a target function f^* with the model $y=f(x; w)$, where w represents the weights of the network. These weights are adjusted during training to minimize the difference between the target function f^* and the model f . The back-propagation method [2] is used to train the weights by iteratively computing gradients and updating the weights to reduce a predefined loss function, which quantifies the difference between the target output and the network's prediction. Thus, the FNN improves its approximation of the target function through training on a dataset of input-output pairs.

Figure 5.1 illustrates the structure of an FNN with two layers (excluding the input layer). In an FNN, neurons in one layer are fully connected to all neurons in the subsequent layer. The use of non-linear activation functions, which was initially inspired by the biological neurons' activation mechanism, is necessary to enable the network to learn complex and non-linear patterns. Without activation functions, a multi-layer network would be reducible to a single-layer network, as it would be capable only of linear transformations. There are many options for activation functions, however, a popular choice is rectified linear units (ReLU), due to its efficiency and ability to effectively backpropagate gradients, aiding in the training of deep NNs [2].

Convolutional Neural Network

While FFNs (Sect. 5.2) are powerful, they face inherent limitations when it comes to handling grid-like and high-dimensional data such as

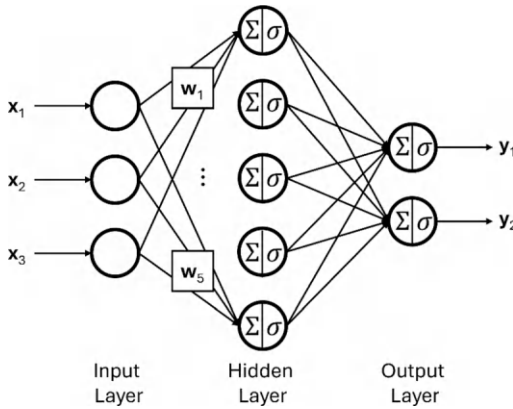


Fig. 5.1 Structure of an FNN with two layers. The input layer only passes the elements of the input vector $\mathbf{x}$ that are shown by x_1 , x_2 , and x_3 in this figure. Then, each neuron in the subsequent layers computes a weighted sum of its inputs followed by an activation function ($\sigma(\cdot)$). For instance, the first neuron of the hidden layer computes $\sigma\left(\sum_{i=1}^3 x_i w_{1,i}\right)$, where w_1 shows the weight vector associated with the first neuron, and $w_{1,i}$ shows the i -th element of it. The output layer performs a similar computation using the activated outputs of the hidden layer. The total number of layers that have weights define the “depth” of the network, while the “width” of a network is determined by the number of neurons in its hidden layers

images. In contrast, convolutional neural networks (CNNs) have been highly successful in analyzing visual data. CNNs are inspired by the biological processes in the visual cortex of animals [2], where simple cells respond to lines and

complex cells respond to certain location-invariant patterns. The convolutional and pooling layers directly mimic the behavior of the simple and complex cells. CNNs can leverage the spatial structure of the data and learn hierarchies of features automatically from input images.

Basic Operations

At the heart of CNNs is the discrete convolution operation (Fig. 5.2) that is a specialized kind of linear operation that involves sliding a filter (also known as a kernel) over the input data to produce a feature map. This process highlights distinct features in the input image, such as edges, textures, and shapes, which are essential for image recognition tasks. The convolution operation is key to the success of CNNs because it reduces the number of parameters compared to fully connected layers, making the network more efficient and less prone to overfitting.

In a typical CNN, after applying convolutional layers to extract features, pooling layers are inserted to progressively reduce the spatial size of the feature maps. For example, a max pooling layer with a 2×2 window and stride of 2 would reduce a 4×4 feature map to 2×2 by taking the maximum value in each 2×2 block (Fig. 5.3). The subsampling of spatial

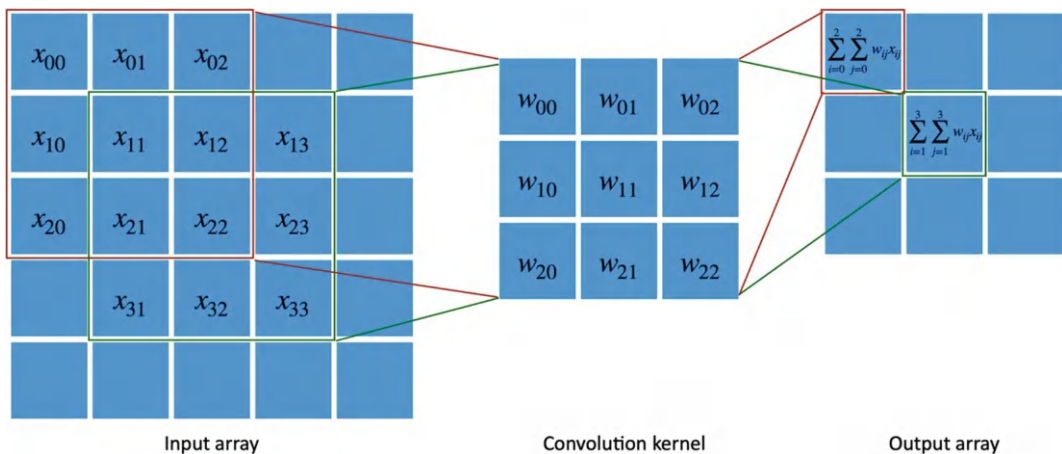


Fig. 5.2 Example of convolution operation with 3×3 kernel size. Each element in the output array is the sum of point-wise multiplication of a 3×3 convolution kernel and a 3×3 corresponding patch in the input array

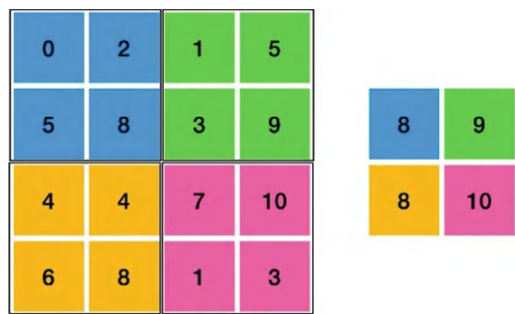


Fig. 5.3 Example of max pooling operation with a stride of 2. Each element in the output array is the maximum value in the corresponding 2×2 region in the input array

information has been shown to improve computational and memory efficiency, preventing overfitting by decreasing the number of trainable parameters, and making the network more robust to small translations of the input.

Properties of CNNs

CNNs leverage sparse connectivity and weight sharing to efficiently process data. Sparse connectivity means each neuron in a convolutional layer connects only to a localized region of the input, reducing the number of parameters. Weight sharing uses the same set of weights (filters) across different regions of the input, enabling the network to detect the same feature in various locations. These mechanisms allow CNNs to capture spatial hierarchies and learn patterns effectively.

One of the powerful properties of CNNs is their ability to achieve translation invariance, meaning they can recognize objects regardless of their position in the image. This is crucial for tasks where the location of the object may vary. However, CNNs are not inherently invariant to other transformations such as rotation and scaling, although techniques like data augmentation can be used to improve their performance in these aspects. These invariant properties make CNNs particularly suitable for a wide range of applications.

Advances in CNNs

In addition to convolutional and pooling layers, batch normalization (BN) and skip connections are often used in modern CNNs. BN stabilizes and accelerates the training process by normalizing the inputs of each layer [11]. This helps in mitigating issues such as internal covariate shift, making the network training more stable and faster. Skip connections, or shortcut connections, allow the gradient to flow through the network more easily, addressing the vanishing gradient problem and enabling the training of much deeper networks such as ResNet architecture [12]. Besides that, skip connections are vital in devising multiscale CNNs, which is particularly important in medical applications [13] because traditional CNNs typically have difficulty in multiscale analysis. Advanced architectures like U-Net [14], feature pyramid networks [15], and Inception [16] models all leverage skip connections to realize their multiscale processing capabilities.

In recent years, significant advances have been made to enhance the computational and memory efficiency of CNNs, making these models better suited for resource-constrained devices. MobileNet-v1 [17] pioneered the use of depthwise separable convolutions, significantly reducing parameters and computational demands. MobileNet-v2 [18] improved upon this by integrating inverted residuals and linear bottlenecks, further boosting performance and efficiency. EfficientNet [19] took a different approach with its compound scaling method, which balances the network’s depth, width, and resolution for optimized resource usage while achieving top-tier accuracy. These innovations have enabled the deployment of powerful CNNs that balance performance and efficiency.

Ophthalmology is a field rich in visual data, and unsurprisingly, CNNs have been employed in numerous ophthalmic applications. These range from low-level image enhancement tasks like image denoising [20, 21] to high-level image classification and segmentation tasks, such as the automatic diagnosis of diseases [22]

and the segmentation of retinal features and biomarkers [7, 8, 23]. By analyzing retinal images, CNNs can identify pathological features with high accuracy, aiding in early diagnosis and treatment, and ultimately improving patient outcomes. In some applications, CNNs have even surpassed human performance [6].

Recurrent Neural Networks

Recurrent Neural Networks (RNNs) are a class of NNs designed to recognize patterns in sequences of data, such as time series, speech, or text. Unlike traditional NNs, which assume that inputs are independent of each other, RNNs leverage sequential information by maintaining a memory of previous inputs in the form of internal states [2]. This capability makes RNNs particularly powerful for tasks where the context of previous elements is crucial for understanding the current element. In ophthalmology, RNNs can analyze patient data like eye movement or retinal images over time to track diseases and improve diagnoses.

RNNs have a loop structure that lets them pass information from one step to the next, acting like a memory (Fig. 5.4). Each neuron in an RNN processes the current input and the hidden state from the previous step to effectively create a form of memory.

Despite their strengths, RNNs face challenges such as difficulty in learning long-term dependencies due to issues like vanishing and exploding gradients. However, they can struggle with remembering information from far back in the sequence. To fix this, Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) networks were created, which are better at handling long-term dependencies and make RNNs more reliable for complex tasks.

Attention Mechanisms

The attention mechanism is inspired by the selective focus ability of the human brain. Attention techniques enable DL models to focus on the most relevant parts of the input or features. It can enhance the representation power of the model to handle tasks with complex dependencies. Attention is basically realized by dynamically assigning different weights to different parts of the input or features. This contrasts with traditional FFNs and CNNs, where fixed weights are always applied to the input.

Attention was initially introduced in the field of natural language processing [24] to help understand relationships between words in a sentence because the process of adaptively adjusting weights based on the context and

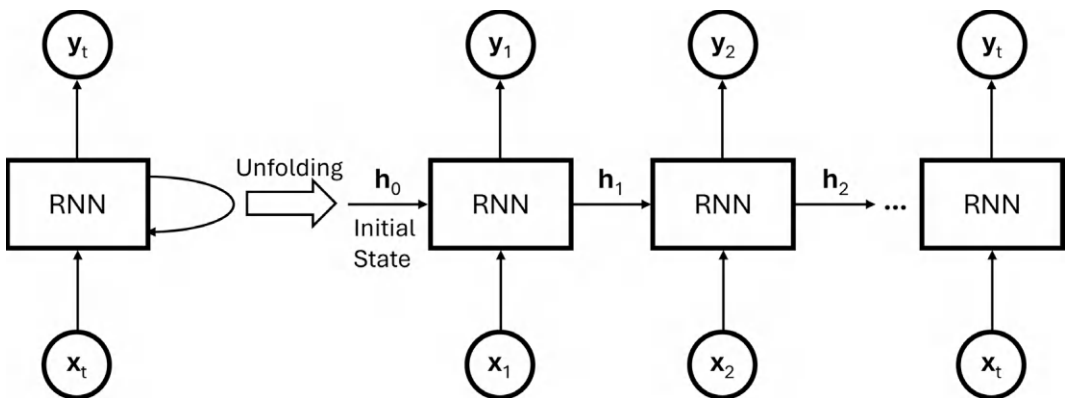


Fig. 5.4 On the left, the figure shows an RNN — a neural network structure with an internal feedback loop. On the right, the RNN is “unfolded” across time steps to illustrate how it processes sequential data. At each time step t , the RNN takes the current input x_t and the hidden state from

the previous step h_{t-1} , producing a new hidden state h_t and an output y_t . The hidden state propagates forward through time, allowing the network to retain and utilize temporal information. Typically, a subset of the outputs is used to produce the final output of the model, depending on the task

relevance to the task can enable the model to capture long range dependencies that may span across the entire input. But later it has proved to be crucial in broader applications such as image processing and computer vision. Common tasks such as image classification, segmentation, and object detection can substantially benefit from enforcing the model to focus on salient regions in the input or features, leading to more precise feature extraction and more accurate object recognition and scene understanding performance.

An attention mechanism can be implemented in various ways in order to determine where to attend (spatial attention), what to attend (channel or feature attention), when to attend (temporal attention), and which to attend (network branch attention in NNs with multiple branches) [25]. For instance, Squeeze-and-Excitation (SE) is a widely used channel attention mechanism that can be used in almost any CNN architecture. SE is designed to explicitly model interdependencies between channels of a convolutional feature map [26]. This method has two main steps: “squeeze” and “excitation.” During the squeeze step, global average pooling followed by a small FFN is applied to capture channel-wise information. In the excitation step, the condensed information obtained using the squeeze operation is further processed through another small FFN to learn channel-wise dependencies and generate a set of weights, which are then used to adaptively re-calibrate the feature maps by scaling each channel. This process allows the network to emphasize important features and suppress less useful ones, enhancing its performance in various tasks, including image classification and object detection.

Transformers

A Transformer is basically an encoder-decoder network architecture that performs several self-attention operations in parallel, known as multi-head attention, to capture relationships between input tokens from different viewpoints [24]. This

parallel attention mechanism allows the model to process sequences more efficiently than traditional methods, such as RNNs.

Self-attention can be briefly described using an example from natural language processing, where it was first introduced. Consider this sentence: “The quick brown fox jumps over the fence.” For each word, the Transformer uses linear transformations to generate the following three different vectors: a query (Q), a key (K), and a value (V). For the word “fox”, the query vector indicates how other words in the sentence are relevant to it. The key vector shows how “fox” is relevant to the other words in the sentence. To obtain the value vector, attention scores are needed. Normalized dot products between query and key vectors are used to obtain the attention scores. Then, these scores are multiplied by the previous value vectors to create a weighted sum, which results in a single, new value vector for each word. This new vector is a refined representation that considers the word’s relationship with all other words in the sequence and summarizes all the information about the word “fox” to be used in the subsequent layer.

A Transformer implements the self-attention mechanism multiple times in parallel. This is called multi-head attention, in which $\text{MultiHead}(Q, K, V) = [\text{head}_1, \dots, \text{head}_h]W_0$, where $\text{head}_i = \text{SelfAttention}(QW_i, KW_i, VW_i)$. All W matrices are learnable parameters. Multi-head attention enables the model to simultaneously focus on different aspects of the sentence. For example, one head might focus on the relevance of “fox” and the verb “jump”, and the other head might focus on the relation between “fox” and its color (“brown”). Also, to account for token positions in the input sequence, Transformers incorporate positional embeddings. These embeddings are added to the input embeddings to provide information about the order of tokens, allowing the model to understand the sequence structure while processing all tokens simultaneously. Unlike RNNs, which process sequences sequentially and can suffer from memory loss,

Transformers process all tokens in parallel, significantly improving efficiency and performance.

Generative Adversarial Networks (GANs)

GANs, which have the capacity to generate data without explicitly modelling the underlying distribution, are one of the most interesting recent innovations in DL. GANs are a special form of neural networks where two networks, a generator, and a discriminator, are trained alternatively. The generator is trained to produce realistic data sample from some random noise while the discriminator is trained to discriminate the fake samples generated by the generator and the real samples. Figure 5.5 shows the structure of GANs. CNNs and RNNs can be easily incorporated into GANs to deal with array data and sequence data.

There are two potential applications of GANs in Ophthalmology. The first one focuses on the generator which can extract the underlying structure of the data and learn to generate new data samples. Plausible OCT scans generated from random noise [27] as well as denoised scans generated from real scans [20] can be achieved by the generator of GANs. The second one focuses on the discriminator which uses the discriminator as a learned prior to detect abnormal samples. For example, Zhou et al. trained a GAN with only healthy data and use the trained discriminator to detect anomalies from OCT [27].

Language Models

Encoder Models: BERT

Encoder models like BERT (Bidirectional Encoder Representations from Transformers) [28] have revolutionized NLP by enabling deep bidirectional understanding of text. BERT processes input text by looking at both the left and right context of each word simultaneously, which allows it to capture the intricate relationships between words in a sentence. The architecture of BERT consists of multiple layers of bidirectional Transformer encoders, each with self-attention mechanisms that focus on different parts of the input text. This design helps BERT to generate contextualized word embeddings that are highly effective for various NLP tasks such as named entity recognition, question answering, and text classification. In ophthalmology, BERT can be employed to extract critical clinical information from unstructured medical records.

Decoder Models: GPT

Decoder models, exemplified by the Generative Pretrained Transformer (GPT) [29], are designed to generate coherent and contextually relevant text based on a given prompt. GPT utilizes an unmodified Transformer decoder architecture, which differs from models like BERT that use Transformer encoder blocks. Unlike

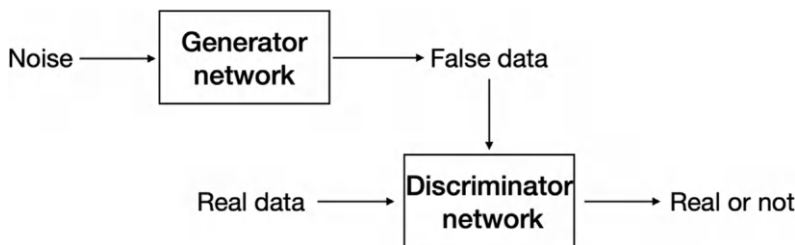


Fig. 5.5 Structure of GAN. GAN is composed of a generator and a discriminator. The generator is trained to generate plausible samples from noise to fool the

discriminator while the discriminator is trained to discriminate between the fake and true samples

encoder-decoder models, GPT lacks the encoder attention part. Take GPT-3, for example, the model is built using 96 attention blocks, each containing 96 attention heads, allowing it to focus on different parts of the input text simultaneously. GPT-3 contains 175 billion parameters and was trained on large text datasets totaling 570GB, including hundreds of billions of words, making it one of the largest neural networks upon its release.

The pre-training of GPT involves a process called “next-word prediction,” a form of unsupervised training where the model predicts the next word in a sentence. The input sequence is fixed to a certain number of tokens, and for shorter sequences, empty values are used to fill the remaining positions. The output is a sequence of probability distributions for each word position, though typically only the final word’s prediction is used when generating text. The model calculates the key, query, and value vectors for each token using dot product matrix multiplication, producing a score matrix. This matrix determines how much attention each word should receive relative to other words in the sequence. The scores are scaled by the square root of the dimension of the key and query vectors to stabilize gradients, followed by a SoftMax operation to obtain attention weights—probability values between 0 and 1. These weights allow the model to focus on the most relevant words when generating text.

GPT’s ability to generate detailed and contextually appropriate text makes it useful in ophthalmology for tasks like drafting clinical notes, generating patient education materials, and providing communication with patients such as chatbot.

Conclusion

In conclusion, this chapter briefly reviewed technical aspects of deep learning that are commonly used in ophthalmic applications. Today, deep learning has profoundly impacted all fields, including ophthalmology, by providing sophisticated tools for analyzing and interpreting

complex visual data. The advancements in neural network architectures, such as CNNs, RNNs, and attention mechanisms, Transformers, and generative adversarial networks have enabled the development of highly accurate models capable of performing a wide range of tasks, from image classification and segmentation to sequence analysis and anomaly detection. The integration of these technologies has not only enhanced diagnostic accuracy but also facilitated early detection and treatment of various ophthalmic conditions, ultimately improving patient outcomes.

As deep learning continues to evolve, the potential for further innovations in ophthalmology remains vast. The emergence of self-supervised learning, GANs, and transformer models opens new avenues for leveraging unlabeled data, generating realistic synthetic images, and extracting valuable insights from unstructured clinical records. The ongoing research and development in this area promise to deliver even more powerful and efficient models, capable of addressing the unique challenges of ophthalmic imaging and providing unprecedented levels of precision and reliability in clinical practice.

References

1. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys.* 1943;5(4):115–35. <https://doi.org/10.1007/BF02478259>.
2. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521(7553):436–44. <https://doi.org/10.1038/nature14539>.
3. LeCun Y, Boser B, Denker JS, et al. Backpropagation applied to handwritten zip code recognition. *Neural Comput.* 1989;1(4):541–51. <https://doi.org/10.1162/neco.1989.1.4.541>.
4. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. ImageNet: a large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. 2009:248–55. <https://doi.org/10.1109/CVPR.2009.5206848>.
5. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ, editors. *Proceedings of the advances in neural information processing systems.* 2012:1097–105.

6. Maslej N, Fattorini L, Perrault R, et al. The AI index 2024 annual report. 2024.
7. Ting DSW, Pasquale LR, Peng L, et al. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol*. 2019;103(2):167–75. <https://doi.org/10.1136/bjophthalmol-2018-313173>.
8. Li T, Bo W, Hu C, et al. Applications of deep learning in fundus images: a review. *Med Image Anal*. 2021;69:101971. <https://doi.org/10.1016/j.media.2021.101971>.
9. Li Z, Wang L, Wu X, et al. Artificial intelligence in ophthalmology: the path to the real-world clinic. *Cell Rep Med*. 2023;4(7):101095. <https://doi.org/10.1016/j.xcrm.2023.101095>.
10. Nath S, Korot E, Fu DJ, et al. Reinforcement learning in ophthalmology: potential applications and challenges to implementation. *Lancet Digit Health*. 2022;4(9):e692–7. [https://doi.org/10.1016/S2589-7500\(22\)00128-5](https://doi.org/10.1016/S2589-7500(22)00128-5).
11. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: *Proceedings of the international conference on machine learning*. 2015:448–456.
12. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016:770–778.
13. Abbasi A, Monadjemi A, Fang L, Rabbani H, Antony BJ, Ishikawa H. Mixed multiscale BM4D for three-dimensional optical coherence tomography denoising. *Comput Biol Med*. 2023;155:106658. <https://doi.org/10.1016/j.compbio.2023.106658>.
14. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: *Lecture notes in computer science (including sub-series lecture notes in artificial intelligence and lecture notes in bioinformatics)*, vol. 9351. 2015. https://doi.org/10.1007/978-3-319-24574-4_28.
15. Lin TY, Dollár P, Girshick RB, He K, Hariharan B, Belongie SJ. Feature pyramid networks for object detection. *CoRR*. <http://arxiv.org/abs/1612.03144>.
16. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. In: *Proceedings of the IEEE computer society conference on computer vision and pattern recognition*. 2015. <https://doi.org/10.1109/CVPR.2015.7298594>.
17. Howard AG, Zhu M, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications. *ArXiv*. 2017;abs/1704.04861.
18. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. MobileNetV2: inverted residuals and linear bottlenecks. In: *Proceedings of the IEEE computer society conference on computer vision and pattern recognition*. 2018. <https://doi.org/10.1109/CVPR.2018.00474>.
19. Tan M, Le Q V. EfficientNet: rethinking model scaling for convolutional neural networks. In: *36th international conference on machine learning, ICMML 2019*. 2019.
20. Halupka KJ, Antony BJ, Lee MH, et al. Retinal optical coherence tomography image enhancement via deep learning. *Biomed Opt Express*. 2018;9(12). <https://doi.org/10.1364/boe.9.006205>.
21. Abbasi A, Monadjemi A, Fang L, Rabbani H, Zhang Y. Three-dimensional optical coherence tomography image denoising through multi-input fully-convolutional networks. *Comput Biol Med*. 2019;108:1–8. <https://doi.org/10.1016/j.compbio.2019.01.010>.
22. Maetschke S, Antony B, Ishikawa H, Wollstein G, Schuman J, Garnavi R. A feature agnostic approach for glaucoma detection in OCT volumes. *PLoS One*. 2019;14(7):0219126.
23. Thompson AC, Jammal AA, Medeiros FA. A review of deep learning for screening, diagnosis, and detection of glaucoma progression. *Transl Vis Sci Technol*. 2020;9(2):42. <https://doi.org/10.1167/tvst.9.2.42>.
24. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. In: *Advances in neural information processing systems*. 2017.
25. Guo MH, Xu TX, Liu JJ, et al. Attention mechanisms in computer vision: a survey. *Comput Vis Media (Beijing)*. 2022;8(3):331–68. <https://doi.org/10.1007/s41095-022-0271-y>.
26. Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-excitation networks. *IEEE Trans Pattern Anal Mach Intell*. 2020;42(8):2011–23. <https://doi.org/10.1109/TPAMI.2019.2913372>.
27. Zheng C, Xie X, Zhou K, et al. Assessment of generative adversarial networks model for synthetic optical coherence tomography images of retinal disorders. *Transl Vis Sci Technol*. 2020;9(2):2–29. <https://doi.org/10.1167/tvst.9.2.29>.
28. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. In: *NAACL HLT 2019—2019 conference of the North American chapter of the association for computational linguistics: human language technologies—proceedings of the conference*, vol. 1. 2019.
29. Radford A, Narasimhan K, Salimans T, Sutskever I. Improving language understanding by generative pre-training. San Francisco, CA, USA; 2018.

Threats to the Validity of the Predictive System Performance for Ophthalmology

6

Michał Woźniak and Andrzej Grzybowski

Abstract

Artificial Intelligence finds application in many areas of everyday life. One of the main fields is medical diagnostics, which plays a key role in society's well-being and the quality and cost of healthcare. However, before allowing such algorithmic solutions, we should realize the dangers of their improper use and design, which could cause great harm to individual patients and cause excessively high costs to a healthcare system. This chapter mainly addresses the problems related to the dangers associated with reliable learning sets, which are the basis for machine learning algorithms, as well as the evaluation of the medical software based on machine learning algorithms. As a result, we can obtain algorithms to support medical diagnosis. We will point out selected aspects that need attention in the design of such systems and

illustrate them with selected problems in ophthalmology.

Keywords

Artificial intelligence · Machine learning · Ophthalmology · Model evaluation · Out-of-distribution · Data bias · Data labeling · Credibility of model evaluation · Experimental protocol

Introduction

The use of artificial intelligence (AI) algorithms has been known for more than fifty years, and the first spectacular success was presented at Stanford University with Edward Shortliffe's development of the MYCIN expert system for diagnosis and therapy of blood infections based on reported symptoms and medical test results [1]. The essential characteristics of the system were the ability to control the dialogue with the user, explain the suggested diagnosis, and recommend therapy. Nowadays, the possibilities of using such systems are increasing, and diagnosis can be based not only on tabular data but also on imaging data, which is typical for ophthalmology [2]. Moreover, designing diagnostic systems based on acquired knowledge

M. Woźniak (✉)
Department of Systems and Computer Networks,
Wrocław University of Science and Technology,
Wrocław, Poland
e-mail: michal.wozniak@pwr.edu.pl

A. Grzybowski
Department of Ophthalmology, University of
Warmia and Mazury, Olsztyn, Poland
Institute for Research in Ophthalmology,
Foundation for Ophthalmology Development,
Poznan, Poland

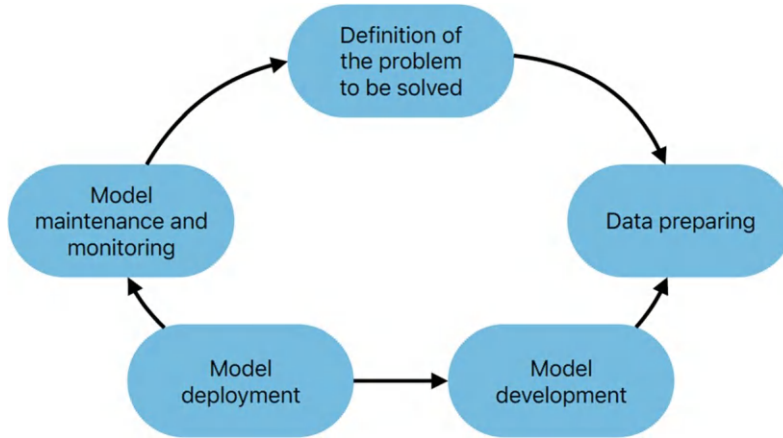


Fig. 6.1 Machine learning-based software lifecycle [6]

from expert physicians has been virtually abandoned, and machine learning (ML) algorithms are increasingly used. However, such systems require human experts to label each learning case (image), while knowledge generalization is performed algorithmically [3]. Such an approach has a number of advantages and often leads to systems with quality that exceeds that of a single expert's diagnostics. They also allow conclusions to be drawn about patient demographic data that did not seem possible to assess on the basis of the information used (e.g., age, sex, race, etc.) based on the images from different parts of the body, including retina [4, 5].

Despite these undoubted advantages, however, it is worth noting the life cycle of such systems and mentioning the critical elements of the production process. The traditional life cycle of ML-based systems consists of five main phases [6], as shown in Fig. 6.1. This cycle is based on the following phases:

- (i) Identify the problem, the value gained by solving it, and the ML tools to solve it. What is required here is a precise definition of the task, i.e., determining what input information we will have, what the output (usually a set of possible diagnoses) is, and the desired quality metrics and their expected values.
- (ii) Data preparation: collection, preparation, feature engineering.
- (iii) Model construction: model selection, learning and tuning, and evaluation using a reliable experimental protocol.
- (iv) Implementation of the model in the production system.
- (v) Monitoring ensures the final ML-based model maintains the desired quality.

Let us pay attention to selected aspects related to the various stages of construction of an example classification system, such as:

- Can we correctly define the required characteristics of the designed system?
- Are the data we have not biased?
- Are we able to correctly label the learning instances?
- Is the size of the learning set sufficient?
- How do we correctly define the evaluation protocol of the designed system?
- Are we sure that the data to be classified by the designed system will come from the same distribution, or can the data distribution change over time?

The chapter presents selected issues related to constructing predictive systems for medical applications. It is not an exhaustive survey but

a summary of important data quality and model evaluation issues partially mentioned in the authors' previous work [7]. It also omits fundamental issues related to the learning stage of the model, as well as problems associated with the choice of its architecture, the problem of using data from other domains for initial learning and then applying transfer learning techniques, its positive or negative impact on the quality of the model. There is also no focus on the problem of the usually small number of available learning objects and the possibility of using semi-supervised learning techniques for this purpose, including those based on self-learning. We also do not concentrate on implementation issues related to running trained models on environments with limited computational resources, which requires, for example, using techniques based on model pruning in the case of deep neural networks.

Correctly Define the Characteristics of the Designed System

As part of the first stage, it is necessary to define the purpose of the designed system correctly and, above all, what characteristics it should have. Incorrect identification of these characteristics can lead to the result not being satisfactory to the end user.

Among the most important features is the granularity of the decision, i.e., what question the proposed classifier is supposed to recognize, e.g., whether it is supposed to answer the question of whether retinopathy is diagnosed or not, or if pathology is found, whether it is supposed to be a more accurate answer using the 5-point International Clinical Diabetic Retinopathy scale [8]. Properly defining the system's response also determines which quality metrics can be considered.

The choice of inadequate metrics may lead to improper prediction quality assessment. The most commonly used metric is accuracy. Unfortunately, although intuitively correct, this metric assessing the quality of a given classification system may lead to erroneous conclusions.

Moreover, seeking to maximize it without assessing its usefulness replicates the so-called Goodhart's Law, which states that when a metric becomes a target, it ceases to be a good metric. This is particularly evident in the case of so-called imbalanced data [9], where significant disparities exist among classes. Consider the case of some rare disease that occurs only for a small fraction of the population—say 0.01% (we will call it the minority class). If an algorithm always indicates that the disease is not present (deterministically indicating that the patient does not suffer from the ailment, i.e., belongs to the majority class), it would achieve an accuracy of 99.99%. The problem with unbalanced data may also be encountered in ophthalmology where, for example, Proliferative Diabetic Retinopathy (PDR) is more than twenty times less common than Nonproliferative Diabetic Retinopathy (NPDR).¹ Therefore, when assessing the quality of systems solving diagnostic problems, including these diagnoses, it is necessary to consider the disproportion in the number of affected populations and the cost associated with their incorrect diagnosis.

Hence, the use of accuracy as a prediction metric is not appropriate since, in the case of rare diagnoses, which, as a rule, have a higher cost of error than the others, we allow for the possibility of error in assigning to a given class those who do not suffer from a given disease (the so-called true negative rate, specificity), but would like to achieve a fairly high rate of correctly classified patients suffering from a given disease (true positive rate, recall, sensitivity). However, this requires defining a correct decision quality criterion, which will be based on a cost analysis of errors made between classes. While this approach looks easy to apply, it allows the formulation of a criterion in the form of the so-called overall risk, i.e., the expected value of the cost of misclassifications [10]. Unfortunately, it is challenging to define in practice because it requires knowledge of the

¹ <https://www.aao.org/education/topic-detail/diabetic-retinopathy-europe>.

so-called loss function, that is, an estimate of the costs associated with misclassifying objects belonging to a given class. In practice, this cost is tough to estimate. For this reason, popular simple metrics are those that allow assessing the quality of a decision from the point of view of a given class (for a two-class classification task, these are, respectively, specificity for the majority class and sensitivity, also called recall for the minority class), or the precision indicating how precise the classification of minority class objects is. However, many prefer that a classification system be evaluated based on a single metric. So, a whole range of aggregate metrics is based on the simple metrics mentioned earlier, such as F_β -measure, Balanced Accuracy, and G-mean [11] have emerged. Unfortunately, this approach results in aggregate metrics that are not unambiguous, i.e., for different values of simple metrics, they could take the same value, and the use of these metrics during the selection of the best model may lead to bias towards the majority class [12]. A partial solution to this problem is using parametric metrics, such as mentioned F_β -measure, in which the importance of each metric could be controlled. However, the problem here may be obtaining information from the user about the expected impact of each simple metric [13].

At the stage of defining the problem, it is also worth noting a crucial characteristic called *explainability* [14], i.e., answering why such a diagnosis was indicated and not another. This need arises from several considerations. In medical applications, there is a trend toward blind acceptance of *black-box* models (e.g., solutions based on neural networks), which will open the door for companies to sell more models to hospitals [15]. So-called *black-box* models, although often the accuracy of their decisions is very high, cannot explain them. So, is it legitimate to have high trust in their quality? One spectacular example is IBM Watson, the first version of which was negatively received by

the medical community, as many examples of unsafe and incorrect treatment recommendations for cancer were identified.²

Equipping AI systems with mechanisms to explain decisions results in staff using a given diagnostic system gaining trust in those decisions by understanding the rationale behind the proposed diagnosis [16]. On the other hand, explainability and interpretability allow for verification of the correctness of the system's operation and its testing [17]. Explainable AI (xAI) is one of Artificial Intelligence's most essential and fastest-growing areas today, whose compatibility should be required for any medical diagnostic support system [18].

Data Quality

The next stage of ML-based systems focuses on data collection. Here, too, we may have to deal with problems that, if not considered carefully, can negatively affect the quality of the obtained solution, e.g., it can lead to bias in the final solution. Bias can have various causes, which include bias in the data or algorithm bias, but also designer bias (human bias) [19]. For example, deep models that provide a high-quality diagnosis in the range of skin lesions [20] are often trained using images of white patients' skin lesion samples, while the fraction of black-derived data generally does not exceed 10% [21]. Jacob et al. [22] noticed a lack of diversity in available ophthalmic datasets, with 45% of the global population lacking readily available representative images, leading to potential misrepresentation of their unique anatomical features and ocular pathology.

One of the reasons for the occurrence of bias in the data is that it is imbalanced or that there is the collection of an unrepresentative sample in terms of the frequency of observations from different categories. In the case of natural data imbalance, one should use data preprocessing

²<https://www.statnews.com/wp-content/uploads/2018/09/IBMs-Watson-recommended-unsafe-and-incorrect-cancer-treatments-STAT.pdf>.

techniques based on over- or undersampling [9] or choose an appropriate method of learning a classifier that considers the disproportion between class sizes [23].

It is more challenging to deal with the case of an unrepresentative learning set, as this may require correcting the data acquisition process. Critical in bias detection are the xAI-based methods mentioned earlier. It is worth noting that one of the important current trends in AI is the so-called *fairness AI*, which, following Mehrabi et al., could be defined as the absence of any bias or favoritism towards an individual or group based on their innate or acquired characteristics when making decisions [24].

Essential from the point of view of ophthalmology is data labeling. Acquiring labels for medical problems is expensive, as it requires specialized knowledge. In ophthalmology, so-called graders assign labels [25], and, of course, the final quality of the learning set strongly depends on their diligence and knowledge [26]. So, the generally accepted model is that individual images are annotated not by a single expert but by a group of experienced ophthalmologists [27]. Unfortunately, significant intergrader variability is usually observed in the labeling process [28, 29]. Hence, label reconciliation by a group of experts is a critical issue. Several techniques are used, generally based on the Delphi method modifications that are used to arrive at a group decision by surveying a panel of experts, using majority voting rules, or $n+1$ annotation rule (i.e., n experts label each image and if they disagree, the labeling decision belongs to a so-called *super-expert*). A process closer to the traditional Delphi method is also possible, i.e., in case of disagreement, the experts discuss disputed cases and then decide on the most probable label. The process is repeated until the desired consistency of responses is achieved as measured with the agreement metrics, such as Cohen's kappa. Label agreements could be supported by dedicated software platforms that do not require online discussion [30].

This consensus-based approach seems to be recommended due to its higher reliability than establishing a final label based on the majority

voting rule. However, it is worth observing here that the assigned labels are not necessarily true. A report on the quality of medical diagnosis in the US, published in 2000, shows that misdiagnoses account for up to 13% of all diagnoses [31], so despite great care during the grading process, we may have to deal with the so-called label noise [32]. On the other hand, we should also pay attention to the property of machine learning methods, which generally have outstanding generalizing ability, i.e., producing high-quality models, even if the learning set is noisy [33].

A separate problem that needs to be addressed is that label inconsistency can be caused by the wrong or improper definition of the label by the adopted different standards in the definition of value thresholding, such as in the definition of normal values for hypertension [34].

Data labeling may also lead to the aforementioned data bias, as little attention is paid to the selection of objects for labeling, and the number of available instances is usually small because of the high annotation cost. The selection of images to be labeled is typically random, without paying attention to ensuring the representativeness of objects within each class and without due attention to the impact of a given object on the quality of the resulting model. Therefore, it should be recommended to develop a procedure to ensure the representativeness of the data selection for labeling, or one may use the active learning approach [35], which is a technique based on the *human-in-the-loop* paradigm, i.e., it is an active learning algorithm indicates which instances are interesting to be annotated for model building.

Of course, not all data collection issues are addressed here. Attention should be paid to the possibility of missing data, data noise, etc. A comprehensive guide to techniques that eliminate data defects is [36]. Additionally, the data received are often unstructured and contain errors. Hence, automatic inconsistency search techniques seem desirable, which may be supported by xAI techniques. Additionally, there is a lack of standardized image acquisition tools,

i.e., different types of fundus cameras whose optical parameters and final image quality may differ substantially. It is increasingly advocated to use smartphones' built-in cameras for ophthalmic examinations, replacing professional fundus cameras [37].

Model Evaluation

Especially in the case of medical applications, evaluating the quality of classification methods is often based on the results of a computer experiment [38]. Unfortunately, such a process, usually done by the authors of a given solution, violates the principle of correct testing. The author of a method should not run the evaluation process (this is contrary to the testing paradigm). It is because if someone is testing a manufactured solution, he or she may introduce unconscious or unintentional biases into the testing process, unjustifiably preferring his or her solution.

This is particularly evident when testing is not conducted reliably. An example of such an inappropriate approach was the first version of the aforementioned IBM Watson system, which, despite internal evaluation, proved useless from a practical point of view. Only a series of subsequent improvements, including training the model on a large number of real data, made it possible to achieve acceptable quality. In addition to being carried out by an independent body, the testing process should also consider the characteristics of the patients for whom it will be used.

Nowadays, in diagnostic software, more and more national or Community (such as EU) agencies certify its use in clinical practice. A given certification indicates that the software is compliant, e.g., with US Food and Drug Administration (FDA) regulations. That means that the outcome of a given software complies with the agency's compliance and quality management criteria, e.g., FDA (known as *FDA software validation*). However, it seems that the certification paths used are based on the results of evaluations performed on the basis of

collections that were collected to build specific solutions. Thus, in ophthalmology, in particular, there have been some reports indicating that independent results of quality testing of market solutions often differ significantly from those reported by manufacturers [39, 40]. This may be due to several reasons, including, among others, the fact that the developed algorithms, implied in the mentioned algorithms, are trained on data from a given geographic region and are presumably biased towards populations from that geographic area. Thus, when testing on data from individuals living in a different geographic region and thus exhibiting differences related to, e.g., eating habits, lifestyle, genetic factors, etc., the algorithms will be biased toward the population of that geographic area. This will de facto test the algorithm on data from a different distribution than that used for learning. Therefore, detecting such a phenomenon based on out-of-distribution detection (OOD) [41, 42] methods is essential. OOD should be part of a given algorithmic solution's certification path and apply it to a geographic region supervised by a given certification body (e.g., in Poland, the Ministry of Health is responsible for this process).

Data Distribution Can Change

The final stage in the life cycle of machine learning-based systems is monitoring their quality to ensure that a guaranteed prediction level is maintained during the exploitation period of the system. Again, it is essential to remember that the models implied in the diagnostic solutions are learned from data from a given distribution, which can change, and new diagnostic cases will come from a different distribution. Changes can be influenced by environmental factors, such as those related to the quality of food or the environment, and cultural changes, such as the increase in time spent using screen displays both at work and for entertainment. Hence, the commonly accepted assumption that training and test data are drawn as independent and identically distributed samples (IID) is not maintained. The change in data distributions is

called *concept drift* [43], and mechanisms for detecting and responding to this phenomenon must be provided at the monitoring stage [44]. The system's validity can be ensured by continuously adapting to changes using continual learning techniques [45]. However, despite the significant progress in this field, it seems that from the point of view of commercial applications, they have not gained widespread acceptance by the industry, and, as a rule, the validity of the algorithm updating is based on re-training the model from scratch.

Another issue related to data quality that should be mentioned is adversarial attacks [46]. Due to the vulnerability of deep neural networks to adversarial examples, hackers can modify a learning set to fool the classifier, i.e., deliberately manipulate machine learning models by inputting carefully crafted input data. Such learning instances fed to the classifier input are intended to deceive it. They are generated using advanced methods, and the change introduced by generating adversarial examples is incomprehensible to humans, e.g., a manipulated image may look like its original. However, it will be recognized differently by the classifier attacked during learning.

Most works analyze adversarial attacks in the context of cybersecurity [47], and it is also worth considering here what advantages the attacker could gain in the case of medical systems. An example application of this type of attack is, for example, altering data to obtain undue benefits from a healthcare system or insurance company [48]. An in-depth analysis of the use of adversarial attacks in medicine is presented in [49].

Conclusion

The chapter aimed to present selected issues related to the life cycle of systems based on machine learning, which may cause the resulting system to be out of line with medical user's expectations in the case of applications in medicine. The paper discussed the fundamental problems in terms of the expected characteristics of

the designed systems, including the classification quality, addresses the data bias problem, issues related to data labeling, and the possibility of using techniques to support the learning process with a small number of labeled data, and finally discussed problems related to data distribution changes, both caused by the issue of the data distribution shift and with intentional manipulation to confuse the classification system.

Failure to consider the aspects discussed in this work may also lead to an increased risk of incorrect use of such systems. Consequently, it may have a negative impact on the quality of medical diagnostics. It seems worthwhile to have an in-depth discussion to discover the real expectations of future users and an in-depth risk analysis of the scope discussed in the paper before starting work on such a solution.

Acknowledgements Michał Woźniak was supported by the CEUS-UNISONO programme, which has received funding from the National Science Center, Poland under grant agreement No. 2020/02/Y/ST6/00037.

References

1. van Melle W. MYCIN: a knowledge-based consultation program for infectious disease diagnosis. *Int J Man-Mach Stud.* 1978;10(3):313–22. <https://www.sciencedirect.com/science/article/pii/S0020737378800492>
2. Tsuji T, Hirose Y, Fujimori K, et al. Classification of optical coherence tomography images using a capsule network. *BMC Ophthalmol.* 2020;20:114. <https://doi.org/10.1186/s12886-020-01382-4>.
3. Hendrycks D, Mazeika M, Wilson D, Gimpel K. Using trusted data to train deep networks on labels corrupted by severe noise; 2018. https://proceedings.neurips.cc/paper_files/paper/2018/file/ad554d8c3b-06d6b97ee76a2448bd7913-Paper.pdf
4. Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, Peng L, Webster DR. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng.* 2018;2(3):158–64.
5. Gichoya JW, Banerjee I, Bhimireddy AR, Burns JL, Celi LA, Chen LC, Correa R, Dullerud N, Ghassemi M, Huang SC, Kuo PC, Lungren MP, Palmer LJ, Price BJ, Purkayastha S, Pyrros AT, Oakden-Rayner L, Okechukwu C, Seyyed-Kalantari L, Trivedi H,

- Wang R, Zaiman Z, Zhang H. AI recognition of patient race in medical imaging: a modelling study. *Lancet Digit Health*. 2022;4(6):e406–14.
6. Machine Learning Best Practices in Healthcare and Life Sciences, AWS Whitepaper, 2021. <https://docs.aws.amazon.com/whitepapers/latest/ml-best-practices-healthcare-life-sciences/ml-best-practices-healthcare-life-sciences.html>
7. Clark Lambert W, Lambert MW, Emamian MH, Woźniak M, Grzybowski A. Artificial intelligence and the scientific method: how to cope with a complete oxymoron. *Clin Dermatol*. 2024. <https://doi.org/10.1016/j.clindermatol.2023.12.021>
8. Wilkinson CP, Ferris FL 3rd, Klein RE, Lee PP, Agardh CD, Davis M, Dills D, Kampik A, Pararajasegaram R, Verdaguer JT, Global Diabetic Retinopathy Project Group. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology* 2003;110(9):1677–82.
9. Krawczyk B. Learning from imbalanced data: open challenges and future directions. *Prog Artif Intell*. 2016;5:221–32. <https://doi.org/10.1007/s13748-016-0094-0>.
10. Duda RO, Hart PE, Stork DG. *Pattern classification*. New York: Wiley; 2001.
11. Branco P, Torgo L, Ribeiro RP. Relevance-based evaluation metrics for multi-class imbalanced domains. In: Kim J, Shim K, Cao L, Lee JG, Lin X, Moon YS, editors. *Advances in knowledge discovery and data mining*. PAKDD 2017. Lecture Notes in Computer Science(), vol. 10234. Cham: Springer; 2017.
12. Brzezinski D, Stefanowski J, Susmaga R, Szczech I. On the dynamics of classification measures for imbalanced and streaming data. *IEEE Trans Neural Netw Learn Syst*. 2020;31(8):2868–78.
13. Hand D, Christen P. A note on using the F-measure for evaluating record linkage algorithms. *Stat Comput*. 2018;28(3):539–47.
14. Biecek P, Burzykowski T. *Explanatory model analysis: explore, explain, and examine predictive models*, 1st ed. Chapman and Hall/CRC; 2021. <https://doi.org/10.1201/9780429027192>
15. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell*. 2019;1:206–15. <https://doi.org/10.1038/s42256-019-0048-x>.
16. Holzinger A, Biemann C, Pattrichs C, Kell D. What do we need to build explainable AI systems for medical domains; 2017. [arXiv:1712.09923v1](https://arxiv.org/abs/1712.09923v1)
17. Lisboa P. Interpretability in machine learning—principles and practice. In: W: 10th International workshop on fuzzy logic and applications. Proceedings of WILF 2013, Springer LNAI, vol. 8256, 2013. pp. 15–21.
18. van der Velden BHM, Kuijf HJ, Gilhuijs KGA, Viergever MA. Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. *Med Image Anal*. 2022;79:102470. <https://doi.org/10.1016/j.media.2022.102470>
19. Norori N, Hu Q, Aellen FM, Faraci FD, Tzovara A. Addressing bias in big data and AI for health care: a call for open science. *Patterns (N Y)*. 2021;2(10):100347. <https://doi.org/10.1016/j.patter.2021.100347>. PMID:34693373;PMCID:PMC8515002.
20. Brinker TJ, Hekler A, Enk AH, Berking C, Haferkamp S, Hauschild A, Weichenthal M, Klode J, Schadendorf D, Holland-Letz T. Deep neural networks are superior to dermatologists in melanoma image classification. *Eur J Cancer*. 2019;119:11–7.
21. Kamulegeya LH. Using artificial intelligence on dermatology conditions in Uganda: a case for diversity in training data sets for machine learning; 2019.
22. Jacoba CMP, Celi LA, Lorch AC, Fickweiler W, Sobrin L, Gichoya JW, Aiello LP, Silva PS. Bias and non-diversity of Big Data in artificial intelligence: focus on retinal diseases. *Semin Ophthalmol*. 2023;38(5):433–41. <https://doi.org/10.1080/08820538.2023.2168486>. Epub 2023 Jan 18. PMID:36651834.
23. Gao L, Zhang L, Liu C, Wu S. Handling imbalanced medical image data: a deep-learning-based one-class classification approach. *Artif Intell Med*. 2020;108:101935. <https://doi.org/10.1016/j.artmed.2020.101935>
24. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A survey on bias and fairness in machine learning. *ACM Comput. Surv*. 2021;54(6):35 p., Article 115 (July 2022). <https://doi.org/10.1145/3457607>
25. Gulshan V, et al. Performance of a deep-learning algorithm versus manual grading for detecting diabetic retinopathy in India. *JAMA Ophthalmol*. 2019;137(9):987–93. <https://jamanetwork.com/journals/jamaophthalmology/fullarticle/2734990>.
26. Ruamviboonsuk P, et al. Deep learning versus human graders for classifying diabetic retinopathy severity in a nationwide screening program. *npj Digital Med*. 2019;2. <https://www.nature.com/articles/s41746-019-0099-8>
27. Grzybowski A, Rao DP, Brona P, Negiloni K, Krzywicki T, Savoy FM. Diagnostic accuracy of automated diabetic retinopathy image assessment softwares: IDx-DR and Medios artificial intelligence. *Ophthalmic Res*. 2023;66(1):1286–92. <https://doi.org/10.1159/000534098>. Epub 2023 Sep 27. PMID:37757777;PMCID:PMC10619585.
28. Liu X, Mitani A, Spitz T, Wu DJ, Ledsam JR. Google and DeepMind: deep learning systems in ophthalmology. In: Grzybowski A, editor. *Artificial*

- intelligence in ophthalmology. Cham: Springer. https://doi.org/10.1007/978-3-030-78601-4_12
29. Grzybowski A, Brona P, Krzywicki T, Gaca-Wysocka M, Berlińska A, Święch A. Variability of grading DR screening images among non-trained retina specialists. *J Clin Med*. 2022;11(11):3125.
 30. Schaeckermann M, et al. Remote tool-based adjudication for grading diabetic retinopathy. *Transl Vis Sci Technol*. 2019;8(40). <http://tvst.arvojournals.org/article.aspx?articleid=2757836>
 31. Institute of Medicine (US) Committee on Quality of Health Care in America. To err is human: building a safer health system. Kohn LT, Corrigan JM, Donaldson MS, editors. Washington (DC): National Academies Press (US); 2000. PMID: 25077248.
 32. Johnson JM, Khoshgoftaar TM. A survey on classifying Big Data with label noise. *J Data Inf Qual*. 2022;14(4):43 pages, Article 23 (December 2022). <https://doi.org/10.1145/3492546>
 33. Patrini G, Rozza A, Menon A, Nock R, Qu L. Making deep neural networks robust to label noise: a loss correction approach 2017;2233–41. <https://doi.org/10.1109/CVPR.2017.240>
 34. Philip R, Beaney T, Appelbaum N, et al. Variation in hypertension clinical practice guidelines: a global comparison. *BMC Med*. 2021;19:117. <https://doi.org/10.1186/s12916-021-01963-0>.
 35. Settles B, Learning A. Springer Cham. 2012. <https://doi.org/10.1007/978-3-031-01560-1>.
 36. Garcia S, Luengo J, Herrera F. Data preprocessing in data mining. *Intell Syst Ref Lib*. 2015;72:1–313.
 37. Mishra C, Tripathy K. Fundus camera. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan. <https://www.ncbi.nlm.nih.gov/books/NBK585111/>
 38. Stapor K, Ksieniewicz P, Garcia S, Woźniak M. How to design the fair experimental classifier evaluation. *Appl Soft Comput*. 2021;104:1–12. <https://doi.org/10.1016/j.asoc.2021.107219>.
 39. Lee AY, Lee CS, Hunt MS, Yanagihara RT, Blazes M, Boyko EJ. Multicenter, head-to-head, real-world validation study of seven automated artificial intelligence diabetic retinopathy screening systems. *Diabetes Care*. 2021;44:e108–9.
 40. Grzybowski A, Brona P. Analysis and comparison of two artificial intelligence diabetic retinopathy screening algorithms in a pilot study: IDx-DR and retinalyze. *J Clin Med*. 2021;10(11):2352.
 41. Teney D, Kafle K, Shrestha R, Abbasnejad E, Kanan C, van den Hengel A. On the value of out-of-distribution testing: an example of Goodhart's law. In: Proceedings of the 34th international conference on neural information processing systems (NIPS'20). 2020; Red Hook, NY, USA: Curran Associates Inc., Article 35, pp. 407–17.
 42. Yang J, Zhou K, Li Y, Liu Z. Generalized out-of-distribution detection: a survey; 2021. [arXiv:2110.11334](https://arxiv.org/abs/2110.11334).
 43. Krawczyk B, Minku LL, Gama J, Stefanowski J, Woźniak M. Ensemble learning for data stream analysis: a survey. *Inf Fusion*. 2017;37:132–56.
 44. Komorniczak J, Zybiewski P, Ksieniewicz P. Statistical drift detection ensemble for batch processing of data streams. *Knowl-Based Syst*. 2022;252:109380. <https://doi.org/10.1016/j.knsys.2022.109380>.
 45. Mai Z, Li R, Jeong J, Quispe D, Kim H, Sanner S. Online continual learning in image classification: an empirical survey. *Neurocomputing*. 2022;469:28–51. <https://doi.org/10.1016/j.neucom.2021.10.021>.
 46. Goodfellow IJ, Shlens J, Szegedy C. Explaining and harnessing adversarial examples. *ICLR (Poster)*; 2015. <https://arxiv.org/abs/1412.6572>
 47. Kozal J, Zwolińska J, Klonowski M, Woźniak M. Defending network IDS against adversarial examples with continual learning. In: *ICDMW proceedings*. 2023, pp. 60–69. <https://ieeexplore.ieee.org/abstract/document/10411601>
 48. Paschali M, Conjeti S, Navarro F, Navab N. Generalizability versus robustness: adversarial examples for medical imaging. *arXiv* 2018; [arXiv: 1804.00504](https://arxiv.org/abs/1804.00504)
 49. Zbrzezny AM, Grzybowski AE. Deceptive tricks in artificial intelligence: adversarial attacks in ophthalmology. *J Clin Med*. 2023;12(9):3266.

Experimental Artificial Intelligence Systems in Ophthalmology: An Overview

7

Luis de Sisternes, Joelle A. Hallak, Kathleen Romond,
and Dimitri T. Azar

Abstract

This chapter provides a comprehensive overview of experimental artificial intelligence (AI) systems in ophthalmology, emphasizing their applications in diagnostics, disease classification, and clinical workflow optimization. Highlighted areas include keratoconus detection, refractive surgery screening, and automated imaging analysis, with advanced techniques such as convolutional neural networks (CNNs), multi-view data integration, and generative models showing promising results. The chapter also explores the potential of reinforcement learning in surgical simulations, the integration of molecular and genomic data, and AI's transformative role in telemedicine. Despite progress, challenges such as model generalization across clinical settings and real-world validation remain critical for broader clinical adoption.

Keywords

Experimental artificial intelligence · Ophthalmology · Keratoconus detection · Refractive surgery screening · Automated imaging analysis · Convolutional neural networks · Multi-view data integration · Generative models · Reinforcement learning

Introduction

Artificial intelligence (AI) research and applications in ophthalmology have been expanding in recent years [1, 2]. With advanced computing, we are able to derive novel insights from multisource data. The majority of AI research includes utilizing deep learning techniques to develop automated methods for disease classification and to help clinicians with identifying progression early. Elsewhere in this book, the various forms of clinically relevant AI techniques have been described. Neural networks consist of digitized inputs, text, images or measurements of varying nature, which proceed through connected layers that progressively detect features, with an ending predictive output (Fig. 7.1) [3]. Learning how to produce this desired predictive output is done automatically through a series optimization, backpropagation and update of layer weights. Deep learning

L. de Sisternes · J. A. Hallak · K. Romond
Department of Ophthalmology and Visual Sciences,
University of Illinois at Chicago, Chicago, IL, USA

K. Romond
e-mail: kromon2@uic.edu

L. de Sisternes · D. T. Azar (✉)
Twenty/Twenty Therapeutics, San Francisco, CA, USA
e-mail: dazar@uic.edu

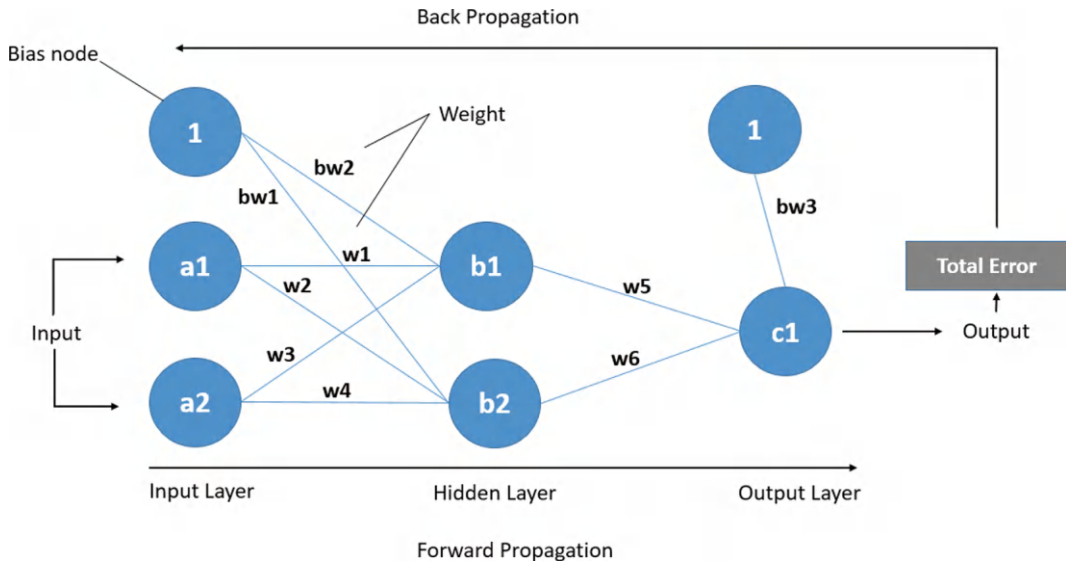


Fig. 7.1 A. Simple neural network and its components. Reconstructed from Taylor M. Neural Networks Math: Visual

(DL) consists of a large-scale network with multiple hidden layers (Fig. 7.2). Researchers are encouraged by performance results from applying convolutional neural nets (CNNs), which analyze pixel-level information for classifying diseases and their severity [4–6]. Natural language processing techniques are also being used to harness data from electronic health records. It is expected that AI will assist ophthalmologists in diagnosing cases at a quicker pace, help with cases that are difficult to diagnose, and in predicting progression for early interventions. Additionally, AI models are currently developed to optimize and develop new features for commercial imaging and non-imaging systems ophthalmologist routinely use, for example, to perform tasks more quickly adapting to the typical use of the system, overcome the hardware limitations in terms of image quality or resolution, optimize clinical workflows, or offer a more comprehensive representation of multitude forms of patient data for faster analysis and review.

AI tasks in ophthalmology include interpretation of imaging techniques (visual fields, [7] optical coherence tomography (OCT), [8] and fundus photos [9]), integration of molecular and genomic data, [10] mapping between

functional and structural information, [11] and data management from electronic health records (Iris Registry) [12]. Many of these applications are now part of routine clinical care, like automated segmentation and quantification of biomarkers in OCT images, with especial interest set in screening applications, such as for diabetic retinopathy and retinopathy of prematurity. Although a high diversity of AI applications for ophthalmology have been proposed and validated in a research setting, most have not yet been translated to therapeutic interventions in ophthalmic clinics. One of the challenges in ophthalmology is that diagnosis relies on information from multiple imaging and clinical data sources. Glaucoma specialists, for example, rely on visual fields for functional changes, fundus photos and OCT for structural changes, and intraocular pressure measurements for clinical changes. Multi-view representation learning is required for ophthalmic diseases that combine information from fundus, optical coherence tomography (OCT), and text/patient metadata. Additionally, validation on real-world data from multiple clinical settings is also a prerequisite requirement prior to translating applications to patient care. AI models can only learn from data that has been presented during a training phase,

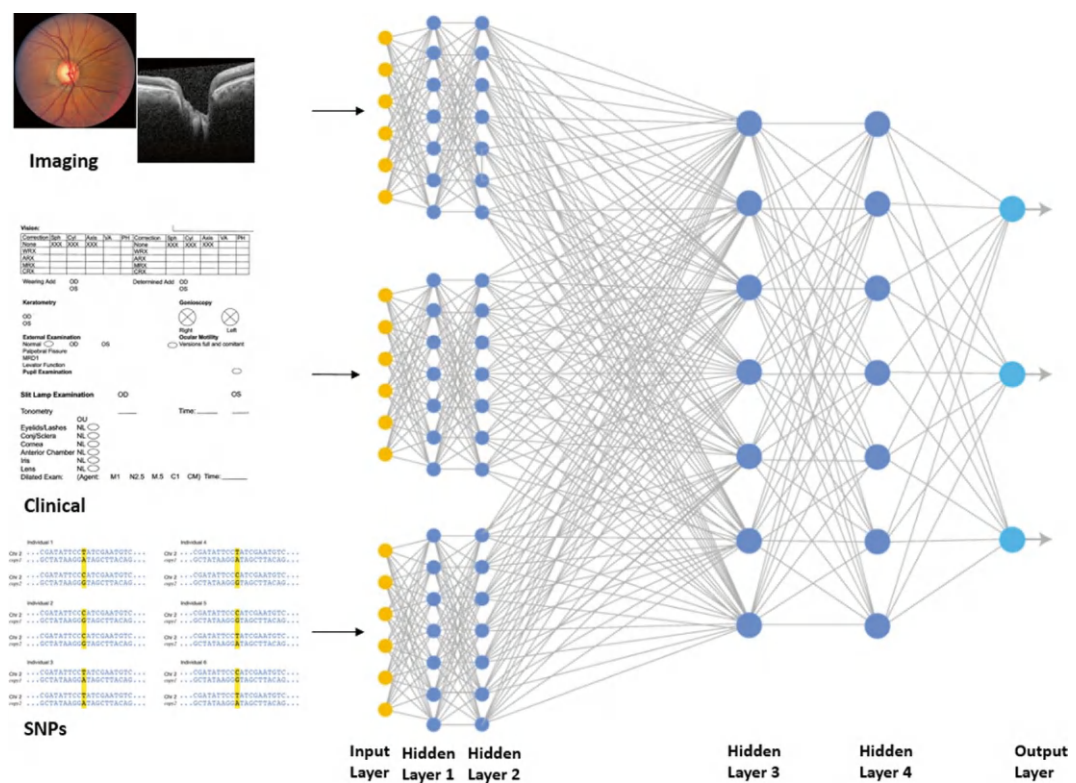


Fig. 7.2 Example of a large-scale network with a variety of ophthalmic data types as input (imaging, clinical and genetic) to determine a certain output. Each data type learns a useful featurization in its lower-level

towers. The data from each tower is then merged and flows through higher levels, allowing the deep neural network to perform inference across data types [3]

while data acquired from separate clinics, with varying settings, population or instrument models can alter the behavior of the model in unexpected ways.

This chapter provides an overview of experimental AI systems in ophthalmology. Specifically, it discusses potential applications in keratoconus, refractive surgery, cataracts, strabismus, retinopathy of prematurity (ROP), age-related macular degeneration prediction (AMD) and neuro-ophthalmology. Reinforcement learning (RL) and inverse reinforcement learning (IRL), and their potential applications for surgical simulation applications are also discussed. Additionally, in light of the modern data-rich hyper-connected world, we conclude by highlighting the impact and potential role of AI systems in ophthalmology for telemedicine and clinical workflow optimization.

AI Applications in Anterior Segment Disorders' Diagnosis

AI applications have a decades-long history in corneal topography interpretation, and their value enhancing clinical decisions for patients with corneal diseases has been demonstrated. Advancements in technology allow us to evaluate corneal curvatures and elevation maps, tissue anatomy, and histology and biomechanical properties. Machine learning (ML) and AI techniques are useful tools to help in corneal disease classification (normal, early suspect keratoconus and keratoconus), allowing early interventions, like collagen cross-linking, to prevent progression and vision loss. Additionally, ML and AI techniques provide useful tools for refractive surgery screening, in vivo corneal morphology exams, and corneal surgeries.

Keratoconus Screening and Classification

Automated classification of keratoconus subjects using ML models (including deep learning) has been an active area of research using information and images acquired from multiple devices, like corneal topographers or OCT devices.

Maeda et al. were one of the earliest groups to apply learning techniques on imaging data for keratoconus automated classification [13]. They used videokeratoscope data and a classification tree combined with a linear discriminant function to distinguish between a keratoconus and a non-keratoconic pattern. Smolek and Klyce were the first to use a classification neural network for keratoconus screening to detect the presence of keratoconus or keratoconus suspects [14]. They reported that the neural networks completely distinguished keratoconus subjects from suspects and from topographies that resembled keratoconus. Additionally, their network approach equaled the sensitivity of currently used tests for keratoconus detection and outperformed them in terms of accuracy and specificity [14].

Viera and Barbosa [15] showed that Zernike polynomials are reliable parameters as inputs in a feed-forward artificial neural network and a discriminant analysis technique, with reported precisions of 94% and 84.8%. Whereas Kovacs et al. [16] used Tomographic data, topographic data, and keratoconus indices from the Scheimpflug camera. They reported that classifiers trained on bilateral data were better than unilateral single parameter data in discriminating fellow eyes of patients with keratoconus from controls (ROC 0.96 versus 0.88). However, this result may be due to including one eye in the training set and the second from the same patient in the test set.

Several studies combined parameters from more than one device. Hwang et al. [17] extracted variables from slit-scan tomography and spectral-domain OCT to differentiate between normal controls and the clinically normal fellow eyes of highly asymmetric keratoconus subjects. The best discrimination was

found when using a combination of variables from both instruments, with spectral-domain OCT corneal thickness measures followed by anterior corneal measures from tomography as being most important. Ambrosio et al. [18] used a tomographic and biomechanical index (TBI), which combined Scheimpflug based corneal tomography and biomechanics. Their multinational retrospective study employed several AI techniques, including logistic regression, a support vector machine (SVM), and random forest (RF), all verified by leave one out cross validation (LOOCV) to distinguish normal versus ectatic cases. The RF model performed the best of all trained and tested methods, and was successful in detecting subclinical (fruste) ectasia, with an area under receiver operating characteristic curve (AUROC) for all ectasia cases of 0.996. Furthermore, the TBI cutoff value of 0.79 showed 100% sensitivity and specificity for detecting clinical ectasia [18].

More recently, some studies have been using deep learning methods for the detection of keratoconus. Kamiya et al. [19], used a deep learning model on colour-coded maps obtained by the anterior segment optical coherence tomography (AS-OCT) to discriminate keratoconus from normal corneas, with classifying the grade of the disease. Similar deep learning methods using OCT data have also shown impressive performance distinguishing between healthy and pathological corneas [20, 21].

Generative AI models also have a role in cornea diagnosis, generating additional data that can be used for further training purposes. A study by Agharezaei et al. [22] evaluated the accuracy of a collection of CNN models distinguishing keratoconus from normal subjects with regular astigmatism using cornea pattern images from a corneal topographer. They employed variational autoencoders (a technique used to generate synthetic image representations that can become nearly indistinguishable from real images) to augment the training data, demonstrating that the utilization of synthesized images during training process can increase classification performance, with the most accurate deep learning models achieving an AUROC of 0.99.

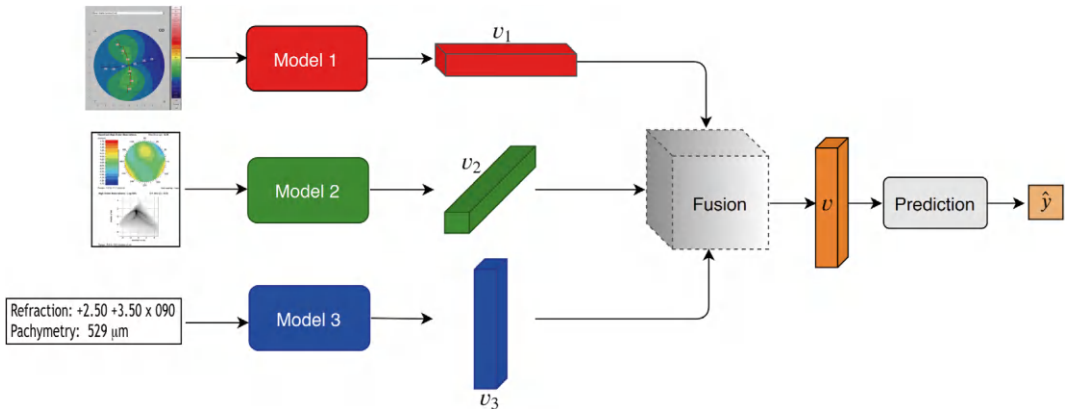


Fig. 7.3 Example of a data integration technique used for multi-view data in refractive surgery for modeling predictions

Corneal Refractive Surgery

Ocular imaging technology has evolved in recent years to address candidacy issues in corneal refractive surgery. A preoperative refractive surgery exam results in imaging, text, and numeric data. These data are a perfect example for multi-view and multi-task learning algorithms, with data integration, feature selection and modeling applications (Fig. 7.3).

Neural networks in refractive surgery applications have been designed to predict surgical outcomes and possible postoperative complications, [23] estimating implantable intraocular lens vault [reference], [24] and for surgery candidate screenings [25, 26]. Balidis et al., developed an algorithm to predict the need for retreatment in patients undergoing refractive surgery for myopia [25]. They used a computerized query to select patients who underwent PRK, LASEK, Epi-LASIK, or LASIK, and investigated 13 factors related to refractive surgery (such as age, room temperature and humidity, astigmatism axis, stromal thickness, laser ablation method, keratometry values, and laser characteristics etc.). After data preprocessing, a learning vector quantizer (LVQ) neural network was employed for classification. LVQ has a non-linear classification property, which is preferred for this task. They reported a sensitivity of 0.88 and specificity of 0.93 [25]. Additional

studies have developed learning algorithms for refractive surgery screening. Using pentacam images as input, Xie et al. developed a deep learning model for screening of candidates for refractive surgery [27]. A total of 6465 corneal tomographic images of 1385 patients from Zhongshan Ophthalmic Center, Guangzhou, China were used to develop the AI model. Their model, Pentacam InceptionResNetV2 Screening System (PIRSS), achieved an overall detection accuracy of 0.947 on the validation data set, and on the independent test data set it achieved an overall detection accuracy of 0.95, which was comparable with that of senior ophthalmologists who are refractive surgeons 0.928 [27]. In another recent study, Yoo et al. developed a machine learning architecture system integrating multi-data from patient demographics, pentacam imaging, and ophthalmic examinations to identify candidates for refractive surgery [26]. Five algorithms were used for predictions, SVM, artificial neural networks, random forests, LASSO, and AdaBoost to classify normal controls versus ectasia-risk patients. An ensemble classifier built from the predictions of these models achieved an AUC of 0.983 and 0.972 in internal and external validation sets, respectively [26].

Recent research efforts involve modifying neural networks to increase AI interpretability. Yoo et al., developed a multiclass machine learning model that provides output selecting

the type of refractive surgery procedure a patient may be eligible for [28]. They constructed a multiclass XGBoost model to classify patients into four categories including laser epithelial keratomileusis, laser in situ keratomileusis, small incision lenticule extraction, and contraindication groups. The model was trained based on clinical decisions of experts and ophthalmic measurements. The SHapley Additive explanations technique was adopted to explain the output of the XGBoost model. They reported an accuracy of 0.81 and 0.789 when testing on the internal and external validation datasets, respectively. The SHapley Additive explanations for the results were consistent with prior knowledge from ophthalmologists [28].

Cataract Diagnosis and Grading

The use of DL algorithms for automatic grading of cataracts has been explored using photographs from slit-lamp microscopes. An early study digitized images before extracting gray-level statistics from within the relevant circular regions of the nucleus [29]. These extracted features were fed into a neural network to produce classifications. Other approaches include Li et al.'s use of support vector machine regression to predict grade of cataract [30]. The authors introduced the first automatic system for nucleus region detection in slit lamp images. Local features were extracted on the basis of anatomical landmarks and validation of more than 5000 images against a clinical ground truth achieved a structure detection success rate of 95% with an average grading difference of 0.36 on a 5.0 scale [30]. Further improvements have since been reported, including Srivastava et al.'s proposal of extracting features representing the visibility, or lack thereof, of lens parts [31]. While previous works had focused on brightness and color metrics of the eye, these authors used features which represented the fading of edges of distinct anatomical regions as the severity of nuclear cataract increased. The integration of these new features with previously used features provided better accuracy than either feature type used alone [31].

Limitations to DL techniques, including incomplete, redundant or noisy representations, have been addressed in more recent work. One study addressed these challenges with their system which first automatically learned features for grading cataract through local filters and then predicted cataract grade using support vector regression [32]. The filters were acquired by clustering patches of images showing lenses from the same grading class and then fed through a convolutional neural network as well as a set of recursive neural networks to produce higher order features. Validation using these selected features on a population-based dataset of over 5000 images showed a mean absolute error of 0.304 [32].

Studies which examine the use of DL algorithms as a means of accurate diagnosis and monitoring of cataract among vulnerable populations have shown similarly promising results, including Liu et al.'s presentation of a convolutional neural network to grade slit-lamp images among pediatric patients [33]. Their algorithm offered excellent mean accuracy, sensitivity and specificity for classification (97.07, 97.28, and 96.83%). Zhang et al. [34] discussed the need of an automatic diagnosis aid for rural populations in China that lack access to expert care. Their study used fundus images, a more accessible imaging modality than slit-lamp, as input to their six-level cataract grading system and achieved an average accuracy of 92.66%. Along this line, advances in technology and AI models are leading to a paradigm shift in cataract patient referral. In work by Wu et al. [35] they established a new tertiary healthcare referral pattern that included three monitoring levels: self-management, primary healthcare, and secondary specialized services. A web-based system is used by patients to upload self-information and suspicious cases are detected and notified to attend community-based facilities, where a slit lamp image can be obtained and uploaded to the AI platform to be automatically classified. Patients and doctors are then immediately informed by the platform for cataracts classified as a "referral" or when a comprehensive examination is needed. The proposed

tertiary referral system was found to be able to increase the ophthalmologist-to-population service ratio by 10.2-fold compared to traditional cataract management [35].

AI Applications in Pediatric Ophthalmology

The most significant advances in AI applications for pediatric ophthalmology involve the automated detection of retinopathy of prematurity. Reid et al. summarize AI applications in pediatric ophthalmology [36]. In addition to ROP, machine learning has also been applied to the classification of pediatric cataracts, prediction of postoperative complications following cataract surgery, detection of strabismus and refractive error, prediction of future high myopia, and diagnosis of reading disability. In addition, techniques have been used for the study of visual development, vessel segmentation in pediatric fundus images, and ophthalmic image synthesis.

Detection of Strabismus

Strabismus is a common affliction, affecting approximately 4% of the population [37]. Often diagnosed in infancy or childhood, the disease presents as misalignment of the eyes so that each eye looks in different directions. Causes include problems with the eye muscles, the nerves of the eye in charge of transmitting information, the area of the brain which controls eye movements, or other eye injuries or disease. Untreated, it may lead to amblyopia, impairment of depth perception and 3D vision, or permanent vision loss, as the brain suppresses the image contributed by the weaker eye. Accurate and timely diagnosis is a topic of interest among researchers of late, as earlier diagnosis will allow for earlier treatment improving quality of life for patients not only in terms of visual outcomes, but through improved self-image and self-confidence. Traditional diagnosis may require several tests performed by and interpreted by a clinician, which may include the

Maddox Rod test, corneal light reflex test, and the commonly used, gold standard prism and alternate cover test (PCT). Level of expertise of the examiners has been shown to affect diagnostic decisions, [38, 39] bringing to light the need for a more objective and accurate approach through automatic methods. Techniques explored thus far have included image-based or video-based methods and eye-tracking methods.

Image and video-based approaches to automatic diagnostic aids have included Yang et al.'s [40] prospective observational pilot study on 30 intermittent exotropia patients, 30 esotropia patients and 30 orthotropic patients who were able to cooperate with a PCT. Two ophthalmologists independently performed the PCT for each subject to examine the angle of deviation. Using an infrared camera, full-face photos were taken with a selective wavelength filter placed over either eye. The resulting images were then fed through a 3D strabismus photo analyzer and correlations between the angles of deviation were compared between the automatic and manual methods which showed an excellent positive correlation ($R=0.900$, $P<0.001$) [40]. Another study selected features and classified images from a digital camera using a supervised support vector machine learning algorithm to mimic the diagnosis process of the Hirschberg test, which calculates the magnitude of yaw in the non-fixating eye by measuring the luminous reflection displacement of the cornea [41]. Their fully automated system was able to find the region of the eyes in images, locate the limbus and area of brightness and classify the images among previously diagnosed cases and controls, achieving 94.14% sensitivity, 95.38% specificity, 98.78% for PPV (Positive predictive value) and 83.07% for NPV (Negative predictive value). Chandna et al. [42] used measures obtained from prisms and cover tests as input to their backpropagation neural networks to produce a differential diagnosis. Despite an average reported accuracy of 100%, their system was limited to vertical strabismus and a series of manual measurements were needed as model input.

Several other studies have presented algorithms with their own limitations [43–45]. These

studies, along with the studies described above, used equipment that would be difficult to obtain or prohibitively expensive at some clinics, only diagnosed one direction of strabismus, or was not sensitive enough to detect unapparent strabismus. Valente et al. [46] attempted to address these limitations in their 2017 study, which used digital video to detect strabismus through the cover test. The only materials needed were a digital camera and regular computer for image and video processing and their methods achieved 87% accuracy while acknowledging measures lower than 1Δ .

Several eye tracking techniques have also been proposed. Corneal light reflection was used in one early animal study to characterize binocular misalignment in macaque monkeys with strabismus through measuring eye alignment errors to fixation targets which were presented throughout the subject's field of gaze at any distance [47]. The method allowed for horizontal and vertical error measurements and showed similar accuracy when compared to standard prism and cover assessments. Another eye tracking system automated a Hirschberg test for infant subjects using a two-camera gaze-estimation system [48]. The Hirschberg ratio and angle Kappa (the angle between the optical and visual axis) were determined through measurements of optical axis direction, corneal reflexes and coordinates of the center of the entrance pupil when the infants looked at images presented on a computer monitor. This method was only tested on five infants and needed further verification [48]. Chen et al. [49] used an eye-tracking system, Tobii X2-60, to collect gaze data. Subjects were asked to look at 9 points on a screen while the tracker detected the gaze points and eye movements. Gaze deviation maps were created and combined to form a gaze deviation image which was fed into a convolutional neural network for categorization. The best performing model of the 6 attempted, achieved 96% sensitivity and 94.1 specificity [49]. Lu et al.'s deep neural network for telemedicine applications was proposed as an automated solution to provide diagnosis for patients living in remote communities and to reduce the burden of on-site specialists

[50]. Their algorithm showed excellent results, with a reported detection performance of 0.933 sensitivity, 0.96 specificity and 0.9865 AUC. Quantitative methods to detect strabismus through an eye-tracking system are since a reality in routine clinical care using a virtual reality headset where stimuli are shown in different eyes alternatively while gaze data is constantly recorded. These devices have shown an intra-test variability measuring prism diopter displacement to alternating cover/uncover response within 1.1Δ [51].

Retinopathy of Prematurity

Retinopathy of Prematurity (ROP) accounts for 6–18% of childhood blindness, worldwide [52]. Early identification of the disease and treatment has shown better visual outcomes, and therefore accurate and timely diagnosis is critical [53]. There is variability in the diagnostic process for ROP and inconsistent classification agreement (plus vs. pre-plus vs. normal) among clinicians has been observed [54, 55]. Challenges which may precipitate variation in clinical diagnoses include geographic differences in training, vagueness of the definition of ROP, differing cut points on the continuous spectrum of vascular abnormality, as judged by clinicians, and a currently used standard published photograph for plus disease from the 1980s showing a smaller and more magnified view compared to currently available images [56]. As discussed elsewhere in this book, AI and ML techniques have been explored as assistive tools to aid with these challenges and improve diagnostic accuracy.

First attempts at automatic and semi-automatic imaging analysis used wide-angle RetCam images and focused on quantification of dilation and vascular tortuosity for plus-disease diagnosis. The diagnostic decisions of these systems were compared to expert diagnosis, and clinical applications were not realized because of either limitations in usability, or lack of agreement with the experts [57–59]. In particular, an ML system proposed by Ataer-Cansizoglu et al. [60]. (Imaging & Informatics in ROP, i-ROP)

showed promising classification capabilities by comparing different cropped shapes and sizes of images, as well as extracted tortuosity and dilation features. Their algorithm showed high diagnostic accuracy (95%) using a large circular 6 disc-diameter crop of wide-angle RetCam images and a metric which combined arterial and venous tortuosity, which was shown to be comparable to 3 experts (96, 94, and 92%) and higher than the mean accuracy of 31 non-experts (81%). Despite the algorithm performing well, manual segmentation of the images limited usability in a real-world setting [60].

Brown et al. [61], presented their fully automated i-ROP DL system in 2018, which was developed to provide a 3-level plus disease diagnosis in ROP patients and performed well on both internal and external validation sets. 5511 RetCam photographs were collected over a 5-year period and U-Net vessel segmentation and a pretrained Inception-V1 technical network were used on the images. The algorithm was compared to gold standard diagnosis performed by an ophthalmic examination by 1 expert and image analysis by 3 experts. A fivefold cross-validation showed AUCs of 0.94 and 0.98 for normal versus pre-plus and plus, and plus versus normal and pre-plus classifications, respectively. External validation on an independent dataset of 100 wide-angle RetCam images continued to show excellent specificity and sensitivity with plus classification exhibiting 93% sensitivity and 94% sensitivity and preplus or worse exhibiting 100% sensitivity and 94% specificity [61].

Building on Brown et al.'s work, Redd et al. [62] used the same DL system to classify images and produce a probability based 9-level disease severity scale. The reference standard diagnosis integrated image-based and ophthalmic diagnoses from expert clinicians, 4861 examinations from 870 infants were included in the analysis and showed an AUC of 0.960 and 94% sensitivity for detecting type-1 ROP and an AUC of 0.910 for clinically significant ROP, showing the system's ability to recognize broad as well as specific categories of the disease. Furthermore, the i-ROP DL severity score correlated with disease severity decided by expert graders, lending

evidence that i) ROP phenotypes present on a continuum of severity from mild to severe and ii) this severity can be measured automatically and accurately with their proposed system [62].

More recently, Mao et al. [63] presented their deep convolutional neural network (DCNN) which provided a diagnostic decision and analyzed pathological features of ROP to generate quantitative metrics for tortuosity, width, fractal dimension and vessel density. Three DCNNs were used in the network architecture. A modified U-net segmented the blood vessels while another segmented the optic disc. Three-level classification was performed by DenseNet. The authors reported a sensitivity and specificity of 95.1 and 97.8% for diagnosis of plus disease and 92.4 and 97.4% for pre-plus or worse [63].

Neuro-Ophthalmology Applications

Identifying abnormalities of the optic nerve can help to reveal vision-related neurological conditions. Detection of papilledema, for example, can alert clinicians to possible elevated intracranial pressure from life-threatening brain tumors or blood clots in the brain. Fundus photographs can capture changes to the optic nerve head, allowing for examination without the use of an ophthalmoscope, a tool which non-ophthalmologists may find difficult to use. AI systems have been used on fundus photo data to detect abnormalities in the optic nerve head and have been proposed as a diagnostic aid for patients with neuro-ophthalmic disease who may be presenting at non-neuro or non-ophthalmic health care clinics.

Optic Disc Abnormalities

Changes to the optic nerve head due to neuro-ophthalmic causes are rare, compared to optic disc abnormalities caused by glaucoma, and the ability of deep learning to aid in the detection of non-glaucomatous versus glaucomatous optic neuropathy has been explored. Yang et al.'s [64] convolutional neural network to perform such a task showed a diagnostic accuracy of 93.4%

sensitivity, 81.8% specificity and 0.874 AUC, with false positive cases exhibiting optic discs that were tilted or extensive peripapillary atrophy. The rarity of neuro-ophthalmic caused changes in optic disc appearance, versus glaucoma caused changes may contribute to neuro-ophthalmic misdiagnosis. Non-ophthalmic physicians may not be confident in identifying markers of neuro-ophthalmic disease when presented with images of optic disks and there is a resulting unmet need for an objective approach to abnormal optic disc diagnosis. Indeed, a 2019 review found high rates of misdiagnosis in neuro-ophthalmology in the literature with patients undergoing unnecessary, costly treatments and test before being seen by a neuro-ophthalmologist [65].

A few studies have attempted automatic detection of non-glaucomatous optic disc abnormalities, using ML techniques to diagnose and stage neuro-ophthalmic conditions by detecting optic disc irregularities, such as papilledema and optic disc atrophy, in fundus photographs. One early study graded papilledema severity using ML and decision tree forest classification to extract optic disc margins features, vascular features and peripapillary texture features from fundus photographs [66]. The system achieved substantial agreement with a severity grading ground truth provided by an expert neuro-ophthalmologist. In a subsequent study, a support vector machine (SVM) was used, alongside a hybrid feature extraction method, to achieve excellent accuracy (92.86%) of papilledema detection on a dataset of 160 fundus images, and even higher accuracy (97.85%) for grading papilledema images as 'mild' or 'severe' [67]. Fatima et al. [68] also presented a computer aided system to detect papilledema. After pre-processing of fundus photos to detect the optic disk and perform vessel segmentation, 26 features were extracted which represented optic disk change. The best features were selected within 4 categories, color, texture, and vascular and disc margin obstruction. Again, an SVM system was tested on 160 fundus images from two publicly available data sets, STARE and AFIO, achieving a 95.6% accuracy for the

STARE dataset, 87.4% for the AFIO dataset, and 85.9% for the combined data [68].

A deep learning system has been introduced to detect papilledema versus other abnormalities versus normal status, using fundus photographs from a multi-site, multiethnic dataset of over 15,000 photos. Reference [69] 14,341 images from 19 sites in 11 countries were used for training and internal testing, and 1505 images from 5 other sites were used for external validation. The internal testing showed a 0.99 AUC for both papilledema detection versus normal and other abnormalities, and normal status detection versus papilledema and other abnormalities. External validation resulted in an AUC of 0.96 for the detection of papilledema, with sensitivity and specificity of 96.4% and 84.7%, respectively [69].

Other studies attempting multi-level classification of optic disk images have included Ahn et al.'s proposed ML techniques for distinguishing between optic neuropathies and pseudo-papilledema, and Yang et al.'s identification of optic disc pallor [70, 71]. A deep learning system with the ability to determine laterality (ocular dominance) from fundus photo features was proposed recently. This algorithm could help with the laborious and time-consuming task of sorting and labeling images and was developed so that determination could be achieved for photos displaying normal or abnormal optic discs [72].

AI in Retina Imaging

Detection and quantification of pathologies in retinal images can help with diagnosis, appropriate disease staging and longitudinal analysis of a subject's retinal health. These pathologies (for example drusen or fluid pockets) can be difficult and tedious to perfectly identify in fundus photographs or OCT data and their quantification (in area or volume) requires precise delineation, which can be extremely time consuming to do by hand especially considering the large amount of data in OCT acquisitions. With this in mind, It has become common practice to use AI to automatically delineate and quantify several retinal landmarks and pathologies, like drusen,

[73] geographic atrophy (GA) [74–76] or edema [77]. Pioneering work by De Fauw et al. [78] [reference] developed a fully automated model for referral suggestion based on disease presence and severity using OCT volumetric images as input. In their approach, two deep learning neural networks were used: One trained to first automatically label the retinal layers and possible pathologies in the volumetric OCT data, and a second one trained with the output of this first network and the original OCT images to classify the acquisition in four referral categories: urgent, semi-urgent, routine and observation. Their work not only demonstrated a performance in making a referral recommendation that reaches or even exceeds that of experts on average, but their two-network approach also has the benefit of providing some explanation regarding how the automated referral decision was made: The results from the first neural network can be used to identify and delineate the pathologies used to make the decision, which can be reviewed by an expert.

Automated classification of retinal disease using imaging data has been one of the main applications of AI in ophthalmology. Work by Abràmoff et al. [79] pioneered the automated classification of diabetic retinopathy (DR) using fundus images and AI in a manner that can be applicable to routine clinical practice. Their approach reported a sensitivity of 0.97 and specificity of 0.87 distinguishing referable DR subjects from non-referable subjects. Work by Venhuizen et al. [80] also proposed for the automated classification of AMD images based on OCT data, reporting an AUC of 0.98. Additionally, one of the most interesting applications of AI in retinal images does not involve the detection of disease present in the images but the prediction of future disease that will be affecting the patient. This is possible by training with large amounts of data accumulated through several years where the outcome of a given patient is known. Applications using this AI predictive power range from the optimization of patient scheduling for routine follow up or possible therapy to making therapeutic decisions based on a subject specific predicted behavior.

Work by de Sisternes et al. was the first pioneering the prediction of a future exudation event in non-exudative AMD eyes at a given time frame [81, 82]. The goal was to be able to identify subjects at higher risk of exudation within a given time frame in order to schedule follow up visits and provide the needed therapy as soon as possible. In their work, they used a combination of quantifiable imaging features automatically extracted from OCT data together with demographic and contralateral eye status to predict the risk of a future exudation. Similar work predicting the appearance of AMD-related neovascularization or future exudation events using AI techniques has also been presented [10, 83].

Within the application of prediction of future retinal disease appearance or progression, work by Mai et al. [84] focused on the prediction of expansion of GA regions in AMD patients, both trying to estimate a rate of expansion and the exact locations within the retina where such expansion would manifest. While there are new developed drugs and treatments for GA, it is not very clear who can best benefit from them as GA can have very varied growth rates and patterns. The main benefits of being able to estimate GA expansion include allowing for optimal patient counselling and personalized treatment and identification of patients with different levels of disease activity and progression for the design and analysis of treatment studies [85]. While this and similar earlier work has been proposed, the prediction of the exact location and time where GA is going to expand for a given subject is still challenging. Using AI techniques, previous studies have been able to link this location of expansion to loss of photoreceptors, concentration of retinal hyperreflective foci, [86] or a pathological choriocapillaris [87]. AI models can be used to delineate and quantify such pathological regions automatically, and later use the collected quantifications over time into a separate model to produce a map of predicted expansion. While an accurate prediction of future atrophy location has not been clearly proven, considerable accuracy has been reported for predicting GA enlargement rates in individual lesions in mm/year [88, 89].

Additional work was presented by Zhou et al. [90], using self-supervised learning (SSL) to train two models, one for fundus imaging data and one for OCT data, that can generalize to a diverse range of complex tasks: Diagnostic classification of several ocular diseases; Prediction of exudation events in non-exudative AMD contralateral ('fellow') eyes within a 1-year time period; and the 3-year prediction of cardiovascular diseases (ischaemic stroke, myocardial infarction and heart failure) and a neurodegenerative disease (Parkinson's disease). Their approach outperformed other dedicated solutions, providing a generalizable solution that alleviates annotation workload while improving model performance to enable broad clinical AI applications from retinal imaging.

Reinforcement and Inverse Reinforcement Learning for Surgical Applications

Reinforcement learning (RL) and inverse reinforcement learning (IRL), or apprenticeship, are AI algorithms that are being explored for healthcare applications. Both methods may have some potential applications to surgery, surgical robotics, surgical training and assessment. In RL the goal is finding the optimal policy based on a defined set of outcomes (rewards). Once an optimal value function is known, it is possible to generate the optimal policy (surgical skill) for a given task from the value function (successful surgery defined in a concrete manner) [91]. RL algorithms learn by trial and error, trying multiple input sequences of interactions (histories) between the decision maker, in this case the surgeon, and their environment [92]. At every decision point, the RL algorithm chooses an action according to its learned policy and receives new observations and immediate outcomes (rewards). In contrast, in IRL, the algorithm tries to recover a reward function based on a defined optimal policy or trajectories given by an expert or another agent (mentor). In other words, in RL an agent attempts to maximize its rewards in an environment by self-learning what optimal

policy to follow, and in IRL an agent attempts to follow a "teacher agent" and does so under the assumption that the teacher agent is maximizing its rewards, even that such rewards cannot be easily defined concretely. IRL works for solving complex problems where the reward cannot be easily defined (in our case, a successful surgery without complications) by mimicking an expert action under given conditions. Once a reward function is learned through IRL, RL can also be used to optimize the actions to achieve the desired reward. RL and IRL may have potential in healthcare when learning requires optimizing a series of complex tasks to achieve a successful outcome or physician demonstration, for example in learning to suture wounds for robotic-assisted surgery [93]. As such, there is potential for applications in robotic-assisted surgery (RAS) in ophthalmology and training.

RL has the potential to optimize a series of motions enhancing RAS. Additionally, segmentation techniques can identify and delineate open wounds from imaging data, and a suturing method can be generated by finding an optimal trajectory taking into consideration external factors, such joint motions and obstacles. Image-trained RNNs can also be developed to tie knots autonomously by learning sequences of events, such as the surgeons' hand movement.

IRL applications may have potential applications for surgical training. These applications assume that the mentor has the same reward function as the observer/trainee and chooses from the same set of actions. The idea is then to infer the reward function of the mentor so as to produce the observed behavior. IRL assumes that the reward function is expressed as a linear function of known features. A prime example is teaching a person how to drive: rather than tell a young unseasoned driver what the reward function is, it is more natural to demonstrate to them how to drive [94]. Supervised methods have been developed to mimic the mentor. However, there are instances where blindly following the mentor's trajectory may not work because the environment encountered is different (consider the example of highway driving and varying traffic patterns). Abeel and Ng were the first to

apply IRL for apprenticeship learning simulation studies and reported that the policy found will have performance comparable to or better than that of an expert, on the expert's unknown reward function [94].

Challenges in applying learning algorithms for surgical applications will be more difficult to overcome and include correctly localizing an instrument's position and orientation in surgical scenes, data collection, adapting to completely unknown and unique situations, which may cause serious surgical errors [91]. Limiting these applications to simulation studies and training may be the safest approach. Ruzicki et al. [95] proposed a system to produce an objective analysis of the surgical skill (between trainees and expertise) of the steps in routine cataract surgery. In their approach, they employed and hand-tuned an existing image classification network (ResNet 152) to automatically identify each unique surgical instrument in microscope video frames acquired during surgery. The formulated encoding in their work was expected to contain information about instrument presence and pose for each frame, which is later used to map to a surgeon skill level based on a Random Forest classifier. While their model achieved high accuracy in tool detection and the corresponding surgical step of the frames, it failed to produce high accuracy results for skill classification. One of the limitations of their work was that skill level was based only on appointment status (either expert or trainee, or both expert and trainee), not including the particularities of the specific surgery frames being analyzed.

Groundbreaking work on applying AI robotics in microsurgery was proposed by Keller et al. [96]. In their work, they demonstrate the feasibility of using a combination of IRL (or learning from demonstration) and RL to train an industrial robot to perform OCT-guided ex vivo corneal needle insertions modeling anterior lamellar keratoplasty (DALK) surgery. DALK surgery requires high precision when positioning a needle tip deep enough within the corneal stroma without puncturing Descemet's membrane. The authors found this challenge perfectly suited for an AI robotics system, first

learning from demonstration of volumetric OCT time series recording of ex vivo DALK needle insertions performed by corneal surgical fellows, and then using a reinforcement learning agent trained on ex vivo human corneas. The result demonstrated that the trained system was able to significantly outperform surgical fellows in reaching a target needle insertion depth in mock corneal surgery trials. The mean and standard deviation of the final perforation-free percent depth obtained by the robot using the learned policy was $84.75\% \pm 4.91\%$ ($N=20$) compared to $78.58\% \pm 6.68\%$ for the surgical fellows using OCT ($N=28$).

Telemedicine and Clinical Workflow Optimization

As evidenced by the multiple AI studies and applications which require large datasets, ophthalmology has been at the forefront of big data utilization in healthcare research. These advancements have been possible due to the specialization of imaging modalities and ophthalmic instruments to provide objective detailed measurements and information that can be easily stored and transmitted through established protocols. Computerized data analysis, fast data transmission systems, and the need of clinical throughput optimization due to an aging population have all been contributing factors to the appearance of telemedicine applications for ophthalmology. In addition, COVID-19 has transformed all aspects of medicine, from clinical care and research to educational and public health activities. In addition, digitalization is taking a large role in all aspects of life, and it is also transforming ophthalmology practice with new digital instruments with possible diagnostic applications. These instruments can be more portable and easy to use, enabling patient care in remote places or optimizing clinical throughput by offering a more automated and objective alternative to measurements traditionally done by an operator in a subjective manner. For example, numerous portable digital fundus cameras and smartphone solutions are now

available, and their performance for automated DR screening has been studied with very promising results [97–99]. Another application where digitalization and AI are promising ophthalmic practice change is the use of virtual headsets to perform functional tests in an objective, automated and precise manner. These tests can range from standard tests typically done manually by operators (potentially decreasing human effort, time and costs while offering quantifiable and repeatable results) to generating full visual field perimetry tests, with some of them using eye telemetry in combination with AI for a more precise, objective and automated function response measurement [51].

We predict that telemedicine will be embraced more as a tool for healthcare delivery in the near future. Home monitors and smart devices, with AI embedded applications, and their connectivity to ophthalmic care providers will allow for immediate intervention. The need to overcome the hurdle of affordability and data privacy will be a priority for many of these applications. Additionally, challenges in translating deep learning applications due to limited explainability and interpretability need to be addressed to ensure targeted representation and to identify potential biases. AI researchers and clinicians should work together to increase interpretability and integrate AI systems in the clinical decision-making process in a responsible manner.

References

1. Ting DSW, Pasquale LR, Peng L, et al. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol*. 2019;103(2):167–75. <https://doi.org/10.1136/bjophthalmol-2018-313173>.
2. Hogarty DT, Mackey DA, Hewitt AW. Current state and future prospects of artificial intelligence in ophthalmology: a review. *Clin Exp Ophthalmol*. 2019;47(1):128–39. <https://doi.org/10.1111/ceo.13381>.
3. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. *Nat Med*. 2019;25:24–9. <https://doi.org/10.1038/s41591-018-0316-z>.
4. Treder M, Lauermaun JL, Eter N. Deep learning-based detection and classification of geographic atrophy using a deep convolutional neural network classifier. *Graefes Arch Clin Exp Ophthalmol*. 2018;256(11):2053–60. <https://doi.org/10.1007/s00417-018-4098-2>.
5. Brown JM, Campbell JP, Beers A, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. *JAMA Ophthalmol*. 2018;136(7):803–10. <https://doi.org/10.1001/jamaophthalmol.2018.1934>.
6. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316(22):2402–10. <https://doi.org/10.1001/jama.2016.17216>.
7. Wang M, Pasquale LR, Shen LQ, et al. Reversal of glaucoma hemifield test results and visual field features in glaucoma. *Ophthalmology*. 2018;125(3):352–60. <https://doi.org/10.1016/j.ophtha.2017.09.021>.
8. Lee CS, Baughman DM, Lee AY. Deep learning is effective for classifying normal versus age-related macular degeneration OCT images. *Ophthalmol Retina*. 2017;1(4):322–7.
9. Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology*. 2017;124(7):962–9. <https://doi.org/10.1016/j.ophtha.2017.02.008>.
10. Schmidt-Erfurth U, Waldstein SM, Klmscha S, et al. Prediction of individual disease conversion in early AMD using artificial intelligence. *Invest Ophthalmol Vis Sci*. 2018;59(8):3199–208. <https://doi.org/10.1167/iovs.18-24106>.
11. Mariottoni EB, Datta S, Dov D, et al. Artificial intelligence mapping of structure to function in glaucoma. *Trans Vis Sci Tech*. 2020;9(2):19.
12. Rich WL 3rd, Chiang MF, Lum F, Hancock R, Parke DW 2nd. Performance rates measured in the American Academy of Ophthalmology IRIS® Registry (Intelligent Research in Sight). *Ophthalmology*. 2018;125(5):782–4. <https://doi.org/10.1016/j.ophtha.2017.11.033>.
13. Maeda N, Klyce SD, Smolek MK, Thompson HW. Automated keratoconus screening with corneal topography analysis. *Invest Ophthalmol Vis Sci*. 1994;35(6):2749–57.
14. Smolek MK, Klyce SD. Current keratoconus detection methods compared with a neural network approach. *Invest Ophthalmol Vis Sci*. 1997;38(11):2290–9.
15. Viera and Barbosa de Carvalho LAV, Barbosa MS. Neural networks and statistical analysis for classification of corneal videokeratography maps based on Zernike coefficients: a quantitative comparison. *Arquivos Brasileiros de Oftalmologia*. 2008;71:337–41. <https://doi.org/10.1590/S0004-27492008000300006>.
16. Kovács I, Miháltz K, Kránitz K, et al. Accuracy of machine learning classifiers using bilateral data from a Scheimpflug camera for identifying eyes with

- preclinical signs of keratoconus. *J Cataract Refract Surg.* 2016;42(2):275–83. <https://doi.org/10.1016/j.jcrs.2015.09.020>.
17. Hwang ES, Perez-Straziota CE, Kim SW, Santhiago MR, Randleman JB. Distinguishing highly asymmetric keratoconus eyes using combined Scheimpflug and spectral-domain OCT analysis. *Ophthalmology.* 2018;125(12):1862–71. <https://doi.org/10.1016/j.optha.2018.06.020>.
 18. Ambrósio R Jr, Lopes BT, Faria-Correia F, Salomão MQ, Bühren J, Roberts CJ, Elsheikh A, Vinciguerra R, Vinciguerra P. Integration of Scheimpflug-based corneal tomography and biomechanical assessments for enhancing Ectasia detection. *J Refract Surg.* 2017;33(7):434–43. <https://doi.org/10.3928/1081597X-20170426-02>.
 19. Kamiya K, Ayatsuka Y, Kato Y, et al. Keratoconus detection using deep learning of colour-coded maps with anterior segment optical coherence tomography: a diagnostic accuracy study. *BMJ Open.* 2019;9(9): e031313. <https://doi.org/10.1136/bmjopen-2019-031313>.
 20. Al-Timemy AH, Alzubaidi L, Mosa ZM, et al. A deep feature fusion of improved suspected keratoconus detection with deep learning. *Diagnostics.* 2023;13(10):1689.
 21. Fassbind B, Langenbucher A, Streich A. Automated cornea diagnosis using deep convolutional neural networks based on cornea topography maps. *Sci Rep.* 2023;13(1):6566.
 22. Agharezaei Z, Firouzi R, Hassanzadeh S, et al. Computer-aided diagnosis of keratoconus through VAE-augmented images using deep learning. *Sci Rep.* 2023;13(1):20586.
 23. Kim J, Ryu IH, Kim JK, et al. Machine learning predicting myopic regression after corneal refractive surgery using preoperative data and fundus photography. *Graefes Arch Clin Exp Ophthalmol.* 2022;60(11):3701–10.
 24. Kamiya K, Ryu IH, Yoo TK, et al. Prediction of phakic intraocular lens vault using machine learning of anterior segment optical coherence tomography metrics. *Am J Ophthalmol.* 2021;226:90–9.
 25. Balidis M, Papadopoulos I, Malandris D, Zachariadis Z, Sakellaris D, Vakalis T, et al. Using neural networks to predict the outcome of refractive surgery for myopia. *4open.* 2019;2:29.
 26. Yoo TK, Ryu IH, Lee G, Kim Y, Kim JK, Lee IS, et al. Adopting machine learning to automatically identify candidate patients for corneal refractive surgery. *npj Digital Med.* 2019;2(1):59. <https://doi.org/10.1038/s41746-019-0135-8>
 27. Xie Y, Zhao L, Yang X, et al. Screening candidates for refractive surgery with corneal tomographic-based deep learning [published online ahead of print, 2020 Mar 26]. *JAMA Ophthalmol.* 2020;e200507. <https://doi.org/10.1001/jamaophthalmol.2020.0507>
 28. Yoo TK, Ryu IH, Choi H, et al. Explainable machine learning approach as a tool to understand factors used to select the refractive surgery technique on the expert level. *Transl Vis Sci Technol.* 2020;9(2):8. <https://doi.org/10.1167/tvst.9.2.8>.
 29. Duncan DD, Shukla OB, West, et al: New objective classification system for nuclear opacification. *J Opt Soc Am A Opt Image Sci Vis.* 1997;14:1197–1204.
 30. Li H, Lim JH, Liu J, et al. An automatic diagnosis system of nuclear cataract using slit-lamp images. *IEEE Trans Biomed Eng.* 2010;57:1690–8.
 31. Srivastava R, Gao X, Yin F, et al. Automatic nuclear cataract grading using image gradients. *J Med Imaging (Bellingham).* 2014;1: 014502.
 32. Gao X, Lin S, Wong TY. Automatic feature learning to grade nuclear cataracts based on deep learning. *IEEE Trans Biomed Eng.* 2015;62:2693–701.
 33. Liu X, Jiang J, Zhang K, Long E, Cui J, et al. Localization and diagnosis framework for pediatric cataracts based on slit-lamp images using deep features of a convolutional neural network. *PLoS ONE.* 2017;12(3): e0168606.
 34. Zhang H, Niu K, Xiong Y, et al. Automatic cataract grading methods based on deep learning. *Comput Methods Programs Biomed.* 2019;182: 104978.
 35. Wu X, Huang Y, Liu Z, et al. Universal artificial intelligence platform for collaborative management of cataracts. *Br J Ophthalmol.* 2019;103:1553–60.
 36. Reid JE, Eaton E. Artificial intelligence for pediatric ophthalmology. *Curr Opin Ophthalmol.* 2019;30(5):337–46. <https://doi.org/10.1097/ICU.0000000000000593>.
 37. Hertle RW. Clinical characteristics of surgically treated adult strabismus. *J Pediatr Ophthalmol Strabismus.* 1998;35(3):138–68.
 38. Anderson HA, Manny RE, Cotter SA, et al. Effect of examiner experience and technique on the alternate cover test. *Optom Vis Sci.* 2010;87:168–75.
 39. Hrynychak PK, Herriot C, Irving EL. Comparison of alternate cover test reliability at near in non-strabismus between experienced and novice examiners. *Ophthalmic Physiol Opt.* 2010;30:304–9.
 40. Yang HK, Seo JM, Hwang JM, Kim KG. Automated analysis of binocular alignment using an infrared camera and selective wavelength filter. *Investig Ophthalmol Vis Sci.* 2013;54:2733–7.
 41. De Almeida JDS, Silva AC, De Paiva AC, et al. Computational methodology for automatic detection of strabismus in digital images through Hirschberg test. *Comput Biol Med.* 2012;42:135–46.
 42. Chandna A, Fisher A, Cunningham I, et al. Pattern recognition of vertical strabismus using an artificial neural network (strabnet). *Strabismus.* 2009;17(4):131–8.
 43. Kim TY, Seo SS, Kim YJ, et al. A new software for quantitative measurement of strabismus based on digital image. *J Korea Multimedia Soc.* 2012;15(5):595–605.
 44. Seo MW, Yang HK, Hwang JM, Seo JM. The automated diagnosis of strabismus using an infrared camera. In: 6th European conference of the international Federation for medical and biological engineering 2015, vol. 45, pp. 142–5.

45. Khumdat N, Phukpattaranont P, Tengtrisorn S. Development of a computer system for strabismus screening. In: 6th Biomedical engineering international conference. IEEE 2013; pp. 1–5.
46. Valente TL, de Almeida JD, Silva AC, et al. Automatic diagnosis of strabismus in digital videos through cover test. *Comput Methods Programs Biomed.* 2017;1(140):295–305.
47. Quick MW, Boothe RG. A photographic technique for measuring horizontal and vertical eye alignment throughout the field of gaze. *Investig Ophthalmol Vis Sci.* 1992;33:234–46.
48. Model D, Eizenman M. An automated Hirschberg test for infants. *IEEE Trans Biomed Eng.* 2011;58:103–9.
49. Chen Z, Fu H, Lo WL, Chi Z. Strabismus recognition using eye-tracking data and convolutional neural networks. *J Healthc Eng.* 2018;2018:1–9.
50. Lu J, Feng J, Fan Z, et al. Automated Strabismus Detection based on Deep neural networks for Telemedicine Applications. 2018; Available at: <https://deepai.org/publication/automated-strabismus-detection-based-on-deep-neural-networks-for-telemedicine-applications>. Accessed 31 July 2020.
51. de Sisternes L, Goyal D, Azar DT, Sinha S. Validation of eye tracking-based visual tests using virtual reality headset. *Invest Ophthalmol Vis Sci.* 2023;64(8):2844.
52. Fleck BW, Dangata Y. Causes of visual handicap in the Royal Blind School, Edinburgh, 1991–2. *Br J Ophthalmol.* 1994;78(5):421.
53. Early Treatment for Retinopathy of Prematurity Cooperative Group. Good WV, Hardy RJ, Dobson V, et al. Final visual acuity results in the early treatment for retinopathy of prematurity study. *Archives Ophthalmol.* 2010;128(6):663.
54. Chan RP, Williams SL, Yonekawa Y, et al. Accuracy of retinopathy of prematurity diagnosis by retinal fellows. *Retina (Philadelphia, PA.).* 2010;30(6):958.
55. Myung JS, Chan RV, Espiritu MJ, et al. Accuracy of retinopathy of prematurity image-based diagnosis by pediatric ophthalmology fellows: implications for training. *J Am Assoc Pediatr Ophthalmol Strabismus.* 2011;15(6):573–8.
56. Ting DS, Peng L, Varadarajan AV, et al. Deep learning in ophthalmology: the technical and clinical considerations. *Prog Retin Eye Res.* 2019;1(72): 100759.
57. Koreen S, Gelman R, Martinez-Perez ME, et al. Evaluation of a computer-based system for plus disease diagnosis in retinopathy of prematurity. *Ophthalmology.* 2007;114(12):e59–67.
58. Wilson CM, Wong K, Ng J, Cocker KD, et al. Digital image analysis in retinopathy of prematurity: a comparison of vessel selection methods. *J Am Assoc Pediatr Ophthalmol Strabismus.* 2012;16(3):223–8.
59. Abbey AM, Besirli CG, Musch DC, et al. Evaluation of screening for retinopathy of prematurity by ROPtool or a lay reader. *Ophthalmology.* 2016;123(2):385–90.
60. Ataer-Cansizoglu E, Bolon-Canedo V, Campbell JP. Computer-based image analysis for plus disease diagnosis in retinopathy of prematurity: performance of the “i-ROP” system and image features associated with expert diagnosis. *Transl Vis Sci Technol.* 2015;4(6):5.
61. Brown JM, Campbell JP, Beers A, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. *JAMA Ophthalmol.* 2018;136(7):803–10.
62. Redd TK, Campbell JP, Brown JM, et al. Evaluation of a deep learning image assessment system for detecting severe retinopathy of prematurity. *Br J Ophthalmol.* 2019;103(5):580–4.
63. Mao J, Luo Y, Liu L, et al. Automated diagnosis and quantitative analysis of plus disease in retinopathy of prematurity based on deep convolutional neural networks. *Acta Ophthalmol.* 2020;98(3):e339–45.
64. Yang HK, Kim YJ, Sung JY, et al. Efficacy for differentiating nonglaucomatous versus glaucomatous optic neuropathy using deep learning systems. *Am J Ophthalmol.* 2020 2 April.
65. Stunkel L, Newman NJ, Bioussé V. Diagnostic error and neuro-ophthalmology. *Curr Opin Neurol.* 2019;32(1):62–7.
66. Echegaray S, Zamora G, Yu H, et al. Automated analysis of optic nerve images for detection and staging of papilledema. *Invest Ophthalmol Vis Sci.* 2011;52:7470–8.
67. Akbar S, Akram MU, Sharif M, et al. Decision support system for detection of papilledema through fundus retinal images. *J Med Syst.* 2017;41:66.
68. Fatima KN, Hassan T, Akram MU, et al. Fully automated diagnosis of papilledema through robust extraction of vascular patterns and ocular pathology from fundus photographs. *Biomed Opt Express.* 2017;8:1005–24.
69. Milea D, Najjar RP, Zhuho J, Ting D, Vasseneix C, Xu X, et al. Artificial intelligence to detect papilledema from ocular fundus photographs. *N Engl J Med.* 2020;382(18):1687–95.
70. Ahn JM, Kim S, Ahn KS, et al. Accuracy of machine learning for differentiation between optic neuropathies and pseudopapilledema. *BMC Ophthalmol.* 2019;19:178.
71. Yang HK, Oh JE, Han SB, et al. Automatic computer-aided analysis of optic & disc pallor in fundus photographs. *Acta Ophthalmol (Copenh).* 2019;97:e519–25.
72. Liu TYA, Ting DSW, Yi PH, Wei J, Zhu H, Subramanian PS, et al. Deep learning and transfer learning for optic disc laterality detection: implications for machine learning in neuro-ophthalmology. *J Neuroophthalmol.* 2020;40(2):178–84.
73. de Sisternes L, Jonna G, Greven MA, et al. Individual drusen segmentation and repeatability and reproducibility of their automated quantification in optical coherence tomography images. *Trans Vis Sci Tech.* 2017;6(1):12.

74. Pramil V, de Sisternes L, Omlor L, et al. A deep learning model for automated segmentation of geographic atrophy imaged using swept-source OCT. *Ophthalmol Retina*. 2023;7(2):127–41.
75. Spaide T, Jiang J, Patil J, et al. Geographic atrophy segmentation using multimodal deep learning. *Trans Vis Sci Tech*. 2023;12(7):10.
76. Zhang G, Fu DJ, Liefers B, et al. Clinically relevant deep learning for detection and quantification of geographic atrophy from optical coherence tomography: a model development and external validation study. *Lancet Digit Health*. 2021;3(10):e665–75.
77. Schlegl T, Waldstein SM, Bogunovic H, et al. Fully automated detection and quantification of macular fluid in OCT using deep learning. *Ophthalmology*. 2018;125(4):549–58.
78. De Fauw J, Ledsam JR, Romera-Paredes B, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med*. 2018;24(9):1342–50.
79. Abràmoff MD, Lou Y, Erginay A, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Invest Ophthalmol Vis Sci*. 2016;57(13):5200–6.
80. Venhuizen FG, van Ginneken B, Bloemen B, et al. Automated age-related macular degeneration classification in OCT using unsupervised feature learning. *Med Imaging* 2015;941411.
81. de Sisternes L, Simon N, Tibshirani R, et al. Quantitative SD-OCT imaging biomarkers as indicators of age-related macular degeneration progression. *Invest Ophthalmol Vis Sci*. 2014;55(11):7093–103.
82. Banerjee I, de Sisternes L, Hallak JA, et al. Prediction of age-related macular degeneration disease using a sequential deep learning approach on longitudinal SD-OCT imaging biomarkers. *Sci Rep*. 2020;10(1):15434.
83. Russakoff DB, Lamin A, Oakley JD, et al. Deep learning for prediction of AMD progression: a pilot study. *Invest Ophthalmol Vis Sci*. 2019;60(2):712–22.
84. Mai J, Lachinov D, Reiter GS, et al. Deep learning-based prediction of individual geographic atrophy progression from a single baseline OCT. *Ophthalmol Sci*. 2024. <https://doi.org/10.1016/j.xops.2024.100466>.
85. Fu DJ, Lipkova V, Liefers B, et al. Evaluating the effects of C3 inhibition on geographic atrophy progression from deep-learning OCT quantification: a split-person study. *Ophthalmol Therapy*. 2023;12(6):3143–58.
86. Schmidt-Erfurth U, Bogunovic H, Grechenig C, et al. Role of deep learning-quantified hyperreflective foci for the prediction of geographic atrophy progression. *Am J Ophthalmol*. 2020;216:257–70.
87. Moulton EM, Shi Y, Zhang Q, et al. Analysis of correlations between local geographic atrophy growth rates and local OCT angiography-measured choriocapillaris flow deficits. *Biomed Opt Express*. 2021;12(7):4573–95.
88. Niu S, de Sisternes L, Chen Q, et al. Fully automated prediction of geographic atrophy growth using quantitative spectral-domain optical coherence tomography biomarkers. *Ophthalmology*. 2016;123(8):1737–50.
89. Anegondi N, Gao SS, Steffen V, et al. Deep learning to predict geographic atrophy area and growth rate from multimodal imaging. *Ophthalmol Retina*. 2023;7(3):243–52.
90. Zhou Y, Chia MA, Wagner SK, et al. A foundation model for generalizable disease detection from retinal images. *Nature*. 2023;622(7981):156–63.
91. Kassahun Y, Yu B, Tibebu AT, Stoyanov D, Giannarou S, Metzen JH, et al. Surgical robotics beyond enhanced dexterity instrumentation: a survey of machine learning techniques and their role in intelligent and autonomous surgical actions. *Int J CARS*. 2016;11(4):553–68. <https://doi.org/10.1007/s11548-015-1305-z>.
92. Gottesman O, Johansson F, Komorowski M, Faisal A, Sontag D, Doshi-Velez F, et al. Guidelines for reinforcement learning in healthcare. *Nat Med*. 2019;25(1):16–8. <https://doi.org/10.1038/s41591-018-0310-5>.
93. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. *Nat Med*. 2019;25(1):24–9. <https://doi.org/10.1038/s41591-018-0316-z>.
94. Abbeel P, Ng AY. Apprenticeship learning via inverse reinforcement learning. In: *Proceedings of the twenty-first international conference on machine learning (ICML '04)*. Banff, Alberta, Canada: Association for Computing Machinery; 2004, p. 1. <https://doi.org/10.1145/1015330.1015430>
95. Ruzicki J, Holden M, Cheon S, et al. Use of machine learning to assess cataract surgery skill level with tool detection. *Ophthalmol Sci*. 2022;3(1): 100235.
96. Keller B, Draelos M, Zhou K, et al. Optical coherence tomography-guided robotic ophthalmic microsurgery via reinforcement learning from demonstration. *IEEE Trans Robot*. 2020;36(4):1207–18.
97. Rajalakshmi R, Subashini R, Anjana RM, et al. Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence. *Eye*. 2018;32(6):1138–44.
98. Rogers TW, Gonzalez-Bueno J, Garcia Franco R, et al. Evaluation of an AI system for the detection of diabetic retinopathy from images captured with a handheld portable fundus camera: the MAILOR AI study. *Eye*. 2021;35(2):632–8.
99. Lupidi M, Danieli L, Fruttini D, et al. Artificial intelligence in diabetic retinopathy screening: clinical assessment using handheld fundus camera in a real-life setting. *Acta Diabetol*. 2023;60(8):1083–8.

Artificial Intelligence in Age-Related Macular Degeneration (AMD)

8

Souvick Mukherjee, Yifan Peng, Qingyu Chen,
Tiarnan D. L. Keenan, Emily Y. Chew, and Zhiyong Lu

Abstract

Age-related macular degeneration (AMD) is the leading cause of legal blindness in developed countries. Based on clinical features, the disease can be classified into early, intermediate, and late stages. In some patients, AMD advances to the late vision-threatening

stage slowly; in others, the disease progresses faster and may quickly lead to a loss of vision in one or both eyes. There is, therefore, a critical need to detect AMD severity accurately. As deep learning has become a rapidly expanding field in imaging-based analysis of AMD, we provide a short overview of recent advances in deep learning applied to AMD severity classification. Our aims are to: (1) give a brief introduction to AMD; (2) indicate how deep learning has been applied to fundus photographs for AMD detection and classification; (3) highlight the potential utility of deep learning methodologies to assist and enhance clinical decision-making in patients with AMD; and (4) provide a starting point for researchers who are interested in this field by providing useful resources such as code and data.

S. Mukherjee · Y. Peng · T. D. L. Keenan ·
E. Y. Chew (✉)
Division of Epidemiology and Clinical Applications,
National Eye Institute (NEI), National Institutes
of Health (NIH), Bethesda, MD, USA
e-mail: echew@nei.nih.gov

S. Mukherjee
e-mail: souvick.mukherjee@nih.gov

Y. Peng
e-mail: yip4002@med.cornell.edu

T. D. L. Keenan
e-mail: tiarnan.keenan@nih.gov

Q. Chen
Biomedical Informatics and Data Science at Yale
School of Medicine, New Haven, USA
e-mail: qingyu.chen@yale.edu

Y. Peng · Z. Lu
National Center for Biotechnology Information
(NCBI), National Library of Medicine (NLM),
National Institutes of Health (NIH),
Bethesda, MD, USA
e-mail: luzh@ncbi.nlm.nih.gov

Y. Peng
Department of Population Health Sciences,
Weill Cornell Medicine, New York, NY, USA

Keywords

Deep learning · Age-related macular
degeneration (AMD) · Geographic atrophy
(GA) · Neovascular AMD · Reticular
pseudodrusen (RPD)

Abbreviations

AMD	Age-related macular degeneration
AREDS	Age-related eye disease study
CGA	Central geographic atrophy
DHA	Docosahexaenoic acid
EPA	Eicosapentaenoic acid
FAF	Fundus autofluorescence
GA	Geographic atrophy
RPE	Retinal pigment epithelium
RPD	Reticular pseudodrusen

Introduction

Age-related macular degeneration (AMD) accounts for approximately 9% of global blindness and is the leading cause of visual loss in developed countries [1, 2]. The number of people with AMD worldwide is projected to be 196 million in 2020, rising substantially to 288 million in 2040 [3]. The prevalence of AMD increases exponentially with age: late AMD in white populations has been estimated by meta-analysis at 6% at 80 years and 20% at 90 years [4]. Over time, increased disease prevalence through changing population demographics may place significant burdens on eye services, especially where retinal specialists are not available in sufficient numbers to perform individual examinations on all patients. It is conceivable that deep learning and/or telemedicine approaches might support future eye services. However, this might only apply when evidence-based systems have undergone extensive validation and demonstrated performance metrics that are at least non-inferior to those of clinical ophthalmologists in routine practice.

AMD arises from a complex interplay between aging, genetics, and environmental risk factors [5, 6]. It is regarded as a progressive, step-wise disease. It is classified by clinical features (based on clinical examination or color fundus photography) into early, intermediate, and late stages [7]. The hallmarks of the intermediate disease are the presence of large drusen

and/or pigmentary abnormalities at the macula (Fig. 8.1). There are two forms of late AMD: neovascular AMD and atrophic AMD, defined by geographic atrophy (GA).

Automated image analysis tools have demonstrated promising results in biology and medicine [8–13]. Early retinal image classification systems of color fundus photographs (CFP) had adopted traditional machine learning with human-engineered features [14]. Recently, deep learning, a subfield of machine learning, has generated substantial interest in the field of ophthalmology [8, 15–20]. Compared to the traditional imaging workflow that relies heavily on human grading, which is highly time-consuming, deep learning enables more accurate and efficient approaches.

Due to the importance of deep learning in the imaging-based analysis of AMD, this chapter aims to discuss the role of medical imaging, empowered by AI, in detecting AMD, which will encourage future practical applications and methodological research.

In the following, we first introduce common data modalities and publicly available datasets and then summarize popular machine learning methods in AMD detection, classification, and prognosis. Finally, we discuss several open problems and challenges. We hope to provide guidance for researchers and ophthalmologists.

Materials

In this section, we briefly introduce common data modalities and publicly available data sets that have been widely used in AI-powered AMD research.

Common Data Modalities

Three imaging modalities are commonly used in AI-powered AMD research.

Color Fundus Photograph (CFP) uses a fundus camera to record color images of the condition of the interior surface of the eye [21]. A fundus camera or retinal camera is a

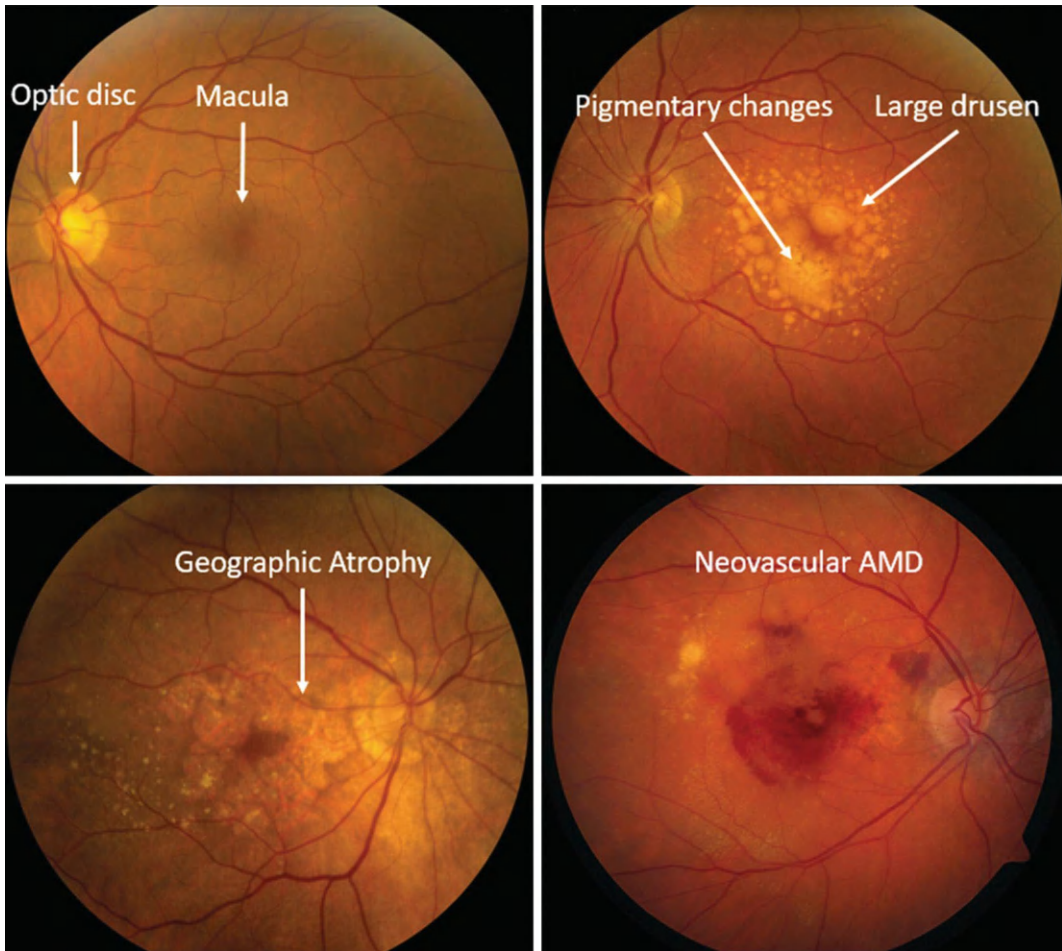


Fig. 8.1 Demonstration of the AMD pathologies, according to the age-related eye disease study (AREDS)

specialized low-power microscope with an attached camera designed to photograph the interior surface of the eye, including the retina, retinal vasculature, optic disc, macula, and posterior pole (i.e., the fundus). The retina is imaged to document conditions such as diabetic retinopathy, age-related macular degeneration, macular edema, and retinal detachment.

Fundus autofluorescence (FAF) is a non-invasive retinal imaging modality used in clinical practice to provide a density map of lipofuscin, the predominant ocular fluorophore, in the retinal pigment epithelium [22]. Classically, FAF utilizes blue-light excitation, then collects emissions within preset spectra to

form a brightness map reflecting the distribution of lipofuscin, a dominant fluorophore located in the retinal pigment epithelium (RPE). Thus, it provides detailed insight into the health of the RPE.

Optical coherence tomography (OCT) is another non-invasive imaging test, which uses light waves to take cross-section pictures of the retina [23]. More recently, advances in OCT technology have made it possible to investigate the feasibility of using OCT for the diagnosis and monitoring of retinal diseases such as glaucoma and AMD. Existing studies have used single or multiple modalities for AMD-related disease diagnosis and prognosis, such as using

CFP and FAF to detect the presence of RPD [24] and using FAF and OCT to detect the presence of GA [25].

AREDS and AREDS2

The Age-Related Eye Disease Study (AREDS), sponsored by the National Eye Institute (National Institutes of Health), was a 12-year multi-center prospective cohort study of the clinical course, prognosis, and risk factors of AMD, as well as a phase III randomized clinical trial to assess the effects of nutritional supplements on AMD progression [26]. In short, 4757 participants aged 55–80 years were recruited between 1992 and 1998 at 11 retinal specialty clinics in the United States. The inclusion criteria were wide, from no AMD in either eye to late AMD in one eye. The participants were randomly assigned to placebo, antioxidants, zinc, or the combination of antioxidants and zinc. The AREDS dataset is publicly accessible to researchers by request at dbGAP (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000001.v3.p1).

Similarly, the AREDS2 was a multi-center phase III randomized clinical trial that analyzed the effects of different nutritional supplements on the course of AMD [27]. 4203 participants aged 50–85 years were recruited between 2006 and 2008 at 82 retinal specialty clinics in the United States. The inclusion criteria were the presence of either bilateral large drusen or late AMD in one eye and large drusen in the fellow eye. The participants were randomly assigned to placebo, lutein/zeaxanthin, docosahexaenoic acid (DHA) plus eicosapentaenoic acid (EPA), or the combination of lutein/zeaxanthin and DHA plus EPA. AREDS supplements were also administered to all AREDS2 participants because they were by then considered the standard of care [28].

AMD Scales

AREDS Simplified Severity Scale and 4-Step Scale

Longitudinal analysis of the AREDS cohort led to the development of the patient-based AREDS Simplified Severity Scale for AMD, based on CFP [29] (Fig. 8.2a). This simplified scale provides convenient risk factors for assessing the risk of progression to late AMD that can be determined by clinical examination or by less demanding photographic procedures than used in the AREDS. The scale combines risk factors from both eyes to generate an overall score for the individual, based on the presence of one or more large drusen (diameter > 125 μm) and/or AMD pigmentary abnormalities at the macula of each eye [29]. The Simplified Severity Scale is clinically useful in that it allows ophthalmologists to predict an individual's 5-year risk of developing late AMD. This 5-step scale (from score 0 to 4) estimates the 5-year risk of the development of late AMD in at least one eye as 0.4%, 3.1%, 11.8%, 25.9%, and 47.3%, respectively [29].

9-Step AREDS Severity Scale

The AREDS 9-step Severity Scale was designed to allow reading center graders to perform comprehensive classification of AMD-related features on fundus photographs, as a research tool [30] (Fig. 8.2b). The detailed nature of the system means that ophthalmologists in clinical practice very rarely use it. In short, the 9-step severity scale combines a 6-step drusen area scale with a 5-step pigmentary abnormality scale. The 5-year risk of advanced AMD increases progressively from less than 1% in step 1 to about 50% in step 9.

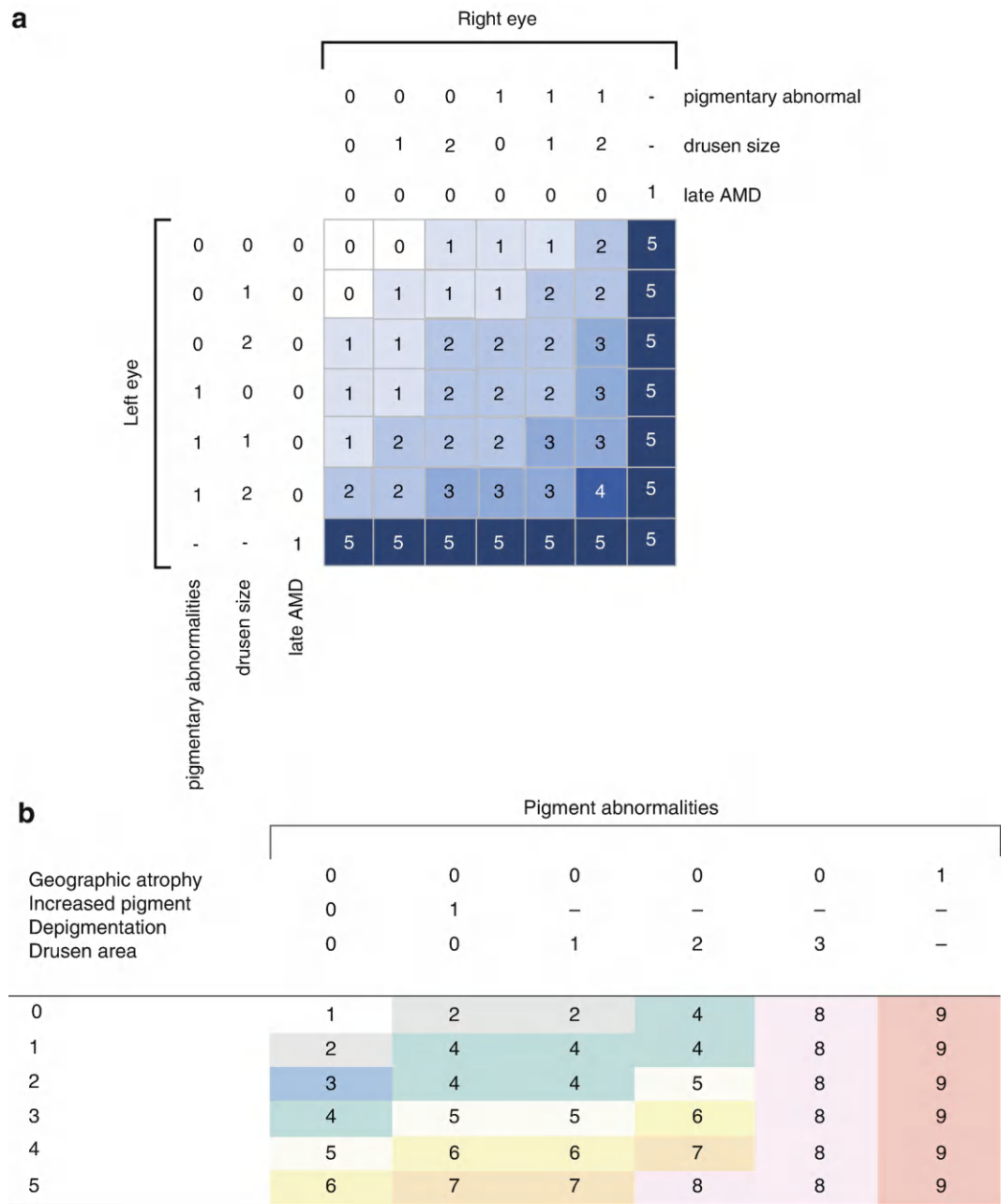


Fig. 8.2 Different AMD scales. **(a)** AREDS Simplified Severity Scale Scoring schematic for participants with and without late age-related macular degeneration (AMD). Pigmentary abnormalities: 0=no, 1=yes; drusen size: 0=small or none, 1=medium, 2=large; late AMD: 0=no, 1=yes. **(b)** AREDS Severity Scale scores 1–9, defined by graders from four categories: geographic atrophy (0/1, i.e., absent/present), increased pigment (0/1, i.e., absent/present), depigmentation (graded 0–3), and drusen area (graded 0–5). The final AREDS Severity Scale score (steps 1–9, shown shaded in different colors) is defined by the combination of findings from these four categories

Methods

Automated image analysis tools based on deep learning have demonstrated promising results in biology and medicine [8–13]. In particular, deep learning, a subfield of machine learning, has recently generated substantial interest in the field of ophthalmology [8, 15–20]. In general, deep learning is the process of training algorithmic models with labeled data (e.g., CFP categorized manually as containing pigmentary abnormalities or not), where these models can then be used to assign labels automatically to new data. Deep learning differs from traditional machine learning methods in that specific image features do not need to be pre-specified by experts in that field. Instead, the image features are learned directly from the images themselves. Past studies have utilized deep learning systems for the identification of various retinal diseases, including diabetic retinopathy [31–36], glaucoma [36–39], and retinopathy of prematurity [40]. In recent years, we have seen the successful use of deep learning on AMD (Fig. 8.3).

Deep Learning in AMD

Challenges

The Collaborative Community on Ophthalmic Imaging published a roadmap for the application of AI in age-related macular degeneration (AMD) [41]. The authors identified several challenges, including the absence of large labeled datasets, lack of metadata for developing generalizable models, and privacy and security restrictions based on data formats. The authors mentioned that these problems needed to be addressed before AI devices in ophthalmology could receive Federal Drug Administration (FDA) approval. Recent advancements, such as the development of foundational models [42], federated learning approaches [43, 44], encouragement of DICOM-based standardization for ocular imaging for future funding by the National Eye Institute-National Institutes of Health [45], and the development of large-scale ophthalmic datasets [46], have generated cautious optimism in the field of AI in ophthalmology.

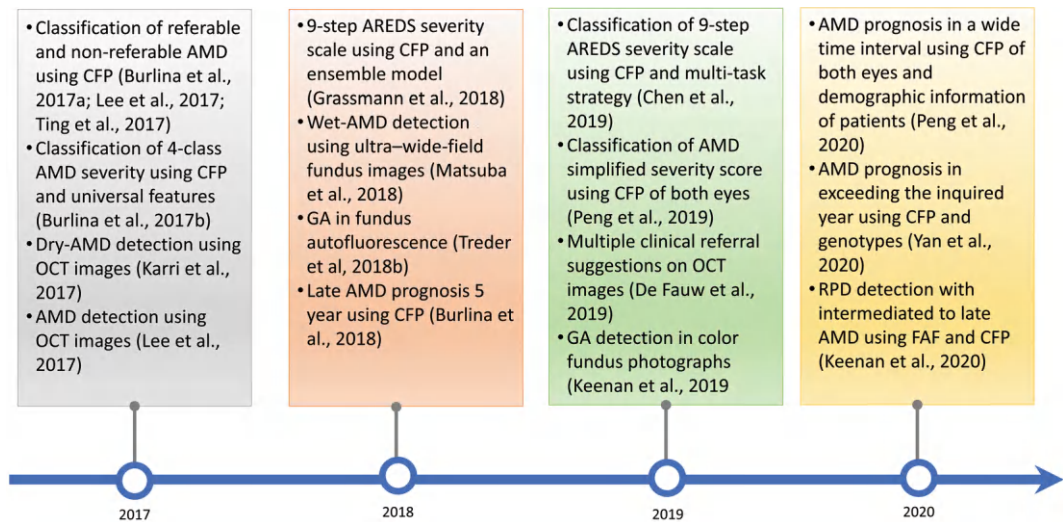


Fig. 8.3 The timeline of selected articles on Deep Learning in AMD

Classification of 2-Class and 2-Class AMD Severity

Recently, several deep learning systems have been developed for the classification of CFP into AMD severity scales, at the level of the individual eye. They applied deep neural networks to a 2-class classification problem in which the task was to distinguish referable AMD (intermediate or advanced disease) from non-referable AMD (no or early disease) [16, 20, 36, 47, 48]. Burlina et al. used transfer learning and universal features to address a 4-class AMD severity classification problem (no, early, intermediate, and advanced disease) and obtained 79.4% vs. 75.8% in accuracy for machine and physician grading, respectively [16].

Classification of 9-Step AREDS Severity Scale

In a more complicated scenario in which the aim was to classify AMD according to the 9-step AREDS Severity Scale, Grassmann et al. trained an ensemble of six different neural net architectures with a random forest approach to directly predict the 9-step AREDS severity scale in an AREDS test set with an overall accuracy of 63.3%, outperforming human graders [17] (Fig. 8.2b). However, 1-step classification of images according to the 9-step AREDS Severity Scale does not reflect regular human grading practice. In reading centers, graders instead first calculate individual scores for four different AMD characteristics (drusen area, geographic atrophy, increased pigmentation, and depigmentation), then combine the scores for these four characteristics into the 9-step non-advanced AREDS severity scale [30]. Hence, a deep learning approach that predicts the overall AREDS severity score directly, without these intermediate steps, may have lower transparency and decreased information content for research and clinical purposes [49, 50]. To address these potential criticisms, Chen et al. designed four deep learning models, each of which is

responsible for classifying an individual characteristic, and trained them separately using a multi-task strategy [51]. Evaluation on both the AREDS and AREDS2 datasets showed that the accuracy of the model exceeded the current state-of-the-art model by >10%.

Classification of AREDS Simplified Severity Scale

Besides AMD classification at the level of the individual eye, it is also helpful to obtain one overall score for the individual from both eyes (Fig. 8.2a). This is particularly relevant because estimates of rates of progression to late AMD are highly influenced by the status of fellow eyes, as the behavior of the two eyes is highly correlated [29]. To this end, Peng et al. proposed a deep learning framework to automatically identify AMD severity from CFP of both eyes [52]. It mimicked the human grading process by first detecting individual risk factors (drusen and pigmentary abnormalities) in each eye and then combining values from both eyes to assign an AMD score to each patient. Thus, the model closely matches the clinical decision-making process, which allows an ophthalmologist to inspect and visualize an interpretable result.

Prediction of Risk of Progression to Late AMD

Besides AMD classification, making accurate time-based predictions of progression to late AMD is also clinically critical. This would enable improved decision-making regarding: (i) medical treatments, especially oral supplements known to decrease progression risk, (ii) lifestyle interventions, mainly smoking cessation and dietary changes, and (iii) intensity of patient monitoring, e.g., frequent reimaging in the clinic or tailored home monitoring programs [53–57]. It would also aid the design of future clinical trials, which could be enriched for participants with a high risk of progression events [58].

Currently, three existing methods are available clinically for using CFP to predict the risk of progression. Of the three methods, the most commonly used is the AREDS Simplified Severity Scale, as described above [29]. The second method is an online risk calculator [59]. Like the Simplified Severity Scale, its inputs include the presence of macular drusen and pigmentary abnormalities; however, it can also receive the individual's age, smoking status, and basic genotype information consisting of two SNPs (when available). The third method is a deep learning-based architecture to predict progression with improved accuracy and transparency in two steps: image classification followed by survival analysis [60]. The model was developed and clinically validated on two datasets from AREDS and AREDS2.

Classification of AMD on Optical Coherence Tomography

Besides CFP, OCT also plays a major role in the detection of AMD [61]. Several recent studies have reported robust performance in the automated classification of AMD from OCT scans. Karri et al. fine-tuned a CNN model to classify OCT images of dry AMD [62]. Lee et al. developed an algorithm to categorize OCT images as either normal or AMD [20]. The images were linked to clinical data points from the electronic health record, and gold-standard labels were extracted using the ICD-9 diagnosis codes. At a patient level, the model achieved an area under the ROC curve of 97.45% with an accuracy of 93.45%. De Fauw et al. further demonstrated expert performance on multiple clinical referral suggestions for two independent test datasets [63].

Deep Learning in GA

Geographic atrophy (GA) is the defining lesion of the atrophic form of late AMD. GA in AMD has been estimated to affect over 5 million people worldwide [3, 4]. Two drugs have been

recently approved by the FDA to slow the progression of GA [64, 65], though there are still no drugs to reverse the condition. Rapid and accurate identification of eyes with GA could improve the recruitment of eligible patients for future clinical trials and eventually to early identification of appropriate patients for proven treatments.

Since the original description of GA by Gass [66], clinical definitions have varied between research groups [67]. In the AREDS, GA was defined as a sharply demarcated, usually circular zone of partial or complete depigmentation of RPE, typically with exposure of underlying large choroidal blood vessels, at least as large as grading circle I-1 (1/8 disc diameter in diameter) [68]. The sensitivity of the retina to light stimuli is markedly decreased (i.e., dense scotomata) in areas affected by GA. The natural history of the disease involves progressive enlargement of GA lesions over time, with visual acuity decreasing markedly as the central macula becomes involved [69].

The identification of GA by ophthalmologists conducting dilated fundus examinations is sometimes challenging. This may be particularly true for cases with early GA, comprising smaller lesions with less extensive RPE atrophy. In addition, the increasing prevalence of GA (through aging populations in many countries) will translate to greater demand for retinal services. As such, deep learning approaches involving retinal images, obtained perhaps using telemedicine-based devices, might support GA detection and diagnosis. However, these approaches would require establishing evidence-based and 'explainable' systems that have undergone extensive validation and demonstrated performance metrics that are at least non-inferior to those of clinical ophthalmologists in routine practice.

In contrast to studies of AMD classification, few studies have explicitly focused on GA. Treder et al. detected and classified GA in FAF images using a deep learning algorithm [70]. Two classifiers were built, one to classify healthy patients and patients with GA and the other to classify patients with GA and other

retinal diseases. Both achieved high accuracy. Keenan et al. conducted an empirical study to investigate the performance of deep learning models on CFP [71]. The first model predicted GA presence from a population of eyes with no AMD to advanced AMD using CFP; the second model predicted CGA presence from the same population, and the third model predicted central GA (CGA) presence from the subset of eyes with GA. Experiments demonstrated that deep learning could achieve high accuracy for the detection of GA, and compared favorably for the detection of CGA to human retinal specialists.

Researchers have developed automatic deep learning GA segmentation algorithms to segment GA from CFPs, enabling the extraction of GA configurations and the investigation of growth rates [72].

Pegcetacoplan, recently approved by the US-FDA, is one of two sanctioned drugs for lowering GA growth rates. Recent studies using AI [73] have shown that eyes treated with pegcetacoplan exhibit substantially slower GA growth rates compared to sham treatments, especially towards the foveal region. The authors suggest that use of AI to identify cohorts with larger growth rates could help optimize pegcetacoplan treatment benefits and have significant clinical implications.

Multitask AI models have been developed to utilize baseline Fundus Autofluorescence (FAF) images and Spectral domain OCT volumes to jointly predict GA lesion areas and annual GA growth rates [74]. The design hypothesis posits that multitask models are less prone to overfitting and that GA lesion area at baseline may be associated with GA progression over time [72, 75].

Additionally, algorithms for segmenting GA from OCT B-scans have reached human expert-level performance and outperformed inter-grader agreements on external validation datasets [76]. With the advent of Notal Vision's FDA-authorized home OCTs [77], such algorithms could provide valuable diagnostic and prognostic benefits outside clinical settings, conserving critical clinical resources.

Deep Learning in RPD

Reticular pseudodrusen (RPD), also known as subretinal drusenoid deposits, have been identified as another disease feature independently associated with an increased risk of progression to late AMD [78]. Unlike soft drusen, which are located in the sub-retinal pigment epithelial (RPE) space, RPD are thought to represent aggregations of material in the subretinal space between the RPE and photoreceptors [78, 79]. Compositional differences have also been found between soft drusen and RPD [79].

The detection of eyes with RPD is important for multiple reasons. Not only is their presence associated with increased risk of late AMD, but the increased risk is weighted towards particular forms of late AMD, including the recently recognized phenotype of outer retinal atrophy [78, 80, 81]. In recent analyses of AREDS2 data, the risk of progression to GA was significantly higher with RPD presence. In contrast, the risk of neovascular AMD was not [57]. Hence, RPD presence may be a powerfully discriminating feature that could be very useful in risk prediction algorithms for the detailed prognosis of AMD progression. The presence of RPD has also been associated with increased speed of GA enlargement [82], which is a critical endpoint in ongoing clinical trials. Finally, in eyes with intermediate AMD, the presence of RPD may be a critical determinant of the reduced efficacy of subthreshold nanosecond laser to slow progression to late AMD [44] as observed in an unplanned subgroup analysis of the clinical trial cohort by stratifying by the presence of RPD.

However, owing to the poor visibility of RPD on clinical examination and CFP [67, 83, 84], they were not incorporated into the traditional AMD classification and risk stratification systems, such as the Beckman clinical classification scale [7] or the AREDS scales [29, 30]. With the advent of more recent imaging modalities, including FAF, near-infrared reflectance, and OCT [79, 80], the presence of RPD may be ascertained more accurately with careful grading at the reading center level [67, 85, 86]. However,

their detection by ophthalmologists (including retinal specialists) in a clinical setting may still be challenging.

Keenan et al. studied a deep learning model to detect RPD in eyes with intermediate to late AMD, using FAF images (FAF model) and CFP images (CFP model), respectively [87]. The gold standard labels were annotated based on the FAF images. Model performance was compared with that of four ophthalmologists using a random subset from the entire test set. Both models achieved a high level of accuracy, equal or superior to four ophthalmologists on the random subset.

Multimodal, multitask, and multiattention classification AI models have been developed to detect reticular pseudodrusen from various imaging modalities, including the challenging CFPs alone [88]. FAF images alone showed better performance than CFP images alone, and the multimodal model combining FAF and CFP demonstrated slightly better performance than FAF alone.

In [89], the authors compared the effectiveness of multimodal images including—color fundus images, fundus autofluorescence images, and near-infrared images—against unimodal images for identifying reticular pseudodrusen (RPD). The presence or absence of RPD was assessed by two experienced graders using a single modality, and their performance was compared to grading using all modalities. The authors found that RPD detection using multimodal grading significantly improved performance and inter-grader agreement compared to unimodal images. They also developed a random forest classifier for automatic RPD detection using multimodal images, demonstrating performance comparable to expert graders.

Additionally, researchers have developed a deep learning network capable of extracting gradable OCT B-scans with drusen or RPD and subsequently trained an AI algorithm to independently segment RPD or drusen lesions from these scans [90]. The model's predictions were compared to expert human performance and were found to be similar.

Deep Learning Based Foundational Model

Another significant advancement in AI includes the development of a foundational model for retinal images called RETFound, which can learn generalizable representations from unlabelled images [41]. This model was trained on approximately 1.6 million unlabelled retinal images from the UK Biobank using an advanced self-supervised learning technique involving a masked autoencoder [91]. RETFound can be finetuned with a limited number of labelled images for various related downstream tasks, such as detecting and prognosticating ocular diseases like diabetic macular edema, age-related macular degeneration, referable diabetic retinopathy and glaucoma. It can also predict the conversion of fellow eyes to neovascular AMD within a year and provide three-year outcome predictions for systemic cardiovascular disorders like ischemic stroke, myocardial infarction, heart failure, and neurodegenerative disorders like Parkinson's disease. The use of multimodal retinal imaging, including both CFPs and OCT images, produced the best results. The researchers demonstrated that RETFound could be efficiently adapted for detecting and predicting a wide range of retinal and other systemic diseases. For downstream tasks, there was substantial performance improvement of pretraining with RETFound, compared to pretraining based on three other state-of-the-art techniques. RETFound's performance was compared to three other models including—(1) supervised learning with ImageNet [92], which uses transfer learning by pretraining the model on 14 million natural images, (2) semi-supervised learning based on masked autoencoders [91] with natural images from ImageNet (1.4 million natural images) and (3) semi-supervised learning based on masked autoencoders [91] with retinal images (1.6 million images). RETFound's ability to detect and predict cardiovascular and neurodegenerative diseases without requiring high-quality labelled datasets should significantly reduce costs.

Discussion and Future Directions

The application of deep learning to AMD research using fundus photographs is just the beginning. As introduced above, attempts have been made to apply deep learning to AMD detection or disease features associated with increased risk of progression to late AMD. However, there is still much work to be conducted in the future.

One limitation arises from the imbalance of cases used for deep learning training, particularly the relatively low proportion of participants with outcomes in the clinical trials. This is likely to have contributed to the relatively lower accuracy of the model. However, this limitation may potentially be addressed by further training using image datasets with a higher proportion of positive cases.

The second limitation of this dataset includes the sole use of CFP, OCT, or FAF. Multi-modal imaging would be desirable. In AMD, some disease features are visualized more clearly to human experts on one than the other modality [93]. For example, macular drusen are typically observed well on CFP but poorly on FAF, while the opposite is true for RPD [67, 83, 84, 93]. Other AMD features are observed in both modalities. For example, pigmentary abnormalities are seen on both though typically classified on CFP [7, 93], while geographic atrophy is seen on both (but typically identified and measured on FAF) [93, 94]. Hence, any techniques that enable the accurate ascertainment of the full spectrum of AMD features would be important for improved disease classification and risk prediction.

Another potential limitation lies in higher levels of image quality for accurate classification. Despite high theoretical accuracy, deep learning models might be impractical in real-world practice. In a recent study, Beede et al. showed that the accuracy of the DL models was highly varied across different clinical setting and locations where eye screenings took place [95]. Further steps need to be taken for the methods to be applied in the clinic.

References

1. Congdon N, O'Colmain B, Klaver CCW, et al. Causes and prevalence of visual impairment among adults in the United States. *Arch Ophthalmol Chic Ill* 1960. 2004;122(4):477–85. <https://doi.org/10.1001/archophth.122.4.477>
2. Quartilho A, Simkiss P, Zekite A, Xing W, Wormald R, Bunce C. Leading causes of certifiable visual loss in England and Wales during the year ending 31 March 2013. *Eye Lond Engl*. 2016;30(4):602–7. <https://doi.org/10.1038/eye.2015.288>.
3. Wong WL, Su X, Li X, et al. Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. *Lancet Glob Health*. 2014;2(2):e106–16. [https://doi.org/10.1016/S2214-109X\(13\)70145-1](https://doi.org/10.1016/S2214-109X(13)70145-1).
4. Rudnicka AR, Jarrar Z, Wormald R, Cook DG, Fletcher A, Owen CG. Age and gender variations in age-related macular degeneration prevalence in populations of European ancestry: a meta-analysis. *Ophthalmology*. 2012;119(3):571–80. <https://doi.org/10.1016/j.ophtha.2011.09.027>.
5. Fritsche LG, Fariss RN, Stambolian D, Abecasis GR, Curcio CA, Swaroop A. Age-related macular degeneration: genetics and biology coming together. *Annu Rev Genomics Hum Genet*. 2014;15:151–71. <https://doi.org/10.1146/annurev-genom-090413-025610>.
6. Ratnapriya R, Chew EY. Age-related macular degeneration-clinical review and genetics update. *Clin Genet*. 2013;84(2):160–6. <https://doi.org/10.1111/cge.12206>.
7. Ferris FL, Wilkinson CP, Bird A, et al. Clinical classification of age-related macular degeneration. *Ophthalmology*. 2013;120(4):844–51. <https://doi.org/10.1016/j.ophtha.2012.10.036>.
8. Ching T, Himmelstein DS, Beaulieu-Jones BK, et al. Opportunities and obstacles for deep learning in biology and medicine. *J R Soc Interface* 2018;15(141). <https://doi.org/10.1098/rsif.2017.0387>
9. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA*. 2017;318(22):2199–210. <https://doi.org/10.1001/jama.2017.14585>
10. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115–8. <https://doi.org/10.1038/nature21056>.
11. Lehman CD, Wellman RD, Buist DSM, et al. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. *JAMA Intern Med*. 2015;175(11):1828–37. <https://doi.org/10.1001/jamainternmed.2015.5231>.
12. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. Chestx-ray8: hospital-scale chest x-ray database

- and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: IEEE Conference on computer vision and pattern recognition (CVPR). IEEE; 2017, pp. 3462–71. <https://doi.org/10.1109/CVPR.2017.369>
13. Wang X, Peng Y, Lu L, Lu Z, Summers RM. TieNet: text-image embedding network for common thorax disease classification and reporting in chest x-rays. In: IEEE Conference on computer vision and pattern recognition (CVPR). IEEE; 2018, pp. 9049–58. <https://doi.org/10.1109/cvpr.2018.00943>
 14. Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: The IEEE conference on computer vision and pattern recognition (CVPR). IEEE; 2005, pp. 886–93. <https://doi.org/10.1109/CVPR.2005.177>
 15. Burlina P, Freund DE, Joshi N, Wolfson Y, Bressler NM. Detection of age-related macular degeneration via deep learning. In: IEEE International symposium on biomedical imaging (ISBI). IEEE; 2016. <https://doi.org/10.1109/isbi.2016.7493240>
 16. Burlina P, Pacheco KD, Joshi N, Freund DE, Bressler NM. Comparing humans and deep learning performance for grading AMD: a study in using universal deep features and transfer learning for automated AMD analysis. *Comput Biol Med.* 2017;82:80–6. <https://doi.org/10.1016/j.combiomed.2017.01.018>
 17. Grassmann F, Mengelkamp J, Brandl C, et al. A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography. *Ophthalmology.* 2018;125(9):1410–20. <https://doi.org/10.1016/j.ophtha.2018.02.037>
 18. Kermany DS, Goldbaum M, Cai W, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell.* 2018;172(5):1122–31.e9. <https://doi.org/10.1016/j.cell.2018.02.010>
 19. Lam C, Yu C, Huang L, Rubin D. Retinal lesion detection with deep learning using image patches. *Invest Ophthalmol Vis Sci.* 2018;59(1):590–6. <https://doi.org/10.1167/jovs.17-22721>
 20. Lee CS, Baughman DM, Lee AY. Deep learning is effective for the classification of OCT images of normal versus age-related macular degeneration. *Ophthalmol Retina.* 2017;1(4):322–7. <https://doi.org/10.1016/j.oret.2016.12.009>
 21. Graham KW, Chakravarthy U, Hogg RE, Muldrew KA, Young IS, Kee F. Identifying features of early and late age-related macular degeneration: a comparison of multicolor versus traditional color fundus photography. *Retina.* 2018;38(9):1751–8. <https://doi.org/10.1097/IAE.0000000000001777>
 22. Holz FG, Bindewald-Wittich A, Fleckenstein M, et al. Progression of geographic atrophy and impact of fundus autofluorescence patterns in age-related macular degeneration. *Am J Ophthalmol.* 2007;143(3):463–72. <https://doi.org/10.1016/j.ajo.2006.11.041>
 23. Fujimoto JG, Pitris C, Boppart SA, Brezinski ME. Optical coherence tomography: an emerging technology for biomedical imaging and optical biopsy. *Neoplasia N Y N.* 2000;2(1–2):9–25. <https://doi.org/10.1038/sj.neo.7900071>
 24. Chen Q, Keenan TDL, Allot A, Peng Y, Agrón E, Domalpally A, Klaver CCW, Luttikhuisen DT, Colyer MH, Cukras CA, Wiley HE, Teresa Magone M, Cousineau-Krieger C, Wong WT, Zhu Y, Chew EY, Lu Z; AREDS2 Deep Learning Research Group. Multimodal, multitask, multiattention (M3) deep learning detection of reticular pseudodrusen: toward automated and accessible classification of age-related macular degeneration. *J Am Med Inform Assoc.* 2021;28(6):1135–48. <https://doi.org/10.1093/jamia/ocaa302>
 25. Arslan J, Samarasinghe G, Benke KK, et al. Artificial intelligence algorithms for analysis of geographic atrophy: a review and evaluation. *Transl Vis Sci Technol.* 2020;9(2):57. <https://doi.org/10.1167/tvst.9.2.57>
 26. Age-Related Eye Disease Study Research Group. The age-related eye disease study (AREDS): design implications. AREDS report no. 1. *Control Clin Trials.* 1999;20(6):573–600. [https://doi.org/10.1016/s0197-2456\(99\)00031-8](https://doi.org/10.1016/s0197-2456(99)00031-8)
 27. AREDS2 Research Group, Chew EY, Clemons T, et al. The age-related eye disease study 2 (AREDS2): study design and baseline characteristics (AREDS2 report number 1). *Ophthalmology* 2012;119(11):2282–9. <https://doi.org/10.1016/j.ophtha.2012.05.027>
 28. American Academy of Ophthalmology Retina/Vitreous Panel. Preferred Practice Pattern® Guidelines. Age-related macular degeneration. *Am Acad Ophthalmol.* 2015.
 29. Ferris FL, Davis MD, Clemons TE, et al. A simplified severity scale for age-related macular degeneration: AREDS Report No. 18. *Arch Ophthalmol Chic Ill 1960.* 2005;123(11):1570–4. <https://doi.org/10.1001/archophth.123.11.1570>
 30. Davis MD, Gangnon RE, Lee L-Y, et al. The age-related eye disease study severity scale for age-related macular degeneration: AREDS report no. 17. *Arch Ophthalmol Chic Ill 1960.* 2005;123(11):1484–98. <https://doi.org/10.1001/archophth.123.11.1484>
 31. Choi JY, Yoo TK, Seo JG, Kwak J, Um TT, Rim TH. Multi-categorical deep learning neural network to classify retinal images: a pilot study employing small database. *PLoS ONE.* 2017;12(11): e0187336. <https://doi.org/10.1371/journal.pone.0187336>
 32. Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology.* 2017;124(7):962–9. <https://doi.org/10.1016/j.ophtha.2017.02.008>
 33. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus

- photographs. *JAMA*. 2016;316(22):2402–10. <https://doi.org/10.1001/jama.2016.17216>.
34. Raju M, Pagidimarri V, Barreto R, Kadam A, Kasivajjala V, Aswath A. Development of a deep learning algorithm for automatic diagnosis of diabetic retinopathy. *Stud Health Technol Inform*. 2017;245:559–63.
 35. Takahashi H, Tampo H, Arai Y, Inoue Y, Kawashima H. Applying artificial intelligence to disease staging: deep learning for improved staging of diabetic retinopathy. *PLoS ONE*. 2017;12(6): e0179790. <https://doi.org/10.1371/journal.pone.0179790>.
 36. Ting DSW, Cheung CY-L, Lim G, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA* 2017;318(22):2211–23. <https://doi.org/10.1001/jama.2017.18152>
 37. Asaoka R, Murata H, Iwase A, Araie M. Detecting preperimetric glaucoma with standard automated perimetry using a deep learning classifier. *Ophthalmology*. 2016;123(9):1974–80. <https://doi.org/10.1016/j.ophtha.2016.05.029>.
 38. Cerentini A, Welfer D, Cordeiro d’Ornellas M, Pereira Haygert CJ, Dotto GN. Automatic identification of glaucoma using deep learning methods. *Stud Health Technol Inform*. 2017;245:318–21.
 39. Muhammad H, Fuchs TJ, De Cuir N, et al. Hybrid deep learning on single wide-field optical coherence tomography scans accurately classifies glaucoma suspects. *J Glaucoma*. 2017;26(12):1086–94. <https://doi.org/10.1097/JIG.0000000000000765>.
 40. Brown JM, Campbell JP, Beers A, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. *JAMA Ophthalmol*. 2018;136(7):803–10. <https://doi.org/10.1001/jamaophthalmol.2018.1934>.
 41. Dow ER, Keenan TDL, Lad EM, Lee AY, et al. From data to deployment: the collaborative community on ophthalmic imaging roadmap for artificial intelligence in age-related macular degeneration. *Ophthalmology*. 2022;129(5):e43–59. <https://doi.org/10.1016/j.ophtha.2022.01.002>.
 42. Zhou Y, Chia MA, Wagner SK, Ayhan MS, et al. A foundation model for generalizable disease detection from retinal images. *Nature*. Accessed 01 July 2024 [Online]. Available: <https://www.nature.com/articles/s41586-023-06555-x>
 43. Lu C, Hanif A, Singh P, Chang K, et al. Federated learning for multicenter collaboration in ophthalmology: improving classification performance in retinopathy of prematurity. *Ophthalmol Retina*. 2022;6(8):657–63. <https://doi.org/10.1016/j.oret.2022.02.015>.
 44. Lo J, Timothy TY, Ma D, Zang P, et al. Federated learning for microvasculature segmentation and diabetic retinopathy classification of OCT data. *Ophthalmol Sci*. 2021;1(4): 100069. <https://doi.org/10.1016/j.xops.2021.100069>.
 45. Goetz KE, Reed AA, Chiang MF, et al. Accelerating care: a roadmap to interoperable ophthalmic imaging standards in the United States—ophthalmology. Accessed: Jul. 08, 2024. [Online]. Available: [https://www.aaojournal.org/article/S0161-6420\(23\)00713-3/fulltext](https://www.aaojournal.org/article/S0161-6420(23)00713-3/fulltext)
 46. Lee AY, Other PIs, Bridge2AI: Salutogenesis data generation project. [Online]. Available: <https://reporter.nih.gov/project-details/10471118>
 47. Matsuba S, Tabuchi H, Ohsugi H, et al. Accuracy of ultra-wide-field fundus ophthalmoscopy-assisted deep learning, a machine-learning technology, for detecting age-related macular degeneration. *Int Ophthalmol*. Published online May 2018. <https://doi.org/10.1007/s10792-018-0940-0>
 48. Treder M, Laueremann JL, Eter N. Automated detection of exudative age-related macular degeneration in spectral domain optical coherence tomography using deep learning. *Graefes Arch Clin Exp Ophthalmol Albrecht Von Graefes Arch Klin Exp Ophthalmol*. 2018;256(2):259–65. <https://doi.org/10.1007/s00417-017-3850-3>.
 49. Doshi-Velez F, Kim B. Towards a rigorous science of interpretable machine learning. *ArXiv Prepr*. Published online 2017. <https://arxiv.org/abs/1702.08608>
 50. Madumal P, Miller T, Vetere F, Sonenberg L. Towards a grounded dialog model for explainable artificial intelligence. *ArXiv Prepr*. Published online 2018. <https://arxiv.org/abs/1806.08055>
 51. Chen Q, Peng Y, Keenan T, et al. A multi-task deep learning model for the classification of age-related macular degeneration. *Proc AMIA Jt Summits Transl Sci*. 2019;2019:505–14. <https://pubmed.ncbi.nlm.nih.gov/31259005>
 52. Peng Y, Dharssi S, Chen Q, et al. DeepSeeNet: a deep learning model for automated classification of patient-based age-related macular degeneration severity from color fundus photographs. *Ophthalmology*. 2018;126(4):565–75. <https://doi.org/10.1016/j.ophtha.2018.11.015>.
 53. Age-Related Eye Disease Study 2 Research Group. Lutein + zeaxanthin and omega-3 fatty acids for age-related macular degeneration: the age-related eye disease study 2 (AREDS2) randomized clinical trial. *JAMA* 2013;309(19):2005–15. <https://doi.org/10.1001/jama.2013.4997>
 54. Age-Related Eye Disease Study Research Group. A randomized, placebo-controlled, clinical trial of high-dose supplementation with vitamins C and E, beta carotene, and zinc for age-related macular degeneration and vision loss: AREDS report no. 8. *Arch Ophthalmol*. 2001;119(10):1417–36. <https://doi.org/10.1001/archophth.119.10.1417>
 55. Areds Home Study Research Group, Chew EY, Clemons TE, et al. Randomized trial of a home monitoring system for early detection of choroidal neovascularization home monitoring of the Eye (HOME) study. *Ophthalmology* 2014;121(2):535–44. <https://doi.org/10.1016/j.ophtha.2013.10.027>

56. Guymer RH, Wu Z, Hodgson LAB, et al. Subthreshold nanosecond laser intervention in age-related macular degeneration: the LEAD randomized controlled clinical trial. *Ophthalmology*. 2019;126(6):829–38. <https://doi.org/10.1016/j.opthta.2018.09.015>.
57. Domalpally A, Clemons TE, Bressler SB, et al. Imaging characteristics of choroidal neovascular lesions in the AREDS2-HOME study: report number 4. *Ophthalmol Retina*. 2019;3(4):326–35. <https://doi.org/10.1016/j.oret.2019.01.004>.
58. Calaprice-Whitty D, Galil K, Salloum W, Zariv A, Jimenez B. Improving clinical trial participant pre-screening with artificial intelligence (AI): a comparison of the results of AI-assisted versus standard methods in 3 oncology trials. *Ther Innov Regul Sci*. 2020;54(1):69–74. <https://doi.org/10.1007/s43441-019-00030-4>.
59. Klein R, Klein BEK, Myers CE. Risk assessment models for late age-related macular degeneration. *Arch Ophthalmol Chic Ill* 1960. 2011;129(12):1605–6. <https://doi.org/10.1001/archophthalmol.2011.372>
60. Peng Y, Keenan TD, Chen Q, et al. Predicting risk of late age-related macular degeneration using deep learning. *NPJ Digit Med*. 2020;3:111. <https://doi.org/10.1038/s41746-020-00317-z>.
61. Ting DSW, Cheung CY, Nguyen Q, et al. Deep learning in estimating prevalence and systemic risk factors for diabetic retinopathy: a multi-ethnic study. *NPJ Digit Med*. 2019;2:24. <https://doi.org/10.1038/s41746-019-0097-x>.
62. Karri SPK, Chakraborty D, Chatterjee J. Transfer learning based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration. *Biomed Opt Express*. 2017;8(2):579–92. <https://doi.org/10.1364/BOE.8.000579>.
63. De Fauw J, Ledsam JR, Romera-Paredes B, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med*. 2018;24(9):1342–50. <https://doi.org/10.1038/s41591-018-0107-6>.
64. “FDA Approves SYFOVRE™ (pegcetacoplan injection) as the First and Only Treatment for Geographic Atrophy (GA), a Leading Cause of Blindness—Apellis Pharmaceuticals, Inc.” Accessed 19 Mar 2024. [Online]. Available: <https://investors.apellis.com/news-releases/news-release-details/fda-approves-syfovretm-pegcetacoplan-injection-first-and-only>
65. “Iveric Bio Announces FDA accepts new drug application and grants priority review for Avacincaptad Pegol for the treatment of geographic atrophy—Iveric Bio.” Accessed 08 July 2024. [Online]. Available: <https://investors.ivericbio.com/news-releases/news-release-details/iveric-bio-announces-fda-accepts-new-drug-application-and-grants>
66. Gass JD. Drusen and disciform macular detachment and degeneration. *Arch Ophthalmol Chic Ill* 1960. 1973;90(3):206–17.
67. Schmitz-Valckenberg S, Sadda S, Staurenghi G, et al. GEOGRAPHIC ATROPHY: semantic considerations and literature review. *Retina Phila Pa*. 2016;36(12):2250–64. <https://doi.org/10.1097/IAE.0000000000001258>.
68. Age-Related Eye Disease Study Research Group. The age-related eye disease study system for classifying age-related macular degeneration from stereoscopic color fundus photographs: the age-related eye disease study Report Number 6. *Am J Ophthalmol*. 2001;132(5):668–81. [https://doi.org/10.1016/s0002-9394\(01\)01218-1](https://doi.org/10.1016/s0002-9394(01)01218-1).
69. Keenan TD, Agrón E, Domalpally A, et al. Progression of geographic atrophy in age-related macular degeneration: AREDS2 report number 16. *Ophthalmology*. 2018;125(12):1913–28. <https://doi.org/10.1016/j.opthta.2018.05.028>.
70. Treder M, Lauermaun JL, Eter N. Deep learning-based detection and classification of geographic atrophy using a deep convolutional neural network classifier. *Graefes Arch Clin Exp Ophthalmol*. 2018;256(11):2053–60. <https://doi.org/10.1007/s00417-018-4098-2>.
71. Keenan TD, Dharssi S, Peng Y, et al. A deep learning approach for automated detection of geographic atrophy from color fundus photographs. *Ophthalmology*. Published online June 2019. <https://doi.org/10.1016/j.opthta.2019.06.005>
72. Liefers B, Colijn JM, González-Gonzalo C, Verzijden T, et al. A deep learning model for segmentation of geographic atrophy to study its long-term natural history. *Ophthalmology*. 2020;127(8):1086–96. <https://doi.org/10.1016/j.opthta.2020.02.009>.
73. Vogl W-D, Riedl S, Mai J, Reiter GS, et al. Predicting topographic disease progression and treatment response of pegcetacoplan in geographic atrophy quantified by deep learning. *Ophthalmol Retina*. 2023;7(1):4–13. <https://doi.org/10.1016/j.oret.2022.08.003>.
74. Anegondi N, Gao SS, Steffen V, Spaide RF, et al. Deep learning to predict geographic atrophy area and growth rate from multimodal imaging. *Ophthalmol Retina*. 2023;7(3):243–52. <https://doi.org/10.1016/j.oret.2022.08.018>.
75. Monés J, Biarnés M. The rate of progression of geographic atrophy decreases with increasing baseline lesion size even after the square root transformation. *Transl Vis Sci Technol*. 2018;7(6):40. <https://doi.org/10.1167/tvst.7.6.40>.
76. Zhang G, Fu DJ, Liefers B, Faes L, et al. Clinically relevant deep learning for detection and quantification of geographic atrophy from optical coherence tomography: a model development and external validation study. *Lancet Digit Health*. 2021;3(10):e665–75. [https://doi.org/10.1016/S2589-7500\(21\)00134-5](https://doi.org/10.1016/S2589-7500(21)00134-5).

77. “SCANLY® Home OCT,” Notal vision. Accessed 08 July 2024. [Online]. Available: <https://notalvision.com/services/{https://notalvision.info/SCANLY}>
78. Spaide RF, Ooto S, Curcio CA. Subretinal drusenoid deposits AKA pseudodrusen. *Surv Ophthalmol*. 2018;63(6):782–815. <https://doi.org/10.1016/j.survophthal.2018.05.005>.
79. Wightman AJ, Guymer RH. Reticular pseudodrusen: current understanding. *Clin Exp Optom*. 2019;102(5):455–62. <https://doi.org/10.1111/cxo.12842>.
80. Sadda SR, Guymer R, Holz FG, et al. Consensus definition for atrophy associated with age-related macular degeneration on OCT: classification of atrophy report 3. *Ophthalmology*. 2018;125(4):537–48. <https://doi.org/10.1016/j.ophtha.2017.09.028>.
81. Spaide RF. Outer retinal atrophy after regression of subretinal drusenoid deposits as a newly recognized form of late age-related macular degeneration. *Retina Phila Pa*. 2013;33(9):1800–8. <https://doi.org/10.1097/IAE.0b013e31829c3765>.
82. Fleckenstein M, Mitchell P, Freund KB, et al. The progression of geographic atrophy secondary to age-related macular degeneration. *Ophthalmology*. 2018;125(3):369–90. <https://doi.org/10.1016/j.ophtha.2017.08.038>.
83. Domalpally A, Agrón E, Pak JW, et al. Prevalence, risk, and genetic association of reticular pseudodrusen in age-related macular degeneration: age-related eye disease study 2 Report 21. *Ophthalmology*. 2019;126(12):1659–66. <https://doi.org/10.1016/j.ophtha.2019.07.022>.
84. Alten F, Clemens CR, Heiduschka P, Eter N. Characterisation of reticular pseudodrusen and their central target aspect in multi-spectral, confocal scanning laser ophthalmoscopy. *Graefes Arch Clin Exp Ophthalmol Albrecht Von Graefes Arch Klin Exp Ophthalmol*. 2014;252(5):715–21. <https://doi.org/10.1007/s00417-013-2525-y>.
85. Ueda-Arakawa N, Ooto S, Tsujikawa A, Yamashiro K, Oishi A, Yoshimura N. Sensitivity and specificity of detecting reticular pseudodrusen in multimodal imaging in Japanese patients. *Retina Phila Pa*. 2013;33(3):490–7. <https://doi.org/10.1097/IAE.0b013e318276e0ae>.
86. van Grinsven MJJP, Buitendijk GHS, Brussee C, et al. Automatic identification of reticular pseudodrusen using multimodal retinal image analysis. *Invest Ophthalmol Vis Sci*. 2015;56(1):633–9. <https://doi.org/10.1167/iovs.14-15019>.
87. Keenan TDL, Chen Q, Peng Y, et al. Deep learning automated detection of reticular pseudodrusen from fundus autofluorescence images or color fundus photographs in AREDS2. *Ophthalmology*. Published online May 21, 2020. <https://doi.org/10.1016/j.ophtha.2020.05.036>
88. Chen Q, Keenan TDL, Allot A, Peng Y, et al. “Multimodal, multitask, multiattention (M3) deep learning detection of reticular pseudodrusen: Toward automated and accessible classification of age-related macular degeneration. *J Am Med Inform Assoc*. 2021;12;28(6):1135–48. <https://doi.org/10.1093/jamia/ocaa302>
89. Van Grinsven MJJP, Buitendijk GHS, Brussee C, Van Ginneken B. Automatic identification of reticular pseudodrusen using multimodal retinal image analysis. *Invest Ophthalmol Vis Sci*. 2015;56(1):633–9. <https://doi.org/10.1167/iovs.14-15019>.
90. Schwartz R, Khalid H, Liakopoulos S, Ouyang Y. A deep learning framework for the detection and quantification of reticular pseudodrusen and drusen on optical coherence tomography. *TVST ARVO J*. Accessed 01 July 2024. [Online]. Available: <https://tvst.arvojournals.org/article.aspx?articleid=2783895>
91. He K, Chen X, Xie S, Li Y, Dollár P, Girshick R. Masked Autoencoders are scalable vision learners. Presented at the proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp 16000–16009. Accessed 01 July 2024. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2022/html/He_Masked_Autoencoders_Are_Scalable_Vision_Learners_CVPR_2022_paper
92. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. ImageNet: a large-scale hierarchical image database. In: 2009 IEEE Conference on computer vision and pattern recognition, June 2009, pp 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>
93. Garrity ST, Sarraf D, Freund KB, Sadda SR. Multimodal imaging of nonneovascular age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 2018;59(4):AMD48–64. <https://doi.org/10.1167/iovs.18-24158>
94. Holz FG, Sadda SR, Staurengi G, et al. Imaging protocols in clinical studies in advanced age-related macular degeneration: recommendations from classification of atrophy consensus meetings. *Ophthalmology*. 2017;124(4):464–78. <https://doi.org/10.1016/j.ophtha.2016.12.002>.
95. Beede E, Baylor E, Hersch F, Iurchenko A, Wilcox L, Ruamviboonsuk P, Vardoulakis LM. A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. In: Proceedings of the 2020 CHI conference on human factors in computing systems (CHI ‘20). 2020, pp. 1–12. <https://doi.org/10.1145/3313831.3376718>

Sowjanya Gowrisankaran, Ashkan Abbasi,
Wei-Chun Lin, Gadi Wollstein, Joel S. Schuman,
and Hiroshi Ishikawa

Abstract

This chapter explores the transformative role of artificial intelligence (AI) in glaucoma diagnosis and management, focusing on its ability to analyze complex structural and functional data for early detection and progression monitoring. Techniques such as convolutional neural networks (CNNs) have advanced the interpretation of imaging

modalities like fundus photography and optical coherence tomography (OCT), as well as functional assessments like visual field tests. Emerging applications include automated disease classification, structure–function mapping, and natural language processing (NLP) for leveraging electronic health records. The chapter also addresses challenges, including data integration, validation, and generalization, while highlighting AI’s potential to improve clinical decision-making and patient outcomes in glaucoma care.

S. Gowrisankaran · A. Abbasi · W.-C. Lin
Department of Ophthalmology, Oregon Health & Science University, Portland, OR, USA
e-mail: gowrisan@ohsu.edu

A. Abbasi
e-mail: abbasi@ohsu.edu

W.-C. Lin
e-mail: linw@ohsu.edu

G. Wollstein · J. S. Schuman
Ophthalmology and Biomedical Engineering, Wills Eye Hospital, Philadelphia, USA
e-mail: gwillstein@willseye.org

J. S. Schuman
e-mail: jschuman@willseye.org

H. Ishikawa (✉)
Department of Ophthalmology, Casey Reading Center, Casey Eye Institute, Portland, USA
e-mail: ishikawh@ohsu.edu

H. Ishikawa
Department of Medical Informatics and Clinical Epidemiology (DMICE), Oregon Health & Science University, Portland, OR, USA

Glaucoma is characterized by progressive loss of retinal ganglion cells (RGC) and their axons, which may result in optic nerve head (ONH) and retinal nerve fiber layer (RNFL) changes that eventually lead to vision loss and irreversible blindness [1–3]. As glaucoma is a slowly progressing disease with irreversible neural damage, early diagnosis and sensitive progression monitoring are essential to glaucoma management. For clinical assessment, structural (e.g., fundus photography (Fig. 9.1), optical coherence tomography (OCT, Fig. 9.2)) and functional (e.g., visual field (VF, Fig. 9.3)) measurements are commonly assessed in addition to conventional observations (e.g., optic disc assessment, intraocular pressure (IOP)). Various longitudinal studies on glaucoma progression reported contradicting non-linear relationships between

structural and functional measurements [4–10]. There are complex, non-linear, asynchronous interactions between them that are not fully understood.

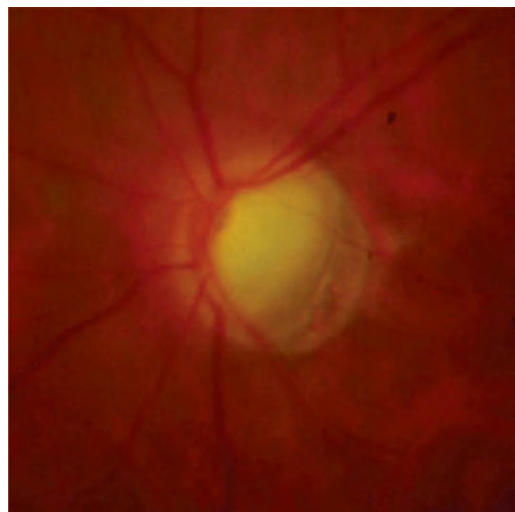


Fig. 9.1 Fundus photography of a left eye with glaucoma. Large cupping and peripapillary atrophy are shown in the image

Recently, the use of artificial intelligence (AI) based approaches for glaucoma diagnosis and monitoring has shown promising results in ophthalmology [11–15]. Deep learning (DL) is a class of state-of-the-art machine learning (ML) algorithms that are especially tailored to extract meaningful features from complex and high-dimensional data. Consequently, AI algorithms, especially DL, have the potential to revolutionize the diagnosis and management of glaucoma based on the interpretation of functional and/or structural information. These algorithms could also improve the understanding of glaucoma by defining the structural features responsible for certain functional damages and identifying phenotypes that follow similar progression patterns.

In addition, natural language processing (NLP), a subtype of deep learning (DL), has shown significant promise in utilizing electronic health records (EHRs) to enhance glaucoma management. EHRs contain vast amounts of structured and unstructured clinical data, and NLP techniques can effectively extract valuable information from these records, aiding in tasks

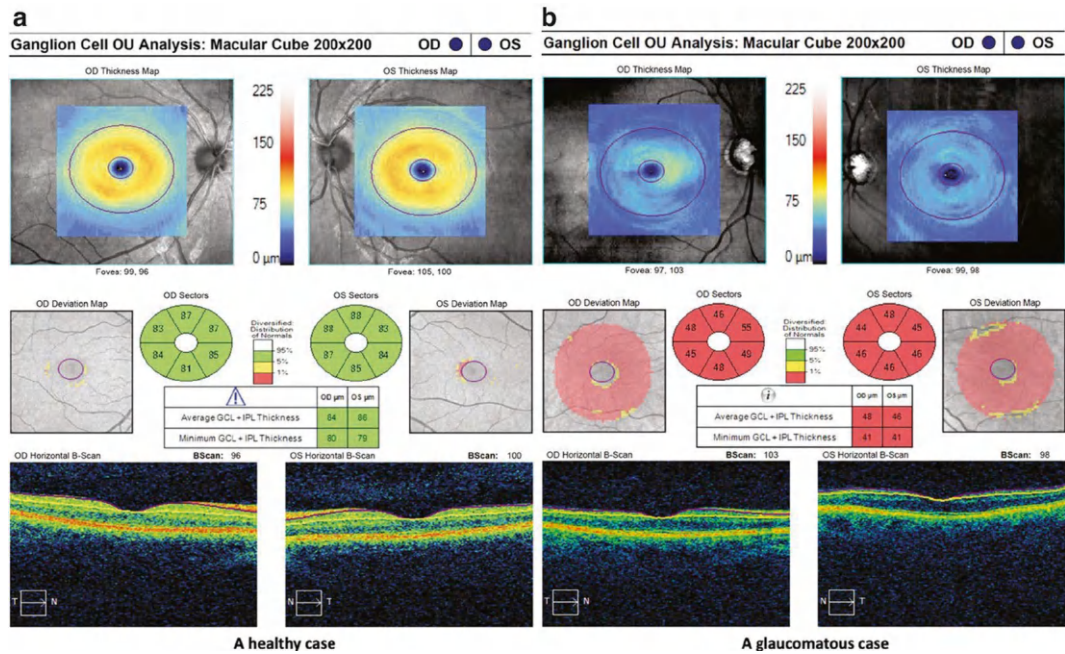


Fig. 9.2 Examples of Cirrus OCT report from a healthy case (a) and a glaucomatous case (b). The image is color-coded (red, orange, and yellow represent thicker areas while green and blue represent thinner areas)

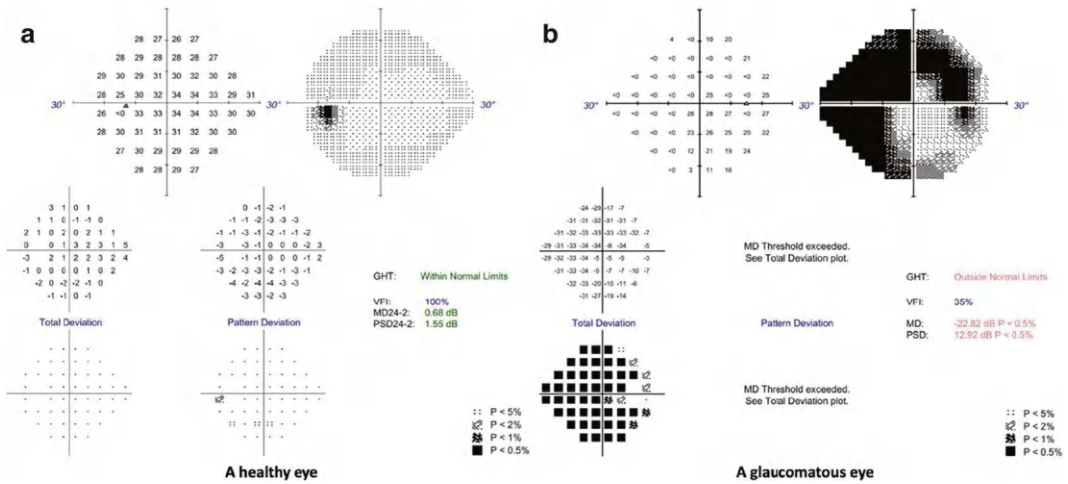


Fig. 9.3 Examples of Humphrey 24-2 VF report from a healthy eye (a) and a glaucomatous eye (b). (b) shows advanced visual field damages with superior nasal

deficit, large inferior nasal step and superior temporal step, and paracentral scotoma

such as data quality improvement, medication reconciliation, and outcome predictions. Recent advancements in large language models (LLMs) have further enhanced the capabilities of NLP in processing complex clinical narratives, thereby supporting more accurate and timely clinical decision-making.

In this chapter, we provide an overview of current AI applications and challenges in glaucoma. Section “[Glaucoma Diagnosis](#)” introduces AI utilization in detecting glaucoma; Section “[Longitudinal Analysis](#)” focuses on role of AI in longitudinal projection; Section “[Structure–Function Correlations](#)” summarizes developments of AI in finding the structure–function relationship; Section “[Other AI Applications in Glaucoma](#)” presents some other applications of AI in glaucoma; and finally, Section [Conclusion](#) describes the challenges related to using DL for glaucoma diagnosis and management.

Glaucoma Diagnosis

The diagnosis of glaucoma can be modeled as a classification problem, where clinical parameters or images are used as input to predict a diagnostic outcome, such as the presence or

severity of glaucoma. This area has been extensively explored with AI.

In 1994, machine learning classifiers were first used to distinguish between normal and glaucomatous eyes based on visual fields [16]. Subsequent studies expanded on this by incorporating more ML methods and data types, demonstrating the effectiveness of these models. Initially, the focus was on using clinical parameters and classical ML classifiers like random forests and support vector machines, which relied on manually designed, problem-specific features that required domain expertise. However, with the advent of deep learning, particularly CNNs, both feature extraction and classification have been automated, significantly improving the efficiency and accuracy of glaucoma diagnosis.

Functional Defects as Input

In clinical practice, visual field (VF) testing is the gold standard for diagnosing and assessing glaucoma. Several classical machine learning classifiers have been developed to automatically detect pre-perimetric glaucoma from VF data, showing promising results [16, 17]. With the development of computational capacity and

the curation of large datasets that are available for training, the implementation of deeper models has become possible. Asaoka et al. [18] proposed a multi-layer feed-forward neural network (FNN) with a stacked denoising autoencoder to classify pre-perimetric glaucoma VFs and healthy VFs and achieved better performance over shallower ML models.

Previous work showed promising performance in the classification of VFs. Yet, these methods treated each VF point as an individual feature and failed to leverage spatial information within VFs. Spatial information is useful to discover VF defect patterns and therefore helps glaucoma diagnosis [19]. Thus, incorporating spatial information into ML classifiers may boost their discrimination abilities. CNN is an evolution of FNN, which replaces matrix multiplication with convolution to process spatial information. Thus, researchers started to implement CNN models to discriminate VFs.

Kucur et al. [20] converted VFs to images using a Voronoi parcellation. A 7-layer CNN, which explicitly took spatial information into account through spatial convolutions, was used to discriminate between healthy and early glaucomatous VFs using the converted images as input. Results demonstrated the supremacy of CNN over FNN that did not consider spatial information (average precision score: 0.874 ± 0.095 vs. 0.843 ± 0.089). By computing the gradient, saliency maps can be obtained to visualize important pixels of the image that contribute most to the outputs of CNN. The saliency maps suggested that CNNs could detect patterns of localized VF defects.

Li et al. [21] took the probability map of pattern deviation (PD) from the VF reports as the input to a CNN. Results showed that CNN achieved higher accuracy compared to ophthalmologists, rule-based methods (Advanced Glaucoma Intervention Study (AGIS) criteria and Glaucoma Staging System (GSS) criteria), and traditional machine learning algorithms (SVM, RF, and k-nearest neighbor).

Structural Damages as Input

Assessment of structural damage has become a practical standard for glaucoma diagnosis. Early studies focused on structural measurements obtained from imaging techniques such as confocal scanning laser ophthalmoscopy (CSLO) and scanning laser polarimetry (SLP) [22, 23]. The promising performance of ML classifiers on structural parameters such as optic disc parameters measured by CSLO and RNFL measurements from SLP was reported. However, due to the popularity of the technologies, recent AI-based studies on structural glaucomatous damages have focused on fundus photography and OCT.

Fundus Photography

Fundus photography is a well-established and cost-effective imaging technique to identify features of the fundus region, including the fovea, macula, optic disc (OD), and optic cup (OC). Glaucoma can be identified by optic nerve cupping. Thus, cup-to-disc ratio (CDR), which measures the vertical diameter of the optic cup to that of the disc, is one of the most important biomarkers for glaucoma diagnosis. A higher CDR value indicates a higher probability of glaucoma. Therefore, many AI-based studies focused on automatic segmentation of OD and OC using deep learning [24]. Segmentation-based methods, however, lack sufficient discriminative representations and are easily affected by noise and low image quality. Moreover, pre-defined clinical parameters lack complex morphological information that might be useful in the diagnosis. Therefore, more recent methods simultaneously learn discriminative representations that optimize classification results directly from fundus images.

In 2015, Chen et al. [25] proposed a six-layer CNN to classify glaucoma and non-glaucoma eyes directly from fundus images from the publicly available ORIGA [26] and SCES [27] datasets. Experimental results showed AUCs of

0.831 and 0.887 on ORIGA and SCES, respectively. In 2018, Fu et al. [28] proposed a Disc-aware Ensemble Network (DENet) which consisted of four streams to integrate the hierarchical context of global fundus images and local OD regions. The first stream used a Residual Network (ResNet) to learn the global representation of the whole fundus image directly and produce a glaucoma classification probability. The second stream adapted the U-shape CNN (U-Net), which is a well-known DL model for image segmentation, to produce the disc probability map and a glaucoma classification probability. The third stream cropped OD region image as input and output a classification probability through a ResNet. The fourth stream applied the pixel-wise polar transformation to transfer the cropped original image to the polar coordinate system to enlarge the cup region and augment the data. Then, a ResNet was trained to output a classification probability. The model was trained on ORIGA dataset and yielded testing results of an accuracy of 0.832 on SCES and 0.666 on SINDI datasets.

Later, Li et al. [29] applied the Inception-v3 model on a private dataset to detect referable glaucomatous optic neuropathy (GON) and achieved an AUC of 0.986 with a sensitivity of 0.956 and a specificity of 0.92. Results also showed that other eye conditions would greatly affect the detection accuracy. Height or pathological myopia contributed most to false-negative results, while physiologic cupping and pathological myopia are the most common reasons for false-positive results.

Though previous methods demonstrated the efficiency of DL in glaucoma diagnosis, DL methods suffer from overfitting problems due to the relatively small dataset available and the many parameters needed for training. In 2018, Chakravarty et al. [30] presented a multi-task CNN that segmented OD and OC on fundus images and jointly classified the image to glaucoma and non-glaucoma. The proposed method was evaluated on the REFUGE dataset to achieve an average dice score (which measures the overlap of segmentations and ground truths) of 0.92 for OD segmentation, 0.84 for OC

segmentation, and AUC of 0.95 for classification. The cross-task design reduced the number of parameters and ensured good generalization of the model on small dataset. In another work, Chai et al. [31] designed a multi-branch neural network (MB-NN) model to leverage domain knowledge includes important measures (e.g., CDR) for glaucoma diagnosis. The first branch extracted hidden features directly from fundus image through a CNN. The second branch used Faster-RCNN which is a deep learning framework for object detection to obtain optic disc region. Then another CNN is used to extract local hidden features. The third branch used a fully convolutional network (FCN) to segment OD, OC, and peripapillary atrophy (PPA), and then calculated measures related to disc, cup, and PPA. RNFL defects, a roughly wedge shape region starting from OD, detected from another CNN and non-image features (e.g., age, IOP, eyesight and symptoms) from case reports were also inputs to the third branch. The proposed framework was verified on a private dataset and achieved an accuracy of 0.915, sensitivity of 0.9233, and specificity of 0.909. More recently, several studies have reported good accuracies for glaucoma diagnosis using fundus images [32–35]. Natarajan et al. [36] used a two-stage DL framework UNet-SNet, where the first stage uses a regularized UNet trained with optic disc regions of interest on the fundus image, and in the second stage a SqueezeNet is fine-tuned with deep features of optic disc to detect glaucoma vs non-glaucoma. The performance accuracy of 0.998, 0.9705 and 1.00 was reported on test sets ACRIMA, Drishti-GS1 and RIM-ONEv1. Although recent advances in DL have significantly improved the accuracy for glaucoma diagnosis using fundus images, real world application of these algorithms is still challenging due to the sub-optimal quality of fundus images in these settings.

Ocular Coherence Tomography (OCT)

OCT is a non-invasive imaging technique that provides micrometer resolution volumetric cross-sectional images of retina. OCT has emerged as the de facto standard in objective

quantification of structural damage in glaucoma. Early studies focused on comparison of various classical ML classifiers using parameters derived from the OCT images [37]. Although classification accuracy of these methods was satisfactory, their dependence on handcrafted features and the need for segmentation of retinal layers were limitations of these approaches.

In 2017, Muhammad et al. [38] used a pre-trained CNN model for feature extraction and a random forest model for classification. Though the proposed model is deep, the images fed into the model were still generated by conventional segmentation methods: (1) retinal ganglion cell plus inner plexiform layer (GCIPL) thickness map; (2) RNFL thickness map; (3) GCIPL probability map; (4) RNFL probability map; (5) *en face* projection. The results showed that the proposed method with the RNFL probability map as input outperformed other OCT and VF inputs but fell short of an experienced human expert.

In 2018, Fu et al. [39] proposed a Multi-Context Deep Network (MCDN) to classify angle-closure and open-angle glaucoma based on Anterior Segment Optical Coherence Tomography (AS-OCT). The anterior chamber angle (ACA) region was first localized by a data-driven AS-OCT structure segmentation method [28] to compute the clinical parameters (e.g., anterior chamber width, lens-vault, chamber Height, iris curvature, and anterior chamber area). A linear SVM was employed to predict an angle-closure probability based on these clinical parameters. Experimental results showed that the proposed method is effective for angle-closure glaucoma screening. Detailed analysis of three input streams showed that DL-based global discriminative features did not work as well as handcrafted visual features (AUC 0.894 vs. 0.924) while DL-based local discriminative features achieved on par performance with handcrafted features (AUC 0.920 vs. 0.924).

RNFL probability maps, which are generated based on swept-source optic coherence tomography (SS-OCT) to superimpose structural changes with VF locations, were also trained with CNNs to discriminate between glaucomatous and healthy eyes [40]. CAM analysis

suggested that anatomical variation in blood vessel or RNFL location caused ambiguity in false positive and false negative. This discovery might be useful for future improvement of DL systems by supplying information about blood vessels.

Thompson et al. [41] used circumpapillary b-scans without retinal layer segmentation and achieved an AUC of 0.96 for differentiating glaucoma and healthy scans using a CNN. Maetschke et al. [42] proposed a 3D CNN to be trained directly on raw spectral domain optical coherence tomography (SD-OCT) volumes of ONH, eliminating the need for segmentation of retinal layers, to classify healthy and glaucomatous eyes. The class activation map (CAM) analysis found neuroretinal rim, optic disc cupping, lamina cribrosa and its surrounding area to be significantly associated with the classification results, which aligned with commonly used clinical markers for glaucoma diagnosis. In 2023, Thierry AH et al. [43], proposed PointNet, a geometric deep learning approach, for glaucoma diagnosis based on 3D point clouds and showed comparable performance to 3DCNNs with an AUC of 0.95. Although this approach requires less information for glaucoma diagnosis compared to the whole volume scan, it requires segmentation of the retinal layers.

Combining Structure and Function

Many studies also developed effective ML classifiers combining structural and functional data. Global VF indices (mean defect, corrected loss variance, and short-term fluctuation) in combination with structural data (CDR, rim area, cup volume, and nerve fiber layer height) analyzed by an ANN was capable of correctly identifying glaucomatous eyes with an accuracy of 88% in an early study [44], which was higher than that of training with only structural or functional data. The development of computational ability accommodated larger models and larger inputs. Bowd et al. [45] took complete VF maps and OCT RNFL thickness measurements of 32 sectors to train multiple ML learning classifiers.

In general, the inclusion of structural, functional, and clinical data outperforms single modality information for glaucoma diagnosis [46]. Many studies have used different types of imaging data such as OCT, RNFL thickness maps and fundus photos as multiple modalities, and some have combined structural and functional information as multiple inputs to the model [47]. Xiong J et al. [46] showed that the use of paired visual field and circumpapillary OCT scans as inputs had higher performance for glaucoma diagnosis (AUC of 0.95 on the validation set and 0.873 on an external test set) compared to using either of those inputs separately [46].

Overall, DL models are capable of learning discriminative representations and identifying glaucoma patients. However, comparing those methods remains challenging because of the variety of training and testing datasets and validation methods. The possibility to extract knowledge that might not be discovered before, such as unknown glaucoma related structures/features that are highly associated with glaucomatous damages and glaucoma phenotyping, is the most exciting part of DL. Therefore, increasing interpretability of DL to visualize learned knowledge will be critical to future development of the use of DL in glaucoma diagnosis.

Longitudinal Analysis

The accelerated retinal ganglion cell loss is a characteristic feature of glaucoma progression together with functional damages. Therefore, identifying progression and estimating the rate of loss either structurally or functionally are crucial to glaucoma management.

The current clinical gold standard for progression analysis is the Guided Progression Analysis (GPA) provided by the commercial software developed by Carl Zeiss [48, 49]. The software allows clinicians to evaluate the patient's functional or structural loss over time compared to his or her own baseline, which is a composite of two initial examinations. Event-based and trend-based analysis are two

approaches to determine progression. Event-based analysis evaluates changes from baseline compared to expected variability. The expected variability is determined by the 95% confidence intervals of the magnitude of fluctuation of stable glaucoma patients from empirical datasets. Progression is defined as the change that exceeds expected variability. Trend-based analysis estimated the rate of change over time using linear regression. While GPA is useful to define and quantify glaucoma progression, GPA does not forecast future progressions, which has the potential to augment clinical decision making.

For VF forecasting, Caprioli et al. [50] projected individual VFs through an exponential model which characterized fast or slow progression rate in VF losses better than linear models. However, both linear model and exponential model assume constant loss rates of VF loss, which usually decay over time [51]. To better depict glaucomatous damage, Chen et al. [52] compared pointwise linear, exponential, logistic functions, and combinations of functions and showed that a combination of exponential and logistic functions predicted future progressions better. Previous methods treated test points as individual points and did not incorporate spatial correlations between VF test points at the time point. Several statistical methods have been proposed to incorporate spatio-temporal correlations in VFs [53–55].

Application of DL in this field of predictive medicine is particularly relevant to management of glaucoma since many factors contributing to the rate or severity of glaucoma progression is still unknown. But unlike the more definitive diagnosis of glaucoma, there have been limited investigation into the potential of DL in predicting future findings. Park et al. [56] developed a recurrent neural network (RNN) to predict the 6th visual field test. The performance of RNN was compared with that of a point-wise linear regression. Results showed that VFs predicted by RNN were more accurate than that by linear regression (root mean square error (RMSE): 4.31 ± 2.54 vs. 4.96 ± 2.76 , $p < 0.001$) and RNN was more robust. However, the proposed method required a large number of VF tests collected

over a long period of time to accurately predict the future VFs. To overcome this problem, Wen et al. [57] trained a deep learning model on the temporal history for a large group of patients to accurately predict future VFs up to 5.5 years given only a single VF test, with a correlation of 0.92 between MD on predicted VFs and MD on actual future VFs.

Herbert P et al. [58], used three baseline VFs along with the OCT RNFL thickness and clinical data corresponding to the first visual field to predict future worsening of visual field sensitivities. Their results showed that incorporating clinical and structural information improved prediction of future visual fields.

Most recently, Ishikawa et al. [59, 60] investigated the question of how far in the future a deep learning model can forecast pointwise visual field using a single visual field test. They developed a hybrid convolution and Transformer approach (CoTrNet) with promising results. These results are further improved with the use of visual field information obtained over two visits as inputs to an RNN and CoTrNet model. The CoTrNet also provides a unified framework that allows the use of different input modalities for predicting future visual field sensitivities. For structural progression forecasting, Song et al. [61] proposed a 2D continuous-time hidden Markov model to predict average circumpapillary RNFL thickness and VFI. Sedai et al. [62] developed a ML regressor to forecast circumpapillary RNFL thickness at the next visit from multimodal data including clinical (age and IOP), structural (circumpapillary RNFL thickness derived from OCT scans and DL-extracted OCT features), and functional (VF parameters) data of 3 prior visits and the inter-visit intervals. Chen et al. [63] also investigated the predictive DL in predicting structural loss. A time-aware long short-term memory network was designed to predict 5th visit of GCIPL thickness map based on 4 prior maps and took uneven intervals between every two visits into account.

Structure–Function Correlations

Relationships between structural loss and functional loss have been a controversial topic, with no consensus. Early work has investigated classical ML models such as LR [64], a Bayesian framework with a radial basis function [65, 66], Bayesian LR [67], and logarithmic regression [66] that map function from structure. However, model performance has been limited and highly dependent on assumptions of a linear relationship or the gaussian distribution of variability in VF measurements, which is not optimal given that it is usually heavily tailed. Given DL's success in identifying and forecasting glaucoma, it may help to improve understanding of the structure–function relationship in glaucoma. In addition, VF tests are subjective, time-consuming, and very noisy. Thus, estimating VF from OCT accurately may help reduce unnecessary VF testing.

In 2017, Uesaka et al. [68] proposed two methods to estimate SITA-Standard mode 10-2 using retinal thickness (RT) data, including GCIPL, RNFL, and RCL thickness maps. The first method, Affine-Structured Non-negative Matrix Factorization (ASNMF), could estimate visual fields while having explainability. The second method was based on CNNs. Results showed that ASNMF worked better for small datasets while CNN was powerful for large datasets. ASNMF achieved an average root mean squared error (RMSE) of 7.27 dB, while CNNs achieved 6.79 dB. In 2018, Sugiura et al. [69] proposed to use Pattern-Based Regularization (PBR) and Pattern-Based Visualization (PBV) with CNNs. PBR was shown to reduce overfitting when CNNs are trained on limited data by first pre-training the network on nonpaired VF data, followed by supervised learning to understand the relationship between RT and VF. PBV helped in visualizing the nonlinear relationships between RT and VF, which is useful for diagnosing glaucoma stages and pinpointing areas of the retina related to VF impairments. Overall, they reported an RMSE of 6.16 for pointwise VF estimation.

The aforementioned research papers employed segmentation-based features, known for their susceptibility to errors in advanced glaucoma and other ocular conditions. Instead of using these features, circumpapillary scans, which capture a circular area around the optic nerve head, can be utilized. In 2020, Christopher et al. [70] compared the performance of using unsegmented Spectralis circle scans versus RNFL thickness maps in estimating both global VF metrics and individual VF test points. They demonstrated significantly improved results using unsegmented circumpapillary scans. For pointwise VF estimation, they reported R2 values ranging from 0.07 to 0.71 dB for 24-2 VF tests and from 0.01 to 0.85 for 10-2 VF tests. Hemelings R et al. [71], estimated overall VF MD and point-wise sensitivities using four different imaging inputs: circumpapillary OCT scans at diameters of 3.5 mm, 4.1 mm, and 4.7 mm, along with scanning laser ophthalmoscope enface images. Their weighted ensemble of the four models achieved a mean absolute error (MAE) of 2.89 dB (2.5–3.3) for MD values and 4.82 dB (4.45–5.22) for point-wise estimates.

While unsegmented circumpapillary scans demonstrated improved results compared to segmentation-based features, leveraging OCT volumes directly may harness even more comprehensive data. In 2019, Maetschke et al. [72] inferred global VF indices (MD and VFI) directly from raw OCT volumes, outperforming classical machine learning methods that rely on segmentation-based features. To estimate pointwise VF tests, Chen et al. [73] demonstrated that a ResNet based method achieves considerably higher accuracy using raw OCT volume scans instead of RNFL thickness maps. They found that using raw OCT volumes reduced the influence of floor effects. Furthermore, they investigated the explainability of their models, demonstrating the models' ability to focus on clinically relevant regions. Building on these findings, in a subsequent study [74], they established a generalized pointwise spatial relationship between raw OCT volumes and VF tests using occlusion analysis of their pointwise VF estimation method. They showed that their

data-driven structure–function maps align with previously published maps, which typically rely on segmentation-based features or anatomical prior knowledge, whereas their new approach is entirely automated, data-driven, and independent of segmentation with higher spatial resolution.

Secondary Use of EHRs in Glaucoma Management

The secondary use of Electronic Health Records (EHRs) holds immense potential for improving glaucoma management through machine learning and natural language processing (NLP) methods. EHRs are a rich source of patient information, capturing longitudinal clinical data with both structured and unstructured components that can be leveraged for predicting glaucoma outcomes and extracting valuable clinical insights. NLP techniques can process unstructured text within EHRs for multiple tasks including information extraction, Named Entity Recognition (NER), question answering, and predictive analytics.

For glaucoma outcome prediction using EHR data, several studies have aimed to predict which glaucoma patients may require surgery or assess surgical outcomes. Baxter et al. [75] developed a multivariable logistic regression model to predict which patients with progressive disease would require surgery, achieving an AUC of 0.67. Additionally, Wang et al. [76] used word embeddings and structured clinical data to develop deep learning models to predict subsequent glaucoma surgeries. The model that incorporated both structured clinical features and input features from clinical notes achieved an AUC of 0.73 and an F1 score of 0.40, compared to using only structured clinical features (AUC, 0.66; F1, 0.34) and only clinical free-text features (AUC, 0.70; F1, 0.42). All models outperformed predictions from a glaucoma specialist's review of clinical notes (F1, 0.295).

On the other hand, Banna et al. [77] developed machine learning models to predict glaucoma surgical success. The random forest model

showed the highest performance with an accuracy of 0.68 and an AUC of 0.74. Furthermore, Lin et al. [78] developed multimodal deep learning models for multiclass glaucoma trabeculectomy surgery outcomes prediction using both structured EHR data and free-text operative notes. The multimodal neural network with transformer blocks had the highest performance, with a macro AUROC of 0.75 and an F1 score of 0.58. It outperformed baseline machine learning models that used structured EHR data alone and unstructured data alone. The study demonstrated that operative notes are an important source of predictive information, and that multiclass surgical outcome prediction can provide valuable insights for clinical decision-making.

In addition to outcome prediction, NLP facilitates comprehensive information extraction from EHRs. Techniques such as NER and relationship extraction can identify and categorize relevant clinical terms, such as medications, symptoms, diagnostic tests, and procedures. Lin et al. [79] presented a medication reconciliation method using an NER model to identify current medications and adherence from progress notes for glaucoma patients. The NER model achieved a micro-averaged F1 score of 0.955 across all categories, indicating that NLP tools can be used to improve patient care and data quality. Additionally, Wang et al. [80] used MedEx to extract glaucoma medications with 85% accuracy. Moreover, Hamid et al. [81] utilized text parsing to identify patients' noncompliance with their glaucoma medications.

More recently, large language models (LLMs) have garnered significant interest in ophthalmology due to their potential to understand clinical knowledge and provide accurate responses. Delsoz et al. [82] evaluated the capabilities of LLMs, such as the Chat Generative Pretrained Transformer (ChatGPT), in diagnosing glaucoma based on specific clinical case descriptions. ChatGPT was correct in 72.7% of cases, performing similarly or better than senior ophthalmology residents. Additionally, Carlà et al. [83] demonstrated the ability of ChatGPT-4 to analyze glaucoma case

descriptions and suggest accurate surgical plans for both routine and complex cases.

These applications highlight the transformative potential of NLP in harnessing the rich data within EHRs to enhance glaucoma management, predict outcomes, and improve the quality of clinical information extraction.

Other AI Applications in Glaucoma

One application of AI is to discover new knowledge in glaucoma. Mendoza et al. [84] developed a DL method to predict age, sex, and race based on Spectralis OCT RNFL circle scans from healthy individuals, glaucoma suspect, and glaucoma patients. A MAE (95% CI) of 4.5 years (3.9, 5.2) and a strong (R^2 (95% CI)) association of 0.73 between predicted and actual age were achieved for predicting age. AUC (95% CI) of predicting race and sex were 0.96 (0.86, 0.99) and 0.70 (0.57, 0.80), respectively. These results suggest that DL can learn demographic features including age, race, and sex that are not apparent to human observers. These results imply that there are still to be discovered patterns and knowledge in retinal OCT scans.

Another application of AI is to enhance OCT scans. Halupka et al. [85] presented a CNN using either mean squared error or a generative adversarial network (GAN) with Wasserstein distance and perceptual similarity to reduce speckle noise of OCT images from both healthy and glaucomatous eyes. The results demonstrated the effectiveness of CNNs to denoising OCT B-scans while preserving structural features of retinal layers. The recently proposed multiscale extension of the BM4D outperforms other denoising methods for 3D OCT in terms of preserving layer/feature information and visual quality of the image [86]. Such denoising methods could be extremely useful in the analysis pipeline and ensure the reliability of the following disease assessment.

Another recent application of DL is related to generating synthetic data, using state of the art generative modals, that could potentially be

useful for training of DL models that require large amounts of data to derive meaningful associations between the inputs and outputs. It is important however to consider ethical generation of such data that is equitable across race, gender etc. to avoid biased training samples.

Conclusion

In this chapter, we discussed the role of AI in glaucoma. Accurate automated diagnosis and prognosis of glaucoma may assist clinicians to increase efficiency, minimize diagnosis errors, and improve overall quality of glaucoma treatments. With its abilities to extract meaningful information from high dimensional and complex multi-modal data, AI may help to discover new biomarkers, patterns, or knowledge to improve the current understanding of glaucoma, which could be useful for promoting research and development of new treatments.

In additional, NLP applications have shown significant promise in the secondary use of EHRs for glaucoma management. By leveraging the advanced NLP techniques like LLMs, valuable insights can be extracted from unstructured clinical text, enabling tasks such as information extraction, named entity recognition, and predictive analytics. These advancements facilitate better outcome predictions, enhance clinical decision-making, and improve data quality, ultimately contributing to more effective glaucoma management.

There are still several challenges for clinical applications of AI in glaucoma. First, datasets used in many studies are small and collected from homogeneous populations while modern AI systems require very large training dataset and are often subject to numerous variabilities. Tremendous efforts would be required to collect a large and general dataset for glaucoma research. Second, the definition of glaucoma is not clear. Disagreements in the definition of the disease phenotypes often happen between experienced ophthalmologists. Therefore, it is hard to obtain high-quality ground-truth labels. Third, despite many efforts in increasing the

interpretability of AI models, AI models are still being considered as “black boxes”, which limits its clinical adoption. Lastly, the quality of EHR data in ophthalmology still needs improvement, and LLMs with domain knowledge in ophthalmology are still under development. Thus, it is crucial to develop more visualization tools for AI algorithms. Despite these challenges ahead, AI will likely have positive impact on research and clinical practice in glaucoma.

References

1. Tan O, Chopra V, Lu ATH, et al. Detection of macular ganglion cell loss in glaucoma by Fourier-domain optical coherence tomography. *Ophthalmology*. 2009;116(12).
2. Quigley H, Broman AT. The number of people with glaucoma worldwide in 2010 and 2020. *Br J Ophthalmol*. 2006;90(3). <https://doi.org/10.1136/bjo.2005.081224>
3. Ramulu P. Glaucoma and disability: which tasks are affected, and at what stage of disease? *Curr Opin Ophthalmol*. 2009;20(2). <https://doi.org/10.1097/ICU.0b013e32832401a9>
4. Hood DC, Tsamis E, Bommakanti NK, et al. Structure-function agreement is better than commonly thought in eyes with early glaucoma. *Invest Ophthalmol Vis Sci*. 2019;60(13). <https://doi.org/10.1167/iovs.19-27920>
5. Rao HL, Zangwill LM, Weinreb RN, Leite MT, Sample PA, Medeiros FA. Structure-function relationship in glaucoma using spectral-domain optical coherence tomography. *Archives Ophthalmol*. 2011;129(7). <https://doi.org/10.1001/archophthalmol.2011.145>
6. Leite MT, Zangwill LM, Weinreb RN, Rao HL, Alencar LM, Medeiros FA. Structure-function relationships using the cirrus spectral domain optical coherence tomograph and standard automated perimetry. *J Glaucoma*. 2012;21(1). <https://doi.org/10.1097/IJG.0b013e31822af27a>
7. Wollstein G, Kagemann L, Bilonick RA, et al. Retinal nerve fibre layer and visual function loss in glaucoma: The tipping point. *Br J Ophthalmol*. 2012;96(1). <https://doi.org/10.1136/bjo.2010.196907>
8. Malik R, Swanson WH, Garway-Heath DF. Structure-function relationship” in glaucoma: past thinking and current concepts. *Clin Exp Ophthalmol*. 2012;40(4). <https://doi.org/10.1111/j.1442-9071.2012.02770.x>
9. Harwerth RS, Wheat JL, Fredette MJ, Anderson DR. Linking structure and function in glaucoma. *Prog Retin Eye Res*. 2010;29(4). <https://doi.org/10.1016/j.preteyeres.2010.02.001>
10. Garg A, Hood DC, Pensec N, Liebmann JM, Blumberg DM. Macular damage, as determined by

- structure-function staging, is associated with worse vision-related quality of life in early glaucoma. *Am J Ophthalmol*. 2018;194. <https://doi.org/10.1016/j.ajo.2018.07.011>
11. Lee A, Taylor P, Kalpathy-Cramer J. Machine learning has arrived! *Ophthalmology*. 2017;124(12). <https://doi.org/10.1016/j.ophtha.2017.08.046>
 12. Rahimy E. Deep learning applications in ophthalmology. *Curr Opin Ophthalmol*. 2018;29(3).
 13. Ting DSW, Pasquale LR, Peng L, et al. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol*. 2019;103(2). <https://doi.org/10.1136/bjophthalmol-2018-313173>
 14. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA—J Am Med Assoc*. 2016;316(22). <https://doi.org/10.1001/jama.2016.17216>
 15. Chen X, Xu Y, Kee Wong DW, Wong TY, Liu J. Glaucoma detection based on deep convolutional neural network. In: Proceedings of the annual international conference of the IEEE engineering in medicine and biology society, EMBS, vol 2015-November; 2015. <https://doi.org/10.1109/EMBC.2015.7318462>
 16. Goldbaum MH, Sample PA, White H, et al. Interpretation of automated perimetry for glaucoma by neural network. *Invest Ophthalmol Vis Sci*. 1994;35(9).
 17. Chan K, Lee TW, Sample PA, Goldbaum MH, Weinreb RN, Sejnowski TJ. Comparison of machine learning and traditional classifiers in glaucoma diagnosis. *IEEE Trans Biomed Eng*. 2002;49(9). <https://doi.org/10.1109/TBME.2002.802012>
 18. Asaoka R, Murata H, Iwase A, Araie M. Detecting preperimetric glaucoma with standard automated perimetry using a deep learning classifier. *Ophthalmology*. 2016;123(9). <https://doi.org/10.1016/j.ophtha.2016.05.029>
 19. Sample PA, Chan K, Boden C, et al. Using unsupervised learning with variational Bayesian mixture of factor analysis to identify patterns of glaucomatous visual field defects. *Invest Ophthalmol Vis Sci*. 2004;45(8). <https://doi.org/10.1167/iovs.03-0343>
 20. Kucur SS, Holló G, Sznitman R. A deep learning approach to automatic detection of early glaucoma from visual fields. *PLoS One*. 2018;13(11). <https://doi.org/10.1371/journal.pone.0206081>
 21. Li F, Wang Z, Qu G, et al. Automatic differentiation of Glaucoma visual field from non-glaucoma visual field using deep convolutional neural network. *BMC Med Imaging*. 2018;18(1). <https://doi.org/10.1186/s12880-018-0273-5>
 22. Bowd C, Chan K, Zangwill LM, et al. Comparing neural networks and linear discriminant functions for glaucoma detection using confocal scanning laser ophthalmoscopy of the optic disc. *Invest Ophthalmol Vis Sci*. 2002;43(11).
 23. Bowd C, Medeiros FA, Zhang Z, et al. Relevance vector machine and support vector machine classifier analysis of scanning laser polarimetry retinal nerve fiber layer measurements. *Invest Ophthalmol Vis Sci*. 2005;46(4). <https://doi.org/10.1167/iovs.04-1122>
 24. Thakur N, Juneja M. Survey on segmentation and classification approaches of optic cup and optic disc for diagnosis of glaucoma. *Biomed Signal Process Control*. 2018;42. <https://doi.org/10.1016/j.bspc.2018.01.014>
 25. Chen X, Xu Y, Yan S, Wong DWK, Wong TY, Liu J. Automatic feature learning for glaucoma detection based on deep learning. In: Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics), vol 9351; 2015. https://doi.org/10.1007/978-3-319-24574-4_80
 26. Zhang Z, Yin FS, Liu J, et al. ORIGA-light: An online retinal fundus image database for glaucoma analysis and research. In: 2010 Annual international conference of the IEEE engineering in medicine and biology society, EMBC'10; 2010. <https://doi.org/10.1109/IEMBS.2010.5626137>
 27. Sng CC, Foo LL, Cheng CY, et al. Determinants of anterior chamber depth: The singapore chinese eye study. *Ophthalmology*. 2012;119(6). <https://doi.org/10.1016/j.ophtha.2012.01.011>
 28. Fu H, Cheng J, Xu Y, et al. Disc-aware ensemble network for glaucoma screening from fundus image. *IEEE Trans Med Imaging*. 2018;37(11). <https://doi.org/10.1109/TMI.2018.2837012>
 29. Li Z, He Y, Keel S, Meng W, Chang RT, He M. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. *Ophthalmology*. 2018;125(8):1199–206.
 30. Chakravarty A, Sivaswamy J. A Deep Learning based Joint segmentation and classification framework for glaucoma assesment in retinal color fundus images. *ArXiv*. 2018;abs/1808.01355.
 31. Chai Y, Liu H, Xu J. Glaucoma diagnosis based on both hidden features and domain knowledge through deep learning models. *Knowl Based Syst*. 2018;161. <https://doi.org/10.1016/j.knosys.2018.07.043>
 32. Aamir M, Irfan M, Ali T, et al. An adoptive threshold-based multi-level deep convolutional neural network for glaucoma eye disease detection and classification. *Diagnostics*. 2020;10(8). <https://doi.org/10.3390/diagnostics10080602>
 33. Islam MT, Mashfu ST, Faisal A, Siam SC, Naheen IT, Khan R. Deep learning-based glaucoma detection with cropped optic cup and disc and blood vessel segmentation. *IEEE Access*. 2022;10. <https://doi.org/10.1109/ACCESS.2021.3139160>
 34. Nawaz M, Nazir T, Javed A, et al. An efficient deep learning approach to automatic glaucoma detection using optic disc and optic cup localization. *Sensors* 2022;22(2). <https://doi.org/10.3390/s22020434>

35. Pascal L, Perdomo OJ, Bost X, Huet B, Otálora S, Zuluaga MA. Multi-task deep learning for glaucoma detection from color fundus images. *Sci Rep*. 2022;12(1). <https://doi.org/10.1038/s41598-022-16262-8>
36. Natarajan D, Sankaralingam E, Balraj K, Karuppusamy S. A deep learning framework for glaucoma detection based on robust optic disc segmentation and transfer learning. *Int J Imaging Syst Technol*. 2022;32(1). <https://doi.org/10.1002/ima.22609>
37. Barella KA, Costa VP, Gonçalves Vidotti V, Silva FR, Dias M, Gomi ES. Glaucoma diagnostic accuracy of machine learning classifiers using retinal nerve fiber layer and optic nerve data from SD-OCT. *J Ophthalmol*. 2013;2013. <https://doi.org/10.1155/2013/789129>
38. Muhammad H, Fuchs TJ, De Cuir N, et al. Hybrid deep learning on single wide-field optical coherence tomography scans accurately classifies glaucoma suspects. *J Glaucoma*. 2017;26(12):1086.
39. Fu H, Xu Y, Lin S, et al. Multi-context deep network for angle-closure glaucoma screening in anterior segment OCT. In: *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bio-informatics)*, vol 11071 LNCS; 2018. https://doi.org/10.1007/978-3-030-00934-2_40
40. Thakoor KA, Li X, Tsamis E, Sajda P, Hood DC. Enhancing the accuracy of glaucoma detection from OCT probability maps using convolutional neural networks. In: *The 41st Annual international conference of the IEEE engineering in medicine and biology society*; 2019, pp. 2036–2040.
41. Thompson AC, Jammal AA, Berchuck SI, Mariottoni EB, Medeiros FA. Assessment of a segmentation-free deep learning algorithm for diagnosing glaucoma from optical coherence tomography scans. *JAMA Ophthalmol*. 2020;138(4). <https://doi.org/10.1001/jamaophthalmol.2019.5983>
42. Maetschke S, Antony B, Ishikawa H, Wollstein G, Schuman J, Garnavi R. A feature agnostic approach for glaucoma detection in OCT volumes. *PLoS ONE*. 2019;14(7):0219126.
43. Thiéry AH, Braeu F, Tun TA, Aung T, Girard MJA. Medical application of geometric deep learning for the diagnosis of glaucoma. *Transl Vis Sci Technol*. 2023;12(2). <https://doi.org/10.1167/tvst.12.2.23>
44. Brigatti L, Hoffman D, Caprioli J. Neural networks to identify glaucoma with structural and functional measurements. *Am J Ophthalmol*. 1996;121(5). [https://doi.org/10.1016/S0002-9394\(14\)75425-X](https://doi.org/10.1016/S0002-9394(14)75425-X)
45. Bowd C, Hao J, Tavares IM, et al. Bayesian machine learning classifiers for combining structural and functional measurements to classify healthy and glaucomatous eyes. *Invest Ophthalmol Vis Sci*. 2008;49(3). <https://doi.org/10.1167/iovs.07-1083>
46. Xiong J, Li F, Song D, et al. Multimodal machine learning using visual fields and peripapillary circular OCT scans in detection of glaucomatous optic neuropathy. *Ophthalmology*. 2022;129(2). <https://doi.org/10.1016/j.ophtha.2021.07.032>
47. Li Y, Han Y, Li Z, Zhong Y, Guo Z. A transfer learning-based multimodal neural network combining metadata and multiple medical images for glaucoma type diagnosis. *Sci Rep*. 2023;13(1). <https://doi.org/10.1038/s41598-022-27045-6>
48. Leung CKS, Cheung CYL, Weinreb RN, et al. Evaluation of retinal nerve fiber layer progression in glaucoma: a study on optical coherence tomography guided progression analysis. *Invest Ophthalmol Vis Sci*. 2010;51(1). <https://doi.org/10.1167/iovs.09-3468>
49. Na JH, Sung KR, Baek S, Lee JY, Kim S. Progression of retinal nerve fiber layer thinning in glaucoma assessed by cirrus optical coherence tomography-guided progression analysis. *Curr Eye Res*. 2013;38(3). <https://doi.org/10.3109/02713683.2012.742913>
50. Caprioli J, Mock D, Bitrian E, et al. A method to measure and predict rates of regional visual field decay in glaucoma. *Invest Ophthalmol Vis Sci*. 2011;52(7). <https://doi.org/10.1167/iovs.10-6414>
51. Otarola F, Chen A, Morales E, Yu F, Afifi A, Caprioli J. Course of glaucomatous visual field loss across the entire perimetric range. *JAMA Ophthalmol*. 2016;134:496–502.
52. Chen A, Nouri-Mahdavi K, Otarola FJ, Yu F, Afifi AA, Caprioli J. Models of glaucomatous visual field loss. *Invest Ophthalmol Vis Sci*. 2014;55(12). <https://doi.org/10.1167/iovs.14-15435>
53. Warren JL, Mwanza JC, Tanna AP, Budenz DL. A statistical model to analyze clinician expert consensus on glaucoma progression using spatially correlated visual field data. *Transl Vis Sci Technol*. 2016;5(4). <https://doi.org/10.1167/tvst.5.4.14>
54. Betz-Stablein BD, Morgan WH, House PH, Hazelton ML. Spatial modeling of visual field data for assessing glaucoma progression. *Invest Ophthalmol Vis Sci*. 2013;54(2). <https://doi.org/10.1167/iovs.12-11226>
55. Anderson AJ. Comparison of three parametric models for glaucomatous visual field progression rate distributions. *Transl Vis Sci Technol*. 2015;4(4). <https://doi.org/10.1167/tvst.4.4.2>
56. Park K, Kim J, Lee J. Visual field prediction using recurrent neural network. *Sci Rep*. Published online 2019
57. Wen JC, Lee CS, Keane PA, et al. Forecasting future Humphrey visual fields using deep learning. *PLoS One*. 2019;14(4):1–14. Vavvas DG, ed.
58. Herbert P, Hou K, Bradley C, et al. Forecasting risk of future rapid glaucoma worsening using early visual field, OCT, and clinical data. *Ophthalmol Glaucoma*. 2023;6(5). <https://doi.org/10.1016/j.ogla.2023.03.005>

59. Abbasi A, Gowrisankaran S, Antony BJ, et al. A unified framework for visual field test estimation and forecasting using convolution and attention networks. *Invest Ophthalmol Vis Sci.* 2024;65(7):2361.
60. Ishikawa H, Abbasi A, Gowrisankaran S, et al. How far in the future can a deep learning model forecast pointwise visual field (VF) data based solely on one VF data input. *Invest Ophthalmol Vis Sci.* 2024;65(7):373.
61. Song Y, Ishikawa H, Wu M, et al. Clinical prediction performance of glaucoma progression using a 2-dimensional continuous-time hidden Markov model with structural and functional measurements. *Ophthalmology.* 2018;125(9). <https://doi.org/10.1016/j.ophtha.2018.02.010>
62. Sedai S, Antony B, Ishikawa H, Wollstein G, Schuman JS, Garnavi R. Forecasting retinal nerve fiber layer thickness from multimodal temporal data incorporating OCT volumes. *Ophthalmol Glaucoma.* 2020;3(1):14–24.
63. Chen Z, Wang Y, Wollstein G, Ramos-Cadena M de los A, Schuman J, Ishikawa H. Macular GCIPL thickness map prediction via time-aware convolutional LSTM. In: 2020 IEEE 17th International symposium on biomedical imaging (ISBI); 2020:1–5. <https://doi.org/10.1109/ISBI45749.2020.9098614>
64. Hood DC, Kardon RH. A framework for comparing structural and functional measures of glaucomatous damage. *Prog Retin Eye Res.* 2007;26(6). <https://doi.org/10.1016/j.preteyeres.2007.08.001>
65. Zhu H, Crabb DP, Schlottmann PG, et al. Predicting visual function from the measurements of retinal nerve fiber layer structure. *Invest Ophthalmol Vis Sci.* 2010;51(11). <https://doi.org/10.1167/iovs.10-5239>
66. Pollet-Villard F, Chiquet C, Romanet JP, Noel C, Aptel F. Structure-function relationships with spectral-domain optical coherence tomography retinal nerve fiber layer and optic nerve head measurements. *Invest Ophthalmol Vis Sci.* 2014;55(5). <https://doi.org/10.1167/iovs.13-13482>
67. Russell RA, Malik R, Chauhan BC, Crabb DP, Garway-Heath DF. Improved estimates of visual field progression using Bayesian linear regression to integrate structural information in patients with ocular hypertension. *Invest Ophthalmol Vis Sci.* 2012;53(6). <https://doi.org/10.1167/iovs.11-7976>
68. Uesaka T, Morino K, Sugiura H, et al. Multi-view learning over retinal thickness and visual sensitivity on glaucomatous eyes. In: Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining, vol Part F129685; 2017. <https://doi.org/10.1145/3097983.3098194>
69. Sugiura H, Murata H, Kiwaki T, Asaoka R, Yousefi S, Yamanishi K. Estimating glaucomatous visual sensitivity from retinal thickness with pattern-based regularization and visualization. In: Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining; 2018. <https://doi.org/10.1145/3219819.3219866>
70. Christopher M, Proudfoot JA, Bowd C, et al. Deep learning models based on unsegmented OCT RNFL circle scans provide accurate detection of glaucoma and high resolution prediction of visual field damage. *Invest Ophthalmol Vis Sci.* 2020;61(7).
71. Hemelings R, Elen B, Barbosa-Breda J, et al. Pointwise visual field estimation from optical coherence tomography in glaucoma using deep learning. *Transl Vis Sci Technol.* 2022;11(8). <https://doi.org/10.1167/tvst.11.8.22>
72. Maetschke S, Antony B, Ishikawa H, Wollstein G, Schuman JS, Garnavi R. Inference of visual field test performance from OCT volumes using deep learning. Published online August 5, 2019.
73. Chen Z, Shemuelian E, Wollstein G, Wang Y, Ishikawa H, Schuman JS. Segmentation-free OCT-volume-based deep learning model improves pointwise visual field sensitivity estimation. *Transl Vis Sci Technol.* 2023;12(6).
74. Chen Z, Ishikawa H, Wang Y, Wollstein G, Schuman JS. Deep-learning-based group pointwise spatial mapping of structure to function in glaucoma. *Ophthalmol Sci.* 2024;4(5): 100523.
75. Baxter SL, Marks C, Kuo TT, Ohno-Machado L, Weinreb RN. Machine learning-based predictive modeling of surgical intervention in glaucoma using systemic data from electronic health records. *Am J Ophthalmol.* 2019;208. <https://doi.org/10.1016/j.ajo.2019.07.005>
76. Wang SY, Tseng B, Hernandez-Boussard T. Deep learning approaches for predicting glaucoma progression using electronic health records and natural language processing. *Ophthalmol Sci.* 2022;2(2). <https://doi.org/10.1016/j.xops.2022.100127>
77. Banna HU, Zanolli A, McMillan B, et al. Evaluation of machine learning algorithms for trabeculectomy outcome prediction in patients with glaucoma. *Sci Rep.* 2022;12(1). <https://doi.org/10.1038/s41598-022-06438-7>
78. Lin WC, Chen A, Song X, Weiskopf NG, Chiang MF, Hribar MR. Prediction of multiclass surgical outcomes in glaucoma using multimodal deep learning based on free-text operative notes and structured EHR data. *J Am Med Inf Assoc.* 2024;31(2). <https://doi.org/10.1093/jamia/ocad213>
79. Lin WC, Chen JS, Kaluzny J, Chen A, Chiang MF, Hribar MR. Extraction of active medications and adherence using natural language processing for glaucoma patients. *AMIA Annu Symp Proc.* 2021;2021.
80. Wang SY, Pershing S, Tran E, Hernandez-Boussard T. Automated extraction of ophthalmic surgery outcomes from the electronic health record. *Int J Med Inform.* 2020;133. <https://doi.org/10.1016/j.ijmedinf.2019.104007>

81. Hamid MS, Valicevic A, Brennenman B, Niziol LM, Stein JD, Newman-Casey PA. Text parsing-based identification of patients with poor glaucoma medication adherence in the electronic health record. *Am J Ophthalmol.* 2021;222. <https://doi.org/10.1016/j.ajo.2020.09.008>
82. Delsoz M, Raja H, Madadi Y, et al. The use of ChatGPT to assist in diagnosing glaucoma based on clinical case reports. *Ophthalmol Ther.* 2023;12(6). <https://doi.org/10.1007/s40123-023-00805-x>
83. Carlà MM, Gambini G, Baldascino A, et al. Large language models as assistance for glaucoma surgical cases: a ChatGPT versus. Google Gemini comparison. *Graefe's Archive Clin Exp Ophthalmol.* Published online 2024. <https://doi.org/10.1007/s00417-024-06470-5>
84. Mendoza L, Christopher M, Belghith A, et al. Deep learning models predict age, sex and race from OCT optic nerve head circle scans. *Invest Ophthalmol Vis Sci.* 2020;61(7).
85. Halupka KJ, Antony BJ, Lee MH, et al. Retinal optical coherence tomography image enhancement via deep learning. *Biomed Opt Express.* 2018;9(12). <https://doi.org/10.1364/boe.9.006205>
86. Abbasi A, Monadjemi A, Fang L, Rabbani H, Antony BJ, Ishikawa H. Mixed multiscale BM4D for three-dimensional optical coherence tomography denoising. *Comput Biol Med.* 2023;155. <https://doi.org/10.1016/j.compbimed.2023.106658>

Artificial Intelligence in Retinopathy of Prematurity

10

Brittni A. Scruggs, Adam M. Hanif, Michael F. Chiang,
and J. Peter Campbell

Abstract

This chapter provides an overview of deep learning (DL), its history, and key innovations that contributed to its success. It reviews diverse approaches of learning from data, along with common DL methods and architectures used in ophthalmic applications, such as image enhancement, segmentation, and disease classification, where DL models

have achieved remarkable accuracy, even surpassing human performance have been reviewed. Beginning with the basic feedforward neural network architecture, the chapter progressively introduces more advanced methods, including convolutional neural networks, recurrent neural networks, transformers, and generative adversarial networks. Recent advancements, including language models and their applications in ophthalmology, are also discussed. The chapter emphasizes how DL has transformed ophthalmic imaging and diagnosis by leveraging various data-driven approaches.

Grant information: This chapter was supported by grants R01 EY019474, R01 EY031331, R21 EY031883, and P30 EY010572, from the National Institutes of Health (Bethesda, MD), the Malcolm M. Marquis, MD Endowed Fund for Innovation, and by unrestricted departmental funding and a Career Development Award (JPC) from Research to Prevent Blindness (New York, NY). AMH is supported by funding from the Heed Ophthalmic Foundation and the VitreoRetinal Surgery Foundation. The funding organizations had no role in the design or conduct of this research.

B. A. Scruggs
Department of Ophthalmology, Mayo Clinic,
Rochester, MN, USA

A. M. Hanif · J. Peter Campbell (✉)
Department of Ophthalmology, Casey Eye Institute,
Oregon Health & Science University,
Portland, OR, USA
e-mail: campbelp@ohsu.edu

M. F. Chiang
National Eye Institute, National Institutes of Health,
Bethesda, MD, USA
e-mail: michael.chiang@nih.gov

Keywords

Deep learning · Retinopathy of prematurity

Introduction

Retinopathy of prematurity (ROP) is a leading cause of preventable childhood blindness worldwide. Extremely preterm infants are at risk of developing ROP given their low gestational age and low birth weight [1, 2]. There are a number of challenges for ROP screening and diagnosis using current technology. ROP screening requires either bedside ophthalmoscopic screening or telemedicine using remote

interpretation of digital fundus imaging. There are several potential challenges to ensuring every at-risk baby is diagnosed accurately and on time. Further, ROP diagnosis is sub-classified by zone, stage, and vascular changes, with each area demonstrating significant intra- and inter-expert subjectivity and disagreement.

Automated image analysis and deep learning systems for ROP have the potential to improve ROP care by improving the efficiency and accuracy of diagnosis and by facilitating quantitative disease monitoring and risk prediction [3]. This chapter focuses on the limitations of current methods for ROP diagnosis and highlights the recent major advances and the clinical and technical challenges of artificial intelligence (AI) for automated diagnosis of ROP in the real world.

Risk Factors and Prevalence

The Multicenter Study of Early Treatment for Retinopathy of Prematurity (ET-ROP) study found that 68% of infants born <1251 g developed mild ROP or worse [4]. The cryotherapy for ROP (CRYO-ROP) study found that every additional gestational week led to a 19% decrease in developing severe ROP requiring treatment (i.e., threshold ROP), whereas each 100 g increase in birth weight was associated with a 27% decrease in threshold ROP [5]. ROP disease and severity are increased, including in mature babies, in countries with lack of oxygen saturation monitors and little to no advanced training of neonatal staff and/or ophthalmologists [6].

Worldwide Impact

Blindness from ROP is often preventable with appropriate primary, secondary, and tertiary prevention [7]. However, approximately 16% of surviving infants born under 32 gestational weeks develop some degree of ROP [8]. A large population study of 184,700 preterm babies with ROP found that 20,000 babies became blind or severely visually impaired by ROP in 2010 [8]. The Global Burden of Disease study estimated

that 257,000 years lived with disability worldwide in 2010 were associated with visual impairment secondary to ROP [9]. The greatest burden of disease currently is in Southeast Asia, Latin America, and North Africa [8]. The worldwide incidence will likely increase due to improved infant survival rates at increasingly younger gestational ages.

Pathophysiology

ROP is a vasoproliferative disease that occurs due to incomplete vascularization and non-physiologic hyperoxia [10]. Fetal retinal vascular development transitions from a process of vasculogenesis (i.e., in situ differentiation and growth of blood vessels) to angiogenesis (i.e., remodeling and expansion of new blood vessels) after 20 weeks gestational age [10, 11]. During the angiogenesis phase, physiologic hypoxia with an oxygen saturation of 60–70% is maintained in utero; however, oxygen saturation increases to >88% after birth with many preterm babies requiring high levels of supplemental oxygen for prevention of serious sequelae, including death [10].

High and variable oxygen levels in the neonatal intensive care unit (NICU) and lack of maternal protection decrease intraocular oxygen-related growth factors, such as insulin growth factor (IGF-1) and vascular endothelial growth factor (VEGF), lead to vaso-obliteration (Phase I ROP) [10]. As the infant ages (usually >32 gestational weeks), the VEGF and IGF-1 drive increases proportionally to the amount of avascular retina present and the degree of hypoxia (Phase II ROP) [10]. This proliferative phase can lead to retinal neovascularization, vitreoretinal traction, retinal detachments, and blindness. Screening of preterm babies prior to phase II ROP (approximately 31 gestational weeks) is imperative for achieving good outcomes as ROP often gets worse prior to improving. The average involution of ROP has been reported as 38.6 weeks [12]. Screening guidelines and current limitations are discussed later in this chapter.

Screening

Timely and accurate ROP diagnosis is essential to prevent blindness from ROP. ROP screening guidelines have been jointly published and updated by the American Academy of Pediatrics, the American Association for Pediatric Ophthalmology and Strabismus, and the American Academy of Ophthalmology [2, 13]. All infants ≤ 1500 g birth weight or ≤ 30 weeks gestational age should be screened at 31 weeks gestational age or four weeks chronological age, whichever is later. Selected infants with higher birth weights and/or gestational age > 30 weeks may benefit from ROP screening; such infants include, but are not limited to, those with increased oxygen requirement and/or severe comorbidities (e.g., lung disease, sepsis, necrotizing enterocolitis, profound anemia).

Limitations in Screening

Only approximately 5–10% of babies within a screening population will develop sight-threatening ROP. Unfortunately, there are often barriers to ensuring consistent screening of at-risk babies especially in low- and middle-income countries (LMIC), including lack of equipment, inadequate training, personnel shortages, and inconsistent examinations between clinicians [3, 14]. The population at risk in these regions is significantly higher due to differences in the level of oxygen regulation. There are also wide disparities in the distribution of ophthalmologists between rural vs. urban settings and between countries. Even when screening is performed in a timely fashion with trained ophthalmologists, the use of indirect ophthalmoscopy remains highly subjective and not interpretable. Preventable blindness due to ROP occurs in an estimated 20,000–50,000 infants annually due to missed, delayed, or incorrect diagnoses [15, 16]. Additional limitations of ROP diagnosis using ophthalmoscopy will be discussed later in this chapter. The use of AI for quantitative ROP diagnosis may enable monitoring of disease

severity between NICUs within a geographic region and over time.

Classification

ROP is classified based the location (zone), extent, and severity of disease (stage and vascular changes) according to the International Classification of Retinopathy of Prematurity (ICROP) guidelines developed in 1984 and subsequently updated in 2005 and 2021 [17–20]. Zone I subtends approximately 30° centered around the optic disc, though the precise definition is a circle centered at the optic disc with radius equal to twice the distance from the fovea to the optic disc. Zone II extends from zone I to the temporal equator and to a tangential point on the nasal ora serrata. The remaining crescent, zone III, extends from zone II anterior to the ora serrata (Fig. 10.1).

Incomplete vascularization (i.e., immature vessels) in zone I or II with persistent avascular retina is considered by some to be “no ROP”, or occasionally “Stage 0” ROP. Stage 1 ROP refers to the presence of thin demarcation line separating the vascularized retina from the anterior avascular region. Progression leads to stage 2 ROP where the demarcation line thickens as a ridge and, ultimately, to stage 3 ROP with extraretinal fibrovascular proliferation along the ridge. Extrafoveal retinal detachments (stage 4A) and sub-total foveal-involving retinal detachments (stage 4B) occur due to the development of neovascularization of the retina with associated tractional membranes [17]. Extraretinal proliferation of nonvascular tissue and further traction can lead to stage 5 ROP, or total retinal detachment. The overall stage is determined by the individual clock hour with the worse stage. Figure 10.1 includes examples of ROP disease in different zones and various stages.

Vascular abnormalities generally increase with more posterior disease and with higher stage and extent of peripheral disease [21]. Plus disease is defined as venous dilation and

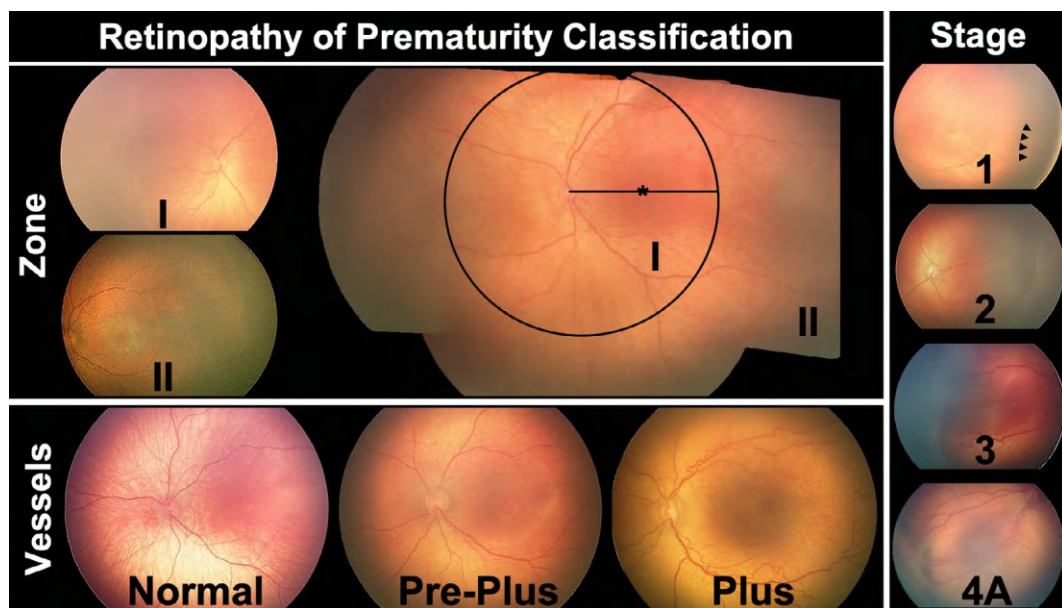


Fig. 10.1 ROP classification by zone, stage, and vessel changes. The top row shows mild ROP disease in zone I (upper panel) and in zone II (bottom panel). The montage photo shows the location of zones I and II. The asterisk indicates the location of the fovea. The bottom row depicts images with a label of normal vessels, pre-plus, or plus disease based on multiple (> 3) expert consensus from ophthalmoscopy and image grading

of vessel tortuosity and dilation. The far-right column depicts representative stages 1 through 4A. Black arrowheads highlight the faint demarcation line in an eye with stage 1 ROP. Note the temporal ridge in stage 2, the neovascularization present in stage 3, and the localized temporal retinal detachment in stage 4A. Stage 4B (sub-total retinal detachment with macula involvement) and stage 5 (total retinal detachment) are not shown

arteriolar tortuosity in two or more quadrants within the posterior pole and greater than a standard published photograph [22]. An intermediate level, pre-plus disease, was introduced to the 2005 ICROP classification suggesting that vascular changes present on a continuum, a concept that was formally acknowledged in the most recent ICROP update in 2021 [20]. In rare cases, infants may develop aggressive ROP (A-ROP) with progressive, often posterior, ROP with marked plus disease; these eyes often manifest flat neovascularization that can be difficult to appreciate.

Treatment

‘Threshold ROP’ is a term for ROP requiring treatment as defined by the CRYO-ROP study as five or more contiguous or eight total clock-hours of stage 3 ROP in zone I or II in

the presence of plus disease [12]. The Early Treatment for ROP (ET-ROP) trial further classified ROP into type 1 and type 2 pre-threshold treatment to guide the treatment of infants with early laser before the development of threshold ROP [4]. Type 1 ROP remains the currently accepted treatment cutoff for ROP and is defined as (1) zone I stage 3 without plus disease; (2) any stage in zone I with plus disease; or (3) zone II stage 2 or 3 with plus disease. Despite these well-defined guidelines, a high degree of variability in treatment paradigms remains between physicians, with disagreement rates as high as 50% between clinicians at the bedside and a masked image-based reading center [23].

Indirect laser photocoagulation ablates retina that is driving VEGF production. Although a mainstay treatment, laser is associated with strabismus and high myopia [24], and incomplete treatment such as skip lesions can lead to ROP progression and retinal detachment

despite treatment. The Bevacizumab Eliminates the Angiogenic Threat of Retinopathy of Prematurity (BEAT-ROP) and Ranibizumab versus laser therapy for the treatment of very low birthweight infants with retinopathy of prematurity (RAINBOW) trials demonstrated the utility of using VEGF inhibitors (intravitreal bevacizumab or ranibizumab, respectively) instead of laser in certain cases, such as zone I stage 3 ROP with plus disease [25, 26]. Despite encouraging results showing fewer unfavorable ocular outcomes than laser therapy, intravitreal therapy for ROP introduces new challenges, including the possibility of end-organ effects from systemic exposure to anti-VEGF medications and the need for increased monitoring, potentially for years, to assess for reactivation. Automated computer systems may soon help clinicians decide which treatment is warranted for an

individual infant and the risk of post-treatment progression.

Limitations in ROP Diagnosis and Management

Assessment of ROP diagnosis and severity depends on the subjective evaluation of zone, stage, and plus disease, and it is well established that there is wide inter-observer variability for all three components despite good relative agreement on ROP disease severity [27, 28]. Clinical images demonstrating high inter-observer variability among ROP experts (stage 1 vs. 2; pre-plus vs. plus disease) are provided in Fig. 10.2. Identifying plus disease remains the most critical finding for diagnosing threshold disease. However, agreement on plus disease

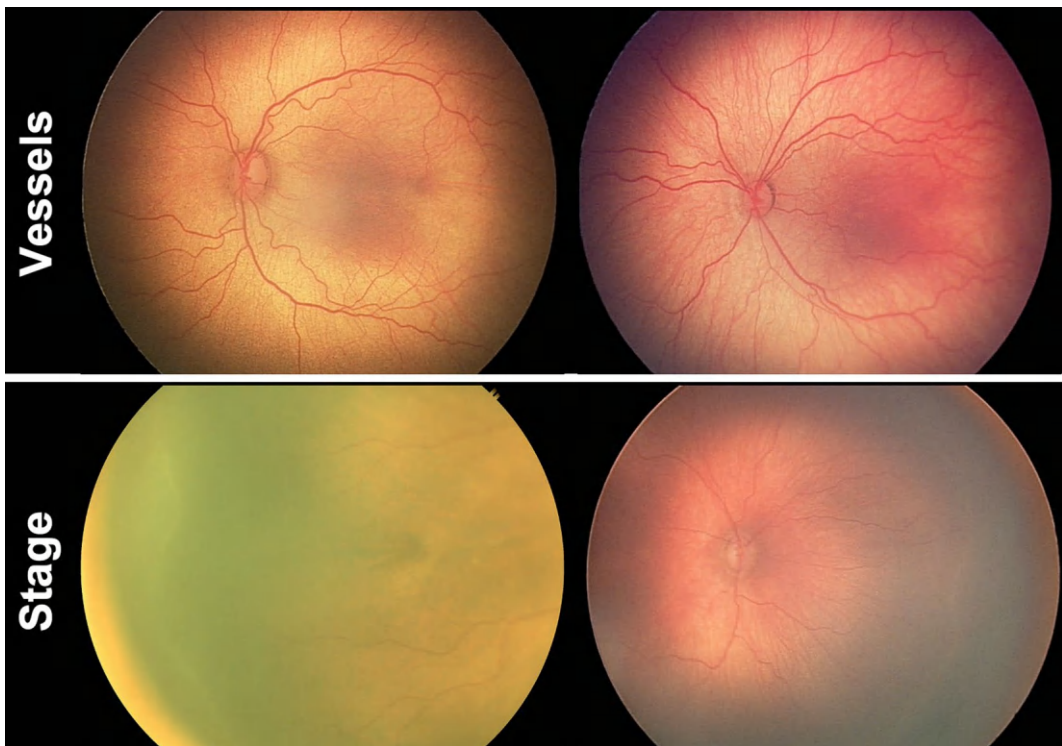


Fig. 10.2 Real world images with high inter-observer variability among ROP experts. The top row shows two fundus photos depicting arteriolar tortuosity and venous dilation in the posterior pole. These are photos from a telemedicine screening program that were graded as

either pre-plus or plus disease with no consensus by ROP experts. Similarly, the bottom row shows two telemedicine photos that some ROP experts documented stage 1 ROP, whereas others documented stage 2 ROP

between experts is imperfect due to varying diagnostic and treatment thresholds for a disease inherently manifesting along a continuum [27]. Gelman et al. found that when diagnosing plus disease, 22 experts had sensitivities and specificities ranging from 0.31–1.00 and 0.57–1.00, respectively, with agreement in only 21% of the images. Cole et al. found that reviewing both single and three serial retinal images from consecutive exams, graders demonstrated wide variability in classifying plus disease, demonstrating a Cohen's kappa from 0.29 to 1.0 [29].

Ghergherehchi et al. proposed that the variability in plus disease diagnosis is partly due to attention to ambiguous criteria for grading vascular features [22]. For example, peripheral vessel tortuosity can be severe, despite appearing normal posteriorly (Fig. 10.3). This is important because the standard photographs for ROP diagnosis are captured within a narrower field of view (FOV) obtained with bedside examination and/or telemedicine photos. Wide FOV allow examiners to focus on different parts of the retina than originally described in ICROP,

and inter-expert agreement is higher in plus disease diagnosis using wide-angle images [30]. Kim et al. found lower accuracy when clinicians diagnosed plus disease one quadrant at a time, suggesting that clinicians subconsciously evaluate the whole eye even when they intend to carefully evaluate plus by quadrant [31]. Figure 10.3 demonstrates the effect of FOV on severity appearance.

Most examiners do not routinely perform photography at the time of examination, and limited objective data may contribute to the significant inter-expert variability across different regions [15, 16, 22]. Increased use of photography offers serial comparisons for monitoring ROP disease and for improving ROP training. The lack of objective diagnosis of ROP and the high rates of inter-observer variability have been hindrances to the development of AI technology in ROP and have contributed to the high medicolegal risk of ROP screening. Such limitations have inspired the development of numerous computer-based systems for ROP, including the i-ROP deep learning (DL) system, which may

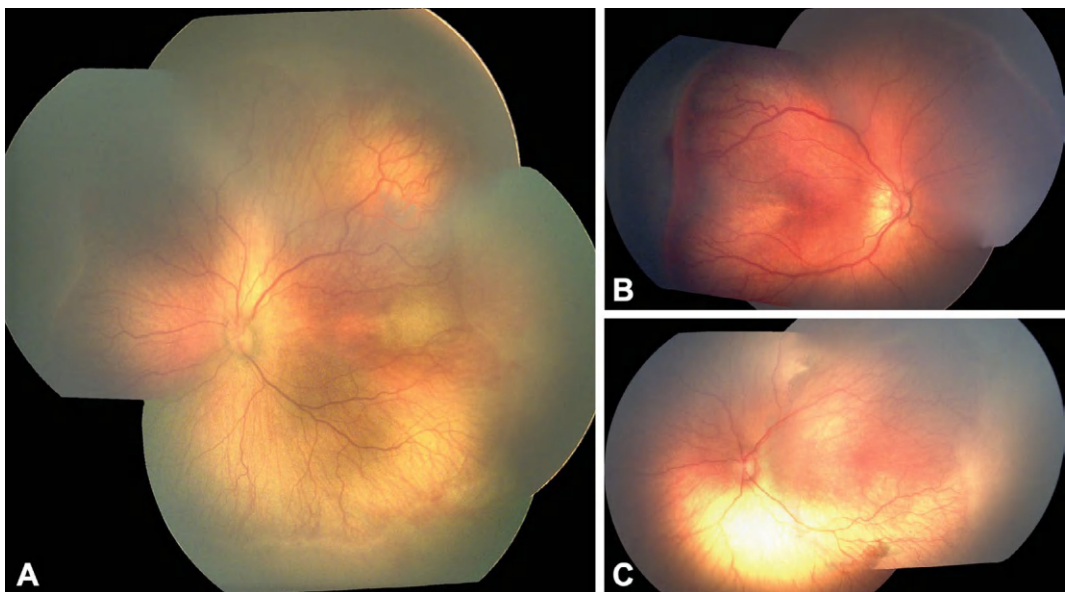


Fig. 10.3 Effect of field of view and peripheral vessel appearance on diagnostic interpretation. **a–c.** Three montage images documenting vascular changes that were more evident in the periphery than in the posterior pole.

Despite expert consensus that all three eyes had pre-plus, not plus, disease, the peripheral vessel tortuosity and dilation correlated with significant peripheral pathology that required ROP treatment with laser in all cases

be a way of standardizing disease severity and will be discussed later in this chapter.

Early AI Systems for ROP Diagnosis

The first computer-based systems for ROP diagnosis utilized manual tracings of dilation and tortuosity to produce an objective metric of severity [32]. Such semi-automated ROP diagnostic systems include ROPTool™ [33], Retinal Image multiScale Analysis (RISA) [34], Computer Assisted *Image Analysis* of the Retina (CAIAR) [35], among others; these systems were reviewed by Wittenberg et al. in 2012 [32]. As feature-extraction-based systems, they all utilized manual or semi-automated systems to quantify dilation and/or tortuosity for correlation with clinical diagnosis of ROP. In contrast to newer machine learning (ML) and DL systems, there was no automated image analysis performed by the computer; instead, feature combinations and diagnostic cut-points were determined manually with clinicians labeling or selecting findings within the images. Comparisons of expert performance to the RISA system demonstrated high diagnostic accuracy for plus disease using the computer-based analysis [34, 36, 37]. However, these systems cannot process large numbers of images and do not correlate well enough with ROP diagnosis to be widely utilized [38].

Automated Detection of Plus Disease

Machine learning utilizes a classifier, such as a support vector machine (SVM), that learns the best relationship between image features and the diagnosis [39]. One approach to have more explainable AI is to combine DL and ML methods with traditional feature extraction, and several groups have attempted this for plus disease [40, 41]. Mao et al. trained a DL network to segment retinal vessels and the optic disc and to diagnosis plus disease based on automated quantitative characterization of pathological features, such as vessel tortuosity, width, fractal dimension, and density [40].

In 2015, a ML model with a trained SVM was developed to determine the combination of features and FOV that best correlated with expert plus disease diagnosis [39]. This automated system diagnosed plus disease as well as experts when incorporating vascular tortuosity from both arteries and veins with the widest FOV [39]. The accuracy was significantly lower using a FOV comparable to that of the standard ICROP photograph; this suggested that experts consider the vascular information from a large area of retina when diagnosing plus disease. The montage images in Fig. 10.3 show examples of peripheral vascular pathology that may influence diagnostic interpretation. Despite expert-level performance, this system was limited in clinical utility as it required manual tracing and segmentation of the vessels as an input [39].

Convolutional neural networks (CNN) incorporate image classification algorithms that differ from traditional feature extraction and ML systems. Using a large database of input images, the CNN uses learnable weights and biases and gives importance to image features (e.g., tortuosity of arterioles, dilation of venules) that best correlate the input image with the diagnosis. The CNN learns these features with or without pre-processing but without explicit human input [40, 42, 43]. The CNN's fully connected 'output' layer classifies the image (e.g., absence or presence of plus disease) with improved performance than feature-extraction-based ML approaches. Worrall et al. reported the first fully automated plus disease diagnosis using a CNN; this study used a real-world dataset that included input image discrepancies across experts [43]. This system's image recognition classifier performed as well as some of the human experts (92% accuracy) [43].

Brown et al. reported the results of a fully automated DL-based system for automated three-level diagnosis of plus disease [42]. This deep CNN, called the i-ROP DL system, was trained and validated on more than 5000 images with a single reference standard diagnosis (RSD) based on the consensus diagnosis of three independent image graders and the clinical diagnosis. The area under the receiver operating

characteristic curve (AUC) for plus disease diagnosis was excellent (0.98). On an independent dataset of 100 images (i.e., not included in the training set), the i-ROP DL system had higher diagnostic agreement with the RSD than seven out of eight of the experts. For diagnosis of plus disease, the sensitivity and specificity of the algorithm were 93% and 94%, respectively. These values increased to 100% and 94%, respectively, when including pre-plus disease or worse [42].

Wagner et al. reported similarly high performance of a DL plus disease classifier developed on over 6000 retinal images from neonates in an ethnically diverse population in London, UK [44]. The model was externally validated on four separate datasets from the USA, Brazil, and Egypt, demonstrating AUCs ranging from 0.975–1.000. Using these datasets, they additionally developed a model using code-free deep learning, a service which may be of particular benefit in resource poor healthcare settings lacking trained data scientists. The code-free model also generalized well to the external validation sets (AUC range: 0.969–0.995), though was inferior to the bespoke model when tested on the USA dataset.

Continuous Scoring for Plus Disease

Vascular disease in ROP presents on a continuum, which likely explains the poor absolute agreement on plus disease classification between experts [27]. This finding motivated the development of a quantitative severity scale using the i-ROP DL system. Redd et al. reported that a scale from 1–9 could accurately detect type 1 ROP with an AUC of 0.95 [45]. Taylor et al. implemented the i-ROP DL algorithm to assign a continuous ROP vascular severity score (VSS) ranging from 1–9 and to classify images based on severity: no ROP, mild ROP, type 2 ROP and pre-plus disease, or type 1 ROP [46]. The continuous ROP VSS was associated with the ICROP category of disease at a single point in time and the clinical progression of ROP over

time [46]. VSS has additionally been found to be highly correlated with both the plus disease classifications and the ophthalmoscopic stage diagnoses ($P < 0.001$) of ICROP committee members, which otherwise demonstrate high intergrader variability [47].

Using the i-ROP dataset, Gupta et al. showed that these continuous scores reflected post-treatment regression in eyes with treatment requiring-ROP [48]. Additionally, eyes requiring multiple treatment sessions (laser or intravitreal injection of bevacizumab) had higher pre-treatment ROP VSS compared with eyes requiring only a single treatment, suggesting that treatment failure may be related to more aggressive disease or disease treated at a later stage [48]. Using fundus images from infants in the Comparing Alternative Ranibizumab Dosages for Safety and Efficacy in ROP (CARE-ROP) trial, Eilts et al. reported VSS assigned by the i-ROP DL system to be significantly decreased up to 4 weeks after injection with ranibizumab, consistent with disease regression. VSS was found to increase during reactivation in cases requiring retreatment. Higher baseline VSS was associated with earlier retreatment and a greater change in VSS in the first week after initial treatment was associated with higher reactivation risk [49].

A study by Yildiz et al. and the iROP Consortium described iROP ASSIST, a fully automated system with CNN-like performance to diagnosis plus vs. not plus disease (0.94 AUC) [50]. Inspired by the algorithms of Ataer-Cansizoglu et al. [39, 51], this system uses handcrafted features with a combined neural network for automatic vessel segmentation, tracing, feature extraction, and classification; it is publicly available for generation of a vessel severity score (0–100) from an input image [50].

Improvement in the feature extraction process will allow clinicians to achieve better performance levels without sacrificing explainability [50]. Ultimately, using similar automated quantitative severity scale for ROP diagnosis may help optimize treatment regimens by better predicting the preterm infants at risk for

treatment failure and disease recurrence [48]. Future clinical trials may use a quantitative scale to help evaluate treatment thresholds.

Automated Classification of ROP Stage and Zone

Most studies have focused on computer-based systems to diagnose plus disease; however, several studies report using DL to grade ROP severity or to classify zone or stage [52, 53]. For example, DeepROP is a different automated ROP detection system that was developed using deep neural networks (DNNs) [54]. An identification DNN model (Id-Net) and a grading DNN model (Gr-Net) directly learned ROP features from big datasets, which were comprised of retinal photographs labeled by ROP experts. Both the identification and the grading DNNs performed better than some of the human experts; impressively, the Id-Net achieved a sensitivity of 96.62% (95%CI, 92.29–98.89%) and a specificity of 99.32% (95%CI, 96.29–9.98%) for ROP identification [54]. Similarly, Hu et al. developed a deep CNN with a novel architecture to determine the presence and severity of ROP disease; a sub-network designed to extract high-level features from images was connected to a second sub-network that predicted ROP severity (mild vs. severe) [55]. Using a feature aggregate operator, this system was found to have a high classification accuracy [55].

Zhao et al. reported the development of a DL system that can automatically draw the border of zone 1 on a fundus image as a diagnostic aid [53]. Mulay et al. first reported the identification of a peripheral ROP ridge directly in a fundus image [52]. A CNN was trained by Coyner et al. in 2018 to automatically assess the quality of retinal fundus images [56, 57]; this would serve well as a prescreening method for telemedicine and computer-based image analysis in ROP. Thus, DL seems to hold promise for automated and objective diagnosis of ROP in digital fundus images. However, none of these systems are yet available for clinical use, and further research is needed.

Optical coherence tomography (OCT) has shown promise as an alternative, more objective modality to fundus photography, which may identify new biomarkers for automated quantification of ROP progression and severity. Nguyen et al. reported an association between OCT-derived retinal thickness measurements at the vascular-avascular junction and ROP stage, implicating ridge thickness as a potential quantitative biomarker for diagnosis and monitoring of ROP [58]. De Oliveira et al. found macular thickness to be associated with persistent avascular retina in patients with previously untreated ROP [59]. A review by Scruggs et al. offers recommendations for future AI research applied to ROP [60], including using OCT and OCT-angiography (OCT-A) to identify the structural signs (e.g., vitreoretinal traction) preceding disease progression [61, 62].

AI for ROP Training

If ROP experts often do not agree on how to diagnose ROP or on the diagnosis of individual babies, it is not surprising that ROP trainees find the task of ROP diagnosis challenging as well. It is well established that ophthalmology graduates complete residency, as well as ophthalmology fellowship programs, without confidence in their ability to diagnose ROP [63–65]. Fewer than one third of learners perform ROP screenings under direct supervision during ophthalmology training [66]. Chan et al. demonstrated that there was significant variability in diagnostic accuracy among retinal fellows when analyzing ROP images compared to RSDs [67]. Both Chan et al. and Myung et al. demonstrated the inconsistent accuracy of detecting type 2 ROP and treatment-requiring ROP by fellows [67, 68].

These studies raise serious concerns for ROP screening performed by inexperienced examiners, and there are no accepted criteria for minimum necessary supervision, exams, treatments, etc. for clinical competency for ROP cares. Improved global education for ROP training is necessary to ensure treatments are performed adequately. The development of AI systems for

automated diagnosis in ROP may facilitate the incorporation of these algorithms within medical training to standardize ROP education and certification [65].

The Intersection of AI and Telemedicine

Telemedicine allows a single expert to screen babies over a large region where screening ophthalmologists may be limited. Such programs also provide objective data (photos) for clinicians to detect change, a task that is difficult when humans rely solely on drawings or chart records. There are multiple examples of telemedicine programs worldwide that are successfully providing efficient screening methods for at risk babies [69–73]. In the 2014 Stanford University Network for Diagnosis of Retinopathy of Prematurity (SUNDRROP) trial, a wide-field lens was employed using the RetCam digital imaging system to obtain five fundus views per eye for 1755 infants over 6995 examinations [74].

The e-ROP Cooperative Group has published numerous studies on telemedicine approaches to evaluating acute-phase ROP [75, 76]. In one e-ROP study, the diagnostic accuracy of reader grading was determined across 5350 image set pairs, and the grading sensitivity and specificity for detecting referral warranted ROP was 90.0% and 87.0%, respectively [75]. In a different telemedicine program, plus disease was detected in 95% of eyes [77]. Large databases of digital fundus images from these ROP telemedicine programs serve as the first step in development of AI for automated image-based ROP diagnosis [78]. Nevertheless, unique telemedicine challenges exist, including varying image quality, fundus pigmentation, prevalence of disease, and ROP phenotype between geographic regions.

The incorporation of AI technology into telemedicine screening programs may enable earlier, more objective detection of ROP and

its progression for healthcare systems without direct access to pediatric ophthalmology care. Greenwald et al. reported the first study evaluating the use of a DL algorithm for ROP screening using the i-ROP algorithm, demonstrating high diagnostic performance for identifying referral-requiring ROP [79]. Coyner et al. reported a risk model for treatment-requiring ROP developed on retinal fundus images collected at the time of first examination of infants in an Indian ROP telemedicine screening program. The model was externally validated on image sets from India, Nepal, and Mongolia, and it demonstrated 100% sensitivity across all three external validation test sets. Aided by telemedicine, use of this risk model could potentially reduce the number of examinations for infants determined to be low-risk without missing cases, and these resources could be redirected to infants deemed high-risk who need closer monitoring [80]. DeCampos-Stairiker et al. applied the i-ROP-derived VSS to an image dataset collected via a tele-ROP screening program in India, drawing epidemiologic trends in disease severity. They observed a significant drop in the proportion of babies in South India developing moderate to severe ROP over a 5-year period, indicating that telemedicine may incorporate AI to remotely explore epidemiologic trends in disease severity [81].

Challenges to AI Implementation

Ting et al. published on the clinical and technical challenges of DL applications in ophthalmology [82]. While AI holds great promise for improving care for ROP, the gap between scientific discovery and clinically useful implementation of technology remains wide. Valikodath et al. surveyed pediatric ophthalmologists regarding their perspectives of AI in ophthalmology, reporting generally positive impressions, though 46% expressed concerns regarding diagnostic accuracy [83]. In addition to performance, other potential challenges hindering the

Table 10.1 Main challenges of AI implementation for ROP diagnosis in clinical practice

	Main challenges	Potential Solutions
Generalizability	• CNNs often do not generalize well to unseen data • Qualitatively different populations and phenotypes being studied, such as in LMIC • Differences in the ways the images were acquired • Technical differences between camera systems • Resolution and quality of input images or labels	• Validation of AI system performance on the target population prior to clinical use using images of varying quality and fields of view • Datasets tested in different populations • Open-access datasets and software • Automated DL-enhanced algorithms integrated into commonly used cameras (e.g., RetCam) or into cloud-based systems
Explainability	• Inability to explain how the algorithm arrived at a conclusion • “Black box” nature of clinical diagnosis, in general [86] • Difficult to develop methodology for understanding the high-level features that CNNs use for discrimination	• Combination of deep learning methods with traditional feature extraction [40, 41, 50] • Correlation of disease specific features with the CNN diagnostic outcome [41] • Rigorous clinical validation demonstrating improvement in outcomes despite lack of complete transparency • Use of activation maps to highlight feature areas on that image that contributed to classification
Regulatory and medico-legal issues	• ROP care is the highest medicolegal risk within ophthalmology • Need to adjudicate liability from care decisions informed by AI [87] • Regulatory requirements will continue to evolve	• Precise indication for use and evidence of effectiveness in a real world population • Innovation of evaluation methods by the Food and Drug Agency (FDA) to ensure safe implementation

deployment of DL systems include generalizability, explainability, and overcoming regulatory and medicolegal issues [82]. While high model performance is prerequisite for inclusion in the clinical toolkit, generalizability must be ensured in development to include ethnic minorities and socioeconomically marginalized individuals who are currently under-represented in clinical AI. Model explainability may be instrumental in ensuring equitable performance across a diverse patient population and gaining trust on both ends of the doctor-patient exchange. To achieve this, protocols for secure exchange of sensitive data and democratization of AI must be considered. To this end, techniques like federated learning and code-free AI are being explored within the context of ROP [44, 84, 85]. Table 10.1 outlines these challenges as they apply to AI for ROP diagnosis.

Conclusions

ROP is a leading cause of preventable childhood blindness worldwide, yet the diagnosis remains both subjective and qualitative. Significant intra- and inter-expert variability limits the efficiency and accuracy of ROP screening and diagnosis. AI-assisted screening may lead to automated, quantifiable, and objective diagnosis in ROP to improve accuracy while lessening the screening burden in LMIC. Providing objectivity to ROP education, AI may improve trainee performance on ROP management. Already, AI has enabled the development of an ROP VSS that correlates with ICROP disease classification and shows promise for quantitative disease monitoring, improved risk prediction, and post-treatment identification of treatment failure and recurrence. Integrated into a telemedicine system,

AI could significantly benefit ROP clinical care and may also improve early identification of severe ROP prior to the development of retinal detachment.

References

1. Flynn JT, Bancalari E, Bachynski BN, et al. Retinopathy of prematurity. Diagnosis, severity, and natural history. *Ophthalmology*. 1987;94(6):620–9.
2. Fierson WM, Ophthalmology AAOPSo, Ophthalmology AAO, Strabismus AAFPOA, Orthoptists AAOC. Screening examination of premature infants for retinopathy of prematurity. *Pediatrics*. 2018;142(6)<https://doi.org/10.1542/peds.2018-3061>
3. Valikodath N, Cole E, Chiang MF, Campbell JP, Chan RVP. Imaging in retinopathy of prematurity. *Asia Pac J Ophthalmol (Phila)*. 2019;8(2):178–86. <https://doi.org/10.22608/APO.201963>
4. Good WV, Group ETfRoPC. Final results of the early treatment for retinopathy of prematurity (ETROP) randomized trial. *Trans Am Ophthalmol Soc*. 2004;102:233–48; discussion 248–50.
5. Schaffer DB, Palmer EA, Plotsky DF, et al. Prognostic factors in the natural course of retinopathy of prematurity. The cryotherapy for retinopathy of prematurity cooperative group. *Ophthalmology* 1993;100(2):230–7. [https://doi.org/10.1016/s0161-6420\(93\)31665-9](https://doi.org/10.1016/s0161-6420(93)31665-9)
6. Chan-Ling T, Gole GA, Quinn GE, Adamson SJ, Darlow BA. Pathophysiology, screening and treatment of ROP: a multi-disciplinary perspective. *Prog Retin Eye Res*. 2018;01(62):77–119. <https://doi.org/10.1016/j.preteyeres.2017.09.002>.
7. Norman M, Hellström A, Hallberg B, et al. Prevalence of severe visual disability among preterm children with retinopathy of prematurity and association with adherence to best practice guidelines. *JAMA Netw Open*. 2019;2(1): e186801. <https://doi.org/10.1001/jamanetworkopen.2018.6801>.
8. Blencowe H, Lawn JE, Vazquez T, Fielder A, Gilbert C. Preterm-associated visual impairment and estimates of retinopathy of prematurity at regional and global levels for 2010. *Pediatr Res*. 2013;74(Suppl 1):35–49. <https://doi.org/10.1038/pr.2013.205>.
9. Vos T, Flaxman AD, Naghavi M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the global burden of disease study 2010. *Lancet*. 2012;380(9859):2163–96. [https://doi.org/10.1016/S0140-6736\(12\)61729-2](https://doi.org/10.1016/S0140-6736(12)61729-2).
10. Smith LE, Hard AL, Hellström A. The biology of retinopathy of prematurity: how knowledge of pathogenesis guides treatment. *Clin Perinatol*. 2013;40(2):201–14. <https://doi.org/10.1016/j.clp.2013.02.002>.
11. Patan S. Vasculogenesis and angiogenesis. *Cancer Treat Res*. 2004;117:3–32. https://doi.org/10.1007/978-1-4419-8871-3_1.
12. Multicenter trial of cryotherapy for retinopathy of prematurity. One-year outcome—structure and function. Cryotherapy for Retinopathy of Prematurity Cooperative Group. *Arch Ophthalmol*. 1990;108(10):1408–16.
13. Fierson WM, Ophthalmology AAOPSo, Ophthalmology AAo, Strabismus AAFPOa, Orthoptists AAoC. Screening examination of premature infants for retinopathy of prematurity. *Pediatrics*. 2013;131(1):189–95. <https://doi.org/10.1542/peds.2012-2996>
14. Gilbert C. Retinopathy of prematurity: a global perspective of the epidemics, population of babies at risk and implications for control. *Early Hum Dev*. 2008;84(2):77–82. <https://doi.org/10.1016/j.earlhumdev.2007.11.009>.
15. Fleck BW, Williams C, Juszczak E, et al. An international comparison of retinopathy of prematurity grading performance within the benefits of oxygen saturation targeting II trials. *Eye (Lond)*. 2018;32(1):74–80. <https://doi.org/10.1038/eye.2017.150>
16. Reynolds JD, Dobson V, Quinn GE, et al. Evidence-based screening criteria for retinopathy of prematurity: natural history data from the CRYO-ROP and LIGHT-ROP studies. *Arch Ophthalmol*. 2002;120(11):1470–6. <https://doi.org/10.1001/archophth.120.11.1470>.
17. Prematurity ICfCoRo. The international classification of retinopathy of prematurity revisited. *Arch Ophthalmol*. 2005;123(7):991–9. <https://doi.org/10.1001/archophth.123.7.991>
18. An international classification of retinopathy of prematurity. II. The classification of retinal detachment. The international committee for the classification of the late stages of retinopathy of prematurity. *Arch Ophthalmol*. 1987;105(7):906–12.
19. An international classification of retinopathy of prematurity. The committee for the classification of retinopathy of prematurity. *Arch Ophthalmol*. 1984;102(8):1130–4. <https://doi.org/10.1001/archophth.1984.01040030908011>
20. Chiang MF, Quinn GE, Fielder AR, et al. International classification of retinopathy of prematurity, 3rd edn. *Ophthalmology* 2021;128(10):e51–68. <https://doi.org/10.1016/j.ophtha.2021.05.031>
21. Campbell JP, Kim SJ, Brown JM, et al. Evaluation of a deep learning-derived quantitative retinopathy of prematurity severity scale. *Ophthalmology*. 2021;128(7):1070–6. <https://doi.org/10.1016/j.ophtha.2020.10.025>.
22. Ghergherehchi L, Kim SJ, Campbell JP, Ostmo S, Chan RVP, Chiang MF. Plus disease in retinopathy of prematurity: more than meets the ICROP? *Asia Pac J Ophthalmol (Phila)*. 2018;7(3):152–155. <https://doi.org/10.22608/APO.201863>

23. Nguyen TP, Young BK, Coyner A, et al. Discrepancies in diagnosis of treatment-requiring retinopathy of prematurity. *Ophthalmol Retina*. 2023. <https://doi.org/10.1016/j.oret.2023.09.001>
24. Geloneck MM, Chuang AZ, Clark WL, et al. Refractive outcomes following bevacizumab monotherapy compared with conventional laser treatment: a randomized clinical trial. *JAMA Ophthalmol*. 2014;132(11):1327–33. <https://doi.org/10.1001/jamaophthalmol.2014.2772>.
25. Mintz-Hittner HA, Kennedy KA, Chuang AZ, Group B-RC. Efficacy of intravitreal bevacizumab for stage 3+ retinopathy of prematurity. *N Engl J Med*. 2011;364(7):603–15. <https://doi.org/10.1056/NEJMoa1007374>
26. Stahl A, Lepore D, Fielder A, et al. Ranibizumab versus laser therapy for the treatment of very low birthweight infants with retinopathy of prematurity (RAINBOW): an open-label randomised controlled trial. *Lancet*. 2019;394(10208):1551–59. [https://doi.org/10.1016/S0140-6736\(19\)31344-3](https://doi.org/10.1016/S0140-6736(19)31344-3)
27. Kalpathy-Cramer J, Campbell JP, Erdogmus D, et al. Plus disease in retinopathy of prematurity: improving diagnosis by ranking disease severity and using quantitative image analysis. *Ophthalmology*. 2016;123(11):2345–51. <https://doi.org/10.1016/j.opthta.2016.07.020>
28. Campbell JP, Ataer-Cansizoglu E, Bolon-Canedo V, et al. Expert diagnosis of plus disease in retinopathy of prematurity from computer-based image analysis. *JAMA Ophthalmol*. 2016;134(6):651–7. <https://doi.org/10.1001/jamaophthalmol.2016.0611>.
29. Cole ED, Park SH, Kim SJ, et al. Variability in plus disease diagnosis using single and serial images. *Ophthalmol Retina*. 2022;6(12):1122–9. <https://doi.org/10.1016/j.oret.2022.05.024>.
30. Rao R, Jonsson NJ, Ventura C, et al. Plus disease in retinopathy of prematurity: diagnostic impact of field of view. *Retina*. 2012;32(6):1148–55. <https://doi.org/10.1097/IAE.0b013e31823ac3c3>.
31. Kim SJ, Campbell JP, Kalpathy-Cramer J, et al. Accuracy and reliability of eye-based vs quadrant-based diagnosis of plus disease in retinopathy of prematurity. *JAMA Ophthalmol*. 2018;136(6):648–55. <https://doi.org/10.1001/jamaophthalmol.2018.1195>.
32. Wittenberg LA, Jonsson NJ, Chan RV, Chiang MF. Computer-based image analysis for plus disease diagnosis in retinopathy of prematurity. *J Pediatr Ophthalmol Strabismus*. 2012;49(1):11–9; quiz 10, 20. <https://doi.org/10.3928/01913913-20110222-01>
33. Wallace DK, Zhao Z, Freedman SF. A pilot study using “ROptool” to quantify plus disease in retinopathy of prematurity. *J AAPOS*. 2007;11(4):381–7. <https://doi.org/10.1016/j.jaapos.2007.04.008>.
34. Gelman R, Jiang L, Du YE, Martinez-Perez ME, Flynn JT, Chiang MF. Plus disease in retinopathy of prematurity: pilot study of computer-based and expert diagnosis. *J AAPOS*. 2007;11(6):532–40. <https://doi.org/10.1016/j.jaapos.2007.09.005>.
35. Shah DN, Wilson CM, Ying GS, et al. Comparison of expert graders to computer-assisted image analysis of the retina in retinopathy of prematurity. *Br J Ophthalmol*. 2011;95(10):1442–5. <https://doi.org/10.1136/bjo.2010.185363>.
36. Chiang MF, Gelman R, Jiang L, Martinez-Perez ME, Du YE, Flynn JT. Plus disease in retinopathy of prematurity: an analysis of diagnostic performance. *Trans Am Ophthalmol Soc*. 2007;105:73–84; discussion 84–5.
37. Koreen S, Gelman R, Martinez-Perez ME, et al. Evaluation of a computer-based system for plus disease diagnosis in retinopathy of prematurity. *Ophthalmology*. 2007;114(12):e59–67. <https://doi.org/10.1016/j.opthta.2007.10.006>.
38. Wilson CM, Wong K, Ng J, Cocker KD, Ells AL, Fielder AR. Digital image analysis in retinopathy of prematurity: a comparison of vessel selection methods. *J AAPOS*. 2012;16(3):223–8. <https://doi.org/10.1016/j.jaapos.2011.11.015>.
39. Ataer-Cansizoglu E, Bolon-Canedo V, Campbell JP, et al. Computer-based image analysis for plus disease diagnosis in retinopathy of prematurity: performance of the “i-ROP” system and image features associated with expert diagnosis. *Transl Vis Sci Technol*. 2015;4(6):5. <https://doi.org/10.1167/tvst.4.6.5>.
40. Mao J, Luo Y, Liu L, et al. Automated diagnosis and quantitative analysis of plus disease in retinopathy of prematurity based on deep convolutional neural networks. *Acta Ophthalmol*. 2019. <https://doi.org/10.1111/aos.14264>.
41. Graziani M, Brown JM, Andrearczyk V, et al. Improved interpretability for computer-aided severity assessment of retinopathy of prematurity. Presented at: SPIE Medical Imaging; 2019; San Diego, California. Session Computer-Aided Diagnosis.
42. Brown JM, Campbell JP, Beers A, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. *JAMA Ophthalmol*. 2018;136(7):803–10. <https://doi.org/10.1001/jamaophthalmol.2018.1934>.
43. Worrall DE, Wilson CM, Brostow GJ. Automated retinopathy of prematurity case detection with convolutional neural networks. Presented at: Deep learning and data labeling for medical applications; 2016; Athens, Greece. Session Lecture Notes in Computer Science.
44. Wagner SK, Liefers B, Radia M, et al. Development and international validation of custom-engineered and code-free deep-learning models for detection of plus disease in retinopathy of prematurity: a retrospective study. *Lancet Digit Health*. 2023;5(6):e340–9. [https://doi.org/10.1016/s2589-7500\(23\)00050-x](https://doi.org/10.1016/s2589-7500(23)00050-x).
45. Redd TK, Campbell JP, Brown JM, et al. Evaluation of a deep learning image assessment system for detecting severe retinopathy of prematurity. *Br J Ophthalmol*. 2018. <https://doi.org/10.1136/bjophthalmol-2018-313156>.

46. Taylor S, Brown JM, Gupta K, et al. Monitoring disease progression with a quantitative severity scale for retinopathy of prematurity using deep learning. *JAMA Ophthalmol*. 2019. <https://doi.org/10.1001/jamaophthalmol.2019.2433>.
47. Campbell JP, Chiang MF, Chen JS, et al. Artificial intelligence for retinopathy of prematurity: validation of a vascular severity scale against international expert diagnosis. *Ophthalmology*. 2022;129(7):e69–76. <https://doi.org/10.1016/j.ophtha.2022.02.008>.
48. Gupta K, Campbell JP, Taylor S, et al. A quantitative severity scale for retinopathy of prematurity using deep learning to monitor disease regression after treatment. *JAMA Ophthalmol*. 2019. <https://doi.org/10.1001/jamaophthalmol.2019.2442>.
49. Eilts SK, Pfeil JM, Poschkamp B, et al. Assessment of retinopathy of prematurity regression and reactivation using an artificial intelligence-based vascular severity score. *JAMA Netw Open* 2023;6(1):e2251512. <https://doi.org/10.1001/jamanetworkopen.2022.51512>
50. Yildiz VM, Tian P, Yildiz I, et al. Plus disease in retinopathy of prematurity: convolutional neural network performance using a combined neural network and feature extraction approach. 2020;9(2)
51. Ataer-Cansizoglu E, You S, Kalpathy-Cramer J, Keck K, Chiang MF, Erdogmus D. Observer and feature analysis on diagnosis of retinopathy of prematurity. In: IEEE Int workshop mach learn signal process; 2012, pp. 1–6. <https://doi.org/10.1109/MLSP.2012.6349809>
52. Mulay S, Ram K, Sivaprakasam M, Vinekar A. Early detection of retinopathy of prematurity stage using deep learning approach. 2019.
53. Zhao J, Lei B, Wu Z, et al. A deep learning framework for identifying Zone I in RetCam Images. *IEEE Access* 2019;10:3530–37.
54. Wang J, Ju R, Chen Y, et al. Automated retinopathy of prematurity screening using deep neural networks. *EBioMedicine*. 2018;35:361–8. <https://doi.org/10.1016/j.ebiom.2018.08.033>.
55. Hu J, Chen Y, Zhong J, Ju R, Yi Z. Automated analysis for retinopathy of prematurity by deep neural networks. *IEEE Trans Med Imaging*. 2019;38(1):269–79. <https://doi.org/10.1109/TMI.2018.2863562>
56. Coyner AS, Swan R, Brown JM, et al. Deep learning for image quality assessment of fundus images in retinopathy of prematurity. *AMIA Annu Symp Proc*. 2018;2018:1224–32.
57. Coyner AS, Swan R, Campbell JP, et al. Automated fundus image quality assessment in retinopathy of prematurity using deep convolutional neural networks. *Ophthalmol Retina*. 2019;3(5):444–50. <https://doi.org/10.1016/j.oret.2019.01.015>.
58. Nguyen TP, Ni S, Ostmo S, et al. Association of optical coherence tomography-measured fibrovascular ridge thickness and clinical disease stage in retinopathy of prematurity. *JAMA Ophthalmol*. 2022;140(11):1121–7. <https://doi.org/10.1001/jamaophthalmol.2022.4173>.
59. de Oliveira TB, Young BK, Ostmo SR, et al. Macular Thickness Is Associated with Persistent Avascular Retina in Patients with a History of Untreated Retinopathy of Prematurity. *Ophthalmol Retina*. Oct 12 2023;<https://doi.org/10.1016/j.oret.2023.10.004>
60. Scruggs BA, Chan RVP, Kalpathy-Cramer J, Chiang MF, Campbell JP. Artificial intelligence in retinopathy of prematurity diagnosis. *Transl Vis Sci Technol* 2020;9(2)
61. Campbell JP. Why do we still rely on ophthalmoscopy to diagnose retinopathy of prematurity? *JAMA Ophthalmol*. 2018;136(7):759–60. <https://doi.org/10.1001/jamaophthalmol.2018.1539>.
62. De Fauw J, Ledsam JR, Romera-Paredes B, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med*. 2018;24(9):1342–50. <https://doi.org/10.1038/s41591-018-0107-6>
63. Patel SN, Martinez-Castellanos MA, Berrones-Medina D, et al. Assessment of a tele-education system to enhance retinopathy of prematurity training by international ophthalmologists-in-training in Mexico. *Ophthalmology* 2017;124(7):953–61. <https://doi.org/10.1016/j.ophtha.2017.02.014>
64. Campbell JP, Swan R, Jonas K, et al. Implementation and evaluation of a tele-education system for the diagnosis of ophthalmic disease by international trainees. *AMIA Annu Symp Proc*. 2015; 2015(2015):366–75.
65. Chan RV, Patel SN, Ryan MC, et al. The global education network for retinopathy of prematurity (Gen-Rop): development, implementation, and evaluation of a novel tele-education system (An American Ophthalmological Society Thesis). *Trans Am Ophthalmol Soc*. 2015;113:T2.
66. Al-Khaled T, Mikhail M, Jonas KE, et al. Training of residents and fellows in retinopathy of prematurity around the world: an international web-based survey. *J Pediatr Ophthalmol Strabismus*. 2019;56(5):282–7. <https://doi.org/10.3928/01913913-20190717-01>.
67. Paul Chan RV, Williams SL, Yonekawa Y, Weissgold DJ, Lee TC, Chiang MF. Accuracy of retinopathy of prematurity diagnosis by retinal fellows. *Retina*. 2010;30(6):958–65. <https://doi.org/10.1097/IAE.0b013e3181c9696a>.
68. Myung JS, Paul Chan RV, Espiritu MJ, et al. Accuracy of retinopathy of prematurity image-based diagnosis by pediatric ophthalmology fellows: implications for training. *J AAPOS*. 2011;15(6):573–8. <https://doi.org/10.1016/j.jaapos.2011.06.011>.
69. Fierson WM, Capone A. *Ophthalmology AAoPSo, American Academy of Ophthalmology AeAoCo*. Telemedicine for evaluation of retinopathy of prematurity. *Pediatrics*. 2015;135(1):e238–54. <https://doi.org/10.1542/peds.2014-0978>
70. Quinn GE, Ying GS, Daniel E, et al. Validity of a telemedicine system for the evaluation of acute-phase retinopathy of prematurity. *JAMA Ophthalmol*. 2014;132(10):1178–84. <https://doi.org/10.1001/jamaophthalmol.2014.1604>.

71. Weaver DT, Murdock TJ. Telemedicine detection of type 1 ROP in a distant neonatal intensive care unit. *J AAPOS*. 2012;16(3):229–33. <https://doi.org/10.1016/j.jaapos.2012.01.007>.
72. Chiang MF, Melia M, Buffenn AN, et al. Detection of clinically significant retinopathy of prematurity using wide-angle digital retinal photography: a report by the American Academy of Ophthalmology. *Ophthalmology*. 2012;119(6):1272–80. <https://doi.org/10.1016/j.ophtha.2012.01.002>.
73. Ells AL, Holmes JM, Astle WF, et al. Telemedicine approach to screening for severe retinopathy of prematurity: a pilot study. *Ophthalmology*. 2003;110(11):2113–7. [https://doi.org/10.1016/S0161-6420\(03\)00831-5](https://doi.org/10.1016/S0161-6420(03)00831-5).
74. Fijalkowski N, Zheng LL, Henderson MT, et al. Stanford university network for diagnosis of retinopathy of prematurity (SUNDROP): five years of screening with telemedicine. *Ophthalmic Surg Lasers Imaging Retina*. 2014;45(2):106–13. <https://doi.org/10.3928/23258160-20140122-01>
75. Quinn GE, Ells A, Capone A, et al. Analysis of discrepancy between diagnostic clinical examination findings and corresponding evaluation of digital images in the telemedicine approaches to evaluating acute-phase retinopathy of prematurity study. *JAMA Ophthalmol*. 2016;134(11):1263–70. <https://doi.org/10.1001/jamaophthalmol.2016.3502>.
76. Ying GS, Pan W, Quinn GE, Daniel E, Repka MX, Baumritter A. Intereye agreement of retinopathy of prematurity from image evaluation in the telemedicine approaches to evaluating of acute-phase ROP (e-ROP) study. *Ophthalmol Retina*. 2017;1(4):347–54. <https://doi.org/10.1016/j.oret.2016.12.005>
77. Schwartz SD, Harrison SA, Ferrone PJ, Trese MT. Telemedical evaluation and management of retinopathy of prematurity using a fiberoptic digital fundus camera. *Ophthalmology*. 2000;107(1):25–8. [https://doi.org/10.1016/s0161-6420\(99\)00003-2](https://doi.org/10.1016/s0161-6420(99)00003-2).
78. Chee RI, Darwish D, Fernandez-Vega A, et al. Retinal telemedicine. *Curr Ophthalmol Rep*. 2018;6(1):36–45. <https://doi.org/10.1007/s40135-018-0161-8>.
79. Greenwald MF, Danford ID, Shahrawat M, et al. Evaluation of artificial intelligence-based telemedicine screening for retinopathy of prematurity. *J AAPOS*. 2020;24(3):160–2. <https://doi.org/10.1016/j.jaapos.2020.01.014>.
80. Coyner AS, Oh MA, Shah PK, et al. External validation of a retinopathy of prematurity screening model using artificial intelligence in 3 low- and middle-income populations. *JAMA Ophthalmol*. 2022;140(8):791–8. <https://doi.org/10.1001/jamaophthalmol.2022.2135>.
81. deCampos-Stairiker MA, Coyner AS, Gupta A, et al. Epidemiologic evaluation of retinopathy of prematurity severity in a large telemedicine program in India using artificial intelligence. *Ophthalmology*. 2023;130(8):837–43. <https://doi.org/10.1016/j.ophtha.2023.03.026>.
82. Ting DSW, Peng L, Varadarajan AV, et al. Deep learning in ophthalmology: the technical and clinical considerations. *Prog Retin Eye Res*. 2019. <https://doi.org/10.1016/j.preteyeres.2019.04.003>.
83. Valikodath NG, Al-Khaled T, Cole E, et al. Evaluation of pediatric ophthalmologists' perspectives of artificial intelligence in ophthalmology. *J aapos*. 2021;25(3):164.e1–164.e5. <https://doi.org/10.1016/j.jaapos.2021.01.011>.
84. Hanif A, Lu C, Chang K, et al. Federated learning for multicenter collaboration in ophthalmology: implications for clinical diagnosis and disease epidemiology. *Ophthalmol Retina*. 2022;6(8):650–6. <https://doi.org/10.1016/j.oret.2022.03.005>.
85. Lu C, Hanif A, Singh P, et al. Federated learning for multicenter collaboration in ophthalmology: improving classification performance in retinopathy of prematurity. *Ophthalmol Retina*. 2022;6(8):657–63. <https://doi.org/10.1016/j.oret.2022.02.015>.
86. Reid JE, Eaton E. Artificial intelligence for pediatric ophthalmology. *Curr Opin Ophthalmol*. 2019;30(5):337–46. <https://doi.org/10.1097/ICU.0000000000000593>.
87. Shah NH, Milstein A, Bagley SC. Making machine learning models clinically useful. *JAMA*. 2019. <https://doi.org/10.1001/jama.2019.10306>.

Artificial Intelligence in Diabetic Retinopathy

11

Andrzej Grzybowski and Piotr Brona

Abstract

Diabetic retinopathy (DR) is a leading cause of blindness among working-age adults globally, with its prevalence rising alongside the increasing incidence of diabetes. This chapter explores the epidemiology of DR, emphasizing disparities between type 1 and type 2 diabetes, regional variations, and its association with other systemic diabetic complications. It discusses advancements in screening initiatives, including telemedicine and national programs, and highlights the transformative potential of artificial intelligence (AI), particularly deep learning, in enhancing DR detection and cost-effective screening. The chapter also addresses the challenges of implementing AI-based solutions, especially in resource-limited settings.

Keywords

Artificial intelligence · Diabetes · Diabetic retinopathy · AI-based medical devices

Epidemiology of Diabetic Retinopathy

Over the last four decades, the number of people living with diabetes has more than quadrupled from 108 million in 1980 to an estimated 422 million in 2014. At the same time, the diabetes prevalence among adults has almost doubled to 8.5% [1]. Future projections estimate that, by 2035, 592 million people will have diabetes, with the largest increase in low- and middle-income regions [2]. There is no doubt that diabetes constitutes a significant problem for global health and well-being. It is a disease that is prevalent worldwide in affluent, resource-rich countries and much poorer developing countries. Diabetes can cause a number of significant complications, each of which is associated with significant morbidity, requiring different, highly qualified medical personnel to diagnose and treat it. This poses a challenge for local health services, which often struggle with either delivering or funding appropriate care.

Diabetic retinopathy is one of the major complications of diabetes and is estimated to be the leading cause of blindness among working-age adults globally [3].

A. Grzybowski (✉)

Department of Ophthalmology, University of Warmia and Mazury, Olsztyn, Poland
e-mail: ae.grzybowski@gmail.com

Institute for Research in Ophthalmology, Foundation for Ophthalmology Development, Poznan, Poland

P. Brona

Department of Ophthalmology, Poznan City Hospital, Poznan, Poland

The prevalence of DR and proliferative DR (pDR) varies between type 1 and 2 diabetes and among different regions of the world. The prevalence of DR among individuals with type 2 diabetes has been reported to be between 20 and 40% in most studies. The reported prevalence of type 1 diabetes in Europe and the USA varies widely between reports ranging from 36.5 to 93.6% [3]. Approximately one-third of those with DR may have vision-threatening DR with either proliferative DR or diabetic macular edema (DME). Overall, the incidence of DR is greater among Western communities than among Asian regions [3]. Singapore is a notable exception to this, with a much higher incidence of DR, but it is comparable to that in developed Western countries.

The incidence of DR among diabetic patients is also increasing in some regions. A study based in Spain revealed that the yearly incidence of DR increased by almost 1% over an 8-year lifespan, from 8.09% in 2007 to 8.99% in 2014, with the incidence of DME also increasing [4]. The increasing worldwide population, coupled with the increasing prevalence of diabetes and increased incidence of DR, has led to an increasing number of patients with ocular complications of diabetes.

In addition to the global burden of pDR and DME, these manifestations appear to be prognostic of other diabetes complications, such as nephropathy, peripheral neuropathy and cardiovascular events [3].

Conventional Screening Initiatives for DR—Telemedicine

There have been many DR screening initiatives throughout the world with varying degrees of coverage and longevity. Nevertheless, only a few countries, most prominently the UK and Singapore, were able to successfully establish and continue DR screening at the national level. It appears such programme is also functioning within Denmark, however very little information regarding it is available in English.

United Kingdom

Each country within the UK established its own national screening programme. The specific protocols and grading methods vary; however, all are based on digital, color fundus photography. The programmes cover all diabetic patients over the age of 12 years old with vision of at least light perception in one eye.

England

The NHS Diabetic Eye Screening Programme (NDESP) is a continuation of an English screening programme established in 2006. Patients are screened on annual visits, with two fundus images per eye—one macula-centered and one disc-centered. Images were taken after mydriasis. These images are later digitally sent to one of the centralized grading centers. The sheer scale of piloting and implementing such an initiative is impressive; in 2015–2016, the programme invited more than 2.5 million diabetic patients to participate in screening, with an increase of 82.8% [5]. This also provides important insight into the epidemiology of DR in the local population. Between 2015 and 2016, screening resulted in just under 8000 urgent referrals with suspected proliferative disease and over 52,000 referrals for suspected maculopathy or preproliferative diabetic retinopathy, with an overall rate of DR of 2.8%.

The aforementioned screening programme was expected to reduce the number of people considered legally blind in England from 4200 to less than 1000. It appears that this goal has been met with a 2014 report showing that DR is no longer the leading cause of certifiable blindness in England and Wales for the first time in 50 years [5].

Wales

The Diabetic Retinopathy Screening Service for Wales (DRSSW), established in 2002, is a mobile screening service. Similar to the English program, two fundus images are taken per eye. Patients with sight-threatening DR are referred to a hospital-based retinal service. Thirty screening

teams serve 220 locations within Wales, accounting for approximately 80% of patients.

Scotland

DR screening started in Scotland in 2006. The qualifying patients were identified automatically using the Scottish Care Information-Diabetes Collaboration database. Screening is based on a single macula-centered image per eye, with mydriasis as needed. Images are later sent to grading centers. Due to automatic patient selection, patient coverage was greater than 99%.

Northern Ireland

The Northern Ireland Diabetic Retinopathy Screening Programme (NIDRSP) was established in 2002. Its function is similar to that of the Welsh programme DRSSW. Patients are referred to the programme by their GPs, with trained readers grading the photographs.

Ireland

DR screening was first identified as a key goal in 2008 but was only introduced in 2013 under the name Diabetic RetinaScreen. The Irish programme screens diabetics 12 and older. The Diabetic RetinaScreen supervises both annual fundus image-based screening and DR treatment and consists of both stationary and mobile community-based screening centers. The grading follows the English system closely both in terms of the grading matrix and the quality assurance protocols. According to the latest report, screening uptake is approximately 67% and has risen considerably since screening was first introduced [6].

Singapore

Singapore began widespread DR screening almost three decades ago in 1991. At that time, Polaroid nonmydriatic fundus photography was chosen because images could be taken by trained staff instead of ophthalmologists. Images were reviewed by a local hospital-based ophthalmologist. At the time, it was the first and

only nationwide DR screening programme. The Singaporean screening initiative was revived in an updated version reflecting technological advancements and possibilities and is now known as the Singapore Integrated Diabetic Retinopathy Programme (SIDRP). Based on primary care clinics equipped with retinal cameras and specialized reading centers employing trained graders, the programme aims for a result within 24 h of screening.

Cost effectiveness analysis has shown that this telemedicine-based model generated \$173 cost savings per patient compared to the previous screening model where family physicians graded the images themselves after special training [7].

Local Screening Initiatives

In addition to the established national screening programmes, there have been a large number of smaller-scale local screening initiatives worldwide. Some of those are similarly long-standing projects that control their population yearly, while most were discontinued or singular screening efforts. Although many screening projects have been attempted, only a few of the aforementioned countries were able to implement nationwide screening, further highlighting the difficulty of such undertakings (Table 11.1).

Automated Screening for Diabetic Retinopathy

The idea of adopting a computer program for assessing fundus images for DR is certainly not new. The first report that we were able to find such an endeavor was published in 1996 by Gardner and colleagues. Almost 30 years ago, the authors established a neural network trained on 147 diabetic and 32 normal fundus images and aimed to train it to recognize particular features of images—aneurysms, hemorrhages and exudates. Due to the many constraints, including computational capacity, each image was divided into small squares 20 or 30 pixels wide and later assessed by an ophthalmologist as containing

Table 11.1 The list of DR screening algorithms available as for June 2024

	Algorithm name	Producer	Country of origin	FDA certification	Registered as a medical device in EU	EU device class	DME
1	IDx-DR	Digital diagnostic	USA	YES	YES	IIa	YES
2	Eyearth	EYENUK	USA	YES	YES	IIa	No data
3	Intelligent Retinal Imaging Systems (IRIS)	Intelligent Retinal Imaging Systems	USA	YES	No data	No data	No data
4	Aeye Diagnostic Screening AEYE-DS	AEYE Health	USA/Israel	YES	No data	No data	YES
5	Retino Scan (TeleEye MD/Rapid DX)	TeleMedC Pty Ltd.	Singapore/Australia	NO	YES	III	No data
6	EyeWisdom	Visionary Intelligence Ltd. or Vistel	China	NO	YES	III	No data
7	Vistel	Visionary Intelligence Ltd. or Vistel	China	NO	YES	III	No data
8	RetmarkerDR	Retmarker	Portugal	NO	YES	IIa	No data
9	SELENA+	eyRIS	Singapore	NO	YES	IIa	No data
10	Medios AI	Remidio	India	NO	YES	IIa	No data
11	RetCAD	Thirona	Netherlands	NO	YES	IIa	No data
12	OphtAI	Evolucare	France	NO	YES	IIa	YES
13	(COMEN) Aireen DR	Aireen	Czech Republic	NO	YES	IIb	No data
14	DeepDee AI	DeepDee AI	Netherlands/Belarus	NO	YES	I	YES
15	MONA DR	Mona Health	Belgium	NO	YES	I	YES
16	CARA	Diagnos	Canada	NO	YES	I	No data
17	Retinalyze	Retinalyze Nordic ApS	Danmark	NO	YES	I	No data
18	Airdoc	Airdoc Technology Inc.	China	NO	YES	I	No data
19	VUNO Med-Fundus AI	VUNO Inc.	Korea	NO	YES	I	No data
20	iPredict	iHealthScreen Inc.	USA	NO	YES	I	No data
21	DeepMind	Google algorithm	USA	NO	NO	NO	No data
22	LuxIA/RetinaAI	RetinaAI	Swiss/Spain	NO	No data	NO	No data
23	retina.help	Brains fro Hire	Germany	NO	NO	NO	No data
24	VeriSee DR	Acer	Taiwan	NO	No data	No data	No data
25	Eyetelligence	Eyetelligence Pty Ltd.	Australia	NO	No data	No data	NO
26	Theia	Toku Eyes	New Zealand	NO	No data	No data	YES
27	CheckEye	CheckEye	Ukraine	NO	NO	NO	No data
28	Recogify	Recogify	Poland	NO	No data	No data	No data

vessels, exudates, hemorrhages, microaneurysms or normal retina without vessels [8].

Another study performed in 2000 described a mixed technique in which algorithms designed to enhance round features in an image were used to select for microaneurysms in a fundus image.

This was later assessed by a neural network to determine the significance of the extracted round feature. This resulted in a sensible detection rate for images containing microaneurysms (81%) compared to that in the opinion of a clinical research fellow [9].

Several studies explored the subject in the early 2000s, without the use of neural networks, relying on various preestablished image-analysis techniques, such as automated detection of anatomical landmarks in fundus images (optic disc, blood vessels, fovea, etc.) coupled with specifically designed algorithms for the detection of DR features. Among those, the first three reports of a system, later known as Retinalyze and discussed in further sections, were published showing relatively good sensitivities of 71–93% and specificities of 72–86%; these were based on small sample sizes reaching 137 patients in the largest study [10, 11].

All of those studies were performed in the predigitization era, in which images, in the form of slides, taken from a fundus camera had to be scanned by hand. This was done using a slide reader or scanner to achieve a workable digital version of the image. The process was time consuming and required specialized equipment, and additional processing steps introduced potential image artifacts and loss of quality. The lack of centralized databases and digital storage of fundus images meant that training and verification images were difficult to acquire. As a consequence, most studies suffer from the use of fewer images than modern models that use tens of thousands of fundus images to establish and validate a system.

Although automated screening was severely limited from a technical standpoint, a number of people have attempted to devise suitable screening methods, recognizing the potential of new technology to enhance or substitute human-based grading.

Deep-Learning Algorithms

In subsequent years, with increasing digitization, new ways of approaching the subject of automated image analysis were made possible. Until the 2010s, experts designed algorithms for the detection of specific features of DR, such as microaneurysm or hemorrhage. In deep learning, the software is presented with a fundus image as a whole and a prespecified result

for that image. Over the course of analyzing many such images—often thousands—it has become possible to distinguish between images with different results. What separates one result from another does not have to be explicitly specified by its designers. The advent of deep learning-based DR detection revealed a significant improvement in the accuracy of the newly developed or improved systems. Abramoff and colleagues reported how the introduction of deep learning techniques allowed a significant improvement to the already established, classically designed, automated DR software—the Iowa Detection Program. Based on the publicly available set of fundus images with/without DR—the Messidor-2 dataset—the sensitivity improved from 94.4 to 96.8%, and the specificity improved from a confidence interval of 55.7%–63% to 87% [12]. For the Iowa Detection Program, deep learning features were added to existing algorithms, and many other initiatives have attempted to establish entirely new deep learning DR detection software. Establishing automated or semiautomated screening, with the use of AI, will require striking a careful balance between sensitivity and specificity, imaging modality, and gradeability of the images, all of which will need to be weighed against the potential cost. The cost–benefit balance is not universal and varies depending on the relationship of those parameters with the relevant population characteristics, such as the prevalence of DR and sight-threatening DR, the availability of treatment, the cost and availability of trained staff, etc. A recent paper explored potential approaches for performing health economic assessments and safety analyses involving the implementation of novel AI DR solutions for widespread screening [13]. Deep learning DR detection has been found to be cost-effective in developed countries, such as Singapore and the United Kingdom. However, there are no published studies on the feasibility of implementing AI DR screening in countries without a robust teleophthalmology screening programme beforehand or in other resource-limited settings [13].

Several significant initiatives for AI-based diabetic retinopathy detection have been described.

IDx-DR

IDx-DR is a combined DR screening solution that incorporates the aforementioned DR screening algorithm with an image quality assessment and feedback system. Submission of images is performed using the IDx-DR client, which is a stand-alone piece of software. The IDx-DR client features a system for the resubmission of images deemed to be of too low quality. The threshold for a positive result was set as ‘more than mild’ diabetic retinopathy according to the International Classification of Disease Reporting Trials (ICDR) grading scale or signs of diabetic macular edema. IDx-DR offers one additional level of vision-threatening DR, indicative of a suspicion of more advanced, possibly proliferative DR. Screening is based on 4 fundus images per patient, 2 from each eye, one macula-centered and one disc-centered, and all images need to be submitted for a result to be produced. The algorithm is able to cope with some quality loss

by utilizing the overlap of the two image fields (Fig. 11.1).

On the front-end, the user is presented with a screening result in one of the 4 categories—no DR, mtmDR, vision-threatening DR and insufficient quality—and on the back-end, IDx-DR produces a numerical value representing its likelihood of mtmDR. Currently, defined cut-offs are used to sort patients into appropriate categories. Theoretically, this means that the IDx-DR output could be adjusted to maximize either sensitivity or specificity depending on the needs of a given screening initiative.

IDx-DR is the first autonomous diagnostic software and one of the first AI-based software programs in medicine to receive Federal Drug and Administration (FDA) approval. In a self-titled pivotal trial, IDx-DR software was studied in a real-world application. A few under 900 patients were screened using IDx-DR coupled with a Topcon NW-400 automatic fundus camera in a primary care setting. The staff operating the IDx-DR client and taking the fundus images were not IDx-DR or clinical trial staff but preexisting employees of those clinics who underwent standardized training. This is important because in a scenario of large-scale DR

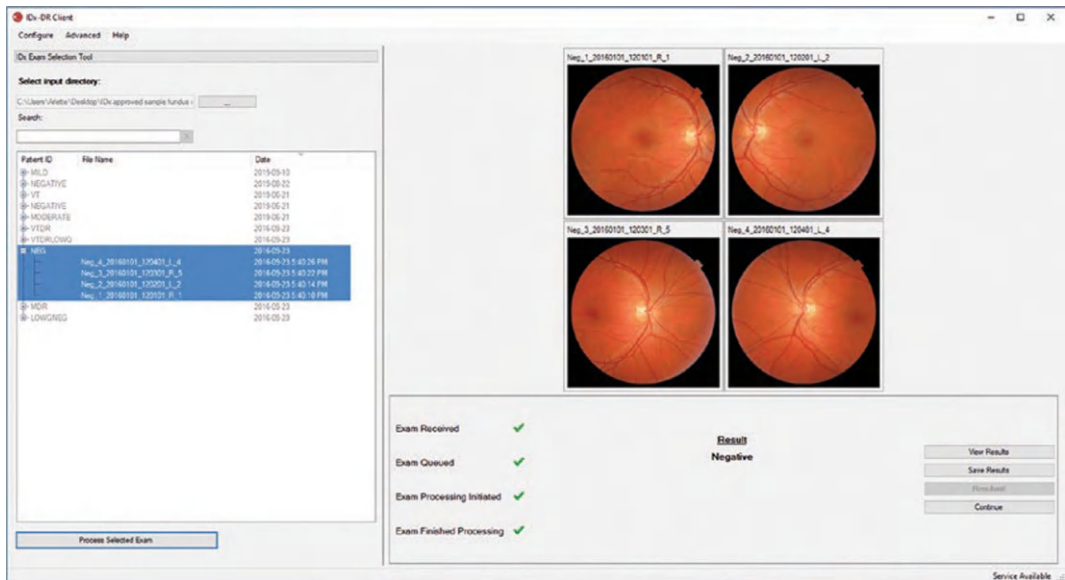


Fig. 11.1 IDx-DR image submission screen

screening deployment, specialized staff, such as those involved in ophthalmology imaging, may find it more difficult to produce and acquire the necessary technical equipment. In previous trials of IDx-DR and other AI algorithms, the performance of the AI algorithm was compared to that of human grading with the same available information, which was mostly fundus images. To strengthen the human grading standard against which the AI was compared, several people graded each image according to consensus grading. This trial took an even more stringent, extreme approach—giving the human graders much more available information while keeping the AI limited to the 4 fundus images taken by relatively inexperienced staff, albeit with an automatic fundus camera and selective mydriasis. This was compared to grading performed on four stereoscopic, widefield fundus images taken by professional technicians and graded by an established, independent reading center—the Wisconsin Fundus Photograph Reading Center. The presence of clinically significant diabetic macular edema (CSME) was also established based on macula OCT imaging, which the algorithm had no access to. With odds stacked against it, the AI was still able to exceed all endpoints set before the trial began, with endpoints at a sensitivity of 87.2% ($> 85\%$), specificity of 90.7% ($> 82.5\%$), and imageability rate of 96.1% (among patients deemed imageable by the reading center). The landmark FDA decision to allow IDx-DR to operate within the United States was largely based on the results of this study [14]. In the US, according to FDA approval, IDx-DR needs to be coupled with the Topcon NW-400 nonmydriatic fundus camera.

Previously, a number of studies on IDx-DR were published, though none were significant. Notably, its performance on the Messidor-2 dataset was significantly better than that in the above-described trial, with 96.8% sensitivity and 87% specificity. In another real-life study performed in the Netherlands, 1410 patients were screened within the Dutch diabetic care system. Three experts graded the resultant images according to the ICDR and EURODIAB grading scales, resulting in significantly different

algorithm performances depending on the scale used. For the EURODIAB IDx-DR, the sensitivity and specificity were 91% and 84%, respectively, whereas for the ICDR, they were 68% and 86%, respectively. The significantly lower performance compared to that of the ICDR could all be attributed to a single aspect of the ICDR—judging a single hemorrhage as at least moderate DR—the authors noted that the sensitivity should change from 68 to 96.1% [15].

This is a great illustration of how important grading criteria are. A number of differing criteria have been used in different studies thus far, including the Eurodiab, ICDR, and ETDRS. Some studies use local grading guidelines, with each being one of the most significant parts affecting the outcome and final performance indicators published. The first question and most important question in establishing DR screening is ‘what is the screening trying to accomplish?’. In the simplest form, the aim of a DR screening initiative should be to identify patients who will require a specialty ophthalmology visit before the next screening episode. This seems to hold true for established traditional screening programmes in developed countries. However, depending on the region and resources available, this can change. In poorer countries with many-fold fewer ophthalmologists and low availability of treatment, one might want to increase the risk of referral. Nevertheless, the scale used to measure and qualify the retinal changes present needs to be backed up by the risk of DR progression and risk to vision at a given level (Fig. 11.2).

Retinalyze

Retinalyze is a DR screening system developed in Denmark. As mentioned, it is one of the first published automated DR analysis programs, with initial reports of its efficacy starting in 2003, based on scanned 35 mm film fundus images. It features a web-based interface, with a per-image result. Images are submitted through the interface utilizing a secure internet protocol. The results are presented in terms of the number/severity of detected retinal changes:

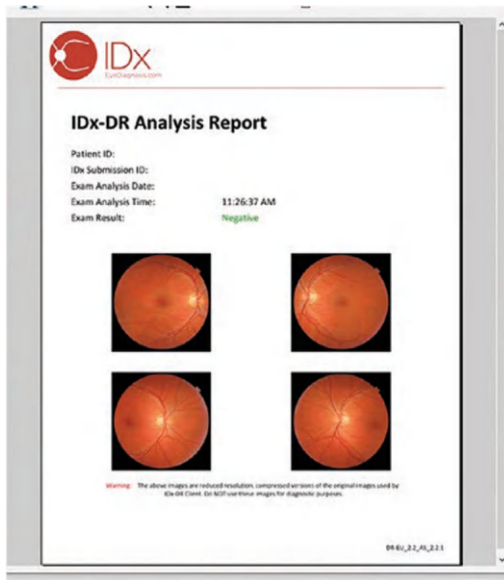


Fig. 11.2 IDx-DR results page

no changes, mild retinal changes, or severe changes. An interesting feature of Retinalyze is that it is able to see an annotated image with the detected retinal changes highlighted; therefore, it is possible to determine what led to the final results of the algorithms. Since its introduction, it went through a period of inactivity and was reintroduced in 2013, with improvements in the modern era of machine learning. It is certified with the Conformité Européenne (CE) level I under the previous regulations. Retinalyze additionally offers screening for AMD and glaucoma from the same fundus photos.

RetmarkerDR

Retmarker is a DR detection system originating in Portugal. It is one of the first AI screening tools successfully implemented in real-life screening, not just for the purpose of a clinical trial. In the central region of Portugal, a long-standing DR screening programme was established in 2001. In 2011, RetmarkerDR was implemented in the existing human-grader-based DR screening programme. This screening is based on several teams of photographers

equipped with mobile fundus cameras. These screening units rotate between different health-care providers covering their whole route over the course of 12 months and then repeat this cycle. Patients with diabetes and no history of DR treatment were invited for screening at a local health center. These images are later collated and sent on a weekly basis to a centralized reading center (Fig. 11.3).

The Retmarker software forms the first line of analysis for those submitted images. Images in which the algorithm detects signs of DR, or progression of DR in case of repeat screening, are sent for human grading, similar to images deemed low quality by the algorithm. In this case, a marker is used in the preliminary ‘disease’ or ‘no disease’ sorting, which then specifies the need for human grader assessment of the ‘disease’ subgroup. For quality assurance, a certain number of negative DR results were also sent for human analysis, with the human graders blinded to the AI decision.

Such implementation of an AI algorithm to detect DR relies on very high sensitivity, as false negatives will rarely be discovered but can compromise specificity. As long as a significant number of images from human analysis are eliminated without missing cases of advanced disease, the process will likely be resource effective, as even a specificity of 50% means almost halving human grader work.

A noteworthy feature that distinguishes Retmarker from other algorithms is its ability to take previous screenings into account. By comparing the fundus images taken at a previous screening visit, the system was able to track retinal changes and determine whether progression occurred. This leads to another interesting avenue—tracking microaneurysms. Microaneurysms disappear over time, and new ones form. Tracking those changes using traditional, human-grader-based methods is very labor intensive but is virtually instantaneous for AI. The rate at which microaneurysms appeared and disappeared was considered the microaneurysm turnover rate. A number of studies have shown that this parameter is a promising predictive factor for future DR progression [9, 16–18].

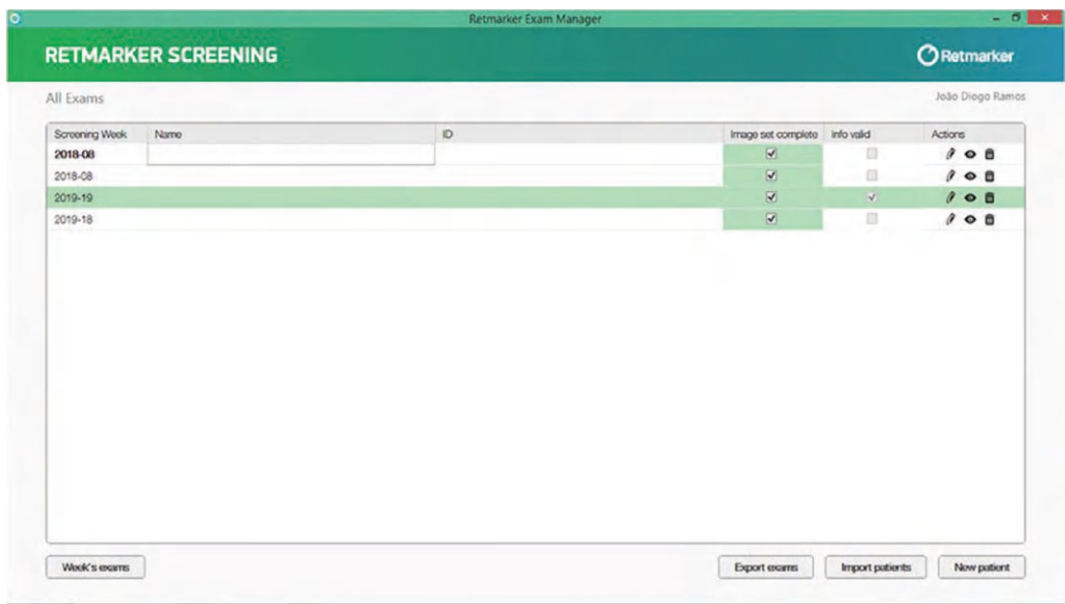


Fig. 11.3 RetmarkerDR exam manager

Although these studies consistently linked increased MA turnover to an increased chance of DR, establishing a clinically significant and actionable link between lesion turnover and diabetic retinopathy progression requires further work (Fig. 11.4).

In addition to being introduced as a part of screening in Portugal, RetmarkerDR was also studied in one of the only head-to-head comparisons of AI DR systems ever published [19]. For the purpose of assessing the potential of introducing autonomous DR detection software into existing English DR screening programs, this study invited AI DR software makers to submit their algorithm for testing. Three systems participated in the testing: RetmarkerDR, Eyeart and iGradingM. Because of technical issues, iGradingM, a DR detection software developed in Scotland, was disqualified from the study, and its parent company has since dissolved. The study involved images taken from consecutive, routine screening visits of more than 20,000 patients to an English DR screening center, which were previously graded as per the national screening protocol and were processed by the systems, and any discrepancies in grading

between the AI and human-graders were sent to an external reading center. Both the efficiency of detecting DR, referable DR and cost-effectiveness were studied [19]. The study concluded with the following sensitivity levels:

- EyeArt accounted for 94.7% of all retinopathies, 93.8% of referable retinopathies (humanly graded as either ungradable, maculopathic, preproliferative, or proliferative), and 99.6% of proliferative retinopathies;
- RetmarkerDR identified 73.0% of patients with any retinopathy, 85.0% of patients with referable retinopathy, and 97.9% of patients with proliferative retinopathy.

Specificity:

- 20% for EyeART for any DR
- 52.3% for RetmarkerDR for any DR

Although the sensitivity levels are much higher for Eyeart, this is equal to the reverse situation occurring in specificity. Of note are the remarkably low specificity levels for both systems compared to more recent reports and estimates of

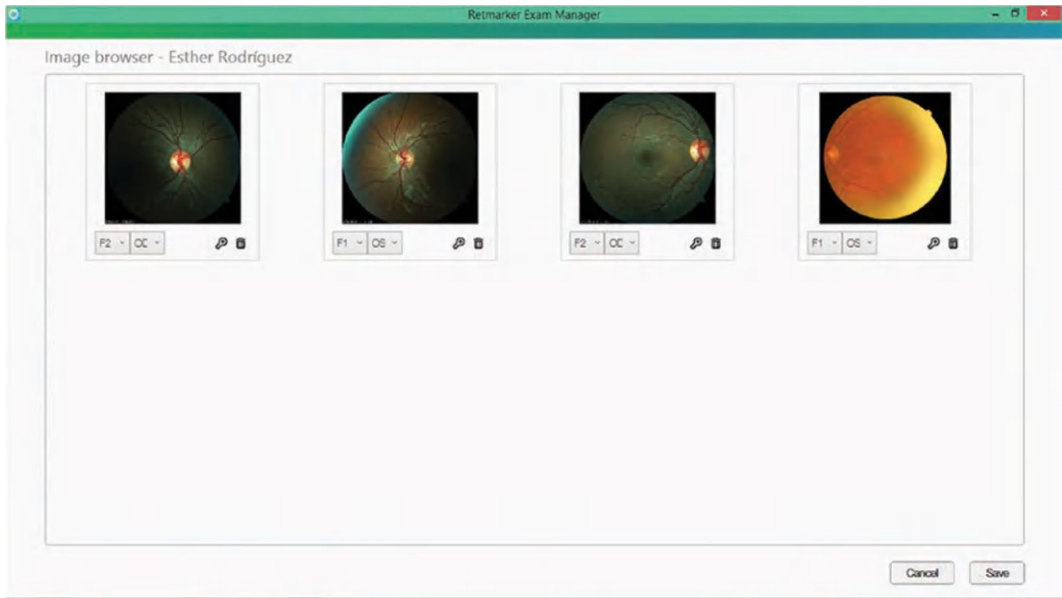


Fig. 11.4 RetmarkerDR image submission

those and other software. It is important to note that although the study was originally published in 2016, it started some years ago. During that period, machine-learning and image analysis methods improved dramatically, and one can assume that the algorithms established for this period improved as well.

Eyeart

Eyeart, the second AI DR screening package to be approved by the FDA, after IDxDR, is being developed by Eyenuk, Inc., which is based in Los Angeles, USA. It also offers another product, Eyemark, for tracking DR progression, which, similar to Retmarker, allows MA turnover measurements. Eyeart is able to take a variable number of pictures per patient, making it suitable for various screening scenarios without the need for further adjustments, in contrast to some of its competitors. This solves a number of issues, as was illustrated by IDx-DR, which had to be specially modified to accept the single image per eye Messidor-2 dataset instead of its typical input of two images.

EyeART was verified retrospectively on a database of 78,685 patient encounters (total of 627,490 images) with a refer/no refer result and a final screening sensitivity of 91.7% and specificity of 91.5%, compared to the Eye Picture Archive Communication System (EyePACS) graders; however, only the abstract for the study was available online. It appears that EyeART has decided to pursue this line of enquiry further with the publication of a full study performed on more than 100,000 consecutive patient visits from the EyePACS database. A total of 850,908 images were analyzed; these images were collected from 404 primary care facilities between 2014 and 2015. Patients generally had 8 images taken, 4 per eye; 1 image of the external eye, a single macula disc-centered image and an image temporal to the disc, although no patient was disqualified because of the number of images taken or their resolution. The images were almost evenly split between nonmydriatic (54%) and mydriatic (46%). The sensitivity and specificity of the final results for detecting referable DR were 91.3% and 91.1%, respectively, which are in line with previous partial results. The sensitivities for detecting higher DRs that

are treatable, either severe or proliferative, were 98.5% and 97.1%, respectively, for detecting CSMEs (as compared to human graders assessing the same fundus pictures). The system accuracy did not seem to change depending on mydriasis, with 98.0% and 98.8% sensitivities for detecting treatable DR in nonmydriatic and mydriatic encounters, respectively. Only 910 patient encounters, less than 1%, were deemed nonscreenable by Eyeart; of those, 198 encounters were previously considered insufficient for full human grading. Nevertheless, of those 910 screening episodes, over one-third had severe or proliferative DR, and the authors noted that the system treats nonscreenable patients as positive for the purpose of patient safety [20].

Eyeart analyzed the whole cohort of over 100,000 screening encounters, almost a million images in less than 2 full days [20]. Assuming an average of 30 s of grading time per image, the

same task would take approximately 7000 work hours or approximately 4 full-time graders working for a whole year, showing how much faster computer analysis can be. Of course, in the actual screening scenario, no one is grading thousands of images at a time, with quick results available within minutes of the screening being much more satisfactory, but AI can do that too 24 h a day every day of the year (Figs. 11.5 and 11.6).

Eyeart achieved similar results in terms of sensitivity to the aforementioned UK study on AI DR screening viability, although there was a considerable discrepancy in specificity between the two studies [19, 21]. As mentioned before, these studies were not performed during the same time period, and further improvements to the system probably account for the increase in its accuracy. Indeed, the authors themselves describe the improvement that the 1.2 version of Eyeart (still based on traditional image analysis

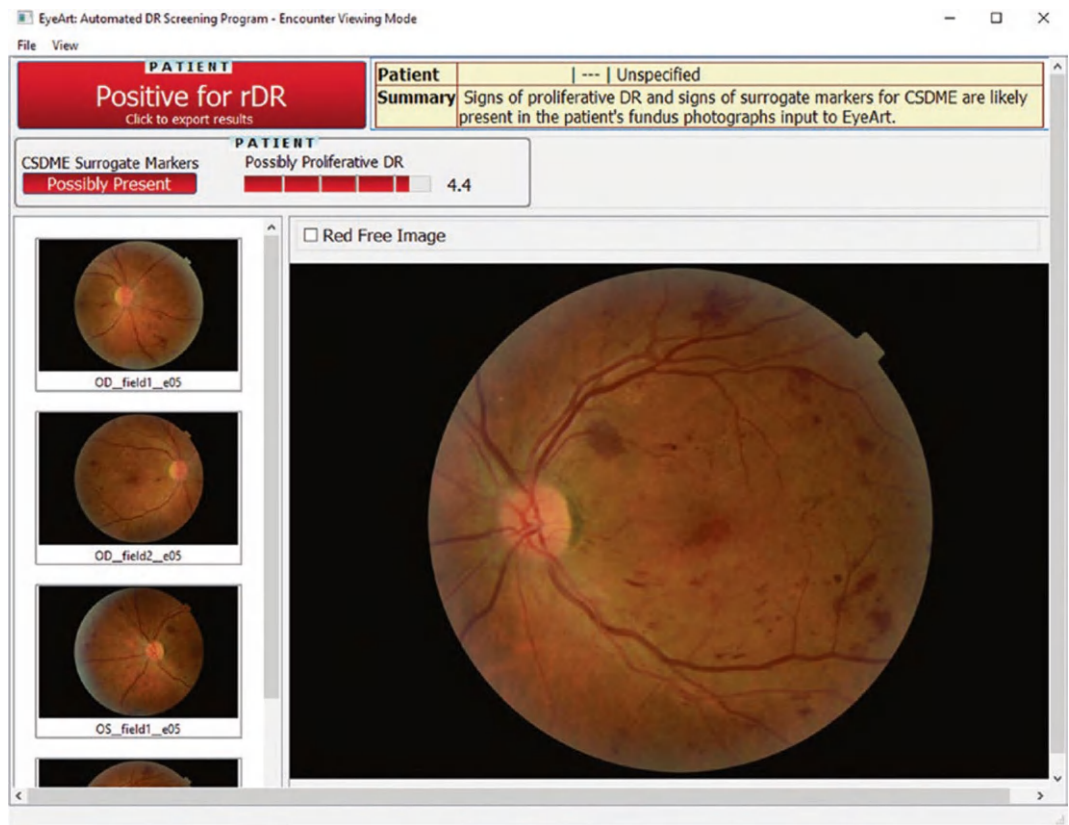


Fig. 11.5 EyeArt results page

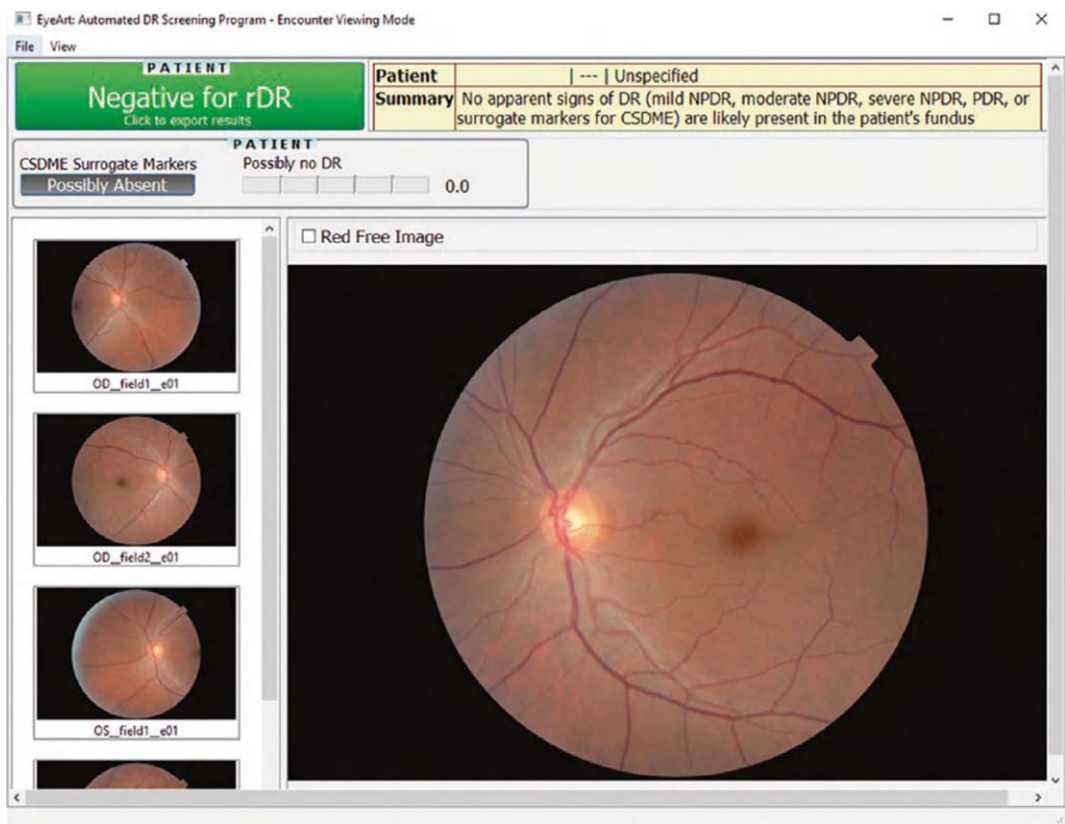


Fig. 11.6 EyeArt results page

techniques) has undergone with the inclusion of multiple convolutional neural networks.

Eyeart was also measured against the Messidor-2 dataset. The referable DR screening sensitivity was 93.8%, and the specificity was 72.2%. Importantly, this dataset does not have a predefined result or grading attached to it, necessitating a separate set of graders to judge it for the standard that the AI is compared against. This grading is separate for each study, further hampering the ability to directly compare any systems involved.

Eyeart has recently published the results of its most robust clinical trial to date. The study was preregistered, as was the IDx-DR pivotal trial, and comprised a similar number of patients—893 patients screened in total. The screening was performed in primary-care clinics with two-field nonmydriatic fundus photography first and 4-field mydriatic imaging

second. This study compared the ability of the EyeART to detect clinically significant DME, moderate nonproliferative DR or higher based on two-field imaging with an external reading center (the Wisconsin Fundus Photograph Reading Center, as was used in the IDx-DR trial) grading decision using 4 wide-field stereoscopic images per eye. For nonmydriatic screening, EyeArt was shown to have high sensitivity (95.5%), good specificity (86%), and gradeability (87.5%). When dilating patients from the initially ungradable group, the overall gradeability of the system increased to 97.4%, while retaining the same sensitivity and increasing the specificity from 0.5 to 86.5%. Although this trial did not involve OCT imaging for the detection of DME, in all other respects, this trial appears similar to the IDx-DR clinical trial, with similar results in terms of both systems' accuracy.

Another result, perhaps even more surprising than the stellar performance of the AI, was a comparison based on a subset of the patients in this trial who underwent dilated ophthalmoscopy after fundus imaging. A total of 497 patients were tested across 10 U.S. clinical centers, with some specialty retinal centers and others general ophthalmology clinics. This was compared against the adjudicated decision of the Wisconsin reading center based on 4 wide-field stereoscopic fundus photographs. Although ophthalmoscope-based examinations had a high specificity of 99.5%, this was coupled with an abysmal sensitivity of 28.1% overall. Even among the retina specialty centers, the sensitivity rate was only 59.1% [22]. This shows that human-based grading using ophthalmoscopy, one of the tools commonly available in primary-care clinics, is very unlikely to be a sensible screening solution if even ophthalmologists struggle with its accuracy.

The most recent study regarding EyeART was performed on 30,000 images taken from the English DR screening programme and followed a very similar protocol and analysis pattern to the only comparative study on AI in DR screening [19, 23]. Images from three different centers were graded according to the established national screening protocol. Among 30,405 screening episodes, EyeART identified all 462 cases of moderate and severe DR. The overall sensitivity for detecting the rDR was 95.7%, and the specificity was 54%. Although the specificity was once again lower than that in other studies, it was still a very significant increase from the 20% specificity in a previous study [19, 23]. The authors concluded that with the introduction of such an AI system into the currently established national screening protocol to replace the primary grader, the overall human grading workload could be halved.

‘Google’ Algorithm

The potential application of new artificial intelligence solutions for the analysis of fundus images, particularly DR, has attracted the

attention of not only smaller, independent teams and companies but also industry giants such as Google. This is not Google’s only foray into medical AI, with teams at Google collaborating to find solutions for automated analysis of histopathology images and other nonimage analysis related publications. A study sponsored by Google Inc. introducing the automated DR screening algorithm was published in 2016 by Gulshan and colleagues. To develop the algorithm, the authors gathered over 128,000 macula-centered images from patients presenting for diabetic screening in India and the US. To validate the resultant algorithm, a random set of images from the same data source was chosen; these images were not used in the creation of the algorithm. The image set for both development and validation consisted of mixed mydriatic and nonmydriatic photos from several different fundus camera models. Additionally, the authors tested the algorithm against the aforementioned French dataset—Messidor-2. The algorithm achieved impressive results, with a sensitivity of 96.1%, a specificity of 93.9% (tuned for high sensitivity), a sensitivity of 87.0%, and a specificity of 98.5% (tuned for specificity). The percentages for the Messidor-2 dataset were 97.5% and 93.4% (high sensitivity) and 90.3% and 98.1% (high specificity), respectively [24]. Although these accuracy results are among the most published and the sample size is considerable, this study emphasized the selection of human graders and their validation. Initially, for the development of the dataset, 54 US-licensed ophthalmologists or ophthalmology trainees aged between 20 and 62,508 images were invited to participate in the study. As a result, each image was graded between 3 and 7 times, looking into both intra- and intergrader agreement. The final DR status and gradeability of the image were set based on majority decisions. A prospective trial was performed to assess the real-world viability of the algorithm. The trial was performed in two hospitals in India on a total of 3049 diabetic patients attending appointments in local general ophthalmology and vitreoretinal clinics, as well as telescreening initiatives. During their

appointments, macula-centered 40–45-degree fundus images were taken mainly with a Forus 3nethra camera, a compact, low-cost fundus camera [25]. All images were nonmydriatic and were not included in further therapeutic decisions for the patients, as they were obtained at their appointments. All images were later graded by a nonphysician trained grader, a retinal specialist. All the images were obtained from one of the two centers (997 patients in total) and were also graded by three retinal specialists via an adjudication process, as described in a previous study. Additionally, any images from the second center with any discrepancies between any of the graders or algorithm outputs (5-point DR grading and DME status) were also adjudicated. The results, in terms of human grading accuracy in detecting rDR, were largely similar to those in the previous study—the four human graders had sensitivities between 73.4 and 88.8%, with specificities between 83.5 and 98.7%. The algorithm had comparable performance, at a sensitivity of 88.9% at the first center and 92.1% at the second center with specificities of 92.2% and 95.2%, respectively [25]. The ‘Google’ DR algorithm was trained on images taken from many different cameras, of which only 0.3% were taken by this specific fundus camera; however, it has shown very good performance on images taken using it, suggesting that the algorithm is able to deal with different equipment being used to take the images [25]. Although the algorithm and its results appear very promising, with good accuracy, it does require further work to be used in a clinical setting, which the authors point out themselves. First, as it currently has no image quality assessment capabilities, only images deemed gradable by the adjudication panel were included in this latest study. Additionally, as with all other algorithms, their place within and the precise protocols of widespread screening and integration into the existing clinical workflow or outside of it remain to be devised and assessed. This recent study was designed specifically because the algorithm did not interfere with the established clinical setup.

SElena+, Singapore Algorithm

Singapore, one of the very few countries that has an established national DR screening programme, is also at the forefront of testing deep learning for DR detection. Ting and colleagues used images from the ongoing Singapore National Diabetic Retinopathy Screening Program (SIDRP), which were additionally graded by 2 senior nonphysician graders and adjudicated by a senior retinal specialist in case of conflicting grading. Overall, 72,610 images taken from 2010 to 2013 were included in the training dataset, and an additional 71,896 from 2014 to 2015 were used for the primary validation dataset. The system was additionally validated using images from multiethnic populations in Singapore and using images taken in screening studies from around the world—China, the African-American Eye Disease Study (US-based), Royal Victoria Eye Hospital (Australia), Mexico and the University of Hong Kong. These studies included between 1052 and 15,798 images for a total validation dataset of 112,618 images and more than 56 thousand patients. Reference standards varied between the different studies, but all included at least 2 graders, with the largest study by image volume ($n=15,798$) also including retinal specialist arbitration.

For the primary validation, that is, the data from the 2014 to 2015 SIDRP, the system demonstrated a sensitivity of 90.5% for detecting referable DR, comparable to that of professional graders on the same dataset at 91.5%, compared to the final retinal specialist arbitration decision. The specificity of this solution was 91.6%, which was lower than that of professional graders (99.3%). Interestingly, the system was better at detecting sight-threatening DR at 100%, with trained graders reporting only 88.6%, again at the cost of lower specificity. As the study included multiple ethnic populations but was devised only on the basis of SIDRP images, the authors analyzed whether it showed racial or other biases. This was made possible by the large racial diversity among the validation

datasets—Malay, Indian, Chinese, White, African-American and Hispanic. The algorithm achieved comparable performance in different subgroups of patients by race; additionally, age, sex, and glycemic control did not affect the accuracy of the algorithm.

Verisee

Verisee, an algorithm was developed based on single-field images taken previously at National Taiwan University with a single fundus camera [26]. The images were graded by two ophthalmologists who were undergoing fellowship training, with an experienced retinal specialist employed for adjudication. The algorithm was trained on approximately 37,000 images, with 1875 images used for validation. The validation dataset was not used for training but was taken with the same camera at the same location. The algorithm achieved 92.2% specificity and 89.5% sensitivity for any DR and 89.2 and 90.1% for the rDR. The algorithm exceeded the sensitivity for detecting the rDR achieved by ophthalmologists in this study, which was calculated at 71.1%, and did much better than that of internal physicians at detecting any DR (64.3% sensitivity, 71.9% specificity, based on diagnoses available in chart records) [26]. Although these results are promising, due to the low volume and homogeneity of the validation dataset, the performance of the algorithm in other scenarios remains uncertain. Nevertheless, the algorithm has been approved by the Taiwanese FDA-equivalent body and is scheduled to be implemented in real-world screening in Taiwan in the near future.

RetCAD

A system developed in the Netherlands allows for the joint detection of DR and AMD from fundus images [27]. This is the only study to show the algorithm's effectiveness at screening for both AMD and rDR at the same time. The validation dataset was rather small, relative

to other studies described here, and comprised 600 images. Nevertheless, the software achieved good accuracy and was able to distinguish between the rDR and referable AMD rather well, with a sensitivity of 90.1% and specificity of 90.6% [27]. Unlike the SELINA software, which can also detect both AMD and DR, both diseases were tested at the same time instead of testing the accuracy against AMD and DR on separate datasets [28]. RetCAD was tested against publicly available datasets of Messidor-2 for DR detection and against the Age-Related Eye Disease Study dataset for AMD and achieved favorable results. However, for all of the above datasets, including the development and validation datasets, only images of good quality were chosen. Together with colleagues we have performed a retrospective comparison of RetCAD and IDxDR using images from real-life screening employing IDxDR deployed in Poland [29]. Image results, two per eye, four total per patient, were analyzed by three graders and a majority decision was taken as the ground truth. The results have shown important differences in calibration between the two systems. The IDx-DR system could detect 141 RDR patients (out of 142 labeled as RDR by graders) with a sensitivity of 99.3%, at a cost of low specificity at 69.0%. The RetCAD system was able to detect 127 patients with referable DR (RDR) with a sensitivity of 89.44%, at a very high specificity of 94.8%. Screening programs should account for such operating parameter differences and resources should be tailored to the model's specific details. Models with higher false-positive rates should have more resources for patient verification. Conversely, a higher false-negative rate should prompt additional expert image grading or earlier rescreening for at-risk patients.

Vistel

Vistel is a Chinese software package that aims at detecting multiple major retinal disorders based on fundus images. It's authors recently published an article describing an analysis of

1493 images from 748 patients while detecting retinal vein occlusion (RVO), retinal artery occlusion (RAO), pathologic myopia, retinal detachment (RD), retinitis pigmentosa (RP), atrophic and neovascular age-related macular degeneration (AMD), epiretinal membrane (ERM), macula hole (MH), central serous chorioretinopathy (CSC), suspect glaucomatous optic neuropathy (GON), and optic nerve atrophy [30]. While sample size for some pathologies like RD ($n=11$) or RP ($n=25$) were extremely low, the results for more common disorders, particularly RDR are promising with 95% sensitivity and 94% specificity in detecting RDR, compared to 87 and 97% achieved on the same dataset by junior ophthalmologists. The article also describes a process in which junior ophthalmologists were presented with the AI's diagnostic output for a given patient and had the opportunity to change their decision, giving an interesting glimpse into what the joint future of ophthalmology and technology might look like [30]. The authors have also previously published an article documenting their progress in incorporating DR staging and specific lesion detection into the algorithm resulting in an improvement of AUC and other accuracy metrics [31].

MonaDR

MONA DR DME (Version 1.0.0) consists of two ensembles: one for diabetic retinopathy (DR) and one for diabetic macular edema (DME). Data was obtained from the EyePACS telemedicine platform, including patient visits from US screening centers [32]. Mona is commercially available in the EU as a Class 1 certified medical device under the European Union Medical Device Directive. The software needs one fundus image per eye centered between the macula and optic disc for algorithmic processing. MONA DR gives a binary output (RDR or no RDR) based on a continuous score between 0 and 4 and a threshold for RDR, reporting three diabetic eye screening results per patient (DR, DME, and a rDR or no RDR result). Mona achieved a robust result when tested

against publicly available datasets Messidor-2, and Messidor-2 Iowa reference with only slight decreases in AUC and sensitivity for DR in Kaggle dataset [32]. Similarly the results from the internal testing dataset were satisfactory with (AUC) of 97.28% for DR and 98.08% for DME.

Other AI DR Solutions

The initiatives described thus far have focused mostly on image analysis. One of the hurdles to overcome with their development is the equipment and technique used to obtain fundus images and how this might affect the system's diagnostic ability or image quality detection protocols. The use of different fundus cameras by many different technicians can introduce considerable variability in picture quality, resolution or sharpness. IDx-DR and Eyeart, for example, are only approved for use in the US when coupled, not only with a single brand of fundus device but also with a specific fundus camera—the Topcon NW-400. Other initiatives employ a number of computational techniques to normalize each image to a standard deemed appropriate for the system. Another line of thinking is that using images from multiple fundus cameras in training the algorithm may help it ignore the non-relevant, fundus camera-dependent changes in the images. This strategy appears to work with most developers reporting that their systems have no significant impact on final accuracy with regard to the fundus camera used. This issue is particularly important in the case of low-cost or mobile fundus cameras. Introducing DR screening in low-resource regions of the world is costly not only in terms of grading but also in terms of equipment cost and portability; establishing permanent, stationary screening points is unlikely to be viable in settings with low population density and low patient mobility. Even in developed, wealthy countries, widespread screening is often performed utilizing mobile screening units, as exemplified by some of the UK-based screening strategies. The rapid development of AI in diabetic retinopathy has not gone unnoticed by companies that already function in the fundus

image field, with companies developing dedicated AI DR screening solutions for their existing fundus imaging hardware.

DR Detection Using Mobile Devices

Another widespread invention of the digital era—the smartphone—and its relatively low cost and ubiquity appear promising in regard to mobile, low-cost screening. In one study, retinal images of 296 patients taken with a smartphone-based add-on and software, the ‘Remidio Fundus on phone’ device, were analyzed by Eyeart software. Although the Eyeart algorithms have not been trained on the use of smartphone-based fundus photography, they achieved a sensitivity of 99.3% for referable DR and 99.1% for sight-threatening DR, with specificities of 68.8% and 80.4%, respectively [33]. Since that study was performed, the authors have developed their own in-house DR analysis software, which was embedded into their current generation Fundus on phone devices (Fig. 11.7).

The software side of Remidio’s DR detection system was named Medios AI. These results have since been replicated in another similar study by Natarajan and colleagues, where 3-field, dilated retinal images taken with a mobile Rhythm camera were compared to the results of a diagnosis by a vitreoretinal specialty resident and specialist based on the same pictures. The images were taken by a healthcare professional with no experience using fundus cameras, and the offline system achieved similarly highly accurate results [34]. In a similar study, the performance of the system was compared with that of 2 vitreoretinal specialists, and the final sensitivity and specificity of the AI for detecting referable DR were 98.8% and 86.7%, respectively [35]. This finding was further corroborated by a study of 900 adult subjects with diabetes in India, where 5 retinal specialists graded images taken with a mobile camera for any DR or rDR. This was later compared to the MediosAI software running offline on an iPhone 6, a 6-year-old mobile device that currently costs less than 200 USD for a refurbished



Fig. 11.7 Remidio FOP device

model. The MediosAI achieved good results, with a sensitivity and specificity pair for any DR of 83.3 and 95.5% and for an rDR of 93 and 92.5% [36]. For the MediosAI, all studies thus far have compared AI and grader performance on the same source material—pictures taken with a mobile camera. We have published a comparison of performance between IDx-DR and MediosAI on a dataset from a polish screening program [37]. The study included 811 diabetic patients from Poland imaged with a stationary Topcon NW-400 fundus camera according to IDx-DR protocol (2 images per eye). Each patient was graded manually by three graders and by ARIAs for diabetic retinopathy (DR) severity. Ground truth results showed no DR in 67% of patients, any DR in 33%, RDR in 22%, and sight-threatening DR in 5%. MediosAI had a 95% sensitivity and 80% specificity for RDR, while IDx-DR had a higher sensitivity at 99% at a cost of lower specificity at 68%. Both ARIAs were accurate with few false negatives, indicating reliable automated DR screening potential [37] (Fig. 11.8).

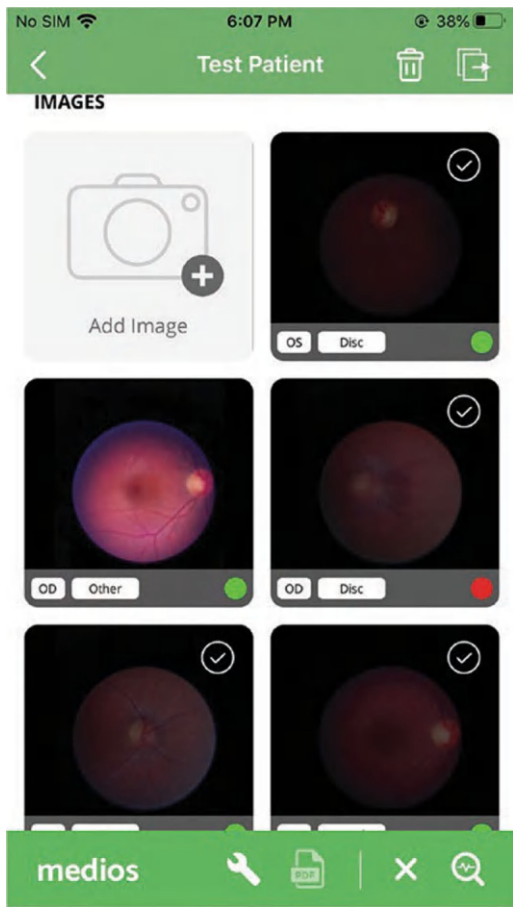


Fig. 11.8 Selection of the MediosAI image

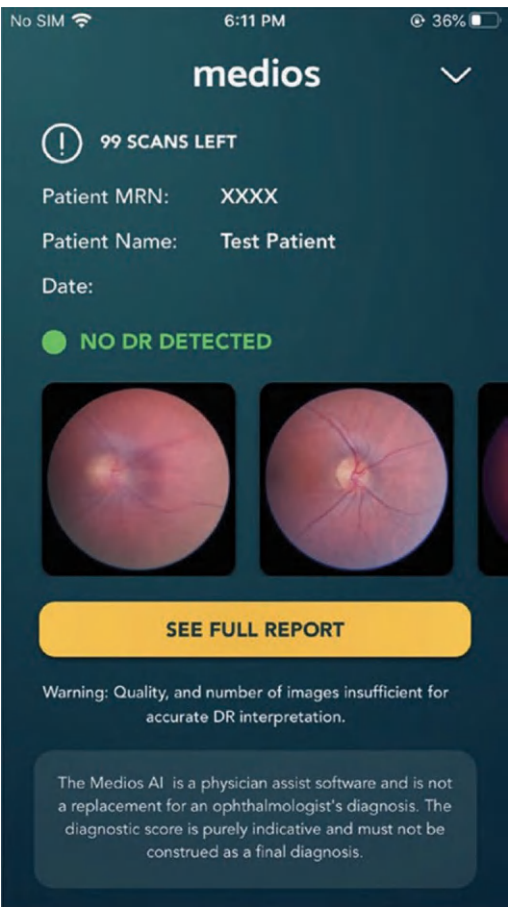


Fig. 11.9 MediosAI report

The major difference in the use of Remidio’s DR screening system, other than its direct implementation in fundus imaging devices, is that the analysis is performed entirely offline, without the need for internet access. Although access to wireless internet sources is spreading all over the world, this can be a very important factor in screening remote and underprivileged communities, where internet access is sometimes not possible and is often unreliable. This approach involves the selection of steam with more mobile, smartphone-based or smartphone-aided fundus imaging solutions being studied and considered for adoption in DR screening. Smartphones coupled with a compatible mobile fundus camera attachment or device provide a low-cost, highly mobile and highly scalable DR

screening solution, especially if the analysis is integrated into the smartphone itself. A recent study conducted in India compared the effectiveness of four such devices for human-based DR grading [38] (Fig. 11.9). It appears that Bosch has also taken a similar approach in improving its ‘Bosch Mobile Eye Care’ fundus camera and developing an in-house DR diagnostic algorithm to be implemented within the fundus camera itself. Single-field images taken with their camera, without pharmacological mydriasis, were analyzed by convolutional neural network-based AI software to assess disease/nondisease or insufficient quality output. The system is cloud-based and requires internet access. Out of the 1128 eyes studied, 44 (3.9%) were deemed inconclusive

by the algorithm, with only 4 out of 568 patients having images from both eyes of insufficient quality. This study compared the performance of AI with that of grading based on 7-field stereoscopic, mydriatic, and ETDRS imaging performed on the same eye. The Bosch DR algorithm achieved good results, with a sensitivity, specificity, PPV, and NPV of 91%, 96%, 94%, and 95%, respectively [39]. However, little is known about the grading criteria employed in this study; in contrast to other similar works, this study employs purely disease-positive/negative criteria rather than the more useful rDR, nonrDR distinction [39]. Unfortunately, no further reports of the effectiveness of these algorithms are available at this time. As of 2024, the screening solution from Bosch remains unavailable in Europe and the USA and appears to be marketed to the Indian market.

After the approval of IDx-DR and Eyeart, the most recent AI DR screening solution to receive FDA approval in April 2024 was a mobile screening device with inbuilt AI-based DR screening named the “AEYE-DS Software Device”, which was developed by US-based AEYE Health. No scientific papers have been published regarding the device thus far; however, there is information available from the preregistered trial at clinicaltrials.gov [40]. The trial analyzed 462 participants and achieved an imageability of 99.1% and sensitivity of 94.7% (95% CI, 85.6–98.2%) while using a Topcon NW-400 camera at two images per eye, similar to the protocol employed by IDx-DR [40]. Unlike IDx-DR, both the mobile and stationary versions of the AEYE health device can be used with a single image per patient protocol. Additionally, an abstract pertaining to the AEYE health device describing the screening of 156 patients with more than mild DR in which a single ophthalmologist served as the ground truth was published, with good sensitivity (93%) and specificity (96%) results [41].

A 2024 Finnish study investigated the performance of 21 ARIAs using images taken with a single handheld retinal camera, Optomed Aurora (Optomed, Tampere, Finland) [42]. The study included images from 106 consecutive patients

from a local DR screening programme and an additional 50 high-morbidity diabetic patients from a hospital ophthalmology outpatient clinic. Macula- and disc-centered color images were taken from each eye (color and red-free, 8 images total per patient), and the color images were analyzed by each of the algorithms. The ground truth was determined by grading by a single ophthalmologist. Although all final results are masked, participating algorithms that consented to their names being published include the aforementioned AEYE health software and RetCad from Thirona. Overall, the algorithms achieved an average of 76% sensitivity and 81% specificity. Nineteen out of 21 algorithms achieved an imageability rate of at least 98%, with the remaining two rejecting 28% and 11% of patient encounters, respectively. The average values for the five best performing algorithms were 89.2% (median 89.3%), 84.3% (median 85.0%) for sensitivity and 91.3% (median 94.8%) for specificity. Similar to previous comparisons, the performance of algorithms varied considerably, with a single algorithm achieving sub20% sensitivity and another achieving 20% specificity [42].

Although mobile screening appears very appealing, and as exemplified above, the results can be very promising, it is conceivable that the lower image quality obtained when using mobile fundus cameras might affect the accuracy of the AI system used to grade it. A recent study compared the performance of a deep learning-based DR detection system against a benchmark, curated image set taken with a desktop camera against its accuracy with images taken with a handheld fundus camera [43]. Although the software, dubbed Pegasus, performed exceptionally well on the curated, desktop dataset, with 93.4% and 94.2% sensitivity and specificity, respectively, this did not translate to an equal detection rate in the handheld camera images, with a statistically significant decrease in accuracy. The parameters for the handheld camera dataset had 81.6% sensitivity and 81.7% specificity—a decrease of more than 10% for each of the parameters [43]. Mobile screening setups and portable cameras are very attractive means

for introducing widespread screening. However, testing on curated, high-quality datasets will overestimate the real-world testing accuracy. Testing the software should be performed in a scenario as close to the desired implementation as possible to achieve accuracy metrics that will be true for real-life screening.

Ground Truth and the Reliability of Human Grading

Setting the ground truth is an essential element required for all stages of developing and verifying a modern diagnostic AI. Although we compare the performance of AI algorithms to that of human graders, as they are the current standard, human grading is not infallible. This is especially true in regions without previous experience in DR screening using telehealth methods, which typically lack an adequately trained workforce for large screening or grading projects. We have previously reported on the variability of grading between retinal specialists who were not purpose trained and who did not participate in previous screening initiatives [44]. Based on the substandard intergrader variability metrics, it seems that every day, ophthalmic practice does not necessarily translate into consistent and adequate image-based grading without additional training and support [44].

In the “Google” study, graders were sometimes shown the images they had previously marked; to measure intragrader reliability or how often given the same image, the grader decided on the same result. Sixteen graders went through enough volume of images for this to be feasibly calculated, and the top 7 or 8 ophthalmologists, based on this measure, were chosen to grade all the images from the validation datasets [24]. Intergrader reliability was also measured for 26 of the ophthalmologists. The mean intragrader reliability for the 16 graders for referable DR was 94%, and the intergrader reliability for the 26 graders was 95.5% [24]. Although 96% reliability sounds impressive and shows good consistency in grading, which still means

that for every 25 images graded, the same person will make a different decision for one of those images upon rereading. Even when the most self-consistent graders were chosen from among several board-certified ophthalmologists, the mean agreement rate for referable DR images was only 77.7% for the EyePacs-1 dataset, with complete agreement among all 8 graders achieving in less than 20% of referable DR images. The grader agreement was much better for nonreferable DR images, with complete agreement for 85.1% of the nonreferable cases [24]. This highlights how many caveats the current universally acceptable grading method and gold standard of certified human grading can have. Out of the 16 graders, on average, 4 out of 100 images were marked differently each time they were assessed by the same person. Out of the 8 most self-consistent graders, only 20% of the referable DR cases were viewed the same by all the graders. This highlights the basic and inherent fallibility of human grading. Although the inconsistencies may be minor to begin with, however they are further compounded when AIs are taught, developed and verified on those imperfect, human-graded datasets.

Human grading issues were further explored in a subsequent 2018 study [45]. This study identified three different ‘grading pools’, namely, EyePac’s graders, certified ophthalmologists and retinal specialists. All the graders had high specificity—97.5, 97.9 and 99.1% for ophthalmologists and 99.1, 99.3, 99.3% for retinal specialists. Sensitivities, however, were much lower for ophthalmologists, who reported sensitivities ranging from 75.2 to 76.4% for individual patients and 83.8% for the majority of patients, than for the adjudicated grading system [45]. Even the majority of retinal specialists in terms of decision grading showed room for improvement, at 88.1%, with individual sensitivities of 74.6, 74.6 and 82.1%. Most cases of discrepancy between the majority grading of ophthalmologists and the adjudicated result stemmed from missed MAs (36%), misinterpreted image artifacts that can be construed as MAs or small hemorrhages (20%), and misclassified hemorrhages (16%).

Cost-Effectiveness of Automated DR Screening

In 2016 in a pioneering paper Tufail and colleagues showed that the both Eyeart and Retmarker could be implemented into the English national DR screening program both as a first-level grader and as a second-level, confirmatory, grader in a cost-effective fashion depending on the specific cost per-patient [19]. This study implemented a relatively early version of AI DR detection algorithms, developed and submitted many months before the paper was published and prior to the current exponential AI sector growth. Recent advancements in the field of AI research promise benefits related to increased accuracy and the effect of more competition and larger scale suggest lower cost. This coupled with all the other arguments commonly touted, whether AI screening is considered, might suggest that cost-effectiveness of AI is apparent. In reality the evidence for cost-effectiveness of AI solutions is conflicting and overall results can depend on even small changes to parameters like accuracy, effectiveness of treatment, cost per patient or the alternative cost of local human grading. One of the arguments made is that the higher cost of human grading in developed countries allows for competitive and cost-effective AI screening pricing in those countries [46]. A 2024 paper modeled cost-effectiveness of AI DR screening using data from 251 thousand patients from a national database with 1100 different scenarios for AI algorithm performance [47]. The study concluded with several interesting results—the most cost-effective algorithm was not the one with overall best accuracy metrics, but one with slightly higher sensitivity at a cost of significant drop in specificity. Overall, under the authors' assumptions only 6 cost-saving and 7 cost-effective screening scenarios were identified [47]. Another study based on Chinese rural population modeled that while AI DR screening is cost-effective compared to a traditional ophthalmologist-based screening, it was not cost-effective in the overall health system context

when all direct and indirect costs were taken into account [48]. Similarly an analysis into an urban Chinese region has showed that in the current iteration an AI-based DR screening would not be cost effective. However, if one factor like increased compliance with referrals is improved by 7.5%, or a 50% increase in on-site screening costs for traditional grading or a 50% decrease in cost of AI grading, the situation would change in favor of AI DR screening [49]. At the same time multiple articles endorsed the economic viability of introducing AI ophthalmic screening in their respective regions, such as Australia [50], Trinidad and Tobago [51], Thailand [52], Brazil [53]. Out of all aforementioned studies, a common unifying conclusion remains that the economic viability and cost-effectiveness of a potential AI DR screening intervention is very sensitive to the AI's cost and assumed treatment effects after diagnosis.

An interesting new avenue is provided by “build-your-own-ai” services, on platforms such as Google Vertex and Amazon Rekognition. These platforms allow for automated machine learning, now named AutoML, making it possible to train advanced image recognition algorithms without coding involved [54]. First reports, detailing an almost step-by-step instruction on how to build your own AI algorithm using those services, have demonstrated considerable success [54, 55]. With the exponential growth of AI these may become a feasible alternative to relatively expensive commercial products providing a cost-effective AI trained on the local population.

New Technologies in Retina Imaging and DR Screening

Although most DR screening efforts are directed toward the analysis of fundus images, there have been significant advancements in employing AI for the analysis of optical coherence tomography (OCT). OCT is commonly used for assessing and monitoring DR and DME on an individual patient basis. Several metrics, such as central

macular thickness, help us establish some objective parameters; nevertheless, the evaluation of OCT scans is still subjective and user dependent, similar to evaluating fundus pictures. The further development of OCT-OCT angiography (OCTA) allows noninvasive tracing of retinal and choroid vasculature; however, the role of OCTA in common ophthalmic practice is not firmly defined, and few objective quantifications are available. The first attempts at using OCTA data for machine learning and automated analysis of DR patients have already been made. OCTA data from 106 patients with type II diabetes and either no DR ($n=23$) or mild nonproliferative DR ($n=83$) were used to train the algorithm to detect DR features from superficial and deep retinal maps [56]. Using a combined approach of using both layers, the system demonstrated an overall accuracy of 94.3%, a sensitivity of 97.9%, a specificity of 87.0%, and an area under the curve (AUC) of 92.4% [56]. Although the relatively high reliability measures are promising, it is important to note that the validation was performed on the training subset. Nevertheless, the study has shown that OCTA can be subjected to deep learning and automated analysis, and we may very well see more such initiatives in the future. The specific computational techniques for detecting DR from OCTA have been further explored in a recent study comparing different neural network approaches to analyzing OCTA and their results. The best performing algorithm achieved an accuracy of 0.90–0.92 [57].

Teaching general practitioners (GPs) to take photos with a mobile fundus camera and subsequently grade them might be an alternative method of widening access to DR screening without the use of AI or automated systems. A recent study investigated training GPs in Sri Lanka to take and grade fundus photos taken with a mobile camera (Zeiss-Visuscout100®). The GPs underwent a training programme delivered by two retinologists; however, of the nine doctors who undertook the training, only two with the best test grading results were chosen for the study. The GPs took and graded nondilated and subsequently mydriatic fundus images,

and their performance was graded against the decision of a retinal specialist after performing a dilated fundus examination using slit lamp biomicroscopy and indirect ophthalmoscopy. Assuming ungradable subjects to be referable, the two GPs achieved sensitivities for detecting the rDR of 85% and 87%, with specificities of 72% and 77%, respectively, for nonmydriatic screening, increasing to 89% and 93%, respectively, specificity and 95% and 96%, respectively, sensitivity after mydriasis. Although training GPs to screen for RDR is theoretically feasible and can achieve good diagnostic accuracy, both the availability of GPs and their ability to take on additional workload are limited. In the aforementioned study, only the two best performing GPs (measured as agreement with the retinal specialist on a test image set) were included; unlike an automated system, the accuracy would likely vary between different GP graders [58].

RetinaRisk, a software developed in Iceland, addresses the issues surrounding DR screening in a different direction. RetinaRisk aims to decrease the overall burden of yearly DR screening by safely extending the time between screening for part of the DR population. Although not explicitly derived from machine learning, it is based on the analysis of extensive datasets. The algorithm includes patient parameters, such as sex, age, HbA1c level, DR status, diabetes type and duration, and blood pressure. As a result, the algorithm presents a recommended time until the next screening interval, which may be longer than the traditionally accepted yearly interval but may also be shorter for a subset of patients at high risk for developing DR complications. In a recent study based on one Norwegian ophthalmic practice between 2014 and 2019, the average screening interval was extended to 23 months compared to the 14-month average for the control group with fixed screening intervals [59]. A recent paper focused on the same issue of safely manipulating DR screening intervals by designing a deep learning algorithm to directly predict retinopathy progression from fundus images, instead of relying on traditional retinopathy grades [60].

The authors retrospectively analyzed patient encounters sourced from a Scottish national DR screening between 2007 and 2019, with over 260,000 patients included overall. The authors demonstrated that the use of deep learning can significantly enhance the prediction of incident referable retinopathy prior to the next screening episode, compared to the current reliance on retinopathy grade alone. This has the potential to positively influence the screening recall interval policy, for instance, by decreasing the anticipated delay of treatment in referable disease cases [60].

Conclusions

The development of deep-learning DR diagnostic software is rapidly increasing. During the last decade, we have seen the concepts surrounding automatic DR screening evolve from a few expert-designed algorithms with varying measures of accuracy to a multitude of different approaches employing the newest developments in deep learning and other fields. We have seen progressively more robust studies emerge, proving the diagnostic or decision-support algorithms to be accurate and reliable, some based on millions of images and others with particularly rigorous setting of their gold standard. During the last two years, a number of software packages have been approved by regulatory bodies around the world and are well on their way to be implemented in widespread screening in the respective countries. Following the general worldwide trend, increasing emphasis is being placed on mobile solutions, which may prove to be a better fit for resource-starved regions. Although the body of evidence for the various algorithms is quite large and constantly increasing, there are significant shortcomings in our current study of AI in DR. Virtually all of the current studies looking into and measuring DR algorithms are sponsored or dependent on the respective algorithm's company. There are very few independent studies, and they are far from one another. The only independent and only robust comparison available, published by

Tufail and colleagues in 2016, compares algorithms tested in 2013. Since then, deep learning and related concepts have progressed almost beyond recognition, and many of the algorithms described here are being constantly updated. Independent studies, particularly comparisons or studies establishing objective criteria through which the respective algorithms could be compared, are missing, with the end-users or consumers left with a considerable dilemma when trying to choose an algorithm for screening their local population.

References

1. Klein BEK. Overview of epidemiologic studies of diabetic retinopathy. *Ophthalmic Epidemiol.* 2007;14(4):179–83.
2. Guariguata L, Whiting DR, Hambleton I, Beagley J, Linnenkamp U, Shaw JE. Global estimates of diabetes prevalence for 2013 and projections for 2035. *Diabetes Res Clin Pract.* 2014;103(2):137–49.
3. Lee R, Wong TY, Sabanayagam C. Epidemiology of diabetic retinopathy, diabetic macular edema and related vision loss. *Eye Vis [Internet].* 2015 Sept 30 [cited 2020 Feb 7];2. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4657234/>
4. Romero-Aroca P, de la Riva-Fernandez S, Valls-Mateu A, Sagarra-Alamo R, Moreno-Ribas A, Soler N. Changes observed in diabetic retinopathy: eight-year follow-up of a Spanish population. *Br J Ophthalmol.* 2016;100(10):1366–71.
5. Scanlon PH. The English National Screening Programme for diabetic retinopathy 2003–2016. *Acta Diabetol.* 2017;54(6):515–25.
6. Pandey R, Morgan MM, Murphy C, Kavanagh H, Acheson R, Cahill M, et al. Irish National Diabetic RetinaScreen Programme: report on five rounds of retinopathy screening and screen-positive referrals (INDEAR study report no. 1). *Br J Ophthalmol.* 2020;Published Online First: 17 Dec 2020.
7. Nguyen HV, Tan GSW, Tapp RJ, Mital S, Ting DSW, Wong HT, et al. Cost-effectiveness of a national telemedicine diabetic retinopathy screening program in Singapore. *Ophthalmology.* 2016;123(12):2571–80.
8. Gardner GG, Keating D, Williamson TH, Elliott AT. Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool. *Br J Ophthalmol.* 1996;80(11):940–4.
9. Hipwell JH, Strachan F, Olson JA, McHardy KC, Sharp PF, Forrester JV. Automated detection of microaneurysms in digital red-free photographs: a diabetic retinopathy screening tool. *Diabet Med.* 2000;17(8):588–94.

10. Hansen AB, Hartvig NV, Jensen MS, Borch-Johnsen K, Lund-Andersen H, Larsen M. Diabetic retinopathy screening using digital non-mydiatic fundus photography and automated image analysis. *Acta Ophthalmol Scand*. 2004;82(6):666–72.
11. Larsen M, Godt J, Larsen N, Lund-Andersen H, Sjølie AK, Agardh E, et al. Automated detection of fundus photographic red lesions in diabetic retinopathy. *Invest Ophthalmol Vis Sci*. 2003;44(2):761–6.
12. Abramoff MD, Lou Y, Erginay A, Clarida W, Amelon R, Folk JC, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Invest Ophthalmol Vis Sci*. 2016;57(13):5200–6.
13. Xie Y, Gunasekaran DV, Balaskas K, Keane PA, Sim DA, Bachmann LM, et al. Health economic and safety considerations for artificial intelligence applications in diabetic retinopathy screening. *Transl Vis Sci Technol*. 2020;9(2):22–22.
14. Abramoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit Med*. 2018;1(1):1–8.
15. Van Der Heijden AA, Abramoff MD, Verbraak F, van Hecke MV, Liem A, Nijpels G. Validation of automated screening for referable diabetic retinopathy with the IDx-DR device in the Hoorn Diabetes Care System. *Acta Ophthalmol (Copenh)*. 2018;96(1):63–8.
16. Haritoglou C, Kernt M, Neubauer A, Gerss J, Oliveira CM, Kampik A, et al. Microaneurysm formation rate as a predictive marker for progression to clinically significant macular edema in nonproliferative diabetic retinopathy. *Retina*. 2014;34(1):157–64.
17. Nunes S, Pires I, Rosa A, Duarte L, Bernardes R, Cunha-Vaz J. Microaneurysm turnover is a biomarker for diabetic retinopathy progression to clinically significant macular edema: findings for type 2 diabetics with nonproliferative retinopathy. *Ophthalmologica*. 2009;223(5):292–7.
18. Pappuru RK, Ribeiro L, Lobo C, Alves D, Cunha-Vaz J. Microaneurysm turnover is a predictor of diabetic retinopathy progression. *Br J Ophthalmol*. 2019;103(2):222–6.
19. Tufail A, Kapetanakis VV, Salas-Vega S, Egan C, Rudisill C, Owen CG, et al. An observational study to assess if automated diabetic retinopathy image assessment software can replace one or more steps of manual imaging grading and to determine their cost-effectiveness. *Health Technol Assess*. 2016;20(92):1–72.
20. Bhaskaranand M, Ramachandra C, Bhat S, Cuadros J, Nittala MG, Sadda SR, et al. The value of automated diabetic retinopathy screening with the EyeArt system: a study of more than 100,000 consecutive encounters from people with diabetes. *Diabetes Technol Ther*. 2019;21(11):635–43.
21. Solanki K, Bhaskaranand M, Bhat S, Ramachandra C, Cuadros J, Nittala MG, et al. Automated diabetic retinopathy screening: large-scale study on consecutive patient visits in a primary care setting. In: *Diabetologia*. New York: Springer; 2016. p. S64–4.
22. Ipp E, Shah VN, Bode BW, Sadda SR. 599-P: diabetic retinopathy (DR) screening performance of general ophthalmologists, retina specialists, and artificial intelligence (AI): analysis from a pivotal multicenter prospective clinical trial. *Diabetes [Internet]*. 2019 June 1 [cited 2020 Feb 26];68(Suppl 1). Available from: https://diabetes.diabetesjournals.org/content/68/Supplement_1/599-P
23. Heydon P, Egan C, Bolter L, Chambers R, Anderson J, Aldington S, et al. Prospective evaluation of an artificial intelligence-enabled algorithm for automated diabetic retinopathy screening of 30 000 patients. *Br J Ophthalmol*. 2020 June 30;bjophthalmol-2020-316594.
24. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316(22):2402–10.
25. Gulshan V, Rajan RP, Widner K, Wu D, Wubbels P, Rhodes T, et al. Performance of a deep-learning algorithm vs manual grading for detecting diabetic retinopathy in India. *JAMA Ophthalmol*. 2019;137(9):987–93.
26. Hsieh YT, Chuang LM, Jiang YD, Chang TJ, Yang CM, Yang CH, et al. Application of deep learning image assessment software VeriSee™ for diabetic retinopathy screening. *J Formos Med Assoc*. 2021;120(1, Part 1):165–71.
27. González-González C, Sánchez-Gutiérrez V, Hernández-Martínez P, Contreras I, Lechanteur YT, Domanian A, et al. Evaluation of a deep learning system for the joint automated detection of diabetic retinopathy and age-related macular degeneration. *Acta Ophthalmol (Copenh)*. 2020;98(4):368–77.
28. Ting DSW, Cheung CYL, Lim G, Tan GSW, Quang ND, Gan A, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA*. 2017;318(22):2211–23.
29. Grzybowski A, Brona P, Krzywicki T, Raumviboonsuk P. Diagnostic accuracy of automated diabetic retinopathy image assessment softwares: IDx-DR and RetCAD. *Manuscr Submitt Publ*. 2024.
30. Li B, Chen H, Yu W, Zhang M, Lu F, Ma J, et al. The performance of a deep learning system in assisting junior ophthalmologists in diagnosing 13 major fundus diseases: a prospective multi-center clinical trial. *npj Digit Med*. 2024;7(1):1–11.
31. Wang Y, Yu M, Hu B, Jin X, Li Y, Zhang X, et al. Deep learning-based detection and stage grading for optimising diagnosis of diabetic retinopathy. *Diabetes Metab Res Rev*. 2021;37(4):e3445.
32. Peeters F, Rommes S, Elen B, Gerrits N, Stalmans I, Jacob J, et al. Artificial intelligence software for

- diabetic eye screening: diagnostic performance and impact of stratification. *J Clin Med*. 2023;12(4):1408.
33. Rajalakshmi R, Subashini R, Anjana RM, Mohan V. Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence. *Eye*. 2018;32(6):1138–44.
 34. Natarajan S, Jain A, Krishnan R, Rogye A, Sivaprasad S. Diagnostic accuracy of community-based diabetic retinopathy screening with an offline artificial intelligence system on a smartphone. *JAMA Ophthalmol*. 2019;137(10):1182–8.
 35. Sosale B, Sosale AR, Murthy H, Sengupta S, Naveenam M. Medios—an offline, smartphone-based artificial intelligence algorithm for the diagnosis of diabetic retinopathy. *Indian J Ophthalmol*. 2020;68(2):391–5.
 36. Sosale B, Aravind SR, Murthy H, Narayana S, Sharma U, Gowda SGV, et al. Simple, mobile-based artificial intelligence algorithm in the detection of diabetic retinopathy (SMART) study. *BMJ Open Diabetes Res Amp Care*. 2020;8(1):e000892.
 37. Grzybowski A, Rao DP, Brona P, Negiloni K, Krzywicki T, Savoy FM. Diagnostic accuracy of automated diabetic retinopathy image assessment softwares: IDx-DR and Medios artificial intelligence. *Ophthalmic Res*. 2023;66(1):1286–92.
 38. Wintergerst MWM, Mishra DK, Hartmann L, Shah P, Konana VK, Sagar P, et al. Diabetic retinopathy screening using smartphone-based fundus imaging in India. *Ophthalmology*. 2020;127(11):1529–38.
 39. Bawankar P, Shanbhag N, K SS, Dhawan B, Palsule A, Kumar D, et al. Sensitivity and specificity of automated analysis of single-field non-mydratric fundus photographs by Bosch DR algorithm—comparison with mydratric fundus photography (ETDRS) for screening in undiagnosed diabetic retinopathy. *PLOS ONE*. 2017;12(12):e0189854.
 40. AEYE Health Inc. Pivotal prospective clinical trial to demonstrate the efficacy and safety of AEYE-DS software device for automated diabetic retinopathy detection from digital fundoscopic images [Internet]. clinicaltrials.gov/; 2023 July [cited 2024 Jan 1]. Report No.: NCT04612868. Available from: <https://clinicaltrials.gov/study/NCT04612868>
 41. Dvey-Aharon Z, Huhtinen P. Screening for diabetic retinopathy in endocrinology clinics by using handheld cameras and applying artificial intelligence algorithms. *J Endocr Soc*. 2021;5(Suppl 1):A419–20.
 42. Kubin AM, Huhtinen P, Ohtonen P, Keskitalo A, Wirkkala J, Hautala N. Comparison of 21 artificial intelligence algorithms in automated diabetic retinopathy screening using handheld fundus camera. *Ann Med*. 2024;56(1):2352018.
 43. Rogers TW, Gonzalez-Bueno J, Franco RG, Star EL, Marín DM, Vassallo J, et al. Evaluation of an AI system for the detection of diabetic retinopathy from images captured with a handheld portable fundus camera: the MAILOR AI study. *Eye*. 2020;1–7.
 44. Grzybowski A, Brona P, Krzywicki T, Gaca-Wysocka M, Berlińska A, Świech A. Variability of grading DR screening images among non-trained retina specialists. *J Clin Med*. 2022;11(11):3125.
 45. Krause J, Gulshan V, Rahimy E, Karth P, Widner K, Corrado GS, et al. Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy. *Ophthalmology*. 2018;125(8):1264–72.
 46. Rajesh AE, Davidson OQ, Lee CS, Lee AY. Artificial intelligence and diabetic retinopathy: AI framework, prospective studies, head-to-head validation, and cost-effectiveness. *Diabetes Care*. 2023;46(10):1728–39.
 47. Wang Y, Liu C, Hu W, Luo L, Shi D, Zhang J, et al. Economic evaluation for medical artificial intelligence: accuracy vs. cost-effectiveness in a diabetic retinopathy screening case. *NPJ Digit Med*. 2024;7(1):43.
 48. Huang XM, Yang BF, Zheng WL, Liu Q, Xiao F, Ouyang PW, et al. Cost-effectiveness of artificial intelligence screening for diabetic retinopathy in rural China. *BMC Health Serv Res*. 2022;22(1):260.
 49. Lin S, Ma Y, Xu Y, Lu L, He J, Zhu J, et al. Artificial intelligence in community-based diabetic retinopathy telemedicine screening in urban China: cost-effectiveness and cost-utility analyses with real-world data. *JMIR Public Health Surveill*. 2023;23(9):e41624.
 50. Pietris J, Lam A, Bacchi S, Gupta AK, Kovoov JG, Chan WO. Health economic implications of artificial intelligence implementation for ophthalmology in Australia: a systematic review. *Asia-Pac J Ophthalmol Phila Pa*. 2022;11(6):554–62.
 51. Ramoutar RR. An economic analysis for the use of artificial intelligence in screening for diabetic retinopathy in Trinidad and Tobago. *Cureus*. 16(3):e55745.
 52. Srisubut A, Kittongsiri K, Sangroongruangsri S, Khemvaranan C, Shreibati JB, Ching J, et al. Cost-utility analysis of deep learning and trained human graders for diabetic retinopathy screening in a nationwide program. *Ophthalmol Ther*. 2023;12(2):1339–57.
 53. Gomez Rossi J, Rojas-Perilla N, Krois J, Schwendicke F. Cost-effectiveness of artificial intelligence as a decision-support system applied to the detection and grading of melanoma, dental caries, and diabetic retinopathy. *JAMA Netw Open*. 2022;5(3):e220269.
 54. Zago Ribeiro L, Nakayama LF, Malerbi FK, Regatieri CVS. Automated machine learning model for fundus image classification by health-care professionals with no coding experience. *Sci Rep*. 2024;14(1):10395.
 55. Jacoba CMP, Doan D, Salongcay RP, Aquino LAC, Silva JPY, Salva CMG, et al. Performance of automated machine learning for diabetic retinopathy

- image classification from multi-field handheld retinal images. *Ophthalmol Retina*. 2023;7(8):703–12.
56. Sandhu HS, Eladawi N, Elmogy M, Keynton R, Helmy O, Schaal S, et al. Automated diabetic retinopathy detection using optical coherence tomography angiography: a pilot study. *Br J Ophthalmol*. 2018;102(11):1564–9.
 57. Heisler M, Karst S, Lo J, Mammo Z, Yu T, Warner S, et al. Ensemble deep learning for diabetic retinopathy detection using optical coherence tomography angiography. *Transl Vis Sci Technol*. 2020;9(2):20–20.
 58. Piyasena MMPN, Yip JL, MacLeod D, Kim M, Gudlavalleti VSM. Diagnostic test accuracy of diabetic retinopathy screening by physician graders using a hand-held non-mydratic retinal camera at a tertiary level medical clinic. *BMC Ophthalmol*. 2019;19(1):89.
 59. Estil S, Steinarrsson ÆP, Einarsson S, Aspelund T, Stefánsson E. Diabetic eye screening with variable screening intervals based on individual risk factors is safe and effective in ophthalmic practice. *Acta Ophthalmol (Copenh)*. 2020;98(4):343–6.
 60. Dai L, Sheng B, Chen T, Wu Q, Liu R, Cai C, et al. A deep learning system for predicting time to progression of diabetic retinopathy. *Nat Med*. 2024;30(2):584–94.

Application of AI in Angle Closure Diagnosis and Management

12

Tin Aung and Xiulan Zhang

Abstract

This chapter explores the application of artificial intelligence (AI) in the diagnosis and management of angle closure, a group of ocular disorders that can lead to elevated intraocular pressure and glaucomatous optic neuropathy. It discusses the anatomical and epidemiological background of angle closure, the current diagnostic modalities such as gonioscopy, ultrasound biomicroscopy, and anterior segment optical coherence tomography, along with their limitations. The chapter highlights how AI advancements have the potential to improve diagnostic accuracy, image analysis, longitudinal disease prediction, and treatment decision-making, addressing global disparities in ophthalmic care.

Keywords

Angle closure · Artificial intelligence · Deep learning · Glaucoma

Background on Angle Closure

Angle closure encompasses a group of ocular anatomical disorders characterized by the appositional or synechial closure of the anterior chamber angle (ACA), which obstructs the outflow of aqueous humor. A common consequence of this condition is elevated intraocular pressure (IOP), potentially leading to glaucomatous optic neuropathy. Angle closure can be primary, resulting from mechanisms such as pupil block, plateau iris, or anterior rotation of the ciliary body, or secondary due to factors like cataracts, iridocyclitis, or prolonged steroid use [1, 2]. In 2002, the International Society of Geographical and Epidemiological Ophthalmology (ISGEO) introduced a classification system that categorizes angle closure into three stages: primary angle closure suspect (PACS), primary angle closure (PAC), and primary angle closure glaucoma (PACG), reflecting the disease's progression [3]. Epidemiological data indicate that the global population of PACG patients aged 40–80 exceeded 20 million in 2013, with projections suggesting an increase to 32 million by 2040. Asians have a higher prevalence of PACG,

T. Aung
Singapore Eye Research Institute, Singapore
National Eye Centre, Singapore, Singapore
e-mail: aung.tin@sneec.com.sg

X. Zhang (✉)
State Key Laboratory of Ophthalmology,
Zhongshan Ophthalmic Center, Sun Yat-sen University,
Guangdong Provincial Key Laboratory
of Ophthalmology and Visual Science,
Guangdong Provincial Clinical Research Center
for Ocular Diseases, Guangzhou, China
e-mail: zhangxl2@mail.sysu.edu.cn

comprising 77% of the global PACG patients, with East Asia having the highest incidence at 1.1% [4, 5]. Although primary open-angle glaucoma (POAG) is the most common type of glaucoma globally, PACG poses a significantly higher risk of severe bilateral visual impairment and blindness [6]. Currently, the diagnosis of angle closure relies heavily on anterior segment imaging examinations such as gonioscopy, ultrasound biomicroscopy (UBM), anterior segment optical coherence tomography (AS-OCT), and Scheimpflug imaging.

The primary goals in managing angle closure are to prevent the progression of the disease, control IOP, and avert optic nerve damage [7]. Treatment options include medical therapy, laser treatments, and surgery. Cholinergic agonists, like pilocarpine, are often used to reopen the drainage angle by contracting the pupillary sphincter. Other IOP-lowering agents used in POAG, such as beta-adrenergic antagonists, and carbonic anhydrase inhibitors, are also applicable in angle closure cases [8]. Laser peripheral iridotomy (LPI) is preferred for treating PAC and early PACG due to its safety and efficacy. For patients who do not respond adequately to medication and laser treatment, surgical options like trabeculectomy, lens extraction [9], and minimally invasive glaucoma surgeries (MIGS) may be necessary [10–29]. Nevertheless, ophthalmologists must consider a variety of complex factors such as patient demographics (age, gender, systemic medication use [30], ocular anatomical features (anterior chamber depth, lens position), and disease severity (degree of elevated IOP [31] and visual field loss [32]) to make tailored treatment decisions and prognosis assessments, which poses a major challenge in ACG management.

In recent years, artificial intelligence (AI) has been increasingly utilized in ophthalmology, benefiting from advancements in medical digitalization and informatization. For angle closure, AI models offer significant clinical value in diagnosis, image processing, longitudinal prediction, and treatment decision-making. They hold promise for addressing the global shortage of ophthalmologists and the uneven distribution of medical resources, thereby enhancing

the diagnosis and treatment of angle closure [33, 34]. Furthermore, the development of large language models (LLMs) also presents considerable potential for advancement [35]. This chapter reviews and summarizes AI research on angle closure, discusses the limitations and challenges of AI models, and proposes future prospects for their application.

Current Diagnostic Modalities for Angle Closure

The evaluation of the ACA using imaging techniques primarily depends on gonioscopy, UBM, and AS-OCT [36]. These methods are crucial for detecting angle closure, significantly enhancing the assessment of the ACA. However, each examination has its limitations and challenges. Every technique offers unique advantages and disadvantages, making it irreplaceable by any single modality.

Gonioscopy

Gonioscopy is a key examination technique for visualizing the ACA dynamically, maintaining its status as the gold standard for angle assessment. This technique is utilized in a variety of applications, including assessing the angle-iris relationship and differentiating between appositional and synechial closures. It allows for the visualization of distinctive signs such as the double-hump sign of plateau iris and the vessels of neovascular angle closure. Additionally, gonioscopy provides important visual information on pigment dispersion, pseudoexfoliation, and elevated episcleral venous pressure. The gonioscopy lens can also be used to examine the optic disc through the central portion of the lens. Beyond initial evaluations in new glaucoma cases or suspects to determine angle status, the three-mirror Goldmann lens facilitates the application of lasers (trabeculoplasty) by allowing simultaneous viewing of all quadrants [37]. MIGS also require intraoperative gonioscopy for angle visualization.

Gonioscopy is divided into static and dynamic examinations. In static examination, the observer refrains from applying pressure to the eye, allowing for an unaltered view of the angle's natural state. This process requires a dark environment and a brief, bright light beam (about 1 mm in length) that avoids passing through the pupil to prevent pupillary constriction [5]. Dynamic examination involves applying lens compression to determine if visible appositional closure indicates permanent peripheral anterior synechia (PAS) and, if so, the extent of such adhesions. It is crucial to distinguish between anatomically closed angles with iridotrabecular contact (ITC) and PAS in eyes with a closed angle. Subtle pressure is applied to the cornea, pushing aqueous humor into the chamber angle. If ITC or apposition is present, the angle will open, making the structures more visible. Conversely, if PAS exists, the angle remains unchanged in the affected areas. A thorough examination of all angle areas is performed and repeated on the other eye.

UBM

UBM is an essential imaging technique for examining the anterior segment of the eye. Initially introduced in the early 1990s by Foster and Pavlin, UBM uses high-frequency ultrasound to create cross-sectional images of the eye [38]. Unlike gonioscopy, which requires specific lighting conditions, UBM can assess angle closure in both light and dark settings [39]. A study by Kong et al. showed that UBM detected larger ITC than gonioscopy, highlighting its ability to visualize transparent and opaque tissues, providing real-time, comprehensive, and objective observations of the living ACA [40]. UBM is the only imaging method that can display the posterior structures of the iris within the anterior segment, such as the ciliary body, suspensory ligament, and anterior choroid, making it invaluable for quantitative analysis of angle morphology [41, 42]. Furthermore, UBM is crucial for understanding the etiology and mechanisms of angle closure, aiding in the decision-making

process for glaucoma surgery by diagnosing conditions like iridociliary cysts, masses, exudates, and plateau iris configuration. It is also instrumental in diagnosing and managing lens dislocation. Notably, UBM is particularly effective in identifying non-pupillary block mechanisms of angle closure, which are prevalent among many Asian patients [43].

UBM serves both qualitative and quantitative purposes in assessing the corneoscleral region, ciliary body, and anterior segment [44–46]. The scleral spur is a crucial structure for qualitative and quantitative examination of the angle structure. Image analysis begins with identifying the scleral spur, a scleral protrusion into the anterior chamber that attaches anteriorly to the trabecular meshwork (TM). It is generally accepted that the area 500 μm anterior to the scleral spur contains all TM, making the identification of the scleral spur key to describing the presence of angle closure.

AS-OCT

AS-OCT provides non-invasive, non-contact, and rapid imaging of the anterior segment, including the cornea, ACA, scleral spur, and the surfaces of the lens and iris. It offers operator-independent measurements and is well-suited for mass screening of angle-closure conditions [47–50].

The principle of AS-OCT relies on low-coherence interferometry to generate cross-sectional images of ocular tissues [51]. The interferometer, which may be based on time domain (TD), spectral domain (SD), or swept source (SS) technologies [52].

AS-OCT is highly effective for the measurements of anterior chamber parameters, precise imaging of the posterior lens capsule, and assessments of corneal shape and functionality [53–56]. Additionally, AS-OCT can measure iris parameters, which gonioscopy cannot, such as iris curvature, area, and thickness—factors independently associated with narrow angles in previous studies [57–62]. With the invention of high resolution AS-OCT, Schlemm's canal imaging *in vivo* is possible [63–65].

An important landmark in interpreting AS-OCT images is the scleral spur, visible as an inward projection of the sclera at the junction between the inner scleral and corneal curvatures. The apposition between the iris and the inner corneo-scleral wall has been used in several studies as a qualitative method for detecting angle closure, although the degree of apposition may vary and does not precisely match the appositional closure defined by gonioscopy [66].

Limitations and Challenges of Current Methods

Existing examination methods each have their limitation. Table 12.1 illustrates the advantages and disadvantages of these methods, highlighting that no single technique can completely replace another. Gonioscopy, though critically important in glaucoma evaluation, is reportedly not performed in about 50% of patients [67, 68]. One main reason for its limited use is its perceived technical difficulty. As a subjective and

somewhat invasive method, it requires topical anesthesia and can cause discomfort. It is also time-consuming, demands skilled operators, and has a steep learning task. UBM also presents significant challenges that limit its clinical use. Being a contact procedure, it can cause discomfort and potentially risk corneal injury. The examination is time-consuming, and the required supine position may not accurately reflect the true state of the ACA [69]. Quality imaging depends on experienced examiners. While prior studies have shown excellent intra-observer reproducibility, they report poor inter-observer reproducibility in assessing the ACA or iris dimensions [70, 71, 72]. AS-OCT, which is based on optical imaging principles, has limitations in visualizing structures beyond the iris, such as the posterior chamber and ciliary body, as well as detecting neovascularization or pigment. Poor agreement between gonioscopic and AS-OCT findings has been noted in the literature. The reproducibility of AS-OCT findings in the inferior quadrant is particularly poor due to variable placement of the scleral spur [73–75].

Table 12.1 Advantages and disadvantages of current diagnostic modalities for angle closure

	Gonioscopy	UBM	AS-OCT
Advantages	Reference standard Portable and cheap tool Evaluation of angle characteristics (e. g. width, iris, insertion, iris profile, pigmentation, vessels) Essential for lasers Compression gonioscopy (discrimination between synechial or appositional closure) Provide detailed 360° view of anterior chamber angle	Visualize structures behind the iris Objective examination Unaffected by refractive media Suitable for quantitative analysis Can image through opaque corneas	Widely available technology High image resolution Easy to perform Fast, non-invasive and non-contact imaging technique Built-in software to perform angle measurements Suitable for deep-learning and automatic classification (not commercially available)
Disadvantages	Steep learning curve Not possible to visualize structures behind the iris No measurements Photography with a photographic slit-lamp is time-consuming Discomfort and are at risk of corneal injury Require topical anesthesia	Need for a water-bath immersion with direct contact to the eye Long image acquisition times Require an experienced operator Discomfort and at risk of corneal injury Small field of view	Landmarks sometimes hard to recognize Partial agreement with slit-lamp gonioscopy Lacking information on essential findings (e.g., pigments, neovessels) Not possible to visualize structures behind the iris Lacking the capacity for dynamic angle examination

AI in Angle Closure Detection

AI is a branch of computer science dedicated to creating algorithms that simulate intelligent behavior [76]. Its subfields, machine learning (ML) and deep learning (DL), have seen rapid advancements in recent years [77, 78]. ML employs statistical methods to help systems enhance their performance through experience, while DL, a deeper layer of ML, uses neural networks to learn from vast datasets [77, 79].

The integration of AI with ophthalmology began with posterior segment diseases such as diabetic retinopathy [80–84]. Recently, innovative AI applications are revolutionizing glaucoma management by introducing significant advancements [85, 86]. For example, iGlaucoma, a sophisticated tool accessible via smartphones and cloud technology, leverages DL to identify visual field defects, marking a significant step forward in making diagnostic tools more accessible [87]. Another pioneering development, FusionNet, integrates visual field and OCT data to enhance the reliability and broad applicability of its analysis, demonstrating markedly improved precision over models that rely on a single modality [88]. Furthermore, AI-driven predictions using baseline color fundus photographs have shown promising results in predicting the onset and progression of glaucoma [89].

There is now emerging evidence demonstrating the application of AI to the diagnosis and management of a variety of anterior segment conditions [50, 90–93]. The combination of AI with various anterior segment imaging techniques, including gonioscopes, UBM, and AS-OCT, has garnered widespread attention for its role in detecting angle closure and related applications.

AI in AS-OCT

The advancement of AI has brought transformative changes across various fields, including medical imaging analysis [94, 95]. Recently, AI techniques have significantly impacted the

analysis of AS-OCT images, revolutionizing how ophthalmologists diagnose and manage glaucoma. AS-OCT imaging provides high-resolution cross-sectional scans of the eye's anterior segment, enabling detailed examinations of critical structures like the cornea, iris, and anterior chamber [59, 66, 96–98].

The complexity and vast amount of information in AS-OCT images pose challenges for manual analysis due to subjectivity, poor intraobserver and interobserver agreement, and a steep learning curve [99, 100]. Variability introduced by mechanical compression during imaging and alterations in angle configuration due to lighting and lens placement complicate analysis further [99]. The integration of AI, including ML and DL techniques, promises to automate and enhance the accuracy of angle closure assessments [101–103].

This section focuses on AI in AS-OCT images, providing a comprehensive overview of the advancements, challenges, and future prospects in this emerging field.

AI in Angle Closure Classification from AS-OCT

AI applications in angle closure detection have significantly advanced, particularly through integration with AS-OCT imaging (Table 12.2). By 2013, ML techniques were already enhancing glaucoma classification, using AS-OCT images to distinguish between open angle and angle-closure glaucoma, with models achieving AUC values exceeding 0.95 [104]. This high accuracy suggests that AI can consistently and reliably identify angle closures, proving invaluable in clinical settings.

The evolution from ML to DL has allowed for significant breakthroughs. DL excels at extracting and transforming image information, which is abundant in AS-OCT images [101, 105]. The development of an automated DL system for angle-closure detection using AS-OCT images represents a significant advancement, featuring a multilevel deep network that focuses on crucial regions like the global anterior segment structure, iris, and ACA [106]. This system has shown

Table 12.2 Summary of AI studies based on AS-OCT images

Year	Author	Output	Dataset	Imaging device
2019	Fu et al.	Open angle closure versus angle closure	7375 images with open-angle and 895 images with angle-closure	Time domain Visante
2019	Fu et al.	Open angle versus angle closure	AS-OCT machine: 7375 images with open-angle and 895 images with angle-closure; HD-OCT machine: 1102 images with open-angle and 300 images with angle-closure	Time domain Visante and Cirrus HD-OCT
2019	Xu et al.	1. Open angle versus angle closure; 2. Yes/PACD versus No PACD	1943 images with open angles, 2093 images with closed angles	Swept-source Casia SS-1000
2019	Hao et al.	Open angle versus narrowed angle versus angle closure	1024 images with open angles, 1792 images with narrowed angle, and 2048 images with angle closure	Swept-source Casia-2
2020	Fu et al.	1. Scleral spur localization; 2. Angle closure classification	4800 images with 1920 angle closure and 7680 open angle	Swept-source Casia SS-1000
2021	Hao et al.	POAG, PACS or PAC/PACG	1584 image sequences with 504 open ACA, 742 appositional ACA and 338 fully closed ACA	Swept-source Casia-2
2021	Li et al.	Open angle, narrow angle, and angle-closure	6000 images, with 2000 open-angle, 2000 narrow-angle, and 2000 angle-closure	Swept-source Casia SS-1000
2016	Niwas et al.	The five various classes of ACG mechanism	156 samples split into five classes	Time domain Visante
2021	Phams et al.	ASOCT segmentation, and Quantification of ASOCT measurements	420 images for training, 200 for validation and another 200 images for testing	Swept-source Casia SS-1000
2021	Wanichwecharungruang et al.	Presence of plateau iris (P) and absence of plateau iris (NP)	218 with the presence of plateau iris (P) and absence, 1200 and 1300 augmented images	Time domain Visante
2020	Xu et al.	X, Y location of the scleral spur	Training of 17,704 images and testing of 921 images	Swept-source Casia SS-1000
2022	Choy et al.	Automatic segmentation of mouse Schlemm's canal	163 images	Speckle variance Enisu R2200
2022	Li et al.	Detection of narrow iridocorneal angles and peripheral anterior synechiae	8694 scans	Swept-source Casia SS-1000

superior performance, with AUC scores increasing from 0.950 to 0.962 on specific clinical datasets. Additionally, a novel method for classifying ACAs in AS-OCT images has enhanced the diagnosis of angle-closure glaucoma, achieving an AUC score of 0.920 [107].

Recent developments include a multi-sequence deep network architecture evaluated on over 16,896 images across 66 eyes [108]. This approach offers a novel classification into open, appositional, and synechial angles, demonstrating superior performance under various lighting conditions, and surpassing traditional methods in terms of applicability, effectiveness, and accuracy. Further advancements include DL classifiers, such as Inception v3 with transfer learning, and ResNet-18 architectures, achieving impressive AUC scores up to 0.933 for detecting angle closure [109]. These AI-driven methods provide precise, automated assessments that enhance clinical diagnosis and could potentially improve clinical practice in glaucoma management [110].

Most AI models based on AS-OCT primarily assess static ACA status. In practice, observing PAS under dynamic gonioscopy is crucial for diagnosing and assessing the severity of angle-closure glaucoma. A significant development in

this area is the 3D DL-based automated (DGS), which has shown robust diagnostic performance in identifying narrow iridocorneal angles and PAS, achieving high AUC, sensitivity, and specificity values [92] (Figs. 12.1 and 12.2). This system not only performs excellently but also provides highly interpretable visual results (Fig. 12.3), demonstrating its effectiveness in identifying PACG cases and holding promise for widespread application in primary eye care, especially for high-risk individuals.

The figure shows the receiver operating curve of angle classification by the 3D digital gonioscopy system with an area under the curve (AUC) of 0.943 (0.933–0.953).

Automated Classification of Angle Closure Mechanisms

The classification of angle closure mechanisms, such as pupillary block and plateau iris, holds significant clinical implications [111–113]. Classifying the mechanism of angle closure, in addition to categorizing the angle itself, allows for a deeper understanding of the etiology and severity of ACG in patients [112]. This knowledge is critical in guiding treatment decisions, including options like peripheral iridotomy or

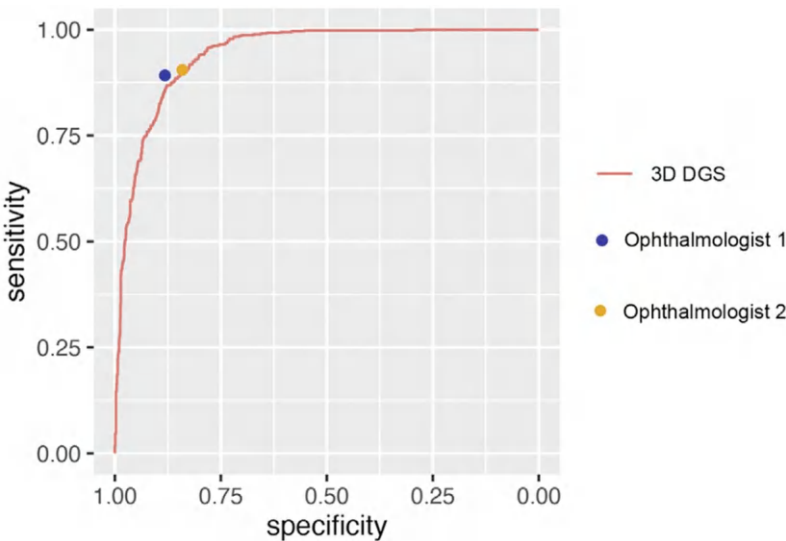


Fig. 12.1 Comparison of diagnostic performance of the 3-dimensional deep-learning based automated digital gonioscopy system in angle classification with ophthalmologists in the external test set [92]

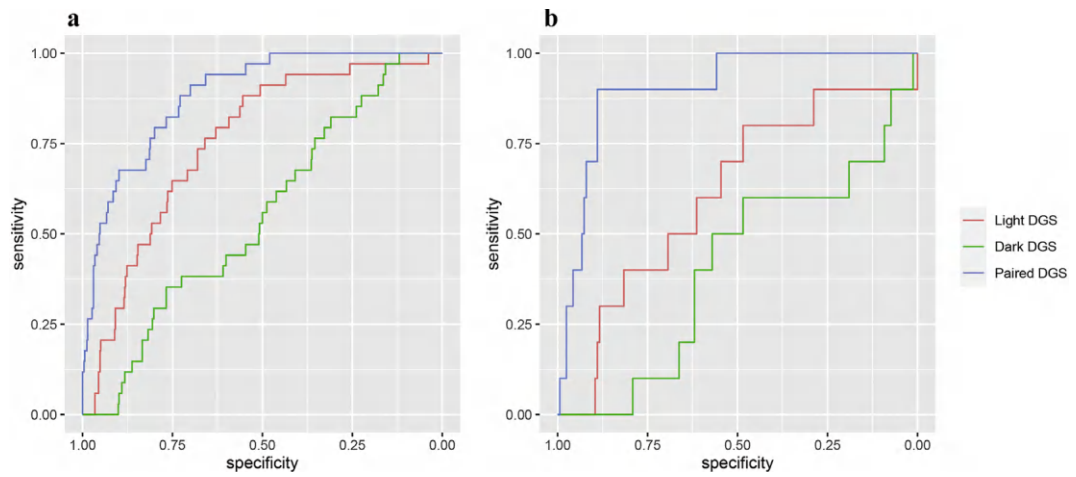


Fig. 12.2 Comparison of diagnostic performance of the deep-learning based automated digital gonioscopy system in peripheral anterior synechia detection in the external test set [92]. The Paired DGS using paired light and dark data achieved the best diagnostic performance with an area under the curve (AUC) of 0.885 (0.836–0.933) at clock-hour level (A) and an AUC of 0.902 (0.818–0.985) at quadrant level (B)

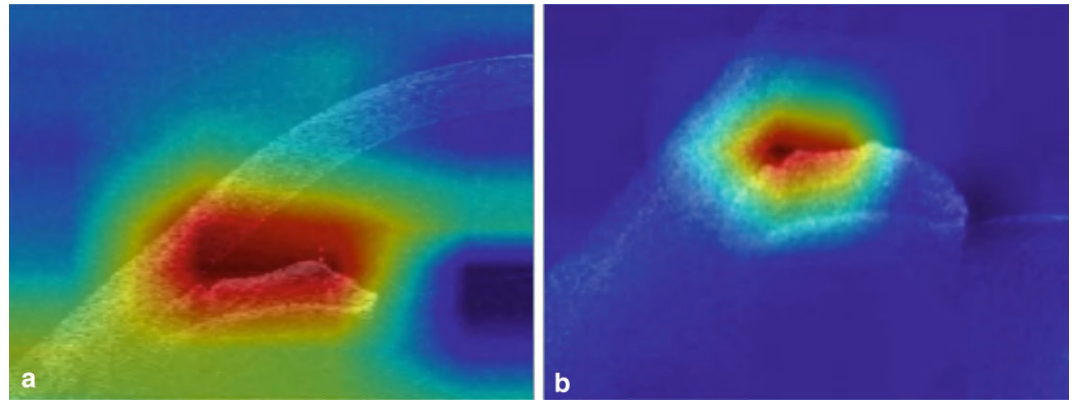


Fig. 12.3 Visualization results of digital gonioscopy [92]. A&B Digital gonioscopy identifies hotspots for narrow angles and adhesions

trabeculectomy [113]. By understanding the underlying mechanisms and severity of angle closure, clinicians can tailor treatment strategies to target the specific pathophysiology, leading to more effective management of angle closure glaucoma [8, 114].

Advancements in AI technology now enable the classification of angles from AS-OCT images and the analysis of underlying mechanisms of angle closure [115]. By leveraging ensemble learning techniques and error-correcting output codes, AI can achieve high

classification accuracy [116]. Its effectiveness is demonstrated in the classification of various mechanisms of ACG with an accuracy of up to 87.65% through advanced validation techniques. Further AI developments have improved the identification of PACG and its specific mechanisms, including pupil block, plateau iris, peripheral iris roll, and exaggerated lens vault [117] (Fig. 12.4). Using multiple feature selection algorithms combined with novel ranking criteria, AI models can pinpoint the most reliable discriminant features from AS-OCT images

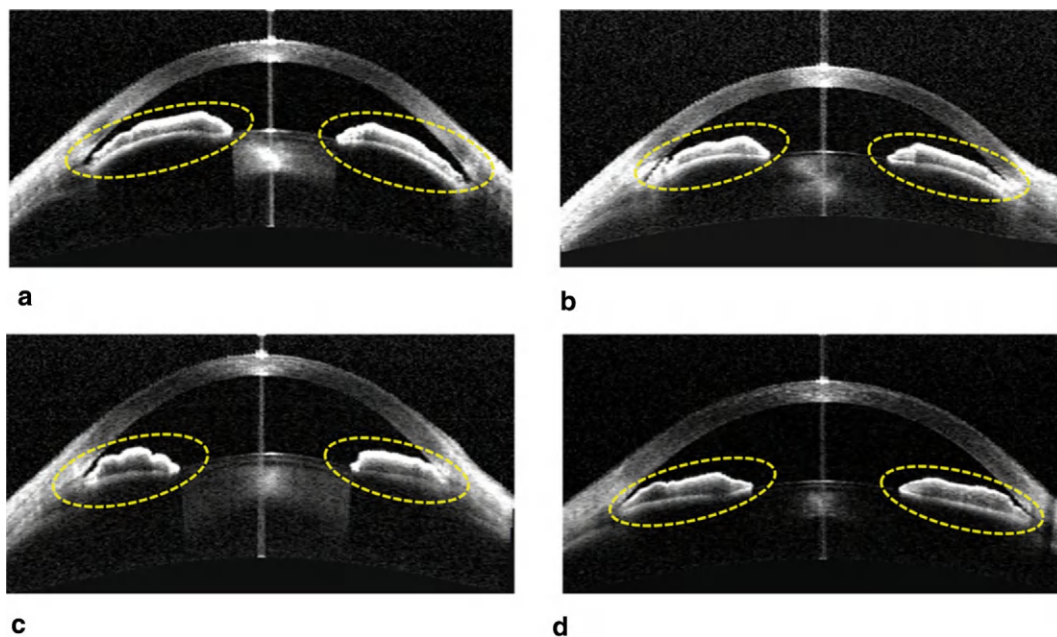


Fig. 12.4 The mechanisms of angle closure involved in AS-OCT images in the series of studies by Niwas et al. [117]. AS-OCT image of an eye with different

mechanisms: A. exaggerated lens vault—LV; B. pupil block—PB; C. thick peripheral iris roll—PIR; D. plateau iris configuration—PL

for PACG detection. This method significantly enhances the accuracy in identifying different PACG mechanisms, showcasing the value of refined feature selection. Additionally, AI's capability extends to utilizing redundant features to improve diagnostic accuracy of ACG [118]. By evaluating different feature selection algorithms, AI models have demonstrated that incorporating redundant features can enhance diagnostic performance, contrasting with methods that prioritize minimal redundancy. Overall, AI provides a robust framework for the automated and precise classification of ACG mechanisms, significantly reducing the subjectivity and manual effort traditionally associated with this process. These advancements hold promise for more personalized and effective glaucoma treatment strategies.

AI in AS-OCT Images Segmentation

The development of AI-based models for anatomical structure recognition and segmentation has become increasingly significant [119–121]. These models have the potential to perform impressive image segmentation and localization

tasks. Moreover, they promise future applications in intelligent navigation systems for anterior segment surgeries, such as MIGS, based on AS-OCT imaging.

A deep convolutional neural network (CNN) was utilized for the localization of the scleral spur and a segmentation approach for AS-OCT images [122]. The neural network has achieved high accuracy in segmenting key structures such as the iris and anterior chamber, with a Dice coefficient of 95.7%. This precision enables the accurate isolation and quantification of intraocular dimensions, critical for diagnosing and managing eye diseases like angle closure glaucoma. Moreover, AI models are capable of performing these segmentations and measurements in less than two seconds per scan. This efficiency supports a comprehensive and objective assessment of the anterior segment, facilitating both diagnosis and treatment. In another application, a CNN based on the ResNet-18 architecture has proven effective in detecting the scleral spur in AS-OCT images [123]. This model has shown precision with mean absolute prediction errors close

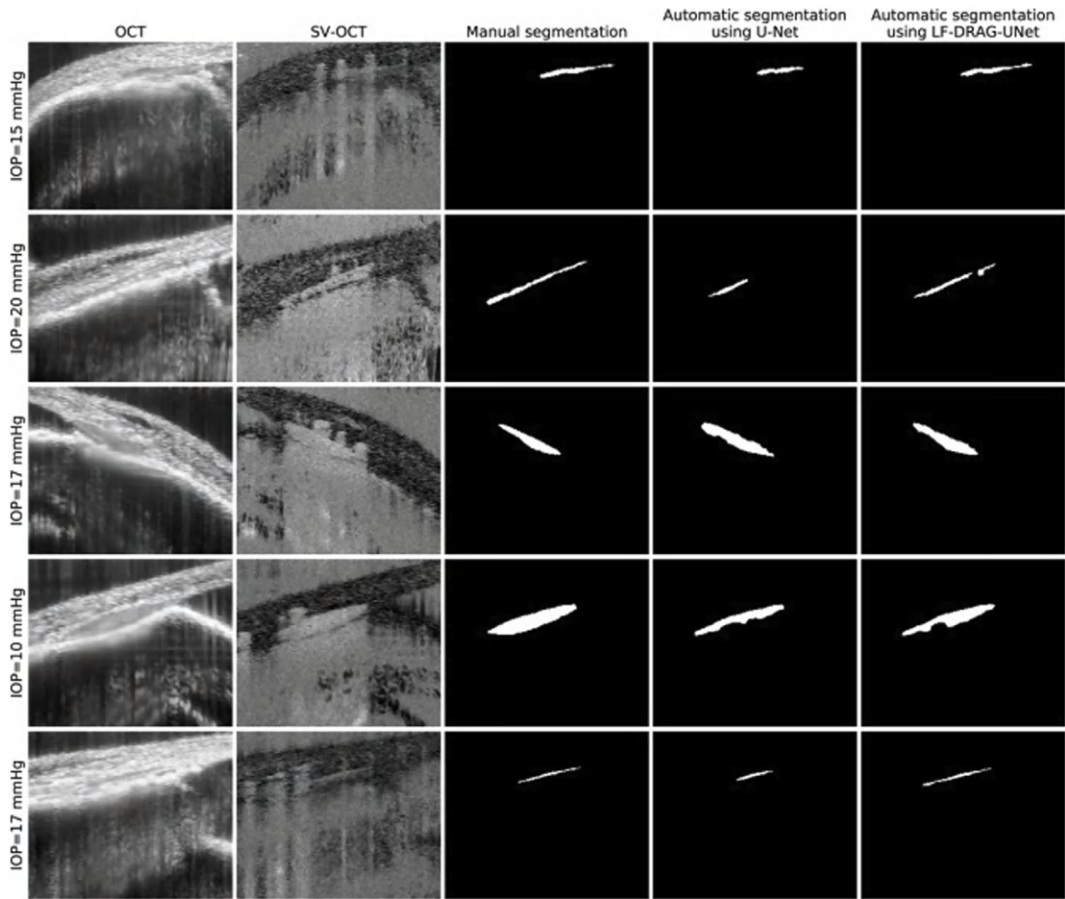


Fig. 12.5 Comparison between automated and manual segmentation of the Schlemm’s canal lumen in AS-OCT scans of living mouse eyes [124]

to those seen in human intra-grader variability for both X and Y coordinates. Such capabilities indicate that deep neural networks can provide reliable, automated assessments of AS-OCT images, enhancing both qualitative and quantitative analysis necessary for optimal ophthalmic care.

DL has also been used to segment the Schlemm’s canal in AS-OCT scans of mouse eyes [124] (Fig. 12.5). The algorithm, based on a U-Net architecture, achieved a mean Dice Similarity Coefficient comparable to manual segmentation by expert graders. This development not only supports more accurate research into IOP and glaucoma but also offers open-source software to facilitate further studies.

Illustrative example showing the superior performance of the proposed LF-Drag-UNet (Late Fusion Dilated Residual Attention-Gated U-Net) compared to baseline U-Net. Columns from left to right depict the intensity OCT image, SV-OCT image, ground truth obtained by manual segmentation, segmentation obtained by the baseline U-Net, and segmentation obtained by LF-Drag-UNet.

Public Database of AS-OCT Images

Automated medical image analysis algorithms, powered by DL techniques, have shown promising results in ophthalmology [80–82]. However, most studies have focused on retinal fundus photographs, with limited attention to AS-OCT

images [125–127]. The lack of a unified evaluation dataset and a large-scale AS-OCT dataset impedes proper comparison and limits the rapid development and deployment of DL techniques for angle closure detection.

A notable initiative was the organization of an international challenge, which focused on tasks such as scleral spur localization and angle closure classification [90]. While many medical image analysis algorithms have been developed for glaucoma diagnosis, there is a lack of focus on AS-OCT imaging. Additionally, the absence of a public AS-OCT dataset hinders the development of automated techniques for angle closure detection and assessment. To address this, Xu et al. organized the Angle closure Glaucoma Evaluation challenge (AGE) in conjunction with MICCAI 2019 (Fig. 12.6). The challenge included tasks such as scleral spur localization

and angle closure classification, with a dataset of 4800 annotated AS-OCT images released for evaluation. Over 200 teams participated, submitting more than 1100 results. Their review paper provides a summary and analysis of the eight onsite challenge methods and their corresponding results. The top-performing approach achieved promising outcomes in both scleral spur localization and angle closure classification. These findings highlight the potential of AI techniques in advancing AS-OCT image analysis and image-based angle closure glaucoma assessment. Moreover, AI algorithms have been used for establishing a normative database for AS-OCT-measured ACA parameters, which mitigate the need for manual scleral spur (SS) annotation and provides more objective and repeatable measurements for angle assessment in clinic [128].

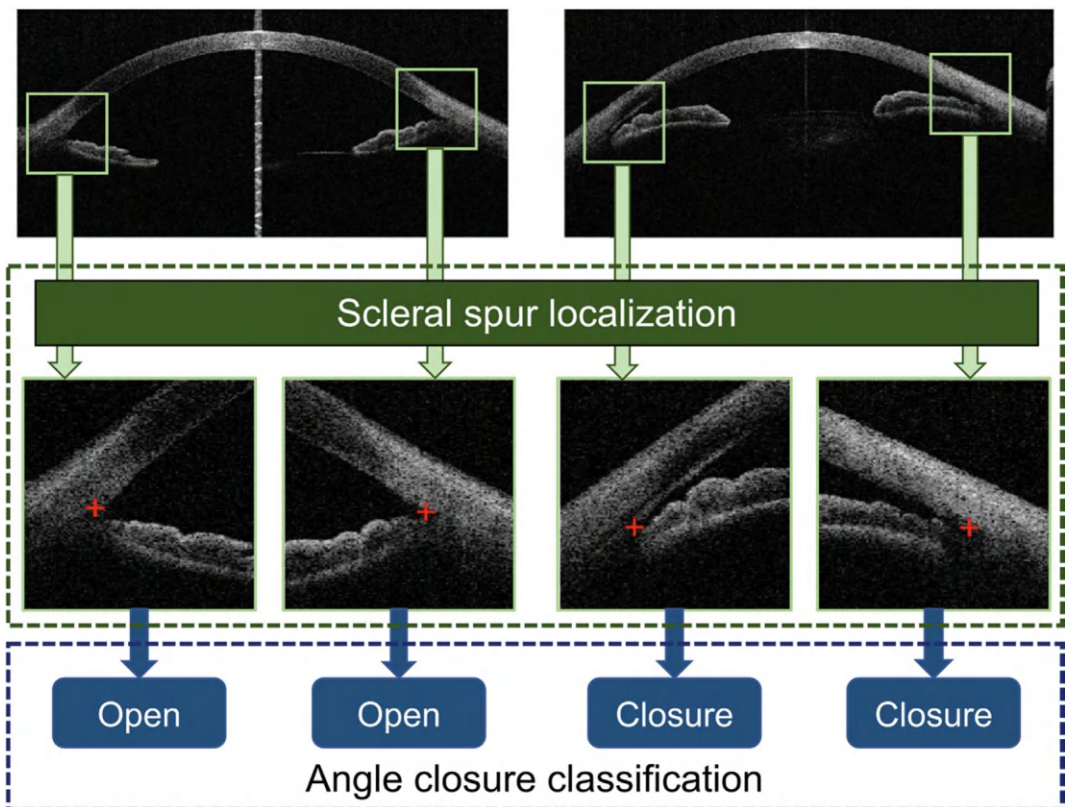


Fig. 12.6 AGE challenge tasks [90]. AGE challenge tasks: scleral spur localization and angle closure classification from AS-OCT images

AI in UBM Images

Like AS-OCT, UBM is also a valuable imaging modality that provides high-resolution cross-sectional images of the anterior segment of the eye [129]. It offers detailed visualization of structures such as the cornea, iris, ciliary body, and ACA, which is particularly useful for assessing conditions like angle-closure glaucoma [130]. The integration of AI algorithms in analyzing UBM images shows great promise in automating and standardizing the interpretation process, thereby enhancing efficiency, accuracy, and reproducibility.

AI has proven effective in automating the localization and classification of the ACA using UBM images, crucial for diagnosing primary angle-closure glaucoma [131]. The AI method achieved a 95% accuracy in identifying the angle recess and an overall accuracy of 77.8% in classifying the ACA. This approach not only improves accuracy but also reduces the time and expertise required compared to manual assessments. These results underscore AI's potential to streamline and enhance the accuracy of glaucoma diagnostics in clinical settings.

DL has demonstrated high proficiency in the automatic detection of angle closure and quantitative analysis of angle parameters in UBM images [132–134]. Using advanced CNNs, the AI system achieved an impressive 98.18% accuracy in classifying the ACA. It also reliably measured various angle parameters, such as the trabecular-iris angle and angle-opening distance, with precision comparable to that of manual methods. This capability indicates that AI can significantly improve the diagnosis and management of primary angle-closure glaucoma by providing accurate and reproducible assessments.

AI in Goniophotos

Goniophotos are crucial for evaluating the angle structures of the eye, which is significant for the diagnosis and management of ocular conditions such as glaucoma [135, 136]. However, the traditional manual interpretation of goniophotos

has limitations [136]. It is a subjective process that can be time-consuming and exhibits considerable variability among clinicians [73]. These factors underscore the need for faster and more convenient assessment methods, highlighting the relevance of integrating AI into goniophoto analysis.

AI in Angle Closure Classification from Goniophotos

Given AI's exceptional capabilities in image feature extraction, especially through DL techniques, its application in this context is highly significant [81, 101]. By leveraging AI, a more efficient and reliable method for angle evaluation can be achieved, streamlining the diagnostic process and enhancing clinical decision-making in ophthalmology. AI-based systems now demonstrate the ability to differentiate between open angle and angle-closure glaucoma in goniophotos with high specificity and sensitivity, achieving rates up to 92.6% and 97.8%, respectively [137]. Novel software developed for assessing angle closure on EyeCam goniophotographs shows high diagnostic consistency with manual methods, achieving a kappa coefficient as high as 0.88 [138]. Recent implementations of CNNs, particularly using the ResNet-50 architecture, have led to superior performance in determining ACA closure, with agreement rates surpassing those of experienced glaucoma experts [139]. These AI systems exhibit high area under the receiver operating characteristic curve (AUC) values, enhancing the capability for remote detection of primary angle-closure glaucoma. Overall, these AI technologies are transforming the approach to diagnosing glaucoma, offering improved diagnostic accuracy and the potential for remote assessment, crucial for early detection and management of the disease.

Automatic Grading of the Pigmentation Distribution of the ACA

The evaluation of ACA pigmentation is clinically important, offering insights into the distinction between pigmentary and non-pigmentary glaucoma, and aiding in the differential diagnosis of secondary glaucoma etiologies [140]. This

section introduces the significance of automatic grading in evaluating the pigmentation distribution within the ACA. Recent developments include a prototype gonioscopy device that captures comprehensive 360° images, enabling automatic and quantitative assessments of angle structures [141]. This automation reduces reliance on subjective interpretations and increases consistency in evaluations. Further advances in AI have enabled accurate classification of pigmentation distribution within the iridocorneal angle [142]. Automated systems have shown substantial agreement with manual grading methods, confirming both their accuracy and reliability. These systems also provide detailed insights into variations in pigmentation across different sectors and quadrants, crucial for thorough analysis.

Semantic Segmentation of the Anatomical Structure of the ACA

AI has significantly advanced the semantic segmentation of anterior chamber structures, achieving notable accuracy in pixel classification [143]. DL-based systems initially demonstrated an average accuracy of approximately 88%, which was enhanced in subsequent developments to about 91%. These improvements were validated through rigorous fivefold cross-validation trials, showing significant enhancements in the segmentation of clinically relevant structures like the ciliary body band and scleral spur, with accuracy improvements of 2.4% and 5.6% respectively [144]. These advancements are crucial for precise identification and delineation of ocular structures, essential for diagnosing various eye conditions. Moreover, the precise segmentation capabilities of AI systems pave the way for their integration into AI-assisted surgical navigation systems, enhancing surgical accuracy and outcomes and marking a significant step forward in both diagnostic and therapeutic applications in ophthalmology.

Real-Time TM Identification

Real-time identification of the TM is immensely significant as it serves as the foundation for future intelligent surgical navigation under

gonioscopy. Accurate and timely localization of the TM during surgery is crucial for minimizing complications and improving outcomes [17, 145]. Advanced technologies such as CNNs and real-time image analysis enable precise and automated localization of TM [146]. The model's performance was evaluated through cross-validation, a separate test set, and intraoperative gonioscopic videos. Results demonstrated remarkable accuracy, with median deviations of 0.8% and worst-case deviations of 4% from expert annotations. When challenged with unseen images, the CNN model outperformed general ophthalmologists by more than two standard deviations. These real-time capable CNNs show great potential for applications in surgical training, automated screenings, and intraoperative guidance, contributing to advancements in clinical practice. Additionally, this study established the first publicly available gonioscopy image bank, fostering future investigations in this field.

Future Prospects

AI in Angle-Closure Grading

Angle-closure grading, essential in assessing the severity of angle closure, plays a crucial role in glaucoma management [52, 66, 147]. It determines the most suitable surgical approach, whether peripheral iridotomy or filtration surgeries such as trabeculectomy [147–149]. Accurate angle grading directly influences treatment decisions and patient outcomes [148, 150].

Current AI applications primarily focus on distinguishing between open and closed angles [92, 107, 108]. However, evaluating the severity of angle closure, or angle grading, is equally vital. As imaging and AI technologies advance, there is significant potential to develop precise AI models tailored for angle-closure grading. These models can analyze detailed angle parameters and morphological characteristics, utilizing DL and advanced image processing to accurately assess the degree of angle closure. This capability allows clinicians to make better-informed

treatment decisions. Furthermore, integrating AI with technologies like AS-OCT could enhance the accuracy of angle-closure grading, as AS-OCT provides detailed cross-sectional images of the anterior segment for improved visualization and measurement.

Surgical Navigation System in MIGS

MIGS distinguishes itself from traditional methods with its precision in lowering IOP, micro-incision technique, minimal trauma, fewer complications, shorter learning curve, faster patient recovery, and outstanding efficacy and safety [10, 17, 20, 24, 145]. In MIGS, goniotomy represents a typical procedure [11, 13, 14, 15, 18, 19, 21, 23, 25–28, 151, 152]. The main challenge of such surgeries lies in accurately identifying the TM intraoperatively and making precise incisions in the correct locations [20]. Advanced surgical navigation systems, enhanced by AI, can significantly improve these operations [153]. These systems utilize real-time localization of the TM and semantic segmentation of the eye's anatomy, which are crucial for precise interventions in MIGS.

In the realm of MIGS, the introduction of the DL-based Goniotomy Navigation System (GNS) represents a significant advancement in intelligent surgical navigation [105]. Developed with comprehensive training from over 5 million frames sourced from 186 anonymized goniotomy procedure videos across three medical institutions, the GNS expertly assists surgeons by identifying key surgical phases and tracking essential instruments and anatomical structures. This system not only delineates each of the eight distinct surgical phases with high precision (99.0% in validation and 97.3% accuracy in external testing) but also accurately tracks ten different surgical instruments and anatomical features with precision rates of 90.9% and 89.7% in validation and external tests, respectively. The GNS enhances surgical precision through real-time, action-oriented audio alerts and visual aids displayed on a third-party screen, closely aligning with the expert

panel's annotations (intraclass correlation coefficient=0.79). A survey among 12 glaucoma surgeons revealed that 83.3% found the GNS extremely beneficial for real-time guidance during goniotomy, highlighting its potential to improve decision-making and minimize risks associated with surgical errors.

This example underscores the transformative impact of integrating AI-driven tools in MIGS, enhancing both the safety and efficacy of glaucoma surgery. However, the current GNS system is trained solely for open-angle glaucoma. In the future, with further breakthroughs in AI technology and its integration with anterior segment imaging techniques, surgical navigation systems designed for angle-closure glaucoma are expected to emerge. These systems will be tailored to the more complex clinical scenarios associated with angle-closure glaucoma, enhancing surgical precision and outcomes.

Large Language Models (LLMs)

LLMs have made substantial progress in ophthalmology, aiding in image interpretation, answering medical queries, and providing clinical decision support [35, 154]. These models use sophisticated natural language processing techniques and DL algorithms to produce human-like text, assisting clinicians in various aspects of patient care [35, 155, 156]. LLMs can analyze extensive data sets, including clinical records and patient outcomes, to provide evidence-based recommendations and support personalized treatment planning [35, 154, 155]. LLMs can also generate comprehensive reports and summaries, enhancing communication between healthcare providers and patients [35, 155, 157].

LLMs are making initial advances in the analysis and interpretation of retinal angiography imaging, including techniques like fundus images, fundus fluorescein angiography (FFA) and indocyanine green angiography (ICGA) [158–160]. DeepDR-LLM system, an integrated image-language system that combines a LLM module with image-based DL

(DeepDR-Transformer), offers personalized, high-quality, and empathetic management suggestions for patients with diabetes, utilizing their retinal images and routine clinical data. This comprehensive digital solution enhances individualized diabetes care and is particularly valuable in low-resource, high-volume settings [161]. Given that gonioscopy, UBM, and AS-OCT all come with textual description reports, LLMs can enhance physicians and patients' understanding of anterior segment images through text, holding the potential to achieve more precise detection of angle closure in glaucoma. This integration of textual information with multimodal imaging data allows for the differentiation between types of angle closure such as appositional and synechial, as well as distinguishing primary from secondary angle closures [162]. Additionally, these systems can assess the extent of angle closure and synechiae, and by synthesizing patient data, provide appropriate treatment decisions. As these models continue to evolve, their applications in the field are expected to grow, revolutionizing patient care and advancing our understanding of angle closure. LLMs hold tremendous potential for future applications [163]: For clinicians, LLMs can offer capabilities for automated medical documentation and, with enhanced inputs and thorough validation, they could eventually diagnose and treat conditions autonomously. For patient care, LLMs can provide valuable triage recommendations, succinct summaries of medical documents, clear explanations of patient conditions, and personalized educational materials adapted to individual comprehension levels [133].

References

1. Patel K, Patel S. Angle-closure glaucoma. *Dis Mon*. 2014;60(6):254–62.
2. Weinreb RN, Aung T, Medeiros FA. The pathophysiology and treatment of glaucoma: a review. *JAMA*. 2014;311(18):1901–11.
3. Foster PJ, Buhrmann R, Quigley HA, Johnson GJ. The definition and classification of glaucoma in prevalence surveys. *Br J Ophthalmol*. 2002;86(2):238–42.
4. Chan EW, Li X, Tham YC, Liao J, Wong TY, Aung T, et al. Glaucoma in Asia: regional prevalence variations and future projections. *Br J Ophthalmol*. 2016;100(1):78–85.
5. Tham YC, Li X, Wong TY, Quigley HA, Aung T, Cheng CY. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology*. 2014;121(11):2081–90.
6. Friedman DS, Foster PJ, Aung T, He M. Angle closure and angle-closure glaucoma: what we are doing now and what we will be doing in the future. *Clin Exp Ophthalmol*. 2012;40(4):381–7.
7. Gedde SJ, Chen PP, Muir KW, Vinod K, Lind JT, Wright MM, et al. Primary angle-closure disease preferred practice pattern®. *Ophthalmology*. 2021;128(1):P30–p70.
8. Sun X, Dai Y, Chen Y, Yu DY, Cringle SJ, Chen J, et al. Primary angle closure glaucoma: what we know and what we don't know. *Prog Retin Eye Res*. 2017;57:26–45.
9. Azuara-Blanco A, Burr J, Ramsay C, Cooper D, Foster PJ, Friedman DS, et al. Effectiveness of early lens extraction for the treatment of primary angle-closure glaucoma (EAGLE): a randomised controlled trial. *Lancet*. 2016;388(10052):1389–97.
10. Chan PPM, Larson MD, Dickerson JE Jr, Mercieca K, Koh VTC, Lim R, et al. Minimally invasive glaucoma surgery: latest developments and future challenges. *Asia Pac J Ophthalmol (Phila)*. 2023;12(6):537–64.
11. Gao X, Lin F, Lu P, Xie L, Tang L, Zhu X, et al. Efficacy and safety of surgical peripheral iridectomy, goniosynechialysis, and goniotomy for advanced primary angle-closure glaucoma without cataract: one-year results of a multicenter study. *J Glaucoma*. 2024.
12. Gao X, Lv A, Lin F, Lu P, Zhang Y, Song W, et al. Efficacy and safety of trabeculectomy versus peripheral iridectomy plus goniotomy in advanced primary angle-closure glaucoma: study protocol for a multicentre, non-inferiority, randomised controlled trial (the TVG study). *BMJ Open*. 2022;12(7):e062441.
13. Lin F, Nie X, Shi J, Song Y, Lv A, Li X, et al. Safety and efficacy of goniotomy following failed surgery for glaucoma. *J Glaucoma*. 2023;32(11):942–7.
14. Lin F, Zhang Y, Zhu X, Yu P, Fan S, Lv A, et al. Impact of peripheral anterior synechiae on the outcome of combined phacoemulsification, goniosynechialysis and goniotomy for primary angle-closure glaucoma and cataract: a multicenter observational study. *J Glaucoma*. 2024.
15. Song Y, Lin F, Lv A, Zhang Y, Lu L, Xie L, et al. Phacogoniotomy versus phacotrabeculectomy for advanced primary angle-closure glaucoma with cataract: a randomized non-inferiority trial. *Asia Pac J Ophthalmol (Phila)*. 2024;13(1):100033.

16. Song Y, Song W, Zhang Y, Zhang H, Xiao M, Zhao X, et al. Efficacy and safety of phacotrabeculectomy versus phacogoniectomy in advanced primary angle-closure glaucoma: study protocol for a multicentre non-inferiority randomised controlled trial (PVP Study). *BMJ Open*. 2021;11(12):e056876.
17. Song Y, Zhang H, Zhang Y, Tang G, Wan KH, Lee JWY, et al. Minimally invasive glaucoma surgery in primary angle-closure glaucoma. *Asia Pac J Ophthalmol (Phila)*. 2022;11(5):460–9.
18. Song Y, Zhang Y, Li F, Zhang Y, Lin F, Lv A, et al. One-year results of a multicenter study: intraocular pressure-lowering effect of combined phacoemulsification, goniosynechialysis, and goniotomy for cases of advanced primary angle-closure glaucoma with cataract. *Asia Pac J Ophthalmol (Phila)*. 2022;11(6):529–35.
19. Song Y, Zhu X, Zhang Y, Shu J, Dang G, Zhou W, et al. Outcomes of partial versus complete goniotomy with or without phacoemulsification for primary open angle glaucoma: a multicenter study. *J Glaucoma*. 2023;32(7):563–8.
20. Zhang X, Lin F, Li F, Lee JWY, Tham CC. Minimally invasive glaucoma surgery: a new era in glaucoma treatment. *Asia Pac J Ophthalmol (Phila)*. 2023;12(6):509–11.
21. Zhang Y, Chen W, Lin TPH, Zhang X, Lam DSC, Chen W. Outcomes of goniotomy with or without secondary intraocular lens implantation in pediatric glaucoma following cataract surgery: a prospective pilot study. *Asia Pac J Ophthalmol (Phila)*. 2023;12(5):444–50.
22. Zhang Y, Song Y, Zhou Y, Bai B, Zhang X, Chen W. A Comprehensive review of pediatric glaucoma following cataract surgery and progress in treatment. *Asia Pac J Ophthalmol (Phila)*. 2023;12(1):94–102.
23. Zhang Y, Yu P, Zhang Y, Sugihara K, Zhu X, Zhang Y, et al. Influence of goniotomy size on treatment safety and efficacy for primary open-angle glaucoma: a multicenter study. *Am J Ophthalmol*. 2023;256:118–25.
24. Fang Z, Song Y, Jin L, Han Y, Zhang X. Phacoemulsification combined with trabecular meshwork-Schlemm canal-based minimally invasive glaucoma surgery in primary angle-closure glaucoma: a systematic review and meta-analysis. *BMC Ophthalmol*. 2025 Apr 3;25(1):168.
25. Song Y, Xie L, Zhu X, Fan S, Lv A, Tang G, et al. Two-year outcome of phacogoniectomy for advanced primary angle closure glaucoma with cataracts: a multicentre study. *Br J Ophthalmol*. 2025 Mar 6;bjo-2024-325375.
26. Lin F, Li L, Lu P, Tang L, Zhang Y, Xie L, et al. Two-year outcomes of combined surgical peripheral iridectomy, goniosynechialysis, and goniotomy for advanced primary angle-closure glaucoma without cataract: a multicenter study. *Asia Pac J Ophthalmol (Phila)*. 2025 Feb 26;100166.
27. Song Y, Fan S, Tang L, Lin F, Li F, Lv A, et al. Two-year outcomes of phacogoniectomy versus phacotrabeculectomy study: a non-Inferiority randomized clinical trial. *JAMA ophthalmol* 2025;in press.
28. Lin F, Lv A, Li F, Song Y, Xie L, Zhu X, et al. Peripheral Iridectomy with goniosynechialysis and goniotomy vs. trabeculectomy for advanced PACG: a randomized controlled trial. *JAMA ophthalmol* 2025;in press.
29. Song Y, Tang L, Fan F, Zhu X, Wen T, Tang G, et al. Three-year outcome of phacogoniectomy for advanced primary angle closure glaucoma with cataracts: a multicenter study. *J Glaucoma* 2025;in revision.
30. Vergroesen JE, Schuster AK, Stuart KV, Asefa NG, Coughard-Grégoire A, Delcourt C, et al. Association of systemic medication use with glaucoma and intraocular pressure: the European eye epidemiology consortium. *Ophthalmology*. 2023;130(9):893–906.
31. Konstas AG, Kahook MY, Araie M, Katsanos A, Quaranta L, Rossetti L, et al. Diurnal and 24-h intraocular pressures in glaucoma: monitoring strategies and impact on prognosis and treatment. *Adv Ther*. 2018;35(11):1775–804.
32. Verma S, Nongpiur ME, Atalay E, Wei X, Husain R, Goh D, et al. Visual field progression in patients with primary angle-closure glaucoma using pointwise linear regression analysis. *Ophthalmology*. 2017;124(7):1065–71.
33. Soh ZD, Tan M, Nongpiur ME, Xu BY, Friedman D, Zhang X, et al. Assessment of angle closure disease in the age of artificial intelligence: a review. *Prog Retin Eye Res*. 2024 Jan;98:101227.
34. Devalla SK, Liang Z, Pham TH, Boote C, Strouthidis NG, Thiery AH, et al. Glaucoma management in the era of artificial intelligence. *Br J Ophthalmol*. 2020;104(3):301–11.
35. Betzler BK, Chen H, Cheng CY, Lee CS, Ning G, Song SJ, et al. Large language models and their impact in ophthalmology. *Lancet Digit Health*. 2023;5(12):e917–24.
36. Friedman DS, He M. Anterior chamber angle assessment techniques. *Surv Ophthalmol*. 2008;53(3):250–73.
37. Raluca M, Mircea F, Andrei F, Carmen D, Miruna N, Grigorios T, et al. Old and new in exploring the anterior chamber angle. *Romanian Journal of Ophthalmology*. 2015;59(4):208–16.
38. Pavlin CJ, Harasiewicz K, Sherar MD, Foster FS. Clinical use of ultrasound biomicroscopy. *Ophthalmology*. 1991;98(3):287–95.
39. Barkana Y, Dorairaj SK, Gerber Y, Liebmann JM, Ritch R. Agreement between gonioscopy and ultrasound biomicroscopy in detecting iridotrabecular apposition. *Arch Ophthalmol (Chicago, Ill: 1960)*. 2007;125(10):1331–5.
40. Kong X, Foster PJ, Huang Q, Zheng Y, Huang W, Cai X, et al. Appositional closure identified by ultrasound biomicroscopy in population-based primary angle-closure glaucoma suspects:

- the Liwan eye study. *Invest Ophthalmol Vis Sci*. 2011;52(7):3970–5.
41. Ishikawa H, Schuman JS. Anterior segment imaging: ultrasound biomicroscopy. *Ophthalmol Clin North Am*. 2004;17(1):7–20.
 42. Gao K, Li F, Li Y, Li X, Huang W, Chen S, et al. Anterior choroidal thickness increased in primary open-angle glaucoma and primary angle-closure disease eyes evidenced by ultrasound biomicroscopy and SS-OCT. *Invest Ophthalmol Vis Sci*. 2018;59(3):1270–7.
 43. He M, Foster PJ, Johnson GJ, Khaw PT. Angle-closure glaucoma in East Asian and European people. Different diseases? *Eye (London, England)*. 2006;20(1):3–12.
 44. Zhu Y, Fang L, Zhong Y, Oatts J, Han Y, Lin S, et al. Clinical and ultrasound biomicroscopic characteristics of congenital fibrovascular pupillary membrane-induced secondary glaucoma. *Front Med*. 2021;8:763137.
 45. El Sanharawi M, Sandali O, Naudet F. Ultrasound biomicroscopy of the ciliary body in ocular/oculodermal melanocytosis. *Am J Ophthalmol*. 2013;156(3):627–8.
 46. Elfalah M, Mohammad M, Toro MD, Abu-Yaghi N, Rejdak R, Yousef YA. Anterior ocular biometrics as measured by ultrasound biomicroscopy. *Healthcare (Basel, Switzerland)*. 2022;10(7).
 47. de Leon JMS, Tun TA, Perera SA, Aung T. Angle closure imaging: a review. *Current Ophthalmology Reports*. 2013;1(2):80–8.
 48. Shabana N, Aquino MC, See J, Ce Z, Tan AM, Nolan WP, et al. Quantitative evaluation of anterior chamber parameters using anterior segment optical coherence tomography in primary angle closure mechanisms. *Clin Exp Ophthalmol*. 2012;40(8):792–801.
 49. Sng CC, Foo LL, Cheng CY, Allen JC Jr, He M, Krishnaswamy G, et al. Determinants of anterior chamber depth: the Singapore Chinese Eye Study. *Ophthalmology*. 2012;119(6):1143–50.
 50. Liu P, Higashita R, Guo PY, Okamoto K, Li F, Nguyen A, et al. Reproducibility of deep learning based scleral spur localisation and anterior chamber angle measurements from anterior segment optical coherence tomography images. *Br J Ophthalmol*. 2023;107(6):802–8.
 51. Ang M, Baskaran M, Werkmeister RM, Chua J, Schmidl D, Aranha Dos Santos V, et al. Anterior segment optical coherence tomography. *Prog Retin Eye Res*. 2018;66:132–56.
 52. Riva I, Micheletti E, Oddone F, Bruttini C, Montescani S, De Angelis G, et al. Anterior chamber angle assessment techniques: a review. *J Clin Med*. 2020;9(12).
 53. Li X, Wang W, Chen S, Huang W, Liu Y, Wang J, et al. Effects of Valsalva maneuver on anterior chamber parameters and choroidal thickness in healthy Chinese: an AS-OCT and SS-OCT study. *Invest Ophthalmol Vis Sci*. 2016;57(9):189–95.
 54. Lee JR, Sung KR, Han S. Comparison of anterior segment parameters between the acute primary angle closure eye and the fellow eye. *Invest Ophthalmol Vis Sci*. 2014;55(6):3646–50.
 55. Li F, Zhou R, Gao K, Jin L, Zhang X. Volumetric parameters-based differentiation of narrow angle from open angle and classification of angle configurations: an SS-OCT study. *Br J Ophthalmol*. 2020;104(1):92–7.
 56. Huang W, Gao X, Li X, Wang J, Chen S, Wang W, et al. Anterior and posterior ocular biometry in healthy Chinese subjects: data based on AS-OCT and SS-OCT. *PLoS ONE*. 2015;10(3):e0121740.
 57. Wang BS, Narayanaswamy A, Amerasinghe N, Zheng C, He M, Chan YH, et al. Increased iris thickness and association with primary angle closure glaucoma. *Br J Ophthalmol*. 2011;95(1):46–50.
 58. Wang B, Sakata LM, Friedman DS, Chan YH, He M, Lavanya R, et al. Quantitative iris parameters and association with narrow angles. *Ophthalmology*. 2010;117(1):11–7.
 59. Huang W, Li X, Gao X, Zhang X. The anterior and posterior biometric characteristics in primary angle-closure disease: data based on anterior segment optical coherence tomography and swept-source optical coherence tomography. *Indian J Ophthalmol*. 2021;69(4):865–70.
 60. Wang W, Li X, Chen S, Huang W, Zhang X. Biometric differences between unilateral chronic primary angle closure glaucoma and fellow non-glaucomatous eyes. *Semin Ophthalmol*. 2018;33(5):595–601.
 61. Li X, Wang W, Huang W, Chen S, Wang J, Wang Z, et al. Difference of uveal parameters between the acute primary angle closure eyes and the fellow eyes. *Eye (Lond)*. 2018;32(7):1174–82.
 62. Zhou R, Li F, Gao K, Zhang X. Smaller anterior chamber volume is associated with higher risk of intraocular pressure elevation after laser peripheral iridotomy: a 1-year follow-up study. *Asia Pac J Ophthalmol (Phila)*. 2020;10(2):188–91.
 63. Gao K, Li F, Aung T, Zhang X. Diurnal variations in the morphology of Schlemm's canal and intraocular pressure in healthy Chinese: an SS-OCT study. *Invest Ophthalmol Vis Sci*. 2017;58(13):5777–82.
 64. Yuan Y, Lin TPH, Gao K, Zhou R, Radke NV, Lam DSC, et al. Aerobic exercise reduces intraocular pressure and expands Schlemm's canal dimensions in healthy and primary open-angle glaucoma eyes. *Indian J Ophthalmol*. 2021;69(5):1127–34.
 65. Gao K, Song S, Johnstone MA, Zhang Q, Xu J, Zhang X, et al. Reduced pulsatile trabecular meshwork motion in eyes with primary open angle glaucoma using phase-sensitive optical coherence tomography. *Invest Ophthalmol Vis Sci*. 2020;61(14):21.

66. Zhang X, Guo PY, Lin C, Li F, Nguyen A, Sakata R, et al. Assessment of iris trabecular contact in eyes with gonioscopic angle-closure. *Ophthalmology*. 2023;130(1):111–9.
67. Hertzog LH, Albrecht KG, LaBree L, Lee PP. Glaucoma care and conformance with preferred practice patterns. Examination of the private, community-based ophthalmologist. *Ophthalmology*. 1996;103(7):1009–13.
68. Quigley HA, Friedman DS, Hahn SR. Evaluation of practice patterns for the care of open-angle glaucoma compared with claims data: the glaucoma adherence and persistency study. *Ophthalmology*. 2007;114(9):1599–606.
69. Bell NP, Nagi KS, Cumba RJ, Chuang AZ, Lee DA, Prager TC, et al. Age and positional effect on the anterior chamber angle: assessment by ultrasound biomicroscopy. *ISRN Ophthalmol*. 2013;2013:706201.
70. Tello C, Liebmann J, Potash SD, Cohen H, Ritch R. Measurement of ultrasound biomicroscopy images: intraobserver and interobserver reliability. *Invest Ophthalmol Vis Sci*. 1994;35(9):3549–52.
71. Balidis MO, Bunce C, Boboridis K, Salzman J, Wormald RP, Miller MH. Intraobserver and interobserver reliability of the R/D score for evaluation of iris configuration by ultrasound biomicroscopy, in patients with pigment dispersion syndrome. *Eye (Lond)*. 2002;16(6):722–6.
72. Zhang Q, Jin W, Wang Q. Repeatability, reproducibility, and agreement of central anterior chamber depth measurements in pseudophakic and phakic eyes: optical coherence tomography versus ultrasound biomicroscopy. *J Cataract Refract Surg*. 2010;36(6):941–6.
73. Sakata LM, Lavanya R, Friedman DS, Aung HT, Gao H, Kumar RS, et al. Comparison of gonioscopy and anterior segment ocular coherence tomography in detecting angle closure in different quadrants of the anterior chamber angle. *Ophthalmology*. 2008;115(5):769–74.
74. Sakata LM, Lavanya R, Friedman DS, Aung HT, Seah SK, Foster PJ, et al. Assessment of the scleral spur in anterior segment optical coherence tomography images. *Arch Ophthalmol (Chicago, Ill: 1960)*. 2008;126(2):181–5.
75. Kim DY, Sung KR, Kang SY, Cho JW, Lee KS, Park SB, et al. Characteristics and reproducibility of anterior chamber angle assessment by anterior-segment optical coherence tomography. *Acta Ophthalmol*. 2011;89(5):435–41.
76. McCarthy J, Minsky ML, Rochester N, Shannon CE. A Proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. *AI Mag*. 2006;27(4):12.
77. Samuel AL. Some studies in machine learning using the game of checkers. *IBM J Res Dev*. 1959;3(3):210–29.
78. Lynch CJ, Liston C. New machine-learning technologies for computer-aided diagnosis. *Nat Med*. 2018;24(9):1304–5.
79. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436–44.
80. Singh A, Jothi Balaji J, Rasheed MA, Jayakumar V, Raman R, Lakshminarayanan V. Evaluation of explainable deep learning methods for ophthalmic diagnosis. *Clin Ophthalmol*. 2021;15:2573–81.
81. Ting DSW, Pasquale LR, Peng L, Campbell JP, Lee AY, Raman R, et al. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol*. 2019;103(2):167–75.
82. Wan KH, Jonas JB. Impact of digital technology on eye diseases during COVID-19. *Asia Pac J Ophthalmol (Phila)*. 2022;11(5):401–2.
83. Ng WY, Zhang S, Wang Z, Ong CJT, Gunasekaran DV, Lim GYS, et al. Updates in deep learning research in ophthalmology. *Clin Sci (Lond)*. 2021;135(20):2357–76.
84. Qamar T, Bawany NZ. Understanding the black-box: towards interpretable and reliable deep learning models. *PeerJ Comput Sci*. 2023;9:e1629.
85. Wang D, Zhang Y, Li Y, Zhou K, Yang Z, Zeng Y, et al. A deep learning-based intraoperative navigation system for minimally invasive glaucoma surgery. *JAMA Ophthalmol*. 2025;in revision.
86. Li F, Wang D, Yang Z, Zhang Y, Jiang J, Liu X, et al. The AI revolution in glaucoma: bridging challenges with opportunities. *Prog Retin Eye Res*. 2024;103:101291.
87. Li F, Song D, Chen H, Xiong J, Li X, Zhong H, et al. Development and clinical deployment of a smartphone-based visual field deep learning system for glaucoma detection. *NPJ Digit Med*. 2020;3:123.
88. Xiong J, Li F, Song D, Tang G, He J, Gao K, et al. Multimodal machine learning using visual fields and peripapillary circular OCT scans in detection of glaucomatous optic neuropathy. *Ophthalmology*. 2022;129(2):171–80.
89. Li F, Su Y, Lin F, Li Z, Song Y, Nie S, et al. A deep-learning system predicts glaucoma incidence and progression using retinal photographs. *J Clin Invest*. 2022;132(11).
90. Fu H, Li F, Sun X, Cao X, Liao J, Orlando JJ, et al. AGE challenge: angle closure glaucoma evaluation in anterior segment optical coherence tomography. *Med Image Anal*. 2020;66:101798.
91. Hao J, Li F, Hao H, Fu H, Xu Y, Higashita R, et al. Hybrid variation-aware network for angle-closure assessment in AS-OCT. *IEEE Trans Med Imaging*. 2022;41(2):254–65.
92. Li F, Yang Y, Sun X, Qiu Z, Zhang S, Tun TA, et al. Digital gonioscopy based on three-dimensional anterior-segment OCT: an international multicenter study. *Ophthalmology*. 2022;129(1):45–53.
93. Olyntho MAC Jr, Jorge CAC, Castanha EB, Gonçalves AN, Silva BL, Nogueira BV, et al.

- Artificial intelligence in anterior chamber evaluation: a systematic review and meta-analysis. *J Glaucoma*. 2024 Sep 1;33(9):658–664.
94. Pinto-Coelho L. How artificial intelligence is shaping medical imaging technology: a survey of innovations and applications. *Bioengineering (Basel)*. 2023;10(12).
 95. Bekbolatova M, Mayer J, Ong CW, Toma M. Transformative potential of AI in healthcare: definitions, applications, and navigating the ethical landscape and public perspectives. *Healthcare (Basel, Switzerland)*. 2024;12(2).
 96. Qian CX, Hassanaly S, Harissi-Dagher M. Anterior segment optical coherence tomography in the long-term follow-up and detection of glaucoma in Boston type I keratoprosthesis. *Ophthalmology*. 2015;122(2):317–25.
 97. Sng CCA, Aquino MCD, Liao J, Ang M, Zheng C, Loon SC, et al. Pretreatment anterior segment imaging during acute primary angle closure: insights into angle closure mechanisms in the acute phase. *Ophthalmology*. 2014;121(1):119–25.
 98. Salim S. The role of anterior segment optical coherence tomography in glaucoma. *J Ophthalmol*. 2012;2012:476801.
 99. Lin AN, Mohammed ISK, Munir WM, Munir SZ, Alexander JL. Inter-rater reliability and repeatability of manual anterior segment optical coherence tomography image grading in keratoconus. *Eye Contact Lens*. 2021;47(9):494–9.
 100. Wanichwecharungruang B, Pattanapongpaiboon W, Kongsomboon K, Parivisutt N, Annopawong K, Seresirikachorn K. Diagnostic performance of anterior segment optical coherence tomography in detecting plateau iris. *BMJ Open Ophthalmol*. 2022;7(1):e000931.
 101. Srivastava O, Tennant M, Grewal P, Rubin U, Seamone M. Artificial intelligence and machine learning in ophthalmology: a review. *Indian J Ophthalmol*. 2023;71(1):11–7.
 102. Mou D, Wang J, Wang Y, Tang X, Dong Z, Wang N, et al. Performance of anterior segment OCT-based algorithms in the opportunistic screening for primary angle-closure disease. *Heliyon*. 2024 Mar 30;10(7):e28885.
 103. Shan J, Li Z, Ma P, Tun TA, Yonamine S, Wu Y, et al. Deep learning classification of angle closure based on anterior segment OCT. *Ophthalmol Glaucoma*. 2024 Jan-Feb;7(1):8–15.
 104. Nongpiur ME, Haaland BA, Friedman DS, Perera SA, He M, Foo LL, et al. Classification algorithms based on anterior segment optical coherence tomography measurements for detection of angle closure. *Ophthalmology*. 2013;120(1):48–54.
 105. Li F, Song D, Wang D, Zhou X, Xing L, Jia H, et al. Predicting visual field pattern deviation probability maps from volumetric OCT scans. *NPD Digit Med*. 2025;in revision.
 106. Fu H, Baskaran M, Xu Y, Lin S, Wong DWK, Liu J, et al. A deep learning system for automated angle-closure detection in anterior segment optical coherence tomography images. *Am J Ophthalmol*. 2019;203:37–45.
 107. Hao H, Zhao Y, Fu H, Shang Q, Li F, Zhang X, et al. Anterior chamber angles classification in anterior segment OCT images via multi-scale regions convolutional neural networks. *Annu Int Conf IEEE Eng Med Biol Soc*. 2019;2019:849–52.
 108. Hao H, Zhao Y, Yan Q, Higashita R, Zhang J, Zhao Y, et al. Angle-closure assessment in anterior segment OCT images via deep learning. *Med Image Anal*. 2021;69:101956.
 109. Li W, Chen Q, Jiang C, Shi G, Deng G, Sun X. Automatic anterior chamber angle classification using deep learning system and anterior segment optical coherence tomography images. *Transl Vis Sci Technol*. 2021;10(6):19.
 110. Yao H, Wang X, Suo Y, He J, Chu C, Yang Z, et al. Primary angle-closed diseases recognition through artificial intelligence-based anterior segment-optical coherence tomography imaging. *Graefes Arch Clin Exp Ophthalmol*. 2024 Dec 16.
 111. Tabatabaei SM, Fakhraie G, Ansari S, Hamzeh N, Safizadeh M, Beikmarzehei A. Plateau iris: a review. *J Curr Ophthalmol*. 2023;35(1):11–6.
 112. Moghimi S, Torkashvand A, Mohammadi M, Yaseri M, Saunders LJ, Lin SC, et al. Classification of primary angle closure spectrum with hierarchical cluster analysis. *PLoS ONE*. 2018;13(7):e0199157.
 113. Gao X, Zhou Y, Zuo C, Chen L, Ren J, Lin H, et al. Predictive equation for angle opening distance at 750 μm after laser peripheral iridotomy in primary angle closure suspects. *Front Med*. 2021;8.
 114. Kwon J, Sung KR, Han S, Moon YJ, Shin JW. Subclassification of primary angle closure using anterior segment optical coherence tomography and ultrasound biomicroscopic parameters. *Ophthalmology*. 2017;124(7):1039–47.
 115. Zhang Y, Zhang X, Zhang Q, Lv B, Hu M, Lv C, et al. Automated classification of angle-closure mechanisms based on anterior segment optical coherence tomography images via deep learning. *Heliyon*. 2024 Jul 26;10(15):e35236.
 116. Bai X, Niwas SI, Lin W, Ju BF, Kwok CK, Wang L, et al. Learning ECOC code matrix for multiclass classification with application to glaucoma diagnosis. *J Med Syst*. 2016;40(4):78.
 117. Niwas SI, Lin W, Bai X, Kwok CK, Sng CC, Aquino MC, et al. Reliable feature selection for automated angle closure glaucoma mechanism detection. *J Med Syst*. 2015;39(3):21.
 118. Niwas SI, Lin W, Kwok CK, Kuo CC, Sng CC, Aquino MC, et al. Cross-examination for angle-closure glaucoma feature detection. *IEEE J Biomed Health Inform*. 2016;20(1):343–54.

119. Bolo K, Apolo Aroca G, Pardeshi AA, Chiang M, Burkemper B, Xie X, et al. Automated expert-level scleral spur detection and quantitative biometric analysis on the ANTERION anterior segment OCT system. *Br J Ophthalmol*. 2024 May 21;108(5):702–709.
120. Kumazu Y, Kobayashi N, Kitamura N, Rayan E, Neculoiu P, Misumi T, et al. Automated segmentation by deep learning of loose connective tissue fibers to define safe dissection planes in robot-assisted gastrectomy. *Sci Rep*. 2021;11(1):21198.
121. Kwan AC, Salto G, Cheng S, Ouyang D. Artificial intelligence in computer vision: cardiac MRI and multimodality imaging segmentation. *Curr Cardiovasc Risk Rep*. 2021;15(9).
122. Pham TH, Devalla SK, Ang A, Soh ZD, Thiery AH, Boote C, et al. Deep learning algorithms to isolate and quantify the structures of the anterior segment in optical coherence tomography images. *Br J Ophthalmol*. 2021;105(9):1231–7.
123. Xu BY, Chiang M, Pardeshi AA, Moghimi S, Varma R. Deep neural network for scleral spur detection in anterior segment OCT images: the Chinese American eye study. *Transl Vis Sci Technol*. 2020;9(2):18.
124. Choy KC, Li G, Stamer WD, Farsiu S. Open-source deep learning-based automatic segmentation of mouse Schlemm's canal in optical coherence tomography images. *Exp Eye Res*. 2022;214:108844.
125. Almazroa A, Alodhayb S, Osman E, Ramadan E, Hummadi M, Dlaim M, et al., editors. *Retinal fundus images for glaucoma analysis: the RIGA dataset. Medical Imaging 2018: Imaging Informatics for Healthcare, Research, and Applications. SPIE*; 2018.
126. Krzywicki T, Brona P, Zbrzezny AM, Grzybowski AE. A global review of publicly available datasets containing fundus images: characteristics, barriers to access, usability, and generalizability. *J Clin Med*. 2023;12(10).
127. Raja H, Akram MU, Khawaja SG, Arslan M, Ramzan A, Nazir N. Data on OCT and fundus images for the detection of glaucoma. *Data Brief*. 2020;29:105342.
128. Soh ZD, Tan M, Lee Z, Yu M, Thakur S, Lavanya R, et al. Deep learning-based normative database of anterior chamber dimensions for angle closure assessment: the Singapore Chinese eye study. *Br J Ophthalmol*. 2025 Mar 20;109(4):497–503.
129. Kottaridou E, Hatoum A. Imaging of anterior segment tumours: a comparison of ultrasound biomicroscopy versus anterior segment optical coherence tomography. *Cureus*. 2024;16(1):e52578.
130. Kumar RS, Baskaran M, Chew PT, Friedman DS, Handa S, Lavanya R, et al. Prevalence of plateau iris in primary angle closure suspects an ultrasound biomicroscopy study. *Ophthalmology*. 2008;115(3):430–4.
131. Yu J, Li W, Chen Q, Deng G, Jiang C, Liu G, et al. Automatic classification of anterior chamber angle based on ultrasound biomicroscopy images. *Ophthalmic Res*. 2021;64(5):732–9.
132. Jiang W, Yan Y, Cheng S, Wan S, Huang L, Zheng H, et al. Deep learning-based model for automatic assessment of anterior angle chamber in ultrasound biomicroscopy. *Ultrasound Med Biol*. 2023 Dec;49(12):2497–2509.
133. Li F, Zhang X, Yang K, Qin J, Lv B, Lv K, et al. Deep learning-based anterior segment identification and parameter assessment of primary angle closure disease in ultrasound biomicroscopy images. *BMJ Open Ophthalmol*. 2025 Jan 20;10(1):e001600.
134. Wang W, Wang L, Wang X, Zhou S, Lin S, Yang J. A deep learning system for automatic assessment of anterior chamber angle in ultrasound biomicroscopy images. *Transl Vis Sci Technol*. 2021;10(11):21.
135. Baskaran M, Perera S, Nongpiur M, Tun T, Park J, Sasikumar R, et al. Angle assessment by EyeCam, goniophotography, and gonioscopy. *J Glaucoma*. 2011;21:493–7.
136. Spaeth GL. Gonioscopy: uses old and new. The inheritance of occludable angles. *Ophthalmology*. 1978;85(3):222–32.
137. Cheng J, Liu J, Lee BH, Wong DW, Yin F, Aung T, et al. Closed angle glaucoma detection in RetCam images. *Annu Int Conf IEEE Eng Med Biol Soc*. 2010;2010:4096–9.
138. Baskaran M, Cheng J, Perera SA, Tun TA, Liu J, Aung T. Automated analysis of angle closure from anterior chamber angle images. *Invest Ophthalmol Vis Sci*. 2014;55(11):7669–73.
139. Chiang M, Guth D, Pardeshi AA, Randhawa J, Shen A, Shan M, et al. Glaucoma expert-level detection of angle closure in goniophotographs with convolutional neural networks: the Chinese American eye study. *Am J Ophthalmol*. 2021;226:100–7.
140. Bustamante-Arias A, Ruiz-Lozano R, Alvarez-Guzmán C, González-Godínez S, Rodríguez-García A. Pigment dispersion syndrome and its implications for glaucoma. *Surv Ophthalmol*. 2021.
141. De Giusti A, Pajaro S, Tanito M, editors. *Automatic pigmentation grading of the trabecular meshwork in gonioscopic images. In: Computational pathology and ophthalmic medical image analysis, vol. 2018. Cham: Springer International Publishing*; 2018.
142. Matsuo M, Pajaro S, De Giusti A, Tanito M. Automated anterior chamber angle pigmentation analyses using 360° gonioscopy. *Br J Ophthalmol*. 2020;104(5):636–41.
143. Peroni A, Cutolo CA, Pinto LA, Paviotti A, Campigotto M, Cobb C, et al., editors. *A deep learning approach for semantic segmentation of gonioscopic images to support glaucoma categorization. In: Medical Image Understanding and Analysis: 24th Annual Conference, MIUA 2020, Oxford, UK, 2020 July 15–17, Proceedings 24. Berlin: Springer*; 2020.

144. Peroni A, Paviotti A, Campigotto M, Abegão Pinto L, Cutolo CA, Gong J, et al. Semantic segmentation of gonio-photographs via adaptive ROI localisation and uncertainty estimation. *BMJ Open Ophthalmol.* 2021;6(1):e000898.
145. Yang SA, Ciociola EC, Mitchell W, Hall N, Lorch AC, Miller JW, et al. Effectiveness of microinvasive glaucoma surgery in the United States: intelligent research in sight registry analysis 2013–2019. *Ophthalmology.* 2023;130(3):242–55.
146. Lin KY, Urban G, Yang MC, Lee LC, Lu DW, Alward WLM, et al. Accurate identification of the trabecular meshwork under gonioscopic view in real time using deep learning. *Ophthalmol Glaucoma.* 2022;5(4):402–12.
147. Halawa OA, Zebardast N, Kolli A, Foster PJ, He M, Aung T, et al. Population-based utility of van Herick grading for angle-closure detection. *Ophthalmology.* 2021;128(12):1779–82.
148. Xu BY, Friedman DS, Foster PJ, Jiang Y, Pardeshi AA, Jiang Y, et al. Anatomic changes and predictors of angle widening after laser peripheral iridotomy: the Zhongshan angle closure prevention trial. *Ophthalmology.* 2021;128(8):1161–8.
149. Smith SD, Singh K, Lin SC, Chen PP, Chen TC, Francis BA, et al. Evaluation of the anterior chamber angle in glaucoma: a report by the American Academy of Ophthalmology. *Ophthalmology.* 2013;120(10):1985–97.
150. Baskaran M, Oen FT, Chan YH, Hoh ST, Ho CL, Kashiwagi K, et al. Comparison of the scanning peripheral anterior chamber depth analyzer and the modified van Herick grading system in the assessment of angle closure. *Ophthalmology.* 2007;114(3):501–6.
151. Yang Z, Wang Z, Fan H, Ge Q, Dang G, Yang X, et al. Outcome of goniotomy over 120°, 240°, and 360° in juvenile open-angle glaucoma: a multicenter study. *J Glaucoma.* 2025 Mar 20.
152. Song Y, Ye X, Yang M, Zhang Z, Zeng L, Fan H, et al. Evaluation of phacogoniotomy for medically-controlled or uncontrolled primary open-angle glaucoma with cataract: a multicenter study. *J Glaucoma* 2025;in revision.
153. Ali MJ, Naik MN, Kaliki S, Dave TV, Dendukuri G. Interactive navigation-guided ophthalmic plastic surgery: the techniques and utility of 3-dimensional navigation. *Can J Ophthalmol.* 2017;52(3):250–7.
154. Tan TF, Thirunavukarasu AJ, Campbell JP, Keane PA, Pasquale LR, Abramoff MD, et al. Generative artificial intelligence through ChatGPT and other large language models in ophthalmology: clinical applications and challenges. *Ophthalmol Sci.* 2023;3(4):100394.
155. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nat Med.* 2023;29(8):1930–40.
156. Salimi A, Saheb H. Large language models in ophthalmology scientific writing: ethical considerations blurred lines or not at all? *Am J Ophthalmol.* 2023;254:177–81.
157. Jin K, Yuan L, Wu H, Grzybowski A, Ye J. Exploring large language model for next generation of artificial intelligence in ophthalmology. *Front Med.* 2023;10:1291404.
158. Chen X, Xu P, Li Y, Zhang W, Song F, He M, et al. ChatFFA: an ophthalmic chat system for unified vision-language understanding and question answering for fundus fluorescein angiography. *iScience.* 2024;27(7).
159. Chen X, Zhang W, Xu P, Zhao Z, Zheng Y, Shi D, et al. FFA-GPT: an automated pipeline for fundus fluorescein angiography interpretation and question-answer. *NPJ Digit Med.* 2024;7(1):111.
160. Chen X, Zhang W, Zhao Z, Xu P, Zheng Y, Shi D, et al. ICGA-GPT: report generation and question answering for indocyanine green angiography images. *Br J Ophthalmol.* 2024.
161. Li J, Guan Z, Wang J, Cheung CY, Zheng Y, Lim L-L, et al. Integrated image-based deep learning and language models for primary diabetes care. *Nat Med.* 2024.
162. Delsoz M, Raja H, Madadi Y, Tang AA, Wirostko BM, Kahook MY, Yousefi S. The use of ChatGPT to assist in diagnosing glaucoma based on clinical case reports. *Ophthalmol Ther.* 2023 Dec;12(6):3121–3132.
163. Yang Z, Wang D, Zhou F, Song D, Zhang Y, Jiang J, et al. Understanding natural language: potential application of large language models to ophthalmology. *Asia Pac J Ophthalmol (Phila).* 2024;100085.

Automatic Retinal Imaging and Analysis: Age-Related Macular Degeneration (AMD) Within Age-Related Eye Disease Studies (AREDS)

T. Y. Alvin Liu and Neil M. Bressler

Abstract

In this chapter, we focus on a series of collaborations between clinicians from the School of Medicine and computer scientists from the Applied Physics Lab at the Johns Hopkins University. The described studies utilized deep learning (DL) and focused on analysis of color fundus photographs (CFP) of patients with age-related macular degeneration (AMD), the leading cause of central vision loss in persons over age 50 in the United States (Bressler in JAMA 291:1900–1901, 2004 [1]) and around the world. The dataset used for training and testing was derived from the Age-Related Eye Disease Studies (AREDS) (Age-Related Eye Disease Study Research G in Control Clin Trials 20:573–600, 1999 [2]), a longitudinal cohort study funded by the National Eye Institute with over 4500 participants and roughly 130,000 CFPs taken with a 30° camera. The ground truth of the deep learning systems (DLS) was based on the annotations (gradings) by trained graders at the University of Wisconsin Fundus Photograph Reading Center, which is the designated reading center for the AREDS.

Keywords

Deep learning · Age-related macular degeneration

In this chapter, we focus on a series of collaborations between clinicians from the School of Medicine and computer scientists from the Applied Physics Lab at the Johns Hopkins University. The described studies utilized deep learning (DL) and focused on analysis of color fundus photographs (CFP) of patients with age-related macular degeneration (AMD), the leading cause of central vision loss in persons over age 50 in the United States [1] and around the world. The dataset used for training and testing was derived from the Age-Related Eye Disease Studies (AREDS) [2], a longitudinal cohort study funded by the National Eye Institute with over 4500 participants and roughly 130,000 CFPs taken with a 30° camera. The ground truth of the deep learning systems (DLS) was based on the annotations (gradings) by trained graders at the University of Wisconsin Fundus Photograph Reading Center, which is the designated reading center for the AREDS.

In the first study [3], a DLS was trained with a combination of transfer learning, using the universal features taken from the fully connected layer of the pre-trained OverFeat [4]

T. Y. Alvin Liu (✉) · N. M. Bressler
Retina Division, Department of Ophthalmology
(Wilmer Eye Institute), Johns Hopkins University School
of Medicine, Baltimore, MD, USA
e-mail: tliu25@jhmi.edu

deep convolutional neural network (DCNN), and a linear support vector machine (LSVM). Only a subset of the AREDS dataset, 5664 images, was used. Three sets of classification experiments were performed: 4-class classification (no AMD vs. early AMD vs. intermediate AMD vs. advanced AMD); 3-class classification (no or early AMD vs. intermediate AMD vs. advanced AMD); 2-class classification (no or early AMD vs. intermediate or advanced AMD). The DLS's performance was compared to that of a human ophthalmologist.

Classification problem	DLS		Ophthalmologist	
	Accuracy (%)	Kappa	Accuracy (%)	Kappa
4-class	79.4	0.6962	75.8	0.6583
3-class	81.5	0.7226	85.0	0.7748
2-class	93.4	0.8482	95.2	0.8897

To achieve the same level of performance, a DLS will need more data for training to perform progressively more fine-grained classifications. Given the size of the training data was fixed in this study, a drop in accuracy and correlation with the gold standard (kappa score) was seen with progressively more fine-grained classifications as expected. The trend was observed for both the DLS and human ophthalmologist. Overall, the DLS's performance was comparable to that of a human ophthalmologist, and the DLS showed robust performance in a 2-class classification task, i.e., detecting AMD that should be referred for management from CFP. This was one of the first DL studies in ophthalmology, and demonstrated that DL was a promising technique in analyzing CFPs from patients with AMD.

The second study [5] aimed to extend the findings of the first study [3], and involved a more complicated study design. A DLS was developed for the 2-class classification task of differentiating between "no or early AMD" and "intermediate or advanced AMD", and its performance was evaluated using fivefold cross validation, in which fourfold were used for training and one fold was used for testing (with

a rotation of the folds). Different experiments were performed using different variations of the AREDS dataset: using the entire AREDS dataset including the stereo pairs (133,821 images) versus using only 1 of the stereo pair for each eye (67,401 images) versus using only 1 of the stereo pair for each eye minus images of poor quality (66,943 images). Two DL approaches were compared: using the AREDS data to retrain only the final LSVM classification stage of the OverFeat DCNN (as it was done in the first study) versus using the AREDS data to optimize all of the DCNN weights over all layers of the AlexNet DCNN [6]. The latter DL approach was more computationally advanced and intensive, and consistently produced a superior result. A human ophthalmologist also independently graded a subset of 5000 images to enable comparison with the DLS. When the entire AREDS dataset was used and the data was split at a patient visit level, the superior DL approach achieved an area-under-receiver-operating-characteristic-curve (AUC-ROC curve) value of 0.96, accuracy of $91.6\% \pm 0.1\%$ and a kappa of 0.829 ± 0.003 . When the entire AREDS dataset was used and the data was split at a patient level, the superior DL approach achieved an AUC-ROC curve value of 0.94, an accuracy of $88.7\% \pm 0.7\%$ and a kappa of 0.770 ± 0.013 . For comparison, the human ophthalmologist achieved an accuracy of 90.2% and kappa of 0.800. This study confirmed that a DLS could be trained to reliably differentiate between non-referable and referable AMD in CFPs (Fig. 13.1), with performance comparable to that of a human ophthalmologist.

The third study [7] explored technical refinements that could improve a DLS's ability in differentiating between non-referable and referable AMD in CFPs, included only one image from each stereo pair in the AREDS dataset (67,401 images in total), and utilized a different DL approach. This different approach used transfer learning, fine-tuned the original ResNet [8] DCNN weights, first performed a 4-step classification [9] (step-1: no or only small drusen ($< 63 \mu\text{m}$) and no pigmentary abnormalities; step-2: multiple small drusen or medium-sized

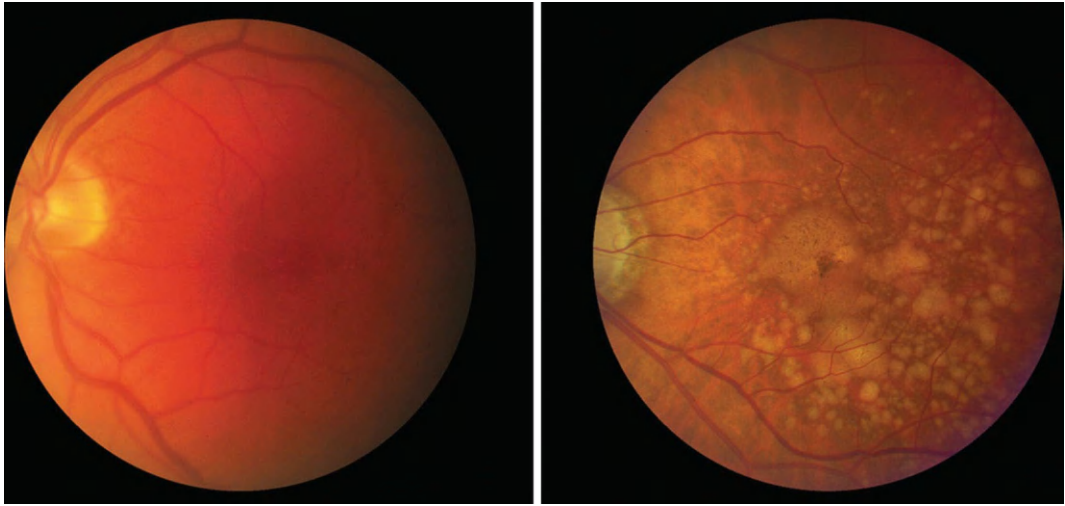


Fig. 13.1 An example of non-referable (left) and referable (right) AMD

drusen (63–125 μm) and/or pigmentary abnormalities; step-3: large drusen ($\geq 125 \mu\text{m}$) or numerous medium-sized drusen and pigmentary abnormalities; step-4: choroidal neovascularization or geographic atrophy), and then fused the four classes into two (non-referable and referable). The resultant DLS, when compared to the AlexNet DLS used in the second study [5], produced a statistically-significant superior performance in terms of accuracy (91.6% vs. 88.4%), sensitivity (89.0% vs. 84.5%) and specificity (93.6% vs. 91.5%).

Based on longitudinal outcomes data, the AREDS 9-step severity scale [10, 11] incorporates detailed quantification of drusen area and pigmentary abnormalities, and provides 5-year risk estimates for progression to advanced AMD, which includes choroidal neovascularization, central geography, or both. While this severity scale provides fairly granular relative risk estimations for developing advanced AMD, the grading is so complex that it is likely no human ophthalmologist uses the scale. Nevertheless, the scale could be useful. For example, an eye with a step-1 score carries 0.3% chance of progression, while an eye with a step-9 score carries a 53% chance [10]. The fourth study [12] had two major contributions: using DL for fine-grained, 9-step severity

classification and for 5-year progression risk calculation inferred directly from a CFP as an input. Only one image from each stereo pair in the AREDS dataset (58,370 images in total) was used. The dataset was split at the patient level, and the training, validation and testing subsets comprised of 88%, 2% and 10% of the images, respectively. The DLS was trained with the ResNet 50 [13] DCNN. The 5-year progression risk inference was performed using three methods: soft prediction, hard prediction and regressed prediction. “Soft prediction” first calculated the 9-step class probabilities for each of the nine classes, using the ResNet 50 derived classifications and SoftMax output values, and computed the 5-year risk estimate as the expected value of class risk under this probability. “Hard prediction” first calculated the 9-step class probabilities for each of the nine classes, using the ResNet 50 derived classifications, and the risk associated with the class with maximum probability was designated as the 5-year risk estimate. “Regressed prediction” skipped the 9-step severity scale prediction step and directly mapped an input CFP image to a 5-year risk estimate by using the ResNet 50 in regression mode.

For the 9-step severity scale classification task, the DLS achieved a linearly weighted

kappa score of 0.738, suggesting a high degree of correlation with the gold standard established at the reading center. However, variation in the DLS's classification accuracy was seen across the nine classes, which was likely driven by the imbalances in sample size available for each class. For example, of the 58,370 images involved in this study, there were 24,411 step-1 images, but only 1160 step-9 images. For the 5-year progression risk estimation task, the "hard prediction" method performed best in most classes and the overall mean estimation error between the three methods ranged from 3.47 to 5.29%, indicating a relatively small estimation error by the DLS.

The fifth study [14] produced by this series of collaboration pivoted from utilizing DL for classification tasks to utilizing DL for generative tasks. Specifically, generative adversarial networks (GANs) [15] were used to create high-resolution CFPs with various stages of AMD (Fig. 13.2). GANs have two main components: "generative" and "discriminative". The "generative" network uses training data to generate synthetic images, which are then presented to the "discriminative" network that is responsible for discriminating between the synthetic and real images. The two networks are "adversarial" in that the "generative" network aims to generate synthetic images that can "fool" the "discriminative" network. These two networks are then trained reiteratively against each other to ultimately maximize the "authenticity" of the

synthetic images. Since the description of GANs by Goodfellow et al. in 2014 [15], many variants of GANs have been developed, e.g. fully-connected GANs, Laplacian Pyramid GANs, boundary equilibrium GANs, self-attention GANs, Big GANs, conditional GANs and progressive GANs (ProGANs) [16]. This study adopted ProGANs, the experiments involved two retinal specialists, and here are the major findings. First, the synthetic images were of high quality. There was no difference in gradeability detected between the synthetic and real images as determined by the two retinal specialists. Second, the synthetic images were judged realistic in most cases. The two retinal specialists only achieved 59.5% and 53.7% accuracy, respectively, in discriminating real AMD images from synthetic images. The accuracies of close to 50% suggested that the retinal specialists' ability in identifying the synthetic images was similar to random chance. Third, the synthetic images generated by the GANs were useful for training a DL algorithm. The authors showed that a DLS, trained entirely with synthetic images, still achieved a respectable AUC-ROC value of 0.9235 and accuracy of 82.92% in differentiating between non-referable versus referable AMD when tested against real images.

In summary, a series of collaboration between retinal specialists at the School of Medicine and computer scientists at the Applied Physics Lab at the Johns Hopkins University produced five publications that explored various

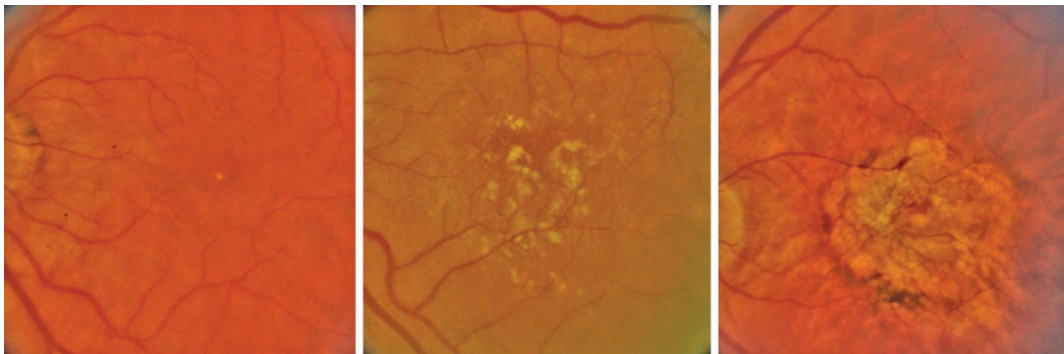


Fig. 13.2 Sample synthetic images generated by GANs: early (left), intermediate (middle) and advanced (right) AMD

aspects of DL applications in AMD using data from CFPs of AREDS: from referable AMD determination to fine-grained AMD severity classification to DCNN technical refinements to synthetic images generation with GANs. Future directions may include the following: using multimodal imaging from the AREDS, such as CFPs and optical coherence tomography imaging, to develop more powerful DLs, using GANs to supplement the under-represented classes of images in the 9-step severity scale to improve progression risk estimations, and testing DLs trained with AREDS images against images collected in clinical practice.

References

1. Bressler NM. Age-related macular degeneration is the leading cause of blindness. *JAMA*. 2004;291(15):1900–1.
2. Age-Related Eye Disease Study Research G. The age-related eye disease study (AREDS): design implications. AREDS report no. 1. *Control Clin Trials*. 1999;20(6):573–600.
3. Burlina P, Pacheco KD, Joshi N, Freund DE, Bressler NM. Comparing humans and deep learning performance for grading AMD: a study in using universal deep features and transfer learning for automated AMD analysis. *Comput Biol Med*. 2017;82:80–6.
4. Razavian AS, Azizpour H, Sullivan J, Carlsson S. CNN features off-the-shelf: an astounding baseline for recognition. Paper presented at: the Institute of Electrical and Electronics Engineers Conference of computer vision and pattern recognition, 12 May 2014, Stockholm, Sweden. <https://arxiv.org/pdf/1403.6382.pdf>
5. Burlina PM, Joshi N, Pekala M, Pacheco KD, Freund DE, Bressler NM. Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks. *JAMA Ophthalmol*. 2017;135(11):1170–6.
6. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. <https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>
7. Burlina P, Joshi N, Pacheco KD, Freund DE, Kong J, Bressler NM. Utility of deep learning methods for referability classification of age-related macular degeneration. *JAMA Ophthalmol*. 2018;136(11):1305–7.
8. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of IEEE conference on computer vision and pattern recognition*, 27–30 June 2016, Las Vegas, NV. p. 771–8.
9. Age-Related Eye Disease Study Research G. The age-related eye disease study system for classifying age-related macular degeneration from stereoscopic color fundus photographs: the age-related eye disease study report number 6. *Am J Ophthalmol*. 2001;132(5):668–81.
10. Davis MD, Gangnon RE, Lee LY, et al. The age-related eye disease study severity scale for age-related macular degeneration: AREDS report no. 17. *Arch Ophthalmol*. 2005;123(11):1484–98.
11. Ying GS, Maguire MG, Alexander J, Martin RW, Antoszyk AN. Complications of age-related macular degeneration prevention trial research G. Description of the age-related eye disease study 9-step severity scale applied to participants in the complications of age-related macular degeneration prevention trial. *Arch Ophthalmol*. 2009;127(9):1147–51.
12. Burlina PM, Joshi N, Pacheco KD, Freund DE, Kong J, Bressler NM. Use of deep learning for detailed severity characterization and estimation of 5-year risk among patients with age-related macular degeneration. *JAMA Ophthalmol*. 2018;136(12):1359–66.
13. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the 2016 IEEE conference on computer vision and pattern recognition (CVPR)*. Piscataway, NJ: Institute of Electric and Electronics Engineers; 2016. p. 771–8.
14. Burlina PM, Joshi N, Pacheco KD, Liu TYA, Bressler NM. Assessment of deep generative models for high-resolution synthetic retinal image generation of age-related macular degeneration. *JAMA Ophthalmol*. 2019.
15. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. *Adv Neural Inf Process Syst*. 2014;2672–80.
16. Karras T, Aila T, Laine S, Lehtinen J. Progressive growing of GANs for improved quality, stability, and variation. 2017. arXiv preprint [arXiv:1710.10196](https://arxiv.org/abs/1710.10196)

Artificial Intelligence for Keratoconus Detection and Refractive Surgery Screening

14

José Luis Reyes-Luis, Ana Silvia Serrano-Ahumada,
and Roberto Pineda

Abstract

Artificial intelligence (AI) has been employed in Keratoconus screening for decades. Since the introduction of Neural Networks, the first machine learning techniques used for topographic classification, AI efforts have focused on detection of nascent forms of corneal ectasias (Forme Frustre Keratoconus and Keratoconus Suspects). Parallel to this, the development of new imaging technology has contributed to the increasing keratoconus data sets, improving the accuracy of algorithms. Currently, AI can assist ophthalmologists with automated interpretation of topographic maps, laser vision correction surgery (particularly in custom planning of the procedure), and prevention of iatrogenic ectasias. This area of medicine will continue to grow in complexity along with new technology promising to make its application routine in the future.

Keywords

Artificial intelligence · Machine learning · Dataset · Forme frustre keratoconus · Keratoconus suspects · Iatrogenic ectasia

Background

Keratoconus (KCN) is a progressive non-inflammatory corneal ectasia that is over 90% bilateral in reported cases [1]. Interestingly, initial unilateral presentations are described in approximately 4.5% of patients; however, 50% of this population will develop signs of keratoconus in the ‘unaffected eye’ within 16 years [2]. The prevalence of keratoconus is also variable, ranging from 1 in 50 people in regions such as Central India [3] to 1 in 2000 people in countries such as the USA [4]. In early disease, patients with keratoconus are typically asymptomatic. Most of them will require glasses later in the course of the disease and, at some point if disease advances, corneal transplant may be necessary [5].

Fortunately, the distinction between mild and advanced keratoconus is not challenging because this disease has characteristic biomicroscopic and retinoscopic manifestations that can be easily recognized. In addition, there are many commercially available diagnostic tools,

J. L. Reyes-Luis · A. S. Serrano-Ahumada ·
R. Pineda (✉)
Massachusetts Eye and Ear, Boston, MA, USA
e-mail: roberto_pineda@meei.harvard.edu

A. S. Serrano-Ahumada
e-mail: cornea@oftalmologosmozo.com

R. Pineda
Harvard Medical School, Boston, USA

including topography, tomography, and anterior segment optic coherence tomography, that are highly accurate in the screening and classification of manifest keratoconus. However, the differentiation of the normal population from keratoconus suspects (KCS) as well as from forme fruste keratoconus (FFKC) patients is still a dilemma for ophthalmologists. Moreover, there is no standardized consensus regarding definitions for KCS and FFKC [5]. In general, for most of the literature, the term “keratoconus suspect” has been applied to patients with abnormal topography who do not have keratoconus in the fellow eye, while “forme fruste keratoconus” has been employed in patients with relatively normal corneal topography in one eye and clinical keratoconus in their fellow eyes [6].

On the other hand, Laser Vision Correction (LVC) procedures are among the most performed surgeries worldwide. Acknowledging the elective nature of the procedure, potential complications must be carefully considered and avoided, as this surgery can induce substantial biomechanical alterations that can result in iatrogenic ectasias [7–9]. The first LVC iatrogenic ectasia was reported by Seiler in 1998. In his report, Seiler described a case of a patient with a previous diagnosis of FFKC, who underwent laser in situ keratomileusis (LASIK) surgery. Since this report, FFKC has been considered a contraindication for LASIK surgery and its identification has become an essential step during refractive surgery screening [10]. However, preventing iatrogenic ectasia is not so easy, as the main challenge is to detect atypical corneas at high risk for developing ectasia while minimizing the number of normal patients who are denied surgery [11]. This must be performed employing high sensitivity and specificity risk factor detection methods [9, 11].

Recently, ectasia scoring systems refined the ability of some devices to detect early forms of keratoconus. For example, the Belin/Ambrosio enhanced ectasia display (BAD-D) by Pentacam (Oculus Wetzlar, Germany) combines elevation-based maps and pachymetry to screen corneal

ectasias with high sensitivity and specificity (Fig. 14.1). This is particularly useful for evaluating patients at risk for post-operative ectasia and identification of early or subclinical keratoconus. A final BAD-D score, also referred to as “final D”, greater than 1.88 has shown a 2.5% false positive rate with less than 1% false negatives [12].

Moreover, the use of Artificial Intelligence (AI) techniques has been recommended for difficult KCN diagnosis (such as differentiation of normal population from KCS and FFKC, as well as refractive surgery screening), as it constitutes a valuable instrument that has been proven to identify early signs of ectasias with high degree of reliability [2].

Artificial Intelligence

One common term used in AI is *big data*. This concept refers to the aggregate of data that has been generated over years in scientific research projects. This information is extremely extensive and complex; therefore, its storage and interpretation with traditional tools is not possible [13]. AI provides the investigator with useful tools that allow a better understanding and analysis of that data [8]. To achieve that objective, big data needs initially to be organized in a dataset. Then, to determine the presence of patterns in an extensive group of information, big data must undergo a process called *data mining*. This pattern recognition is very important and cannot be achieved manually (due to the complexity of the information). Data mining can be performed using two approaches: (1) *Supervised classification*, which is assisted by prototypes or examples; or (2) *Unsupervised classification* or *clustering*, which examines relationships between the properties of the objects [13].

The first study involving the use of AI for corneal pathologies was performed by Maeda in 1994 [14]. Since then several AI techniques have been employed for corneal topography interpretation and early keratoconus detection [11, 15].

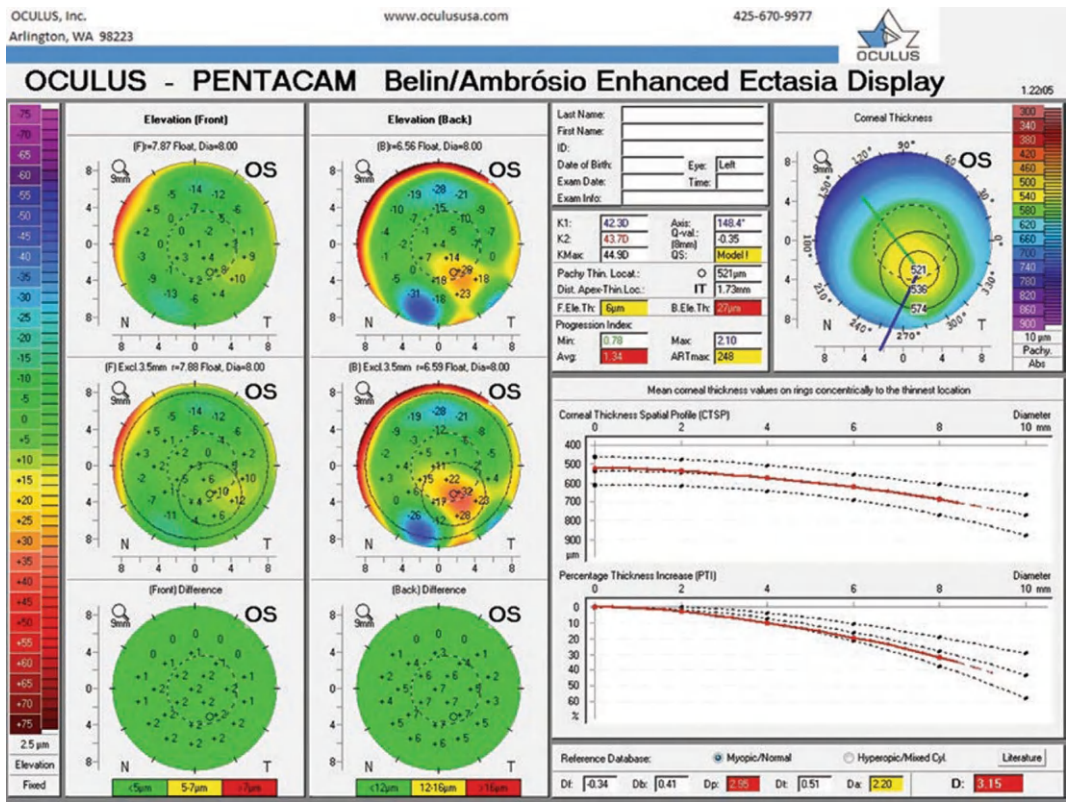


Fig. 14.1 The Belin-Ambrosio enhanced ectasia display (BAD-D) in a case with early keratoconus. Abnormal reported indices are represented in yellow and red

Artificial Intelligence Techniques

“Machine learning” is a popular AI technique which has the objective of mimicking human learning. This method has the advantage of giving computers the aptitude to ‘learn’ without being programmed to do so [5]. This is accomplished through machine programs, that allow the boosting of output using a previously known information dataset, and using multiple types of algorithms, e.g. the classifiers algorithm *See Box 14-1*. Machine learning algorithms are “trained” on a guidance data set, and afterward are “validated” on a different set of test or verification data, to ensure the external validity of this process [5].

Classifier

In machine learning, a classifier is an algorithm that implements classification to find patterns within a dataset. As previously exposed and according to the way new information is obtained, these algorithms can be divided into two: supervised or unsupervised. Classifiers can be used, for example, in the classification of topographic maps as normal, astigmatic, or with KCN [15].

Although most of the AI studies for KCN detection and refractive screening are based on

supervised machine learning techniques; other types of AI, including unsupervised machine learning technique, expert decision tree [15], and multivariate logistic regression analysis [16], have also proven to be useful.

Studies

Supervised Machine Learning Classifiers

Neural Networks

“Neural network” (NN) refers to the AI method that simulates human neurological processing abilities with the objective of interpolating or estimating complicated data [17]. This method has the capacity of detecting characteristics, that are concealed within the input data, without requiring additional information about logic structures between input and output data. Therefore, developing a successful interpretation system and obtaining an accurate response, the establishment of input data is fundamental in this technique.

To date, several studies have used NN for corneal topography classification. The first NN study was reported in 1995 by Maeda and collaborators. In their publication, topographic maps were classified (using 11 indices of the computer-assisted videokeratoscope TMS-1 [ComputedAnatomy, New York, NY]) into normal, astigmatism with the rule, KCN mild to advanced, post-photorefractive keratectomy and post penetrating keratoplasty. The NN backpropagation model (which is a supervised, multilayer, feed-forward network) was used, obtaining a correct classification in 80% of cases (with >90% specificity, and 44–100% sensitivity) [17].

In 1997, Smolek and Klyce published their first NN study. Using 10 indices of the computer-assisted videokeratoscope TMS-1, they were able to detect KCS and KCN severity with 100% accuracy, 100% sensitivity, and 100% specificity [18]. Four years later, these two authors also reported their NN findings using wavelet data from the computer-assisted

videokeratoscope TMS-1. The NN was able to detect prior refractive surgery with 99.3% accuracy, 99.1% sensitivity and 100% specificity. These outcomes were considerably higher than those obtained by clinicians (who only achieved 65% sensitivity and 95% specificity) [19].

Accardo and Pensiero, on the other hand, used NN to evaluate unilateral and bilateral indices, as well as early KCN maps obtained with the videokeratoscope EyeSys (EyeSys Vision, Houston, Texas). This method allowed the authors to differentiate early KCN from non-KCN corneas with 94.1% sensitivity and 97.6% specificity [20].

In 2008, Vieira de Carvalho and Barbosa employed NN and discriminant analysis to classify corneal shapes using Zernicke coefficients. Eyesys videokeratoscope data was used as inputs, achieving a 94% accuracy with NN and 84.8% accuracy with discriminant analysis [21].

Six years later, Silverman and collaborators compared the ability of NN and linear discriminant analysis to differentiate between normal and KCN corneas. This was done by using maps of epithelial and stromal thickness obtained from Artemis 1 (ArcScan, Morrison, CO), digital ultrasound scanner. Both classifiers obtained an Area Under the Receiver Operating Characteristics (AUROC) of 1.0 (indicative of complete separation of groups); however, NN sensitivity and specificity were slightly better when compared to linear discriminant analysis (98.9% and 99.5% versus 94.6% and 99.2%, respectively) [11].

Kovacs et al., in turn, analyzed unilateral and bilateral indices obtained with Scheimpflug imaging (Pentacam corneal topographer). NN differentiated normal and KCN corneas with an AUROC of 0.99, 100% sensitivity, and 95% specificity; however, differentiation between normal and KFFK corneas was slightly less accurate, as it only achieved an AUROC of 0.97, 90% sensitivity, and 90% specificity [2].

Most recently, Feng et al. developed a convolutional neural network called KerNet. This novel AI model uses end-to-end deep learning to process raw data from the Pentacam HR system for keratoconus and subclinical keratoconus detection [22].

This model was compared with four machine-learning algorithms (eXtreme Gradient Boosting XGBoost, Light Gradient Boosting Machine LGBM, Logistic Regression LR and Random Forest RF) showing better diagnostic performance for early keratoconus discrimination with overall accuracies of 95.24% in the test set (AUC 0.984) and 94.12% in the validation set (AUC 0.983) for asymmetric keratoconus eyes [23].

Mohammadpour in 2022 compared AI-based KCN classifiers including BADD and TKC of Pentacam, and OPD Scan III corneal navigator, and Sirius Phoenix, they found this last one had superior accuracy for differentiating KCN (91.24%) and KCS (88.68%) compared to the others [24].

Artificial intelligence models have been researched not only to improve the diagnosis of KCN, but also as a screening tool for KCN or refractive surgery candidates. Ahn et al. [25] investigated how to select patients who would best benefit from a corneal topography as KCN screening using a total of five AI models (three fully connected neural network and two ensemble models) for training and validation of clinical information obtained for basic ophthalmic examinations such as best corrected visual acuity (BCVA), intraocular pressure (IOP), autokeratometry values, mean corneal power (Km) and corneal astigmatism, and differences in BCVA, IOP and spherical equivalent (SE) between eyes. These models performed relatively well when tested with external validation, demonstrating a sensitivity of over 90% but a low positive predictive value (PPV).

Automated Decision Tree Classification

The “automated decision tree” classification method is a machine learning algorithm that was generated employing a sample of training data with a known class assignment. This adaptable investigation tool is suitable for nonparametric data analysis and has been used for data mining [26]. In this technique, cutoff values of discriminant variables are designated, giving place to successive ‘nodes’, ‘branches’ (which divide into 2 mutual subgroups), and ‘leaves’ (the final decision of class assignment).

Smajda and collaborators have used the automated decision tree classification method to discriminate inputs from the Galilei corneal topographer (Ziemer Ophthalmic Systems, Port, Switzerland). Posterior surface asymmetry index and corneal volume were the two most important discriminant variables in this study. With this method, normal and KCN corneas were classified with 100% sensitivity and 93.6% specificity; while normal and FFKC corneas were categorized with 93.6% sensitivity and 97.2% specificity [26].

WEKA Software Classifiers (SVM, Random Forest, Bayes Network)

More recently, an open-source machine learning software was created at the University of Waikato, New Zealand. This software, called Waikato Environment for Knowledge Analysis (WEKA) workbench, contains a compilation of machine learning algorithms and data processing tools for data mining [27]. An example of the learning algorithms offered by WEKA are the supervised classifiers: Support Vector Machine (SVM), Bayes Network, Multi-Layer Perceptron (MLP), Naive Bayes, Random Forest, and Radial Basis Function Neural Network (RBNFNN) [13].

In 2010, Souza et al. evaluated the performance of SVM, MLP and RBNFNN on KCN detection using indices from the Orbscan II (Bausch & Lomb, Quebec, Canada), scanning slit topographer. The results showed high KCN detection indices for all three machine learning algorithms, however, no-difference between SVM and MLP efficiency (with 0.99 AUROC, and 100% sensitivity), and RBNFNN high values (0.98 AUROC and 98% sensitivity) [28].

Later, Ruiz Hidalgo and collaborators evaluated the KCN detection effectiveness of the SVM algorithm using 22 parameters obtained from the Pentacam corneal topographer. Their results demonstrated a KCN versus non-KCN classification accuracy of 98.9%, sensitivity of 99.1%, and specificity of 98.5%; while a FFKC vs non-KCN classification achieved an accuracy of 93.1%, sensitivity of 79.1%, and specificity 97.9%. Finally, the authors were able to classify

5 groups [KCN, FFKC, Astigmatic, PR (post-refractive surgery) and Normal] with an 88.8% accuracy, sensitivity 89%, and 95.2% specificity [29].

More recently, in a study published by Lopes et al. intended to detect corneal ectasia susceptibility by analyzing Pentacam tomographic data of patients with unilateral asymmetric ectasia, clinical KCN, and stable LASIK. The authors compared five classifiers (Random Forest, Naive Bayes, NN, SVM, and regularized Discriminant Analysis) of which the Random Forest (RF) yielded the highest accuracy (80–85.2% sensitivity and 96.6% specificity) [8].

Similarly, a recent study by Kundu et al. analyzed curvature, wavefront aberrations using Zernike polynomials and three-dimensional (3D) thickness distributions from images obtained with corneal tomography combining Placido and OCT (MS-39, CSO, Italy) with a RF AI model. These images were classified as healthy, very asymmetric ectasia (VAE) and keratoconus (KC) with an AUC of 0.994, 0.976 and 0.997 respectively [30].

In a current literature review, the RF classifier is one of the most widely adopted AI models and has been used not only as previously mentioned as an aid in the diagnosis and classification of healthy, KCS or KCN; but also, to identify high-risk factors of progression in KCN patients in a first visit [31], as well as to identify progression based on topographic changes such as maximum anterior curvature (Kmax) with good results [32].

Unsupervised Machine Learning

Contrary to supervised machine learning techniques, the unsupervised algorithms do not need to pre-label data for training [33]. This allows investigators to use a non-biased approach that employs multiple comprehensive parameters for analysis.

Although to date most of the machine learning techniques for automatic detection of KCN are supervised, Yousefi et al. conducted an unsupervised machine learning study using a

big dataset of corneal images obtained with the corneal swept-source 1000 CASIA OCT. The classification was able to detect KCN cases with 97.7% sensitivity and 94.1% specificity. To the best of our knowledge, this is the first and only study that employs unsupervised machine learning in corneal ectasia detection [33].

Non-machine Learning Classifiers

Linear Discriminant Analysis

“Linear Discriminant Analysis” (LDA) is a technique commonly used for data reduction that can help as a pre-processing step for pattern classification applications and machine learning. Its goal is to display the original data in a smaller dimensional space while maximizing the separation between categories [34].

Saad and Gatinel used LDA to differentiate KCN, FFKC, and Non-KCN using 51 Orbscan II topographic indices. Percentage of thickness increase and maximum posterior corneal elevation were the most important contributors and, with this technique, the differentiation capacity between normal and FFKC groups and between normal and KCN groups reached an AUROC of 0.98 and 0.99, respectively. These results suggest that it is plausible to accurately differentiate normal from FFKC eyes using topography indices [35].

Two years later, the same investigators, used the OPD-Scan corneal analyzer to obtain Zernicke coefficients and differentiate between normal, FFKC and KCN corneas using LDA. An AUROC of 0.98 for the distinction between normal and FFKC corneas was reported, while an AUROC 0.96 between normal and KCN corneas was reached [36].

Expert System Classifier

The “Expert System Classifier” is a non-machine learning AI technique that through an extensive set of decision rules, deductive decisions, and step-by step logical operations resolutions are reached [17]. In 1994, Maeda and collaborators reported the first Expert System Classifier results and combined them with LDA to differentiate KCN and non-KCN. Using

indices from TMS-1 maps, this method resulted in 96% accuracy, 89% sensitivity and 99% specificity [15].

Multivariate Logistic Regression Analysis

Although multivariate logistic regression analysis (MLRA) is not a machine learning technique, the classification results that have been obtained with this classic statistical method are worth including in this chapter. This method is convenient for models with dichotomous dependent variables and uses logistic regression coefficients to estimate odds ratios for each independent variable [37].

In 2018, Hwang and collaborators used a multivariate logistic regression analysis to differentiate normal from FFKC corneas. They combined 5 Scheimpflug Pentacam indices with 11 anterior segment OCT RT-Vue-100 (Optovue, Fremont, CA) indices, obtaining a classification system with 100% sensitivity and 100% specificity. The mayor contributors in this analysis were the epithelial thickness variability, the total focal corneal thickness variability from OCT, the anterior curvature and the topometric indices from Scheimpflug tomography [16]. In Fig. 14.2, anterior segment OCT and Scheimpflug images, like the ones used in Hwang study, are displayed.

More recently, Almeida et al. used MLRA to analyze 22 Scheimpflug corneal tomography (Pentacam HR) variables to distinguish between KCN, very asymmetric ectasia and normal corneal topography and tomography (VAE-NTT) and healthy corneas. The analysis of this variables and AI model training led to the creation of the Boosted Ectasia Susceptibility Tomography Index (BESTi), which was then validated. This index showed good performance with an AUC of 0.91, 86.02% sensitivity and 83.97% specificity when classifying between control group and VAE-NTT, which was statistically significant compared to Belin-Ambrosio Deviation Index (BAD-D) and Pentacam Random Forest Index (PRFI) [38]. This promises to be a useful tool when screening for subclinical KCN before refractive surgery.

Corneal Biomechanics

Recently, a spotlight has fallen on the evaluation of corneal Biomechanics not only due to its capability of diagnosing masked corneal ectasias, but also for guiding therapeutic interventions such as crosslinking. Nowadays, there are few devices for in vivo evaluation of corneal biomechanics which include the ocular response analyzer (ORA) and the Corvis ST [39–41]. Biomechanics have contributed to the expansion of corneal imaging and more studies are expected to improve datasets for AI in keratoconus and refractive screening.

Corvis ST (Oculus Optikgeräte GmbH, Wetzlar, Germany) uses an ultra-high-speed Scheimpflug camera to record corneal deformation after an air-puff recording three states to produce dynamic response parameters that correspond to corneal biomechanical information. AI has been used since the introduction of this device to better understand the output data and enhance its diagnostic value.

This was the case of the Corneal Biomechanical Index (CBI), which included several dynamic corneal response parameters and was developed by Vinciguerra et al. to distinguish between normal corneas and KCN through logistic regression with an AOC of 0.983 and a specificity of 100% and sensitivity of 98.2% with a cut-off value of 0.5 [42].

The Tomographic and Biomechanical Index (TBI) was later introduced and validated by Ambrosio et al. This index combined information obtained from Pentacam HR and Corvis ST that was analyzed and combined using three different AI methods. The random forest with leave-one-out cross-validation (RF/LOOCV) had the best performance regarding KCN detection with an AUROC of 0.996 and a 100% sensitivity and 100% specificity with a cut-off value of 0.79 [43].

In Fig. 14.3 the results of a patient with a very asymmetric ectasia (VAE-E) evaluated with the Ambrosio, Roberts and Vinciguerra (ARV) biomechanics and tomographic assessment, performed with the Corvis ST device.

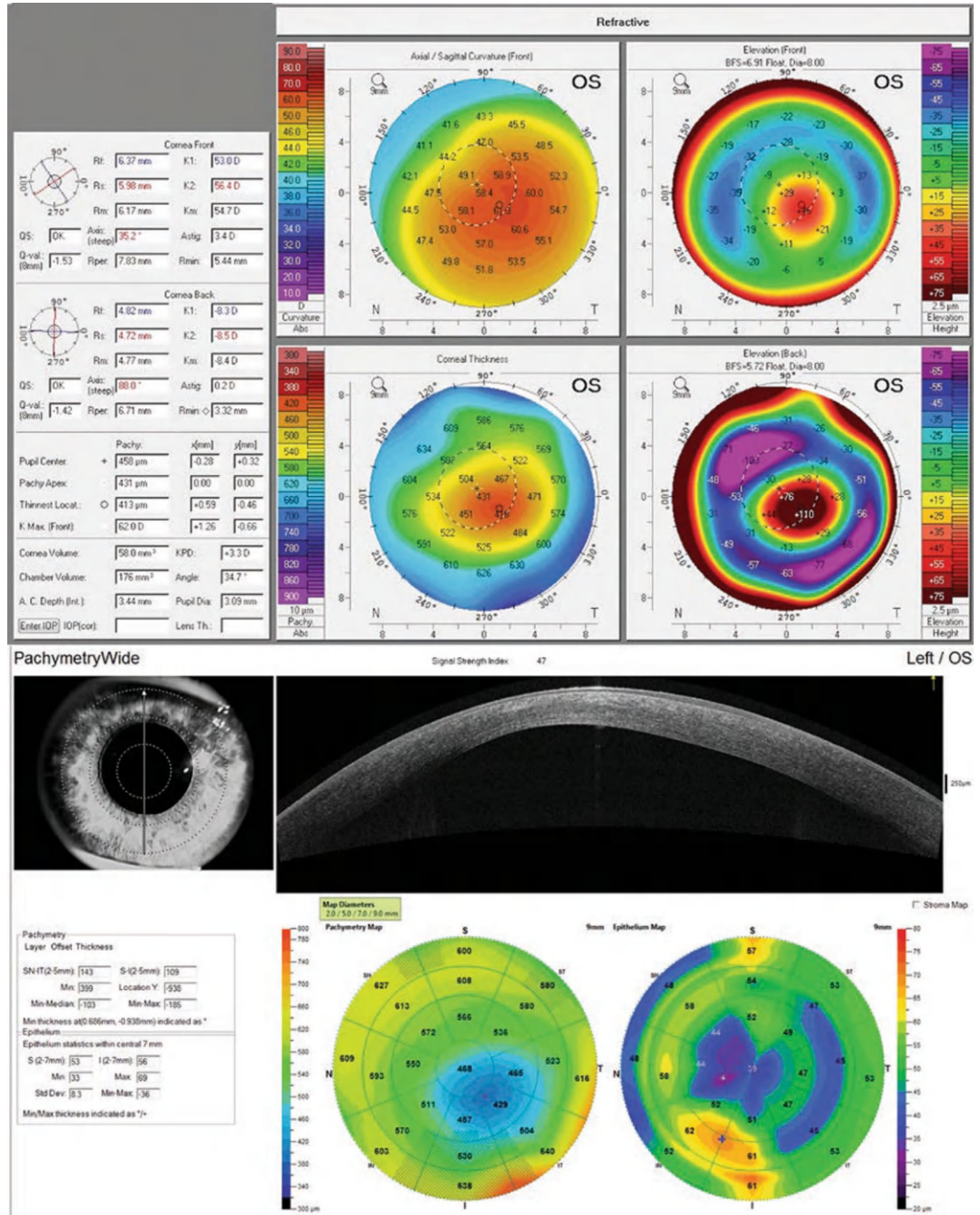


Fig. 14.2 Combined image of anterior segment OCT and Scheimpflug tomography from a patient with KCN in the left eye. The Scheimpflug tomography shows the anterior elevation and curvature maps (top), and the corneal thickness and the posterior elevation maps (middle). The total corneal thickness and epithelial thickness map can be found in the anterior segment OCT report (bottom)

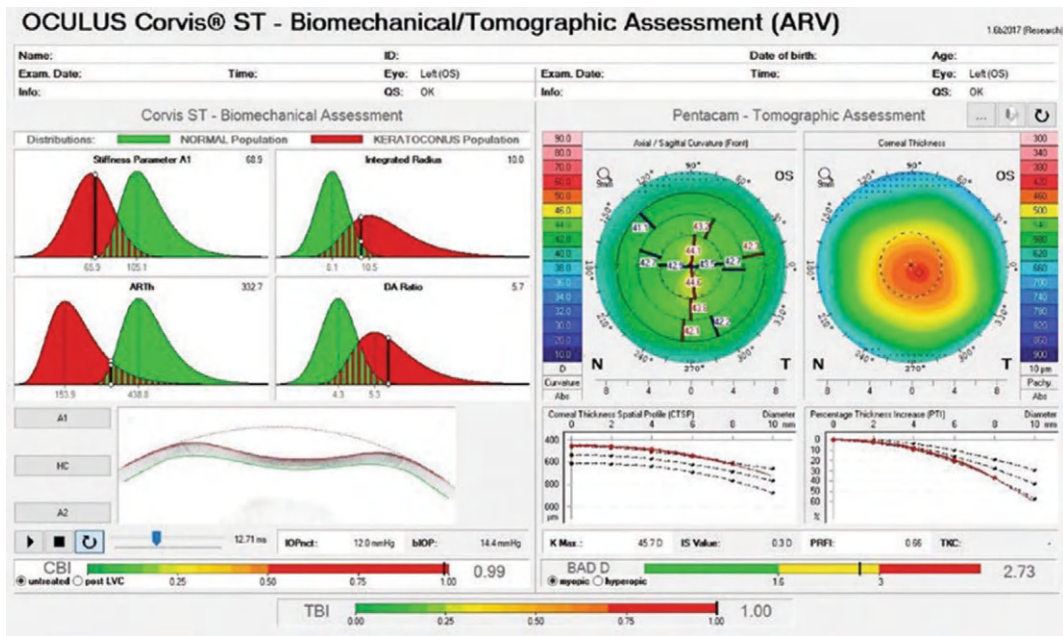


Fig. 14.3 The ARV biomechanical and tomographic display in a patient left eye with keratoconus. We can see below the Corvis biomechanical index (CBI), tomographic biomechanical index (TBI) and Belin Ambrosio

enhanced ectasia display (BAD-D). Abnormal reported indices are represented in yellow and red. Image courtesy of Dr. Renato Ambrosio MD Ph.D.

More recently, a study from Tan et al. was the first to use machine learning (ML) to analyze raw corneal dynamic deformation video parameters using Corvis ST without correlation with corneal topography to diagnose KCN solely on biomechanical parameters. This group trained a feedforward neural network model (5-FNN) to distinguish between normal corneas and KCN based on four corneal biomechanical parameters with an 98.7% accuracy, 99.3% sensitivity, 100% specificity and 100% precision [44].

Limitations

The term “ground truth” is used to depict the algorithm result that the AI technique is programmed to obtain; however, reaching that ground truth is not a simple task. Large datasets can be more reliable for designating the ground truth, as cutoffs can be set to include only the advanced forms of the disease (e.g. manifest corneal ectasias). However, in cases which there

is not an expert consensus on the definition of a specific condition, such as early ectasias following refractive surgery, a reliable classification is difficult to achieve [5].

In addition, the lack of large sample sizes for low-incidence conditions (such as FFKC and early post-refractive surgery ectasia) and the scarcity of exposure to a sufficient variety of presentations of the same pathology and its differential diagnoses (e.g. ectasias following refractive surgery, KCN, and pellucid marginal degeneration), undermine the results of the algorithms and reduce their external validity [14].

Advantages of AI in Cornea

The advantages of using AI in corneal ectasias include:

- Automated interpretation of topographic maps. This may help unskilled observers

enhance their decision-making performance to a more skilled level [17].

- Early Keratoconus detection. KCS and FFKC detection by AI models have achieved good results, allowing ophthalmologists to diminish progression by the administration of early treatments. One of these treatments is Riboflavin-induced UVA cross-linking, which seems to be an effective and enduring treatment when the detection is done at early stages [45].
- Customized LASIK ablations. Recently AI capacity to combine data has been focused on optimizing refractive results. This led to the design of LASIK ablations based on axial length measurements, total eye wavefront input, and detailed corneal and anterior segment topography data, this first module was called Innoveyes by Alcon. With this information, mathematical models of raytracing, prediction of biomechanical changes of the cornea and the anticipated epithelial remodeling can also be obtained [46].

Conclusions

Due to the lack of a validated datasets upon techniques as well as the absence of a consistent definition for KCN or its early manifestations, it is difficult to truly determine whether one AI technique offers a significant advantage over the others (e.g. FFKC and KCS). Despite this, AI techniques have proven to be fruitful in KCN detection and refractive surgery screening, with promising capacity for continued progress as imaging instrumentation and methods become more advanced [5].

References

1. Burns DM, Johnston FM, Frazer DG, Patterson C, Jackson AJ, et al. Keratoconus: an analysis of corneal asymmetry. *Br J Ophthalmol*. 2004;88(10):1252–5. <https://doi.org/10.1136/bjo.2003.033670>.
2. Kovács I, Miháltz K, Kránitz K, Juhász É, Takács Á, Dienes L, et al. Accuracy of machine learning classifiers using bilateral data from a Scheimpflug camera for identifying eyes with preclinical signs of keratoconus. *J Cataract Refract Surg*. 2016;42(2):275–83. <https://doi.org/10.1016/j.jcrs.2015.09.020>.
3. Jonas JB, Nangia V, Matin A, Kulkarni M, Bhojwani K. Prevalence and associations of keratoconus in rural Maharashtra in central India: the central India eye and medical study. *Am J Ophthalmol*. 2009;148(5):760–5. <https://doi.org/10.1016/j.ajo.2009.06.024>.
4. Kennedy RH, Bourne WM, Dyer JA. A 48-year clinical and epidemiologic study of keratoconus. *Am J Ophthalmol*. 1986;101(3):267–73.
5. Lin SR, Ladas JG, Bahadur GG, Al-Hashimi S, Pineda R. A review of machine learning techniques for keratoconus detection and refractive surgery screening. *Semin Ophthalmol*. 2019;34(4):317–26. <https://doi.org/10.1080/08820538.2019.1620812>.
6. Klyce SD. Chasing the suspect: keratoconus. *Br J Ophthalmol*. 2009;93(7):845–7. <https://doi.org/10.1136/bjo.2008.147371>.
7. Ambrósio R, Randleman JB. Screening for ectasia risk: what are we screening for and how should we screen for it? *J Refract Surg*. 2013;29(4):230–2. <https://doi.org/10.3928/1081597X-20130318-01>.
8. Lopes BT, Ramos IC, Salomão MQ, Guerra FP, Schallhorn SC, Schallhorn JM, et al. Enhanced tomographic assessment to detect corneal ectasia based on artificial intelligence. *Am J Ophthalmol*. 2018;195:223–32. <https://doi.org/10.1016/j.ajo.2018.08.005>.
9. Santhiago MR, Smadja D, Gomes BF, Mello GR, Monteiro ML, Wilson SE, et al. Association between the percent tissue altered and post-laser in situ keratomileusis ectasia in eyes with normal preoperative topography. *Am J Ophthalmol*. 2014;158(1). <https://doi.org/10.1016/j.ajo.2014.04.002>.
10. Seiler T, Quirke AW. Iatrogenic keratectasia after LASIK in a case of forme fruste keratoconus. *J Cataract Refract Surg*. 1998;24(7):1007–9.
11. Silverman RH, Urs R, Roychoudhury A, Archer TJ, Gobbe M, Reinstein DZ. Epithelial remodeling as basis for machine-based identification of keratoconus. *Invest Ophthalmol Vis Sci*. 2014;55(3):1580. <https://doi.org/10.1167/iiov.13-12578>.
12. Belin MW, Villavicencio OF, Ambrósio RR. Tomographic parameters for the detection of keratoconus. *Eye Contact Lens*. 2014;40(6):326–30. <https://doi.org/10.1097/ICL.0000000000000077>.
13. Amancio DR, Comin CH, Casanova D, Travieso G, Bruno OM, Rodrigues FA, et al. A systematic comparison of supervised classifiers. *PLoS ONE*. 2014;9(4). <https://doi.org/10.1371/journal.pone.0094137>.
14. Maeda N, Klyce SD, Smolek MK, Thompson HW. Automated keratoconus screening with corneal topography analysis. *Invest Ophthalmol Vis Sci*. 1994;35(6):2749–57.
15. Klyce SD. The future of keratoconus screening with artificial intelligence. *Ophthalmology*. 2018;125(12):1872–3. <https://doi.org/10.1016/j.ophttha.2018.08.019>.

16. Hwang ES, Perez-Straziota CE, Kim SW, Santhiago MR, Randleman JB. Distinguishing highly asymmetric keratoconus eyes using combined Scheimpflug and spectral-domain OCT analysis. *Ophthalmology*. 2018;125(12):1862–71. <https://doi.org/10.1016/j.ophtha.2018.06.020>.
17. Maeda N, Klyce SD, Smolek MK. Neural network classification of corneal topography. Preliminary demonstration. *Invest Ophthalmol Vis Sci* 1995;36(7):1327–35.
18. Smolek MK, Klyce SD. Current keratoconus detection methods compared with a neural network approach. *Invest Ophthalmol Vis Sci*. 1997;38(11):2290–9.
19. Smolek MK, Klyce SD. Screening of prior refractive surgery by a wavelet-based neural network. *J Cataract Refract Surg*. 2001;27(12):1926–31.
20. Accardo P, Pensiero S. Neural network-based system for early keratoconus detection from corneal topography. *J Biomed Inform*. 2002;35(3):151–9.
21. Carvalho LAVD, Barbosa MS. Neural networks and statistical analysis for classification of corneal videokeratography maps based on Zernike coefficients: a quantitative comparison. *Arq Bras Oftalmol*. 2008;71(3):337–41.
22. Feng R, Xu Z, Zheng X, Hu H, Jin X, Chen DZ, Yao K, Wu J. KerNet: a novel deep learning approach for keratoconus and sub-clinical keratoconus detection based on raw data of the Pentacam HR system. *IEEE J Biomed Health Inform*. 2021;25(10):3898–910. <https://doi.org/10.1109/JBHI.2021.3079430>.
23. Xu Z, Feng R, Jin X, Hu H, Ni S, Xu W, Zheng X, Wu J, Yao K. Evaluation of artificial intelligence models for the detection of asymmetric keratoconus eyes using Scheimpflug tomography. *Clin Experiment Ophthalmol*. 2022;50(7):714–23. <https://doi.org/10.1111/ceo.14126>.
24. Mohammadpour M, Heidari Z, Hashemi H, Yaseri M, Fotouhi A. Comparison of artificial intelligence-based machine learning classifiers for early detection of keratoconus. *Eur J Ophthalmol*. 2022;32:1352–60. <https://doi.org/10.1177/11206721211073442>
25. Ahn H, Kim NE, Chung JL, Kim YJ, Jun I, Kim TI, Seo KY. Patient selection for corneal topographic evaluation of keratoconus: a screening approach using artificial intelligence. *Front Med*. 2022;9:934865. <https://doi.org/10.3389/fmed.2022.934865>.
26. Smadja D, Touboul D, Cohen A, Doveh E, Santhiago MR, Mello GR, et al. Detection of subclinical keratoconus using an automated decision tree classification. *Am J Ophthalmol*. 2013;156(2). <https://doi.org/10.1016/j.ajo.2013.03.034>
27. Witten IH, Frank E, Hall MA, Pal CJ. Data mining: practical machine learning tools and techniques. 4th ed. Amsterdam: Elsevier; 2016. p. 7–15.
28. Souza MB, Medeiros FW, Souza DB, Garcia R, Alves MR. Evaluation of machine learning classifiers in keratoconus detection from Orbscan II examinations. *Clinics (Sao Paulo)*. 2010;65(12):1223–8. <https://doi.org/10.1590/s1807-59322010001200002>.
29. Hidalgo IR, Rodriguez P, Rozema JJ, Dhuhghaill SN, Zakaria N, Tassignon M-J, et al. Evaluation of a machine-learning classifier for keratoconus detection based on Scheimpflug tomography. *Cornea*. 2016;35(6):827–32. <https://doi.org/10.1097/ICO.0000000000000834>.
30. Kundu G, Shetty R, Khamar P, et al. Universal architecture of corneal segmental tomography biomarkers for artificial intelligence-driven diagnosis of early keratoconus. *Br J Ophthalmol*. 2023;107(5):635–43. <https://doi.org/10.1136/bjophthalmol-2021-319309>.
31. Kundu G, Shetty N, Shetty R, Khamar P, D'Souza S, Meda TR, Nuijts RMMA, Narasimhan R, Roy AS. Artificial intelligence-based stratification of demographic, ocular surface high-risk factors in progression of keratoconus. *Indian J Ophthalmol*. 2023;71(5):1882–8. https://doi.org/10.4103/IJO.IJO_2651_22.
32. Shetty R, Kundu G, Narasimhan R, Khamar P, Gupta K, Singh N, Nuijts RMMA, Sinha Roy A. Artificial intelligence efficiently identifies regional differences in the progression of tomographic parameters of keratoconic corneas. *J Refract Surg (Thorofore, N.J.: 1995)*. 2021;37(4):240–8. <https://doi.org/10.3928/1081597X-20210120-01>
33. Yousefi S, Yousefi E, Takahashi H, Hayashi T, Tampo H, Inoda S, et al. Keratoconus severity identification using unsupervised machine learning. *Plos One*. 2018;13(11). <https://doi.org/10.1371/journal.pone.0205998>
34. Tharwat A, Gaber T, Ibrahim A, Hassanien AE. Linear discriminant analysis: a detailed tutorial. *AI Commun*. 2017;30(2):169–90. <https://doi.org/10.3233/AIC-170729>.
35. Saad A, Gatinel D. Topographic and tomographic properties of forme fruste keratoconus corneas. *Invest Ophthalmol Vis Sci*. 2010;51(11):5546. <https://doi.org/10.1167/iovs.10-5369>.
36. Saad A, Gatinel D. Evaluation of total and corneal wavefront high order aberrations for the detection of forme fruste keratoconus. *Invest Ophthalmol Vis Sci*. 2012;53(6):2978. <https://doi.org/10.1167/iovs.11-8803>.
37. Alexopoulos EC. Introduction to multivariate regression analysis. *Hippokratia*. 2010;14(Suppl 1):23–8.
38. Almeida GC Jr, Guido RC, Balarin Silva HM, Brandão CC, de Mattos LC, Lopes BT, Machado AP, Ambrósio R Jr. New artificial intelligence index based on Scheimpflug corneal tomography to distinguish subclinical keratoconus from healthy corneas. *J Cataract Refract Surg*. 2022;48(10):1168–74. <https://doi.org/10.1097/j.jcrs.0000000000000946>.
39. Yuan A, Pineda R. Developments in imaging of corneal biomechanics. *Int Ophthalmol Clin*. 2019;59(4):1–17. <https://doi.org/10.1097/IIO.0000000000000286>.
40. Gokul A, Vellara HR, Patel DV. Advanced anterior segment imaging in keratoconus: a review. *Clin Exp Ophthalmol*. 2018;46(2):122–32. <https://doi.org/10.1111/ceo.13108>.

41. De Stefano VSD, Dupps WJ. Biomechanical diagnostics of the cornea. *Int Ophthalmol Clin*. 2017;57(3):75–86. <https://doi.org/10.1097/HIO.0000000000000172>.
42. Vinciguerra R, Ambrósio R Jr, Elsheikh A, Roberts CJ, Lopes B, Morenghi E, Azzolini C, Vinciguerra P. Detection of keratoconus with a new biomechanical index. *J. Refract Surg (Thorofare, N.J.: 1995)*. 2016;32(12):803–10. <https://doi.org/10.3928/1081597X-20160629-01>
43. Ambrósio R Jr, Lopes BT, Faria-Correia F, Salomão MQ, Bühren J, Roberts CJ, Elsheikh A, Vinciguerra R, Vinciguerra P. Integration of Scheimpflug-based corneal tomography and biomechanical assessments for enhancing ectasia detection. *J Refract Surg (Thorofare, N.J.: 1995)*. 2017;33(7):434–43. <https://doi.org/10.3928/1081597X-20170426-02>
44. Tan Z, Chen X, Li K, Liu Y, Cao H, Li J, Jhanji V, Zou H, Liu F, Wang R, Wang Y. Artificial intelligence-based diagnostic model for detecting keratoconus using videos of corneal force deformation. *Transl Vis Sci Technol*. 2022;11(9):32. <https://doi.org/10.1167/tvst.11.9.32>.
45. Keating A, Roberto Pineda II, Colby K. Corneal cross linking for keratoconus. *Semin Ophthalmol*. 2010;25(5–6):249–55. <https://doi.org/10.3109/08820538.2010.518503>.
46. Sightmap & InnovEyes. YouTube [Internet] [cited 2020 Mar 24]. Available from: <https://www.youtube.com/watch?v=CPcRoH0qcPM>

Artificial Intelligence in the Treatment of Keratoconus: The Use of Intracorneal Ring Segments

15

C. Fariselli, F. Versace and J. L. Alio

Abstract

The chapter analyzes the clinical results of an artificial neural network (ANN) that has been processed in order to improve the predictability of intracorneal ring segments (ICRS) implantation in keratoconus. The ANN to guide ICRS provides an increase in the visual acuity, reduction in the spherical equivalent and improvement in the optical quality of keratoconus patients. Artificial intelligence gives better results when compared with the manufacturer's nomograms in terms of better corrected vision and reduction of the coma-like aberrations.

Keywords

Keratoconus · Intracorneal ring segments · Artificial neural network

Keratoconus is a progressive, non-inflammatory corneal ectatic disease, characterized by corneal thinning and steepening [1]. Changes associated with the progression of the disease are not limited to anatomical and morphological alterations of the cornea, but they are also responsible of a significant decrease in quality and quantity of vision, resulting from increased ocular aberrations and loss of corneal transparency, which can affect quality of life [2]. The grading system “RETICS classification” is based on clinical signs and corneal biomechanical characteristics, but also on optical and visual function parameters [3]. Moreover, the measurement of corrected distance visual acuity (CDVA), which is objective and readily available in daily practice, can be correlated with other continuous variables, allowing a greater practical use [3].

Intracorneal ring segments (ICRS) implantation is a minimally invasive and reversible surgery [4, 5], which can delay, and sometimes avoid, the need for corneal transplantation [6]. It represents an additive surgical procedure, which is able to reshape highly distorted corneal surfaces [7] and to redistribute the asymmetrical corneal stress caused by the biomechanical

C. Fariselli
Department of Ophthalmology, Imola Hospital,
Bologna, Italy

F. Versace
CSO, Scandicci, Firenze, Italy

J. L. Alio (✉)
Department of Cornea, Cataract and Refractive
Surgery, Alicante, Spain
e-mail: jlalio@vissum.com

Department of Research and Development,
VISSUM, Alicante, Spain

School of Medicine, Universidad Miguel Hernández,
Alicante, Spain

decompensation [8], producing a remodeling of the anterior and posterior corneal topography. Placement of ICRS deep into the stroma of selected patients [9] improves the optical quality of the cornea and reduces optical aberrations, with consequent better visual acuity and contact lens tolerance [10, 11] (Figs. 15.1 and 15.2).

Traditionally, the characteristics of ICRS to be implanted (as number, arc length and thickness) are chosen according to the manufacturer's nomogram. However, this treatment planning is based on parameters which are subjective, empirical and not predictable in keratoconus

cases, such as sphero-cylindrical refraction and the evaluation of the topographic pattern of the cone by the surgeon (decentered or not decentered cones). Moreover, the surgical results are influenced by several factors [12], including the geometric assessment of the cornea [11] and the preoperative visual impairment [13]. The structure of the cornea is really unpredictable in keratoconic eyes [11], due to the loss of the well-organized lamellar distribution typical of normal corneas. As a consequence of the degeneration of the corneal tissue, there is not a linear relationship between the thickness of the implanted ring segment and the degree of central corneal flattening. The modeling effect induced by the ICRS in keratoconic eyes may be different from the effect induced in normal corneas and also really different from each other [14]. Another factor that may affect the surgical results is the preoperative visual impairment [13]. Advanced keratoconic corneas are affected by large amounts of corneal higher order aberrations [15], which are closely related to the corneal biomechanical status [16]. The predictability of the postoperative visual outcome is limited if only the preoperative manifest refraction or best corrected visual acuity (BCVA) are considered [17], but can be improved if also corneal

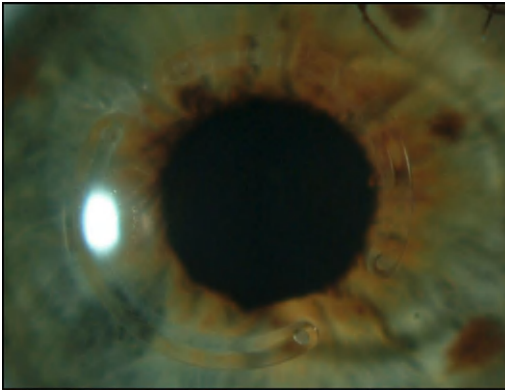


Fig. 15.1 Keraring (Mediphacos)

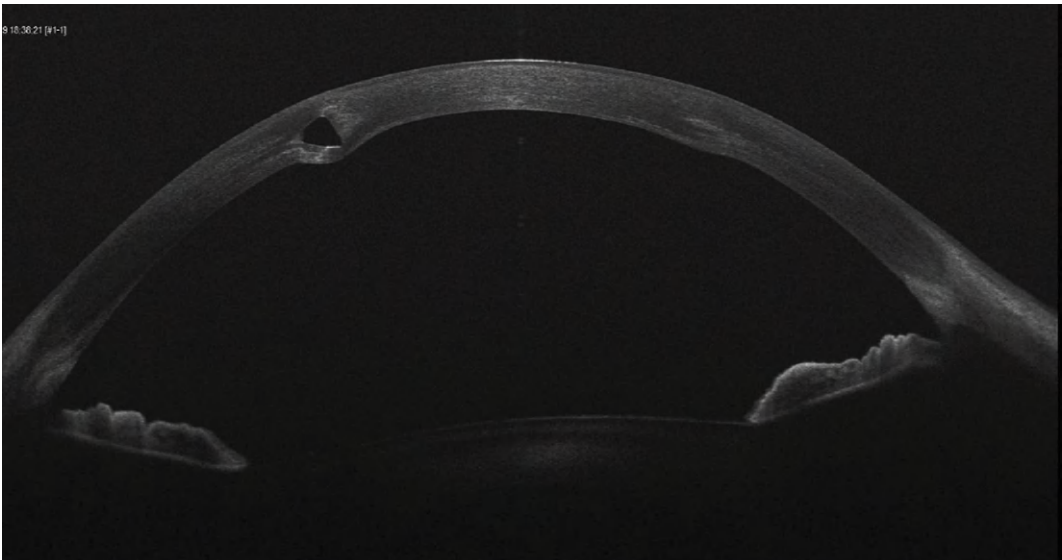


Fig. 15.2 AS-OCT of a Keraring (Mediphacos)

aberrometry is included [16]. Therefore, contrary to the traditional treatment planning based on manufacturer's nomogram, the selection of the ring segment to implant should not be based only on refraction and the subjective appearance of the corneal topographic pattern, but should be based also on corneal aberrometry [18], in order to obtain more predictable results after ICRS implantation in highly distorted corneas.

The need to create a new optimized nomogram for the clinical practice has induced our team to set up a software based on artificial intelligence (Artificial Neural Network Version 1), built only with successful cases of ICRS implantation, where the combinations of rings had induced a gain in BCVA. An artificial neural network (ANN) is a computational program which simulates the decision process in networks of nerve cells of the biological central nervous system [19]. The software receives an input, is able to change its internal state (activation) according to that input, and to produce an output depending on the input and the activation. The software "learns" to perform tasks by considering examples and is able to generate identifying characteristics from the learning material that it processes [20]. The input is represented by previous eyes implanted with ICRS, evaluating the outcomes obtained with different types of rings for a similar ectasia. In particular, the information the neural network needs are the corneal higher order aberrations (preoperative and postoperative, until the seventh Zernike order), the segment position and the segment shape and thickness. *The output consists in the suggestion for the ICRS implantation in a specific patient* (i.e. the type, number, thickness and location of the incision), and the information the software needs are the topography of that patient and the corneal thickness in the place where the segments will be implanted. With all this information, ANN provides a topographic and aberrometric simulation in terms of tangential curvature, sagittal curvature, elevations, Strehl ratio (OPD) and Point Spread Function (PSF) with visual simulation. Predicting the outcome according to the PSF, in fact, allows to simulate the visual outcome by convoluted image of

the ETDRS chart. Thus, ANN simulates which combinations of segments could provide the best topographic outcome, but also the best corneal optical quality, and consequently the best quality of vision, as a function of the Strehl ratio, for the patient.

In 2020 our team has published on Eye and Vision Journal the results of a retrospective, comparative, nonrandomized, pilot, clinical study designed in order to analyze the clinical results of the Artificial Neural Network

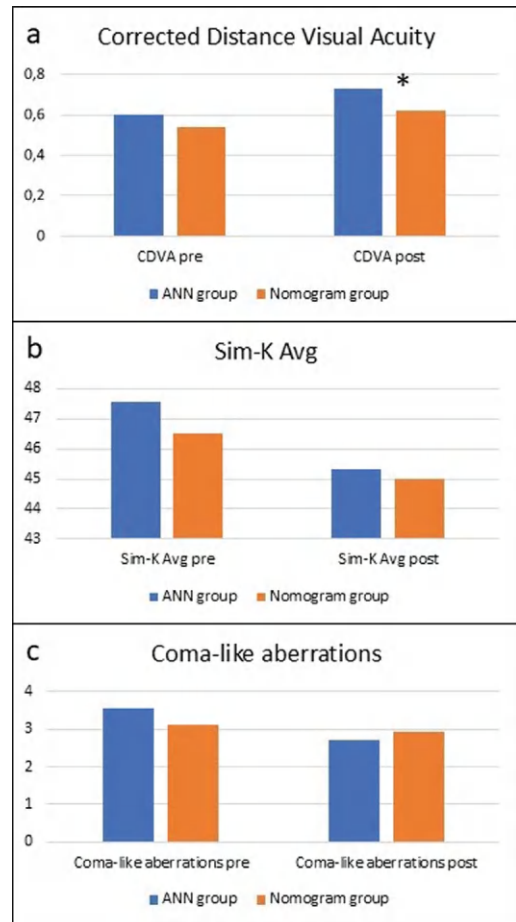


Fig. 15.3 a) Preoperative and postoperative CDVA in the "ANN group" and in the "Nomogram group", with a statistically significant difference (asterisk) in the final visual acuity between the two groups ($p < 0.05$). b) Preoperative and postoperative simulated mean keratometry in the 3-mm central zone in the "ANN group" and in the "Nomogram group". c) Preoperative and postoperative coma-like aberrations in the "ANN group" and in the "Nomogram group"

processed to improve the predictability of the results after ICRS implantation (ANN Version 1) [21]. Corneal topography, tomography, aberrometry, pachymetry and volume analysis were performed with Sirius System (CSO, Firenze, Italy). All ICRS (KeraRing, Mediphacos, Belo Horizonte, Brazil) were implanted using femtosecond laser technology (60-kHz IntraLase femtosecond system: IntraLase Corp, Irvine, California, USA) and the selection of the number (1 or 2), arc length, and thickness of ICRS was performed following the Artificial Neural Network Version 1 (“ANN group”) or the manufacturer’s nomograms (“Nomogram group”). Some patients received only the ICRS implantation, while others underwent the combination of ICRS with corneal collagen CXL, which has been demonstrated to be an effective and safe treatment for keratoconus correction [22–26], with similar improvement rates in visual acuity, spherical and cylindrical errors and mean keratometry values [27]. In both groups, the “ANN group” and the “Nomogram group”, a decrease in keratometric values and spherical equivalent was observed, with an improvement in uncorrected and corrected distance visual acuity; but the eyes implanted according to the ANN Version 1 reached a significantly better CDVA when compared to the cases operated according to the manufacturer’s nomogram. Because of the success or the failure of ICRS implantation is

closely related to the degree of visual limitation [13], and because of the decrease in visual acuity is the real responsible of the visual disability of keratoconus patients [3], the results suggest that the combination of segments chosen by the artificial intelligence, providing a greater improvement in patient’s visual acuity, can better reduce the disability induced by keratoconus. Several previous studies have shown that the regularization of the corneal tissue induced by the ICRS implantation produces a decrease in anterior corneal high order aberrations (HOAs), especially the asymmetric aberrations (coma and coma-like) [10, 13, 28], and that the patients with the largest decrease in aberrations were those with the most advanced disease [13]. In our study, in the “Nomogram group”, a significant reduction in total aberrations was associated to a significant reduction in astigmatism aberrations, which could be justified by the fact that the manufacturer’s nomogram is based on refraction and astigmatism; on the other hand, a not statistically significant reduction in the coma-like aberrations was observed. In the “ANN group”, total anterior corneal aberrations, high order, astigmatism and coma-like aberrations decreased after the ICRS implantation, with, in particular, a statistically significance in the decrease of coma-like aberrations. The artificial intelligence, in fact, works suggesting which combination of segments will provide the best quality of vision, as a function

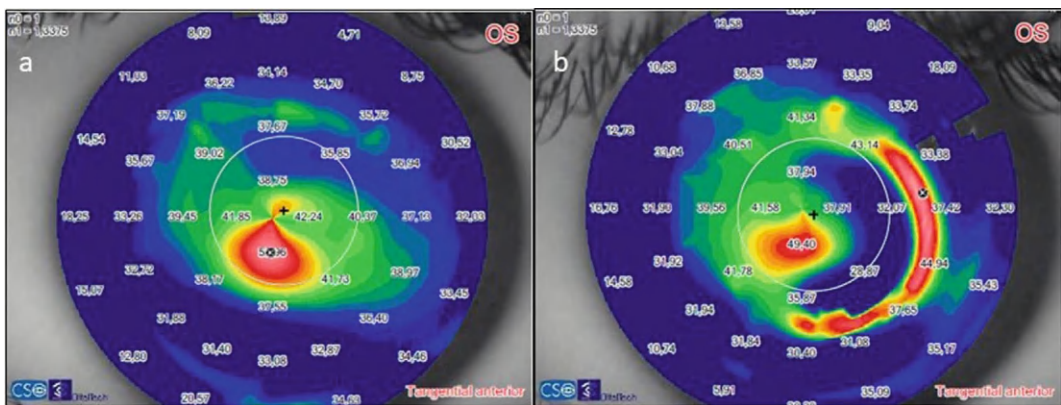


Fig. 15.4 a) Preoperative tangential anterior map of a keratoconic eye of the “ANN group”. b) Postoperative tangential anterior map, after the ICRS implantation following the ANN Version 1 suggestion

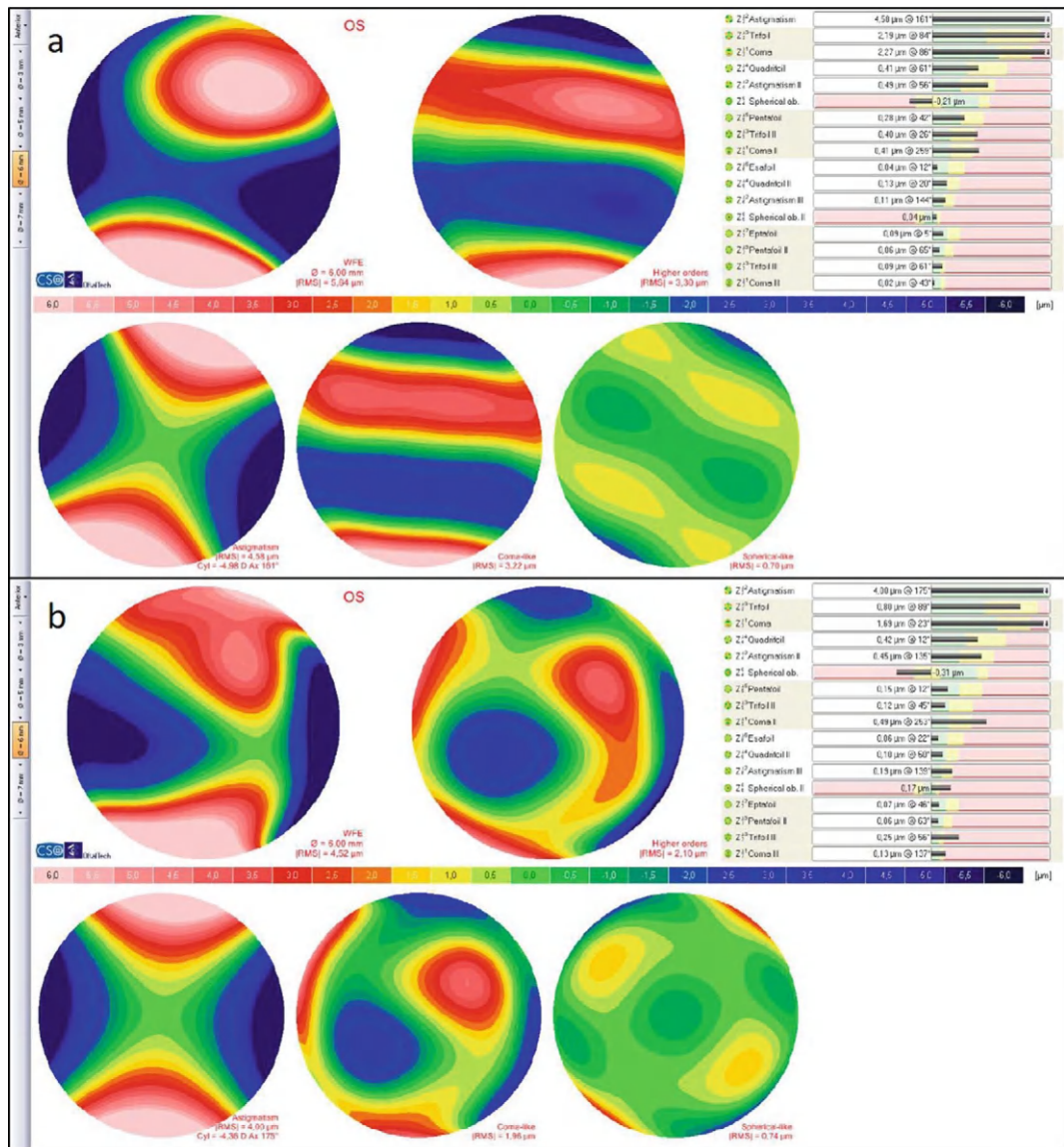


Fig. 15.5 a) Preoperative aberrometry. b) Postoperative aberrometry, with a decrease in high order aberrations and in coma-like aberrations

of the Strehl ratio (OPD), through a reduction in corneal higher order aberrations. As a consequence, the improvement in the optical quality justifies the significantly higher post-operative CDVA found in the ANN group (Figs. 15.3, 15.4, 15.5 and 15.6).

In conclusion, the results of the study suggest that the artificial intelligence is able to simulate the combination of segments which

could provide the best topographic outcome and the best corneal optical quality, through a decrease in corneal high order aberrations, with consequently the best vision for the patient. Furthermore, implanting ICRS according to ANN Version 1 suggestion can improve the predictability of the results. Keratoconic corneas, in fact, are so structurally abnormal that the modeling effect induced by the ICRS implantation is

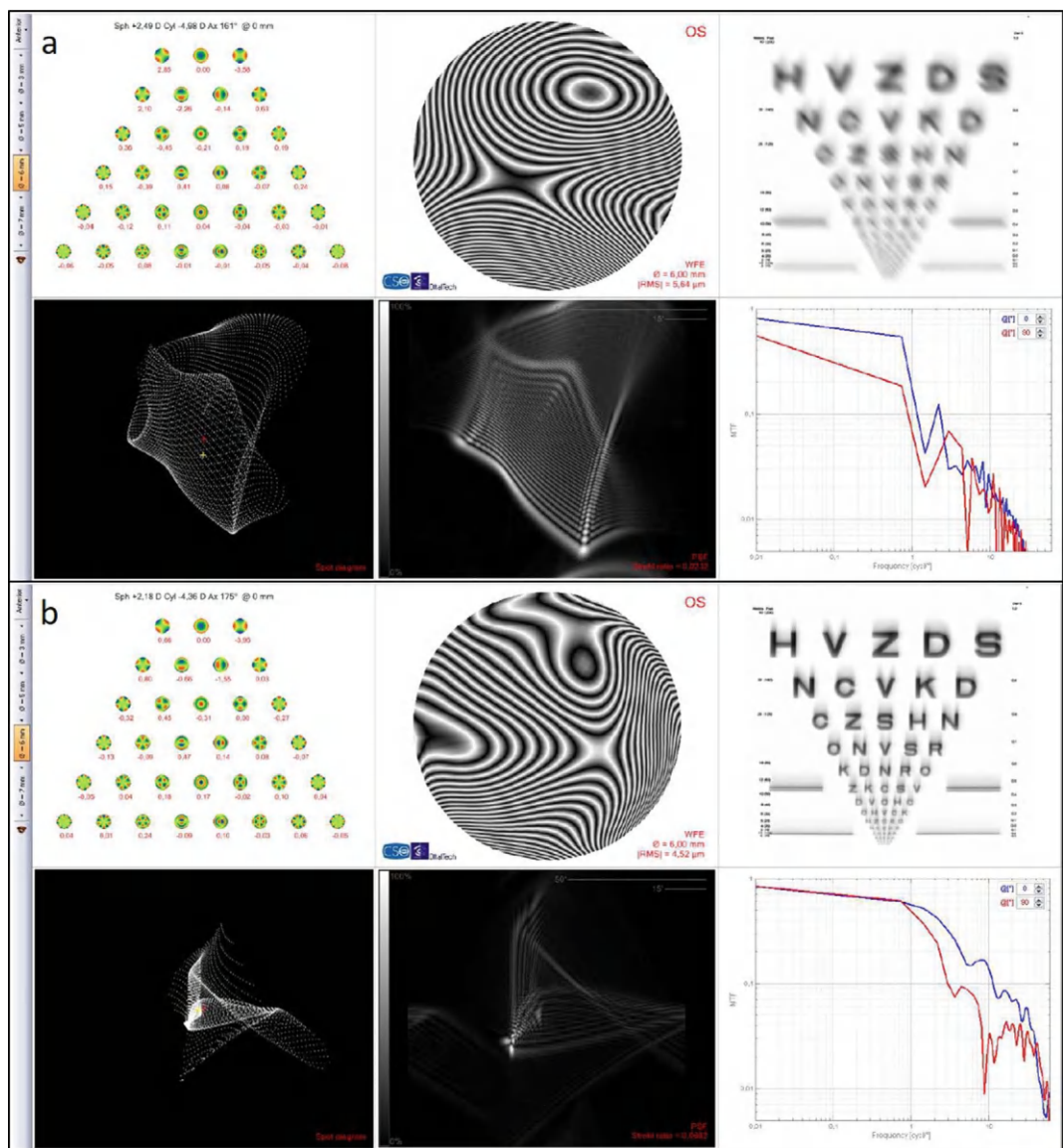


Fig. 15.6 a) Preoperative quality of the retinal image. b) Postoperative quality of the retinal image, with an increase of the Point Spread Function and a significant improvement of the quality of the retinal image

really unpredictable. As a consequence, to plan the surgery upon the results of a large number of previous cases (input) is extremely useful, because the activation of the neural network can

be enhanced by a learning process and increasing the number of cases used as input makes the output more predictable and accurate, improving the surgical outcome.

References

1. Rabinowitz YS. Keratoconus. *Surv Ophthalmol*. 1998;42(4):297–319.
2. Kandel H, Pesudovs K, Watson SL. Measurement of quality of life in keratoconus. *Cornea*. 2020;39:386–93.
3. Alió JL, Piñero DP, Alesón A, Teus MA, Barraquer RI, Murta J, et al. Keratoconus-integrated characterization considering anterior corneal aberrations, internal astigmatism, and corneal biomechanics. *J Cataract Refract Surg*. 2011;37:552–68.
4. Alió JL, Shabayek MH, Artola A. Intracorneal ring segments for keratoconus correction: long-term follow-up. *J Cataract Refract Surg*. 2006;32:978–85.
5. Burris TE, Ayer CT, Evensen DA, Davenport JM. Effects of intrastromal corneal ring size and thickness on corneal flattening in human eyes. *Refract Corneal Surg*. 1991;7:46–50.
6. Ertan A, Colin J. Intracorneal rings for keratoconus and keratectasia. *J Cataract Refract Surg*. 2007;33(7):1303–14.
7. Fleming JF, Wan WL, Schanzlin DJ. The theory of corneal curvature change with the intrastromal corneal ring. *CLAO J*. 1989;15(2):146–50.
8. Roberts CJ, Dupps WJ Jr. Biomechanics of corneal ectasia and biomechanical treatments. *J Cataract Refract Surg*. 2014;40(6):991–8.
9. Peña-García P, Alió JL, Vega-Estrada A, et al. Internal, corneal and refractive astigmatism as prognostic factor for ICRS implantation in mild to moderate keratoconus. *J Cataract Refract Surg*. 2014;40:1633–44.
10. Shabayek MH, Alió JL. Intrastromal corneal ring segment implantation by femtosecond laser for keratoconus correction. *Ophthalmology*. 2007;114(9):1643–52.
11. Colin J, Cochener B, Savary G, Malet F. Correcting keratoconus with intracorneal rings. *J Cataract Refract Surg*. 2000;26(8):1117–22.
12. Flecha-Lescún J, Calvo B, Zurita J, Ariza-Gracia MÁ. Template-based methodology for the simulation of intracorneal segment ring implantation in human corneas. *Biomech Model Mechanobiol*. 2018;17:923–38.
13. Vega-Estrada A, Alió JL, Brenner LF, et al. Outcome analysis of intracorneal ring segments for the treatment of keratoconus based on visual, refractive, and aberrometric impairment. *Am J Ophthalmol*. 2013;155:575–84.
14. Daxer A, Fratzl P. Collagen fibril orientation in the human corneal stroma and its implication in keratoconus. *Invest Ophthalmol Vis Sci*. 1997;38:121–9.
15. Alió JL, Shabayek MH. Corneal higher order aberrations: a method to grade keratoconus. *J Refract Surg*. 2006;22:539–45.
16. Piñero DP, Alió JL, El Kady B, et al. Refractive and aberrometric outcomes of intracorneal ring segments for keratoconus: mechanical versus femtosecond-assisted procedures. *Ophthalmology*. 2009;116:1675–87.
17. Levinger S, Pokroy R. Keratoconus managed with intacs: one-year results. *Arch Ophthalmol*. 2005;123:1308–14.
18. Piñero DP, Alió JL, Teus MA, Barraquer RI, Uceda-Montañés A. Modeling the intracorneal ring segments effect in keratoconus using refractive, keratometric, and corneal aberrometric data. *Invest Ophthalmol Vis Sci*. 2010;51:5583–91.
19. Graupe D. Principles of artificial neural networks. 3rd ed. World Scientific Publishers; 2013.
20. Russell S, Norvig P. Artificial intelligence: a modern approach. Prentice Hall; 2010. p. 578.
21. Fariselli C, Vega-Estrada A, Arnalich-Montiel F, Alió JL. Artificial neural network to guide intracorneal ring segments implantation for keratoconus treatment: a pilot study. *Eye Vis (Lond)*. 2020;9(7):20.
22. Kohlhaas M, Spoerl E, Schilde T, Unger G, Wittig C, Pillunat LE. Biomechanical evidence of the distribution of cross-links in corneas treated with riboflavin and ultraviolet A light. *J Cataract Refract Surg*. 2006;32(2):279–83.
23. Raiskup-Wolf F, Hoyer A, Spoerl E, Pillunat LE. Collagen crosslinking with riboflavin and ultraviolet-A light in keratoconus: long-term results. *J Cataract Refract Surg*. 2008;34(5):796–801.
24. Wollensak G, Spoerl E, Seiler T. Riboflavin/ultraviolet-a-induced collagen crosslinking for the treatment of keratoconus. *Am J Ophthalmol*. 2003;135(5):620–7.
25. Ferenczy PA, Dalcegio M, Koehler M, Pereira TS, Moreira H, Luciane BM. Femtosecond-assisted intrastromal corneal ring implantation for keratoconus treatment: a comparison with crosslinking combination. *Arq Bras Oftalmol*. 2015;78(2):76–81.
26. Ibrahim O, Elmassry A, Said A, Abdalla M, El Hennawi H, Osman I. Combined femtosecond laser-assisted intracorneal ring segment implantation and corneal collagen cross-linking for correction of keratoconus. *Clin Ophthalmol*. 2016;10:521–6.
27. Cakir H, Pekel G, Perente I, Genç S. Comparison of intrastromal corneal ring segment implantation only and in combination with collagen crosslinking for keratoconus. *Eur J Ophthalmol*. 2013;23(5):629–34.
28. Vega-Estrada A, Alió JL, Brenner LF, Burguera N. Outcomes of intrastromal corneal ring segments for treatment of keratoconus: five-year follow-up analysis. *J Cataract Refract Surg*. 2013;39:1234–40.

Artificial Intelligence for Cataract Management

16

Haotian Lin, Xiaohang Wu, Lixue Liu,
and Zizheng Cao

Abstract

Cataracts are the leading cause of blindness worldwide. Early diagnosis and timely treatment of cataracts can improve patients' quality of life and reduce healthcare burdens. However, imbalanced distribution of medical resources has limited the coverage of cataract management. Besides, the current standard of management of cataract patients still needs improvement for better visual outcomes. With the advent of artificial intelligence (AI) technologies, AI applications in medicine and healthcare-related fields have

been intensively investigated in recent years. In contrast to other ocular disorders, cataracts hold promise for the management by AI agents considering their apparently uniform lesion areas and pathological bases. In the following section, we summarize the state-of-the-art AI research in each clinical scenario for cataract patients and give the prospects about AI-based cataract management.

Keywords

Artificial intelligence · Cataract management · Cataract detection · Cataract grading · Postoperative management · Pediatric cataract

H. Lin (✉) · X. Wu · L. Liu · Z. Cao
State Key Laboratory of Ophthalmology, Guangdong
Provincial Key Laboratory of Ophthalmology and
Visual Science, Guangdong Provincial Clinical
Research Center for Ocular Diseases, Zhongshan
Ophthalmic Center, Sun Yat-sen University,
Guangzhou, China
e-mail: linht5@mail.sysu.edu.cn

H. Lin
School of Biomedical Engineering, Sun Yat-sen
University, Guangzhou, China

Hainan Eye Hospital and Key Laboratory of
Ophthalmology, Zhongshan Ophthalmic Center,
Sun Yat-sen University, Haikou, Hainan, China

Centre for Precision Medicine and Department
of Genetics and Biomedical Informatics,
Zhongshan School of Medicine, Sun Yat-sen University,
Guangzhou, Guangdong, China

Artificial Intelligence for Detection and Grading of Cataracts

Cataract was estimated responsible for 52.6 million (18.2–109.6 million) visually impaired patients in 2015 [1]. Considering the association between cataracts and aging, the prevalence of cataracts is expected to increase with the global trend of population aging [2]. However, it's expensive in terms of both time and costs for ophthalmologists to inspect cataracts, particularly in low-income and middle-income

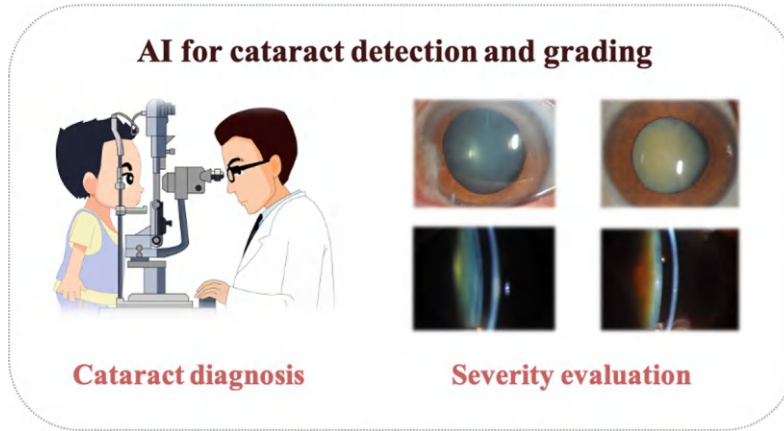


Fig. 16.1 AI for cataract detection and grading

countries given the high interregional disparity of medical resources [3]. Therefore, many researchers have focused on the use of AI for automated identification and grading of cataracts, which can be helpful for large-population screening [4–7] (Fig. 16.1).

With the universalization of cataract surgery, an increasing number of cataract patients have had their eyes operated. Many researchers have included these postoperative eyes in datasets for better differentiation of lens status. Acharya et al. proposed an artificial neural network classifier for the identification of normal, cataract and postoperative slit-lamp photographs in 2010 [4]. Fuzzy K-means clustering algorithm, a kind of traditional machine learning techniques, was applied to detect the features specific to each class. By using 140 photographs, this system exhibited an average rate of 93.3% in detecting different lens condition. Although this system produced acceptable results, it was not appropriate for clinical adoption since it didn't identify referable cases in each class.

A universal artificial intelligence platform for collaborative management of cataracts was later developed by Wu et al. [6]. For each slit-lamp photograph, this AI platform was designed to perform three steps: mode recognition, cataract diagnosis and severity evaluation. After that, a management decision of referral or follow-up was given. The three-step strategy extended the context this AI agent to multiple capture modes

and cataracts of different etiologies and phenotypes. Accordingly, this platform showed great potential to be implemented in clinical scenarios and improve the efficiency as well as coverage of ophthalmology care, which was later proved in this study. In traditional healthcare system, an ophthalmologist was estimated to serve 4000 patients in a year. In the novel tertiary healthcare system proposed by the authors, however, an ophthalmologist could serve as many as 40,806 patients in a year with the assistance of mobile devices, community-based healthcare facilities and this AI platform. The results suggested this collaborative platform and referral pattern were of high efficiency and could be expanded to the management of other ophthalmic diseases, with updated user-accessible mobile devices and automatic examination instruments.

Artificial Intelligence for Preoperative Assessment of Cataract Surgery

The current standardized management of a visually impairing cataract is surgical removal of the cataractous lens and implantation with an intraocular lens instead [8]. Cataract surgery is one of the most commonly performed surgical procedures today, with more than 11 million eyes undergoing intraocular lens (IOL) implantation worldwide every year [9–11]. The postoperative vision of patients depends largely on

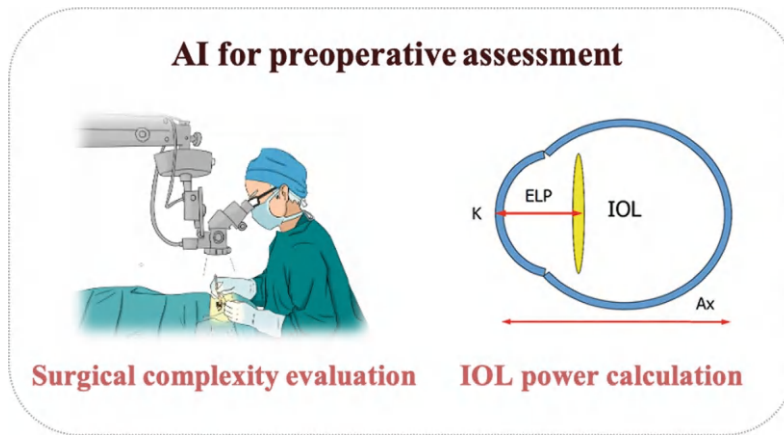


Fig. 16.2 AI for preoperative assessment

an accurate calculation of the IOL optical power, which represents a challenge for ophthalmologists [12]. Researchers have integrated artificial intelligence in this process to achieve better visual outcomes for patients [13, 14] (Fig. 16.2).

Currently, the most commonly used formulas for IOL power calculation are from the Vergence formula category such as SRK/T formula [15]. However, their accuracy is only able to achieve a ± 0.5 diopter from the intended target refraction in 60–80% of eyes. Findl et al. tried improving IOL power calculation through optimization of prediction of the pseudophakic anterior chamber depth (pACD), which is an important yet hard-to-estimate variable in SRK/T formula, one of Vergence formula category [13]. The performance of multilayer perceptron, a type of artificial neural networks, was compared with that of state-of-the-art linear regression for the prediction of pACD from a variety of preoperative biometric parameters. It turned out there was no significant improvement from linear regression to multilayer perceptron in terms of correlation coefficient in prediction of pACD. Replacing the SRK/T pACD prediction with the neural network pACD prediction enhanced performance of the formula by only 0.01 D in mean absolute error in IOL power calculation. This may be a result of the rather small dataset of only 77 eyes. Sramka et al. made another trial in a dataset containing information about up to 2194 eyes [14]. In this

study, IOL power was calculated from clinical data through two machine learning models: Support Vector Machine Regression model and Multilayer Neural Network Ensemble model. The prediction errors of the two machine learning models were significantly lower than those of SRK/T formula, indicating a strong potential of machine learning models for improving IOL calculations.

Artificial Intelligence for Intraoperative Management of Cataract Surgery

Despite as one of the most commonly performed surgical procedures worldwide, cataract surgical techniques vary greatly between surgeons, as do surgical outcomes. The establishment of a consistent evaluation system that enables objective supervision and assessment of performance during cataract surgeries is of utmost importance. Wang et al. developed an intelligent model called DeepSurgery, which utilized the three-dimensional convolutional neural network (3D CNN) algorithm to analyze cataract surgery (CS) videos [16]. This model successfully recognized the chronological order of the 12 surgical steps involved in the cataract surgery procedure, providing timely alerts for surgeons about incorrect workflow sequences. Furthermore, DeepSurgery automatically

quantified and provided real-time scoring for the six major steps of the surgery, including main incision formation, capsulorhexis formation, hydrodissection, phacoemulsification, cortical material removal and IOL implantation. During real-time comparative testing, DeepSurgery exhibited a remarkable step recognition accuracy of 90.30%, showcasing performance on par with the expert panel in surgical scoring (with a kappa range of 0.58–0.77). The development of DeepSurgery paves the way for AI to serve as a training system for novice surgeons, enabling precise assessment of their surgical performance and skills, highlighting the steps that require further improvement. Additionally, it has the potential to establish a standardized and efficient workflow for cataract surgery.

Artificial Intelligence for Postoperative Management of Cataract Surgery

Generally, cataract surgery is safe and effective, with 84–94% of eyes achieving best-corrected visual acuity of 20/30 or better at 6 months after surgery [17, 18]. Still, postoperative complications could arise, the most common one is posterior capsule opacification (PCO) [8]. It is the consequence of proliferation of retained lens

epithelial cells and can cause decreased visual acuity, blurred vision, or glare. Various factors affect the occurrence of PCO, including patient age at the surgery [19], sex [20], diabetes mellitus [21], IOL material [20] and so on. Because of the complex nature of PCO pathogenesis, AI models are expected to be useful in predicting its occurrence (Fig. 16.3).

Mohammadi et al. applied artificial neural network to predict the occurrence of clinically significant PCO 2 years after phacoemulsification [22]. A total of 10 input variables were selected to develop the model. Trained and tested on a dataset of 352 eyes, this model produced a reasonable accuracy of 87%. Another attempt was made by Jiang et al. [23]. They used slit-lamp images rather than variables for the prediction of PCO progression. For each of 1015 patients, slit-lamp images of six consecutive postoperative reexamination stages were collected: the 3rd, 6th, 9th, 12th, 18th and 24th month. A temporal sequence network was then constructed to predict the existence of PCO in 24th month postoperatively. This proposed model offered exceptional performance with a sensitivity of 88.55%, a specificity of 94.31% and an area under curve of 0.9718. Prediction models of this kind help to plan treatment strategies and to provide early warning for the patients.

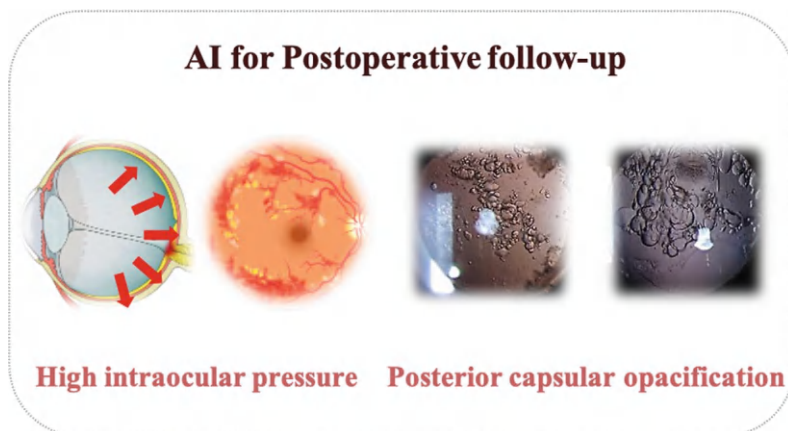


Fig. 16.3 AI for postoperative follow-up

Artificial Intelligence for Pediatric Cataract Management

Congenital/pediatric cataracts (CCs) affect 2.2–13.6 per 10,000 children globally, representing a common congenital anomaly occurring before or during the critical stage of visual development, which has become the leading cause of childhood blindness worldwide [1, 24]. Pediatric cataracts differ from age-related cataracts in three main aspects, which pose new requirements and challenges for artificial intelligence-based disease management. **Firstly**, pediatric cataract is a rare disease with reported incidence of 0.32–22.9 per 10,000 children globally [25]. The scarcity of patients and uneven distribution of data hinder both the accumulation of clinical experience among medical practitioners and training of AI models, resulting in significant challenges for accurate screening and diagnosis of pediatric cataracts. **Secondly**, disparities exist in the clinical presentation of congenital cataract images when compared to age-related cataracts. The diverse degrees of opacification and morphological variations in the lenses of congenital cataracts dictate distinct grading criteria compared to age-related cataracts [26]. As a result, AI models tailored for age-related cataracts are unable to be generalized for application to congenital cataracts. **Furthermore**, the visual impairment caused by congenital cataracts and postoperative complications in children can have long-term consequences. The early assessment of congenital cataracts and the subsequent follow-up plan, which entails significant healthcare resources and can be effectively addressed through the utilization of artificial intelligence. Therefore, many researchers have dedicated to developing an artificial intelligence model and disease management platform for congenital cataracts [27–31].

In underdeveloped regions with limited healthcare resources, congenital cataracts (CC) may not be included in routine neonatal screening. To address this situation, Lin et al. employed random forest (RF) and adaptive boosting methods to establish CC identification

models by incorporating birth conditions, family medical history, and family environmental factors [28]. By conducting a four-fold cross-validation for internal and external validation, the results indicated that the random forest (RF) demonstrates promising receiver operating characteristic (ROC) area under the curve (AUC) values of 0.91 and 0.95, respectively, in the internal and external validation. To evaluate the identification models in a real-world clinical setting with a relatively low prevalence of CC, the CC identification models were tested with gradually decreasing proportions of CC patients, achieving stable performance and maintaining acceptable accuracy and AUC values of 0.92 and 0.96, respectively. Furthermore, RF analysis revealed that family history of CCs, parental education, and comorbidity exhibited the greatest relevance to CCs. The models have the potential to address the deficiency in current CC screening, particularly in underdeveloped areas.

The low prevalence of congenital cataracts can make diagnosis and management difficult for unexperienced doctors. An artificial intelligence platform for the multihospital collaborative management of congenital cataracts was developed by Long et al. [31]. Deep-learning algorithms were utilized to create CC-Cruiser, an AI agent consisting of three functional networks: population screening, risk stratification and decision making for treatment. They conducted the silico test, a multihospital clinical trial, a website-based study, and a ‘finding a needle in a haystack’ test, all of which demonstrated promising accuracy and efficiency. In particular, they conducted a test to compare the real-world performance of CC-Cruiser with that of individual ophthalmologists. The results indicated that the AI agent’s performance surpassed that of competent and novice doctors and was comparable to that of experts. Furthermore, they have established a cloud-based AI platform for rare diseases and put forward an operational mechanism for multi-hospital collaboration. Jiang et al. further explored the multi-center generalizability of infantile cataracts detection [27]. By using deep learning to develop lens partition

strategies (LPSs), the accuracy of the AI agent was improved to 97.35%, and the accuracy of congenital cataract three-degree evaluation increased to 92.86, 92.77 and 91.57%.

Pediatric cataracts have an early onset age, indicating the crucial nature of postoperative follow-up. Therefore, Long et al. additionally developed an intelligent agent specifically designed to address two common complications in patients with congenital cataracts: high intraocular pressure (IOP) and visual axis opacification (VAO). This innovative agent effectively identified individuals at a higher risk of

developing complications, tailored personalized follow-up schedules, and made informed intervention decisions based on telehealth [30]. Their further endeavors involved constructing a CC-Guardian web-based smartphone application and establishing a telehealth cloud platform for clinical implementation, both of which significantly alleviated the socioeconomic burden of patients when compared to traditional in-person follow-up methods. Specifically, patients can save up to 928.6 miles per year, reduce expenditures by \$1324.1 per year, and save 24.9 h per year in travel time (Fig. 16.4).

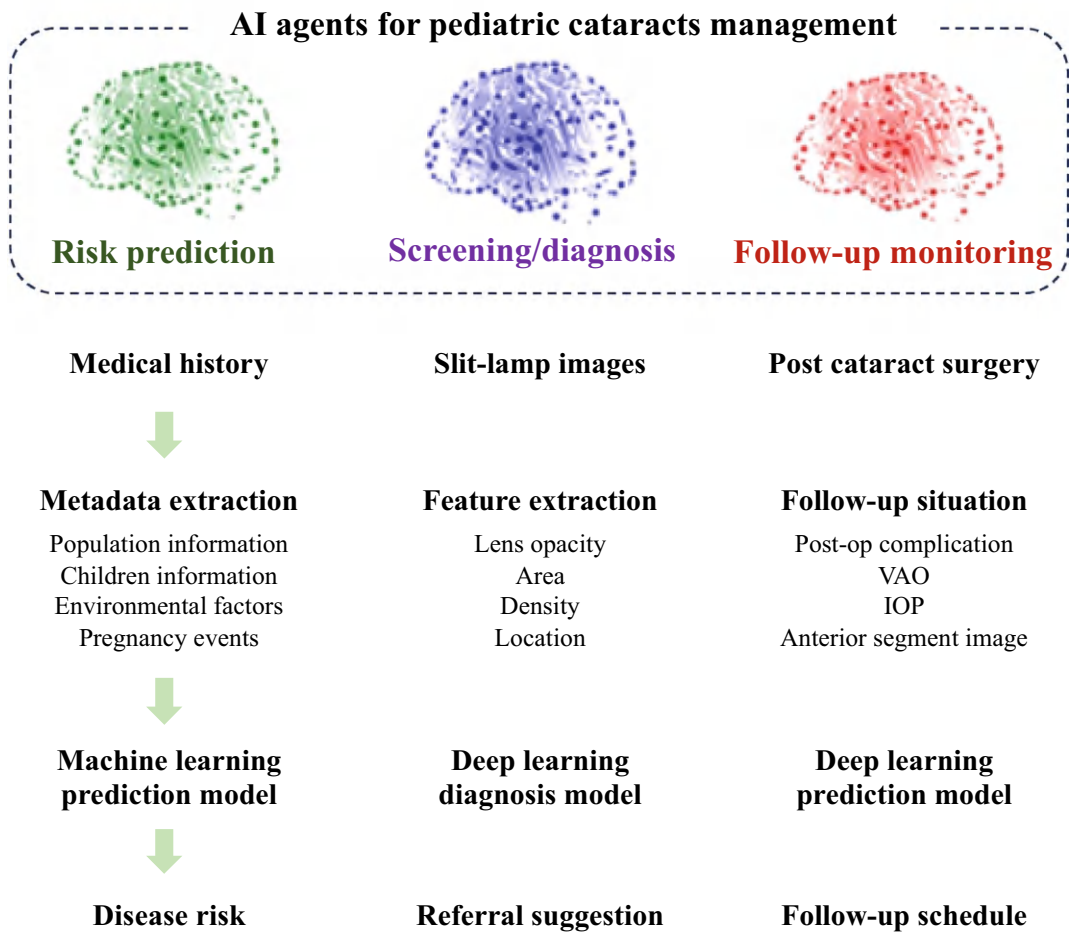


Fig. 16.4 AI for pediatric cataract management

Future of AI-Based Cataract Management

As is discussed above, AI technologies have been applied to multiple aspects of cataract management. From a visionary perspective, AI applications may be extended to other cataract-related areas such as smartphone screening for high-risk populations, as the first step of digital hierarchical referral pattern; prediction of post-operative visual and refractive outcomes; patient assessment for implants such as multifocal or accommodative IOLs. It is becoming apparent that AI technologies will play a crucial role in the revolution of healthcare service delivery and public health economics within the field of ophthalmology. AI will serve as a powerful addition to the current diagnostic and therapeutic strategy for cataracts.

References

1. Flaxman SR, Bourne RRA, Resnikoff S, et al. Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis. *Lancet Glob Health*. 2017;5:e1221–34. [https://doi.org/10.1016/S2214-109X\(17\)30393-5](https://doi.org/10.1016/S2214-109X(17)30393-5).
2. Song P, Wang H, Theodoratou E, et al. The national and subnational prevalence of cataract and cataract blindness in China: a systematic review and meta-analysis. *J Glob Health*. 2018;8: 010804. <https://doi.org/10.7189/jogh.08.010804>.
3. Ramke J, Zwi AB, Lee AC, et al. Inequality in cataract blindness and services: moving beyond unidimensional analyses of social position. *Br J Ophthalmol*. 2017;101:395–400. <https://doi.org/10.1136/bjophthalmol-2016-309691>.
4. Acharya RU, Yu W, Zhu K, et al. Identification of cataract and post-cataract surgery optical images using artificial intelligence techniques. *J Med Syst*. 2010;34:619–28. <https://doi.org/10.1007/s10916-009-9275-8>.
5. Gao X, Lin S, Wong TY. Automatic feature learning to grade nuclear cataracts based on deep learning. *IEEE Trans Biomed Eng*. 2015;62:2693–701. <https://doi.org/10.1109/TBME.2015.2444389>.
6. Wu X, Huang Y, Liu Z, et al. Universal artificial intelligence platform for collaborative management of cataracts. *Br J Ophthalmol*. 2019;103:1553–60. <https://doi.org/10.1136/bjophthalmol-2019-314729>.
7. Xu X, Zhang L, Li J, et al. A hybrid global-local representation CNN model for automatic cataract grading. *IEEE J Biomed Health Inform*. 2020;24:556–67. <https://doi.org/10.1109/JBHI.2019.2914690>.
8. Liu Y-C, Wilkins M, Kim T, et al. Cataracts. *Lancet*. 2017;390:600–12. [https://doi.org/10.1016/S0140-6736\(17\)30544-5](https://doi.org/10.1016/S0140-6736(17)30544-5).
9. Abell RG, Vote BJ. Cost-effectiveness of femtosecond laser-assisted cataract surgery versus phacoemulsification cataract surgery. *Ophthalmology*. 2014;121:10–6. <https://doi.org/10.1016/j.ophtha.2013.07.056>.
10. Frampton G, Harris P, Cooper K, et al. The clinical effectiveness and cost-effectiveness of second-eye cataract surgery: a systematic review and economic evaluation. *Health Technol Assess*. 2014;18(1–205):v–vi. <https://doi.org/10.3310/hta18680>.
11. Wang W, Yan W, Müller A, He M. A global view on output and outcomes of cataract surgery with national indices of socioeconomic development. *Invest Ophthalmol Vis Sci*. 2017;58:3669–76. <https://doi.org/10.1167/iovs.17-21489>.
12. Olsen T. Sources of error in intraocular lens power calculation. *J Cataract Refract Surg*. 1992;18:125–9. [https://doi.org/10.1016/s0886-3350\(13\)80917-0](https://doi.org/10.1016/s0886-3350(13)80917-0).
13. Findl O, Struhal W, Dorffner G, Drexler W. Analysis of nonlinear systems to estimate intraocular lens position after cataract surgery. *J Cataract Refract Surg*. 2004;30:863–6. <https://doi.org/10.1016/j.jcrs.2003.08.027>.
14. Sramka M, Slovak M, Tuckova J, Stodulka P. Improving clinical refractive results of cataract surgery by machine learning. *PeerJ*. 2019;7: e7202. <https://doi.org/10.7717/peerj.7202>.
15. Melles RB, Holladay JT, Chang WJ. Accuracy of intraocular lens calculation formulas. *Ophthalmology*. 2018;125:169–78. <https://doi.org/10.1016/j.ophtha.2017.08.027>.
16. Wang T, Xia J, Li R, et al. Intelligent cataract surgery supervision and evaluation via deep learning. *Int J Surg*. 2022;104: 106740. <https://doi.org/10.1016/j.ijsu.2022.106740>.
17. Ewe SYP, Abell RG, Oakley CL, et al. A comparative cohort study of visual outcomes in femtosecond laser-assisted versus phacoemulsification cataract surgery. *Ophthalmology*. 2016;123:178–82. <https://doi.org/10.1016/j.ophtha.2015.09.026>.
18. Ruit S, Tabin G, Chang D, et al. A prospective randomized clinical trial of phacoemulsification vs manual sutureless small-incision extracapsular cataract surgery in Nepal. *Am J Ophthalmol*. 2007;143:32–8. <https://doi.org/10.1016/j.ajo.2006.07.023>.
19. Apple DJ, Solomon KD, Tetz MR, et al. Posterior capsule opacification. *Surv Ophthalmol*. 1992;37:73–116. [https://doi.org/10.1016/0039-6257\(92\)90073-3](https://doi.org/10.1016/0039-6257(92)90073-3).
20. Ando H, Ando N, Oshika T. Cumulative probability of neodymium: YAG laser posterior capsulotomy after phacoemulsification. *J Cataract Refract Surg*. 2003;29:2148–54. [https://doi.org/10.1016/s0886-3350\(03\)00353-5](https://doi.org/10.1016/s0886-3350(03)00353-5).

21. Ebihara Y, Kato S, Oshika T, et al. Posterior capsule opacification after cataract surgery in patients with diabetes mellitus. *J Cataract Refract Surg.* 2006;32:1184–7. <https://doi.org/10.1016/j.jcrs.2006.01.100>.
22. Mohammadi S-F, Sabbaghi M, Z-Mehrjardi H, et al. Using artificial intelligence to predict the risk for posterior capsule opacification after phacoemulsification. *J Cataract Refract Surg.* 2012;38:403–8. <https://doi.org/10.1016/j.jcrs.2011.09.036>.
23. Jiang J, Liu X, Liu L, et al. Predicting the progression of ophthalmic disease based on slit-lamp images using a deep temporal sequence network. *PLoS ONE.* 2018;13: e0201142. <https://doi.org/10.1371/journal.pone.0201142>.
24. X W, E L, H L, Y L. Prevalence and epidemiological characteristics of congenital cataract: a systematic review and meta-analysis. *Sci Rep.* 2016;6. <https://doi.org/10.1038/srep28564>.
25. Sheeladevi S, Lawrenson JG, Fielder AR, Suttle CM. Global prevalence of childhood cataract: a systematic review. *Eye.* 2016;30:1160–9. <https://doi.org/10.1038/eye.2016.156>.
26. Amaya L, Taylor D, Russell-Eggitt I, et al. The morphology and natural history of childhood cataracts. *Surv Ophthalmol.* 2003;48:125–44. [https://doi.org/10.1016/s0039-6257\(02\)00462-9](https://doi.org/10.1016/s0039-6257(02)00462-9).
27. Jiang J, Lei S, Zhu M, et al. Improving the generalizability of infantile cataracts detection via deep learning-based lens partition strategy and multicenter datasets. *Front Med.* 2021;8: 664023. <https://doi.org/10.3389/fmed.2021.664023>.
28. Lin D, Chen J, Lin Z, et al. A practical model for the identification of congenital cataracts using machine learning. *EBioMedicine.* 2020;51: 102621. <https://doi.org/10.1016/j.ebiom.2019.102621>.
29. Lin H, Li R, Liu Z, et al. Diagnostic efficacy and therapeutic decision-making capacity of an artificial intelligence platform for childhood cataracts in eye clinics: a multicentre randomized controlled trial. *EClinicalMedicine.* 2019;9:52–9. <https://doi.org/10.1016/j.eclinm.2019.03.001>.
30. Long E, Chen J, Wu X, et al. Artificial intelligence manages congenital cataract with individualized prediction and telehealth computing. *npj Digit Med.* 2020;3:112. <https://doi.org/10.1038/s41746-020-00319-x>.
31. Long E, Lin H, Liu Z, et al. An artificial intelligence platform for the multihospital collaborative management of congenital cataracts. *Nat Biomed Eng.* 2017;1:0024. <https://doi.org/10.1038/s41551-016-0024>.

Artificial Intelligence in Refractive Surgery

17

Yan Wang, Haohan Zou, Pinghui Wei,
Mohammad Alzogool, and Xuan Chen

Abstract

Artificial intelligence (AI) has emerged as an indispensable tool in refractive surgery, providing significant support to surgeons across various stages of the procedure. Its applications span preoperative stages (diagnostic screening and patient selection), intraoperative parameter optimization, and postoperative complication prevention. This chapter focuses on the utilization of AI in corneal refractive surgery and implantable collamer lens implantation. With the continuous advancements in AI technology, its role in enhancing the safety, precision, and efficiency of refractive surgery is expected to become increasingly pivotal. The ultimate objective is to achieve personalized and accurate vision correction for each

individual, tailoring treatments to their specific requirements and optimizing long-term visual outcomes. The incorporation of AI into refractive surgery represents a promising pathway towards realizing this vision.

Keywords

Artificial intelligence · LASIK · SMILE · ICL · Refractive surgery · Assisted treatment · Precision medicine · Coupling effects

Refractive surgery has undergone rapid advancements in the last decades with good visual effects and long-term safety [1]. There are several refractive surgery types available both in the cornea and lens. The present chapter focuses on application of AI in corneal refractive surgery and implantable collamer lens implantation. Corneal refractive surgery with laser vision correction is most popular worldwide, including the earliest application of excimer laser photorefractive keratectomy (PRK), the most popular laser refractive surgery performed laser-assisted in situ keratomileusis (LASIK), and small incision lenticule extraction (SMILE) also known as Keratorefractive Lenticule Extraction (KLEx) that use femtosecond laser. With constantly updating surgical technology and ophthalmic

Y. Wang (✉) · H. Zou · P. Wei · M. Alzogool · X. Chen
Tianjin Eye Hospital, Tianjin Key Lab of Ophthalmology and Visual Science, Tianjin Eye Institute, Nankai University Affiliated Eye Hospital, Tianjin, China
e-mail: wangyan7143@vip.sina.com

Y. Wang · H. Zou · P. Wei · M. Alzogool · X. Chen
Nankai University Eye Institute, Nankai University, Tianjin, China

Y. Wang
Clinical College of Ophthalmology, Tianjin Medical University, Tianjin, China

examination equipment, more clinical data related to the surgery are being generated and more accuracy for the preoperative assessment, screening, and surgical type determination are required. Therefore, the use of artificial intelligence (AI) to assist diagnosis and surgical design is indispensable.

Application of AI in Refractive Surgery

Application of AI for Surgical Screening and Suspected Keratoconus Detection

Surgical safety is the basis for refractive surgery development worldwide. Screening the susceptible populations of keratoconus or keratectasia is an important step before surgery. Due to the occult onset of keratoconus, its early stages are often difficult to detect. Furthermore, it greatly affects the patient's vision with a risk of blindness; thus far, clinical diagnosis remains difficult.

Accordingly, many ways from different perspectives have proposed the use of machine learning (ML) technology to assist in the research and diagnosis of keratoconus [2]. Mainly includes the following:

(1) AI diagnostic algorithm based on a single detection device

With the continuous updating of corneal topography technology, from the earliest Placido disk imaging topographers based on the anterior corneal surface to the development of tomography devices that can map both the anterior and posterior corneal surfaces, the amount of measured data produced has exponentially increased. Also, from focusing on the anterior corneal surface parameter analysis then changed to focus on the comprehensive analysis of full corneal parameters. By incorporating AI algorithms and producing effective ML models, significant improvement in the diagnostic rate of keratoconus and different classification tasks such as keratoconus, subclinical keratoconus, high astigmatism cornea, and cornea after refractive surgery have been achieved [3].

(2) AI diagnostic algorithm for combined Multimodal data

In order to improve the diagnostic accuracy of keratoconus, keratectasia, and other related diseases, different screening equipment are often combined during the clinical diagnosis process to develop algorithms for multi-source diagnostic methods, which includes anterior segment OCT devices, optical aberration measuring instruments, confocal microscopy, and in vivo measurement of corneal biomechanics. However, because the results of different functional instruments have different meanings, and there are large differences in the results of different instruments with the same functional category, analyzing these inspection parameters is difficult and complicated. To make the model more compatible and robust, and also convenient to use for research and application between different devices and clinics, a retrospective analysis was performed to develop diagnostic algorithms from a single corneal topographic device to the cross-platform data of three various topographic device sources [4]. Realize the analysis and evaluation of maps obtained from a variety of topographic devices. Meanwhile, a diagnostic model combined with corneal biomechanical measurement has also proven to have high diagnostic performance [5, 6].

(3) Improve diagnostic efficiency for suspected patients in different populations with AI algorithms

In order to identify the diagnostic differences among different ethnic groups, different population-based studies have been performed, showing that ethnic origin influences the keratoconus incidence and corneal physiological parameters of people in different regions also show regional distribution characteristics [7]. Recently, our team completed a study among 2000 participants based on the Scheimpflug corneal tomography parameters to establish a model that can diagnose subclinical keratoconus with high accuracy. We used the support vector machine (SVM) and gradient boosted decision tree (GBDT), an iterative ML composed of multiple

decision trees that screens attribute features with larger weights, to construct a subclinical keratoconus diagnosis model, and performed a 10-fold cross-validation to verify the accuracy. The model achieved a 95.53% diagnostic accuracy. The accuracy of the model in distinguishing subclinical keratoconus from the normal cornea was 96.67%, and the accuracy in distinguishing keratoconus from the normal cornea was 98.91%. In particular, suspected patients have a high diagnostic accuracy [8].

(4) **AI in preoperative screening using corneal biomechanics**

The morphological changes in potential high-risk patients (keratoconus suspect, subclinical keratoconus, forme fruste keratoconus) before refractive surgery are atypical, and it is difficult to screen the potential diseases solely based on topography and tomography. Studies have shown that the corneal biomechanics property of high-risk patients is weaker than that of normal corneas; [9, 10] therefore, relying on the assistance of AI to monitor the biomechanical properties of patients can effectively screen out high-risk surgical populations [11]. Tan and Wang et al. [12] analyzed rudimentary corneal dynamic deformation videos, extracting corneal contours to calculate independent biomechanical parameters. They successfully developed an intelligent diagnostic model for keratoconus, achieving exceptional accuracy rates: 99.6% overall accuracy, 99.3% sensitivity, 100% specificity, and 100% precision. Notably, this diagnostic accuracy markedly surpassed that of the Corneal Biomechanical Index (CBI), thereby advancing the sphere of intelligent diagnosis and treatment for keratoconus. Building upon their groundwork, this group delved into more intricate details and observed that A1dArc length and several corneal biomechanical indices exhibit remarkable sensitivity, enabling the detection of early-stage keratoconus, even in cases lacking topographic abnormalities [10]. Furthermore, they employed an AI-based model to embark on localized biomechanical assessments. The study focused on detecting nuanced

biomechanical changes by computing parameters based on 72-pixel points within 1 mm ($576 \text{ px}/8 \text{ mm}=72 \text{ px/mm}$). This innovative approach introduced a novel method for characterizing local corneal biomechanics, distinctly different from the analysis of whole corneal biomechanical information. Most notably, this model capable of early keratoconus (KC) diagnosis with remarkable precision rates: 95.73% precision, 95.61% recall, and a 95.58% F1 score. This achievement stands superior to current widely accepted approaches (ranging from 86.28% to 93.36%, all $P<0.01$), thereby significantly enhancing diagnostic accuracy and mitigating the risk of misdiagnosis [13].

(5) **Finding new medical rules or connections and provides more clinical clues**

In addition to being able to predict and diagnose the diseases, which get the better outcomes than conventional methods. AI can also discover new medical laws or connections that have never been noticed or discovered before. Our study found that the diagnostic features of keratoconus include not only common clinical indexes such as central corneal astigmatism, index of surface variance, asymmetry index, corneal thickness, and posterior corneal surface height (posterior corneal surface elevation), but also clarified the significance of the aspheric parameters for the suspicious keratoconus diagnosis. From a purely biomechanical standpoint, our findings offer clinical physicians a novel avenue for diagnosing keratoconus. The detection of localized corneal biomechanical alterations through the analysis of dynamic deformation videos, coupled with ML, has proven to be instrumental in the early diagnosis of keratoconus. By emphasizing localized biomechanical changes, our approach stands to assist ophthalmologists in promptly identifying early-stage keratoconus, thereby positively impacting patients' vision and facilitating timely interventions for optimal outcomes. Moreover, collecting large sample data from different centers and different populations to establish and train machine-learning algorithms can further enhance the universality and generalizability of diagnostic models [14].

(6) Analysis of images and big data

The emergence of numerous ML models has provided more possibilities for the research and application concerning the assisted diagnosis of keratoconus detection, through testing and comparison of various algorithms to find the optimal model, thereby improving the ability of disease diagnosis. The research mainly focuses on the analysis of images and data. Some of the most commonly used methods are SVM [15], decision tree (DT) [14], multilayer perceptron (MLP), radial basis function (RBFN) [16], and convolutional neural networks (CNN) [17]. Generally, the dataset includes a training set and one or more validation or test sets. The training set is mainly used to build and train the model and the test or validation set is used to evaluate the model. K-fold, leave-one-out, and other cross-validation methods are used to perform proper internal validation of the training dataset, and it is better to use another dataset derived from clinical data to verify it again. However, the validation of most studies is still based on the data set itself and lacks clinical data validation (Fig. 17.1).

AI Application to Improve Surgical Accuracy and Personalized Design

(1) ML algorithm for surgical nomogram

To ensure the accuracy and predictability of corneal refractive surgery, the risk of overcorrection or undercorrection is reduced [18, 19]. Previous nomogram reports of adjusting magnitudes of spherical equivalent and astigmatism before PRK [20] and LASIK [21] have shown that the use of multiple regression analysis to establish a nomogram model that can consider numerous factors including age, diopter, and corneal curvature, improves the accuracy of PRK and LASIK surgery [22]. However, with the development of new surgical techniques and multi-source clinical examination equipment, more factors influencing the outcome of surgery have been observed, such as temperature, humidity [23], wind speed, and air pressure [24]. Unlike PRK and LASIK, the nomogram adjustment of SMILE surgery needs to consider more factors and depends more on the experience of the surgeon. The former analysis of data using only a

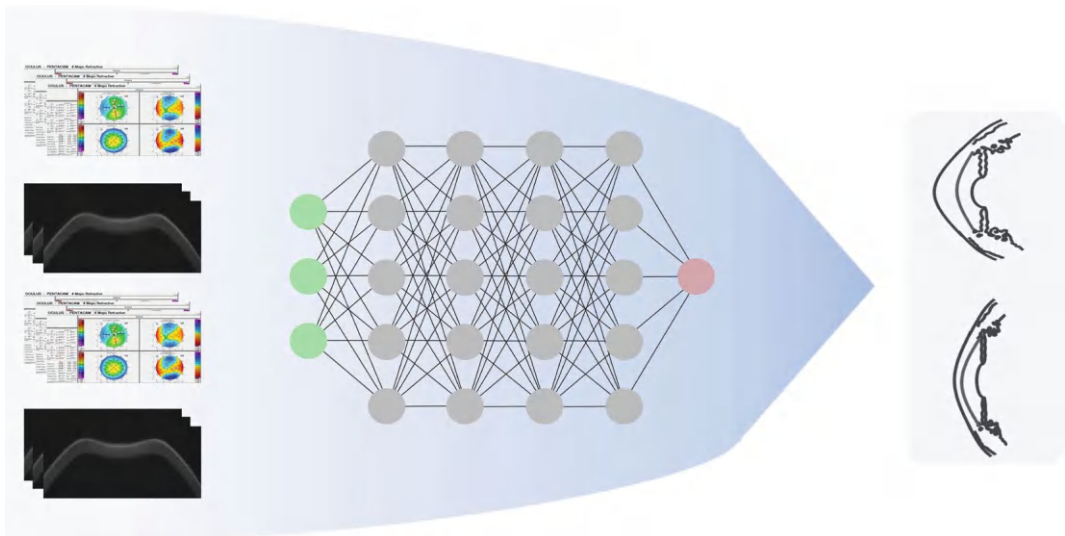


Fig. 17.1 AI assists in the diagnosis of keratoconus and other related ectatic corneal disorders. Input corneal topographic map or corneal morphological parameters, the

model is processed to achieve classification and grading for different cases

single factor or a small sample is insufficient for newly arising needs.

With the help of association analysis, information gain, classifier, and other algorithms, it is possible to realize the correlation between these factors and analyze its impact on refractive surgery. Based on AI technology, research and development of new intelligent refractive surgery platforms assist doctors in completing the entire process from preoperative screening and parameter design to result prediction.

Our team selected 1146 cases that underwent SMILE surgery with ideal postoperative results. From these samples, the nominal features were transformed into binary ones, and the numeric features were normalized into range [0, 1]. The critical features affecting the nomogram values were resolved according to information gain analysis. The multilayer perceptron algorithm was used to train the artificial neural network model to predict the SMILE nomogram and conduct clinical control experiments for validation. Moreover, we compared the outcomes of the surgeon group with the ML group in terms of safety, efficacy, and predictability. Significant results showed that the efficacy index in the ML group (1.48 ± 1.08) was significantly higher than that in the surgeon group (1.3 ± 0.27) ($t = -2.17$, $P < 0.05$), and 83% of the eyes in the surgeon group and 93% of the eyes in the ML group were within ± 0.50 D. The error of SE correction was -0.09 ± 0.024 and -0.23 ± 0.021 for the ML and surgeon groups, respectively. The outcomes of all aspects of the ML group have reached the level of experienced surgeons or even better [25] (Fig. 17.2). This study proves the AI feasibility in the design of refractive surgery treatment strategies. However, it is worth noting that when building a nomogram model, it is not that the more data attributes, the better the model will be. Including too much information will lead to over fitting and model accuracy reduction.

- (2) Analyzing and quantify the influencing factors of surgery

ML extracts and organizes raw data to form valuable feature vectors. Quantifying the importance of different features with the help of weighting analysis or interpretable methods to explain the model output is important for the clinical application of refractive surgery. Analyzing and comparing the clinical data of different surgeons from different surgical centers, screening and comparing the features that have a greater impact on surgery in the data of each center, thus discovering the variability of clinical experience and treatment outcomes. Based on these techniques, it was clarified that preoperative corneal curvature, corneal thickness, and preoperative equivalent spherical lens degree are the main influencing factors of SMILE surgery [26].

- (3) Standardized database construction

In addition, it is necessary to create a worldwide refractive surgery database. The clinical data of refractive surgery is growing at an unprecedented rate all over the world. The establishment of a standardized public database is fundamental for the in-depth AI development in this field, and it is also the key to the development and evaluation of algorithm models. In addition to the multi-center clinical research development, which will greatly enhance the safety and efficiency of the model, reduce misdiagnosis, and miss diagnosis as well as overtreatment during the refractive surgery treatment process, and ensure precision medicine is highly beneficial.

AI Application in ICL Implantation

Implantable Collamer Lens (ICL) implantation has been an alternative for patients who are not suitable candidates for corneal refractive surgery [27–29]. To date, the procedure is widely accepted as safe and effective, with positive visual and optical outcomes [30, 31]. With

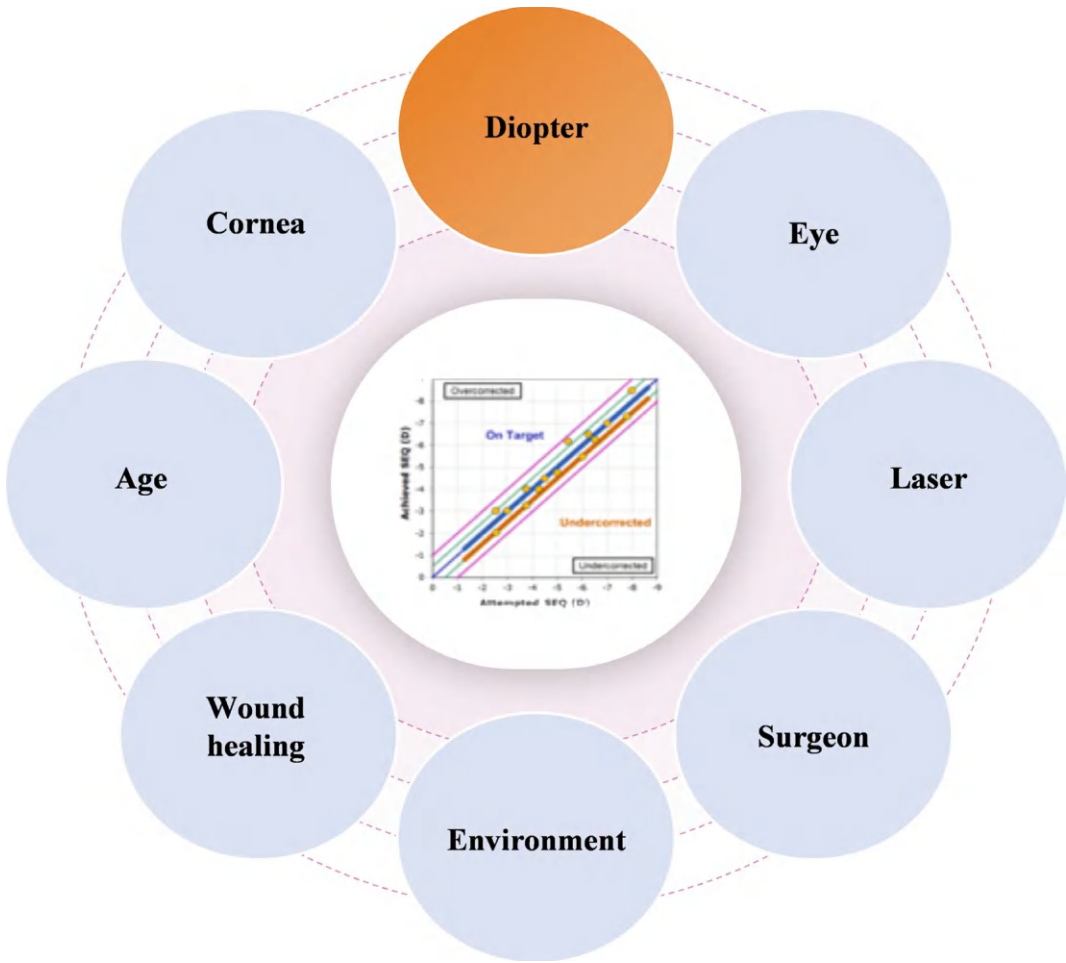


Fig. 17.2 Various factors affect the accuracy and predictability of SMILE surgery outcomes. Using AI can achieve comprehensive analysis and control of influencing factors

the advancement of technology, the use of AI in ICL implantation has significantly improved the precision and safety of the procedure. A crucial factor in achieving satisfactory postoperative results is the accurate calculation of lens power. Several ocular parameters, including anterior chamber depth and corneal curvature, have been identified as factors that could affect postoperative refractive error [32]. Additionally, the position of the ICL after surgery could also have an impact, with higher vault possibly leading to hyperopia and lower vault tending to cause cataract[33]. Therefore, developing a prediction model to improve postoperative refractive correction accuracy is essential. Jiang and his team

developed ML models based on the preoperative dimensional ocular parameters of 1678 patients who underwent ICL implantation [34]. Their ML models showed lower numerical errors and comparable percentages of eyes within certain prediction errors, providing a more accurate lens power calculation.

In addition to ICL lens power, choosing the right size is more crucial for successful ICL implantation and long-term results. Research indicates that approximately 1.2% of individuals may require ICL removal due to improper size selection [35]. AI has been found to be helpful in selecting the appropriate size, with studies showing high accuracy rates

in predicting ICL size and vault height [36, 37]. Combining data from different ophthalmic devices and algorithms could enhance the safety and accuracy of ICL implantation [38]. Ultrasound biomicroscopy (UBM) has particularly shown promise in providing more accurate measurements for predicting postoperative vault and optimal ICL size selection. These findings contribute to the advancement of AI in ophthalmology for improving ICL implantation outcomes. In summary, the application of AI in ICL implantation has greatly enhanced the accuracy and safety of the procedure. AI technology has the potential to further advance and expand its applications in the medical field in the future.

Trends and Challenges of AI in Refractive Surgery

There are many factors affecting refractive surgery, but most of the previous studies predominantly analyzed singular factors. The challenge of addressing various factors has long been difficult to overcome. The application of AI has solved many current real-world intractable problems that are difficult to solve even with ideal experimental models, and refractive surgery is a good example that goes beyond traditional statistical methods to provide a good solution for this type of surgery.

Coupling Effects and Advantages of AI

Cornea has optical, biomechanical, and biological properties. These factors have internal influence on each other when the surgery acts on corneal tissue. We call this phenomenon the coupling effect, that is, there is mutual influence between the optics and mechanics of the human eye, and finally another result, we believe that optics, biomechanics and biological tissue interact. We also call it optical-mechanical-biological coupling. It's a relatively complicated process. At least it is not the result of a single action, there is a process of transformation,

especially corneal refractive surgery. Utilizing the capacity of artificial intelligence to process complex, high-dimensional data holds the potential to surpass the limitations of traditional clinical research reliant on biostatistics and prior knowledge. This approach enables the analysis and discovery of interrelationships and coupled effects among multiple factors, thereby devising more personalized and precise treatment strategies. Particularly, with the continuous generation and accumulation of heterogeneous clinical data sources, including images, audio, text, and more, artificial intelligence demonstrates significant potential in both diagnosis and treatment within the realm of refractive surgery. It also presents notable advantages in outcome prediction and patient management.

Expansion of Application and New Challenges

The use of artificial intelligence in refractive surgery will not only for assisting in diagnosis but also assisting in treatment and health management (Fig. 17.3). There are more and more scenarios for AI application, and the scope of application is gradually expanding. However, AI has been shown to be effective in clinical diagnosis and decision-making, artificial intelligence, particularly deep learning, is still a black box for us until now, which makes it difficult to explain what is going on inside. In the process of using AI to assist decision-making, must be combined with clinical practice and disease characteristics, doctors need a deeper understanding of the relationship between the two, otherwise it will cause a mistake. Meanwhile, large-scale clinical applications also need to consider ethical and legal issues. The interpretation of algorithms and the establishment of relevant specifications will also be a focal issue that needs to be addressed in the future. For the discipline of vision correction, which is inseparable from images and data, refractive surgeons should actively embrace the convenience brought by artificial intelligence to help this discipline develop faster and more accurately.

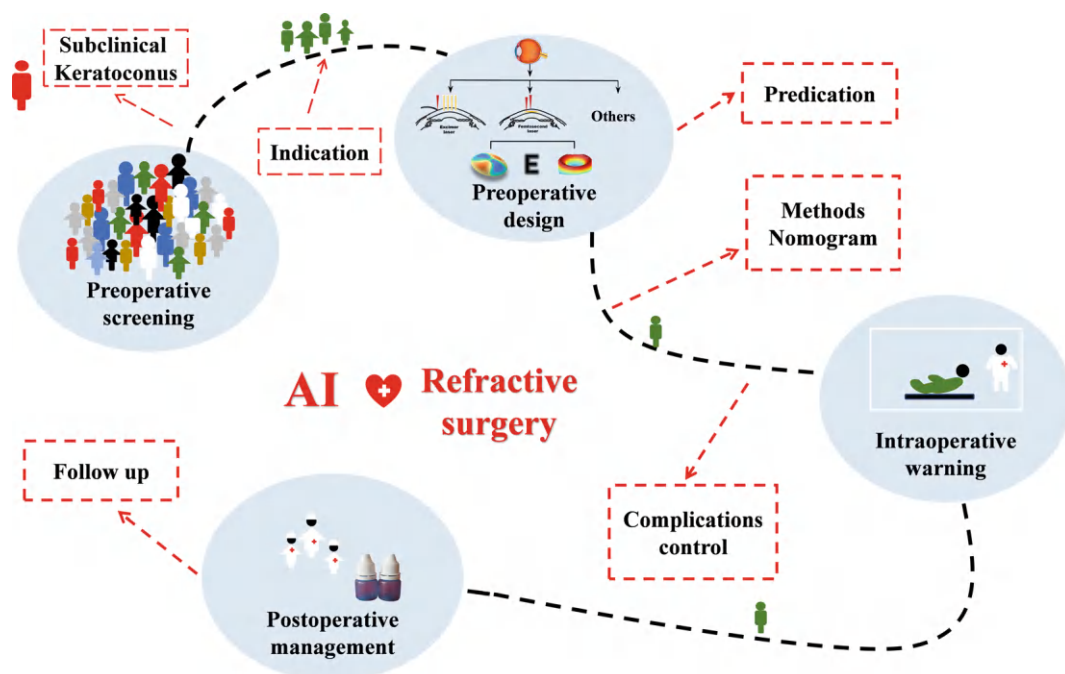


Fig. 17.3 AI runs through the entire process of refractive surgery, assisting physicians in accomplishing all tasks from preoperative screening, surgery design, intraoperative control, and postoperative management

References

- Kim TI, Alio Del Barrio JL, Wilkins M, Cochener B, Ang M. Refractive surgery. *Lancet*. 2019;393(10185):2085–98.
- Lin SR, Ladas JG, Bahadur GG, Al-Hashimi S, Pineda R. A review of machine learning techniques for keratoconus detection and refractive surgery screening. *Semin Ophthalmol*. 2019;34(4):317–26.
- Ruiz Hidalgo I, Rodriguez P, Rozema JJ, Ni Dhubbghaill S, Zakaria N, Tassignon MJ, Koppen C. Evaluation of a machine-learning classifier for keratoconus detection based on Scheimpflug tomography. *Cornea*. 2016;35(6):827–32.
- Mahmoud AM, Roberts C, Lembach R, Herderick EE, McMahon TT, Clek Study G. Simulation of machine-specific topographic indices for use across platforms. *Optom Vis Sci*. 2006;83(9):682–93.
- Machado AP, Lyra JM, Ambrósio R, Ribeiro G, Araújo LPN, Xavier C, Costa E. Comparing machine-learning classifiers in keratoconus diagnosis from ORA examinations. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011. p. 90–5.
- Ambrosio R Jr, Lopes BT, Faria-Correia F, Salomao MQ, Bühren J, Roberts CJ, Elsheikh A, Vinciguerra R, Vinciguerra P. Integration of Scheimpflug-based corneal tomography and biomechanical assessments for enhancing ectasia detection. *J Refract Surg*. 2017;33(7):434–43.
- Ma R, Liu Y, Zhang L, Lei Y, Hou J, Shen Z, Yi X, Wang Y. Distribution and trends in corneal thickness parameters in a large population-based multicenter study of young Chinese adults. *Invest Ophthalmol Vis Sci*. 2018;59(8):3366–74.
- Zou HH, Xu JH, Zhang L, Ji SF, Wang Y. Assistant diagnose for subclinical keratoconus by artificial intelligence. *Zhonghua Yan Ke Za Zhi*. 2019;55(12):911–5.
- Huo Y, Chen X, Cao H, et al. Biomechanical properties analysis of forme fruste keratoconus and subclinical keratoconus. *Graefes Arch Clin Exp Ophthalmol*. 2023;261(5):1311–20.
- Chen X, Cao H, Huo Y, et al. Screening of sensitive in vivo characteristics for early keratoconus diagnosis: a multicenter study. *Front Bioeng Biotechnol*. 2023;11:1158299.
- Huo Y, Chen X, Khan GA et al. Corneal biomechanics in early diagnosis of keratoconus using artificial intelligence. *Graefes Arch Clin Exp Ophthalmol*. 2023.
- Tan Z, Chen X, Wang Y, et al. Artificial intelligence-based diagnostic model for detecting keratoconus using videos of corneal force deformation. *Transl Vis Sci Technol*. 2022;11(9):32.

13. Chen X, Tan Z, Wang Y, et al. Localized corneal biomechanical alteration detected in early keratoconus based on corneal deformation using artificial intelligence. *Asia Pac J Ophthalmol* (Phila). 2023 Nov 15.
14. Lopes BT, Ramos IC, Salomao MQ, Guerra FP, Schallhorn SC, Schallhorn JM, Vinciguerra R, Vinciguerra P, Price FW Jr, Price MO, Reinstein DZ, Archer TJ, Belin MW, Machado AP, Ambrosio R Jr. Enhanced tomographic assessment to detect corneal ectasia based on artificial intelligence. *Am J Ophthalmol*. 2018;195:223–32.
15. Ruiz Hidalgo I, Rozema JJ, Saad A, Gatinel D, Rodriguez P, Zakaria N, Koppen C. Validation of an objective keratoconus detection system implemented in a Scheimpflug tomographer and comparison with other methods. *Cornea*. 2017;36(6):689–95.
16. Souza MB, Medeiros FW, Souza DB, Garcia R, Alves MR. Evaluation of machine learning classifiers in keratoconus detection from orbscan II examinations. *Clinics* (Sao Paulo). 2010;65(12):1223–8.
17. Lavric A, Valentin P. Kerato detect: keratoconus detection algorithm using convolutional neural networks. *Comput Intell Neurosci*. 2019;2019:8162567.
18. Jin HY, Wan T, Wu F, Yao K. Comparison of visual results and higher-order aberrations after small incision lenticule extraction (SMILE): high myopia vs. mild to moderate myopia. *BMC Ophthalmol*. 2017;17(1):118.
19. Zhang J, Wang Y, Wu W, Xu L, Li X, Dou R. Vector analysis of low to moderate astigmatism with small incision lenticule extraction (SMILE): results of a 1-year follow-up. *BMC Ophthalmol*. 2015;15:8.
20. Shapira Y, Vainer I, Mimouni M, Sela T, Munzer G, Kaiserman I. Myopia and myopic astigmatism photorefractive keratectomy: applying an advanced multiple regression-derived nomogram. *Graefes Arch Clin Exp Ophthalmol*. 2019;257(1):225–32.
21. Moniz N, Fernandes ST. Nomogram for treatment of astigmatism with laser in situ keratomileusis. *J Refract Surg*. 2002;18(3 Suppl):S323–326.
22. Liyanage SE, Allan BD. Multiple regression analysis in myopic wavefront laser in situ keratomileusis nomogram development. *J Cataract Refract Surg*. 2012;38(7):1232–9.
23. Seider MI, McLeod SD, Porco TC, Schallhorn SC. The effect of procedure room temperature and humidity on LASIK outcomes. *Ophthalmology*. 2013;120(11):2204–8.
24. Neuhaus-Richard I, Frings A, Ament F, Görsch IC, Druchkiv V, Katz T, Linke SJ, Richard G. Do air pressure and wind speed influence the outcome of myopic laser refractive surgery? Results from the Hamburg weather study. *Int Ophthalmol*. 2014;34(6):1249–58.
25. Cui T, Wang Y, Ji S, Li Y, Hao W, Zou H, Jhanji V. Applying machine learning techniques in nomogram prediction and analysis for SMILE treatment. *Am J Ophthalmol*. 2020;210:71–7.
26. Liang S, Ji S, Liu X, Wang Y. Applying information gain to explore factors affecting small-incision lenticule extraction: a multicenter retrospective study. *Front Med* (Lausanne). 2022;9: 837092.
27. Pazo EE, McNeely RN, Arba-Mosquera S, Palme C, Moore JE. Unilateral ectasia after small-incision lenticule extraction. *J Cataract Refract Surg*. 2019;45(2):236–41.
28. Galvis V, Villamil JF, Acuna MF, Camacho PA, Merayo-Llodes J, Tello A, Zambrano SL, Rey JJ, Espinoza JV, Prada AM. Long-term endothelial cell loss with the iris-claw intraocular phakic lenses (Artisan(R)). *Graefes Arch Clin Exp Ophthalmol*. 2019;257(12):2775–87.
29. Jiao XL, Li J, Yu Z, Wei PH, Song H. Comparison of the visual performance of iris-fixed phakic lens and implantable collamer lens to correct high myopia. *BMC Ophthalmol*. 2021;21(1):244.
30. Chen X, Shen Y, Xu H, Wang X, Zhou X. One-year natural course of corneal densitometry in high myopic patients after implantation of an implantable collamer lens (model V4c). *BMC Ophthalmol*. 2020;20(1):50.
31. Kamiya K, Shimizu K, Igarashi A, Kitazawa Y, Kojima T, Nakamura T, Oka Y, Matsumoto R. Posterior chamber phakic intraocular lens implantation: comparative, multicentre study in 351 eyes with low-to-moderate or high myopia. *Br J Ophthalmol*. 2018;102(2):177–81.
32. Sayed KM, Alsamman AH. Interchangeability between Pentacam and IOLMaster in phakic intraocular lens calculation. *Eur J Ophthalmol*. 2015;25(3):202–7.
33. Lee J, Kim Y, Park S, Bae J, Lee S, Park Y, Lee J, Lee JE. Long-term clinical results of posterior chamber phakic intraocular lens implantation to correct myopia. *Clin Exp Ophthalmol*. 2016;44(6):481–7.
34. Jiang Y, Shen Y, Chen X, Niu L, Li B, Cheng M, Lei Y, Xu Y, Wang C, Zhou X, Wang X. Artificial intelligence-based refractive error prediction and EVO-implantable collamer lens power calculation for myopia correction. *Eye Vis* (Lond). 2023;10(1):22.
35. Kamiya K, Igarashi A, Hayashi K, Negishi K, Sato M, Bissen-Miyajima H. A multicenter prospective cohort study on refractive surgery in 15,011 eyes. *Am J Ophthalmol*. 2017;175:159–68.
36. Russo A, Filini O, Savini G, Festa G, Morescalchi F, Boldini A, Semeraro F. Predictability of the vault after implantable collamer lens implantation using OCT and artificial intelligence in White patient eyes. *J Cataract Refract Surg*. 2023;49(7):724–31.
37. Shen Y, Wang L, Jian W, Shang J, Wang X, Ju L, Li M, Zhao J, Chen X, Ge Z, Wang X, Zhou X. Big-data and artificial-intelligence-assisted vault prediction and EVO-ICL size selection for myopia correction. *Br J Ophthalmol* 2021.
38. Chen X, Ye Y, Yao H, Liu C, He A, Hou X, Zhao K, Cui Z, Li Y, Qiu J, Chen P, Yang Y, Zhuang J, Yu K. Predicting post-operative vault and optimal implantable collamer lens size using machine learning based on various ophthalmic device combinations. *Biomed Eng Online*. 2023;22(1):59.

Artificial Intelligence in Cataract Surgery Training

18

Nouf Alnafisee, Waverly Rose Brim, Sidra Zafar,
Bassell Hammoud, Kristen Park, S. Swaroop Vedula,
and Shameema Sikder

Abstract

Surgery is a major source of error in patient care. Preventing complications from surgical errors in the general operating room is estimated to lead to the reduction of up to 41,846 readmissions and save \$620.3 million per year. It has been established that poor technical skill is associated with an increased risk of severe adverse events after surgery, and traditional models to train surgeons are being challenged by rapid advances

in technology, an intensified patient-safety culture, and a need for value-driven health systems. Cataract surgery is one of the most common surgical procedures performed across the world, and though this surgery has distinctive benefits such as improved vision and a reduced risk of falls, failing to learn the procedure to mastery is consequential. This chapter discusses the current methods available for evaluating technical skill in cataract surgery as well as the recent technological advancements that have enabled the capture and analysis of large amounts of complex surgical data for automated objective assessment of surgical skills.

N. Alnafisee
Faculty of Biology, Medicine and Health,
University of Manchester, Manchester, UK

W. R. Brim
Johns Hopkins University School of Medicine,
Baltimore, MD, USA

S. Zafar · S. Sikder (✉)
Wilmer Eye Institute, Johns Hopkins University
School of Medicine, Baltimore, MD, USA
e-mail: ssikder1@jhmi.edu

B. Hammoud
Cole Eye Institute, Cleveland Clinic, Cleveland,
OH, USA

K. Park
School of Medicine, University of Pennsylvania,
Philadelphia, PA, USA

S. Swaroop Vedula · S. Sikder
Malone Center for Engineering in Healthcare,
Johns Hopkins University Whiting School
of Engineering, Baltimore, MD, USA

Keywords

Artificial intelligence · Cataract surgery ·
Surgical skill · Surgical training

Cataract surgery is one of the most common surgical procedures performed across the world. It is estimated that by 2050, almost 50 million individuals will need cataract surgery in the United States (U.S.) alone [1]. Given the growing incidence of cataracts, surgical training has become an increasingly important component of ophthalmology residency, and failing to learn the procedure to mastery can be consequential.

Moreover, surgeons must often sustain high volumes of procedures even after Board certification to retain their skill in cataract surgery, minimize the risk of complications, and optimize patient care.

In addition to the six core competencies mandated by the Accreditation Council for Graduate Medical Education, the American Board of Ophthalmology (ABO) has also identified surgical proficiency as a competence that should be met by ophthalmology training programs [2]. However, there is minimal guidance from certifying agencies on the best methods to assess for surgical skill [3]. In the case of cataract surgery, graduating ophthalmology residents in the United States are required to complete a minimum of 86 cataract surgeries as primary surgeon [4]. A case log, however, is both insufficient for assessing competence and not useful for providing feedback to the surgeon. Other conventional methods for skill assessment including direct observation and subjective evaluation in a traditional apprenticeship model of teaching surgery lack standardization and objectivity. As an alternative to subjective evaluation, structured rating scales for skill assessment have been developed in recent years [5]. However, feedback from structured rating scales is typically not timely or incomplete. Therefore, new approaches to reliable, valid, and universally accessible methods to assess skill are necessary to improve care for cataract surgery patients.

Recent advances in surgical data science have enabled novel methods for surgical skill assessment. These include techniques to directly analyse instrument motion, video, or other data from the surgical field as well as manual approaches such as crowdsourcing. However, the validity of these approaches for skill assessment in the operating room (OR) remains to be seen [6, 7]. This brings us to the following questions: (1) Where does ophthalmology stand in relation to other surgical specialties regarding the use of new techniques for the assessment of surgical technical skill? (2) What do we have to do to bridge this gap and begin implementing automated methods for technical skill

assessment in ophthalmology training, specifically in cataract surgery?

Broadly, measures of technical skill in cataract surgery may be derived from three data sources: (I) surgeon performance data, (II) the usage of a phacoemulsification machine or other devices [8], and (III) clinical outcomes [9–11] (Fig. 18.1). Figure 18.2 illustrates different sources of data, the ease with which they can be obtained, and current methods to use them to assess technical skill. Each source of data and the methods that have been utilized to analyze them for technical skill are discussed below.

Sources of Data for Technical Skill Assessment in Cataract Surgery

I. Surgeon performance data

(a) Direct observation

The most prevalent and the most subjective method of assessment involves the direct observation of the surgeon performing the surgery in the operating room. Direct observation currently plays a critical role in training surgeons because it allows for feedback both during and after the procedure. However, it requires the presence of an experienced surgeon and/or educator and is not readily reproducible, owing to large inter-observer variation in their assessments and the feedback they provide.

(b) Instrument motion data (sensor-based)

Motion analysis techniques can be used to assess a surgeon's dexterity based on the movements of their hands, fingers, or instruments. To date, sensor-based motion analysis in ophthalmology has only been used for assessing skill in corneal suturing [12], oculoplastics [13], and for guiding robot-assisted membrane peeling in vitreoretinal surgeries [14]. Motion analysis has been shown to stratify performance based on surgical experience with more experienced surgeons using less time, movement, and distance to complete the given task [12, 13]. However,

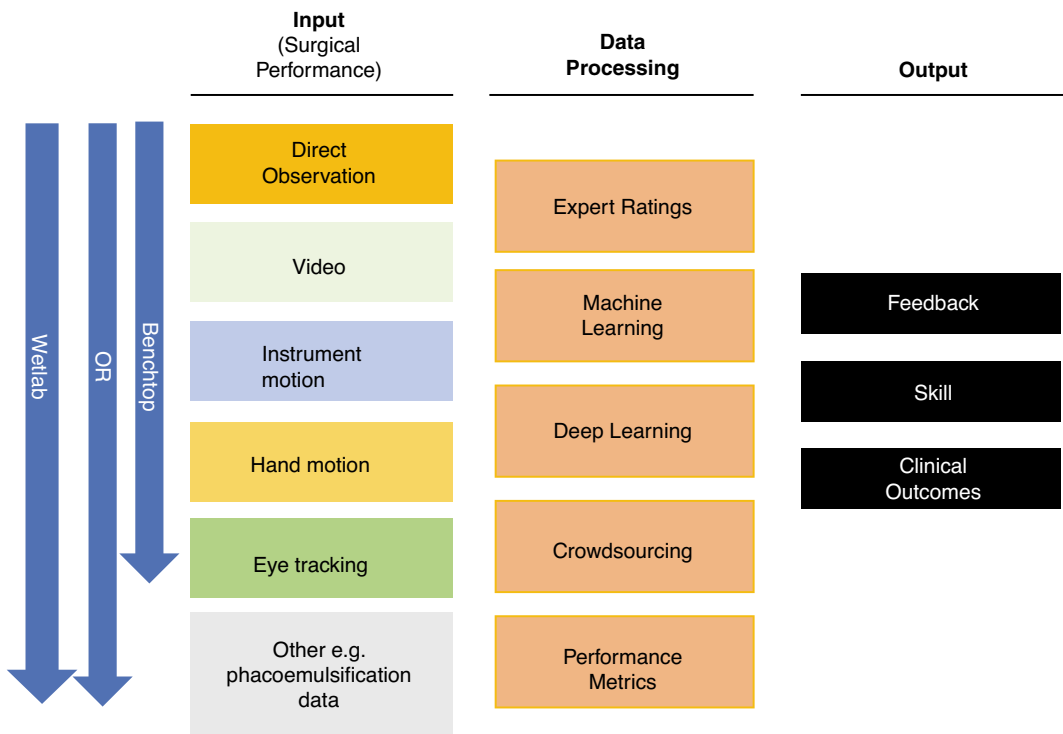


Fig. 18.1 An overview of the scope for sources of data for surgical skill assessment and the training context in which they may be obtained

to date, this has only been applied in a wet lab environment, and the first-year costs of the required equipment sum to \$19,060 [15].

(c) Video-based

Video recordings are common during cataract surgery and can be a rich source of data for analyzing surgeon performance. Phacotracking [15], a computer vision-based method, can measure instrument motion during cataract surgery to provide objective metrics of performance. A study by Smith et al. analyzed 20 intraoperative surgical videos and determined metrics provided by this methodology (the total time taken, the total path length, and the number of movements made by the surgeon) to successfully discriminate between expert and novice surgeons [16]. Similar results were also reported by Din et al. and Balal et al. [17, 18]. Balal et al. observed higher levels of variability among junior surgeons and updated the previous tracking method by adding stable feature points to the

frames and tracking their movement to correctly identify surgical instrument movements.

II. Surgical device usage

Ogawa et al. developed the Surgical Media Center (SMC) (Abbott Medical Optics, Inc.), a cataract surgery recording device capable of measuring changes in multiple objective parameters, including phacoemulsification power, vacuum level, aspiration rate over time, and foot pedal position [8]. By doing so, the SMC can detect inappropriate phacoemulsification techniques by analyzing usage over time in both OR and wet lab settings. Ogawa et al. found significant differences in performance metrics measured by the SMC between expert (n=3) and novice surgeons (n=3) [8]. Findings from the study suggest that the time to reach maximum vacuum and the speed of increase in vacuum may be regarded as indicators for the skillfulness of handling the phacoemulsification device [8] Additional tools available include *Touch*

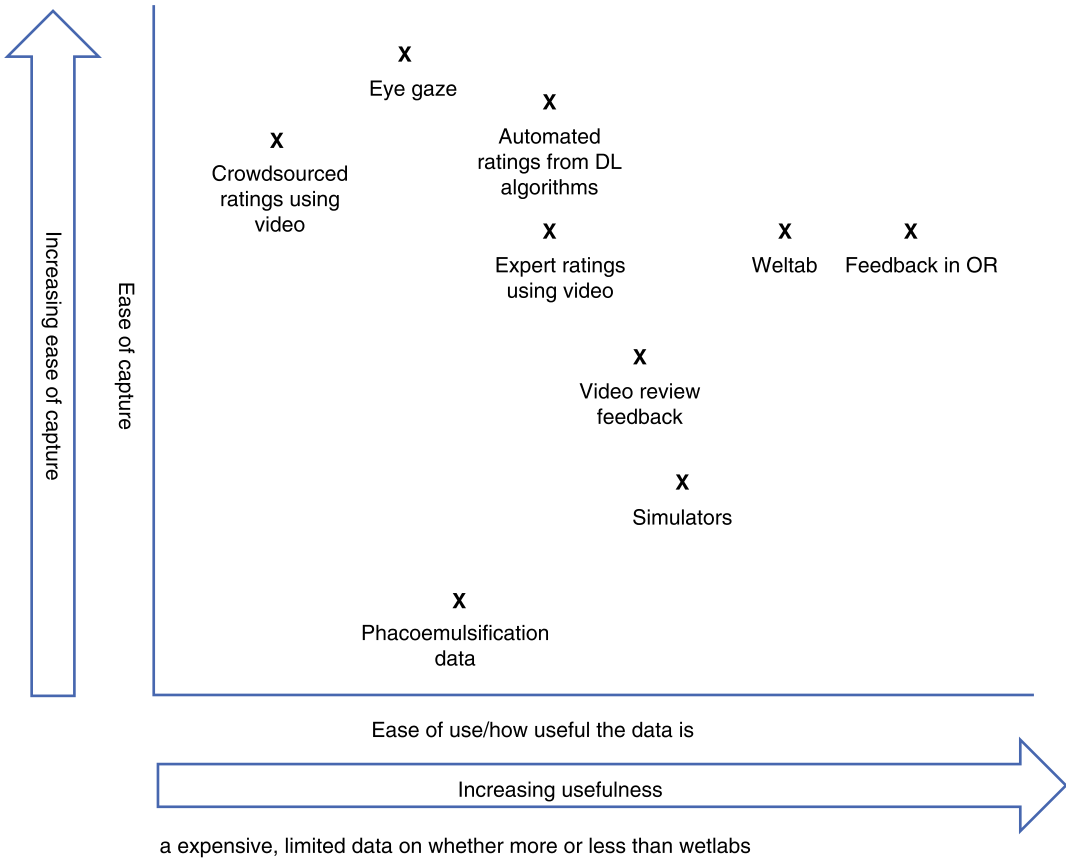


Fig. 18.2 Overview of different sources of data, ease with which they can be obtained, and access to current methods to use them to assess technical skill

Surgery Enterprise (Medtronic, Inc.), which is a video management and analytics platform that also allows recording of surgeries, uploading them to cloud storage, and the application of AI-powered analytics to identify phases and provide insights on surgical workflow [19].

III. Clinical outcomes

The technical skill of the operating surgeon has been proposed to be an important determinant of postoperative outcomes [20]. While tracking resident complication rates may allow for monitoring progress and ensuring patient safety, the rates of clinically significant complications, such as posterior capsule rupture and vitreous loss, remain low. Furthermore, outcomes may not be readily mapped to specific feedback that can

drive learning and improvements in surgeons' performance.

Approaches to Surgical Skill Assessment

(a) Structured rating scales

Structured rating scales are one of the earliest attempts to bring objectivity to surgical skill assessment. The scales are structured, which allows raters to provide consistent assessments and feedback. However, the rating scales are completed by humans, who are subject to rater effects from several sources. Rater effects introduce subjectivity and variation across raters.

Multiple structured rating scales are presently available for the assessment of technical skill in cataract surgery. These include the Objective Assessment of Skills in Intraocular Surgery (OASIS) [21], Global Rating Assessment of Skills in Intraocular Surgery (GRASIS) [22], Subjective Phacoemulsification Skills Assessment (SPESA) [23], Objective Structured Assessment of Cataract Surgical Skill (OSACSS) [24], International Council of Ophthalmology–approved Ophthalmology Surgical Competency Assessment Rubrics (ICO-OSCAR) [25, 26], and the Objective Structured Assessments of Technical Skills (OSATS) [27]. Although rubrics remain the most affordable assessment method with an annual cost of approximately \$4000 [15], their widespread use and implementation has been limited because it takes time and expert surgeons to complete them. Other rubrics have been developed for assessing skill in the wet lab, for example, the Iowa Department of Ophthalmology Objective Wet Laboratory Structured Assessment of Skill and Technique (OWLSAT) [28] and the Eye Surgical Skills Assessment Test (ESSAT) [29, 30]. However, wet lab costs may reach up to approximately \$370,000, which make it infeasible for widespread implementation [17].

Assessments using structured rating scales may be obtained from individual experts or from a crowd, i.e., through crowdsourcing. In crowdsourcing methodology, a randomly selected sample of unrelated individuals who are not necessarily domain experts, but who are incentivized, are recruited to perform repetitive tasks. On average, the responses from the crowd highly agree with those from domain experts. Findings from multiple studies provide evidence of validity of ratings obtained using crowdsourcing methodology in robotic surgery, urology, laparoscopic surgery, and gynaecology [31–33].

(b) **Objective instrument motion data (sensor-based)**

Sensors provide time series data on instrument motion (e.g., position, orientation). The time series data are typically summarized into

metrics of instrument motion. Examples of metrics include duration for which instruments are moved, distance traveled by the instruments, number of times the instrument or the surgeon's hand moved, amplitude of instrument tremor, instrument velocities, and forces [12–14]. Early studies show differences in the metrics between surgeons with more and less experience. The metrics may also be used as features to regress on other assessments of surgical skill such as scores on structured rating scales.

(c) **Objective video analysis**

In cataract surgery, videos of the surgical field can be readily captured through the operating microscope. Videos are complex data because they are time series of images made of tens of thousands of pixels. Until recent advances in machine learning (ML) and deep learning (DL) techniques, only manual review of videos was possible. ML and DL techniques have enabled completely data-driven analysis of surgical videos. Surgical skill assessments based on data-driven analysis of videos using ML or DL techniques developed in an appropriate manner are objective and don't have the same subjectivity seen in assessments from human raters.

ML and DL techniques allow training models using methods to learn patterns in and across frames in videos [34]. DL is a subset of ML techniques that take as input high-dimensional video images and regress them on labels of interest, e.g., scores on structured rating scales. DL techniques involve artificial neural networks, which are inspired by the interconnected neurons in the human brain. On the other hand, conventional ML techniques require extensive pre-processing of the video data into features of a fixed dimension. Using the features from video images as input, the ML techniques can be trained to find patterns and boundaries that allow accurate predictions of labels of interest.

Surgeons' use of algorithms that apply ML and DL techniques is an active area of research. Research on algorithms for cataract surgery has focused on segmenting anatomy (semantic segmentation) in video images, recognition of surgical instruments in video images, identification

of the surgical activity given frames in video images, clipped videos, or videos of the entire surgical procedure, and assessment of surgical skill given videos of the surgical field. Information on anatomy, instruments, and activity can enhance the specificity of skill assessment, for example, an assessment for each step of the procedure. Therefore, this chapter includes a summary of literature on algorithms for segmenting structures, identifying instruments, and recognizing activities in cataract surgery in addition to those for objectively assessing surgical skill.

Semantic Segmentation of Cataract Surgery Video Images

Semantic segmentation of images involves drawing outlines around anatomical structures and other objects visible in the images. Anatomical structures and objects such as instruments in images of cataract surgery have considerable information for downstream tasks such as identifying the ongoing activity and skill assessment. In cataract surgery, the limbus is the anatomical region of most interest because nearly all information about the surgical procedure is constrained within this space. When the iris is fully dilated during cataract surgery, the difference between the pupil and the limbus is obscured. Consequently, segmenting the pupil is practically equivalent to segmenting the limbus. Using conventional methods from computer vision, Bouget, et al., initially processed images from cataract surgery videos to generate a mask around the pupil and then applied Hough transforms to detect the circular region within the mask [35, 36]. More recently, Sokolova, et al., used a masked region-based convolutional neural network (Mask R-CNN) to segment the pupil in cataract surgery video images [37]. A Mask R-CNN is designed to learn to predict segmentation masks for regions of interest in the images [38]. Both studies reported accurately segmenting the pupil. Bouget, et al., reported accurately segmenting the pupil in 95% of video images [35]. Sokolova, et al., reported 100% intersection over union (IoU) for the pupil in 90% of

images. IoU is a metric that compares the area of the predicted mask and the ground truth annotation of the pupillary margin. Predicting segmentation masks is an active topic of research in the general computer vision community with new foundational models such as the segment anything model, being developed that can be used for a variety of problems. However, fine-tuning the generic foundational model to the specific context of cataract surgery video images may be necessary [39].

Analysis of Surgical Instruments in Cataract Surgery Video Images

In cataract surgery, the surgical instruments provide substantial amounts of information on which step of the procedure is being performed and how skillfully the surgeon is performing it. Consequently, multiple studies have been done to analyze surgical instruments in cataract surgery videos with three objectives: (1) recognize instruments, i.e., label which instruments are visible in the images; (2) detect, segment, or localize instruments, i.e., predict a mask around the instruments or identify all pixels in the image that show the instruments; and (3) track location of one or more instruments across video images (i.e., across time).

Recognition of instruments is non-trivial because of similarities in appearance between the instruments or differences in instrument scale (altered by zooming the camera in or out), color gradient, or orientation. In early research, conventional computer vision methods were used to recognize instruments. Bouget, et al., used methods to compute descriptors of key points in images [35]. The descriptors are then modeled using a bag of visual words clustering to obtain a vector of cluster labels for each image. The video is then summarized as a histogram of the visual words across all its images. The histograms are then used to train a ML algorithm to predict labels of instruments visible in the image. Using this approach, Bouget, et al., reported an accuracy of 84% while recognizing seven instrument labels [35].

On the other hand, using DL techniques, Al Hajj, et al., used a CNN based framework in which optical flow for three contiguous subsequences of the video is computed [40]. By fusing information from the three subsequences within the CNN, the method shows an area under the receiver operating characteristic curve (AUC) of at least 0.95 to recognize 10 instruments in cataract surgery videos. In a different study, Zisimopoulos, et al., achieved 99% accuracy to recognize 21 instruments in a small dataset called CATARACTS [41]. They trained a ResNet-152 architecture for a convolutional neural network (CNN) using a dataset of 25 videos.

Algorithms to recognize instruments are useful when they are generalizable to other datasets. Paranjape, et al., used two datasets to test generalizability and estimated that model performance drops by up to 43% when it is trained on one dataset and tested on another dataset [42]. They also proposed a new method to train the models to learn information that is generalizable to other datasets. Using two novel loss functions, called the Barlow Feature Alignment Loss and the Correlation Alignment Loss, an unsupervised approach was developed to train the model to learn features from the source dataset that were relevant for predicting instrument labels in the target dataset. Using the new loss functions allowed a 6% improvement in accuracy compared with conventional methods for training generalizable models.

Analysis of Surgical Activity Using Cataract Surgery Video Images

Cataract surgery is a procedure made up of several steps. In turn, steps are composed of activities of shorter durations. Analysis of activities in surgical videos may be formulated in terms of different questions or problem statements. While there is no consensus on all terminologies used in this research domain, we use the following definitions in this chapter. Predicting an activity label given videos that are trimmed such that each video shows one step is called activity

recognition. Identifying the frame(s) where one activity transitions to another in untrimmed procedure videos is called temporal activity localization. On the other hand, predicting a bounding box across time to encapsulate each step in the procedure in an untrimmed video is referred to as spatiotemporal activity detection. Finally, another analysis of activity involves predicting a label for each frame in a video (per-frame activity recognition). In the literature on surgical activity analysis, terms such as phase recognition or step recognition are commonly used. However, care must be taken to verify what the terms mean given the problem addressed in the study.

Studies on algorithms to analyze activities in cataract surgery have focused on steps. However, the studies differ in the specific steps in the procedure they analyzed and the methods they used. In early research using conventional computer vision methods, Bouget, et al., used descriptors of key points in images represented as a bag of visual words for recognition of eight steps (preparation, betadine injection, corneal incision, capsulorhexis, phacoemulsification, cortical aspiration, IOL implantation, IOL adjustment and wound sealing) [35]. The study reported correctly labeling the step in up to 94.4% of videos that were previously trimmed to include one step each using manual annotations [35].

Lalys, et al., similarly proposed a high-level task recognition system based on application-dependent visual cues and time series analysis, using either a Dynamic Time Warping or Hidden Markov Model algorithm [36]. Five subsystems based on visual cues were implemented: color-oriented visual cues (simple histogram intersection), texture-oriented visual cues (bag-of-visual-words-approach), shape-oriented visual cues (Haar classifier trained for instrument categorization and bag-of-visual-words approach used to refine features for instrument detection), and all other visual cues. Using this framework, the study reported an accuracy of almost 94% for recognition of 12 steps in cataract surgery (preparation, betadine injection, paracentesis, main incision, viscoelastic

injection, capsulorhexis, phacoemulsification, cortical aspiration of big lens pieces, cortical aspiration of the reminiscent lens, expansion of main incision, IOL insertion, IOL adjustment and wound sealing).

Building upon their initial work, Lalys, et al., subsequently developed methods to identify surgical activities at a lower level of granularity, which they defined as “the use of one surgical tool for one surgical action performed on one anatomical structure” [43]. Using a dataset of 20 cataract surgeries, 18 activities were identified, containing up to 25 pairs of activities. A per frame recognition accuracy of 64.5% was achieved using image analysis techniques [44].

Some methods have been developed for “online” use to identify activities in cataract surgery videos. Online application refers to predicting activity at a given time using images and information up until that point in time. For online application, Quellec, et al., proposed a content-based video retrieval (CBVR) method [45]. In this method, cataract surgery video images are represented as color features and motion features (computed from optical flow). Subsequences in features are extracted and indexed for each video. For a given video until a certain time in the procedure, the subsequences from all previous frames in the video are compared with those in a database. Videos with matching subsequences are retrieved from the database and their activity label is applied to the given video. CBVR using subsequence matching is computationally intensive. Consequently, Quellec, et al., developed an alternative method in which motion features are modeled using complex spatiotemporal polynomials [43]. The polynomials are used for matching videos in the CBVR framework. For activity recognition, the spatiotemporal polynomial based CBVR framework had an AUC of 0.83 in a dataset of 186 procedures performed by 10 surgeons [43].

In another study, Charrière, et al., evaluated a CBVR based method using instrument and motion data [46]. In this method, each video is segmented into fixed length subsequences of a certain duration. Each subsequence is then represented as a bag of visual words and motion

histograms to generate features for each video. Several models, such as conditional random fields, Bayesian networks, and hidden Markov models, are then trained to use the features to classify the surgical activity. Using this method on a small dataset of 30 cataract surgery videos, Charrière, et al., reported an AUC of 0.83 to recognize seven steps.

The availability of deep learning methods to model surgical videos has shifted the emphasis away from conventional computer vision methods. Zisimopoulos, et al., extended their CNN for instrument recognition and modeled the CNN features from images across time using a recurrent neural network [40]. Using the CATARACTS dataset, the CNN-RNN framework had 78.3% accuracy for per-frame recognition of steps in cataract surgery. In a different approach, using only a CNN but modeling time as a channel in addition to three (RGB) color channels for each pixel, Primus, et al., developed a network to recognize 11 steps in 21 videos of cataract surgery [47]. The method showed an average precision of 0.74 and recall (sensitivity) of 0.72.

In another study using a dataset of 99 cataract surgery videos, Yu, et al., evaluated models to analyze information on instruments visible in images (manually annotated) and the video images themselves [48]. The models were trained and tested to recognize 10 steps in cataract surgery. A support vector machine using information on which instrument is visible in an image had an AUC of 0.74. Modeling instrument usage over time using a RNN had an AUC of 0.77. Analyzing each image in the with a CNN and using the features to classify the step showed an AUC of 0.71. On the other hand, temporal modeling of the CNN features using a RNN had an AUC of 0.75. Adding instrument motion to the spatiotemporal CNN-RNN model did not improve the AUC (0.74). In summary, the evaluation by Yu, et al., showed that identifying instruments may provide useful information to recognize steps in cataract surgery, but spatiotemporal modeling of images in an end-to-end fashion may improve accuracy.

Recently, Shah, et al., evaluated DL models based on the transformer architecture for

recognition of steps in cataract surgery videos [49]. In addition to evaluating existing DL models that were developed for activity segmentation in other surgical procedures, Shah, et al., developed a new model. In their model, video images are converted into tokens. Attention mechanisms are used to learn how the tokens are related to each other in space and across time. The attention across time is learnt at both short- and long-term scales. Evaluation of this model, called gated long short transformer (GLSFormer), showed a precision and recall of 90% and 89%, respectively, on one dataset and 70% and 56%, respectively on a second dataset [49].

Objective Assessment of Surgical Skill

Methods for objective surgical skill assessment for cataract surgery have relied upon information on instruments and their motion or end-to-end modeling of video images. In most early studies, instruments were tracked over time and metrics of instrument motion, e.g., duration, speed, movements, were computed [12, 13, 16]. The metrics were compared between surgeons with different amounts of experience, and any differences were interpreted as evidence of validity of the metrics. However, early studies did not evaluate concurrent validity of the metrics by using them to predict assessments of surgical skill using alternative tools or scales.

An alternative evaluation of skill is to focus on step-specific metrics, for example, using circularity of capsulorhexis. Zhu, et al., analyzed videos of surgeons performing capsulorhexis on a benchtop physical simulator (Kitaro, Inc.), and computed metrics such as size, centrality, and circularity of the rhexis, and motion features [50]. Using a support vector machine, the metrics were regressed on expert gradings of rhexis size, centrality, and circularity. The reported accuracy of the method was between 60% for centrality and 86% for size.

Other studies evaluated correlation between metrics of motion and scores on structured rating scales. Yeh, et al., developed a platform

called PhacoTrainer® to compute metrics of motion from cataract surgery videos [51]. The platform relies upon a DL algorithm called YOLACT to segment the limbus, pupil, and instruments, which are then used to identify and track key points across images. The tracking allows computation of instrument usage metrics such as path length, maximum instantaneous speed, and coverage area (area traversed across images by the key point). Correlation of the metrics with scores on the OSACSS rating scale ranged between -0.24 and -0.85 . The negative correlations indicate that high values of the metrics correspond to poor skill.

Other studies developed and evaluated DL algorithms to predict scores on ICO-OSCAR for phacoemulsification. Kim, et al., used manual annotations of the tips of instruments and the points of entry into the eye in images of capsulorhexis to train a temporal convolutional network (TCN) to predict scores on ICO-OSCAR [52]. The models took as input either the location (x, y pixel coordinates) or the velocity as input. The study also evaluated an additional model using optical flow as input to the DL algorithm. The AUC was 0.67, 0.86 and 0.80 for the algorithms using instrument tip positions, velocities, and optical flow, respectively, using a dataset of 99 videos of capsulorhexis.

Using the same dataset as in Kim, et al. [52], Hira, et al., evaluated a variety of methods to predict scores on ICO-OSCAR for the capsulorhexis step [53]. They included five conventional computer vision methods—bag of words, augmented bag of words, discrete transforms, sequential motion textures, and approximate entropy. In addition, the study also evaluated two DL algorithms. The first DL algorithm involved predicting the instrument tips and points of insertion into the eye and then using the predicted locations as input to a TCN. The second DL algorithm involved a CNN-RNN framework enhanced with spatial and temporal attention modules. The attention modules allow the network to learn which segments of images and which images are important for predicting the skill scores. The AUCs for the conventional computer vision methods were between

0.49 for sequential motion textures and 0.76 for augmented bag of words. On the other hand, the AUC was 0.79 and 0.78 for the DL algorithms using predicted keypoints and attention modules, respectively [53].

Other Applications of DL Algorithms to Cataract Surgery

Tabuchi, et al., combined recognition of two critical steps in cataract surgery an assessment of risk of complications in each step [54]. In this method, one DL algorithm is trained to recognize the capsulorhexis and phacoemulsification steps. A separate DL algorithm is trained to predict incidence of complications such as vitreous prolapse, rupture of capsule, iris damage or prolapse, rupture of lens zonules, dropped nucleus, discontinuity in rhexis, runaway tears, and damage to the wound. However, training DL models to predict risk of complications is difficult because the video datasets typically do not include enough cases with intraoperative complications.

On a similar note, Wang, et al., developed a 3D-CNN to analyze videos of cataract surgery recognize 12 steps in addition to predicting scores on ICO-OSCAR for phacoemulsification [55]. The predicted sequence of steps was compared with a pre-specified canonical sequence of steps and deviations from the canonical sequence were flagged as alerts to the surgeon. The DL model trained on a dataset of 200 cataract surgery videos had an accuracy between 74.7 and 98.8% to recognize different steps. The predicted skill scores had a 60% agreement with expert annotated ground truth.

Future Directions

The future of cataract surgery skills assessment is moving towards autonomous techniques for more effective and efficient surgical skill acquisition. This push stems from the need to ensure both quality and safety in surgery. A goal for interventions is tailoring training programs,

addressing specific needs based on individual skill levels. It also enables a more refined assessment of the impact of training interventions, allowing for a more precise evaluation of skill enhancement and proficiency development over time.

In the future, deep learning algorithms may advance to a point where they can provide autonomous feedback to surgeons, thus, supplanting the traditional apprenticeship model for surgical training. However, research on AI algorithms for multiple technical problems is necessary to enable autonomous feedback for surgeons. One research direction is to incorporate additional sources of data such as eye gaze. For example, Nespolo, et al., used a wearable eye tracker to collect data on surgeons' eye gaze during surgery [56]. By analyzing the video images using DL algorithms to recognizing landmarks in the surgical field, it becomes possible to compare them with the area of interest from the surgeons' eye gaze. Another research direction is algorithms that not only provide an assessment but also feedback for the surgeon. Only when algorithms analyze the granular concepts can they learn higher-level concepts for feedback. Implementation of such algorithms can subsequently provide a scalable, rapid, and objective method for the assessment of technical skill across all surgical disciplines. Generalizability is a requirement for all AI algorithms. However, generalizability is a non-trivial problem to solve even for instrument identification in cataract surgery video images [42]. Finally, the ultimate goal of technological advances is to improve surgeons' learning curve and to ensure that patients are able to navigate their surgical care. Research on effectiveness of the AI-based technologies is necessary to inform training curricula, certification processes, and monitoring outcomes of surgical care.

References

1. National Eye Institute. Cataract data and statistics [Internet]. 2019. Available from: <https://www.nei.nih.gov/learn-about-eye-health/resources-for->

- health-educators/eye-health-data-and-statistics/cataract-data-and-statistics
2. Bartley GB. Verifying surgical competence: our fiduciary responsibility. *Ophthalmology*. 2020;127(80):997–9.
 3. Lee AG, Volpe N. The impact of the new competencies on resident education in ophthalmology. *Ophthalmology*. 2004;111(7):1269–70.
 4. ACGME review committee for ophthalmology. Case log information: ophthalmology. https://www.acgme.org/globalassets/pfassets/programresources/oph_caseloginfo.pdf. Accessed January 6, 2024.
 5. Puri S, Sikder S. Cataract surgical skill assessment tools. *J Cataract Refract Surg* [Internet]. 2014;40(4):657–65. Available from: <https://doi.org/10.1016/j.jcrs.2014.01.027>.
 6. Vedula SS, Ishii M, Hager GD. Objective assessment of surgical technical skill and competency in the operating room. *Annu Rev Biomed Eng* [Internet]. 2017;19(1):301–25. Available from: <http://www.annualreviews.org/doi/https://doi.org/10.1146/annurev-bioeng-071516-044435>.
 7. Levin M, McKechnie T, Khalid S, Grantcharov TP, Goldenberg M. Automated methods of technical skill assessment in surgery: a systematic review. *J Surg Educ*. 2019;76(6):1629–39.
 8. Ogawa T, Shiba T, Tsuneoka H. Usefulness of surgical media center as a cataract surgery educational tool. *J Ophthalmol*. 2016;2016.
 9. Gauba V, Tsangaris P, Tossounis C, Mitra A, McLean C, Saleh GM. Human reliability analysis of cataract surgery. *Arch Ophthalmol*. 2008;126(2):173–7.
 10. Cox A, Dolan L, MacEwen CJ. Human reliability analysis: a new method to quantify errors in cataract surgery. *Eye*. 2008;22(3):394–7.
 11. Finn AP, Borboli-Gerogiannis S, Brauner S, Peggy Chang HY, Chen S, Gardiner M, Greenstein SH, Kloek C, Miller JW, Chen TC. Assessing resident cataract surgery outcomes using medicare physician quality reporting system measures. *J Surg Educ*. 2016;73(5):774–9.
 12. Saleh GM, Voyazis Y, Hance J, Ratnasothy J, Darzi A. Evaluating surgical dexterity during corneal suturing. *Arch Ophthalmol*. 2006;124(9):1263–6.
 13. Saleh GM, Sim D, Lindfield D, Borhani M, Ghoussayni S, Gauba V. Motion analysis as a tool for the evaluation of oculoplastic surgical skill: evaluation of oculoplastic surgical skill. *Arch Ophthalmol*. 2008;126(2):213–6.
 14. Zheng Y, Yang X, Mo B, Qi Y, Yang Y, Lin C, Han S, Wang N, Guang C, Liu W. Evaluation of the hand motion and peeling force in inner limiting membrane peeling. *Transl Vis Sci Technol*. 2023;12(3):32.
 15. Nandigam K, Soh J, Gensheimer WG, Ghazi A, Khalifa YM. Cost analysis of objective resident cataract surgery assessments. *J Cataract Refract Surg*. 2015;41(5):997–1003.
 16. Smith P, Tang L, Balntas V, Young K, Athanasiadis Y, Sullivan P, et al. “PhacoTracking” an evolving paradigm in ophthalmic surgical training. *JAMA Ophthalmol*. 2013;131(5):659–61.
 17. Balal S, Smith P, Bader T, Tang HL, Sullivan P, Thomsen ASS, et al. Computer analysis of individual cataract surgery segments in the operating room. *Eye* [Internet]. 2019;33(2):313–9. Available from: <http://www.nature.com/articles/s41433-018-0185-1>.
 18. Din N, Smith P, Emeriewen K, Sharma A, Jones S, Wawrzynski J, et al. Man versus machine: software training for surgeons—an objective evaluation of human and computer-based training tools for cataract surgical performance. *J Ophthalmol*. 2016;2016.
 19. Low SAW, Braga-Mele R, Yan DB, El-Defrawy S. Intraoperative complication rates in cataract surgery performed by ophthalmology resident trainees compared to staff surgeons in a Canadian academic center. *J Cataract Refract Surg*. 2018;44(11):1344–9.
 20. Add reference for Medtronic Touch Surgery.
 21. Cremers SL, Ciolino JB, Ferrufino-Ponce ZK, Henderson BA. Objective assessment of skills in intraocular surgery (OASIS). *Ophthalmology*. 2005;112(7):1236–41.
 22. Cremers SL, Lora AN, Ferrufino-Ponce ZK. Global rating assessment of skills in intraocular surgery (GRASIS). *Ophthalmology*. 2005;112(10):1655–60.
 23. Feldman BH, Geist CE. Assessing residents in phacoemulsification. *Ophthalmology*. 2007;114(8):1586–1586.e2.
 24. Saleh GM, Gauba V, Mitra A, Litwin AS, Chung AKK, Benjamin L. Objective structured assessment of cataract surgical skill. *Arch Ophthalmol*. 2007;125(3):363–6.
 25. Swaminathan M, Ramasubramanian S, Pilling R, Li J, Golnik K. ICO-OSCAR for pediatric cataract surgical skill assessment. *J AAPOS*. 2016;20(4):364–5.
 26. Golnik KC, Beaver H, Gauba V, Lee AG, Mayorga E, Palis G, et al. Cataract Surgical Skill Assessment. *Ophthalmology*. 2011;118(2):427–427.e5.
 27. RCOphth. Objective assessment of surgical and technical skills (OSATS) [Internet]. Available from: <https://www.rcophth.ac.uk/curriculum/ost/assessments/workplace-based-assessments/objective-assessment-of-surgical-and-technical-skills-osats/>
 28. Lee AG, Greenlee E, Oetting TA, Beaver HA, Johnson AT, Boldt HC, et al. The Iowa ophthalmology wet laboratory curriculum for teaching and assessing cataract surgical competency. *Ophthalmology*. 2007;114(7):21–6.
 29. Taylor JB, Binenbaum G, Tapino P, Volpe NJ. Microsurgical lab testing is a reliable method for assessing ophthalmology residents’ surgical skills. *Br J Ophthalmol*. 2007;91(12):1691–4.
 30. Fisher JB, Binenbaum G, Tapino P, Volpe NJ. Development and face and content validity of an eye surgical skills assessment test for ophthalmology residents. *Ophthalmology*. 2006;113(12):2364–70.
 31. Dai JC, Lendvay TS, Sorensen MD. Crowdsourcing in surgical skills acquisition: a developing technology in surgical education. *J Grad Med Educ*. 2017;9(6):697–705.

32. Polin MR, Siddiqui NY, Comstock BA, Hesham H, Brown C, Lendvay TS, et al. Crowdsourcing: a valid alternative to expert evaluation of robotic surgery skills. *Am J Obstet Gynecol*. 2016;215(5):644.e1-644.e7.
33. Kowalewski TM, Comstock B, Sweet R, Schaffhausen C, Menhadji A, Averch T, et al. Crowdsourced assessment of technical skills for validation of basic laparoscopic urologic skills tasks. *J Urol*. 2016;195(6):1859–65.
34. Sheikh AY, Fann JI. Artificial intelligence. *Thorac Surg Clin* [Internet]. 2019;29(3):339–50.
35. Bouget D, Lalys F, Jannin P, Bouget D, Lalys F, Jannin P, et al. Surgical tools recognition and pupil segmentation for cataract surgical process modeling. In: *Medicine meets virtual reality—NextMed*. 2012. p. 78–84.
36. Lalys F, Riffaud L, Bouget D, Jannin P. A framework for the recognition of high-level surgical tasks from video images for cataract surgeries. *IEEE Trans Biomed Eng*. 2012;59(4):966–76.
37. Sokolova N, Schoeffmann K, Taschwer M, Sarny S, Putzgruber-Adamitsch D, El-Shabrawi Y. Automatic detection of pupil reactions in cataract surgery videos. *PLoS ONE*. 2021;16(10): e0258390.
38. He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. In: *2017 IEEE international conference on computer vision (ICCV)*, Venice, Italy; 2017. p. 2980–8.
39. Paranjape JN, Nair NG, Sikder S, Vedula SS, Patel V. AdaptiveSAM: towards efficient tuning of SAM for surgical scene segmentation. *arXiv:2308.03726 [cs.CV]*.
40. Al Hajj H, Lamard M, Charrière K, Cochener B, Quéllec G. Surgical tool detection in cataract surgery videos through multi-image fusion inside a convolutional neural network. *Proc Annu Int Conf IEEE Eng Med Biol Soc EMBS*. 2017;2002–5.
41. Zisimopoulos O, Flouty E, Luengo I, Giataganas P, Nehme J, Chow A, et al. DeepPhase: surgical phase recognition in CATARACTS Videos. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)*. 2018;11073 LNCS:265–72.
42. Paranjape JN, Sikder S, Patel VM, Vedula SS. Cross-dataset adaptation for instrument classification in cataract surgery videos. In: Greenspan H, et al. editors. *Medical image computing and computer assisted intervention—MICCAI 2023*. MICCAI 2023. Lecture notes in computer science, vol. 14220. Cham: Springer; 2023.
43. Quéllec G, Lamard M, Cochener B, Cazuguel G. Real-time task recognition in cataract surgery videos using adaptive spatiotemporal polynomials. *IEEE Trans Med Imaging*. 2015;34(4):877–87.
44. Lalys F, Bouget D, Riffaud L, Jannin P. Automatic knowledge-based recognition of low-level tasks in ophthalmological procedures. *Int J Comput Assist Radiol Surg*. 2013;8(1):39–49.
45. Quéllec G, Charrière K, Lamard M, Droueche Z, Roux C, Cochener B, et al. Real-time recognition of surgical tasks in eye surgery videos. *Med Image Anal*. 2014;18(3):579–90.
46. Charrière K, Quéllec G, Lamard M, Martiano D, Cazuguel G, Coatrieux G, et al. Real-time analysis of cataract surgery videos using statistical models. *Multimed Tools Appl*. 2017;76(21):22473–91.
47. Primus MJ et al. Frame-based classification of operation phases in cataract surgery videos. In: Schoeffmann K, et al. editors. *MultiMedia modeling. MMM 2018. Lecture notes in computer science*, vol. 10704. Cham: Springer; 2018.
48. Yu F, Silva Croso G, Kim TS, Song Z, Parker F, Hager GD, et al. Assessment of automated identification of phases in videos of cataract surgery using machine learning and deep learning techniques. *JAMA Netw open*. 2019;2(4): e191860.
49. Shah NA, Sikder S, Vedula SS, Patel VM. GLS former: gated - long, short sequence transformer for step recognition in surgical videos. In: Greenspan H, et al. editors. *Medical image computing and computer assisted intervention—MICCAI 2023*. MICCAI 2023. Lecture notes in computer science, vol. 14228. Cham: Springer; 2023.
50. Zhu J, Luo J, Soh JM, Khalifa YM. A computer vision-based approach to grade simulated cataract surgeries. *Mach Vis Appl*. 2014;26(1):115–25.
51. Yeh HH, Jain AM, Fox O, Sebov K, Wang SY. PhacoTrainer: deep learning for cataract surgical videos to track surgical tools. *Transl Vis Sci Technol*. 2023;12(3):23.
52. Kim TS, O'Brien M, Zafar S, Hager GD, Sikder S, Vedula SS. Objective assessment of intraoperative technical skill in capsulorhexis using videos of cataract surgery. *Int J Comput Assist Radiol Surg*. 2019;14:1097–105.
53. Hira S, Singh D, Kim TS, et al. Video-based assessment of intraoperative surgical skill. *Int J Comput Assist Radiol Surg*. 2021;17:1801–11.
54. Tabuchi H, Morita S, Miki M, Deguchi H, Kamiura N. Real-time artificial intelligence evaluation of cataract surgery: a preliminary study on demonstration experiment. *Taiwan J Ophthalmol*. 2022;12(2):147.
55. Wang T, Xia J, Li R, Wang R, Stanojcic N, Li JO, Long E, Wang J, Zhang X, Li J, Wu X, Liu Z, Chen J, Chen H, Nie D, Ni H, Chen R, Chen W, Yin S, Lin D, Yan P, Xia Z, Lin S, Huang K, Lin H. Intelligent cataract surgery supervision and evaluation via deep learning. *Int J Surg*. 2022;104: 106740.
56. Nespolo RG, Cole E, Wang D, Yi D, Leiderman YI. A platform for tracking surgeon and observer gaze as a surrogate for attention in ophthalmic surgery. *Ophthalmol Sci*. 2022;3(2): 100246.

Deep Learning Applications in Ocular Oncology

19

T. Y. Alvin Liu, Neslihan Dilruba Koseoglu,
and Zelia M. Correa

Ocular oncology is a subspecialty within ophthalmology that primarily focuses on the diagnosis and treatment of intraocular malignancies and ocular surface malignancies. Examples of intraocular malignancy include uveal melanoma (UM), retinoblastoma (Rb), and metastatic tumor from systemic dissemination. Examples of ocular surface malignancy include conjunctival squamous cell carcinoma (SCC), conjunctival melanoma and conjunctival lymphoma. Currently, deep learning (DL) is the cutting-edge machine learning (ML) technique for medical image analysis in various disease including cancer. However, given the relative rarity of intraocular and ocular surface malignancies and the need for a large amount of data in order to train a deep learning system (DLS), there are only a limited number of studies published to date discussing the application of DL in the field of ocular oncology. In this chapter, we will highlight these publications and discuss the future directions of DL applications in ocular oncology.

T. Y. Alvin Liu (✉) · N. D. Koseoglu
Retina Division, Wilmer Eye Institute,
Johns Hopkins University, Baltimore, MD, USA
e-mail: tliu25@jhmi.edu

Z. M. Correa
Ocular Oncology Service, Bascom Palmer Eye Institute,
Sylvester Comprehensive Cancer Center, University of
Miami, Miller School of Medicine, Miami, FL, USA

Intraocular Tumors

Uveal Melanoma

UM is the most common primary intraocular malignancy in adults [1]. This neoplasm arises from the uveal tract that is divided into iris, ciliary body, and choroid. Although UM is relatively rare in the United States with an incidence of 5.2 cases per million, [2] it is a potentially lethal disease, with an overall 5-year survival rate of 80.9% that has not improved in the past 4 decades, [2] despite the advancement in local treatment, which typically involves plaque brachytherapy and proton beam irradiation.

Sun et al. [3] used DL techniques to detect the expression of BRCA1-associated protein 1 (BAP1) in histopathology slides. The BAP1 protein is involved in tumor suppression and is produced by the BAP1 gene, which is located on chromosome 3p21.1 [4]. Inactivation mutation of the BAP1 gene has been shown to increase the metastatic potential of UM [5] and to be present in 81% to 84% of metastatic UM [6–8]. In this study, 47 enucleated eyes from 47 patients were included. The paraffin blocks were cut into 4-micron-thick sections that were incubated with mouse monoclonal antibodies against BAP1 and counterstained with hematoxylin eosin (H&E). The resultant glass slides were digitally scanned, and the region of interest containing the UM tumor was further

cropped into numerous 256×256 pixel image patches. In total, 8,176 histopathology image patches were generated in this fashion. Each image patch was annotated twice by an ophthalmic pathologist, who established the ground truth, and each image patch was classified as 1 of the 4 categories: positive (positive for BAP1 expression, 2576 patches), negative (negative for BAP1 expression, 4720 patches), blurred (too blurred to be classified, 560 patches) and excluded (lack of UM cells, 320 patches). The 8176 image patches were randomly split into a training (6800 image patches) and testing (1376 image patches) subset. The authors applied transfer learning to a pre-trained DenseNet-121 network, and achieved a sensitivity of 97.09%, specificity of 98.12%, and overall accuracy of 97.10% in predicting nuclear BAP1 expression. While this study represents the first DL study in the field of ocular oncology, the resultant DLS ultimately only emulates what a human pathologist is capable of performing, the task of identifying images with positive BAP1 staining. In addition, the reported methodology has limited clinical practicalities, as it requires histopathology slides obtained from enucleated eyes and the current standard of care for most eyes with UM is globe-preserving local therapy with either plaque brachytherapy or proton beam irradiation.

Similarly, Zhang et al. [9] aimed to predict BAP1 expression in UM tumors. However, a key difference was that the DLS developed by Zhang et al. was trained to directly predict BAP1 expression from H&E-stained digital slide images. A total of 184 whole slide images (WSIs), acquired from enucleation or local resections, were included in this study. The WSIs were divided into training (66 BAP1⁺ and 74 BAP1⁻) and testing (16 BAP1⁺ and 28 BAP1⁻) sets. Each WSI was further split into smaller image patches. The ground truth (BAP+ or BAP-) for each H&E-stained image patch was based on its counterpart BAP1-stained image patch. Different versions of the DLS were trained, with varying image patch dimensions and with and without region

correlation (an image patch was more likely to be BAP1+ if its surrounding image patches were also BAP1+). Their best performing model, which used 1024×1024 pixel image patches for training and incorporated region correlation, achieved an AUC of 0.90 in patch-level predictions.

In another UM study, Liu et al. [10] extended the application of DL techniques in digital pathology slide analysis in UM, and aimed to train a DLS to perform a task that is impossible for a human pathologist—predicting gene expression profile (GEP) in smeared slides stained with H&E alone obtained from fine needle aspiration biopsy (FNAB) of uveal melanomas. The underlying hypothesis is that cancer cell morphology reflects the underlying genetics and careful analysis of cytopathology images will provide helpful prediction of the biological behavior of the tumor and clinical course of the patient. UM is unique among malignancies in that GEP, independent of other clinicopathological parameters, has been shown to be the most robust method currently available to predict long-term metastasis risk and survival. UM patients can be divided into two classes by GEP: class 1 and class 2, and there is a stark contrast in long-term survival between the two classes. The 92-month survival probability in class 1 patients is 95%, versus 31% in class 2 patients [11, 12].

In this study, 20 de-identified FNAB cytology slides from 20 patients with UM underwent H&E staining. Whole-slide scanning was performed for each cytology slide at a magnification of $40\times$, and native resolution crops containing melanoma cells were saved. Each snapshot image measured 1716 pixels (width) $\times$ 926 pixels (height), and was further split into 8 tiles of equal size. The tiles were then screened and selected for further processing only if at least one melanoma cell was present. Typically, each slide generated hundreds of $40\times$ snapshot images, and out of the 20 slides, a total of 26,351 unique image tiles were generated. Schematic representation for data processing is shown in Fig. 19.1. The GEP ground

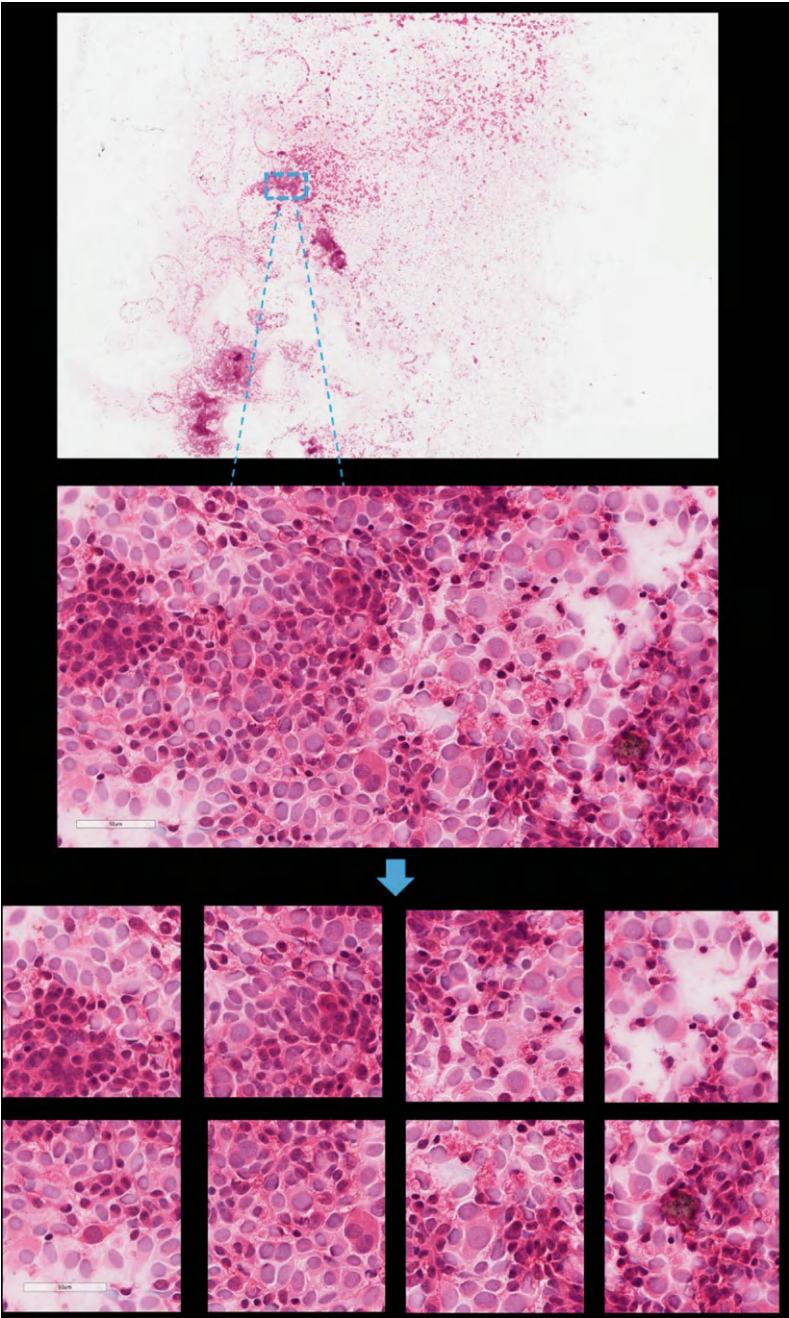


Fig. 19.1 Schematic representation of data processing. (Top Panel) Whole slide scanning; one slide per patient. (Middle Panel) Snapshot image manually captured at 40 $\times$; multiple 40 $\times$ images were captured from each slide. (Bottom Panel) Each 40 $\times$ image was further divided into eight tiles of equal sizes

truth at the slide level was established by the commercially-available DecisionDx-UM[®] test [Friendswood, Texas], and the GEP label designated to a particular slide was propagated to all the image tiles generated from that slide. I.e. if “slide 1” was determined to be GEP class 1 by the DecisionDx-UM[®] test, then all the image tiles generated from “slide 1” were labeled as “class 1.”

The authors applied transfer learning to a pre-trained ResNet-152 network for the underlying binary classification problem of distinguishing between class 1 and class 2 image tiles. Due to the low amount of data (patient) variation, the validation slides would have a strong effect on the model performance, so “leave-one-out” cross-validations were performed to evaluate the DLS’s performance. To test each of the 20 slides/patients, 10 models using different training/validation split were trained. Specifically, for each of the leave-one-out cross-validations, 10 random samplings for the validation subset selection were performed. If “slide 1” was used as the testing slide, then the other 19 slides were used for model development: 17 slides for training and 2 slides for validation (one from class 1 and one from class 2). “Slide 1” was then tested 10 different times by 10 different models that were generated by 10 random and different combinations of training and validation slides. For example, model #1 would use “slide 2” and “slide 11” for validation. Model #2 would use “slide 3” and “slide 12” for validation. Model #3 would use “slide 4” and “slide 13” for validation etc. Eventually, 10 models were generated, and the mean accuracy of these 10 models was obtained. If the lower 95 confidence interval (CI) value exceeded 50%, then it was concluded that the GEP of “slide 1” was correctly predicted. This process was repeated for all 20 slides/patients, such that each slide/patient was evaluated 10 times by 10 different models. Out of this cohort of 20 UM patients, the DLS achieved a point estimate of 75% (15/20 patients) accuracy in predicting the GEP at the patient level.

Given that GEP is the most robust prognostic test available with a high correlation

with survival, this pilot study by Liu et al. [10] showed that survival prognostication could be predicted from H&E cytopathology slides alone in UM using DL.

In a follow up study by the same group with a larger cohort (89 whole-slide images from 82 patients), Liu et al. [13] presented a DLS with a dual-attention feature extraction mechanism able to directly predict GEP from digital cytopathology images. The FNABs of the tumors and digital WSI generation were performed using the same protocols as the pilot study, but the image tiles were automatically extracted using an original algorithm. The data was partitioned on a patient level. In contrast to the pilot study in which model evaluation was performed using leave-one-out cross validations, model performance in this study was evaluated using a more conventional hold-out test set (73% training; 27% testing). The model achieved an AUC of 0.944, an accuracy of 91.7%, a sensitivity of 91.7%, and a specificity of 91.7% for GEP prediction on a slide-level analysis.

Retinoblastoma

Rb is the most common primary intraocular tumor in children [14]. Current advances in screening and multidisciplinary management of Rb have increased survival and globe salvage, especially in developed countries with dedicated centers [15]. However, in developing countries, the relative lack of access to healthcare resources proposes a big challenge in the diagnosis and referral of these patients.

To address this problem, Zhang et al. [16] developed a DLS to detect referable Rb patients from fundus images. A total of 36,623 color fundus images were obtained from 713 patients during examinations under anesthesia. For each eye, multiple images were taken, capturing the posterior pole and 12 clock hours of the peripheral retina. Images were further divided into 3 groups; normal fundus (n=19,045), stable Rb (n=2918) and active Rb (n=14,660) by 6 ocular oncology experts. Their model achieved excellent results at distinguishing normal fundus

images from images with active Rb at an image level [AUC of 0.9982 (95%CI 0.986–1.000)], detecting active Rb at an eye level [AUC of 0.991 (95%CI 0.970–1.000)] and distinguishing active Rb from stable Rb at an eye level [AUC of 0.962 (95%CI 0.915–1.000)].

Ocular Surface Tumors

Conjunctival melanoma is a rare but lethal malignancy of the ocular surface [17]. Although commonly presenting as pigmented lesions, on rare occasion they may also present as amelanotic lesions, making the diagnosis more challenging. Yoo et al. [18] developed a low-shot DL model to classify ocular surface lesions and ultimately detect conjunctival melanoma in ocular surface images. Low-shot DL models require only a small number of annotated images for training. The training data was collected from publicly available images through search engines and divided by diagnosis: conjunctival melanoma (n=136), nevus or melanosis (n=93), pterygium (n=75) and normal images (n=94). In addition to these real images, synthetic images of normal conjunctiva (n=100), pterygium (n=100), nevus and melanosis (n=100) and melanoma (n=100) were generated using cycle consistency GAN (CycleGAN) and progressive growing of GAN (PGGAN) to augment the training dataset. Generative adversarial networks (GAN) is a subtype of DL techniques that can be used for data augmentation [19]. In this study, The authors trained 5 different DL architectures to detect conjunctival melanoma: MobileNetV2 performed best (AUC:0.976, accuracy:96.5%), followed by NasNet (AUC:0.972, accuracy: 95.7%), GoogleNet (AUC:0.970, accuracy:95.7%), ResNet-50 (AUC:0.968, accuracy:93.0%), and InceptionV3 (AUC:0.954, accuracy:92.0%). Incorporation of synthetic images generated by GANs during training improved the performance for all models.

Orbital and Adnexal Tumors

Ocular adnexal lymphoma (OAL) is the most common primary orbital malignancy and can arise from the conjunctiva, lacrimal gland, the orbit and the eyelids or may present secondary to systemic lymphoma [20, 21]. OAL may present with clinical and imaging features similar to those of idiopathic orbital inflammation (IOI), which requires different treatments [21].

Xie et al. [21] developed a DL model using clinical and imaging features combined with multimodal radiomics extracted from 3 magnetic resonance imaging (MRI) sequences to differentiate between OAL and IOI. Imaging features included: affected orbital area, involved quadrants, shape and borders of the lesions and MRI signal characteristics (homogeneity, intensity on T1W1 and T2W1 and enhancement pattern on T1W1+C). Clinical features included age, gender and laterality. Region of interest (ROI) areas were manually segmented from axial MRIs for radiomic analysis, and features were extracted from these ROIs. Three different DL approaches were tested: using only radiological features (DL-rad), only clinical features (DL-clin) and both radiological and clinical features (DL-both). DL-both performed best in differentiating OAL and IOI at both patient and image level [AUC 0.953 (95% CI, 0.895–1.000)/0.843 (95% CI, 0.786–0.898), respectively].

Basal cell carcinoma (BCC) and squamous cell carcinoma (SCC) are the two most common eyelid malignancies. However, other malignant eyelid tumors, such as melanoma, sebaceous carcinoma (SC) and metastatic tumors may also be seen. Location, appearance and growth rate of the lesion should raise suspicion for malignancy [22].

Luo et al. [23] aimed to detect BCC and SC in H&E stained WSIs using a DL model (DenseNet). The dataset included 116 WSIs with BCC (n=116 patients) and 180 WSIs with SC (n=129 patients), and the ground truth was

established by 2 expert pathologists, based on H&E and immunohistochemistry staining. Their model achieved a high accuracy of 98.3% in differentiating between BCC and SC on a slide level using an integration module constructed by average probability aggregation strategy, and significantly outperformed both junior and senior human pathologists.

Using a similar approach, Wang et al. [24] developed a DLS to detect eyelid melanoma in H&E stained WSIs. The authors trained 3 different models (VGG16, Google Inception V1 and V2) on 105,415 images patches acquired from 38 WSIs ($n=11$ melanoma and $n=27$ non-malignant tissue that included normal tissue between malignant lesions, typical eyelid nevi and atypical eyelid nevi). The testing set included 142,104 image patches from 79 WSIs ($n=55$ melanoma and $n=24$ non-malignant lesion). VGG16 had the best performance in distinguishing melanoma from non-malignant lesions with an AUC of 0.989 (95% CI 0.989 to 0.991) at patch level and an AUC of 0.998 (95% CI 0.994 to 1.000) at a slide level.

Conclusion

Although intraocular and ocular surface malignancies are relatively rare, the number of published DL applications for these diseases has increased substantially in the past few years. However, most of these studies lacked independent external validations and virtually all relied on retrospectively collected data. Further work is required, which includes prospective validation of the DLSs with larger patient sample sizes from diverse populations.

References

1. Singh AD, Turell ME, Topham AK. Uveal melanoma: trends in incidence, treatment, and survival. *Ophthalmology*. 2011;118(9):1881–5. <https://doi.org/10.1016/j.ophtha.2011.01.040>.
2. Aronow ME, Topham AK, Singh AD. Uveal melanoma: 5-year update on incidence, treatment, and survival (SEER 1973–2013). *Ocul Oncol Pathol*. 2018;4(3):145–51. <https://doi.org/10.1159/000480640>.
3. Sun M, Zhou W, Qi X, et al. Prediction of BAP1 expression in uveal melanoma using densely-connected deep classification networks. *Cancers (Basel)*. 2019;11(10). <https://doi.org/10.3390/cancers11101579>.
4. Murali R, Wiesner T, Scolyer RA. Tumours associated with BAP1 mutations. *Pathology*. 2013; 45(2):116–26. <https://doi.org/10.1097/PAT.0b013e32835d0efb>.
5. Stalhammar G, See TRO, Phillips SS, Grossniklaus HE. Density of PAS positive patterns in uveal melanoma: correlation with vasculogenic mimicry, gene expression class, BAP-1 expression, macrophage infiltration, and risk for metastasis. *Mol Vis*. 2019; 25:502–16.
6. Griewank KG, van de Nes J, Schilling B, et al. Genetic and clinico-pathologic analysis of metastatic uveal melanoma. *Mod Pathol*. 2014;27(2):175–83. <https://doi.org/10.1038/modpathol.2013.138>.
7. Harbour JW, Onken MD, Roberson ED, et al. Frequent mutation of BAP1 in metastasizing uveal melanomas. *Science*. 2010;330(6009):1410–3. <https://doi.org/10.1126/science.1194472>.
8. Koopmans AE, Verdijk RM, Brouwer RW, et al. Clinical significance of immunohistochemistry for detection of BAP1 mutations in uveal melanoma. *Mod Pathol*. 2014;27(10):1321–30. <https://doi.org/10.1038/modpathol.2014.43>.
9. Zhang H, Kalirai H, Acha-Sagredo A, Yang X, Zheng Y, Coupland SE. Piloting a deep learning model for predicting nuclear BAP1 immunohistochemical expression of uveal melanoma from hematoxylin-and-eosin sections. *Transl Vis Sci Technol*. 2020;9(2):50. <https://doi.org/10.1167/tvst.9.2.50>.
10. Liu TYA, Zhu H, Chen H, et al. Gene expression profile prediction in uveal melanoma using deep learning: a pilot study for development of an alternative survival prediction tool. *Ophthalmol Retina*. 2020. <https://doi.org/10.1016/j.oret.2020.06.023>.
11. Onken MD, Worley LA, Ehlers JP, Harbour JW. Gene expression profiling in uveal melanoma reveals two molecular classes and predicts metastatic death. *Cancer Res*. 2004;64(20):7205–9. <https://doi.org/10.1158/0008-5472.CAN-04-1750>.
12. Onken MD, Worley LA, Char DH, et al. Collaborative ocular oncology group report number 1: prospective validation of a multi-gene prognostic assay in uveal melanoma. *Ophthalmology*. 2012;119(8):1596–603. <https://doi.org/10.1016/j.ophtha.2012.02.017>.
13. Liu TYA, Chen H, Gomez C, Correa ZM, Unberath M. Direct gene expression profile prediction for uveal melanoma from digital cytopathology images via deep learning and salient image region identification. *Ophthalmol Sci*. 2023;3(1): 100240. <https://doi.org/10.1016/j.xops.2022.100240>.

14. Pandey AN. Retinoblastoma: an overview. *Saudi J Ophthalmol.* 2014;28(4):310–5. <https://doi.org/10.1016/j.sjopt.2013.11.001>.
15. Fabian ID, Onadim Z, Karaa E, et al. The management of retinoblastoma. *Oncogene.* 2018;37(12):1551–60. <https://doi.org/10.1038/s41388-017-0050-x>.
16. Zhang R, Dong L, Li R, et al. Automatic retinoblastoma screening and surveillance using deep learning. *Br J Cancer.* 2023;129(3):466–74. <https://doi.org/10.1038/s41416-023-02320-z>.
17. Wong JR, Nanji AA, Galor A, Karp CL. Management of conjunctival malignant melanoma: a review and update. *Expert Rev Ophthalmol.* 2014;9(3):185–204. <https://doi.org/10.1586/17469899.2014.921119>.
18. Yoo TK, Choi JY, Kim HK, Ryu IH, Kim JK. Adopting low-shot deep learning for the detection of conjunctival melanoma using ocular surface images. *Comput Methods Programs Biomed.* 2021;205:106086. <https://doi.org/10.1016/j.cmpb.2021.106086>.
19. Frid-Adar M, Diamant I, Klang E, Amitai MM, Goldberger J, Greenspan H. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing.* 2018;321:321–31.
20. Hsu CR, Chen YY, Yao M, Wei YH, Hsieh YT, Liao SL. Orbital and ocular adnexal lymphoma: a review of epidemiology and prognostic factors in Taiwan. *Eye (Lond).* 2021;35(7):1946–53. <https://doi.org/10.1038/s41433-020-01198-y>.
21. Xie X, Yang L, Zhao F, et al. A deep learning model combining multimodal radiomics, clinical and imaging features for differentiating ocular adnexal lymphoma from idiopathic orbital inflammation. *Eur Radiol.* 2022;32(10):6922–32. <https://doi.org/10.1007/s00330-022-08857-6>.
22. Rana H, Stokkermans TJ, Purt B, Chou E. Malignant eyelid lesions. StatPearls. StatPearls Publishing Copyright© 2024, StatPearls Publishing LLC. 2024.
23. Luo Y, Zhang J, Yang Y, et al. Deep learning-based fully automated differential diagnosis of eyelid basal cell and sebaceous carcinoma using whole slide images. *Quant Imaging Med Surg.* 2022;12(8):4166–75. <https://doi.org/10.21037/qims-22-98>.
24. Wang L, Jiang Z, Shao A, et al. Self-supervised learning mechanism for identification of eyelid malignant melanoma in pathologic slides with limited annotation. *Front Med (Lausanne).* 2022;9:976467. <https://doi.org/10.3389/fmed.2022.976467>.

Artificial Intelligence Applied to Neuro-Ophthalmic Conditions

20

Samy Zaher, Raymond P. Najjar, and Dan Milea

Abstract

Recent deep learning methods have shown high accuracy to predict neuro-ophthalmic and neurological disease, based on analysis of retinal images alone. Despite these important recent discoveries, more prospective, real-life studies are needed to validate DL algorithms as potential diagnostic aids in relevant clinical settings. Emerging Large language models have shown variable performance in the subspecialized area of neuro-ophthalmology, requiring more validation prior to their integration into daily practice.

Keywords

Artificial intelligence · Deep learning · Machine learning · Detection · Optic neuropathies

Introduction

The current, extraordinary boom of AI applied to medicine and ophthalmology [1–4] has impacted the world of Neuro-Ophthalmology as well. Various deep learning (DL) models have been recently developed for prediction of neuro-ophthalmic and neurological conditions, based on retinal images alone. Before the era of DL, early computer-aided diagnostic systems have already attempted to automatically detect optic disc abnormalities on retinal color fundus images [5], focusing mainly on papilledema detection [6].

In 2020, a first large, multicentre, multi-ethnic, international investigation aimed to develop, train and test a dedicated algorithm for detection of optic disc abnormalities related to neurological and neuro-ophthalmic conditions, the BONSAI study (**B**rain and **O**ptic **N**erve **S**tudy with **A**rtificial **I**ntelligence) [7]. This new algorithm suggested that potentially life-threatening papilledema due to raised intracranial pressure could be accurately identified by a deep learning system (DLS), when tested vs. normal discs and abnormal discs due to other conditions. This large collaborative study has included populations from 24 neuro-ophthalmology sites in 15 countries on three continents, using fundus images obtained with a large set of fundus camera brands. The training and validation data sets from 6779 patients included 14,341

S. Zaher
Neuro-Ophthalmology Department and BRAINLab,
Rothschild Foundation Hospital, Paris, France

R. P. Najjar
Singapore Eye Research Institute, Duke-NUS
Medical School, Department of Ophthalmology
and Yong Loo Lin School of Medicine, National
University of Singapore, Singapore, Singapore

D. Milea (✉)
Rothschild Foundation Hospital, Paris, France
e-mail: visual.neurosciences@gmail.com

photographs of abnormal and normal optic discs and the external validation testing was performed in 1505 images from five separate international centers. In this retrospective *in silico* dataset, papilledema detection (Fig. 20.1) was achieved with high AUC (0.96), high sensitivity (96%) and good specificity (85%). Despite its initial success, the study had several limitations, including its retrospective design, using images that were highly selected by the participating centers. However, this was an important step ahead, showing that computer programs examining digital retinal images may accurately classify disorders of the optic discs related to serious vision and life-threatening conditions. The potential practical and innovative value of such a tool was potentially big, since many neurologists, emergency doctors or other non-ophthalmology trained physicians are often not comfortable with diagnosing papilledema on ophthalmoscopy alone. Obtaining assistance from a machine in this context was a promising avenue.

The next question was whether the performance of such DL algorithms was comparable with that of humans. For this purpose, a dedicated study evaluated the diagnostic performance of two expert neuro-ophthalmologists

with that of the BONSAI-DLS, for discriminating various optic disc abnormalities [8]. Unsurprisingly, the algorithm required a significantly shorter time (only 25 s), to classify a large sample of 800 optic disc images, compared to the two neuro-ophthalmology experts. Also, in terms of performance, the DLS reached a comparable accuracy with that of the two expert neuro-ophthalmologists (85% versus 80–84%). It is however important to notice that the human evaluations were performed purely on retinal images, without taking into account other clinical signs (visual loss, headache, tinnitus, etc.), which are of paramount importance for accurate evaluations in everyday clinics.

The use of a dedicated DLS to accurately detect papilledema was an important conceptual first step, but a new question raised rapidly: could a DLS also evaluate the severity of papilledema, with important clinical and therapeutic consequences? For this purpose, the BONSAI Group developed a new model aiming to automatically classify papilledema severity on images, based on a simplified classification of the 5-grade Frisen system. The algorithm reached good classification performance (88%), comparable with that attained by three independent neuro-ophthalmologists [9].

Despite these first encouraging results obtained in highly curated retrospective datasets, it was not clear whether such algorithms could be also applied in real life conditions, at best evaluated in prospective randomized studies. As an intermediate step, the BONSAI group has retrospectively tested a modified version of the DLS in a large dataset collected prospectively in an Emergency Department data in patients presenting with neurological symptoms, visual loss or high blood pressure. The DLS had an excellent accuracy (98%) to detect papilledema, which was considerably better than that of non-expert humans, when initially tested. This study suggested the great potential of DL models as a diagnostic aid in detecting papilledema in non-ophthalmology clinics, therefore reducing unnecessary workups and urgent referral to ophthalmology consultations [10].

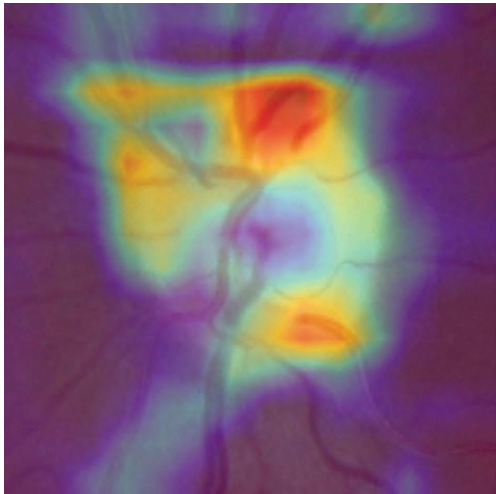


Fig. 20.1 Example of class activation probability (heatmap), displayed on an optic disc image with confirmed papilledema

In the paediatric population, presence of papilledema associated with neurological symptoms can suggest severe underlying neurological disorders, including potentially life-threatening brain tumours or cerebral vascular abnormalities. Diagnosing accurately papilledema in this setting is particularly difficult, due to various clinical limitations such as lack of child cooperation as well as high prevalence of differential diagnosis (i.e. benign pseudopapilledema due to buried, optic disc drusen, hyperopia, etc.). In addition, the availability of ophthalmologists is becoming limited and non-ophthalmic physicians (paediatricians, neurologists, general practitioners) are often not sufficiently trained to identify papilledema on ophthalmoscopy. As a consequence, it is not uncommon that such situations prompt expensive, potentially invasive, but often unnecessary investigations, such as neuroimaging, lumbar punctures, etc. The emergence of AI-based interpretation of retinal imaging has recently created hope that affordable, standard retinal photographs can be used in order to distinguish between potentially life-threatening papilledema and benign aspects of the optic nerve head in paediatric populations. For this purpose, the BONSAI team has developed an updated version of the BONSAI-DLS, able to accurately distinguish papilledema from normal optic discs and other disc abnormalities in children (overall accuracy of 90%). The performance of the DLS was similar in prepubertal children (i.e. with higher prevalence of pseudopapilledema) and postpubertal children (i.e. with higher prevalence of idiopathic intracranial hypertension) [11].

Distinguishing between true papilledema and false papilledema (i.e. pseudopapilledema due to optic disc drusen—ODD) can be very challenging in adults as well, even by expert neuro-ophthalmologists, especially when ODD are deeply buried or when papilledema is mild. The currently accepted method to identify ODD is optical coherence tomography (OCT), requiring costly machines and high human expertise. Recently, a new OCT-based DL algorithm was developed with the purpose to

automatically identify ODD on OCT data. The system achieved a very good classification performance between ODD, true papilledema and healthy optic discs, though measured on a small sample size, collected retrospectively for test assessment [12]. Given the fact that OCT is an expensive device, requiring expert operators for good quality imaging, an interesting idea rapidly emerged: could trivial, affordable color retinal images suffice, if analysed with high quality DL methods, in order to discriminate between true papilledema and pseudo-papilledema? In a recent study performed in a large population (2,180 patients and 4,508 color fundus images obtained in 30 neuro- ophthalmology centers from 19 countries), a DLS was able to discriminate with an overall accuracy of 0.9 between true and pseudopapilledema due to ODD. Unsurprisingly, the accuracy of the DLS was higher for discriminating severe papilledema from visible drusen, but dropped only marginally for the difficult task of discriminating between deep, buried drusen and mild-to-moderate papilledema [13].

Numerous other DL models applied to retinal images have been used with various success in order to discriminate between glaucomatous and non-glaucomatous optic neuropathies, predict multiple sclerosis, distinguish between various optic neuropathies, predict stroke and other purposes, recently summarized in a dedicated subspeciality Review [14]. Interestingly, the numerous emerging large language models, widely successful in other areas of ophthalmology seem to still require additional developments in Neuro-Ophthalmology, due to possible confabulations [14].

AI-Based Imaging of the Retina to Identify Neurodegenerative Diseases

Recent AI-based investigation methods applied to retinal imaging have aimed to identify patients with neurodegenerative disorders, including Alzheimer's disease (AD), the most

common cause of dementia. Diagnosing early stage dementia and mild cognitive impairment is difficult, but important, given the newly emerging treatments in this context. Accurate diagnosis requires a combination of clinical findings and various biomarkers, including brain magnetic resonance imaging (MRI), amyloid positron emission tomography (PET), cerebrospinal fluid (CSF) or blood biomarkers. The sensitivity and specificity of these methods vary a lot, prompting the need of other potential biomarkers, including the retina. Indeed, several studies have reported retinal deposits of A β protein and phosphorylated tau proteins in patients with AD. Various retinal imaging studies have also shown an association between retinal thinning with ageing and brain atrophy. More recently, DL algorithms based on retinal imaging have gained interest as a possible diagnostic aid for AD. In a large study including 12,949 retinal photographs from 648 patients with confirmed Alzheimer's disease dementia and 3240 controls without the disease, a model achieved good performance, with accuracy ranging from 80 to 92% to predict amyloid PET positivity. Interestingly, the model was more accurate in patients with eye disease such as age-related macular degeneration and glaucoma (90%), and those with diabetes (82%), compared to those without eye disease (72%) and without diabetes (72%), respectively [15].

Research is also underway regarding Parkinson's disease (PD), the second most common neurodegenerative disorder. Retinal features have been described on OCT in patients with PD, primarily involving parafoveal ganglion cell-inner plexiform layer (GCIPL) atrophy, as well as potentially more specific inner nuclear layer (INL) thinning, possibly related with degeneration of dopaminergic amacrine cells. OCT-based DL algorithms have the potential to aid in early diagnosis of this disease in preclinical PD [16].

Summary

In summary, AI applied to retinal images has gained tremendous interest in Neuro-ophthalmology and Neurology; in this systemic context it may well be that fundus images might interest in the future more non-ophthalmologists than ophthalmologists [17]. The ability of AI to accurately detect abnormal optic discs may soon be implemented for patients' triaging in non-ophthalmic set-ups. Future prospective studies, including large clinical datasets are needed to establish if such systems may become a safe, effective, attractive and affordable solution for non-ophthalmologists. Early results regarding the relationship of fundus imaging and neurodegenerative diseases and stroke are also promising but require more validation in prospective studies. Other neuro-ophthalmic disorders affecting eye movements and pupil recordings, are currently under evaluation using AI methods. There is no doubt that AI-based methods applied to the analysis of ocular structures or function will translate into improved knowledge and disease detection of the brain.

References

1. Cruz Rivera S, Liu X, Chan AW, Denniston AK, Calvert MJ. SPIRIT-AI and CONSORT-AI working group; SPIRIT-AI and CONSORT-AI steering group; SPIRIT-AI and CONSORT-AI consensus group. Guidelines for clinical trial protocols for intervention involving artificial intelligence: the SPIRIT-AI extension. *Nat Med.* 2020;26(9):1351–63.
2. Ting DSW, et al. Deep learning in ophthalmology: the technical and clinical considerations. *Prog Retin Eye Res.* 2019;72: 100759.
3. Dai L, Sheng B, Chen T, Wu Q, Liu R, Cai C, Wu L, Yang D, Hamzah H, Liu Y, Wang X, Guan Z, Yu S, Li T, Tang Z, Ran A, Che H, Chen H, Zheng Y, Shu J, Huang S, Wu C, Lin S, Liu D, Li J, Wang Z, Meng Z, Shen J, Hou X, Deng C, Ruan L, Lu F, Chee M, Quek TC, Srinivasan R, Raman R, Sun X, Wang YX, Wu J, Jin H, Dai R, Shen D, Yang X, Guo M, Zhang

- C, Cheung CY, Tan GSW, Tham YC, Cheng CY, Li H, Wong TY, Jia W. A deep learning system for predicting time to progression of diabetic retinopathy. *Nat Med*. 2024;30(2):584–94.
4. Hood DC, La Bruna S, Tsamis E, Thakoor KA, Rai A, Leshno A, de Moraes CGV, Cioffi GA, Liebmann JM. Detecting glaucoma with only OCT: implications for the clinic, research, screening, and AI development. *Prog Retin Eye Res*. 2022;90: 101052.
5. Medeiros FA, Jammal AA, Thompson AC. From machine to machine: an OCT-trained deep learning algorithm for objective quantification of glaucomatous damage in fundus photographs. *Ophthalmology*. 2019;126(4):513–21.
6. Milea D, Singhal S, Najjar RP. Artificial intelligence for detection of optic disc abnormalities. *Curr Opin Neurol*. 2020;33(1):106–10.
7. Echegaray S, et al. Automated analysis of optic nerve images for detection and staging of papilledema. *Invest Ophthalmol Vis Sci*. 2011;52:7470–8.
8. Milea D, Najjar RP, Zhuho J, Ting D, Vasseneix C, Xu X, Aghsaie Fard M, Fonseca P, Vanikieti K, Lagrèze WA, La Morgia C, Cheung CY, Hamann S, Chiquet C, Sanda N, Yang H, Mejico LJ, Rougier M-B, Kho R, Thi Ha Chau T, Singhal S, Gohier P, Clermont-Vignal C, Cheng C-Y, Jonas JB, Yu-Wai-Man P, Fraser CL, Chen JJ, Ambika S, Miller NR, Liu Y, Newman NJ, Wong TY, Biousse V, BONSAI Group. Artificial intelligence to detect papilledema from ocular fundus photographs. *N Engl J Med*. 2020b;382(18):1687–95.
9. Biousse V, Newman NJ, Najjar RP, Vasseneix C, Xu X, Ting DS, Milea LB, Hwang JM, Kim DH, Yang HK, Hamann S, Chen JJ, Liu Y, Wong TY, Milea D, BONSAI (brain and optic nerve study with artificial intelligence) study group. Optic disc classification by deep learning versus expert neuro-ophthalmologists. *Ann Neurol*. 2020.
10. Vasseneix C, Najjar RP, Xu X, Tang Z, Loo JL, Singhal S, Tow S, Milea L, Ting DSW, Liu Y, Wong TY, Newman NJ, Biousse V, Milea D; BONSAI group. Accuracy of a deep learning system for classification of papilledema severity on ocular fundus photographs. *Neurology*. 2021;97(4):e369–e377.
11. Biousse V, Najjar RP, Tang Z, Lin MY, Wright DW, Keadey MT, Wong TY, Bruce BB, Milea D, Newman NJ; BONSAI study group. Application of a deep learning system to detect papilledema on nonmydriatic ocular fundus photographs in an emergency department. *Am J Ophthalmol*. 2024;261:199–207.
12. Lin MY, Najjar RP, Tang Z, Cioplean D, Dragomir M, Chia A, Patil A, Vasseneix C, Peragallo JH, Newman NJ, Biousse V, Milea D; BONSAI (brain and optic nerve study with artificial intelligence) group. The BONSAI (brain and optic nerve study with artificial intelligence) deep learning system can accurately identify pediatric papilledema on standard ocular fundus photographs. *J AAPOS*. 2024;28(1):103803.
13. Girard MJA, Panda S, Tun TA, Wibroe EA, Najjar RP, Aung T, Thiéry AH, Hamann S, Fraser C, Milea D. Discriminating between papilledema and optic disc drusen using 3D structural analysis of the optic nerve head. *Neurology*. 2023;100(2):e192–202.
14. Sathianvichitr K, Najjar RP, Zhiqun T, Fraser JA, Yau CWL, Girard MJA, Costello F, Lin MY, Lagrèze WA, Vignal-Clermont C, Fraser CL, Hamann S, Newman NJ, Biousse V, Milea D; BONSAI group. A deep learning approach for accurate discrimination between optic disc drusen and papilledema on fundus photographs. *J Neuroophthalmol*. 2024. <https://doi.org/10.1097/WNO.0000000000002223>. Epub ahead of print. PMID: 39090774.
15. Kenney RC, Requarth TW, Jack AI, Hyman SW, Galetta SL, Grossman SN. AI in neuro-ophthalmology: current practice and future opportunities. *J Neuroophthalmol*. 2024. <https://doi.org/10.1097/WNO.0000000000002205>. Epub ahead of print. PMID: 38965655.
16. Cheung CY, Ran AR, Wang S, Chan VTT, Sham K, Hilal S, Venketasubramanian N, Cheng CY, Sabanayagam C, Tham YC, Schmetterer L, McKay GJ, Williams MA, Wong A, Au LWC, Lu Z, Yam JC, Tham CC, Chen JJ, Dumitrascu OM, Heng PA, Kwok TCY, Mok VCT, Milea D, Chen CL, Wong TY. A deep learning model for detection of Alzheimer's disease based on retinal photographs: a retrospective, multicentre case-control study. *Lancet Digit Health*. 2022;4(11):e806–15.
17. Wagner SK, Romero-Bascones D, Cortina-Borja M, Williamson DJ, Struyven RR, Zhou Y, Patel S, Weil RS, Antoniadis CA, Topol EJ, Korot E, Foster PJ, Balaskas K, Ayala U, Barrenechea M, Gabilondo I, Schapira AHV, Khawaja AP, Patel PJ, Rahi JS, Denniston AK, Petzold A, Keane PA; for UK biobank eye & vision consortium. Retinal optical coherence tomography features associated with incident and prevalent Parkinson disease. *Neurology*. 2023;101(16):e1581–e1593.

Artificial Intelligence Using the Eye as a Biomarker of Systemic Risk

21

Jinyuan Wang, Rachel Marjorie Wei Wen Tseng, Tyler Hyungtaek Rim, Carol Y. Cheung, and Tien Yin Wong

Abstract

The eye is the unique organ in the body which allows for the direct observation and imaging of the neurological and vascular system. In recent years, researchers have harnessed the noninvasive nature of colour fundus photographs (CFPs) to examine changes of the retina as possible biomarkers of the systemic diseases' risks. Building on large-scale epidemiological and clinical studies

that have reported relationships of retinal features such as retinal vascular calibre with systemic diseases, the application of artificial intelligence (AI) technology, specifically in deep learning (DL), on ocular images, especially CFPs, is advancing new research that focuses on retina-systemic disease relationships. In this relatively new field, current studies fall into 3 basic groups: (1) cross-sectional studies that use AI-DL technology on CFPs to detect or estimate systemic risk factors (e.g., age, blood pressure, smoking); (2) cross-sectional studies that use AI-DL technology on CFPs to replace established

J. Wang · T. Y. Wong (✉)
School of Clinical Medicine, Beijing Tsinghua
Changung Hospital, Tsinghua Medicine, Tsinghua
University, Beijing, China
e-mail: wongtienyin@tsinghua.edu.cn

J. Wang
e-mail: jinyuanwang@tsinghua.edu.cn

R. M. W. Tseng · T. H. Rim · T. Y. Wong
Singapore Eye Research Institute, Singapore
National Eye Centre, Singapore, Singaporee-mail:
rachel.marjorie.tseng.w.w@seri.com.sg

T. H. Rim
e-mail: eyedoc.htrim@gmail.com

T. H. Rim
Ophthalmology and Visual Sciences Academic
Clinical Program (Eye ACP), Duke-NUS Medical
School, Singapore, Singapore

C. Y. Cheung
Department of Ophthalmology and Visual Sciences,
The Chinese University of Hong Kong, Hong Kong,
China
e-mail: carolcheung@cuhk.edu.hk

J. Wang · T. Y. Wong
Beijing Visual Science and Translational Eye
Research Institute (BERI), Beijing Tsinghua
Changung Hospital Eye Center, Tsinghua
Medicine, Tsinghua University, Beijing, China

T. Y. Wong
Beijing Key Laboratory of Intelligent Diagnostic
Technology and Devices for Major Blinding Eye
Diseases, Tsinghua University, Beijing, China

T. Y. Wong
Department of Ophthalmology, Peking Union
Medical College Hospital, Beijing, China

T. Y. Wong
State Key Laboratory of Ophthalmology,
Zhongshan Ophthalmic Center, Sun Yat-sen University,
Guangdong Provincial Key Laboratory
of Ophthalmology and Visual Science,
Guangdong Provincial Clinical Research Center
for Ocular Diseases, Guangzhou, China

examination results, such as carotid ultrasound, cardiac computerized tomography, and magnetic resonance imaging, to detect the systemic biomarkers (e.g., coronary artery calcium); (3) longitudinal studies that use AI-DL technology on CFPs to predict the incidence or risk of systemic disease (e.g., cardiovascular event or mortality). The range of systemic factors studied from CFPs via AI-DL approaches is reviewed based on these cross-sectional and longitudinal studies, and areas of future research are discussed while acknowledging the limitations that AI-DL on CFPs presents. In addition, with the updates and developments of eye imaging, more advanced ocular imaging like optical coherence tomography (OCT), OCT-angiography, ultra-wide field images, adaptive optics and external eye photos have also been applied in AI-DL technology to predict systemic diseases.

Keywords

Artificial intelligence · Deep learning · Colour fundus photography · Ocular imaging · Systemic biomarker · Systemic disease · Cardiovascular disease · Risk stratification

Introduction

Artificial intelligence (AI) technology is transforming healthcare. One major feature in ophthalmology is attributable to the ability to directly visualize retinal blood vessels, measuring 100–300 μm in size, via digital colour fundus photographs (CFPs). The use of CFPs offers a unique and easily accessible opportunity to study the human microcirculation [1] and characterize systemic diseases that manifest themselves in the retina. Identifying systemic disease risk factors through such means (i.e., via the eye) is an area of emerging research with major epidemiological and clinical studies that show

changes in the retinal vasculature can mirror systemic microcirculation changes.

To date, there are a number of large population-based studies that assessed the association between such retinal microvascular abnormalities and systemic diseases. This includes retinal vascular changes that are commonly found in hypertensive and/or diabetic retinopathy cases and their association with systemic diseases such as cardiovascular disease (CVD) and neurological diseases [2–8]. One example is that patients with Alzheimer's disease (AD) have some retinal features which may be early biomarkers of AD (i.e. narrower retinal vein calibre [9], more tortuous retinal vessels [10], and reduced blood flow [11], etc.), which may facilitate the early detection of AD [12]. The calibre of the retinal vessels (i.e. widths of retinal arterioles and venules) has been established as a promising biomarker for CVD risk estimation. For example, narrower retinal arterioles are associated with CVD and hypertension, a finding more common in females [13], while wider retinal venules was associated with an increased risk of stroke [14–16] and diabetes [17].

Building on such large-scale epidemiological and clinical studies, the application of AI technology, specifically in deep learning (DL), on ocular images, especially CFPs is advancing new research directions in retina-systemic disease relationships. This review provides a comprehensive summary of the various systemic-related outcomes that can be predicted via CFPs using AI-DL technology. The current studies fall into 3 basic groups: (1) cross-sectional studies that use AI-DL technology on CFPs to detect or estimate systemic risk factors (e.g., age, blood pressure, smoking); (2) cross-sectional studies that use AI-DL technology on CFPs to replace established examination results, such as carotid ultrasound, cardiac computerized tomography, and magnetic resonance imaging, to detect the systemic biomarkers (e.g., coronary artery calcium); (3) longitudinal studies that use AI-DL technology on CFPs to predict the incidence or risk of systemic disease (e.g., cardiovascular diseases event or mortality) [18] (Fig. 21.1).

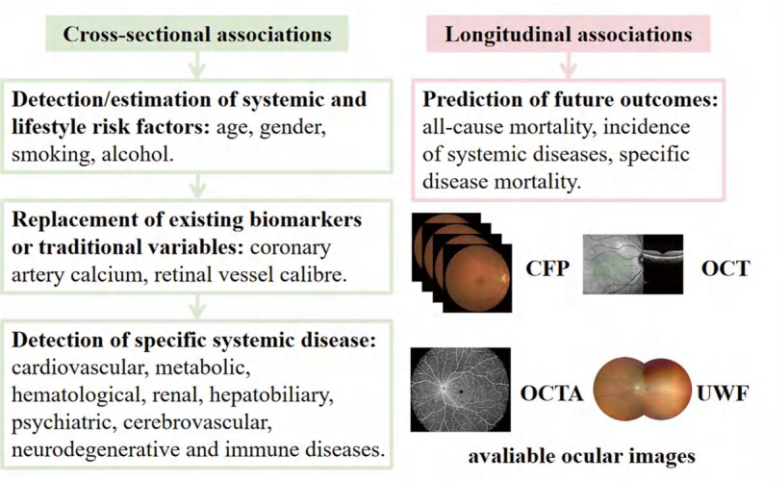


Fig. 21.1 Framework for artificial intelligence to evaluate systemic diseases via ocular images

Cross-Sectional Studies

Prediction of Demographic and Lifestyle Factors

Table 21.1 illustrates key studies in which AI-DL on CFPs have been used for the prediction of demographic and lifestyle factors. Among the identified studies, most of them investigated age as a predictable variable from CFPs. Chronological age is reliable at portraying growth milestones accurately [19], and the retina is considered the “window” to the whole body. Therefore, predicting age from CFPs via AI-DL could provide valuable information about the status of a target organ and/or the body [20]. In addition to age as a predictor, the ability to identify sex with high confidence from CFPs via AI-DL has been demonstrated in similar studies. For example, Rim et al. [21], showed fair results in their external multi-ethnic test sets for both age (coefficient of determination, $R^2=0.36\text{--}0.63$) and sex (area under a curve [AUC] = 0.80–0.91) predictions, demonstrating reasonable generalizability on predicting sex and age from CFPs. Nonetheless, although many interpretability work have been done to show the heat-maps during AI-DL predictions, it still remains unclear as to which part of the retina contributes to the prediction.

In terms of lifestyle factors, smoking status is commonly assessed because of the direct link between CVD and smoking habits. The effect of smoking on retinal vasculature was previously reported with studies showing that cigarette smoking was linked with a wider retinal venular caliber [25, 26]. Other studies have also demonstrated that one’s smoking status is associated with CVD due to the dual effects on both the retinal and systemic circulation, as suggested by visual changes in the retinal vascular structure [26–28]. Along with the fair results obtained from internal test sets of 3 unique studies (AUC = 0.71–0.86) [22–24], this allowed ophthalmic researchers to figure out with good confidence that smoking status prediction was predominantly predicted using the retinal vessels via AI-DL.

Prediction of Body Composition Factors

Table 21.2 summarizes the 3 main studies that used AI-DL to predict body composition factors based on CFPs. The association between an increased body-mass index (BMI) and mortality, in terms of stroke [29], cancer [30, 31], etc., has been established because BMI is a common measure of adiposity [32]. However, just like many other systemic factors, BMI prediction in

Table 21.1 Applications of AI in determining demographic and lifestyle factors

	Variables	Author, Year	Results	
			Internal test set	External test set
Demographic factors	Age	Poplin et al., 2018 [22]	MAE = 3.26 years CI = 3.22–3.31 $R^2 = 0.74$	NA
		Kim et al., 2020 [19]	Overall: MAE = 3.06 years $R^2 = 0.92$ Subgroup with HTN: MAE = 3.46 years $R^2 = 0.74$ Subgroup with DM: MAE = 3.55 years $R^2 = 0.75$ Subgroup with smokers: MAE = 2.65 years $R^2 = 0.86$	NA
		Rim et al., 2020 [21]	MAE = 2.43 years $R^2 = 0.83$	MAE = 3.38–4.50 years $R^2 = 0.36–0.63$
		Gerrits et al., 2020 [23]	MAE = 2.78 years $R^2 = 0.89$	NA
		Zhu et al., 2023 [20]	NA	MAE = 3.55 years $R^2 = 0.81$
	Sex	Poplin et al., 2018 [22]	AUC = 0.97 CI = 0.966–0.971	NA
		Kim et al., 2020 [19]	Overall: AUC = 0.97 HTN: AUC = 0.96 DM: AUC = 0.96 Smoker: AUC = 0.98	NA
		Rim et al., 2020 [21]	AUC = 0.96 Accuracy = 0.91	AUC = 0.80–0.91 Accuracy = 0.70–0.85
		Gerrits et al., 2020 [23]	AUC = 0.97 CI = 0.96–0.98 Accuracy = 0.93	NA
Environmental factors	Smoking status	Poplin et al., 2018 [22]	AUC = 0.71 CI = 0.70–0.73	NA
		Vaghefi et al., 2019 [24]	AUC = 0.86 Accuracy = 88.9%	NA
		Gerrits et al., 2020 [23]	AUC = 0.78 Accuracy = 0.81	NA

MAE mean absolute error, CI confidence interval, R^2 R-Square, HTN hypertension, DM diabetes mellitus, AUC area under the receiver operating characteristic curve, NA results not available

CFPs has not been suitable for clinical application yet, due to the great variability in mean absolute error (MAE) with a low generalisability across the ethnic groups in cohort studies [21, 22].

Recently, researchers discovered that body muscle mass is a more reliable measure of cardiometabolic risk than BMI. Using body muscle mass as a variable allowed for the detection of an age-related condition which reflects skeletal

muscle loss called sarcopenia [21]. Rim et al., developed an AI-DL model that enabled the quantification of muscle mass from CFPs. The MAE (6.09 kg) was high, and the coefficient of determination ($R^2 = 0.33$) was low in the external testing set, reiterating the need for further validation studies before assessing whether CFPs could be used as an alternative screening tool for sarcopenia.

Table 21.2 Applications of AI in determining body composition factors

	Variables	Author, Year	Results	
			Internal test set	External test set
Body composition factors	Body mass index	Poplin et al., 2018 [22]	MAE=3.29 units R ² =0.13 CI=3.24–3.34	NA
		Rim et al., 2020 [21]	MAE=2.15 kg/m ² CI=2.12–2.19 R ² =0.17	MAE=2.37–3.52 kg/m ² R ² =0.01–0.14
	Body muscle mass	Rim et al., 2020 [21]	MAE=5.11 kg CI=5.04–5.19 R ² =0.52	MAE=6.09 kg CI=5.96–6.23 R ² =0.33
	Height	Rim et al., 2020 [21]	MAE=5.20 cm CI=5.13–5.28 R ² =0.42	MAE=5.48–7.09 cm R ² =0.08–0.28
	Body weight	Rim et al., 2020 [21]	MAE=7.69 kg CI=7.57–7.81 R ² =0.36	MAE=8.28–11.81 kg R ² =0.04–0.19
	Relative fat mass/ percentage body fat	Gerrits et al., 2020 [23]	MAE=5.68 units R ² =0.43	NA
		Rim et al., 2020 [21]	MAE=4.71 kg CI=4.64–4.78 R ² =0.23	MAE=4.50 kg CI=4.39–4.60 R ² =0.08

MAE mean absolute error, CI confidence interval, R2 R-Square, NA results not available

Prediction of Neurological and Psychiatric Diseases

The retina and the brain share a special relationship since both structures develop from the neural tube and are part of the central nervous system [33]. Embryonic origin aside, visual changes have also been reported as the first few symptoms in many patients diagnosed with AD [33, 34], with some studies showing the aggregation of amyloid beta monomers in the retina in these patients [35]. In the realm of AI-DL, existing studies vary in the type of ocular imaging technologies used to explore the association between neurological diseases and the retina [35–37]. In particular, Lim et al., evaluated the potential of an AI-DL model as an ischemic stroke risk assessment from CFPs, and this resulted in a varying AUC of 0.685–0.994 for 6 different datasets [38]. Additionally, the team also found that retinal vessel calibre could be predictive of ischemic stroke in patients although there was low generalizability when the model was tested in unseen images [38]. Recent years have also witnessed the AI-DL

models in detecting stroke [39] and ischemic stroke [40], however, due to the data limitation of cerebrovascular diseases compared to other major systemic disease, there is still much room to explore. Considering the emerging potential of retinal imaging as a non-invasive strategy, employing AI-DL in CFPs could act as opportunistic screening for neurological and psychiatric diseases, and ultimately increase screening adherence rates in the community (Table 21.3).

Prediction of Cardiovascular and Circulatory Disorders

Table 21.4 details the AI-DL studies focused on predicting systemic risk factors and specific diseases (e.g., anaemia and hypertension) of the cardiovascular and circulatory system. Blood pressure (BP) is an important indicator for CVD, but it is also a means to maintain homeostasis and therefore fluctuates based on body and emotional status. Using the retina as a biomarker instead is preferred because the retina shows accumulated damage due to high

Table 21.3 Applications of AI in predicting systemic diseases involved with the neurological and psychiatric system

	Systemic diseases/ variables	Author, Year	Results	
			Internal test set	External test set
AD	AD amyloid β positive AD+ amyloid β positive	Cheung et al., 2022 [41]	AD: AUC = 0.73–0.93 Amyloid β positive: AUC = 0.68–0.86 AD+ amyloid β positive: AUC = 0.73–0.85	NA
Schizophrenia	Schizophrenia	Appaji et al., 2022 [42]	AUC = 0.98	NA
PD	PD (H-Y scale and UPDRSIII score)	Ahn et al., 2023 [43]	H-Y scale: AUC = 0.77 UPDRS-III score: AUC = 0.77	AUC = 0.67
MMD screening	MMD screening	Hong et al., 2024 [44]	MMD screening: AUC = 94.6%	NA
MMD staging	MMD staging		MMD staging 1/2/3/4/5: AUC = 78.9%/80.6%/93.6%/91.8%/88.1%	

AD Alzheimer's disease, PD Parkinson's disease, MMD Moyamoya disease, H-Y Hoehn and Yahr, UPDRS-III Unified Parkinson's Disease Rating Scale part III, AUC area under the receiver operating characteristic curve, NA results not available

blood pressure and experiences comparably less fluctuation compared to conventional BP stable marker, rendering it a more stable marker. The results of applying AI-DL to date show that unlike other body factors (e.g., height and weight), there is generalisability across the different ethnic groups for blood pressure prediction [6]. However, the R^2 value is somewhat low with ranges of 0.24–0.40. Apart from using blood pressure as a biomarker for hypertension, disease prediction of hypertension was also reported by Dai et al. [45], and Zhang et al. [46], but modest predictability was observed with a combined AUC of 0.651–0.766.

For the prediction of anaemia and/or related biomarkers, biomarkers including hemoglobin (Hb), haematocrit, and red blood cell were predicted from CFPs in 3 separate studies [21, 23, 47]. Anaemia was also predictable using an AI-DL model developed by Mitani et al., with a modest AUC of 0.88 using a combined model of systemic risk factors and CFPs [47]. Rim et al., tested these hematologic factors in external datasets with varying ethnicities but generalizability was limited among other ethnic groups [21].

Prediction of Metabolic and Endocrinological Diseases

Table 21.5 details the studies on the endocrine system. Of the different systemic diseases and/or variables that were tested using the AI-DL model, the prediction of biomarkers related to diabetes, including glucose and HbA1c, showed somewhat low predictive performance in both the internal and external test sets.

Testosterone (MAE = 3.76 nmol/L, R^2 = 0.54) was predictable from CFPs but Gerrits et al., discovered that the AI-DL model that was trained to predict testosterone levels, was indirectly predicting sex as well. The team additionally found that the performance of the model was affected when it was trained on solely males or females, indicating that sex had an indirect effect on the performance of the model when predicting for testosterone [23]. Apart from systemic risk factors and related biomarkers, endocrine system-related diseases were also reported, including a machine learning system created by Benson et al., to predict diabetic peripheral neuropathy, which showed relatively good performance (AUC = 0.89) [49]. The prediction of other

Table 21.4 Applications of AI in predicting systemic diseases involved with cardiovascular and circulatory system

	Systemic diseases/ variables	Author, Year	Results	
			Internal test set	External test set
Hypertension or related biomarkers	SBP, DBP/ mmHg	Poplin et al., 2018 [22]	SBP: MAE = 11.23 mmHg CI = 11.18–11.51 R ² = 0.36 DBP: MAE = 6.42 mmHg CI = 6.33–6.52 R ² = 0.32	NA
		Rim et al., 2020 [21]	SBP: MAE = 9.29 mmHg CI = 9.16–9.43 R ² = 0.31 DBP: MAE = 7.20 mmHg CI = 7.09–7.30 R ² = 0.35	SBP: MAE = 10.55–13.95 mmHg R ² = 0.17–0.21 DBP: MAE = 7.14–8.09 mmHg R ² = 0.16–0.27
		Gerrits et al., 2020 [23]	SBP: MAE = 8.96 mmHg R ² = 0.40 DBP: MAE = 6.84 mmHg R ² = 0.24	NA
	Hypertension	Dai et al., 2020 [45]	AUC = 0.651	NA
		Zhang et al., 2020 [46]	AUC = 0.766	NA
Anaemia or related biomarkers	Anaemia	Mitani et al., 2020 [47]	Metadata/fundus/combined: AUC = 0.73/0.87/0.88 AUC = 0.89 (combined in diabetes subgroup)	NA
	Hemoglobin levels	Mitani et al., 2020 [47]	Metadata/fundus/combined: MAE = 0.73/0.67/0.64 g/dL CI = 0.72–0.74/0.66–0.68/0.62–0.64 AUC = 0.74/0.87/0.88	NA
		Rim et al., 2020 [21]	MAE = 0.79 g/dL CI = 0.78–0.80 R ² = 0.56	MAE = 0.93–0.98 g/dL R ² = 0.06–0.33
		Gerrits et al., 2020 [23]	MAE = 0.61% R ² = 0.34	NA
	Haematocrit levels	Mitani et al., 2020 [47]	Metadata/fundus/combined: MAE = 2.10/1.94/1.83% CI = 2.07–2.13/1.91–1.97/1.80–1.86	NA
		Rim et al., 2020 [21]	MAE = 2.03% CI = 2.00–2.06 R ² = 0.57	MAE = 2.62–2.81% R ² = 0.09–0.26
	Red blood cell count	Mitani et al., 2019 [47]	Metadata/fundus/combined: MAE = 0.26/0.26/0.25·10 ¹² /L CI = 0.26–0.27/0.25–0.26/0.25–0.25	NA
		Rim et al., 2020 [21]	MAE = 0.26·10 ¹² /L CI = 0.25–0.26 R ² = 0.45	MAE = 0.33–0.37·10 ¹² /L R ² = -0.02–0.14

SBP systolic blood pressure, DBP diastolic blood pressure, MAE mean absolute error, CI confidence interval, R² R-Square, AUC area under the receiver operating characteristic curve, NA results not available

Table 21.5 Applications of AI in predicting systemic diseases involved with the metabolic and endocrinological system

	Systemic diseases/variables	Author, Year	Results	
			Internal test set	External test set
Diabetes or related biomarkers	Diabetes/blood glucose control	Rim et al., 2020 [21]	Fasting blood glucose: MAE = 8.55 mg/dL CI = 8.40–8.71 R ² = 0.11	MAE = 10.10 mg/dL CI = 9.83–10.36 R ² = 0.05
		Babenko et al., 2020 [48]	AUC = 0.702	NA
	Diabetic peripheral neuropathy	Benson et al., 2020 [49]	Accuracy = 89%	NA
	Hyperglycemia	Zhang et al., 2020 [46]	AUC = 0.880	NA
	HbA1c	Poplin et al., 2018 [22]	MAE = 1.39% CI = 1.29–1.50 R ² = 0.09	NA
		Rim et al., 2020 [21]	MAE = 0.33% CI = 0.32–0.33 R ² = 0.13	MAE = 0.35 CI = 0.34–0.36 R ² = 0.07
		Gerrits et al., 2020 [23]	MAE = 0.61% R ² = 0.34	NA
Lipid related biomarkers	Dyslipidaemia	Zhang et al., 2020 [46]	AUC = 0.703	NA
	HDL cholesterol	Rim et al., 2020 [21]	MAE = 9.45 mg/dL R ² = 0.13	MAE = 9.46 mg/dL R ² = 0.08
Other biomarkers	Testosterone	Gerrits et al., 2020 [23]	MAE = 3.76 nmol/L R ² = 0.54	NA

HbA1c hemoglobin A1C, *HDL* high-density lipoprotein, *MAE* mean absolute error, *CI* confidence interval, *R²* R-Square, *AUC* area under the receiver operating characteristic curve, *NA* results not available

conditions such as dyslipidaemia and diabetes showed moderate predictive performance.

Prediction of Kidney Disease

Among systemic diseases, chronic kidney disease (CKD) is infamously described by the American Society of Nephrology to be a silent killer. The traditional way of screening for CKD is invasive and includes collecting serum creatinine levels [50]. As such, screening for CKD is limited and a tough challenge for most communities. Applying AI to CFPs for the detection of CKD would not only act as an adjunct screening tool for CKD but could also be a gamechanger in increasing the detection rate and lowering the mortality rate of patients with CKD.

Despite the huge potential impact, few studies have explored the prediction of CKD from CFPs. Of note, Sabanayagam et al. [51], predicted CKD (AUC = 0.73) with modest

generalisability and a separate AI-DL system developed by Kang et al. [52], achieved an AUC of 0.81 although no external validation was conducted (Table 21.6). In particular, the performance of the model developed by Sabanayagam et al. [51], was generally stable (AUC > 0.9 for all models) across the different models (i.e. CFPs, risk factors, combined) that were trained, suggesting that risk factor information was not required for CKD risk assessment in patients. In the study, CKD was strictly defined as an estimated glomerular filtration rate of less than 60 units. This estimation was based on factors such as creatinine level, age, sex, and bodyweight. Given that predicting age and sex from AI-DL in CFPs is well-established and highly accurate, this suggests the high potential of using AI-DL in CFPs to predict CKD in the future as an opportunistic screening method that can be integrated into CFPs and an eventual replacement of traditional invasive methods.

Table 21.6 Applications of AI in predicting systemic diseases involved with the renal system

	Systemic diseases/ variables	Author, Year	Results	
			Internal test set	External test set
CKD or related biomarkers	CKD	Sabanayagam et al., 2020 [51]	AUC (CFP/RF/ Combined) = 0.911/0.916/0.938 Subgroup of DM patients: AUC = 0.889/0.899/0.925 Subgroup of HTN patients: AUC = 0.889/0.889/0.918	AUC (CFP/RF/ Combined) = 0.733–0.835/0.829–0.887/0.810–0.858
		Kang et al., 2020 [52]	Overall AUC = 0.81 AUC = 0.81–0.87 as HbA1c levels increased from <6.5% to >10%	NA
	CKD early CKD eGFR incident CKD	Zhang et al., 2021 [53]	CKD: AUC = 0.930 early CKD: AUC = 0.864 eGFR: $R^2 = 0.507$, PCC = 0.716, MAE = 11.1 mL/min/1.73 m ² incident CKD: HR = 1.87	CKD: AUC = 0.898/0.897 (A/B) early CKD: AUC = 0.848/0.845 eGFR: A: $R^2 = 0.481$, PCC = 0.700, MAE = 12.9 mL/min/1.73 m ² eGFR: B: $R^2 = 0.327$, PCC = 0.577, MAE = 11.8 mL/min/1.73 m ² incident CKD: HR = 2.21 (A: Guangdong province, B: COACS)
	DKD	Betzler et al., 2023 [54]	AUC(CFP/RF/ Combined) = 0.826/0.847/0.866	SEED: AUC = 0.764/0.802/0.828 SMART2D: AUC = 0.726/0.701/0.761
	Creatinine	Rim et al., 2020 [21]	MAE = 0.11 CI = 0.11–0.11 $R^2 = 0.38$	MAE = 0.11–0.17 $R^2 = 0.01–0.26$

CKD chronic kidney disease, eGFR estimated glomerular filtration rate, DKD diabetic kidney disease, AUC area under the receiver operating characteristic curve, CFP color fundus photos, RF risk factor, DM diabetes mellitus, HTN hypertension, HbA1c hemoglobin A1C, R^2 R-Square, PCC Pearson correlation coefficient, MAE mean absolute error, HR hazard ratio, CI confidence interval, NA results not available, COACS China suboptimal health cohort study, SEED Singapore Epidemiology of Eye Diseases, SMART2D Singapore Macroangiopathy and Microvascular Reactivity in Type 2 Diabetes

Other Retinal Biomarkers

Table 21.7 details the studies focused on other retinal imaging biomarkers. There is strong evidence from epidemiological studies that changes in the retinal vasculature mirror systemic microcirculation changes. However, the process for assessing retinal vascular changes is time-consuming and requires professional training, which has limited the expansion and wider application of these traditional methods in other primary care settings outside of ophthalmology [55]. To address these challenges, semi-automated software, such as the Singapore I Vessel Assessment deep-learning system (SIVA-DLS), was created

and applied on CFP, allowing for a more efficient, objective, and quantifiable way of assessing width of retinal vessels from CFP through the use of heat maps (Fig. 21.2). The SIVA-DLS study reported high intra-class correlation coefficients (0.82–0.95) between the SIVA-DLS and validated human measurements.

Coronary artery calcium (CAC) is a pre-clinical marker of atherosclerosis, a cardiovascular condition that could implicate the circulatory system, and is strongly associated with risk of clinical CVD [59]. Measurement of CAC scores has increasingly been used for stratification of CVD risk. Recently, Son et al. created an AI-DL model to predict abnormal CAC from CFP both

Table 21.7 Applications of AI in predicting other established imaging biomarker

	Systemic diseases/variables	Author, Year	Results	
			Internal test set	External test set
Other imaging Biomarkers	Retinal vessel calibre	Cheung et al., 2020 [56]	ICC=0.88–0.95	ICC=0.69–0.92
	Coronary artery calcification	Son et al., 2020 [57]	AUC=0.823–0.832	NA
	Sonographically confirmed carotid artery atherosclerosis	Chang et al., 2020 [58]	AUC=0.713	NA

ICC intraclass correlation coefficient, AUC area under the receiver operating characteristic curve, NA results not available

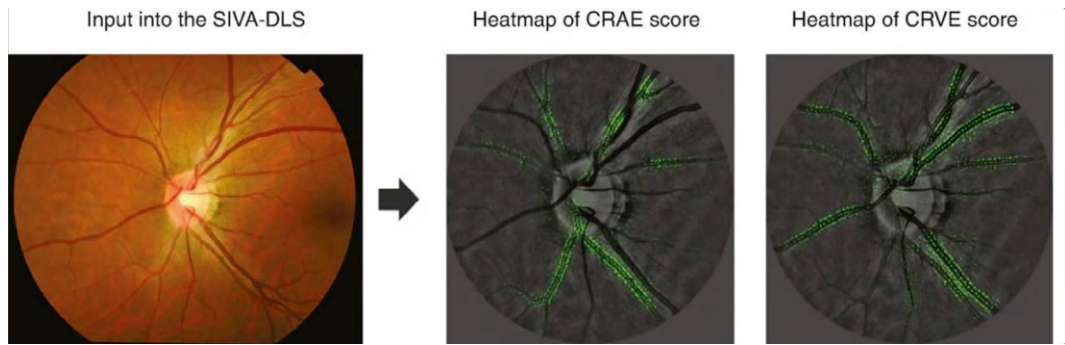


Fig. 21.2 Use of SIVA-DLS to assess width of retinal vessels in a more efficient, objective, and quantifiable way

unilaterally and bilaterally and the performance (AUC=0.823–0.832) was promising [57]. In addition, Carotid Intima Media Thickness (CIMT) was measured using ultrasonography by averaging three measurements made 10 mm proximal to the bifurcation and was used as the proxy marker for atherosclerosis. Chang et al. developed an AI-DL model to predict carotid artery atherosclerosis, and the model was able to predict the sonographically confirmed carotid artery atherosclerosis with an AUC of 0.713 [58]. Both CAC and CIMT models were not tested externally and further validation is required prior to assessing the clinical applicability of these models.

Longitudinal Studies

Leveraging the predictive values and cross-sectional outcomes that are estimated using AI-DL and CFPs, research in the retinal biomarkers field has also been expanded towards predicting

future events (Table 21.8). Since applying AI-DL in CFPs is sufficient for risk factors prediction of systemic diseases, there is a high possibility that CFPs could also be directly associated with and therefore a good predictor of the incidence of CVD events. Recent work includes the survival analysis of risk stratification for incident CVD events and mortality using the predicted probability of CVD occurrence from CFPs at baseline. The conventional Cox proportional hazard model is an extension of current efforts at creating an optimal predictive model using deep features of CFPs as an input. In the past, variables such as, age, sex, socioeconomic status, and other CVD risk factors were manually or statistically selected for survival analysis, but presently, the Cox model is generated by using the deep features of CFPs that are observed based on the association between CFPs and various risk factors via DL. Apart from this hybrid model, other papers have used different methods ranging from neural networks to machine learning techniques. This includes

Cox-nnet [60], DeepSurv [61], and Nnet-survival [62]. Currently, no studies that have applied this recent network with CFPs to predict systemic-related event outcome using time-series data, suggesting that there is room to explore the relationship between the incidence of systemic disease outcomes and the retina.

Of the various studies that investigate the applications of AI-DL in predicting longitudinal outcomes of systemic diseases, a few should be mentioned. Poplin et al. [22] predicted CVD risk factors from CFPs and thereafter used the results to predict Major Adverse Cardiac Events (MACE) over 5 years in the UK Biobank. The performance of the AI-DL model was of similar to the performance of the European Systematic Coronary Risk Evaluation (SCORE) risk calculator [22]. Another study by Chang et al. used proxy markers such as CIMT and existence of carotid artery plaque to train an AI-DL model to predict atherosclerosis [49]. The study demonstrated that the retinal biomarkers was significantly associated with an increased risk, represented by hazard ratio, for CVD mortality after adjusting for the Framingham risk score (FRS). In addition, Cheung et al. looked at CVD and its association to retinal vessel calibre [47, 56]. The team found that a narrow central retinal arteriolar equivalent measured by SIVA-DLS was associated with incident CVD and all-cause mortality in 2 prospective cohorts [47]. Subsequently, the author then found the correlation between narrower retinal vessel calibre and incident cognitive decline and dementia [56]. Lastly, Zhu et al. used data from the UK biobank to demonstrate that the difference between one's age and the predicted age from CFP (retinal age gap) independently predicted one's mortality risk and that the mortality risk was positively associated with the retinal age gap [20]. Later, the team continued the study in retinal age gap and various systemic diseases' correlations, and found significant results that retinal age gap increase the risks of diseases, in CVD [57], stroke [58], Parkinson's disease [55], kidney failure [59], CKD and subsequent CVD complications [60]. These studies show the potential of CFPs as a screening tool

for risk stratification and delivery of tailored interventions.

Areas of Future Research

Table 21.9 shows the challenges and future considerations on AI-DL models in predicting systemic diseases via ocular images in various aspects [18, 69].

When assessing the performance of AI-DL models, the predictability and generalisability of the models must be evaluated appropriately. Predictability refers to how accurate the models are at predicting the desired variables. The AI-DL model that predicts age from CFPs is an example given the relatively high coefficient of determination in comparison to the somewhat low coefficient of determination that was produced by the AI-DL model that predicts BP [22]. There are no specific guidelines or cut-off values for coefficient of determination, but the predictability of AI-DL systems on specific biomarkers can be inferred from the performance of these systems on internal test sets. Predictability, however, does not guarantee generalisability. In terms of generalisability, it is crucial to determine how well the AI-DL models will perform in different clinical settings and different populations with varying ethnicities. Rim et al., has demonstrated this by testing their AI-DL models on external multi-ethnic testing datasets [21]. In this case, the AI-DL model that predicts BP from CFPs demonstrates good generalisability even though the coefficient of determination is not as high as it is for age prediction. Taking into account these 2 performance indices when developing AI-DL models would substantiate the use of retinal markers as a surrogate marker and proxy for conventional markers in the future.

Additional challenges that could implicate the performance of AI-DL models on CFPs include the limited number of datasets available for systemic diseases, the adoption of retinal examinations into CVD guidelines, as well as gaining acceptance from physicians, patients, and the public. In particular, the collation of

Table 21.8 Applications of AI-DL in predicting longitudinal outcomes of systemic diseases

Outcome	Author, Year	Ground truth (Data)	Datasets	Results
CVD events	Poplin et al., 2018 [22]	Age, gender, SBP, BMI, current smoker	UK Biobank	AUC = 0.70 CI = 0.648–0.740
	Cheung et al., 2020 [56]	Retinal vessel calibre (the SEED study)	SEED, SP2, Dunedin, HKCES, AHES, RICP, IRED, CUHK-STDTR, GUSTO, SIDRP, CVD screening study, BES, UK Biobank, KSH, Austin Health study	After adjusting factors including demographic factors, BMI, total cholesterol level, etc., the narrower CRAE _B (HR = 1.12/OR = 1.88) and narrower CRAE _C (HR = 1.13/OR = 1.67) were independently associated with incident CVD events and 10-year all-cause mortality in SEED study and BES study respectively (HR value for SEED study, OR value for BES study)
CVD mortality	Chang et al., 2020 [58]	Carotid artery atherosclerosis	HPC-SNUH	AUC = 0.713 for predicting carotid artery atherosclerosis Patients with a DL-FAS greater than 0.66 has a significantly higher risk of CVD mortality compared to those with score below 0.33 (HR = 8.33) Risk association between DL-FAS and CVD mortality was significant in patients with intermediate to high Framingham risk scores
All-cause mortality	Zhu et al., 2023 [20]	Age	UK Biobank	The correlation between retinal age and chronological age (retinal age gap) was strong (0.81, $P < 0.001$, MAE = 3.55 years) Cox regression models showed that each 1 year increase in the retinal age gap, the mortality risk increased by 2% (HR = 1.02)
Incident CVD	Zhu et al., 2022 [63]	Age	UK Biobank	Each 1-year increase in retinal age gap was associated with a 3% increase in the risk of incident CVD (HR = 1.03)

(continued)

Table 21.8 (continued)

Outcome	Author, Year	Ground truth (Data)	Datasets	Results
Incident cognitive decline and dementia	Cheung et al., 2022 [64]	Retinal vessel calibre	National University Hospital and St. Luke's Hospital in Singapore	Narrower retinal arteriolar calibre (HR per standard deviation decrease = 1.258) and wider retinal venular calibre (HR per standard deviation increase = 1.204) were associated with increased risk of cognitive decline Narrower retinal arteriolar calibre was also associated with incident dementia (HR per standard deviation decrease = 1.624)
Incident stroke	Zhu et al., 2022 [65]	Age	UK Biobank	Each 1-year increase in the retinal age gap was associated with a 4% increase in the risk of stroke (HR = 1.04)
Incident Parkinson's disease	Hu et al., 2022 [66]	Age	UK Biobank	1-year increase in retinal age gap was associated with a 10% increase in risk of Parkinson's disease (HR = 1.10)
Incident kidney failure	Zhang et al., 2023 [67]	Age	UK Biobank	Each 1-year greater retinal age gap at baseline was independently associated with a 10% increase in the risk of incident kidney failure (HR = 1.10)
Incident CKD and subsequent CVD complications	Wu et al., 2024 [68]	Age	UK Biobank	Each 1-year increase in retinal age gap was associated with a 3% increase in the risk of CKD incidence (HR = 1.03) Each 1-year increase in retinal age gap was associated with a 10% increase in the risk of incident CVD (HR = 1.10)

CVD cardiovascular disease, CKD chronic kidney disease, SBP systolic blood pressure, BMI body mass index, SEED Singapore Epidemiology of Eye Diseases, SP2 Singapore Prospective study program, HKCES Hong Kong children eye study, AHES Australian heart eye study, RICEP retinal imaging in chest pain study, IRED retinal imaging in renal disease study, CUHK-STR The Chinese University of Hong Kong sight-threatening diabetic retinopathy study, GUSTO growing up in Singapore towards healthy outcomes birth cohort, SiDRP Singapore integrated diabetic retinopathy program, BES Beijing Eye Study, KSH Kangbuk Samsung health study, HPC-SNUH Health Promotion Center of Seoul National University Hospital, Korea, AUC area under the receiver operating characteristic curve, CI confidence interval, CRAE_B central retinal artery equivalent at zone B, HR hazard ratio, OR odds ratio, CRAE_C central retinal artery equivalent at zone C, DL-FAS deep-learning fundoscopic atherosclerosis score, MAE mean absolute error

Table 21.9 The challenges and future considerations on AI-DL models in predicting systemic diseases via ocular images in various aspects

Challenges	Opportunities
Limited datasets, limited real-world studies in rare diseases, limited longitudinal cohort and comprehensive biobanks	Call for more real-word studies in local hospitals and healthcare centers. The comprehensive biobank construction and data integration are important, including consortium actions to combine data together to support big data based AI-DL models training and testing
Limited modalities, including easily accessible images (i.e. external eye photos) and advanced high-resolution images	More modalities like SS-OCT, OCTA, UWF, Vis-OCT, PS-OCT, PD-OCT, AO-OCT, oximetry and etc. can be candidates in futural studies. Even more functional examination also can be candidates, such as microperimetry, visual acuity, intraocular pressure and etc.
Limited multi-modality algorithms and limited paired multi-modality data	The combination of more than 1 modality in training AI-DL algorithm may increase the efficacy
The division of specialties	Interdisciplinary integration will benefit patients and public
Controversial use of synthetic images	The exploding development of generative AI may contribute to the data limitation and direct the AI-DL models toward easily implimentable vision-language foudation models
Controversial predictability (i.e. accuracy, interpretability) and generalisability (i.e. usability and reliability)	Call for large-scale studies to strengthen AI-DL models' performance, including internal test, external test and interpretability studies
FDA/EMA approval difficulties, facing of ethics, liability, regulations and costs	Call for more RCT clinical trials over the world through evidence, validation, mechanistic understanding and human-machine interaction testing

FDA United States Food and Drug Administration, *EMA* European Medicines Agency, *AI* artificial intelligence, *DL* deep learning, *SS-OCT* sweep source optical coherence tomography, *OCTA* optical coherence tomography angiography, *UWF* ultra-wide field, *Vis-OCT* visible wavelength OCT, *PS-OCT* polarization-sensitive OCT, *PD-OCT* polarization-diversity OCT, *AO-OCT* adaptive optics OCT, *RCT* randomized controlled trial

CFPs datasets according to systemic diseases is tough given the division of specialties, with ophthalmology being commonly separated from other departments. The division of specialties also makes the incorporation of retinal examinations into other clinical pathways and therefore CVD guidelines more challenging. Additionally, when a new AI-DL system is created and has the potential of being enforced in clinical practice, it is imperative for researchers to prove and simultaneously convince different stakeholders, including physicians, patients, and the public, that the specific AI-DL model benefits healthcare systems, addresses current unmet needs, and would add substantial value to existing clinical settings and technologies prior to its employment in real-world settings.

Recently, RETFound, the first AI foundation model in ophthalmology, which was developed using millions of ocular images from

the National Health Service (NHS) by Zhou et al. [70], enabled AI-DL techniques to predict ischaemic stroke, myocardial infarction, heart failure, Parkinson’s disease, via retinal images. Notably, with the rapid updating of generative AI technologies, especially various popular vision language models, Chen et al. [71] constructed Fundus fluorescein angiography (FFA)-GPT, the world’s first multimodal integrated intelligent system of visual-large language model for ophthalmic image interpretation. The trend of the completion and generalisability of AI application in ophthalmic area will be of more importance.

Conclusion

This review presents systemic diseases factors that have been predicted from the retina using ocular images (Fig. 21.1). Various studies have

expressed the versatility of AI-DL technology in assessing systemic diseases and risk factors via ocular images, especially CFPs. Recent years, there were also several reviews concluded the studies in this area and summarized study methods of AI-DL application and further limitations and considerations, which gave us inspirations in this prospective area [18, 69, 72–75]. Further efforts to discover other systemic risk factors and biomarkers that could be predicted from the ocular images are underway. To date, there remains insufficient prospective studies and lack of evidence in real-world settings and therefore the clinical application and implementation of AI-DL models using ocular images is limited. Moving forward, prospective studies conducted in real-world settings will be needed as evidence that the AI-DL models are beneficial and to ensure that AI-DL models are safe and cost-effective.

References

- Wagner SK, Fu DJ, Faes L, Liu X, Huemer J, Khalid H, et al. Insights into systemic disease through retinal imaging-based oculomics. *Transl Vis Sci Technol*. 2020;9(2):6.
- Rim TH, Teo AWJ, Yang HHS, Cheung CY, Wong TY. Retinal vascular signs and cerebrovascular diseases. *J Neuroophthalmol*. 2020;40(1):44–59.
- McGeechan K, Liew G, Macaskill P, Irwig L, Klein R, Klein BE, et al. Prediction of incident stroke events based on retinal vessel caliber: a systematic review and individual-participant meta-analysis. *Am J Epidemiol*. 2009;170(11):1323–32.
- Wong TY, McIntosh R. Systemic associations of retinal microvascular signs: a review of recent population-based studies. *Ophthalmic Physiol Opt*. 2005;25(3):195–204.
- Lim M, Sasongko MB, Ikram MK, Lamoureux E, Wang JJ, Wong TY, et al. Systemic associations of dynamic retinal vessel analysis: a review of current literature. *Microcirculation*. 2013;20(3):257–68.
- Sabanayagam C, Lye WK, Klein R, Klein BE, Cotch MF, Wang JJ, et al. Retinal microvascular calibre and risk of diabetes mellitus: a systematic review and participant-level meta-analysis. *Diabetologia*. 2015;58(11):2476–85.
- Kim DH, Chaves PHM, Newman AB, Klein R, Sarnak MJ, Newton E, et al. Retinal microvascular signs and disability in the Cardiovascular Health Study. *Archiv Ophthalmol* (Chicago, Ill: 1960). 2012;130(3):350–6.
- Wong TY, McIntosh R. Hypertensive retinopathy signs as risk indicators of cardiovascular morbidity and mortality. *Br Med Bull*. 2005;73–74:57–70.
- Kesler A, Vakhapova V, Korczyn AD, Naftaliev E, Neudorfer M. Retinal thickness in patients with mild cognitive impairment and Alzheimer's disease. *Clin Neurol Neurosurg*. 2011;113(7):523–6.
- Cheung CY, Ong YT, Ikram MK, Ong SY, Li X, Hilal S, et al. Microvascular network alterations in the retina of patients with Alzheimer's disease. *Alzheimers Dement*. 2014;10(2):135–42.
- Feke GT, Hyman BT, Stern RA, Pasquale LR. Retinal blood flow in mild cognitive impairment and Alzheimer's disease. *Alzheimers Dement* (Amst). 2015;1(2):144–51.
- Frost S, Kanagasalingam Y, Sohrabi H, Vignarajan J, Bourgeat P, Salvado O, et al. Retinal vascular biomarkers for early detection and monitoring of Alzheimer's disease. *Transl Psychiatry*. 2013;3(2): e233.
- McGeechan K, Liew G, Macaskill P, Irwig L, Klein R, Klein BEK, et al. Meta-analysis: retinal vessel caliber and risk for coronary heart disease. *Ann Intern Med*. 2009;151(6):404–13.
- Cheung CY, Tay WT, Ikram MK, Ong YT, De Silva DA, Chow KY, et al. Retinal microvascular changes and risk of stroke: the Singapore Malay Eye Study. *Stroke*. 2013;44(9):2402–8.
- Kawasaki R, Xie J, Cheung N, Lamoureux E, Klein R, Klein BE, et al. Retinal microvascular signs and risk of stroke: the Multi-Ethnic Study of Atherosclerosis (MESA). *Stroke*. 2012;43(12):3245–51.
- Wong TY, Klein R, Couper DJ, Cooper LS, Shahar E, Hubbard LD, et al. Retinal microvascular abnormalities and incident stroke: the atherosclerosis risk in communities study. *Lancet*. 2001;358(9288):1134–40.
- Nguyen TT, Wang JJ, Sharrett AR, Islam FMA, Klein R, Klein BEK, et al. Relationship of retinal vascular caliber with diabetes and retinopathy. The multi-ethnic study of atherosclerosis (MESA). *Diabetes Care*. 2008;31(3):544–9.
- Jinyuan, Wang Ya Xing, Wang Dian, Zeng Zhuoting, Zhu Dawei, Li Yuchen, Liu Bin, Sheng Andrzej, Grzybowski Tien Yin, Wong. Artificial intelligence-enhanced retinal imaging as a biomarker for systemic diseases. *Theranostics*. 2025;15(8):3223–3233. <https://doi.org/10.7150/thno.100786>.
- Kim YD, Noh KJ, Byun SJ, Lee S, Kim T, Sunwoo L, et al. Effects of hypertension, diabetes, and smoking on age and sex prediction from retinal fundus images. *Sci Rep*. 2020;10(1):4623.
- Zhu Z, Shi D, Guankai P, Tan Z, Shang X, Hu W, Liao H, Zhang X, Huang Y, Yu H, Meng W, Wang W, Ge Z, Yang X, He M. Retinal age gap as a predictive biomarker for mortality risk. *Br J Ophthalmol*. 2023;107(4):547–54.
- Rim TH, Lee G, Kim Y, Tham YC, Lee CJ, Baik SJ, et al. Prediction of systemic biomarkers from

- retinal photographs: development and validation of deep-learning algorithms. *Lancet Digit Health*. 2020;2(10):e526–36.
22. Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng*. 2018;2(3):158–64.
 23. Gerrits N, Elen B, Craenendonck TV, Triantafyllidou D, Petropoulos IN, Malik RA, et al. Age and sex affect deep learning prediction of cardiometabolic risk factors from retinal images. *Sci Rep*. 2020;10(1):9432.
 24. Vaghefi E, Yang S, Hill S, Humphrey G, Walker N, Squirrell D. Detection of smoking status from retinal images; a convolutional neural network study. *Sci Rep*. 2019;9(1):7180.
 25. Kifley A, Liew G, Wang JJ, Kaushik S, Smith W, Wong TY, et al. Long-term effects of smoking on retinal microvascular caliber. *Am J Epidemiol*. 2007;166(11):1288–97.
 26. Ikram MK, de Jong FJ, Vingerling JR, Witteman JC, Hofman A, Breteler MM, et al. Are retinal arteriolar or venular diameters associated with markers for cardiovascular disorders? The Rotterdam study. *Invest Ophthalmol Vis Sci*. 2004;45(7):2129–34.
 27. Sun C, Wang JJ, Mackey DA, Wong TY. Retinal vascular caliber: systemic, environmental, and genetic associations. *Surv Ophthalmol*. 2009;54(1):74–95.
 28. Kifley A, Wang JJ, Cugati S, Wong TY, Mitchell P. Retinal vascular caliber, diabetes, and retinopathy. *Am J Ophthalmol*. 2007;143(6):1024–6.
 29. Song YM, Sung J, Davey Smith G, Ebrahim S. Body mass index and ischemic and hemorrhagic stroke: a prospective study in Korean men. *Stroke*. 2004;35(4):831–6.
 30. Reeves GK, Pirie K, Beral V, Green J, Spencer E, Bull D. Cancer incidence and mortality in relation to body mass index in the Million Women study: cohort study. *BMJ*. 2007;335(7630):1134.
 31. Calle EE, Rodriguez C, Walker-Thurmond K, Thun MJ. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. *N Engl J Med*. 2003;348(17):1625–38.
 32. Shah NR, Braverman ER. Measuring adiposity in patients: the utility of body mass index (BMI), percent body fat, and leptin. *PLoS ONE*. 2012;7(4):e33308.
 33. Chiquita S, Rodrigues-Neves AC, Baptista FI, Carecho R, Moreira PI, Castelo-Branco M, et al. The retina as a window or mirror of the brain changes detected in Alzheimer's disease: critical aspects to unravel. *Mol Neurobiol*. 2019;56(8):5416–35.
 34. Sadun AA, Borchert M, DeVita E, Hinton DR, Bassi CJ. Assessment of visual impairment in patients with Alzheimer's disease. *Am J Ophthalmol*. 1987;104(2):113–20.
 35. Hart NJ, Koronyo Y, Black KL, Koronyo-Hamaoui M. Ocular indicators of Alzheimer's: exploring disease in the retina. *Acta Neuropathol*. 2016;132(6):767–87.
 36. Jiang H, Wei Y, Shi Y, Wright CB, Sun X, Gregori G, et al. Altered macular microvasculature in mild cognitive impairment and Alzheimer disease. *J Neuroophthalmol*. 2018;38(3):292–8.
 37. Harju M, Tuominen S, Summanen P, Viitanen M, Pöyhönen M, Nikoskelainen E, et al. Scanning laser Doppler flowmetry shows reduced retinal capillary blood flow in CADASIL. *Stroke*. 2004;35(11):2449–52.
 38. Lim G, Lim ZW, Xu D, Ting DSW, Wong TY, Lee ML, et al. Feature isolation for hypothesis testing in retinal imaging: an ischemic stroke prediction case study. *Proc AAAI Conf Artif Intell*. 2019;33(01):9510–5.
 39. Pachade S, Coronado I, Abdelkhalq R, Yan J, Salazar-Marioni S, Jagolino A, Green C, Bahrainian M, Channa R, Sheth SA, Giancardo L. Detection of stroke with retinal microvascular density and self-supervised learning using OCT-A and fundus imaging. *J Clin Med*. 2022;11(24):7408.
 40. Li H, Gao M, Song H, Wu X, Li G, Cui Y, Li Y, Xie Z, Ren Q, Zhang H. Predicting ischemic stroke risk from atrial fibrillation based on multi-spectral fundus images using deep learning. *Front Cardiovasc Med*. 2023;1(10):1185890.
 41. Cheung CY, Ran AR, Wang S, Chan VTT, Sham K, Hilal S, Venkatasubramanian N, Cheng CY, Sabanayagam C, Tham YC, Schmetterer L, McKay GJ, Williams MA, Wong A, Au LWC, Lu Z, Yam JC, Tham CC, Chen JJ, Dumitrascu OM, Heng PA, Kwok TCY, Mok VCT, Milea D, Chen CL, Wong TY. A deep learning model for detection of Alzheimer's disease based on retinal photographs: a retrospective, multicentre case-control study. *Lancet Digit Health*. 2022;4(11):e806–15.
 42. Appaji A, Harish V, Korann V, Devi P, Jacob A, Padmanabha A, Kumar V, Varambally S, Venkatasubramanian G, Rao SV, Suma HN, Webers CAB, Berendschot TTJM, Rao NP. Deep learning model using retinal vascular images for classifying schizophrenia. *Schizophr Res*. 2022;241:238–43.
 43. Ahn S, Shin J, Song SJ, Yoon WT, Sagong M, Jeong A, Kim JH, Yu HG. Neurologic dysfunction assessment in Parkinson disease based on fundus photographs using deep learning. *JAMA Ophthalmol*. 2023;141(3):234–40.
 44. Hong J, Yoon S, Shim KW, Park YR. Screening of Moyamoya Disease from retinal photographs: development and validation of deep learning algorithms. *Stroke*. 2024 Jan 23.
 45. Dai G, He W, Xu L, Pazo EE, Lin T, Liu S, et al. Exploring the effect of hypertension on retinal microvasculature using deep learning on East Asian population. *PLoS ONE*. 2020;15(3):e0230111.
 46. Zhang L, Yuan M, An Z, Zhao X, Wu H, Li H, et al. Prediction of hypertension, hyperglycemia and dyslipidemia from retinal fundus photographs via deep

- learning: a cross-sectional study of chronic diseases in central China. *PLoS ONE*. 2020;15(5): e0233166.
47. Mitani A, Huang A, Venugopalan S, Corrado GS, Peng L, Webster DR, et al. Detection of anaemia from retinal fundus images via deep learning. *Nat Biomed Eng*. 2020;4(1):18–27.
 48. Boris Babenko AM, Traynis I, Kitade N, Singh P, Maa A, Cuadros J, Corrado GS, Peng L, Webster DR, Varadarajan A, Hammel N, Liu Y. Detecting hidden signs of diabetes in external eye photographs. *arXiv*. 2020.
 49. Benson J, Estrada T, Burge M, Soliz P, editors. Diabetic peripheral neuropathy risk assessment using digital fundus photographs and machine learning. In: 2020 42nd annual international conference of the IEEE engineering in medicine & biology society (EMBC), 20–24 July; 2020.
 50. Wong TY, Xu D, Ting D, Nusinovici S, Cheung C, Shyong TE, Cheng C-Y, Lee ML, Hsu W, Sabanayagam C. Artificial intelligence deep learning system for predicting chronic kidney disease from retinal images. *Invest Ophthalmol Vis Sci*. 2019;60(9):1468.
 51. Sabanayagam C, Xu D, Ting DSW, Nusinovici S, Banu R, Hamzah H, et al. A deep learning algorithm to detect chronic kidney disease from retinal photographs in community-based populations. *Lancet Digit Health*. 2020;2(6):e295–302.
 52. Kang EY HY, Li C, Huang Y, Kuo C, Kang J, Chen K, Lai C, Wu W, Hwang Y. A deep learning model for detecting early renal function impairment using retinal fundus images: model development and validation study. *JMIR Med Inf*. 2020.
 53. Zhang K, Liu X, Xu J, Yuan J, Cai W, Chen T, Wang K, Gao Y, Nie S, Xu X, Qin X, Su Y, Xu W, Olvera A, Xue K, Li Z, Zhang M, Zeng X, Zhang CL, Li O, Zhang EE, Zhu J, Xu Y, Kermany D, Zhou K, Pan Y, Li S, Lai IF, Chi Y, Wang C, Pei M, Zang G, Zhang Q, Lau J, Lam D, Zou X, Wumaier A, Wang J, Shen Y, Hou FF, Zhang P, Xu T, Zhou Y, Wang G. Deep-learning models for the detection and incidence prediction of chronic kidney disease and type 2 diabetes from retinal fundus images. *Nat Biomed Eng*. 2021;5(6):533–45.
 54. Betzler BK, Chee EYL, He F, Lim CC, Ho J, Hamzah H, Tan NC, Liew G, McKay GJ, Hogg RE, Young IS, Cheng CY, Lim SC, Lee AY, Wong TY, Lee ML, Hsu W, Tan GSW, Sabanayagam C. Deep learning algorithms to detect diabetic kidney disease from retinal photographs in multiethnic populations with diabetes. *J Am Med Inform Assoc*. 2023;30(12):1904–14.
 55. Walsh JB. Hypertensive retinopathy. Description, classification, and prognosis. *Ophthalmology*. 1982;89(10):1127–31.
 56. Cheung CY, Xu D, Cheng CY, Sabanayagam C, Tham YC, Yu M, et al. A deep-learning system for the assessment of cardiovascular disease risk via the measurement of retinal-vessel calibre. *Nat Biomed Eng*. 2020.
 57. Son J, Shin JY, Chun EJ, Jung K-H, Park KH, Park SJ. Predicting high coronary artery calcium score from retinal fundus images with deep learning algorithms. *Transl Vis Sci Technol*. 2020;9(2):28.
 58. Chang J, Ko A, Park SM, Choi S, Kim K, Kim SM, et al. Association of cardiovascular mortality and deep learning-fundusoscopic atherosclerosis score derived from retinal fundus images. *Am J Ophthalmol*. 2020;217:121–30.
 59. Detrano R, Guerci AD, Carr JJ, Bild DE, Burke G, Folsom AR, et al. Coronary calcium as a predictor of coronary events in four racial or ethnic groups. *N Engl J Med*. 2008;358(13):1336–45.
 60. Ching T, Zhu X, Garmire LX. Cox-nnet: An artificial neural network method for prognosis prediction of high-throughput omics data. *PLoS Comput Biol*. 2018;14(4): e1006076.
 61. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med Res Methodol*. 2018;18(1):24.
 62. Gensheimer MF, Narasimhan B. A scalable discrete-time survival model for neural networks. *PeerJ*. 2019;7: e6257.
 63. Zhu Z, Chen Y, Wang W, Wang Y, Hu W, Shang X, Liao H, Shi D, Huang Y, Ha J, Tan Z, Kiburg KV, Zhang X, Tang S, Yu H, Yang X, He M. Association of retinal age gap with arterial stiffness and incident cardiovascular disease. *Stroke*. 2022;53(11):3320–8.
 64. Cheung CY, Wong WLE, Hilal S, Kan CN, Gyanwali B, Tham YC, Schmetterer L, Xu D, Lee ML, Hsu W, Venketasubramanian N, Tan BY, Wong TY, Chen CPLH. Deep-learning retinal vessel calibre measurements and risk of cognitive decline and dementia. *Brain Commun*. 2022;4(4):fcac212.
 65. Zhu Z, Hu W, Chen R, Xiong R, Wang W, Shang X, Chen Y, Kiburg K, Shi D, He S, Huang Y, Zhang X, Tang S, Zeng J, Yu H, Yang X, He M. Retinal age gap as a predictive biomarker of stroke risk. *BMC Med*. 2022;20(1):466.
 66. Hu W, Wang W, Wang Y, Chen Y, Shang X, Liao H, Huang Y, Bulloch G, Zhang S, Kiburg K, Zhang X, Tang S, Yu H, Yang X, He M, Zhu Z. Retinal age gap as a predictive biomarker of future risk of Parkinson's disease. *Age Ageing*. 51(3):afac062.
 67. Zhang S, Chen R, Wang Y, Hu W, Kiburg KV, Zhang J, Yang X, Yu H, He M, Wang W, Zhu Z. Association of retinal age gap and risk of kidney failure: a UK biobank study. *Am J Kidney Dis*. 2023;81(5):537–544.e1.
 68. Wu G, Zhang X, Borchert GA, Zheng C, Liang Y, Wang Y, Du Z, Huang Y, Shang X, Yang X, Hu Y, Yu H, Zhu Z. Association of retinal age gap with chronic kidney disease and subsequent cardiovascular disease sequelae: a cross-sectional

- and longitudinal study from the UK Biobank. *Clin Kidney J.* 2024;17(7):sfaf088.
69. Zhuoting, Zhu Yueye, Wang Ziyi, Qi Wenyi, Hu Xiayin, Zhang Siegfried K., Wagner Yujie, Wang An Ran, Ran Joshua, Ong Ethan, Waisberg Mouayad, Masalkhi Alex, Suh Yih Chung, Tham Carol Y., Cheung Xiaohong, Yang Honghua, Yu Zongyuan, Ge Wei, Wang Bin, Sheng Yun, Liu Andrew G., Lee Alastair K., Denniston Peter van, Wijngaarden Pearse A., Keane Ching-Yu, Cheng Mingguang, He Tien Yin, Wong. Oculomics: Current concepts and evidence. *Prog. Retin. Eye Res.* 2025;106:101350. <https://doi.org/10.1016/j.preteyeres.2025.101350>.
 70. Yukun, Zhou Mark A., Chia Siegfried K., Wagner Murat S., Ayhan Dominic J., Williamson Robbert R., Struyven Timing, Liu Moucheng, Xu Mateo G., Lozano Peter, Woodward-Court Yuka, Kihara Naomi, Allen John E. J., Gallacher Thomas, Littlejohns Tariq, Aslam Paul, Bishop Graeme, Black Panagiotis, Sergouniotis Denize, Atan Andrew D., Dick Cathy, Williams Sarah, Barman Jenny H., Barrett Sarah, Mackie Tasanee, Braithwaite Roxana O., Carare Sarah, Ennis Jane, Gibson Andrew J., Lotery Jay, Self Usha, Chakravarthy Ruth E., Hogg Euan, Paterson Jayne, Woodside Tunde, Peto Gareth, Mckay Bernadette, Mcguinness Paul J., Foster Konstantinos, Balaskas Anthony P., Khawaja Nikolas, Pontikos Jugnoo S., Rahi Gerassimos, Lascaratos Praveen J., Patel Michelle, Chan Sharon Y. L., Chua Alexander, Day Parul, Desai Cathy, Egan Marcus, Fruttiger David F., Garway-Heath Alison, Hardcastle Sir Peng T., Khaw Tony, Moore Sobha, Sivaprasad Nicholas, Strouthidis Dhanes, Thomas Adnan, Tufail Ananth C., Viswanathan Bal, Dhillon Tom, Macgillivray Cathie, Sudlow Veronique, Vitart Alexander, Doney Emanuele, Trucco Jeremy A., Guggenheim James E., Morgan Chris J., Hammond Katie, Williams Pirro, Hysi Simon P., Harding Yalin, Zheng Robert, Luben Phil, Luthert Zihan, Sun Martin, McKibbin Eoin, O'Sullivan Richard, Oram Mike, Weedon Chris G., Owen Alicja R., Rudnicka Naveed, Sattar David, Steel Irene, Stratton Robyn, Tapp Max M., Yates Axel, Petzold Savita, Madhusudhan Andre, Altmann Aaron Y., Lee Eric J., Topol Alastair K., Denniston Daniel C., Alexander Pearse A., Keane. A foundation model for generalizable disease detection from retinal images. *Nature* 2023;622(7981):156–163. <https://doi.org/10.1038/s41586-023-06555-x>
 71. Xiaolan, Chen Weiye, Zhang Pusheng, Xu Ziwei, Zhao Yingfeng, Zheng Danli, Shi Mingguang, He. FFA-GPT: an automated pipeline for fundus fluorescein angiography interpretation and question-answer. *npj Digit. Med.* 2024;7(1). <https://doi.org/10.1038/s41746-024-01101-z>.
 72. Bjorn Kaijun, Betzler Tyler Hyungtaek, Rim Charumathi, Sabanayagam Ching-Yu, Cheng. Artificial Intelligence in Predicting Systemic Parameters and Diseases From Ophthalmic Imaging *Frontiers in Digital Health* 2022;4. <https://doi.org/10.3389/fdgh.2022.889445>
 73. Louis, Arnould Fabrice, Meriaudeau Charles, Guenancia Clément, Germanese Cécile, Delcourt Ryo, Kawasaki Carol Y., Cheung Catherine, Creuzot-Garcher Andrzej, Grzybowski. Using Artificial Intelligence to Analyse the Retinal Vascular Network: The Future of Cardiovascular Risk Assessment Based on Oculomics? *A Narrative Review Ophthalmology and Therapy* 2023;12(2):657–674. <https://doi.org/10.1007/s40123-022-00641-5>
 74. Huimin, Li Jing, Cao Andrzej, Grzybowski Kai, Jin Lixia, Lou Juan, Ye. Diagnosing Systemic Disorders with AI Algorithms Based on Ocular Images *Healthcare* 2023;11(12):1739. <https://doi.org/10.3390/healthcare11121739>
 75. Andrzej, Grzybowski Kai, Jin Jingxin, Zhou Xiangji, Pan Meizhu, Wang Juan, Ye Tien Y., Wong. Retina Fundus Photograph-Based Artificial Intelligence Algorithms in Medicine: A Systematic Review *Ophthalmology and Therapy* 2024;13(8):2125–2149. <https://doi.org/10.1007/s40123-024-00981-4>

Overview of Intraocular Lens Power Calculation Formulas Based on Artificial Intelligence

22

Wiktor Stopyra  and Andrzej Grzybowski 

Abstract

Accurate intraocular lens power calculation is a very important aspect of phacoemulsification because patients' expectations for perfect vision after cataract surgery are still increasing. However, the European Registry of Quality Outcomes reported that the percentage of patients with the prediction error within $\pm 0.5D$ after cataract surgery was 73.7%. Therefore with the development of cataract surgery so many new intraocular lens power calculation formulas have been launched in recent years. They are divided according different classifications but the most practical one is based on logical approach. The latest formulas use artificial intelligence to varying degrees. The most common are methods based on optics and incorporate artificial intelligence components to refine their prediction. However,

some formulas use only artificial intelligence. The rarest are universal artificial intelligence algorithms, which can improve the accuracy of pre-existing formulas.

Keywords

Phacoemulsification · Intraocular lens power calculation formula · Artificial intelligence based formula · The Kane formula · The Hill-RBF formula · The PEARL-DGS formula · The Hoffer QST formula

Introduction

Cataract surgery is the most commonly performed ophthalmological surgical procedure [1]. Initially, cataract surgery was only an extraction technique but now it is also a refractive treatment [2]. Therefore, patients' expectations for perfect vision after phacoemulsification are still increasing. With the technology development, the accuracy of measurements of biometric parameters required for intraocular lens power calculation has improved significantly. However, recent studies reported that only 70–80% of eyes achieved a postoperative refraction within $\pm 0.50 D$ of the predicted value [3]. This forces cataract surgeons to look for new, more precise intraocular lens (IOL) power calculation formulas.

W. Stopyra (✉)
MW-Med Eye Centre, Krakow, Poland
e-mail: wiktorstopyra@gmail.com

W. Stopyra
University of Applied Sciences, Nowy Targ, Poland

A. Grzybowski
Institute for Research in Ophthalmology,
Foundation for Ophthalmology Development,
Poznan, Poland

History of IOL Implantation

One of the first reports about intraocular lenses comes from Giacomo Casanova's diaries from 1776. There he describes a meeting in Warsaw with a traveling ophthalmologist Felice Tadini, who demonstrated lenses made of beautiful crystal, intended to be implanted inside the eye. The first attempt to implant an artificial glass lens was made by Dresden ophthalmologist Johan Virgilius Casamaata in 1795, but the procedure ended in failure due to the imperfections of the lens. The too heavy implant fell to the fundus [4].

Sir Nicolas Harold Lloyd Ridley was the first person to successfully implant an IOL on 29 November 1949, at St Thomas' Hospital in London. The lens was manufactured by the Rayner company of Brighton, East Sussex, England from Perspex (PMMA) made by Imperial Chemical Industries. Ridley had observed that Royal Air Force pilots who sustained eye injuries during World War II involving PMMA windshield material did not show any rejection or foreign body reaction, and deduced that the transparent material was inert and useful for implantation in the eye. The first patient obtained high postoperative myopia [5].

Intraocular lenses implantation has become more and more common with the development of phacoemulsification in 1967 by Charles Kelman, an American eye surgeon, jazz musician and also Broadway producer [6]. Then, accuracy of intraocular lens power calculation began to be carefully considered. Svyatoslav Nikolayevitsch Fyodorov, a Russian ophthalmologist, launched the first widely known intraocular lens power calculation formula in 1967 [7].

Intraocular Lens Power Calculation Formulas Division

Since Fyodorov introduced his formula many others have been developed over the years. This was due to the desire to obtain the most precise

postoperative refractive outcomes. However, according to European Registry of Quality Outcomes the percentage of the prediction error within $\pm 0.5D$ after cataract surgery is 73.7% [8]. The accuracy of refractive power of the IOL implant depends not only on accuracy of the preoperative biometric data such as axial length (AL), keratometry (K), anterior chamber depth (ACD) where inaccuracies in measurement can contribute to 36%, 22%, 42% of errors, respectively but most of all on effective lens position (ELP) estimation resulting from the selection of appropriate formula [9, 10]. The basic biometric parameters necessary for correct IOL power calculation are marked on optical model of pseudophakic eye and shown in Fig. 22.1.

Traditionally intraocular lens power calculation formulas were classified by generations, as shown in Table 22.1 [11].

However, categorizing formulas based on logical approach is much more useful (Table 22.2) [12].

In addition, there are also formulas that are used in special cases, e.g. phakic lens formulas (Olsen-Feingold, Holladay and Linz-Homburg-Castrop) [13] or post-refractive formulas listed in Table 22.3 [14, 15].

Formulas Based on AI

The calculation of the IOL power using the optical formula is based on a simple scheme shown in Fig. 22.2.

However, this scheme is much more extensive when using the artificial intelligence (AI) formula (Fig. 22.3).

There are 2 possible AI-based formulas approaches. The first, when AI formula can be designed to either predict the postoperative spherical equivalent (SE) for a given IOL power and set of biometric parameters and the second, which can predict the IOL power implanted given the postoperative SE and the biometric parameters.

Intraocular lens power calculation formulas based on AI can be divided into 2 groups

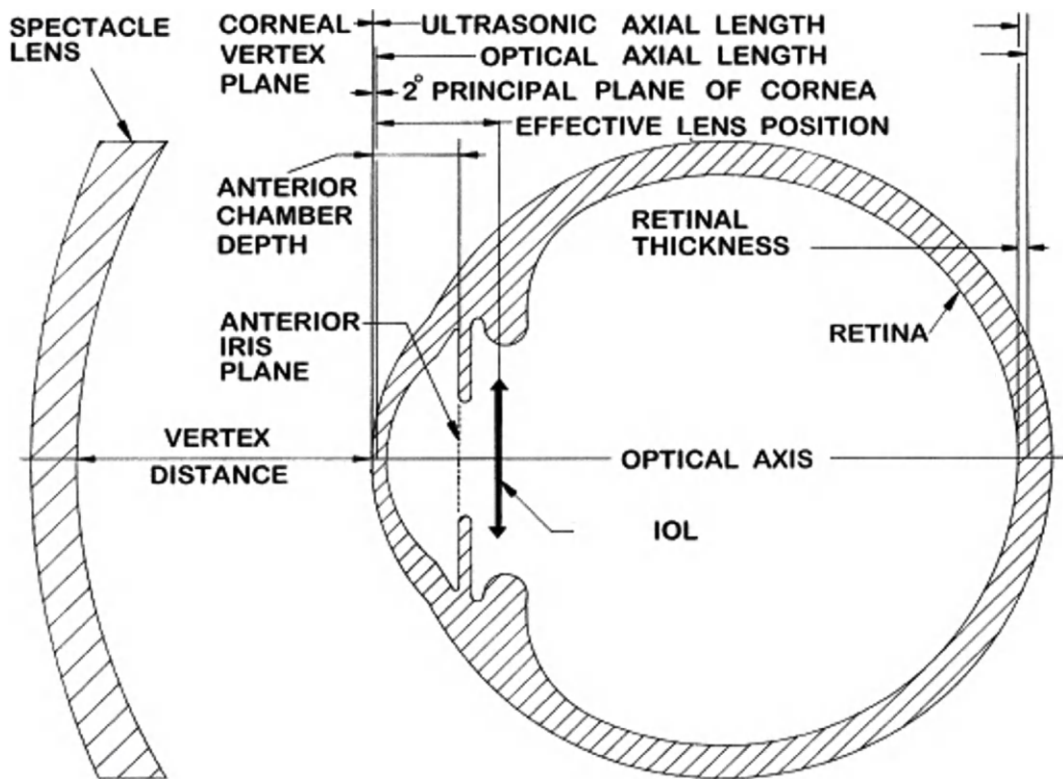


Fig. 22.1 Optical model of pseudophakic eye

Table 22.1 Generations of intraocular lens power calculation formulas

Generation				
1st	2nd	3rd	4th	Newer formulas
Fyodorov Binkhorst SRK Colenbrander Hoffer van der Heijde	SRK II Binkhorst II	Hoffer Q Holladay 1 SRK/T	Haigis Olsen Holladay 2 Barrett Universal II	Hill-RBF 3.0 Kane PEARL-DGS EVO Ladas SF AI K6 Hoffer QST

i.e. hybrids (FullMonte, Ladas SF AI, PEARL-DGS, Kane, Hoffer QST, Zhu-Lu and Zeiss AI) and pure AI methods (Hill-RBF, Karmona, Nallasamy). In addition, there are universal algorithms using AI to improve the accuracy

of 3rd or 4th generation formulas e.g. Sramka approach, the Precision Ladas Universal Superalgorithm (PLUS) method. However, these are not separate formulas. Table 22.4 shows the use of AI in IOL power calculation.

Table 22.2 Intraocular lens power calculation formulas

Data-driven		Optical approach			Hybrid
Simple regression	Artificial intelligence	Simple thin lens vergence	Interactive thick lens vergence	Ray tracing	Combination
SRK SRK II	Hill-RBF 2.0 Karmona Hill-RBF 3.0 Nallasamy	Binkhorst Holladay 1 Hoffer Q SRK/T Holladay 2 Haigis T2 Castrop K6 Panacea	EVO Naeser 2	Olsen Okulix O Barrett Universal	Kane FullMonte Ladas SF AI PEARL-DGS Hoffer QST Zhu-Lu

Table 22.3 Post refractive intraocular lens power calculation formulas

Only historical data	Combination of historical and current corneal data	No historical data
Feinz-Mannis Corneal bypass	Barrett True-K Masket	R-Factor Wang-Koch-Maloney Shammas Haigis-L Potvin-Hill Pentacam Barrett True-K No History OCT-based PhacoOptic Advanced lens measurement approach

preoperative AL, both K values, ACD and LT) from 200 consecutive cases of one surgeon’s results with one IOL model. The neural network was trained to produce the postoperative refractive error, and the Holladay surgeon factor was continuously refined using the same results. After the network was successfully trained with the clinical data, it was utilized to compute IOL power in a double-masked study [17]. Fullmonte was a hybrid formula. Clarke and Kalpener proposed a combined method between SRK/T and an AI algorithm called Bayesian Additive Regression Trees, optimized with Markov Chain Monte Carlo simulator to produce its refractive predictions [18]. Unpublished data have shown a reduction in the standard deviation compared with the SRK/T formula. Unfortunately, the website of FullMonte is no longer functional.

AI Hybrid Formulas

The FullMonte Method

The history of FullMonte is interesting and inspiring, as Gerald Clarke was the pioneer proposing with Jeanne Burmeister in 1997 the first neural network for biometric computations [16]. They developed and trained a neural network to predict IOL powers utilizing a personalized Holladay program and clinical data (included

The Ladas Super Formula AI

In 2015, John G. Ladas, Albert Jun, Aazim Siddiqui, and Uday Devgan proposed the concept of an IOL ‘super formula’ which combined the benefits of existing formulas [19]. They considered the third-generation formulas (Hoffer



Fig. 22.2 The scheme of IOL power calculation using the optical formula

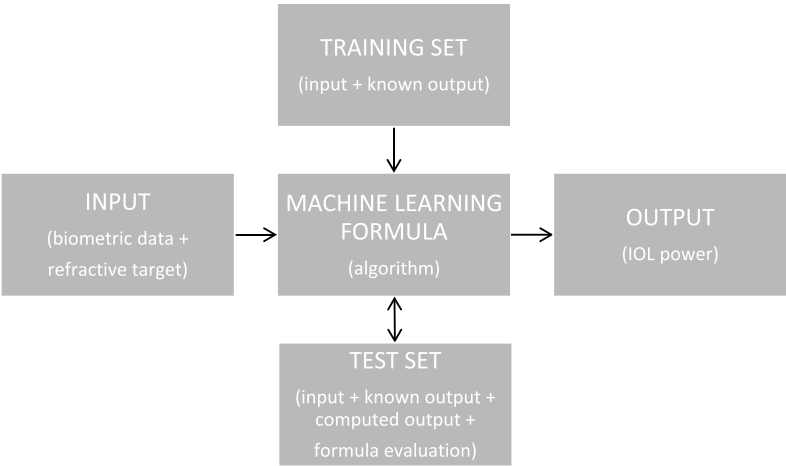


Fig. 22.3 The scheme of IOL power calculation using the AI formula

Table 22.4 The use of AI in IOL power calculation

AI hybrid formulas	Pure AI formulas	Universal AI algorithms
FullMonte Ladas SF AI Kane PEARL_DGS Hoffer QST Zhu-Lu Zeiss AI	Hill-RBF 2.0 Karmona Hill-RBF 3.0 Nallasamy	Sramka approach PLUS method

Q, Holladay 1, Holladay 1 with Koch adjustment, SRK/T) and a fourth generation formula i.e. Haigis [19]. Although past formulas were thought of as two-dimensional algebraic equations, Ladas et al. found a novel methodology of depicting these formulas as mathematical equations with the potential to be graphed on the x, y and z axes and rendered them in three dimensions. In this way they provided a framework to analyze these formulas in three dimensions and observe areas of nuance where they were similar or different. Based on this and peer-reviewed literature, the identified best portions of each of the mentioned IOL formulas were chosen and an IOL ‘super surface’ was developed [20]. Retrospective analyses have shown that while the SRK/T tends to be better for longer eyes [11, 21], older studies have shown that very short eyes yield better outcomes with the Hoffer Q [22]. Some eyes require more nuance

and optimization. For example, shorter eyes are difficult to calculate as the smallest changes in ELP can alter the lens power calculations significantly. From this super surface, the ‘super formula’ was derived [19, 20]. Then they used software (Python 3.7 with scikit-learn package) to refine a baseline formula. Further, steps to prevent overfitting of a model are done by performing a fivefold cross validation within the training set. This was all done after randomly separating the data into ten equal parts and subsequently using nine of the ten to train the algorithm and test on the remaining tranche.

The Kane Formula

Jack X Kane, MD from Australia presented his IOL power calculation formula in September 2017 using about 30,000 cases [23]. The Kane formula was created utilizing several large data sets from selected high-volume surgeons. It uses a combination of theoretical optics, thin lens formulas and ‘big data’ techniques to make its predictions [24]. As a hybrid formula it is based on optics and incorporates both regression and AI components to further refine its prediction. The formula was developed using cloud-based computing. A focus of the formula was to reduce the errors seen at the extremes of the various ocular

dimensions where the current formulas display larger errors. A-constant, AL, K, ACD and sex are compulsory variables used in the Kane formula. Although adding LT and CCT improve predictions, they are optional variables. This allows owners of older biometers to use the formula. The formula has been developed to have an A-constant very similar to the SRK/T A-constant. If the surgeon has an optimized A-constant, then that is recommended for use [25].

The PEARL-DGS Formula

In 2017, Guillaume Debellemanni re, Damien Gatinel and Alain Saad launched a new formula, developed using optical biometry and machine learning modeling called PEARL-DGS (DGS named after the formula developers) [26, 27]. This is a thick lens formula based on the prediction of the theoretical internal lens position (TILP) which is the theoretical distance between the posterior corneal surface and the anterior IOL surface, back-calculated from postoperative data [28]. TILP is independent of both the lens principal plane positions and the corneal thickness. It is predicted using various machine learning algorithms comprising regular multiple regression, support vector regression, gradient boosted trees and neural networks. Additionally, Jupyter was utilized in developing of the PEARL-DGS formula. This is an open-source application allowing sharing of Python code in the form of a succession of cells, called Notebooks, facilitating research sharing and reproducibility. The refractive indices of the Atchinson model eye were used, except for the corneal index which was determined empirically during the formula development process. The axial length was adjusted by the Cooke modified AL formula (CMAL) [29]. Data inputs of the PEARL DGS formula include AL, K, ACD, LT, WTW, and CCT [26]. Output linearization, adjustable IOL constants (same as for third generation formulas) and no re-training for new IOL models are characteristic for PEARL-DGS formula [28].

The Hoffer QST Formula

In 2021 Kenneth Hoffer (USA), Giacomo Savini (Italy) and Leonardo Taroni (Italy) proposed the Hoffer QST formula as improving the third generation vergence-based Hoffer Q by using AI. Above all, they tried to increase accuracy in long eyes. First, they aimed to improve the prediction of the ELP. For this purpose, they selected a totally independent training sample of 537 eyes. They entered into Linear Machine Learning model the following variables: K, AL, ACD measured from corneal epithelium to lens, corneal radius and gender. The Python language with the Scikit-learn library was adopted and the best regressor of their cohort of data was the TheilSen model. As a second step, they developed a customized AL adjustment. Exploiting a reliable training dataset of 375 eyes longer than 25.0 mm they optimized the individual AL using Excel. They then generated a new machine learning algorithm based on above mentioned inputs to find the ideal AL modification. This time the best regressor was the Huber model [30]. The first results presented during 2021 Annual Meeting of American Society of Cataract and Refractive Surgery showed that the Hoffer QST formula improved the outcomes of Hoffer Q significantly [31].

The Zhu-Lu Formula

In June 2023, Yi Lu, Xiangjia Zhu & others from China published the Zhu-Lu formula based on 1828 highly myopic eyes (1462 eyes training set and 366 eyes test set) [32]. They constructed 2 sets of learning feature (set 1 incorporated results from Haigis and SRK/T formulas and set 2 contained results only from Haigis formula). Two supervised machine learning models (the eXtreme Gradient Boosting and support vector regression) were trained with each of 2 sets of learning features. Therefore they obtained 4 sub-models each of which could function independently for IOL power calculation in highly myopic eyes. To increase the model robustness,

they adopted the weighted average of the calculation results from four sub-models and generated an assembled prediction formula. Python 3.7 with the scikit-learn package was the software used in model construction. The formula needs following variables: AL, K, ACD, LT, WTW and sex [32].

The Zeiss AI Formula

Peter Kenny and Douglas Koch, Warren Hill, Li Wang published the Zeiss AI formula in July 2023 [33]. This is a new AI-based approach that first learns physical optics of the human eye and then optimizes itself on thousands of cases for each IOL model covering the entire population (short, medium, and long eyes) [33]. AL, ACD, LT age and sex are compulsory variables used in the Zeiss AI formula. Unfortunately, its online IOL calculator is available only in the United States [34].

Pure AI Formulas

The Hill-RBF Formula

Hill-RBF was launched in 2016 by Warren E. Hill, MD and was based on radial basis function (RBF), which is similar to a neural network. The Hill-RBF formula is a pure data-driven IOL calculation approach and therefore it is free of the limitations and advantages of a lens-position estimation [8]. It employs pattern recognition developed by Matlab and a sophisticated form of data interpolation. Starting with a large number of cases where the biometry and the outcomes are known, RBF is capable of finding distinct patterns in the apparently random cloud of data-points. It has been optimized for use with biometry data from the LENSTAR LS 900 (Haag-Streit, Köniz, Switzerland) optical biometer in combination with the Alcon SN60WF (Alcon Laboratories, Fortworth, TX, USA) biconvex IOL for powers from +6.00 D to +30.00 D. It has separately been optimized for IOL powers up to +35.00 D based on a

similar biconvex IOL design. For IOL powers from +5.00 D to −5.00 D, it performs best with this combination of biometry device and the Alcon MA60MA (Alcon Laboratories, Fortworth, TX, USA) extended range IOL. The Hill-RBF on-line calculator may also be used with data from other optical biometers, which provide clinically equivalent biometry data as compared to the LENSTAR LS 900. It may also be used with other biconvex IOL models in the power range of +6.00 D to +35.00 D and other meniscus design IOL models in the power range of +5.00 D to −5.00 D. It uses a large dataset—more than 12,000 eyes in version 2.0 and over 30,000 eyes in version 3.0 which has been available since September 2020. However, Hill-RBF will not provide a refractive prediction if the preoperative parameters are out of bounds, as predictions are likely to be inaccurate. In addition, early versions did not allow input of a target refraction other than plano. Data inputs of the Hill-RBF 3.0 formula include AL, K, ACD, LT, WTW, central corneal thickness (CCT) and sex [35].

The Karmona Formula

This is one of the latest power calculation formulas has been designed and programmed in Shiny-RStudio Version 1.1.423 (R Foundation, Boston, USA) by David Carmona González (Madrid, Spain, 2018–2021). It is a result of pure artificial intelligence without Gaussian theoretical components. It is based on two combined models of machine learning called regularized Bayesian neural network and cubist decision tree. These were selected from eleven nonlinear regression models (k-Nearest Neighbors, linear discriminant analysis, Support Vector Machines with RBF or with linear kernel, neural networks testing several hidden layers, random forest, decision trees, Least Absolute Shrinkage and Selection Operator regression, generalized linear model, multivariate adaptive regression spline, and stochastic gradient boosting) as the best efficient. Then they were tuned, thereby improving their

metrics. Finally, an ensembled model was generated via machine learning method that combined prediction from separate strategies [36]. It needs following data: AL, ACD, K, WTW, and corneal anterior surface central radius. Optional variables for Karmona formula are LT, posterior corneal central radius, anterior segment depth (the sum of ACD and LT), ratio of the back to front corneal surface central radius and gender. It is optimized for anterior keratometry with IOL Master® 700 (Carl Zeiss Meditec, Jena, Germany) and posterior keratometry with Pentacam® (Oculus Optikgeräte GmbH, Wetzlar, Germany), however it can apparently be used with measurements from any other device. It uses the A constant as the IOL constant [37]. As a new method, Karmona formula has been developed with a database of only 386 eyes, with axial lengths that ranged mainly between 22.0 and 25.0 mm. As any other AI based formula, it is constantly learning, so its predictive capacity is expected to continuously improve. Using unique data input such as anterior segment depth and ratio of the back to front corneal surface central radius as a separate parameter, Karmona formula has potentially been designed to increase the accuracy in extreme and atypical eyes.

The Nallasamy Formula

In 2022 TingyangLi, Joshua Stein and Nambi Nallasamy published formula based on dataset of 5016 eyes [38]. The preoperative biometry records were obtained from Lenstar LS 900 optical biometers (Haag-Streit USA, EyeSuite software V.i9.1.0.0) at University of Michigan's Kellogg Eye Center. Nallasamy et al. utilized ensemble machine-learning model which was consisted of two layers. In the first layer a group of level-1 learners was trained based on the preoperative patient data and IOL power as input data and the postoperative refraction as the target value. The second layer consists of the meta-model which uses the output of the

level-1 learners as the input features. Therefore, the number of input qualities for the level-2 model equals the number of level-1 models. The output from the level-2 meta-model is the final predicted refraction. As input data the formula uses 5 compulsory variables such AL, K, ACD, WTW, LT and 1 optional i.e. CCT [38].

Universal AI Algorithms

Sramka Approach

Martin Sramka, Martin Slovak, Jana Tuckova and Pavel Stodulka published in 2019 their method of improvement SRK/T formula utilizing AI. They developed two structures, the first based on the support vector machine regression model (SVM-RM) and the second using multilayer neural network ensemble model (MLNN-EM). Their SVM-RM was designed and trained in MATLAB 2017a (MathWorks, Natick, MA, USA) utilizing the fitsvm method. The optimal hyperparameters of the model, which are essential, were found through the optimize hyperparameters (MathWorks 2017a) method that searched for optimal kernel function, epsilon, kernel scale, polynomial order and box constraint. The selection set was used in model training by the sequential minimal optimization algorithm with 30% of randomly selected data used for holdout validation. Their MLNN was designed and trained in MATLAB 2017a by fitnet (MathWorks 2017b). It had one hidden layer with five neurons and one output layer with one neuron with linear transfer function. The Hyperbolic Tangent Sigmoid transfer function was used as a transfer function in the hidden layer. The ensemble median factor was set to 10 i.e. 10 MLNN were trained by the selection set in order to produce a desired output taken as a median of all outputs. The selection set was randomly divided into three groups, for network training, validation, and testing by a 70:15:15 ratio. MLNN training was stopped when the network performance in the validation

group failed to improve or remained the same for 20 epochs. The optimal number of neurons in the MLNN hidden layer was found iteratively, testing all available combinations of neurons from 1 up to 350 neurons in a hidden layer. Both SVM and MLNN yielded similar results much better than original SRK/T formula [39].

The Precision Ladas Universal Superalgorithm (PLUS) Method

John Ladas and the team at Advanced Euclidean Solutions proposed method which allowed application of advanced AI to any 3rd or 4th generation formula [40]. The most state-of-the-art AI algorithms (e.g. ensemble learning with gradient boosting) are applied to their method to focus on the validation and quality control of postoperative refractive data and improve workflow efficiencies related to acquiring data and updating the model. The most important advantages of the method are: the model's ability to be refined and retrained indefinitely; the ability to add input parameters (e.g., age) to the model as they become available without having to create a model from scratch; and the ability to apply the methodology to specific patient populations (e.g., those with a history of corneal refractive surgery).

Availability of AI-Based Formulas

All IOL power calculation formulas except FullMonte are available on-line. In addition, from 2022 European Society of Cataract and Refractive Surgeons presents calculators of 7 popular IOL formulas including 4 AI-based (Kane, Hill-RBF, PEARL-DGS and Hoffer QST). This is beneficial both for formulas developers and eye surgeons and their patients. On the one hand, it gives the opportunity of formulas widespread use, on the other hand thanks to outcomes of successive patients the formulas are becoming more and more precise. Detailed web pages with calculators of AI-based formulas are listed in Table 22.5.

Table 22.5 AI based formulas on-line calculators

Formula	On-line calculator
Ladas super formula AI	www.iolcalc.com
Hill-RBF 3.0	www.rbfcaculator.com
Kane	www.iolformula.com
PEARL-DGS	www.iolsolver.com
Hoffer QST	www.hofferqst.com
Karmona	www.karmona-iol.com
Nallasamy	www.lenscalc.com
Zhu-Lu	www.hm-zlf.com
Zeiss AI	www.zcalc.meditec.zeiss.com
Hill-RBF/HofferQST/ Kane/PEARL-DGS	www.iolcalculator.esrcs.org

Variables Utilized in IOL Formulas Based on AI

The aforementioned formulas differ in logical approach (pure AI-based and hybrids), their calculation principles and input data. Among variables utilized in AI-based formulas AL and ACD are always used. Some formulas (Hill-RBF 3.0, Kane, Karmona and Nallasamy) to improve their accuracy utilize optional data in addition to mandatory data Table 22.6.

Accuracy of IOL Power Calculation AI-Based Formulas

There are currently no publications that equate the accuracy of all AI-based formulas. On the other hand, many articles have been published that compare the correctness of various IOL power calculation formulas including AI-based ones.

Many different tools could be employed to assess the accuracy of the chosen formula. Most studies in this area are based on observation of mean absolute error (MAE) and percentage of eyes within ± 0.5 D, less often of median absolute error (MedAE). Hoffer et al. recommended to utilize MedAE as a primary outcome in this type of studies due to the not-normal distribution of absolute refractive prediction error [41]. In turn, Cooke et al. proposed mean rank score

Table 22.6 Variables utilized in formulas based on artificial intelligence

Formula	Logical approach classification	Variables utilized									
		AL	K	KmA	KmP	ACD	LT	CCT	WTW	Age	G
FullMonte	Hybrid	C	C			C	C				
LSF AI	Hybrid	C	C			C					
Hill-RBF	1.0 AI	C	C			C					
		C	C			C					
		C	C			C	O	O	O		O
PEARL-DGS	Hybrid	C	C			C	C	C	C		
Kane	Hybrid	C	C			C	O	O			C
Karmona	AI	C		C	O	C	O		C		C
Hoffer QST	Hybrid	C	C			C					C
Nallasamy	AI	C	C			C	C	O	C		
Zhu-Lu	Hybrid	C	C			C	C		C		C
Zeiss AI		C				C	C			C	C

AL axial length; K keratometry; KmA keratometry-mean anterior; KmP keratometry-mean posterior; ACD anterior chamber depth; LT lens thickness; CCT central corneal thickness; WTW white to white; G gender; RBF radial basis function; PEARL prediction enhanced by artificial intelligence and output linearization; DGS Debellemanière, Gatinel, Saad; Hoffer QST Hoffer Q, Savini, Taroni; C compulsory; O optional, AI artificial intelligence

for subgroup analysis (e.g. in terms of AL) [42], and Haigis developed his own rank [41]. Numerical error was defined as the difference between the real postoperative refractive outcome expressed as spherical equivalent (equal to the sum of spherical power and half of cylindrical power) and the refraction predicted by each formula. A positive value indicated a hyperopic error and a negative value referred to myopic error while absolute value means absolute error. Detailed outcomes of last studies (MAE, percentage of eyes within ± 0.5 D and ± 1.0 D) are summarized in Table 22.7.

Overall, there is still no agreement among cataract surgeons regarding the choice of the ideal IOL power calculation formula. However, out of AI based formulas, Kane seems to be the most accurate. Many recent studies have shown that the Kane formula achieved the lowest MAE, and/or the highest percentage of eyes within ± 0.5 D [23, 24, 43, 45–48]. Hill-RBF, which is often regarded highly among IOL

formulas, obtained mostly second outcome [17, 23, 28, 43, 49]. Only few studies have shown the superiority of PEARL-DGS [50, 51], however, recently its online calculator was updated with a new version of the formula so its results may be better in the future.

Conclusions

The machine learning method can be used as a stand-alone formula as well as incorporated into various existing IOL power formulas to refine their accuracy [52]. Overall, IOL power calculation formulas based on AI are promising and have the potential to positively affect the precision of postoperative refractions after cataract surgery. It can be expected that newer and more precise AI-based methods will be developed. And certainly, the most frequently used formulas today will be upgraded into new versions in the future.

Table 22.7 Outcomes of MAE, percentage of eyes within ± 0.5 D and ± 1.0 D in several last studies

Study		Ladas SF AI	Hill- RBF	Kane	PEARL- DGS	Hoffer QST	Karmona	Nallasamy	Zhu-Lu	Zeiss AI
Kane 2017 (3122 eyes) [17]	MAE	0.40	0.41							
	± 0.5 D (%)	69.8	69.6							
	± 1.0 D (%)	94.3	94.3							
Connell 2019 (864 eyes) [23]	MAE		0.35	0.33						
	± 0.5 D (%)		75.3	77.9						
	± 1.0 D (%)		96.4	96.6						
Darcy 2020 (10,930 eyes) [43]	MAE		0.39	0.38						
	± 0.5 D (%)		71.2	72.0						
	± 1.0 D (%)		94.9	95.2						
Carmona-González 2020 (481 eyes) [36]	MAE		0.30				0.24			
	± 0.5 D (%)		78.9				90.4			
	± 1.0 D (%)		100				100			
Debellemanière 2021 (4242 eyes) [28]	MAE		0.30		0.29					
	± 0.5 D (%)				87.4					
	± 1.0 D (%)				97.8					
Shammas 2022 (595 eyes) [44]	MAE		0.31	0.30	0.29	0.31				
	± 0.5 D (%)		81.5	80.3	81.7	80.7				
	± 1.0 D (%)		99.5	99.7	99.8	99.8				
Guo 2023 (1828 eyes) [32]	MAE		0.38	0.49	0.55				0.34	
	± 0.5 D (%)		72.9	59.6	53.2				80.6	
	± 1.0 D (%)		92.0	88.0	86.0				96.0	
Kenny 2023 (278 eyes) [33]	MAE		0.43	0.44	0.45	0.50				0.40
	± 0.5 D (%)		68.0	67.0	67.0					73.0
	± 1.0 D (%)		93.0	92.0	91.0					93.0
Taroni 2023 (1259 eyes) [30]	MAE		0.38	0.36		0.36				
	± 0.5 D (%)		74.7	76.3		76.3				
	± 1.0 D (%)		95.0	96.0		96.7				
Li 2023 (6893 eyes) [38]	MAE				0.33			0.31		
	± 0.5 D (%)				77.7			80.2		
	± 1.0 D (%)				97.4			97.6		

MAE mean absolute error

Acknowledgements

Financial support No financial support was received for this submission.

Conflict of Interest The author declares that there is no conflict of interest with this submission.

References

1. Aristodemou P, Sparrow JM, Kaye S. Evaluating refractive outcomes after cataract surgery. *Ophthalmology*. 2019;126:13–8. <https://doi.org/10.1016/j.ophtha.2018.07.009>.
2. Alio JL, Grzybowski A, Aswad AE et al. Refractive lens exchange. *Surv Ophthalmol*. 2014;59(6):579–98. <https://doi.org/10.1016/j.survophthal.2014.04.004>.
3. Melles RB, Holladay JT, Chang WJ. Accuracy of intraocular lens calculation formulas. *Ophthalmology*. 2018;125(2):169–78. <https://doi.org/10.1016/j.ophtha.2017.08.027>.
4. Grzybowski A, Kanclerz P. The first idea of IOL implantation. *Acta Ophthalmol*. 2019;97(1):e142–4. <https://doi.org/10.1111/aos.13888>.
5. Williams HP. Sir Harold Ridley's vision. *Br J Ophthalmol*. 2001;85(9):1022–3. <https://doi.org/10.1136/bjo.85.9.1022>.
6. Oransky I, Charles Kelman. *Lancet*. 2004;364(9429):134. [https://doi.org/10.1016/S0140-6736\(04\)16646-4](https://doi.org/10.1016/S0140-6736(04)16646-4).

7. Fedorov SN, Kolinko AI. A method of calculating the optical power of the intraocular lens. *Vestn Oftalmol.* 1967;80(4):27–31. [Article in Russian].
8. Nemeth G, Modis L Jr. Accuracy of the Hill-radial basis function method and the Barrett Universal II formula. *Eur J Ophthalmol.* 2021;31(2):566–71. <https://doi.org/10.1177/1120672120902952>.
9. Stopyra W. The exactness of intraocular lens power calculation formulas for short eyes and correlation between method accuracy and eyeball axial length. *Cesk Slov. Oftal.* 2022 Summer;78(5):236–40. <https://doi.org/10.31348/2022/24>.
10. Lee AC, Quazi MA, Pepose JS. Biometry and intraocular lens power calculation. *Curr Opin Ophthalmol.* 2008;19(1):13–7. <https://doi.org/10.1097/ICU.0b013e3282f1c5ad>.
11. Stopyra W. Comparison of the accuracy of six intraocular lens power calculation formulas for eyes of axial length exceeding 25.0 mm. *J Fr Ophthalmol.* 2021;44(9):1332–9. <https://doi.org/10.1016/j.jfo.2021.04.009>.
12. Koch D, Hill W, Abulafia A et al. Pursuing perfection in intraocular lens calculations: I. Logical approach for classifying IOL calculation formulas. *J Cataract Refract Surg.* 2017;43(6):717–8. <https://doi.org/10.1016/j.jcrs.2017.06.006>.
13. Wendelstein JA, Hinterberger S, Hoffmann PC, Hirschnall N, Koss MJ, Langenbucher A, Jirak P, Bolz M. Evaluation of Phakic IOL power calculation using the new LHC formula and comparison with four conventional methods. *J Cataract Refract Surg.* 2022. <https://doi.org/10.1097/j.jcrs.0000000000001055>. Online ahead of print.
14. Li H, Nan L, Li J, Song H. Accuracy of intraocular lens power calculation formulas after laser refractive surgery in myopic eyes: a meta-analysis. *Eye Vis (Lond).* 2020;1(7):37. <https://doi.org/10.1186/s40662-020-00188-1>.
15. Wang L, Koch DD. Intraocular lens power calculations in eyes with previous corneal refractive surgery: review and expert opinion. *Ophthalmology.* 2021;128(11):e121–31. <https://doi.org/10.1016/j.optha.2020.06.054>.
16. Clarke GP, Burmeister J. Comparison of intraocular lens computations using a neural network versus the Holladay formula. *J Cataract Refract Surg.* 1997;23(10):1585–9. [https://doi.org/10.1016/s0886-3350\(97\)80034-x](https://doi.org/10.1016/s0886-3350(97)80034-x).
17. Kane JX, Van Heerden A, Atik A, et al. Accuracy of 3 new method for intraocular lens power selection. *J Cataract Refract Surg.* 2017;43(3):333–9. <https://doi.org/10.1016/j.jcrs.2016.12.021>.
18. Cassidy J, Nouri A, Betz V, et al. High-performance, robustly verified Monte Carlo simulation with FullMonte. *J Biomed Opt.* 2018;23(8):1–11. <https://doi.org/10.1117/1.JBO.23.8.085001>.
19. Ladas JG, Siddiqui AA, Devgan U, et al. A 3-D “super surface” combining intraocular lens formulas to generate a “super formula” and maximize accuracy. *JAMA Ophthalmol.* 2015;133(12):1431–6. <https://doi.org/10.1001/jamaophthalmol.2015.3832>.
20. Siddiqui AA, Juthani V, Kang J et al. The future of intraocular lens calculations: Ladas super formula. *Ann Eye Sci.* 2019;4:19. <https://doi.org/10.21037/aes.2019.04.02>.
21. Stopyra W. The accuracy of intraocular lens power calculation formulas for eyes longer than 25.00 mm and the correlation between method accuracy and eye axial length. *Klin Oczna.* 2021;123(3):129–33. <https://doi.org/10.5114/ko.2021.107849>.
22. Stopyra W. Effectiveness, sensitivity and specificity of intraocular lens power calculation formulas for short eyes. *Turk J Ophthalmol.* 2022;52(3):201–7. <https://doi.org/10.4274/tjo.galenos.2021.83704>.
23. Connell BJ, Kane JX. Comparison of the Kane formula with existing formulas for intraocular lens power selection. *BMJ Open Ophthalmol.* 2019;4(1): e000251. <https://doi.org/10.1136/bmjophth-2018-000251>.
24. Voytsekhivskyy OV, Hoffer KJ, Savini G, et al. Clinical accuracy of 18 IOL power formulas in 241 short eyes. *Curr Eye Res.* 2021;46(12):1832–43. <https://doi.org/10.1080/02713683.2021.1933056>.
25. Kane J. Kane calculator [Internet] <https://iolformula.com>. Accessed 31 Dec 2023.
26. Gatinel D, Debellemanière G, Saad A, et al. Determining the theoretical effective lens position of thick intraocular lenses for machine learning-based IOL power calculation and simulation. *Transl Vis Sci Technol.* 2021;10(4):27. <https://doi.org/10.1167/tvst.10.4.27>.
27. Debellemanière G, Saad A, Gatinel D. PEARL DGS calculator [Internet] IOL Solver <https://iolsolver.com>. Accessed 31 Dec 2023.
28. Debellemanière G, Dubois M, Gauvin M, et al. The PEARL-DGS formula: the development of an open-source machine learning-based thick IOL calculation formula. *Am J Ophthalmol.* 2021;232:58–69. <https://doi.org/10.1016/j.ajo.2021.05.004>.
29. Cooke DL, Cooke TL. Approximating sum-of-segments axial length from a traditional optical low-coherence reflectometry measurement. *J Cataract Refract Surg.* 2019;45(3):351–4. <https://doi.org/10.1016/j.jcrs.2018.12.026>.
30. Taroni L, Hoffer KJ, Pellegrini M, et al. Comparison of the new Hoffer QST with 4 modern accurate formulas. *J Cataract Refract Surg.* 2023;49(4):378–84. <https://doi.org/10.1097/j.jcrs.0000000000001126>.
31. Hoffer Q, Savini G. Uptadint the Hoffer Q formula using algorithms and machine learning. [Internet] <https://ascrs.org/clinical-education/abstracts/2021/>. Accessed 31 Dec 2023.
32. Guo D, He W, Wei L, Song Y, Qi J, Yao Y, Chen X, Huang J, Lu Yi, Zhu X. The Zhu-Lu formula: a machine learning-based intraocular lens power calculation formula for highly myopic eyes. *Eye Vis (Lond).* 2023;10(1):26. <https://doi.org/10.1186/s40662-023-00342-5>.

33. Kenny PI, Kozhaya K, Truong P, Weikert MP, Wang L, Hill WE, Koch DD. Efficacy of segmented axial length and artificial intelligence approaches to intraocular lens power calculation in short eyes. *J Cataract Refract Surg.* 2023;49(7):697–703. <https://doi.org/10.1097/j.jcrs.0000000000001185>.
34. Zeiss AI calculator. [Internet] <https://zcalc.meditec.zeiss.com>.
35. Hill W. Hill-RBF calculator [Internet] <https://rbfcalculator.com>. Accessed 31 Dec 2023.
36. Carmona-González D, Palomino-Bautista C. Accuracy of a new intraocular lens power calculation method based on artificial intelligence. *Eye (Lond).* 2021;35(2):517–22. <https://doi.org/10.1038/s41433-020-0883-3>.
37. Carmona-González D. Karmona calculator [Internet] <https://karmona-iol.com>. Accessed 31 Dec 2023.
38. Li T, Stein J, Nallasamy N. Evaluation of the Nallasamy formula: a stacking ensemble machine learning method for refraction prediction in cataract surgery. *Br J Ophthalmol.* 2023;107(8):1066–71. <https://doi.org/10.1136/bjophthalmol-2021-320599>.
39. Sramka M, Slovak M, Tuckova J, Stodulka P. Improving clinical refractive results of cataract surgery by machine learning. *PeerJ.* 2019;2(7): e7202. <https://doi.org/10.7717/peerj.7202>.
40. De Rojas J, Ladas J. AI models for IOL calculation. *CRSToday*. Sept/Oct 2023.
41. Hoffer KJ, Savini G. Update on intraocular lens power calculation study protocols: the better way to design and report clinical trials. *Ophthalmology.* 2021;128(11):e115–20. <https://doi.org/10.1016/j.optha.2020.07.05>.
42. Cooke DL, Cooke TL. Comparison of 9 intraocular lens power calculation formulas. *J Cataract Refract Surg.* 2016;42(8):1157–64. <https://doi.org/10.1016/j.jcrs.2016.06.029>.
43. Darcy K, Gunn D, Tavassoli S, Sparrow J, Kane JX. Assessment of the accuracy of new and updated intraocular lens power calculation formulas in 10930 eyes from the UK National Health Service. *J Cataract Refract Surg.* 2020;46(1):2–7. <https://doi.org/10.1016/j.jcrs.2019.08.014>.
44. Shammas HJ, Taroni L, Pellegrini M et al. Accuracy of never IOL power formulas in short and long eyes using sum-of-segment biometry. *J Cataract Refract Surg.* 2022;48(10):1113–20. <https://doi.org/10.1097/j.jcrs.0000000000000958>.
45. Kane JX, Melles RB. Intraocular lens formula comparison in axial hyperopia with a high-power intraocular lens of 30 or more diopters. *J Cataract Refract Surg.* 2020;46(9):1236–9. <https://doi.org/10.1097/j.jcrs.0000000000000235>.
46. Cheng H, Wang L, Kane JX, et al. Accuracy of artificial intelligence formulas and axial length adjustments for highly myopic eyes. *Am J Ophthalmol.* 2021;223:100–7. <https://doi.org/10.1016/j.ajo.2020.09.019>.
47. Savini G, Di Maita M, Hoffer KJ, et al. Comparison of 13 formulas for IOL power calculation with measurements from partial coherence interferometry. *Br J Ophthalmol.* 2021;105(4):484–9. <https://doi.org/10.1136/bjophthalmol-2020-316193>.
48. Chang P, Qian S, Wang Y, Li S, Yang F, Hu Y, Liu Z, Zhao Y. Accuracy of new-generation intraocular lens calculation formulas in eyes with variations in predicted refraction. *Graefes Arch Clin Exp Ophthalmol.* 2022. <https://doi.org/10.1007/s00417-022-05748-w>.
49. Hou M, Ding Y, Liu L, et al. Accuracy of intraocular lens power calculation in primary angle-closure disease: comparison of 7 formulas. *Graefes Arch Clin Exp Ophthalmol.* 2021;259(12):3739–47. <https://doi.org/10.1007/s00417-021-05295-w>.
50. Rocha-de-Lossada C, Colmenero-Reina E, Flikier D et al. Intraocular lens power calculation formula accuracy: comparison of 12 formulas for trifocal hydrophilic intraocular lens. *Eur J Ophthalmol.* 2021;31(6):2981–88. <https://doi.org/10.1177/1120672120980690>.
51. Hipólito-Fernandes D, Luís ME, Serras-Pereira R, et al. Anterior chamber depth, lens thickness and intraocular lens calculation formula accuracy: nine formulas comparison. *Br J Ophthalmol.* 2022;106(3):349–55.
52. Li T, Stein J, Nallasamy N. AI-powered effective lens position prediction improves the accuracy of existing lens formulas. *Br J Ophthalmol.* 2022;106(9):1222–6. <https://doi.org/10.1136/bjophthalmol-2020-318321>.

Relevance and Impact of AI in the Pharmaceutical Industry—Focus on Clinical Development

23

Andreas Maunz, Yun Yvonna Li, Carl Glittenberg, and Sascha Fauser

Abstract

This article outlines the role of image analysis using AI in the Pharmaceutical Industry, especially for clinical research and development. The focus is on relevance and impact image-based AI may have in the industry, particularly how it may help solving challenges in clinical development, taking into account recent achievements and trends in machine learning. The state-of-the-art in computational analysis of optical coherence tomography (OCT) involves deep learning techniques to segment OCT images into relevant anatomical structures and quantify tissue phenotypes to help characterize disease status. With the introduction of these techniques, OCT quantification has been improved, leading to more reliable diagnosis and treatment outcomes, as well as as well training of predictive models on the quantified features, or directly on the input images.

Keywords

Artificial intelligence · Machine learning · Clinical development · Retinal segmentation · Optical coherence tomography

Introduction

In clinical practice and in research, Optical Coherence Tomography (OCT), especially Spectral Domain Optical Coherence Tomography (SD-OCT), has become an extremely important image modality. It uses light to capture two- and three-dimensional images of the retina with near-cellular resolution [1], allowing for fast and non-invasive acquisition of high-resolution images of the retina and choroid. It therefore plays a highly significant role in clinical diagnosis as well as in treatment decisions, for example in diagnosing and monitoring ocular conditions, such as age-related macular degeneration (AMD), diabetic macular edema (DME), or glaucoma. It has also been used to study the effects of various drugs on the retina, and to evaluate the efficacy of treatments such as anti-vascular endothelial growth factor (VEGF) therapy for neovascular age-related macular degeneration (nAMD).

A. Maunz (✉) · Y. Y. Li · C. Glittenberg · S. Fauser
F. Hoffmann-La Roche AG, Basel, Switzerland
e-mail: andreas.maunz@roche.com

Many features hitherto have been assessed manually and mostly qualitatively from OCTs, for example, a standard question in clinical practice and clinical trials is the presence (“wet”) or absence (“dry”) of intra- or subretinal fluid. The fact that the data is digitally acquired provides a great foundation for automated image analysis techniques, which in turn allows practitioners with all levels of experience to screen for signs of diseases and detect them in an efficient and objective manner. Among the different computerized analytical approaches, AI-based methods (“deep learning”) have shown superior performance to earlier image processing [2].

In the following, we highlight some unique advantages for the use of AI in clinical research and development.

AI Provides Fast and Cost Effective Assessment of Features

AI-based image analysis allows faster and more cost effective assessment of features, eliminating human error, as well as the noise due to human intervention. Once established, an image analysis system can measure various features quantitatively (drusen volume, fluid volume, etc.) at high speed, with constant bias, and at almost no additional cost. It can also easily be repeated or done as soon as new images are available.

AI Analyzes and Produces High-Dimensional, Quantitative Data

High-dimensional data refers to data that has many features or dimensions, for example an image can have millions of pixels, each of which is a feature or dimension [3]. Data is quantitative if captured on a continuous scale, such as volumes, areas, distances, etc. Quantitative data contains more information than qualitative data, e.g. presence or absence of a pathology. Labeling pixels, or taking measurements on images, however, can only be done to a limited extent by human workers in

a slow, labor-intensive and error-prone process. Examples include drusen, intra- and subretinal fluid, geographic atrophy lesions.

AI Adds a Data-Driven Perspective

Hypothesis testing is a method based on existing knowledge, and is used to generate new knowledge by testing theories. Clinical trials use this framework to test for differences between treatment arms. As such, the method is fundamentally limited by preformulated hypotheses. AI, on the other hand, can analyze large amounts of data and identify patterns that humans may not be able to see, allowing for a wider discovery range than hypothesis testing by design. For example, studies have shown that sex can be predicted from retinal fundus photographs, whereas clinicians are currently unaware of distinct retinal feature variations between males and females [4]. Similar findings were made for age, smoking status, and various systemic biomarkers [5]. However, it is important to note that AI can also be biased if the training data is, which makes it important to subsequently confirm its findings using other means.

AI Makes More Data Accessible

AI can unlock data types with highly sophisticated structure, such as audio, images, videos (sensor data in general), but can also work well on unstructured data, such as free text. For example, AI can be used to analyze medical images or videos, as well as health record data in free text form. Prior to the advent of deep learning, access to such data using automated analyses has been very limited. Working with such data previously required human intervention and expert knowledge, at the cost of increase in level of noise, higher financial burden, narrower scope, and slower progress. On top of that, many analyses are not even feasible or prohibitively expensive using human workers.

Together, these aspects form a new data-driven, fully quantitative space of data analysis, deeper and more comprehensive than previously possible.

The rest of this chapter is structured as follows: In the first part we focus on characteristic properties of Optical Coherence Tomography (OCT) data, highlighting specific challenges, forcing practitioners to take difficult decisions, but also how to leverage unlabeled data, which may alleviate some of the same challenges. In the second part we outline the impact of OCT-based AI may have in ophthalmology, particularly how it may help to solve challenges in clinical development.

Properties of Clinical Image Datasets and Impact on Their Analysis

Fundamental decisions regarding dataset composition and image processing have to be taken when working with actual image datasets. This section illustrates some frequent challenges and decision points encountered by practitioners of computer vision on SD-OCT.

Image Defects in SD-OCT

In foveal OCT scans, frequently found problems include random translations and rotations, parts of the scan being cut off (Fig. 23.1), or off-center images. Motion artifacts due to eye movement, shadowing artifacts due to the presence of blood vessels, or other structures that block the light beam, and signal attenuation due

to scattering and absorption of light by ocular media, are also frequently encountered [1].

There are methods for signal enhancement to address the challenge introduced by the limitation of imaging quality, however additional defects may be introduced in the process. For example, some device manufacturers improve the signal-to-noise ratio by averaging multiple B-scans to enhance the signal-to-noise ratio. A study by the University of Vienna found that, for Heidelberg Spectralis SD-OCT scans, real time averaging using 16 B-Scans provided the best tradeoff for pathology segmentation in patients with clear optical media [6]. While frequently helpful, this method also introduces new defects, such as contrast and brightness differences within the same B-Scan (Fig. 23.2).

Such artifacts are frequently found in clinical image datasets: in one of our recent studies including approximately 200 patients with either nAMD or diabetic macular edema (DME), noticeable artifacts were found in 15% of the OCT scans. In case of fluid detection using deep learning, low brightness on certain B-Scans can lead to false positive detection of subretinal fluid (SRF) when using convolutional neural networks (Fig. 23.3).

“Small” Versus “Big” Data

More data often leads to models that produce better results if the data is representative of the population. At the same time, generating clinical image data (images and annotations) is an expensive task. At the same time, the amount of images required for robust performance in different AI tasks is hard to predict upfront; it

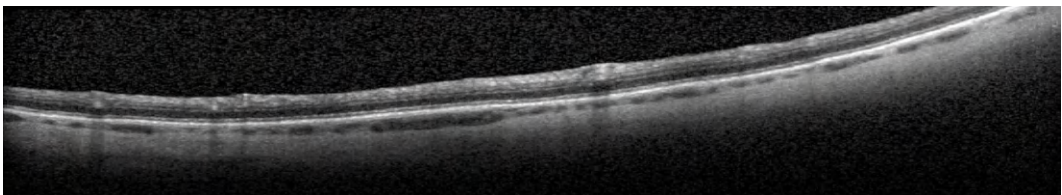


Fig. 23.1 B-Scan cut off due to rotation (top right). Such cases may limit the application of segmentation of retinal layers, and the calculation of thicknesses and volumes

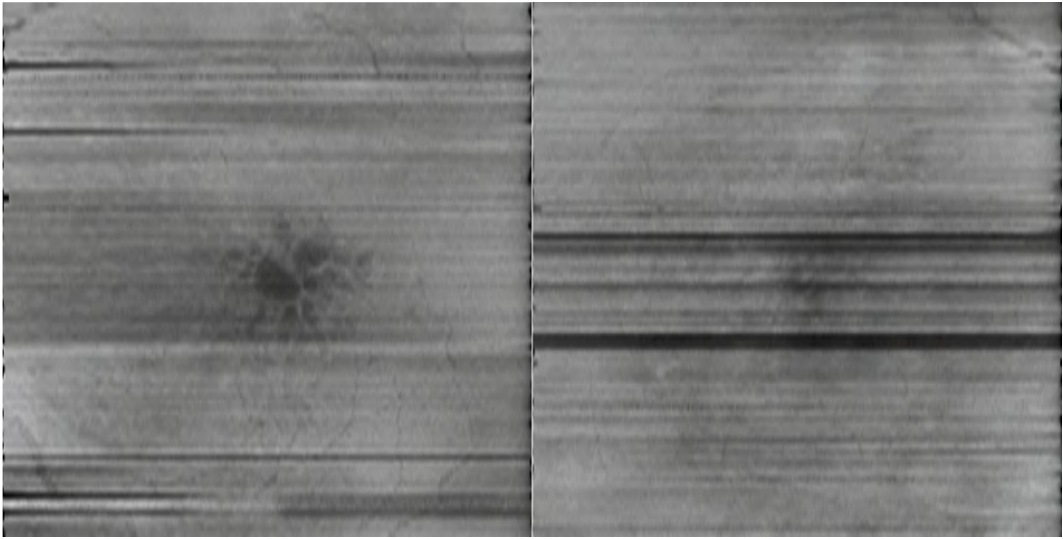


Fig. 23.2 En-face projections of Heidelberg Spectralis OCT scans with brightness and contrast differences, using 16 B-Scans for averaging. For the case on the left

these properties vary within a given B-Scan, whereas the scan on the right shows very few intra-B-Scan fluctuations

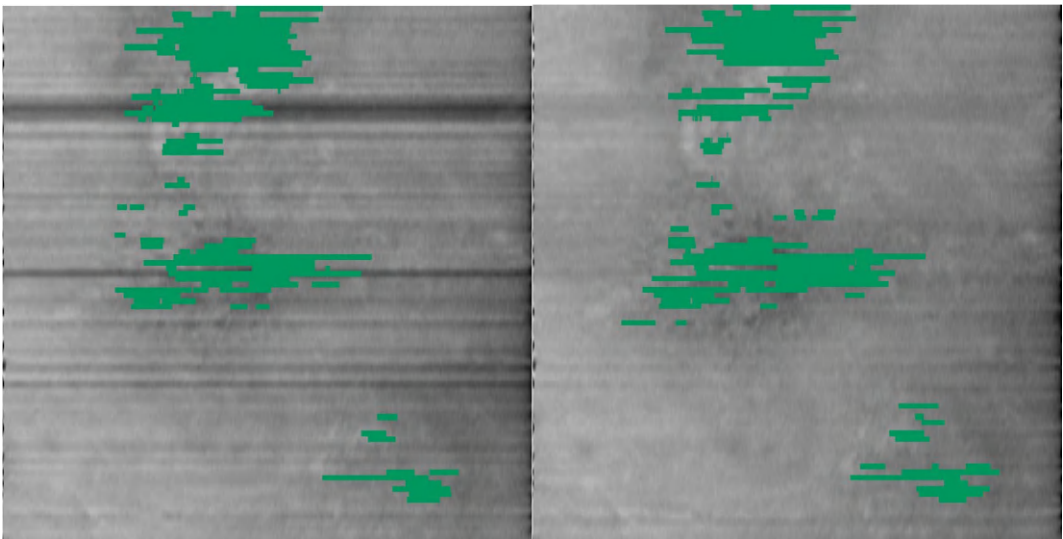


Fig. 23.3 En-face projections of OCT volumes, overlaid with detected subretinal fluid. More SRF is detected in the darker area on the original scan (left), where less SRF is detected after histogram equalization (right)

depends on the complexity of the task and the type of algorithm used, as well as on image quality, e.g. absence of defects, and the signal-to-noise ratio.

Large amounts of high-quality images would be optimal, but in practice it is often helpful to explore techniques for enhancing the effective

image set size, helping to enhance model robustness without producing more data. The following practical questions may be relevant:

- **Small set of high-quality images versus large set of mixed-quality images:** Is it more important to strive for the best image

quality possible, or is it more important to have a larger image set size, increasing the variety of images seen during training and extending the scope of the model?

- **Enhancing image quality through preprocessing versus letting the neural network deal with quality issues:** Is it worthwhile investing into preprocessing to enhance the signal-to-noise ratio or to remove artifacts (as with histogram equalization in the previous section), possibly at the cost of introducing new problems, or should the investment be made to develop AI algorithms that can deal flexibly with varying quality, similarly to how a human would?

Depending on the use case for an AI model, different levels of robustness may be required. For example, when applied to real-world data, robustness requirements may be higher, compared to controlled conditions in clinical trials.

The above considerations will influence the selection of images, as well as image preprocessing techniques. Since no general rules can be applied, it will be necessary to try out various options in practice. Besides signal enhancement and artifact removal, preprocessing can also include “boosting” of the dataset by adding normalized, resized, cropped, or otherwise perturbed versions of the images. Such image augmentations, which are readily available in deep learning libraries, can help improve the robustness of the neural network, and make it less likely to overfit to intricacies of the specific set of images.

Leveraging Unannotated Data

One argument for not overly investing into image pre-processing can be found in recent technical progress made in machine learning methods that enable the use of unannotated data alongside with annotated data, which engenders much more potent and practical applications, combining image sets across devices and quality levels. This trend in methodology fits well with clinical practice, where images are acquired

for a diverse set of patient population, pathological presentation, and image acquisition settings, including imaging devices and variation in image quality. Such a setting makes the annotation of a representative set of data points more difficult. Without a representative set of annotations, supervised learning methods can have difficulties to achieve good performance, especially for out-of-distribution data, i.e. data not represented in the training set. Annotations are also expensive to obtain on a large scale, especially when the goal is to be able to generalize across imaging devices.

One answer to this problem of diversity in data is self-supervised and semi-supervised learning (SSL). Self-supervised learning aims to learn from images without any annotation and semi-supervised learning aims to learn from a dataset containing both unannotated and annotated images [7].

Experiments with Contrastive Learning

Out of the many different SSL approaches for incorporating information from both labeled and unlabeled images, contrastive learning based on the SimCLR framework [8] presents a simple yet powerful option. The SimCLR framework was originally designed for image classification tasks. It defines contrastive loss between similar pairs of images (which “attract” each other) and between dissimilar pairs of images (which “repel” each other). The approach was adapted for image segmentation and applied to a variety of OCT images taken from different “domains”, i.e. different diseases and OCT device models that were available for analysis. Supervised segmentation loss applied on labeled images was trained in combination with the contrastive loss on both labeled and unlabeled images. Additionally, the 3D nature of OCT images was leveraged to generate positive pairs from slices near each other in 3D. The method (coined “SegCLR”, and depicted in Fig. 23.4) was tested on a dataset acquired by two different types of OCT scanners, Heidelberg Spectralis and Zeiss

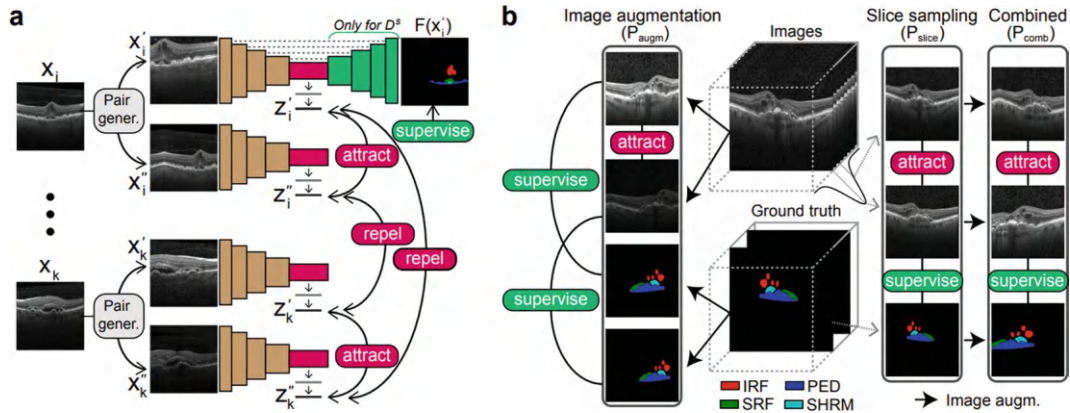


Fig. 23.4 Overview of the SegCLR framework. Panel **a** shows the architecture schematic where slices from the same domain attract each other and slices from different

domains repel each other. Panel **b** shows details on positive pair generation based on image augmentation and nearness of slices along the 3rd dimension

Cirrus. Annotated Spectralis scans formed the “source domain” scans and unannotated scans from Cirrus the “target domain”. The proposed method not only outperformed the baseline methods on the target domain, it also was proven to be advantageous on the source domain [9].

Extending the idea of training on contrastive loss from both the target and source domain, we trained a so-called zero-shot contrastive learning model using labeled and unlabeled data from only a single OCT scanner (Spectralis) and model performance was tested on a dataset from a different scanner (Topcon). Figure 23.5 shows qualitative results, indicating the ability to generalize to an unseen domain while the supervised model trained on the same domain is not able to generalize very well to the target domain and significantly over-segments.

We find that contrastive learning as one example of SSL shows promise to reduce reliance on resource-intensive and slow manual annotation effort. The recent emergence of foundation models, which expands on the idea of SSL with the ambition to incorporate diverse data sources, further point to the potential of learning from unlabeled data in conjunction with labeled data [10]. Nevertheless, validating a model that generalizes to out-of-distribution data will always require a significant amount of annotated data.

Use Cases for AI-Based Image Analysis in the Pharmaceutical Industry

In clinical practice, AI is mostly used to improve disease diagnosis and assist decision-making in diseases such as diabetic retinopathy (DR), age-related macular degeneration (AMD) or glaucoma [11, 12]. In research and development, on the other hand, more insights into disease biology is the primary focus. Specifically, in clinical development AI-based quantitative analyses can play a role in prescreening, profiling and subgrouping of patients, as well as detailed comparison of retinal morphology across arms. In Fig. 23.6 we outline this in more detail for “disease clusters”, namely diabetic Indications and retinal vein occlusion, neovascular AMD, as well as intermediate AMD and Geographic Atrophy.

As shown in Fig. 23.6, SD-OCT is the most frequently used imaging modality to be applied across disease areas, and quantitative measurements derived from OCT segmentations form a pivotal element in quantification of disease pathology. OCT images can be used to detect intra- and subretinal fluids [13, 14], which are indicators of disease activity in nAMD and DME, and used for treatment decisions and patient monitoring. In nAMD, OCT can also be used to detect Subretinal HypeReflective

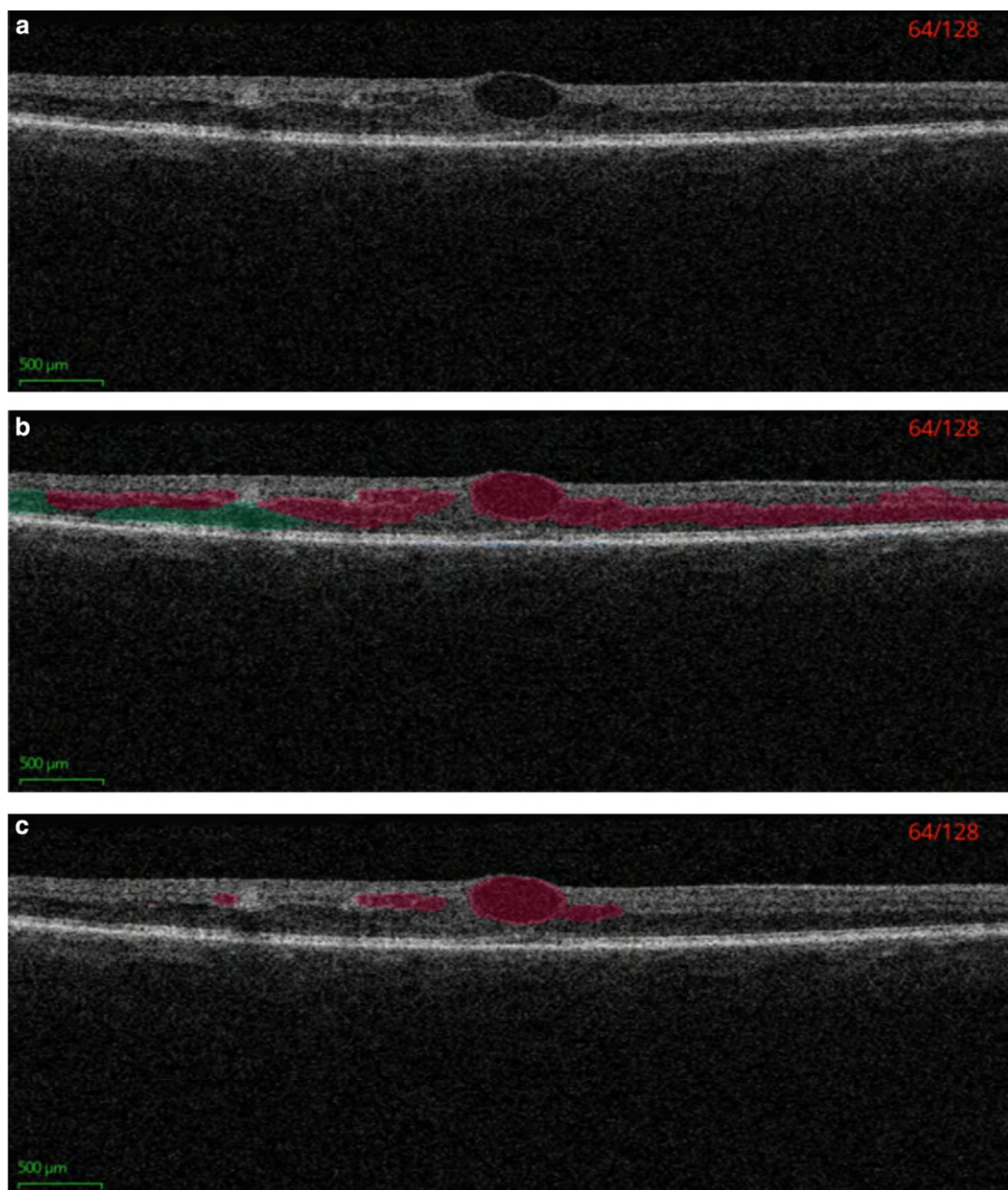


Fig. 23.5 **a** Original Topcon image. **b** Topcon image segmented with supervised model trained with Spectralis images. **c** Topcon image segmented with semi-supervised model trained with Spectralis images

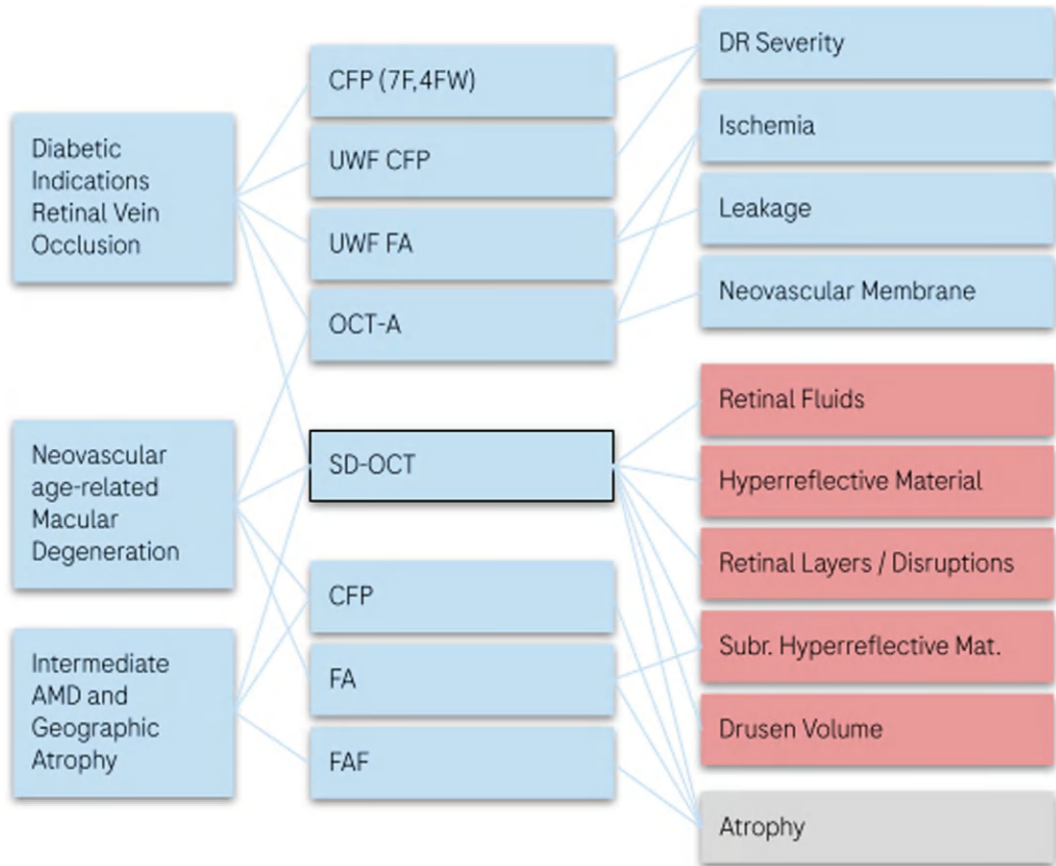


Fig. 23.6 “Web of Modalities” in clinical research, highlighting frequently used modalities per disease area and their connection to pathology/retinal landmarks. SD-OCT is the most densely connected node

Material (SHRM) and Pigment Epithelial Detachment (PED) [15].

Patterns of leakage and ischemia play another major role in diabetic indications, and are analyzed using CFP, FA, and OCT-A. In the remainder of this chapter, we detail some use cases of OCT-based AI in clinical development.

Monitoring of Eyes Under Treatment

This use case involves detecting changes in morphology or function that indicate response to therapy, which further enables personalized and adaptive treatment plans for patients with eye diseases such as diabetic macular edema or age-related macular degeneration [16]. Recent

advances in deep learning have enabled automated segmentation of various retinal pathologies and hyperreflective material on OCT images. Corresponding OCT segmentation models (see Fig. 23.7) have been shown to be effective in detecting, segmenting and providing quantitative metrics for different types of retinal fluids or hyperreflective material, for example, to monitor changes in retinal pathology levels over time in patients with nAMD [17]. Going forward, patients may be enabled to monitor the disease status daily from home using a portable OCT device, promising a significantly higher rate of image sampling, compared to clinical studies. In such a situation AI-based analysis [18] may be even more impactful, and also required from a feasibility perspective.

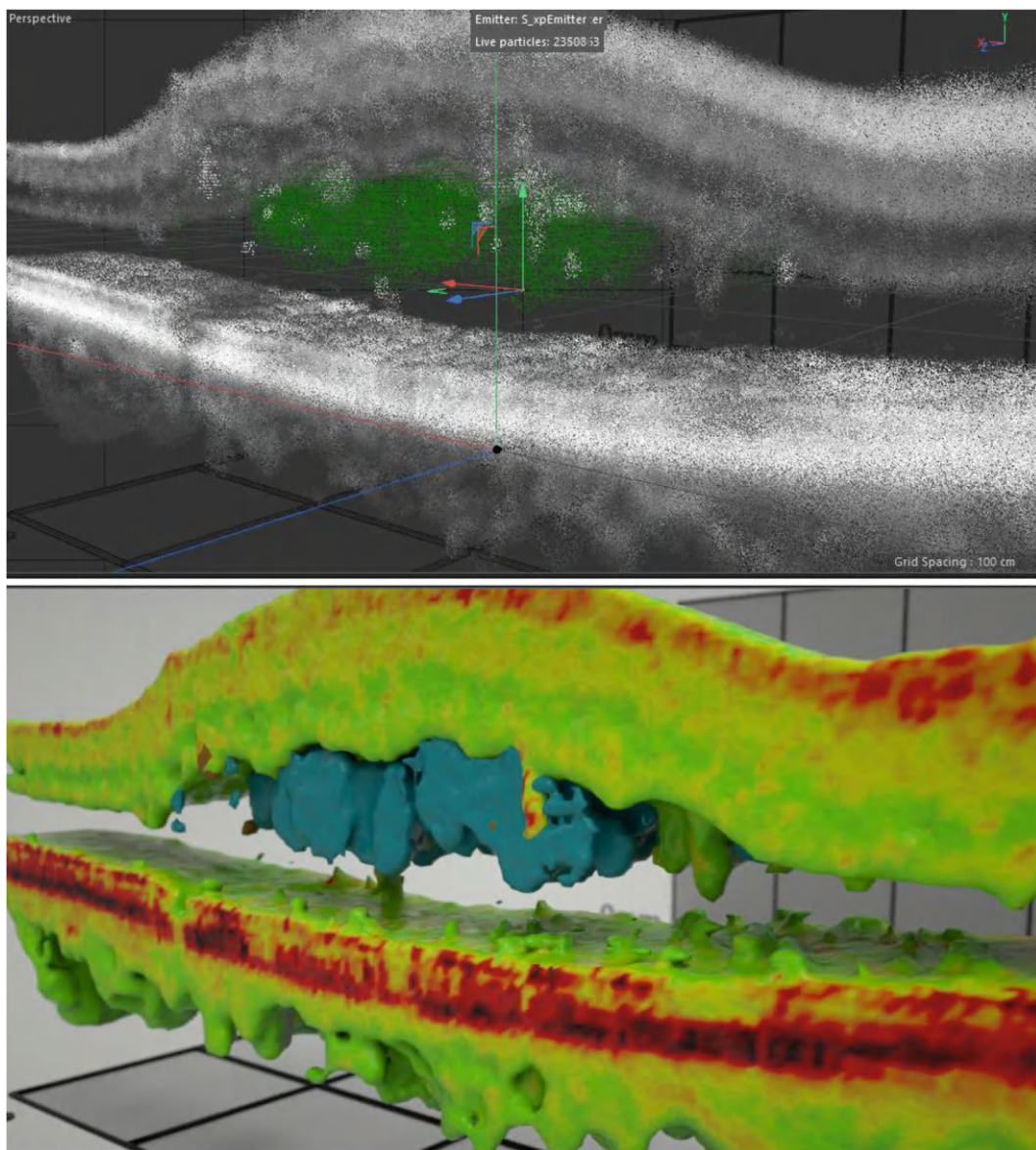


Fig. 23.7 Rendering of volumetric SD-OCT data. Original scan (top) and segmentation by a deep learning-based OCT segmentation model (bottom)

Digitally-Enhanced Screening Through Automated, Real-Time, Quantitative Measurements

This use case involves using machine learning to transform raw images into a preclinical screening assay that can capture the effects of chemical, genetic or pathological perturbations

on cells or tissues [19], or to use retinal images to profile and subgroup patients [16]. This can enable equally accurate, but faster and cheaper identification of potential drug candidates or targets, as well as the design of safer and more cost-effective trials.

In a clinical example, patient screening for enrollment into a DME study typically involves

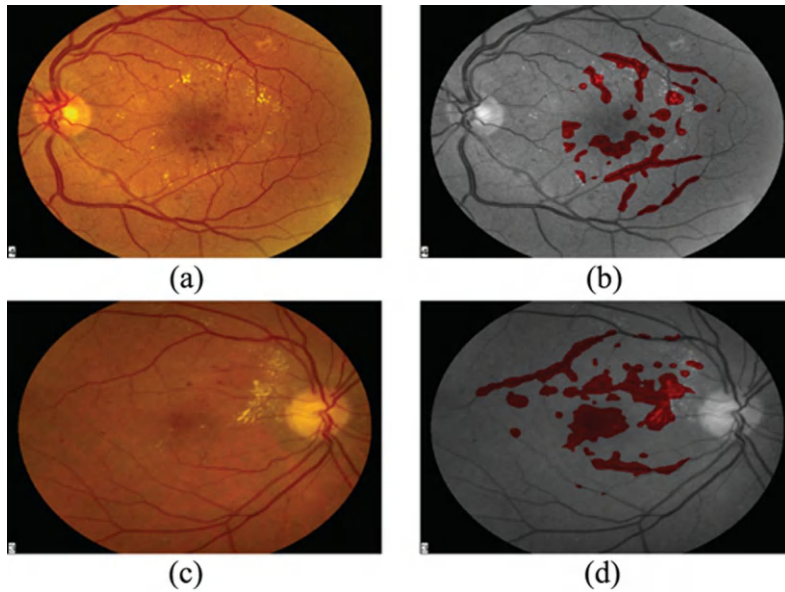


Fig. 23.8 Examples of attribution maps or “hot spots” on CFPs generated by the deep learning (DL) classification models. **a, b** Example of map created by the DL

model to detect MT with CFT > 250 μm . **c, d** Example of map created by the DL model to detect MT with CFT > 400 μm

assessment of central subfield thickness. This has been done using a reading center, where human graders assess the thickness based on OCT, which is a costly, tedious and error-prone process. It has been shown, however, that the required assessment can be done with a deep learning model trained on 2-D fundus images without explicit depth information [20], see Fig. 23.8. Besides avoiding the aforementioned downsides, the model can also be applied in rural settings or in developing countries, where OCT devices are not available.

Enrichment of Study Arms

This use case analyzes retinal morphology at screening to designate patients to study arms, either by selecting for imaging biomarker profiles of interest, or by predicting the response to a drug or progression of the disease [21]. This approach can enable enrichment of study arms with patients who are more likely to benefit from a drug and reduce the sample size and cost of clinical trials.

For example, a deep learning model trained on OCT images taken from eyes with untreated neovascular AMD at baseline was developed to predict the onset of fibrosis after 12 months, which is clinically detected by Fluorescein Angiography (FA). Such a model could ultimately be used to prevent fibrosis by treating patients before the onset, or to design a study with an “enriched” set of patients who are more likely to develop fibrosis soon.

The specific type of choroidal neovascularization (CNV), which is currently clinically determined using FA, could be replaced by predictive models trained on OCT data, which showed the same level of performance as FA-based models. Replacing FA acquisition with OCT not only shortens acquisition time, OCT is also non-invasive and does not require as much specialized training compared to FA [22].

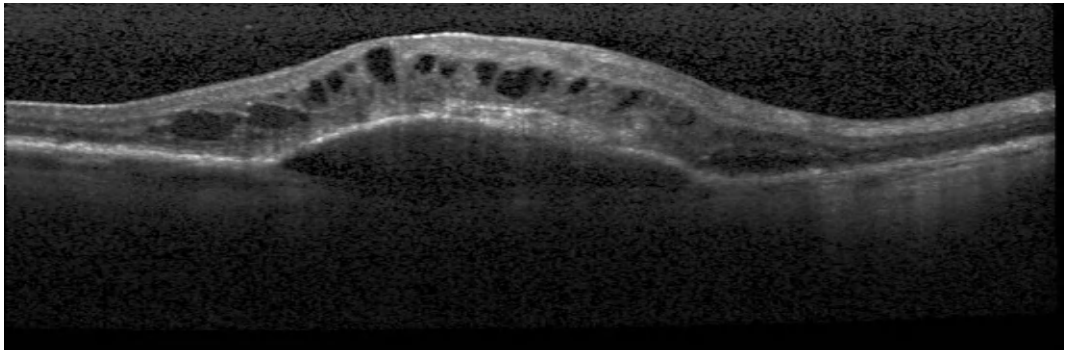
An additional insight discovered from the fibrosis use cases shows that intermediate fluid and layer segmentation derived from OCT images produced models that converge much quicker and yielded better performance,

compared to training an end to end deep learning model directly on the original OCT images (Fig. 23.9).

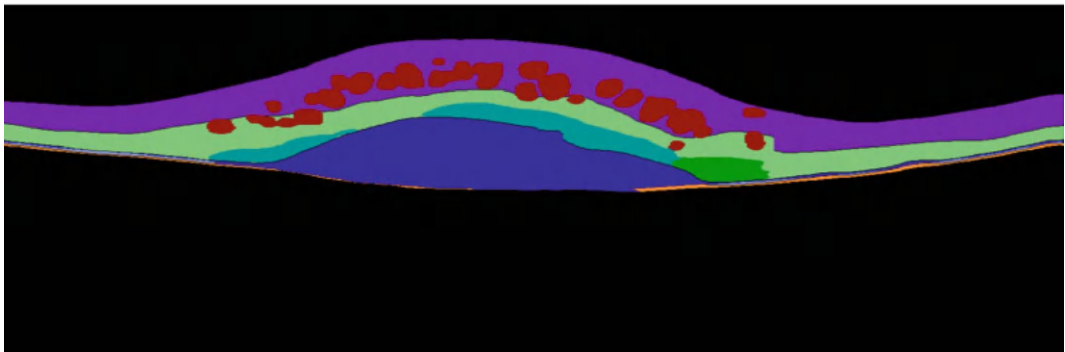
Another example is a machine learning model to predict treatment response in AMD patients, more specifically, change in visual acuity in response to ranibizumab therapy, by using features extracted from the segmented fluid and layer representation. Baseline data from eyes with untreated neovascular AMD at baseline was used to train a model that predicts future response to ranibizumab treatment better than standard clinical variables, e.g. baseline visual acuity and age alone, with statistical significance. However, further developments will be needed to realize clinical utility [23].

Establish Secondary, Digital, Quantitative Outcomes Describing Morphology

This use case uses machine learning to extract quantitative features from images which represent the amount, volume, or morphology of cells, pathologies, or tissues [16]. As of today, these features are mostly relegated to exploratory outcomes in trials [24], and serve to provide additional insights into the mechanism of action or efficacy of a drug. However, digital endpoints also have potential to fundamentally modernize the way clinical trials measure efficacy. Implementation barriers remain high, though. Currently most applications are not imaging-based, but frequently use wearable sensors



Raw representation: Preserves full image but may extrapolate less well to other scanners.



Segmented representation: Influenced by expert knowledge but retains spatial information and may extrapolate better to other OCT scanner types.

Fig. 23.9 End-to-end Deep Learning input for fibrosis prediction: Raw (top) and segmented (bottom, assembled from DL-based segmentations) representations

[25]. These can range from mobility measures for patients with amyotrophic lateral sclerosis (ALS) to smartphone-based Parkinson's disease severity assessment [26].

Post-hoc Analysis of Clinical Studies for R&D

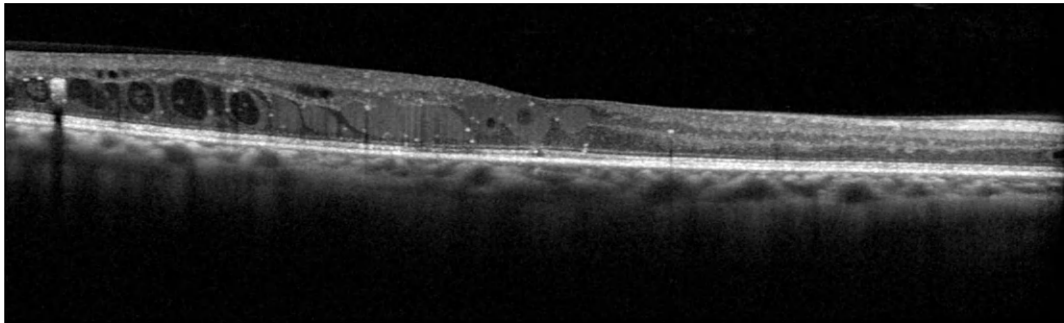
This use case involves post-hoc analysis of clinical studies [16, 27] to enable molecule differentiation using imaging biomarkers. It aims to discover new biomarkers, subgroups, mechanisms or indications that were neither anticipated nor planned in the original study design.

For example, hyperreflective foci (HRF) can be seen on OCT in various retinal diseases. They appear as small bright dots and are linked to disease severity (see Fig. 23.10). Their histopathological correlates are hypothesized to be displaced retinal pigment epithelium cells, macrophages or extravascular material. While their assessment is in principle possible by manual

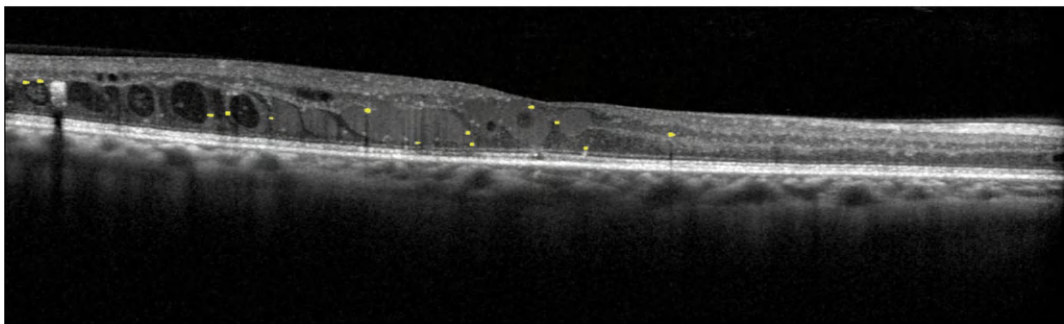
counting, it is usually too time consuming, and volume measurements are impossible. A deep learning model that can automatically identify, count and measure the volume of HRF was developed to better understand the association between disease severity and distribution and change of HRF under therapy.

One recent analysis confirmed the link of HRF to disease severity on a much larger data set than before, by using baseline data from approximately 1500 patients pooled from two recent studies, YOSEMITE (NCT03622580) and RHINE (NCT03622593). It was found that HRF volume was significantly negatively correlated to visual acuity and diabetic retinopathy severity score (DRSS).

In another analysis, data from patients with DME treated with two different medications, faricimab and aflibercept, was analyzed for differences in HRF reduction. Faricimab showed a significantly greater reduction of HRF in the inner retina than aflibercept [28], where HRF volumes reduced more markedly from baseline



OCT Scan with HRF.



The same scan with HRF detected.

Fig. 23.10 B-Scan without (top) and with automated segmentation of HRF (bottom; yellow)

to week 48 for both faricimab administered every 8 weeks (-118.29 pL), and faricimab administered using a treat-and-extend regimen (-130.05 pL), than aflibercept administered every 8 weeks (-58.67 pL; $p=0.0006$ and $p<0.0001$). Similar significant results were obtained for the outer retina. This effect could potentially be explained by an anti-inflammatory effect of faricimab, and therefore might also point towards the origin of HRFs [29].

Conclusions

In this chapter we outlined the relevance and impact image-based AI may have in clinical development of new therapies. The applications are numerous and span the whole drug development process from biomarker identification and disease understanding, to improving the design and lowering the cost of clinical studies, and to automated patient monitoring. Although the number of applications is still limited, the potential is enormous and can help to improve drug development in ophthalmology considerably. We highlighted particularly attractive examples, but also challenges to overcome in AI model design. These challenges include potentially conflicting requirements for increasing dataset size, dataset quality, and the associated cost. Recent developments in the machine learning community to leverage additional unannotated data may provide help here.

Bibliographys

1. Aumann S, Donner S, Fischer J, Müller F. Optical Coherence Tomography (OCT): principle and technical realization. In: Bille JF, editor. High resolution imaging in microscopy and ophthalmology. Springer International Publishing; 2019. p. 59–85. https://doi.org/10.1007/978-3-030-16638-0_3.
2. Tong Y, Lu W, Yu Y, Shen Y. Application of machine learning in ophthalmic imaging modalities. *Eye and Vis.* 2020;7:22.
3. Muller J, et al. Integrated dual analysis of quantitative and qualitative high-dimensional data. *IEEE Trans Visual Comput Graphics.* 2021;27:2953–66.
4. Korot E, et al. Predicting sex from retinal fundus photographs using automated deep learning. *Sci Rep.* 2021;11:10286.
5. Poplin R, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng.* 2018;2:158–64.
6. Podkowinski D, et al. Impact of B-scan averaging on spectralis optical coherence tomography image quality before and after cataract surgery. *J Ophthalmol.* 2017;2017:1–8.
7. Chaitanya K, Erdil E, Karani N, Konukoglu E. Local contrastive loss with pseudo-label based self-training for semi-supervised medical image segmentation. *Med Image Anal.* 2023;87:102792.
8. Chen T, Kornblith S, Norouzi M, Hinton G. A Simple framework for contrastive learning of visual representations. 2020. Preprint at <http://arxiv.org/abs/2002.05709>.
9. Gomariz A, et al. Unsupervised domain adaptation with contrastive learning for OCT segmentation. 2022. Preprint at <http://arxiv.org/abs/2203.03664>.
10. Wang D, et al. A real-world dataset and benchmark for foundation model adaptation in medical image classification. *Sci Data.* 2023;10:574.
11. Jin K, Ye J. Artificial intelligence and deep learning in ophthalmology: Current status and future perspectives. *Adv Ophthalmol Pract Res.* 2022;2:100078.
12. Ting DSW, et al. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol.* 2019;103:167–75.
13. Hsu H-Y, et al. Automatic segmentation of retinal fluid and photoreceptor layer from optical coherence tomography images of diabetic macular edema patients using deep learning and associations with visual acuity. *Biomedicine.* 2022;10:1269.
14. Trichonas G, Kaiser PK. Optical coherence tomography imaging of macular oedema. *Br J Ophthalmol.* 2014;98:ii24–ii29.
15. Siggel R, Spital C, Lentzsch A, Liakopoulos S. Optical coherence tomography angiography for the detection of macular neovascularization—comparison of en face versus cross-sectional view. *Eye.* 2023;37:256–62.
16. Shah P, et al. Artificial intelligence and machine learning in clinical development: a translational perspective. *npj Digit Med.* 2019;2:69.
17. Lin M, Bao G, Sang X, Wu Y. Recent advanced deep learning architectures for retinal fluid segmentation on optical coherence tomography images. *Sensors.* 2022;22:3055.
18. Chakravarthy U, et al. Automated identification of lesion activity in neovascular age-related macular degeneration. *Ophthalmology.* 2016;123:1731–6.
19. Chandrasekaran SN, Ceulemans H, Boyd JD, Carpenter AE. Image-based profiling for drug discovery: due for a machine-learning upgrade? *Nat Rev Drug Discov.* 2021;20:145–59.
20. Arcadu F, et al. Deep learning predicts OCT measures of diabetic macular thickening from color

- fundus photographs. *Invest Ophthalmol Vis Sci.* 2019;60:852.
21. Vamathevan J, et al. Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov.* 2019;18:463–77.
 22. Maunz A, et al. Accuracy of a machine-learning algorithm for detecting and classifying choroidal neovascularization on spectral-domain optical coherence tomography. *JPM.* 2021;11:524.
 23. Maunz A, et al. Machine learning to predict response to Ranibizumab in neovascular age-related macular degeneration. *Ophthalmology Science.* 2023;3:100319.
 24. Landers M, Dorsey R, Saria S. Digital endpoints: definition, benefits, and current barriers in accelerating development and adoption. *Digit Biomark.* 2021;5:216–23.
 25. Breaking the mold: How digital endpoints could transform trials. <https://www.clinicaltrialsarena.com/features/features-digital-endpoints/>.
 26. Lipsmeier F, et al. Reliability and validity of the Roche PD mobile application for remote monitoring of early Parkinson's disease. *Sci Rep.* 2022;12:12081.
 27. Mallon A-M, et al. Advancing data science in drug development through an innovative computational framework for data sharing and statistical analysis. *BMC Med Res Methodol.* 2021;21:250.
 28. Maunz A, et al. Automated segmentation of hyperreflective foci in diabetic macular Edema shows greater volume reduction by Faricimab vs Aflibercept in Phase 3 YOSEMITE and RHINE. *Invest Ophthalmol Vis Sci.* 2023;64:PB0039–PB0039.
 29. Vujosevic S, et al. Imaging retinal inflammatory biomarkers after intravitreal steroid and anti-VEGF treatment in diabetic macular oedema. *Acta Ophthalmol.* 2017;95:464–71.

Artificial Intelligence in the Diagnosis of Dry Eye

24

Mohammad Hassan Emamian, Roqayeh Aliyari,
Carla Lanca, and Andrzej Grzybowski

Abstract

Dry eye is a multifactorial disease, and its prevalence reaches 50% in some regions of the world. The disease is characterized by several signs and symptoms, and there are various diagnostic methods. However, diagnostic methods available have limitations with some tests having poor reliability and reproducibility. So far, there is no gold standard test for diagnosing dry eye disease (DED) as there is no linear association between

signs and symptoms, and often patients who have multiple signs of the disease report few symptoms or even have no symptoms. The above issues have led to difficulties in diagnosing DED. New imaging modalities, such as anterior segment optical coherence tomography (OCT), infrared meibography, in vivo confocal microscopy, tear interferometry, and non-invasive tear break-up time (TBUT), have emerged to allow objective measurements. However, the interpretation of results is based on subjective judgment. Artificial intelligence (AI) can help to solve these problems and contribute to decision making by interpreting the results objectively. During the past years and especially after 2014, several research studies on the use of AI in the diagnosis, classification, and monitoring of DED have been published, and the results seem promising. Machine learning and deep learning methods have shown high sensitivity, specificity, and accuracy. Implementation of AI has significantly increased the speed of diagnosing DED and its causes, however, it has not yet been integrated into clinical practice. AI will play an important role in diagnosing and management of dry eye soon. AI can also be used for analysis of big data, which may predict estimates of DED prevalence and its risk factors. This will be a big step to identify individual risk factors upscaling precision medicine in DED.

M. H. Emamian · R. Aliyari
Ophthalmic Epidemiology Research Center,
Shahrood University of Medical Sciences,
Shahrood, Iran
e-mail: emamian@shmu.ac.ir

R. Aliyari
e-mail: aliyari@shmu.ac.ir

C. Lanca
Escola Superior de Tecnologia da Saúde de Lisboa
(ESTeSL), Instituto Politécnico de Lisboa,
Lisboa, Portugal
e-mail: carla.rita.costa@gmail.com

Comprehensive Health Research Center (CHRC), Escola
Nacional de Saúde Pública, Universidade Nova de
Lisboa, Lisboa, Portugal

A. Grzybowski (✉)
Institute for Research in Ophthalmology, Foundation
for Ophthalmology Development, Poznan, Poland
e-mail: ae.grzybowski@gmail.com

Keywords

Dry eye disease · Artificial Intelligence · Machine learning · Diagnostic imaging modalities

Introduction

Dry eye disease (DED), keratoconjunctivitis sicca, or dysfunctional tear syndrome is a common and multifactorial eye disease, that affects hundreds of millions of people throughout the world. According to the last definition by the Tear Film & Ocular Surface Society in Dry Eye Workshop II, (TFOS DEWS II), “*Dry eye is a multifactorial disease of the ocular surface characterized by a loss of homeostasis of the tear film, and accompanied by ocular symptoms, in which tear film instability and hyperosmolarity, ocular surface inflammation and damage, and neurosensory abnormalities play etiological roles*” [1]. Aqueous deficient, and evaporative due to impaired function of the meibomian glands are the two primary forms of DED [1], although a mix of these forms have been reported [2]. The evaporative is the most common form of DED [1].

The common signs of DED include loss of homeostasis in the tear film, tear film instability and hyperosmolarity, ocular surface inflammation and damage, as well as neurosensory abnormalities [1]. Ocular discomfort and visual disturbance are the main symptoms of DED. Other non-specific symptoms include redness, burning, stinging, foreign body sensation, pruritus, photophobia, and blurred vision [2–4]. Although there are different subjective tests for the diagnosis of DED, perform diagnosis is still difficult and remains challenging [4]. The main causes for difficulties in the diagnosis of DED include a weak correlation between clinical signs and symptoms [1, 5, 6], the low reproducibility and the subjective nature of diagnostic tests. Most diagnostic tests for DED vary with season, time of day, and by examiners [7].

Therefore, new imaging-based techniques may help to overcome these issues. AI potentials uses are automatic detection, classification of DED, and investigation of the etiology, risk factors and biomarkers of DED.

Risk Factors and Prevalence

The prevalence of DED was reported between 5 and 50% in meta-analysis studies and reach to as high as 75% when its definition is based on DED signs [8]. The prevalence of DED is influenced by its definition, diagnostic measures, and the geographical regions, and in general increases with age. It should be noted that majority of research on DED have been focused on adults. Based on a systematic review and meta-analysis study with a total sample size of 1,488,935 subjects from 23 primary studies, the pooled prevalence of DED in Asian population was 20.1% [9]. A systematic review and meta-analysis study showed that the pooled prevalence of DED was 8.1% in the US [10]. Africa had the highest DED prevalence in the world with 47.9% [11]. In the UK the DED prevalence was 32.1% (95% CI 25.5–37.7%) by the TFOS DEWS II diagnostic criteria [12]. In the European population dry eye prevalence was 13.7% [11]. For signs, North America had the lowest regional prevalence (3.5%), and Eastern Asia with a prevalence of 42.8%, was the highest [11].

Most studies on the DED prevalence and incidence are in adult populations, and young people have been receiving less attention. Recent evidences indicate high prevalence of DED (57.1%) in Polish university students [13], which encourage more research and interventions in young adults and adolescents with widespread and growing use of smart phones and computers.

DED is more common in the female sex especially in adults and elderly [9, 11, 14]. However it was more common in European male than female sex when only DED signs were used for its diagnosis [11]. DED prevalence increases with increasing age and is more

common in Asian ethnicity [9, 14]. Some consistent risk factors for DED include meibomian gland dysfunction (MGD), connective tissue disease, Sjogren syndrome, androgen deficiency, computer use, contact lens wear, estrogen replacement therapy, pollution, low humidity, sick building syndrome, and using medications such as antihistamines, antidepressants, anxiolytics, and isotretinoin [14]. Diabetes, thyroid disease, psychiatric conditions, pterygium, and refractive surgery are among the probable risk factors of DED [14]. DED was more common and severe in people with depression [15]. A recent meta-analysis study has confirmed that dyslipidemia and using lipid lowering medications increase the odds of DED [16]. In another systematic review, 48 variables were evaluated as DED risk factors and seven of them (female gender, contact lenses, use of computers, thyroid abnormalities, hypertension, antidepressant and antihistamine) were identified as strong risk factors for DED [17].

Dry Eye Diagnosis

Many diagnostic tests and imaging methods are available for the diagnosis of ocular surface diseases and dry eye [18–20], which are briefly introduced below.

Tear Examination

Lipid, aqueous, and mucin are the three layers of normal tear. Any disturbances in quality or quantity of these layers may cause DED. There are many tests for qualitative and quantitative evaluation of tears, which are briefly described below [20].

Schirmer's Test

Schirmer's test has been used since 1903. A strip is inserted inside the lower eyelid and the amount of moisture content is measured on the

strip after 5 min. It may be used with or without local anesthesia. Although it is popular and simple to use, patients' discomfort and low sensitivity, specificity and repeatability restrict its usage.

Strip Meniscometry

In this test a thin and scaled strip (SMTube) is immersed into the tear meniscus of the lower eyelid for 5 s to absorb tears and the amount of tear meniscus volume is measured. This test which is easier than Schirmer's test has acceptable repeatability and reproducibility and specific correlations with tear meniscus height (TMH), TBUT, and Schirmer's test.

Tear Meniscus Volume, Height

Tear Meniscus volume (TMV) is positively associated with lacrimal secretory rate. TMV is reduced in DED. TMH has good accuracy to measure TMV. Both invasive and non-invasive methods may be used for measuring TMH. Optical coherence tomography (OCT) is a reliable non-invasive method for measuring TMH.

Tear Ferning Test

This test is an inexpensive and quick test for biochemical investigation of tear and can be used for diagnosis of DED, keratoconjunctivitis and cystic fibrosis. As a limitation, the crystal pattern of ferns depends on tear composition and the environmental conditions.

Tear Osmolarity

This method is one of the most accurate methods to diagnosis DED. TearLab (TearLab Corporation, Escondido, CA) and I-PEN (I-MED Pharma Inc., Dollard-des-Ormeaux, Canada) are two recent techniques to measure tear osmotic pressure.

Tear Inflammatory Factor

It is evident that inflammation has an important role in DED. InflammDry® and TearScan™ are two commercial options to measure inflammatory biomarkers in the tear film.

Tear Biomarkers and Proteomics

It has been shown that tear proteome such as lipocalin-1, lactoferrin, lysozyme, lacrimal proline-rich protein, and the prolactin-induced protein, decrease in DED. Alpha enolase is one of the best biomarkers to diagnose DED with an accuracy of 85%. Elevation of matrix metalloproteinase (MMP)-9 expression is one of the most potent biomarkers for DED [21, 22]. In a recent review study upregulation of apolipoprotein, haptoglobin, annexin 1, Glutathione S-transferase, and downregulation of lipocalin-1, prolactin inducible protein (PIP), lysozyme C, lactotransferrin, cystatin S, mammaglobin-b, and proline rich protein were associated with DED [23]. Although several potential biomarkers have been identified in DED, more research is needed to diagnose DED by proteomics and differentiate it from other ocular diseases [24, 25].

Tear Breakup Time (TBUT)

TBUT which is defined as the interval between the last blink and the first dry spot in the tear film. It is the most frequently used test for the diagnosis of DED with a sensitivity of 72% and specificity of 61%. In the invasive form of TBUT, the cornea is stained with liquid or a paper strip of fluoresceine and breakup time is measured under a cobalt blue filter.

Non-invasive Tear Film Break-Up Time (NITFBUT)

Despite the routine TBUT test, which requires fluorescein to be instilled into the

eye, fluorescein is not used in NITFBUT. In this method, tear film stability is measured by observing the images of the Placido disk over time, and analysis of inter-blink changes. Many corneal topography measurement systems or keratography modalities can be used for NITFBUT with higher accuracy and repeatability than TBUT. The sensitivity and specificity of NITFBUT for detecting DED were between 82–84% and 76–94% respectively [18].

Tear Film Thickness

Lipid layer thickness (LLT) is commonly used for the diagnosis of MGD. Interferometry is a recent method for direct measurement of LLT. LipiView II interferometer (TearScience Inc, Morrisville, NC) measures LLT, blink rate and blinking quality through the videos recorded.

Interferometry

Lipid layer thickness (LLT) can be measured quantitatively by interferometry on a nanometer scale. In this method images from the ocular surface are recorded using a CCT camera and after illumination with white light source. LLT then can be measured based on different optical interference colors of tear film. Interferometry can differentiate different types of MGD. It may also be used for measurement of tear meniscus height and then detection of DED [18].

Non-contact Meibography

In non-contact meibography, loss of glands or alterations in meibum composition can be differentiated from healthy meibum when illuminated with infrared light. The images can be captured by infrared Charge-coupled device (CCD) camera, equipped in slit lamp or mobile systems. Real-time videos can also be used for monitoring of meibomian glands and their functions [18, 26].

Eyelid Margin Examination

Blunt, irregular, and hyperemic eyelid margins may be seen in patients with MGD. Neovascularization, a shift of the mucocutaneous junction, telangiectasia, and plugging are among other, signs of MGD in eyelid margins. Lid wiper epitheliopathy which may be detected in eyelid margins after staining with fluorescein or lissamine green, is another sign of DED. Demodex mite infection, which is correlated with DED, and MGD may also be detected in examination of eyelid margins.

Ocular Surface Staining

Ocular surface cell staining can be used to evaluate integrity of epithelial cells. Staining is commonly done with Fluorescein sodium, lissamine green, or rose bengal dyes. Corneal epithelial cells are stained with fluorescein when they are defective. Therefore, this test is useful for diagnosing keratitis and corneal abrasions or erosion and it is not a specific test for diagnosis of DED. If the tear film does not provide sufficient protection, the ocular surface turns rose red after staining with rose bengal dye. The sensitivity and specificity of rose bengal staining in diagnosing dry eye is not as high as fluorescein staining. Staining with lissamine green is similar to rose bengal and because it has less cytotoxicity, it is used more frequent in eye clinics. Fluorescein and lissamine green may be used concurrently as a mixed dye.

Slit Lamp Biomicroscopy

Slit lamp examination is globally available in eye clinics with low costs. It may be used with corneal staining and for the measurement of TBUT. Despite its benefits, it is a totally subjective method, unless slit lamp photographs or videos are taken.

Impression Cytology

This method is based on using cellulose acetate filter paper to the ocular surface to collect superficial cells of ocular surface and then analysis of cells by histological, immunohistological, or molecular techniques. Impression cytology is a very useful in differential diagnosis of DED and other ocular surface disorders and neoplasms [27].

Anterior Segment Optical Coherence Tomography (AS-OCT)

Although expensive, AS-OCT is widely used in ophthalmology clinics for assessing ocular surface diseases. This non-invasive imaging modality can be used for diagnosis and classification of DED by measuring tear meniscus height, depth, and area [28, 29]. AS-OCT also may be use for differential diagnosis of DED from other ocular surface diseases [30]. The Swept-source OCT as a recent evolution of OCT, has also shown high repeatability in diagnosis of ocular surface diseases [31] and with the ability to measure tear meniscus volume [32]. AS-OCT has also been used for measurement of pre-corneal tear film thickness and its layers, recognition of lid-parallel conjunctival folds, and Meibomian gland imaging [18]; all of which are useful in diagnosing and management of DED.

In Vivo Confocal Microscopy (IVCM)

This non-invasive technique allows the evaluation of DED at cellular level. It may be used for DED-related ocular surface damage by identification of conjunctival goblet cells. IVCM is also a valuable tool in the diagnosis and management of MGD [18, 19].

Light Scatter

Loss of integrity of the tear film can lead to light scattering in DED. In DED with punctate keratopathy, backward light scattering can be detected by Scheimpflug imaging. The quantification of light scattering is done by the Objective Scatter Index (OSI), based on the opacification of media and the quality of the tear [26]. A significant association was observed between DED and ocular forward and backward light scattering [33].

Ocular Thermography

Thermography, which is a non-invasive modality can be done by thermographic camera operating within the infrared range. In DED, where the inflammation is present, ocular surface has higher temperature with more fluctuations and greater variance [18]. This method showed good results in the screening of DED using a handheld smartphone-based infrared thermal camera [34]. Glaucoma, diabetic retinopathy, age-related macular degeneration, retinal vascular occlusions, intraocular tumors as well as scleritis, and other inflammatory eye diseases have thermographic features [35].

Tear Film Imager

The tear film imager (TFI) can directly and dynamically measure tear physiology, mucous layer thickness and the lipid layer thickness lipid break-up time. An acceptable sensitivity (87%) and specificity (88%) were reported by this method for the diagnosis of DED [18].

Blink Pattern Analysis

Prolonged eyelid-closed time and shorter blink intervals were reported in DED. This pattern of blinking which is correlated with TBUT, can be detected by video camera recording and may be used for the diagnosis of DED [18, 36].

Wavefront Aberrometry

Wavefront aberrometry is commonly used for corneal disorders and refractive surgery. However, changes in tear film can also be detected with dynamic analysis of wavefront. When measured over approximately 30 s, higher-order aberrations can be seen in DED [18, 26].

Bulbar Redness Assessment

Bulbar redness (BR) is a non-specific ocular condition in several ocular conditions including DED, conjunctivitis, contact lens wearing, and adverse effects of glaucoma medications. Recently many computer-based image analysis techniques and AI-based automated image analysis were developed. The results are promising, however the non-specific nature of BR, limits its application in DED diagnosis [26].

Dry Eye Analyzer

There are new devices that can measure a variety of dry eye examinations such as TMH, NIBUT, TFT, infrared photography of the meibomian glands, and conjunctival hyperemia analysis, conveniently. LipiView II ocular surface interferometer, LipiScan dynamic meibomian imager, OCULUS Keratograph 5 M, and Kanghua dry eye analyzer are some examples of these state-of-the-art devices.

Questionnaires for Diagnosing Symptoms of DED

Symptom investigation is critical not only for diagnosis of DED but also to monitor the disease and response to treatment [19]. There are many patient-reported questionnaires for the diagnosis of DED. A recent study found that among 49 questionnaires, only 22 were validated [37]. Ocular Surface Disease Index (OSDI), Dry Eye Questionnaire (DEQ-5), Impact of Dry Eye on Everyday Living (IDEEL), National Eye Institute's Visual Function Questionnaire (NEI VFQ-25), Dry Eye-Related Quality-of-Life Score (DEQS), Computer-vision symptom scale (CVSS17), Symptoms Analysis in Dry Eye (SANDE),

Symptoms Assessment in Dry Eye (SANDE), Ocular Comfort Index (OCI), Instant Ocular Symptoms Survey (IOSS), University of North Carolina Dry Eye Management Scale (UNC DEMS), Chinese Version of Dry Eye-Related Quality of Life (CDERQOL), and Numerical Rating Scale (NRS), are common patient-reported questionnaires, used for diagnosis and monitoring of DED both in clinical practice and research [19, 38]. All the above questionnaires have limitations and different sensitivity, specificity, and precision in the diagnosis of DED. In another study, six of the twenty-four questionnaires (OSDI, IDEEL, DEQS, UNC DEMS, CDERQOL, and NEI VFQ-25) had questions on health-related quality of life and were evaluated properly for psychometric properties [39]. The probability of a true diagnosis of DED depends on DED prevalence and it has been shown that the best results are obtained by a combination of ocular and conjunctival staining with DEQ-5, TBUT, and osmolarity test [38]. The DEQ-5 and IOSS were recommended for young children because of the relative simplicity and repeatability of these tests [40].

Dry Eye Diagnosis by AI

Many studies on AI using machine learning methods to detect DED have been published and most of them were published after 2014. Storås et al. [41], investigated all published articles until May 21, 2021, and reported the current evidence about AI in DED. TBUT, interferometry and slit-lamp images, IVCN, meibography, tear osmolarity, proteomics analysis, OCT, population surveys, and other clinical tests were the input sources of data in machine learning models. However, most studies used interferometry, slit lamp, and meibography images. The authors noticed that deep learning techniques have increased significantly over the recent years and have replaced image pre-processing and feature extraction techniques. Although the performance results were promising, more work is still needed before using AI models in clinics

[41]. In other four review studies, Ji et al. [42], Yang et al. [43], Wu et al. [44], Zhang et al. [45] investigated AI in the diagnosis of ocular surface diseases. The authors reported high accuracy of AI methods in different studies for the diagnosis of MGD and concluded that AI technology can be used effectively and more efficiently to diagnose MGD and DED. AI analyses using bioinformatics databases and proteomics have been used widely in the diagnosis and treatment of DED [23] and MGD [46]. The results were promising with a sensitivity and specificity of approximately 90% [47]. Nevertheless, more research is required before widely application to clinical practice [46].

Table 24.1 shows a summary of studies using AI to diagnose DED. The results are promising, and in some studies, the accuracy of AI for the diagnosis DED was 90% [48] and even close to 100% [49].

Researchers at Bascom Palmer Eye Institute, University of Miami, Florida, USA, developed an automated tool for the detection of dry eye disease (DED) and Fuchs' endothelial corneal dystrophy (FECD) by using images from ultra-high resolution anterior segment optical coherence tomography (AS-OCT). AI and deep learning networks were used to identify features in the corneal epithelial layer and achieved a 96.9% sensitivity and 95.8% specificity to detect DES and 100% sensitivity and 97.8% specificity to detect FECD [50].

Recently a group of Japanese researchers used a smart eye camera to collect videos from the anterior segment of the eye. The authors found acceptable accuracy of tear film breakup time estimation (0.789) sensitivity (0.778) and specificity (0.857) of an AI model for the diagnosis of DED [51]. A fully automated system with high performance was developed to measure TMH, which is important in the diagnosis of DED [52]. Another study suggests that AS-OCT is interchangeable with K5M in subjective real-time grading of meibomian gland loss [53]. Numerous studies also find high accuracy and promising results for the diagnosis of MGD using AI [54–65].

Table 24.1 The recent studies on using AI for the diagnosis of dry eye

Publication year	First authors	Imaging modality	Sample size (eye)	Tests for diagnosis of DEED	AI algorithm	Diagnostic performance
2023	Wang [52]	Oculus camera	510 (Images)	Meniscus height measured by ophthalmologists	Neural network model based on UNet	IoU: 0.86 DSC: 0.92 (ACC: 0.96, Sen: 100%, Spe: 91.3%, F1-Score: 96.4%, Pre: 93.1%, for TMH<0.2 mm)
2023	Shimizu [51]	Videos of the anterior segment of the eye	158	TBUT OSDI	CNN	ACC: 0.789 F-score: 74.0% AUROC: 0.877
2023	Wan Cheng [66]	Images of tear meniscus height by Keratography 5 M	305 (images)	Meniscus height	DeepLab V3 ResNet50, and GoogleNet, and FCN	CC: 0.94 ICC: 0.90
2023	Kikukawa [67]	Tear film images by KOWA DR-1 α	350	TBUT	CNN (ResNet50 model)	ACC: 0.813 Sen: 84.3% Spe: 83.3% AUROC: 0.898
2023	Li [60]	Meibography for diagnosis of dry eye subtypes	280	TBUT, OSDI, Schirmer's test, corneal staining	SimCLR neural network	
2023	Abdelmotaal [48]	Video keratotomy	244	TBUT, OSDI, Schirmer's test, corneal staining	CNNs	AUROC: 0.99 ACC: 0.90 F1-score: 82%
2023	Qu [68]	Cornea images by photo slit-lamp	1046 (images)	Ophthalmologists	Fully convolutional networks (FCNs) U-Net, D-LinkNet, ResNet34	IoU: 0.937 ACC: 76.5% AUROC: 0.940 CC: 0.908
2023	Pan [69]	Keratography 5 M Functional visual acuity Objective scatter index	57	OSDI NIBUT	Unary linear regression, logarithmic regression	Sen: 0.950 Spe: 0.838
2023	Feng [70]	Fluorescein staining images by photo slit-lamp	382 (images)	Ophthalmologist examination	SVM, NB, DT, BT, KNN, RF	ACC: 82.7% AUROC: 96.6

(continued)

Table 24.1 (continued)

Publication year	First authors	Imaging modality	Sample size (eye)	Tests for diagnosis of DEED	AI algorithm	Diagnostic performance
2022	Edorth [71]	AS-OCT	228		Random forest and logistic regression	AUROC: 0.87 Sen: 86.4% Spe: 91.7%
2022	Zhang [72]	Blink Video	509	TBUT, Schirmer II test, corneal staining, lipid layer thickness	UNet RezNet	ACC: 96.3%
2022	Vyas [73]	TBUT video	1235 (images)	TBUT	CNNs AlexNet, VGG-16, GoogleNet inception V3, ResNet-50	ACC: 0.83
2022	Jing [74]	Keratography, confocal microscopy, and Pentacam for detection of corneal sub-basal nerve morphology and corneal intrinsic aberrations	269	OSDI	Deep learning CNS-Net	Significant CCs
2022	Zheng [36]	Blink videos by Keratography 5 M	100	OSDI, DEQ-5, Tear osmolality, NIBUT, TMH	Deep learning U-Net	IoU: 0.89, DSC: 0.93, BAC: 0.96, Sen: 0.99
2022	Fineide [49]	Clinical features Meibographic images	431 (subjects)	TBUT, OSDI	Decision trees Random forest Multilayer perceptron, ZeroR Adaptive boosting LogitBoost	ACC: 99.77% F-score and Precision>0.99 for the Random Forest technique
2021	Chase [75]	AS-OCT	151	TBUT, OSDI, Schirmer's test, corneal staining, conjunctival staining	CNN (VGG19)	ACC: 84.6% Sen: 86.4% Spe: 82.4%
2021	Dai [76]	Meibographic images for MG morphology	120 (subjects)	TBUT, OSDI, TMH, CFS, lid margin score, meibum expressibility	CNN U-Net	IoU: 0.908
2021	Deng [77]	Keratography 5 M for measuring tear meniscus height	217 (subjects)		CNN	IoU: 82.5% F-score: 90.1%
2021	Elsawy [78]	AS-OCT	879	Ophthalmologist examination	CNN (VGG19)	AUROC: 0.98 to 0.99 F1 Score: 86% to 90%

(continued)

Table 24.1 (continued)

Publication year	First authors	Imaging modality	Sample size (eye)	Tests for diagnosis of DED	AI algorithm	Diagnostic performance
2020	Su [79]	Fluorescent-stained images	101	Superficial punctate keratitis, diagnosed by an ophthalmologist	CNN	Sen: 0.94 Spe: 0.79 CC: 0.85
2020	Nam [80]	DED risk factors	4391	Ophthalmologist examination and clinical symptoms	DT, lasso, Logistic regression, Network graphs	AUROC: 0.70
2020	Yang [81]	Delphi method to determine DED risk factors	250	OSDI, NIBUT, Tear Osmolarity, Schirmer's test, CFS	ANN, Logistic regression	AUROC: 0.741 to 0.786 Sen: 81.9% to 83.1% Spe: 62.0% to 72.4%
2020	da Cruz [82, 83]	Interferometry	106 (images)		Random forest	ACC: 0.99 AUROC: 0.999 F-score: 99.6%
2019	Cartes [84]	Tear osmolarity	40	TBUT, OSDI, Schirmer's test, CFS	LR ML	ACC: 85%
2019	Yabusaki [85]	Interferometry	138 (images)	–	SVM (ML)	F-score = 98.1% for ADDE and 0.815 for EDE
2018	Szyperski [86]	Interferometry	110	–	Linear regression	AUROC: 78.6%
2016	Remeseiro [87, 88]	Tearscope plus	178	–	Web-based tear film classification and map	ACC: 0.961 Sen: 89.7% Spe: 98.0% F-score: 91.2%
2014	Kanellopoulos [89]	AS-OCT	140	–	Linear regression	AUROC: 0.908 to 0.932
2014	Gonzalez [90]	Tear electrophoresis for peptide/proteome profiles	93		PCA, NN	ACC: 89.3%

MG Meibomian gland, ADDE Aqueous deficient dry eye, EDE Evaporative dry eye, SVM Support vector machine, NB Naive Bayes, DT Decision tree, BT Boosting tree, KNN K-nearest neighbors, RF Random Forest, IoU Intersection over Union, DSC Dice similarity coefficient, CFS Corneal fluorescein staining, ACC Accuracy, Pre Precision, TMH Tear meniscus height, AUROC Area under the ROC curve, CC Correlation coefficient, BAC Balanced accuracy, ANN Artificial neural networks

AI Algorithms

Figure 24.1 shows a simple process of using AI in dry eye diagnosis. The two main AI algorithms include machine learning and deep learning models. In the machine learning algorithms, the input variables are selected by experts and then investigated in models such as Linear Regression, Logistic Regression, Support Vector Machine, Decision Tree, Naive Bayes algorithm, KNN algorithm, K-mean, and Random Forest. In deep learning algorithms, which are based on the concept of artificial neural network (ANN) different sources of data including images, videos, and sequential and tabular data may be used without manipulation by experts. Other common DL algorithms are Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs) [44]. Deep learning tasks include classification, object detection, and semantic segmentation. Popular algorithms are Visual Geometry Group (VGG),

ResNet, Inception, and InceptionResNet for classification; FasterRCNN, You Only Look Once (YOLO), and Single Shot Multibox Detector (SSD), for object detection; and Fully Convolutional Network (FCN), U-Net and Google’s DeepLab, for semantic segmentation [44].

The core of a CNN network is the convolutional layer, which accounts for the highest percentage of computations. A set of filters makes each convolution layer, and its output is made up of the convolution between the filters and the input layer. The pooling layer reduces the spatial size of feature maps resulting from the convolution layer. The most common examples of a pooling layer are max pooling and average pooling. The fully connected layer classifies images according to the extracted features of the previous layers (Fig. 24.2). VGG is a specific convolutional network designed for classification and localization and accepts 224 * 224-pixel images as input. Complex shapes and objects can be

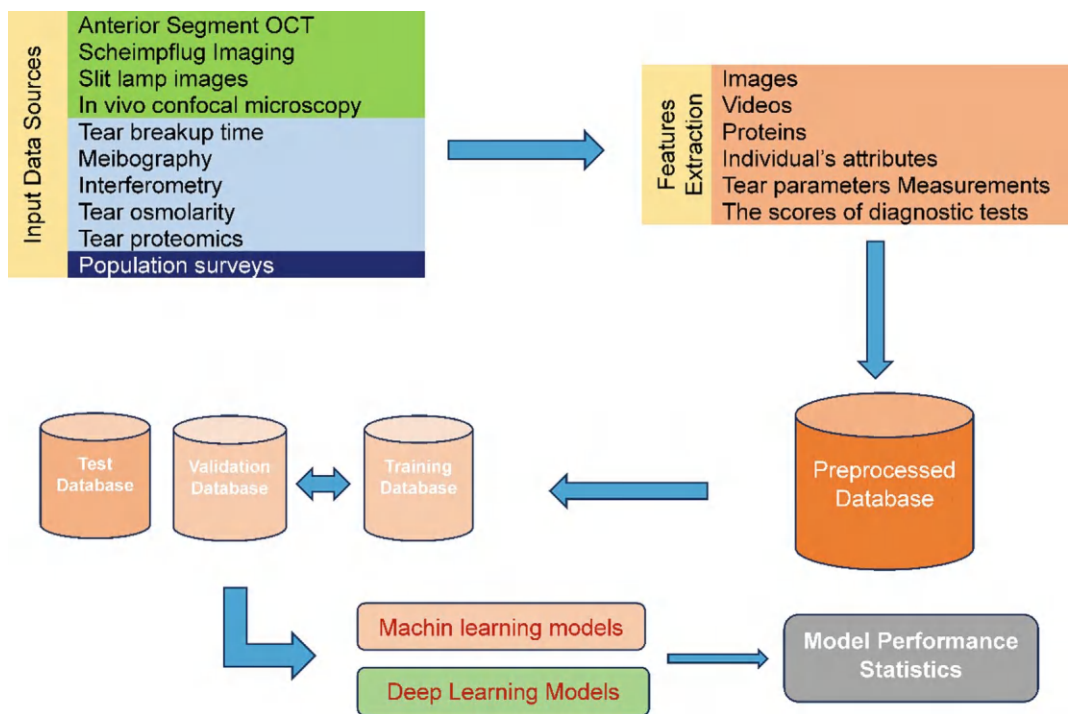


Fig. 24.1 The process of using AI in diagnosis of DED

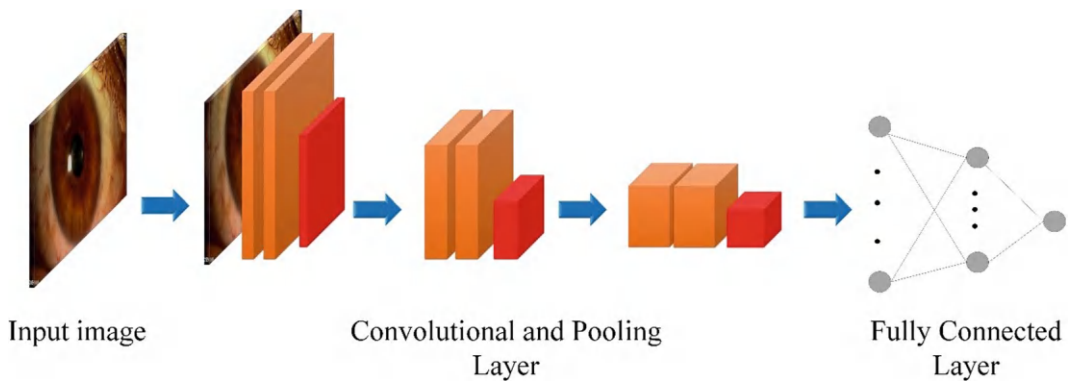


Fig. 24.2 Convolutional neural network architecture

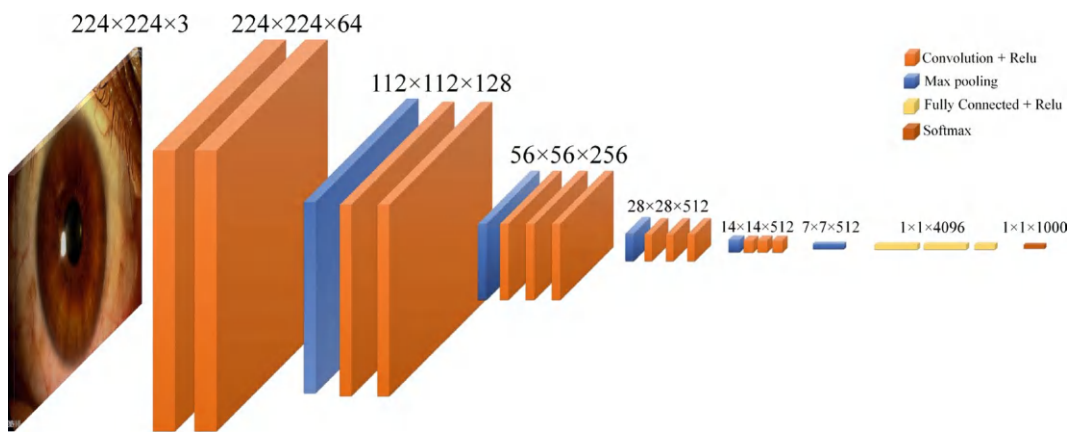


Fig. 24.3 Visual Geometry Group (VGG) architecture

diagnosed in VGG by large number of filters (Fig. 24.3). VGG16 and VGG19, are two common VGG, which are differ based on the number of layers.

Conclusion

The diagnosis of DED relies on various subjective tests and methods, and there is no gold standard test. The correlation between the symptoms reported by the patients and the symptoms of the disease in clinical examinations is weak. Despite various diagnostic tests, the above issues have caused difficulties in diagnosing

DED and classifying its severity. Objective methods based on AI can fill this gap.

In recent years, very promising and successful results have been obtained in the diagnosis of DED using AI. These methods are generally non-invasive, with high speed and acceptable accuracy, sensitivity, and specificity. However, AI-based methods still need development, and more work is needed to assess real-world performance, generalizability, and interpretability, data sharing, customization of AI algorithms, continual learning techniques, RCTs, and prospective studies [91].

It seems that the time has come for these techniques to be integrated into diagnostic

devices such as OCT and other imaging tools. Even home-based diagnosis of DED may be possible with new techniques. Scientific associations should also introduce AI-based methods as a standard for diagnosis and screening of DED, and these new methods should be used to classify, monitor DED and to evaluate its treatment process.

References

- Craig JP, Nichols KK, Akpek EK, Caffery B, Dua HS, Joo C-K, et al. TFOS DEWS II Definition and Classification Report. *Ocul Surf.* 2017;15(3):276–83.
- Messmer EM. The pathophysiology, diagnosis, and treatment of dry eye disease. *Dtsch Arztebl Int.* 2015;112(5):71–81; quiz 2.
- The definition and classification of dry eye disease: report of the Definition and Classification Subcommittee of the International Dry Eye WorkShop. *Ocul Surf.* 2007;5(2):75–92.
- Mohamed HB, Abd El-Hamid BN, Fathalla D, Fouad EA. Current trends in pharmaceutical treatment of dry eye disease: a review. *Eur J Pharm Sci.* 2022;175:106206.
- Kyei S, Dzasimatu SK, Asiedu K, Ayerakwah PA. Association between dry eye symptoms and signs. *J Curr Ophthalmol.* 2018;30(4):321–5.
- Tawfik A, Pistilli M, Maguire MG, Chen Y, Yu Y, Greiner JV, et al. Association of dry eye symptoms and signs in patients with dry eye disease. *Ophthalmic Epidemiol.* 2023;1–9.
- Zeev MS, Miller DD, Latkany R. Diagnosis of dry eye disease and emerging technologies. *Clin Ophthalmol.* 2014;8:581–90.
- Stapleton F, Alves M, Bunya VY, Jalbert I, Lekhanont K, Malet F, et al. TFOS DEWS II epidemiology report. *Ocul Surf.* 2017;15(3):334–65.
- Cai Y, Wei J, Zhou J, Zou W. Prevalence and incidence of dry eye disease in Asia: a systematic review and meta-analysis. *Ophthalmic Res.* 2022;65(6):647–58.
- McCann P, Abraham AG, Mukhopadhyay A, Panagiotopoulou K, Chen H, Rittiphairoj T, et al. Prevalence and incidence of dry eye and Meibomian gland dysfunction in the United States: a systematic review and meta-analysis. *JAMA Ophthalmol.* 2022;140(12):1181–92.
- Papas EB. The global prevalence of dry eye disease: a Bayesian view. *Ophthalmic Physiol Opt.* 2021;41(6):1254–66.
- Vidal-Rohr M, Craig JP, Davies LN, Wolffsohn JS. The epidemiology of dry eye disease in the UK: the Aston dry eye study. *Cont Lens Anterior Eye.* 2023;46(3):101837.
- Wróbel-Dudzińska D, Osiał N, Stępień PW, Gorecka A, Żarnowski T. Prevalence of dry eye symptoms and associated risk factors among university students in Poland. *Int J Environ Res Public Health.* 2023;20(2):1313.
- Craig JP, Nelson JD, Azar DT, Belmonte C, Bron AJ, Chauhan SK, et al. TFOS DEWS II report executive summary. *Ocul Surf.* 2017;15(4):802–12.
- Zhou Y, Murrough J, Yu Y, Roy N, Sayegh R, Asbell P, et al. Association between depression and severity of dry eye symptoms, signs, and inflammatory markers in the DREAM Study. *JAMA Ophthalmol.* 2022;140(4):392–9.
- Li Y, Xie L, Song W, Chen S, Cheng Y, Gao Y, et al. Association between dyslipidaemia and dry eye disease: a systematic review and meta-analysis. *BMJ Open.* 2023;13(11):e069283.
- Hasan ZA. Dry eye syndrome risk factors: a systemic review. *Saudi J Ophthalmol.* 2022;35(2):131–9.
- Schmidl D, Schlatter A, Chua J, Tan B, Garhöfer G, Schmetterer L. Novel Approaches for Imaging-Based Diagnosis of Ocular Surface Disease. *Diagnostics [Internet].* 2020; 10(8).
- Wolffsohn JS, Arita R, Chalmers R, Djalilian A, Dogru M, Dumbleton K, et al. TFOS DEWS II diagnostic methodology report. *Ocul Surf.* 2017;15(3):539–74.
- Wu Y, Wang C, Wang X, Mou Y, Yuan K, Huang X, et al. Advances in dry eye disease examination techniques. *Front Med.* 2022;8.
- Winiarczyk M, Biela K, Michalak K, Winiarczyk D, Mackiewicz J. Changes in Tear Proteomic Profile in Ocular Diseases. *Int J Environ Res Public Health.* 2022;19(20).
- Chester T, Garg SS, Johnston J, Ayers B, Gupta P. How can we best diagnose severity levels of dry eye disease: current perspectives. *Clin Ophthalmol.* 2023;17:1587–604.
- Pur DR, Krance SH, Pucchio A, Miranda RN, Felfeli T. Current uses of artificial intelligence in the analysis of biofluid markers involved in corneal and ocular surface diseases: a systematic review. *Eye (Lond).* 2023;37(10):2007–19.
- Ma JYW, Sze YH, Bian JF, Lam TC. Critical role of mass spectrometry proteomics in tear biomarker discovery for multifactorial ocular diseases (Review). *Int J Mol Med.* 2021;47(5).
- Jung JH, Ji YW, Hwang HS, Oh JW, Kim HC, Lee HK, et al. Proteomic analysis of human lacrimal and tear fluid in dry eye disease. *Sci Rep.* 2017;7(1):13363.
- Di Cello L, Pellegrini M, Vagge A, Borselli M, Ferro Desideri L, Scorgia V, et al. Advances in the noninvasive diagnosis of dry eye disease. *Appl Sci.* 2021;11(21):10384.
- Nichols KK, Mousavi M. Chapter 2—Clinical assessments of dry eye disease: tear film and ocular surface health. In: Galor A, editor. *Dry eye disease.* Elsevier; 2023. p. 15–23.

28. Wang SB, Cornish EE, Grigg JR, McCluskey PJ. Anterior segment optical coherence tomography and its clinical applications. *Clin Exp Optom*. 2019;102(3):195–207.
29. Lim SH. Clinical applications of anterior segment optical coherence tomography. *J Ophthalmol*. 2015;2015:605729.
30. Lamazales LL, Binotti W, Cox S, Balbuena-Pareja A, Hamrah P. Anterior segment-optical coherence tomography (AS-OCT) epithelial thickness mapping may provide a novel tool to differentiate neurotrophic keratopathy from dry eye disease. *Invest Ophthalmol Vis Sci* 2023;64(8):3604.
31. Recchioni A, Rauz S, Venkataraman AP, Dominguez-Vicent A. Swept-source anterior segment OCT in dry eye, ocular surface disease and healthy eyes: a repeatability study of quantitative measurements. *Invest Ophthalmol Vis Sci*. 2023;64(8):4020.
32. Gaudenzi D, Mori T, Crugliano S, Grasso A, Frontini C, Carducci A, et al. AS-OCT and ocular hygrometer as innovative tools in dry eye disease diagnosis. *Appl Sci*. 2022;12(3):1647.
33. Koh S, Maeda N, Ikeda C, Asonuma S, Mitamura H, Oie Y, et al. Ocular forward light scattering and corneal backward light scattering in patients with dry eye. *Invest Ophthalmol Vis Sci*. 2014;55(10):6601–6.
34. Zhang Q, Wu Y, Song Y, Qin G, Yang L, Talwar SS, et al. Screening evaporative dry eyes severity using an infrared image. *J Ophthalmol*. 2021;2021:8396503.
35. Gulias-Cañizo R, Rodríguez-Malagón ME, Botello-González L, Belden-Reyes V, Amparo F, Garza-Leon M. Applications of infrared thermography in ophthalmology. *Life*. 2023;13(3):723.
36. Zheng Q, Wang L, Wen H, Ren Y, Huang S, Bai F, et al. Impact of incomplete blinking analyzed using a deep learning model with the Keratograph 5M in dry eye disease. *Transl Vis Sci Technol*. 2022;11(3):38.
37. Sánchez-Brau M, Seguí-Crespo M, Cantó-Sancho N, Tauste A, Ramada JM. What are the dry eye questionnaires available in the scientific literature used for? A scoping review. *Am J Ophthalmol*. 2023;246:174–91.
38. Papas EB. Diagnosing dry-eye: Which tests are most accurate? *Cont Lens Anterior Eye*. 2023;46(5):102048.
39. Okumura Y, Inomata T, Iwata N, Sung J, Fujimoto K, Fujio K, et al. A review of dry eye questionnaires: measuring patient-reported outcomes and health-related quality of life. *Diagnostics (Basel)*. 2020;10(8).
40. Chidi-Egboka NC, Golebiowski B, Lee SY, Vi M, Jalbert I. Dry eye symptoms in children: can we reliably measure them? *Ophthalmic Physiol Opt*. 2021;41(1):105–15.
41. Storås AM, Strümke I, Riegler MA, Grauslund J, Hammer HL, Yazidi A, et al. Artificial intelligence in dry eye disease. *Ocul Surf*. 2022;23:74–86.
42. Ji Y, Liu S, Hong X, Lu Y, Wu X, Li K, et al. Advances in artificial intelligence applications for ocular surface diseases diagnosis. *Front Cell Dev Biol*. 2022;10:1107689.
43. Yang HK, Che SA, Hyon JY, Han SB. Integration of artificial intelligence into the approach for diagnosis and monitoring of dry eye disease. *Diagnostics (Basel)*. 2022;12(12).
44. Wu X, Liu L, Zhao L, Guo C, Li R, Wang T, et al. Application of artificial intelligence in anterior segment ophthalmic diseases: diversity and standardization. *Ann Transl Med*. 2020;8(11):714.
45. Zhang Z, Wang Y, Zhang H, Samusak A, Rao H, Xiao C, et al. Artificial intelligence-assisted diagnosis of ocular surface diseases. *Front Cell Dev Biol*. 2023;11:1133680.
46. Storås AM, Fineide F, Magnø M, Thiede B, Chen X, Strümke I, et al. Using machine learning model explanations to identify proteins related to severity of meibomian gland dysfunction. *Sci Rep*. 2023;13(1):22946.
47. Grus FH, Podust VN, Bruns K, Lackner K, Fu S, Dalmasso EA, et al. SELDI-TOF-MS ProteinChip array profiling of tears from patients with dry eye. *Invest Ophthalmol Vis Sci*. 2005;46(3):863–76.
48. Abdelmotaal H, Hazarbasanov R, Taneri S, Al-Timemy A, Lavric A, Takahashi H, et al. Detecting dry eye from ocular surface videos based on deep learning. *Ocul Surf*. 2023;28:90–8.
49. Fineide F, Storås AM, Chen X, Magnø MS, Yazidi A, Riegler MA, et al. Predicting an unstable tear film through artificial intelligence. *Sci Rep*. 2022;12(1):21416.
50. Euro Times. Diagnosing OSD with AI 2020. Available from: <https://www.esrcs.org/eurotimes/diagnosing-osd-with-ai>.
51. Shimizu E, Ishikawa T, Tanji M, Agata N, Nakayama S, Nakahara Y, et al. Artificial intelligence to estimate the tear film breakup time and diagnose dry eye disease. *Sci Rep*. 2023;13(1):5822.
52. Wang S, He X, He J, Li S, Chen Y, Xu C, et al. A fully automatic estimation of tear meniscus height using artificial intelligence. *Invest Ophthalmol Vis Sci*. 2023;64(13):7.
53. Edgar AK, Connor HRM, Kamarelddin S, Musich J, McLouta S, Choi E, et al. Anterior segment optical coherence tomography meibography compared with keratograph meibography. *Ophthalmic Physiol Opt*. 2024;44(1):115–23.
54. Zhang Z, Lin X, Yu X, Fu Y, Chen X, Yang W, et al. Meibomian gland density: an effective evaluation index of meibomian gland dysfunction based on deep learning and transfer learning. *J Clin Med*. 2022;11(9).
55. Wang J, Li S, Yeh TN, Chakraborty R, Graham AD, Yu SX, et al. Quantifying Meibomian gland morphology using artificial intelligence. *Optom Vis Sci*. 2021;98(9):1094–103.

56. Yu Y, Zhou Y, Tian M, Zhou Y, Tan Y, Wu L, et al. Automatic identification of Meibomian gland dysfunction with meibography images using deep learning. *Int Ophthalmol*. 2022;42(11):3275–84.
57. Saha RK, Chowdhury AMM, Na KS, Hwang GD, Eom Y, Kim J, et al. Automated quantification of Meibomian gland dropout in infrared meibography using deep learning. *Ocul Surf*. 2022;26:283–94.
58. Zhou YW, Yu Y, Zhou YB, Tan YJ, Wu LL, Xing YQ, et al. An advanced imaging method for measuring and assessing Meibomian glands based on deep learning. *Zhonghua Yan Ke Za Zhi*. 2020;56(10):774–9.
59. Swiderska K, Blackie CA, Maldonado-Codina C, Morgan PB, Read ML, Fergie M. A deep learning approach for meibomian gland appearance evaluation. *Ophthalmol Sci*. 2023;3(4):100334.
60. Li S, Wang Y, Yu C, Li Q, Chang P, Wang D, et al. Unsupervised learning based on meibography enables subtyping of dry eye disease and reveals ocular surface features. *Invest Ophthalmol Vis Sci*. 2023;64(13):43.
61. Wang Y, Shi F, Wei S, Li X. A deep learning model for evaluating Meibomian glands morphology from meibography. *J Clin Med*. 2023;12(3).
62. Wang J, Yeh TN, Chakraborty R, Yu SX, Lin MC. A deep learning approach for Meibomian gland atrophy evaluation in meibography images. *Transl Vis Sci Technol*. 2019;8(6):37.
63. Yeh CH, Yu SX, Lin MC. Meibography phenotyping and classification from unsupervised discriminative feature learning. *Transl Vis Sci Technol*. 2021;10(2):4.
64. Swiderska K, Read ML, Blackie CA, Maldonado-Codina C, Morgan PB. Latest developments in meibography: a review. *Ocul Surf*. 2022;25:119–28.
65. Setu MAK, Horstmann J, Schmidt S, Stern ME, Steven P. Deep learning-based automatic Meibomian gland segmentation and morphology assessment in infrared meibography. *Sci Rep*. 2021;11(1):7649.
66. Wan C, Hua R, Guo P, Lin P, Wang J, Yang W, et al. Measurement method of tear meniscus height based on deep learning. *Front Med*. 2023;10.
67. Kikukawa Y, Tanaka S, Kosugi T, Pflugfelder SC. Non-invasive and objective tear film breakup detection on interference color images using convolutional neural networks. *PLoS ONE*. 2023;18(3):e0282973.
68. Qu JH, Qin XR, Li CD, Peng RM, Xiao GG, Cheng J, et al. Fully automated grading system for the evaluation of punctate epithelial erosions using deep neural networks. *Br J Ophthalmol*. 2023;107(4):453–60.
69. Pan A-P, Ma Y, Hu R, Cao X, Wu Y, Zhou K, et al. Simultaneous real-time analysis of tear film optical quality dynamics and functional visual acuity in dry eye disease. *Eye Vis*. 2023;10(1):16.
70. Feng J, Ren Z-K, Wang K-N, Guo H, Hao Y-R, Shu Y-C, et al. An automated grading system based on topological features for the evaluation of corneal fluorescein staining in dry eye disease. *Diagnostics*. 2023;13(23):3533.
71. Edorh NA, El Maftouhi A, Djerada Z, Arndt C, Denoyer A. New model to better diagnose dry eye disease integrating OCT corneal epithelial mapping. *Br J Ophthalmol*. 2022;106(11):1488–95.
72. Zhang ZZ, Kuang RF, Wei ZY, Wang LY, Su GY, Ou ZH, et al. Detection of the spontaneous blinking pattern of dry eye patients using the machine learning method. *Zhonghua Yan Ke Za Zhi*. 2022;58(2):120–9.
73. Vyas AH, Mehta MA, Kotecha K, Pandya S, Alazab M, Gadekallu TR. Tear film breakup time-based dry eye disease detection using convolutional neural network. *Neural Comput Appl*. 2022.
74. Jing D, Liu Y, Chou Y, Jiang X, Ren X, Yang L, et al. Change patterns in the corneal sub-basal nerve and corneal aberrations in patients with dry eye disease: An artificial intelligence analysis. *Exp Eye Res*. 2022;215:108851.
75. Chase C, Elsawy A, Eleiwa T, Ozcan E, Tolba M, Abou SM. Comparison of autonomous AS-OCT deep learning algorithm and clinical dry eye tests in diagnosis of dry eye disease. *Clin Ophthalmol*. 2021;15:4281–9.
76. Dai Q, Liu X, Lin X, Fu Y, Chen C, Yu X, et al. A novel Meibomian gland morphology analytic system based on a convolutional neural network. *IEEE Access*. 2021;9:23083–94.
77. Deng X, Tian L, Liu Z, Zhou Y, Jie Y. A deep learning approach for the quantification of lower tear meniscus height. *Biomed Signal Process Control*. 2021;68:102655.
78. Elsawy A, Eleiwa T, Chase C, Ozcan E, Tolba M, Feuer W, et al. Multidisease deep learning neural network for the diagnosis of corneal diseases. *Am J Ophthalmol*. 2021;226:252–61.
79. Su TY, Ting PJ, Chang SW, Chen DY. Superficial punctate keratitis grading for dry eye screening using deep convolutional neural networks. *IEEE Sens J*. 2020;20(3):1672–8.
80. Nam SM, Peterson TA, Butte AJ, Seo KY, Han HW. Explanatory model of dry eye disease using health and nutrition examinations: machine learning and network-based factor analysis from a national survey. *JMIR Med Inform*. 2020;8(2):e16153.
81. Yang WJ, Wu L, Mei ZM, Xiang Y. The application of artificial neural networks and logistic regression in the evaluation of risk for dry eye after vitrectomy. *J Ophthalmol*. 2020;2020:1024926.
82. da Cruz LB, Souza JC, de Sousa JA, Santos AM, de Paiva AC, de Almeida JDS, et al. Interferometer eye image classification for dry eye categorization using phylogenetic diversity indexes for texture analysis. *Comput Methods Programs Biomed*. 2020;188:105269.
83. da Cruz LB, Souza JC, de Paiva AC, de Almeida JDS, Junior GB, Aires KRT, et al. Tear film

- classification in interferometry eye images using phylogenetic diversity indexes and Ripley's K function. *IEEE J Biomed Health Inform.* 2020;24(12):3491–8.
84. Cartes C, López D, Salinas D, Segovia C, Ahumada C, Pérez N, et al. El ojo seco está relacionado con un aumento intrasujeto de la variabilidad de osmolaridad lagrimal confirmado por tecnología de aprendizaje de máquinas. *Arch Soc Esp Oftalmol.* 2019;94(7):337–42.
 85. Yabusaki K, Arita R, Yamauchi T. Automated classification of dry eye type analyzing interference fringe color images of tear film using machine learning techniques. *Model Artif Intel Ophthalmol.* 2019;2(3):28–35.
 86. Szyperski PD. Comparative study on fractal analysis of interferometry images with application to tear film surface quality assessment. *Appl Opt.* 2018;57(16):4491–8.
 87. Remeseiro B, Mosquera A, Penedo MG. CASDES: a computer-aided system to support dry eye diagnosis based on tear film maps. *IEEE J Biomed Health Inform.* 2016;20(3):936–43.
 88. Remeseiro B, Barreira N, García-Resúa C, Lira M, Giráldez MJ, Yebra-Pimentel E, et al. IDEAS: A web-based system for dry eye assessment. *Comput Methods Programs Biomed.* 2016;130:186–97.
 89. Kanellopoulos AJ, Asimellis G. In vivo 3-dimensional corneal epithelial thickness mapping as an indicator of dry eye: preliminary clinical assessment. *Am J Ophthalmol.* 2014;157(1):63-8.e2.
 90. González N, Iloro I, Soria J, Duran JA, Santamaría A, Elortza F, et al. Human tear peptide/protein profiling study of ocular surface diseases by SPE-MALDI-TOF mass spectrometry analyses. *EuPA Open Proteom.* 2014;3:206–15.
 91. Li Z, Wang L, Wu X, Jiang J, Qiang W, Xie H, et al. Artificial intelligence in ophthalmology: the path to the real-world clinic. *Cell Reports Medicine.* 2023;4(7):101095.

Artificial Intelligence in Refractive Errors

25

Carla Lanca, Mohammad Hassan Emamian,
and Andrzej Grzybowski

Abstract

Uncorrected refractive errors (URE) are a leading cause of visual impairment (VI) and blindness. Big data acquisition and computer technology analysis are expanding quickly. By combining traditional techniques with artificial intelligence (AI) based tools several studies have shown great potential in the detection, prediction, and risk stratification of URE, especially myopia. Continued improvement of imaging techniques and methods to measure refraction coupled with the application of new AI algorithms may be the future

of clinical practice and standard of care for the diagnosis and prognosis of refractive errors. This chapter reviews how AI can be applied to the field of URE and discuss challenges to implementation into clinical practice and future directions.

Keywords

Artificial Intelligence (AI) · Uncorrected refractive errors · Visual Impairment (VI) · Myopia detection · Myopia prediction · AI-based risk stratification · AI algorithms in eye care

C. Lanca

Escola Superior de Tecnologia da Saúde de Lisboa (ESTeSL), Instituto Politécnico de Lisboa, Lisboa, Portugal

C. Lanca

Comprehensive Health Research Center (CHRC), Escola Nacional de Saúde Pública, Universidade Nova de Lisboa, Lisboa, Portugal

M. H. Emamian

Ophthalmic Epidemiology Research Center, Shahrour University of Medical Sciences, Shahrour, Iran

A. Grzybowski (✉)

Department of Ophthalmology, University of Warmia and Mazury, Olsztyn, Poland
e-mail: ae.grzybowski@gmail.com

Institute for Research in Ophthalmology, Foundation for Ophthalmology Development, Poznan, Poland

Uncorrected Refractive Errors

Uncorrected refractive errors (URE) are a leading cause of visual impairment (VI) and blindness, contributing to 42% of cases of VI in adults aged 50 years or over [1, 2]. In 2010, 6.8 million were blind and 101.2 million had VI due to URE [3]. VI prevalence in the working-age population continues to increase globally in developed regions, although there was a mild decrease in less-developed countries [4]. According to the World Health Organization (WHO) only about 36% of individuals with distance VI due to refractive error have received

access to an appropriate intervention [5]. The burden of URE can be effectively treated with spectacles to avoid VI. Early detection of URE in children is of paramount importance, as if left untreated URE can lead to amblyopia and permanent VI. A meta-analysis estimated that the global pooled prevalence amblyopia was 1.44% (95% confidence intervals [CI] 1.17–1.78%), with higher rates in Europe and North America, compared with Asia and Africa [6].

URE are associated with reduced vision related quality of life, which can be exacerbated in low-income communities [7]. Additionally higher burden of URE is associated with older age and female sex [8]. The global economic productivity loss associated with URE was estimated at 268.8 billion international dollars [9]. Thus, the development of measures to prevent and treat URE are imperative. According to Steinmetz et al. [2] more refractive services should be developed to provide affordable spectacles and better management of URE. Other strategies can be implemented to address the burden of URE. For example, the development of automated screening strategies can increase the detection at early stages with lower cost and higher cost–benefit for the populations. The member states of WHO have endorsed the 2030 global target to increase effective coverage of refractive error by 40% [5]. Artificial intelligence (AI) based approaches may help to accomplish this goal.

Among all refractive errors, myopia has deserved special attention from the eye care community. Myopia prevalence is highest in southeast Asian children. Alarming reports and projections of myopia worsening in United States of America (USA) and Europe have also been published. In 2050 in the USA 222 million are expected to be myopic and 48 million to have high myopia (-5.0 D or worse) [10]. According to the projections, it is expected that 43% of VI will be attributed to myopia. Additionally, loss of productivity from uncorrected myopia associated with VI in 2025 was estimated at US\$244 billion (95% CI, US\$49 billion–US\$697 billion) with higher burden in

East Asia [11]. Thus, it is important to design and implement screening strategies and offer adequate and timely treatment.

The aetiology of myopia has been extensively studied in the past decades. Myopia development and progression are the product of a complex interaction between genetic and environmental risk factors (Fig. 25.1) [12]. Risk factors may differ in regions with different ethnicity and lower prevalence of Myopia [13]. Associations between educational pressure and myopia have been consistently observed [14–16]. A recent study concluded that there is an intricate relationship between education and less time outdoors with half of the relationship between education and myopia being mediated by children not being outdoors during schooling [17]. Thus, risk factor analysis is important to develop and support interventions to control myopia.

The World Council of Optometry (WCO) adopted a resolution, an evidence-based standard of care for myopia management, that encourages optometrists to be more proactive in the management of myopia [18]. In a recent editorial Chen et al [19] proposed that such resolution also applies to other eye care professionals, such as ophthalmologists, and even other stakeholders to change the paradigm in the management of childhood myopia. Further guidelines to manage childhood myopia have been proposed by expert consent to identifying the risk profile based on child's history (Fig. 25.2) [20].

The consensus is based on the thought that the greater the number of risk factors, the higher will be the risk of progression. However, some of the factors may be difficult to collect in a clinical environment with good accuracy, such as outdoor time and near work. Those risk factors have traditionally been measured by questionnaires that are not standardized and are subject to recall and reporting biases. Past progression may also imply that some time will be needed to observe the child and action to control myopia will obviously be delayed by the period of observation. Many of the traditional measures used in clinical practice are based on

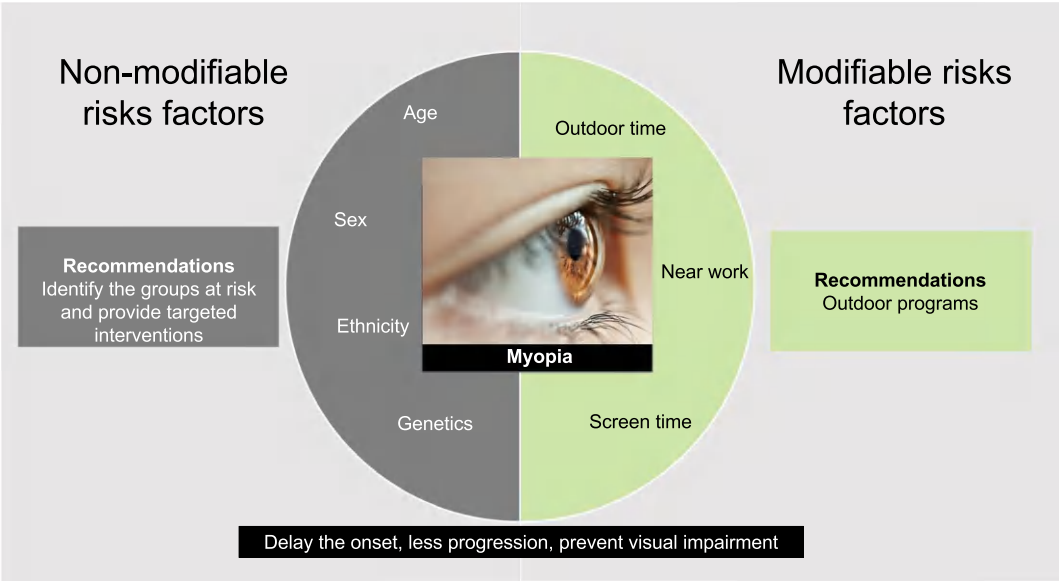


Fig. 25.1 Non-modifiable and modifiable risk factors for myopia

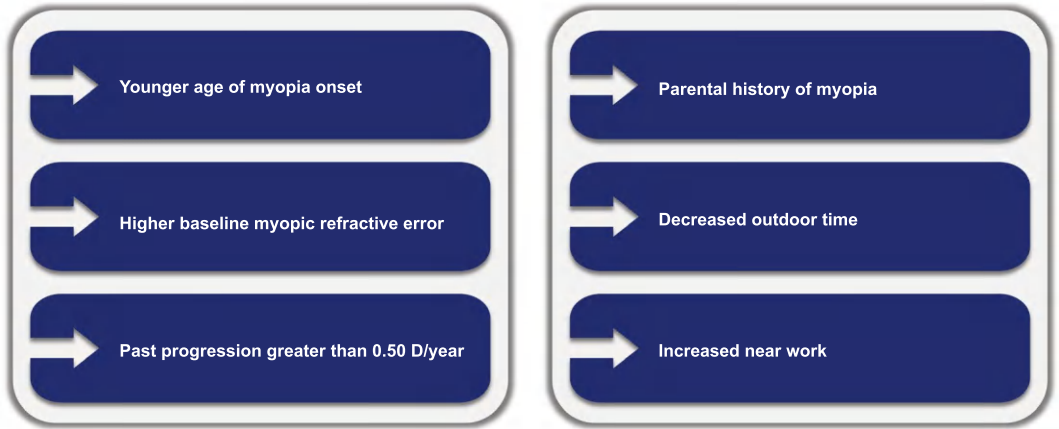


Fig. 25.2 Guidelines to manage childhood myopia based on expert consent

case-to-case clinicians’ decision and objective tools with high accuracy to aid clinicians may be further developed such as cloud-based tools with the use of AI to predict progression or risk of progression on children, helping to provide timely treatment.

Further challenges are emerging in myopia management, as some myopic adults continue to exhibit clinically significant degree of progression in adulthood (Fig. 25.3) [21]. Therefore,

myopia represents an increasing global health problem, affecting children and adults with significant public economic repercussions.

Further resolutions directed to URE are important. A three-step approach in the management of refractive errors is necessary, starting by a public health approach, advocating, promoting, and protecting the eye health of individuals and community (Fig. 25.4). Furthermore, as URE may go unnoticed by children and parents,

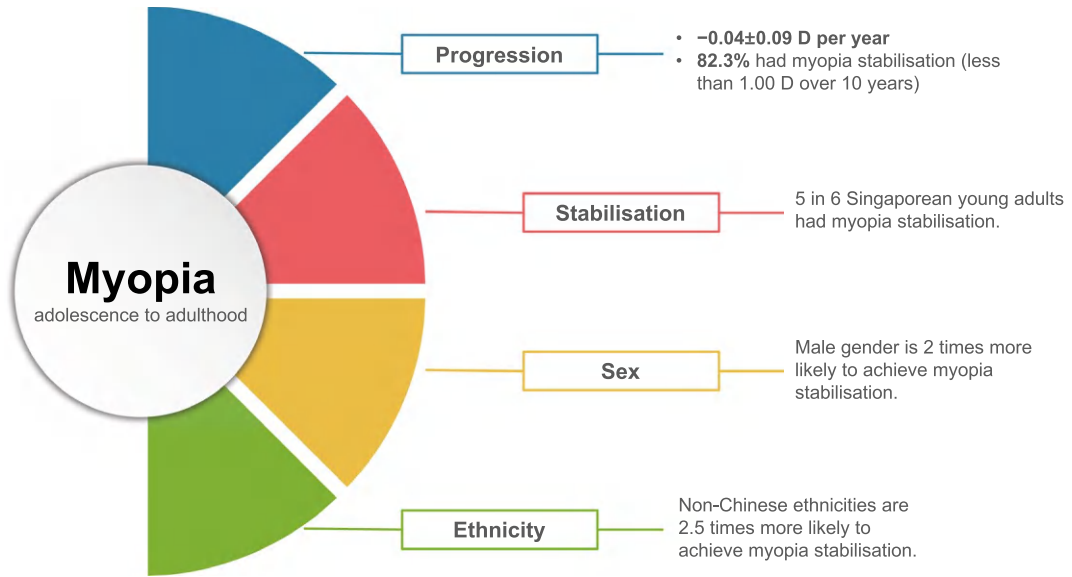


Fig. 25.3 Factors influencing myopia stabilisation in adulthood (adapted from Foo et al. [21])

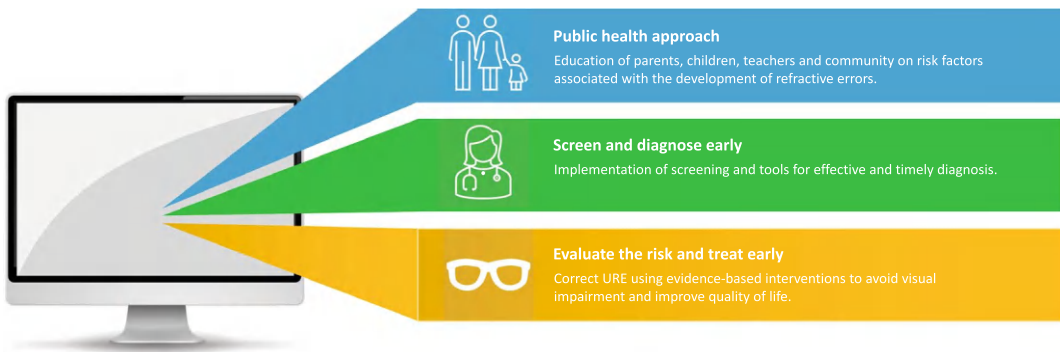


Fig. 25.4 A three-step approach in the management of refractive errors

screening is advisable to avoid delaying care. Strategies to increase public awareness of the URE burden are important for health policy making. The detection and diagnosis of URE is usually made by using traditional visual acuity and refraction measures. Those measures rely on experienced eye care professionals and are time consuming, especially in populations with communication disabilities or uncooperative children. In remote areas or regions with lack of eye care professionals, URE remain undetected, resulting in irreversible VI.

Applications of Artificial Intelligence

Applications of AI in the medical field have grown exponentially and ophthalmology is no exception. AI models to detect features of diabetic retinopathy, age-related macular degeneration, glaucoma, ocular surface diseases and disease-related VI have been developed [22–26]. More recently, in the past 5 years, the interest in AI algorithms for refractive error detection and follow-up, especially in the myopia field, is flourishing [27–29]. In the myopia control

arena, the use of new AI technologies may play an essential role to help in the prediction of the onset and progression of myopia in children.

Application of AI in the diagnosis and prognosis of refractive error may substantially reduce the time needed to observe the patient and be useful for clinical decision making, which reduces time and human effort. The use of AI models enables the analysis of large-scale datasets and the identification of patterns that precede URE development. In future with improved performance, AI based algorithms may be deployed into clinical practice, especially in the screening context, without requiring a trained operator. Accurate and easily accessible, open-source algorithms to detect URE, will enable clinicians and researchers to easily determine the risk of developing URE and determine treatment and follow-up much faster and with less subjectivity than before.

A further area of development are screening programmes where challenges emerge by the need of trained healthcare providers and economic costs. Additionally, screening programmes usually rely on non-invasive tests, such as uncorrected visual acuity tests with low sensitivity and high false positive referrals [30, 31]. A previous study found that the most accurate test for screening of significant refractive error was noncycloplegic retinoscopy (74%), followed by the Retinomax Autorefractor (66%), the SureSight Vision Screener (63%), and the Lea Symbols VA test (58%) [32]. Accurate AI models may be emergent tools with great potential in the screening of URE. The aims of AI in URE are summarized in Fig. 25.5.

Early and timely diagnosis is especially important in children. With the new emergent AI tools, the stratification of children by risk groups with higher risk profiles allows the identification of progression patterns (e.g. myopia progression) that are essential for the choice of the right treatment plan. Research on the application of deep learning to refractive error prediction is growing and several studies have been published on the use of imaging of the posterior segment of the eye (fundus photos and optical coherence

tomography, OCT), ocular appearance photos from smart-based phones and photorefracton [28, 29, 33–38]. Additionally, AI models have been also applied to traditional measures of refractive error and in treatment and follow-up of patients with myopia [27, 39].

The potential of AI based tools is tremendous, especially in highly populated countries with low numbers of eye care professionals or in regions with remote accessibility. Cost-effective AI screening may be a future standard of care to identify and correct refractive errors in areas with limited resource settings. Reliable assessment of risks factors for the development of predictions of refractive error is also another area that needs further improvement. Studies have been published on the use of percentile curves and nomograms to predict SE progression or axial length progression [40, 41]. However, the use of AI may be a promising tool, potentially to allow more accurate measures without the need to instil cycloplegic drops during the eye examination, avoiding side effects, diminishing time spent on examination and the need for additionally collection of biometrical measures, such as AL.

Prediction of Refractive Error with Imaging

Fundus photography and posterior segment OCT are widely spread across clinical practices around the world and are mainly used to diagnose and monitor retinal diseases. Previous studies have shown that fundus photography and OCT-based deep learning are useful tools to identify high myopia and vision-threatening conditions such as retinoschisis, macular hole, retinal detachment, and choroidal neovascularisation [35, 42, 43]. The results from those studies showed high accuracy of the models and with better performance than retinal specialists. As knowledge evolve, we should strive to find better tools to determine the risk of developing retinal changes and other early features before pathological changes emerge. One of the

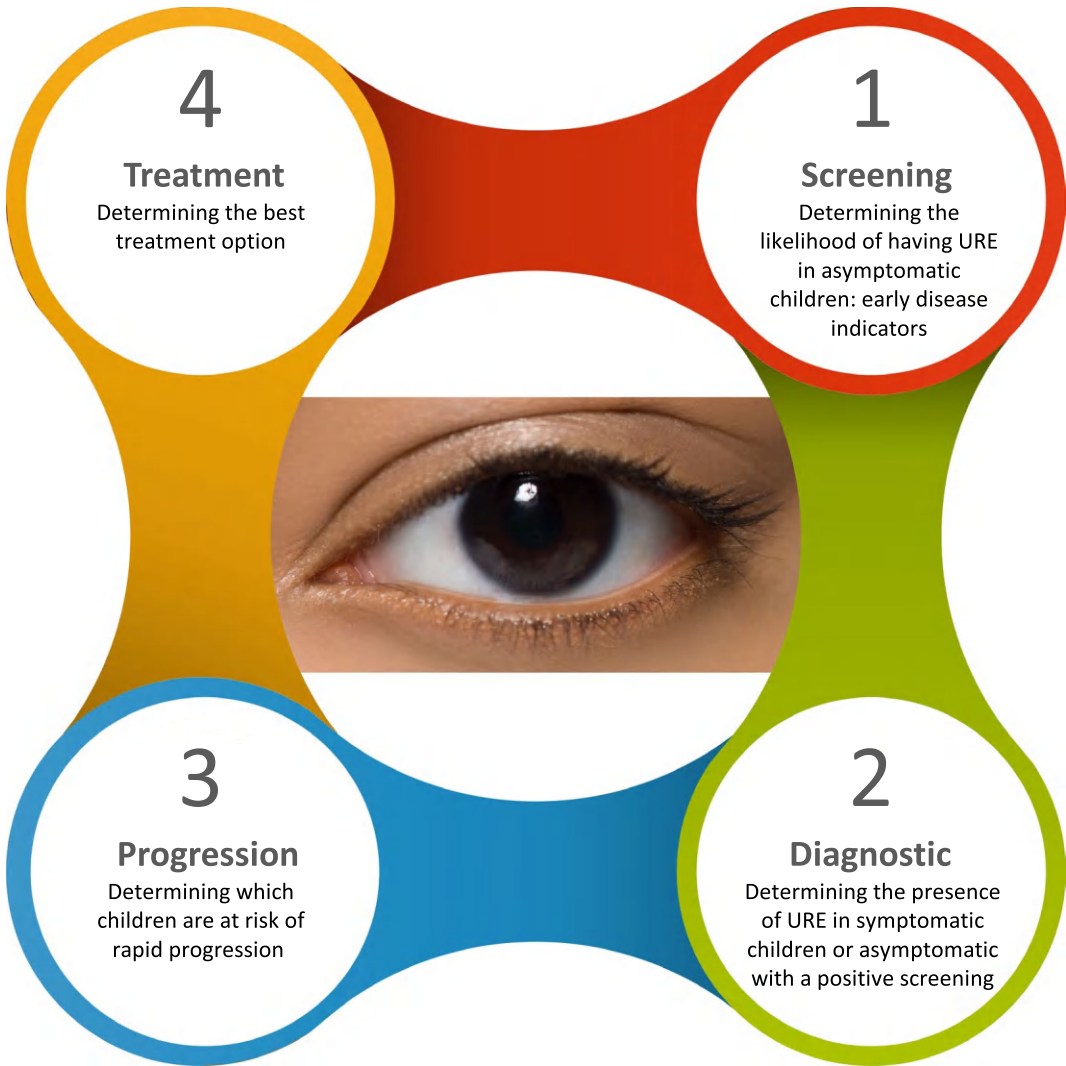


Fig. 25.5 Aims of artificial intelligence in uncorrected refractive errors. URE, uncorrected refractive errors

advantages of using imaging to identify URE would be to eliminate the need for cycloplegic refraction while maintaining good accuracy in the diagnosis. Deep learning algorithms to detect refractive error may help on those predictions and may in future be used as screening devices, reducing time and workload to improve patients' outcomes.

Deep learning is a subset of machine learning algorithm that derives from conventional neural network and employs graph technologies to build a multi-layer learning model [44].

It enables a machine to learn from raw datasets, such as retinal imaging, and classify features accordingly. An example of a deep learning model showing the training and knowledge transfer process to predict refractive errors is shown in Fig. 25.6.

So far, only a few studies have applied AI to fundus photographs for the prediction of refractive error mostly in adult populations (Table 25.1) [29, 33–36]. In a previous study conducted by a Google team, Varadarajan et al. [29] showed that a model trained to predict spherical

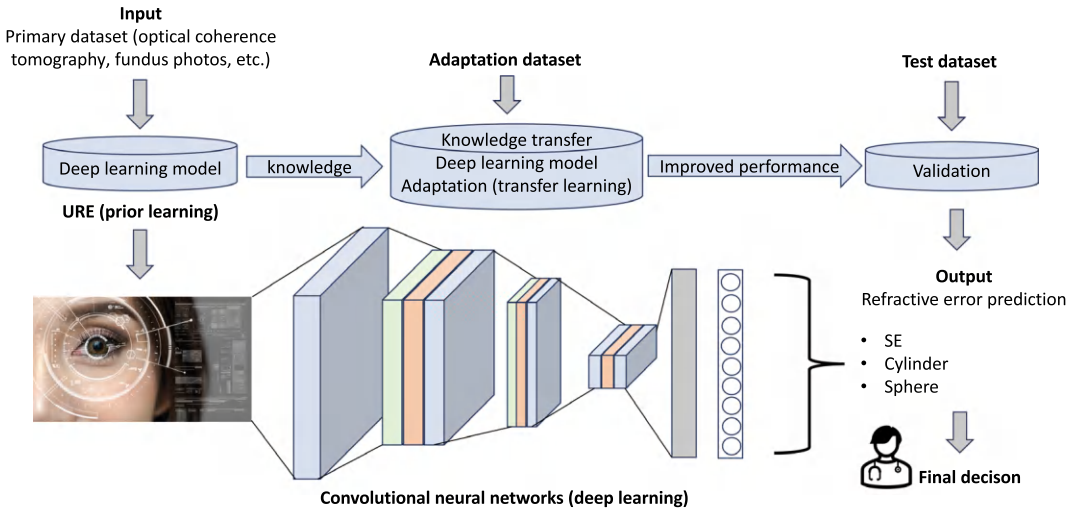


Fig. 25.6 Example of an imaging-deep learning model showing the training and knowledge transfer process to predict refractive errors. URE, uncorrected refractive errors; SE: Spherical equivalent

equivalent (SE) from retinal fundus images was accurate in the prediction of the spherical component (mean absolute error [MAE]=0.63 D, 95% CI, 0.63–0.64; $R^2=0.88$, 95% CI: 0.88, 0.89), but not the cylindrical component (Table 25.1). The accuracy of the model in both the UK biobank (MAE=0.56 D) and Age-Related Eye Disease Study (AREDS) datasets (MAE=0.91 D) [29] was higher than reported by Yoo et al. [38] (OCT-based deep learning; MAE=2.66 D), but lower than reported by Zou et al. [36] (retinal photographs; MAE=0.50 D for the fusion model-based deep learning system [FMDLS] and MAE=0.31 D for the cycloplegic refraction). Varadarajan et al. [29] commented that it may not be possible to predict astigmatism entirely by fundus photos as toricity of the cornea is an important factor involved in the astigmatism development. Nevertheless, Zou et al. [36] were able to identify the cylinder and axis and showed good agreement with cycloplegic refraction with vascular regions of the retina being highlighted by the AI model. Curiously, the with-the-rule astigmatism was identified by areas parallel to the retinal blood vessels, whereas against-the-rule astigmatism was identified on areas perpendicular to the vessels. Further studies maybe necessary to understand

the hypothesis behind the relationship between retinal features and predictability of astigmatism.

The authors of the google study have shown that attention maps highlighted the foveal area consistently, which may indicate that the role of the fovea in myopia development should be further explored. It is known that axial length (AL) drives axial ametropia [29]. Thus, retinal features used by the deep learning model in the google study may be primarily linked to axial ametropia and predictions may not be as accurate for lens ametropia. However, AL was not evaluated in the google study and this hypothesis needs further testing.

Tan et al. [35] used 226,686 retinal photographs from nine multi-ethnic cohorts and six regions to build a deep learning algorithm to predict high myopia and myopic macular degeneration (MMD). The model identified accurately eyes with high myopia by both SE and AL criteria and outperformed six human experts in the detection of high myopia and MMD (Table 25.1). Most notably the algorithm was able to identify 99% of choroidal neovascularisation cases in myopic eyes. The performance of those models to detect high myopia was higher (area under the receiver operating characteristic curve [AUC]=0.97) than the performance reported

Table 25.1 Studies on the prediction of refractive error using retinal photography and artificial intelligence-based algorithms

Study	City and country	Age (years)	Design	Sample size	Methods	Outcome measure	Performance of the best model
Varadarajan et al. 2018 [29]	UK and USA	40–80	UK Biobank is an ongoing observational study AREDS was a clinical trial in the United States	Training dataset n = 226,870 images Validation n = 24,007 UK Biobank and 15,750 AREDS images	Retinal photograph-based DLA (ResNet) Model used the “attention” method to identify features correlated with refractive error	SE Autorefractive in UK Biobank and subjective refraction in AREDS	MAE: UK Biobank = 0.56 D (95% CI, 0.55–0.56); AREDS = 0.91 D (95% CI, 0.89–0.93) Baseline expected MAE UK Biobank = 1.81 D (95% CI: 1.79–1.84); AREDS = 1.63 (95% CI: 1.60–1.67)
Zou et al. 2022 [36]	Tianjin, China	6–40	Population-based comparative study (retrospective dataset)	7873 images Training set: 2769 patients; Validation set: 791 patients; Test set: 394 patients	FMDLS via retinal fundus photographs compared with cycloplegic refraction (CNN + CNN-based transfer learning; XGBoost)	Refractive error (sphere and cylinder)	Sphere and cylinder MAE and r FMDLS = 0.50 D and 0.95 Cyclo-refraction = 0.31 D and 0.81 Axis Accuracy = 0.89; Specificity = 0.94 Sensitivity = 0.88; AUC = 0.81
Tan et al. 2021 [35]	Singapore, China, Taiwan, India, Russia and UK	Mean varied from 54.1 to 58.4	Retrospective multicohort	226,686 images MMD: 25,678 images High myopia: 225,671 images	Retinal photograph-based DLA (ResNet101) Comparison of DLA with 6 human experts	MMD and high myopia (SE and AL)	DLA outperformed the experts: AUC for MMD: 0.978 (95% CI, 0.957–0.994) AUC for high myopia: 0.973 (95% CI, 0.941–0.995)

(continued)

Table 25.1 (continued)

Study	City and country	Age (years)	Design	Sample size	Methods	Outcome measure	Performance of the best model
Shi et al. 2021 [34]	China	1–57	Hospital based and cross-sectional	Training set: 22,692 images; Validation set: 600 images; Test set: 600 images	Optos fundus images-based DLA (DenseNet and ResNet)	SE Measures: cycloplegic refraction	MAE: 1.12 D
Yang et al. 2022 [33]	Fudan, China	16–55	Hospital based and cross-sectional	Training dataset 987 patients External validation data set 133 images	UWF fundus images-based DLA (ResNet-50, Inception-v3, Inception-ResNet-v2)	SE Measures: subjective refraction	R ² : similar in 3 DLA (0.96) MAE: 1.72–1.76 D Deviation from the “true value”: – by 0.75 D (29.44% to 31.47%) – by: 2.00 D (65%)
Foo et al. 2023 [48]	Singapore	6–12	Retrospective population-based study	Train and validation 7456 images (1878 eyes) External validation 821 images (189 eyes)	Fundus images + clinical data (age, sex, race, parental myopia, and baseline cycloplegic SE) based DLA (DenseNet-121)	High myopia (SE ≤ −6.00 D) during teenage years (5 years later, age 11–17)	AUCs for the different models varied from 0.90 to 0.98 Mixed models (fundus image + baseline SE): AUC = 0.97
Dong et al. 2021 [49]	Beijing, China	≥ 50	Population-based, cross-sectional	6394 images 3468 participants	DLA based on colour fundus photographs and SFCT by OCT (CNN architecture as Inception-Resnet-v2)	AL and SFCT	MAE AL = 0.56 mm [95% CI: 0.53,0.61] SFCT = 49.20 μm (95% CI: 45.83,52.54) R ² AL = 0.59 (95% CI: 0.50,0.65) SFCT = 0.62 (95% CI: 0.57,0.67)

(continued)

Table 25.1 (continued)

Study	City and country	Age (years)	Design	Sample size	Methods	Outcome measure	Performance of the best model
Jeong et al. 2021 [50]	Anam, Korea	NA	Hospital based, retrospective	1296 images	DLA based on fundus images (ResNeT)	AL	MAE = 0.9 mm (95% CI: 0.85–0.91) $R^2 = 0.67$ (95% CI: 0.58–0.87) Accuracy = 90% within an error margin of ± 1 mm
Oh et al. 2023 [51]	Seoul, Korea	5–95	Hospital based, retrospective	8657 images 3829 patients	UWF fundus images-based DLA (transfer learning-based on EfficientNet B3)	AL	MAE and R^2 0.744 mm (95% CI, 0.709–0.779 mm) and 0.815 (95% CI, 0.785–0.840) Accuracy was 73.7% with error margins of ± 1.0 mm

AREDS Age-Related Eye Disease Study, *NA* Not available, *DLA* Deep learning algorithm, *SE* Spherical equivalent, *AL* Axial length, *OCT* Optical coherence tomography, *AUC* Area under the receiver operating characteristic curve, *CI* Confidence intervals, *MMD* Myopic macular degeneration, *UWF* Ultrawide field, *CNN* Convolutional neural network, *MAE* Mean absolute error, *ResNet* Residual network, *FMDLS* Fusion model-based deep learning system, *r* Correlation coefficient, R^2 R-squared, *SFCT* Subfoveal choroidal thickness

by Yoo et al. using OCT-based deep learning (AUC=0.81) and similar to the Jain et al. study, albeit a bit smaller (AUC=0.96) [38, 45]. Although the model was able to detect MMD with high accuracy, further identification of MMD features and categories is also an important area.

Further studies such as the Du et al. [46] were developed to understand if deep learning algorithms were able to identify MMD features on retinal photos. The authors found the model to have high sensitivity (S) and AUC to detect diffuse (S:84%; AUC: 0.97), patchy (S:87%; AUC: 0.98) and macular (S:85%; AUC: 0.98) atrophy. Similar results were found by Li et al. [43] using macular OCT images with AUCs of 0.961–0.999 and similar or higher performance of the AI model compared to retina specialists. Those results are promising as with further development it may be possible to implement screening for pathological myopia, decreasing the risk of visual impairment development.

Segmentation of myopia-induced lesions from fundus images using deep learning have also been implemented and analysed by Hemelings et al. [47]. By using convolutional neural networks the authors reported good performance of the model in the segmentation of lesions on the optic disc, retinal atrophy, and retinal detachment, which opens an avenue for differential diagnosis of glaucoma and high myopia [47].

More recently, Foo et al. [48] developed a deep learning system to predict the development of high myopia in teenagers in Singapore. Similarly, to the Tan et al. [35] study the authors of the Singaporean study showed that the model had high performance to identify which children are at risk of high myopia and with great potential as a tool to support clinical decisions.

Traditional fundus images (30 or 45° view) have great potential in predicting URE, but how about ultrawide field (UWF) fundus images (up to 200° view) that provide more information on peripheral areas of the retina? Yang et al. [33] used UWF fundus images to predict refractive error. Although the results show that it may be a feasible approach in future, the models

trained by UWF need further improvement in the accuracy before being applicable to clinical practice. Compared to the Varadarajan et al. [29] (MAE=0.56–0.91 D) and Shi et al. [34] (MAE=1.12 D) studies, the MAE was higher (MAE: 1.72–1.76 D). However, it was inferior to that reported by the Yoo et al. [38] study using OCT-based AI learning (MAE=2.66 D). Additionally, the authors of UWF study found that higher MAE was related with older age, which may be associated with the loss of the foveal reflection due to the natural process of aging [33]. Similar, to previous studies the macular area seems to be one of the most important anatomical areas for refractive error prediction.

Further studies have been also developed using deep learning models to predict AL. For example, Dong et al. [49] reported a MAE of 0.56 mm and a coefficient of determination (R^2) of 0.59 (95% CI, 0.50–0.65) by using fundus images. While Jeong et al. [50] found an R^2 of 0.67 and accuracy of 90%, within an error margin of ± 1 mm. Oh et al. [51] developed a deep learning model to predict AL using UWF fundus photography and reported a MAE of 0.74 mm and an R^2 value of 0.82 (95% CI, 0.785–0.840). The authors found that in some eyes with higher AL, peripheral regions were more highlighted than the macula or disc. However, the model showed lower accuracy (74% within an error margin of ± 1 mm), and further improvement of the model is necessary before application into clinical practice. Similarly, Foo et al. [48] were not able to achieve a clinical acceptable performance of AI models in the prediction of AL. A consistent feature of the models reported in previous studies, seems to be an overestimation of AL in shorter eyes and an underestimating in longer eyes. Based on the results of the studies so far, biometry is still the best measurement for AL and deep-learning algorithms are far from good accuracy to be applicable in clinics. Nevertheless, further studies should continue to explore and improve accuracy as portable fundus cameras associated with such algorithms may be used in future as cost-effective ways of measuring AL and other outcomes.

By using deep learning models, two previous studies have confirmed that there is a relationship between the shape of the optic nerve and refractive errors that can be identified by OCT based AI models (Table 25.2) [38, 45].

Yoo et al. [38] developed a retrospective study including healthy participants and validated their results by using OCT based deep learning models to estimate refractive errors. The authors found that the model yielded an AUC of 0.813 (95% CI, 0.744–0.881) and an accuracy of 71.4% (95% CI, 65.3–76.9%) to detect high myopia. The study had some limitations, such as lack of external dataset for validation and use of refractive error measures without cycloplegia. In another study, Jain et al. [45] built a deep transfer learning model with adaptation training to predict URE using posterior OCT images of the macula and the optic nerve. The images were used to train the algorithm in the classification of eyes with high myopia, moderate myopia, or hyperopia. The authors initially trained the algorithm on an East Asian dataset and validated the results on an Indian population. The model with the best performance to classify the refractive error achieved an accuracy of only 75%.

Interestingly, Jain et al. [45] generated saliency maps with the gradient-weighted class activation mapping (Grad-CAM) technique to visualize the areas on which the deep-learning model was focused on. Based on those results the authors highlighted the role of Bruch's membrane opening. Previous studies have suggested that a larger Bruch's membrane opening diameter is associated high myopia [52–54]. A previous study have also shown that the peripapillary retinal nerve fibre layer (RNFL) thickness patterns are different in myopic and hyperopic eyes and ocular magnification effects need to be taken in consideration when reading OCT measures [55]. Several limitations remain in the identification and grading of regions of interest from OCT images, especially in patients with pre-myopia or low myopia. Huang et al. [56] proposed an attention-based multi-scale feature fusion network for myopia grading to improve the performance of the models.

Recently, Wang et al. [57] proved that machine learning models can also predict long-term visual acuity from clinical information in highly myopic eyes (AUC of 0.870; 95% CI, 0.816–0.912). Those findings support the role of AI in assisting retinal specialists to monitor high risk patients and predict future poor visual acuity.

Yang, Y et al. [28] in an innovate approach used data from photography of ocular appearance derived from the Myopia AI program in China and applied facial recognition with deep convolution neural networks. The results of the study suggest that deep learning can detect myopia through subtle morphological features with high accuracy. Based on the saliency maps the authors were able to find that the deep learning model focused on eyes, especially the temporal sclera, outperforming ophthalmologists in the accuracy of the diagnosis.

Can imaging help to predict refractive error? It seems that AI models can be used and feed with images to estimate refractive error. Nevertheless, future studies are necessary to improve the accuracy of the models. Patients with poor image quality are often excluded from those studies. Thus, further tools are necessary to predict URE in children unable to perform OCT or fundus photo imaging. Cost-effectiveness should also be considered as performing autorefraction measures may be more cost-effective and of easier application. However, further studies are necessary to understand the best cost-effectiveness options to screen refractive error using deep learning models.

Deep Learning-Based Photorefraction Images

Eccentric photorefraction has been proposed as a method for screening due to its simplicity and speed, requiring less cooperation from children. Nevertheless, accuracy is a concern especially as ethnicity and individual-specific defocus calibration may be necessary [58, 59]. In the search for the development of low-cost and rapid screening

Table 25.2 Studies on the prediction of refractive error using optical coherence tomography and face photography and artificial intelligence-based algorithms

Study	City and Country	Age (years)	Design	Sample size	Methods	Outcome measure	Performance of the best model
<i>OCT</i>							
Yoo et al. 2021 [38]	Seoul, South Korea	21–50	Retrospective + prospective validation design	Development (n = 688 eyes) Test (n = 248 eyes)	OCT-based DLA (regression model based on a pretrained ResNet50 architecture)	SE	MAE = 2.66 D; $r^2 = 0.59$ ($P < 0.001$) AUC for high myopia 0.813 (95% CI, 0.744–0.881) Accuracy 71.4% (95% CI, 65.3–76.9%)
Jain et al. 2023 [45]	Seoul, South Korea and Maharashtra, India	> 18	Cross-sectional Multicentre	Pre-trained Korean dataset (n = 2602 eyes) Indian dataset - adaptation training data (n = 60 eyes) for transfer learning—test data (n = 142 eyes) for validation	DLA to learn the features of OCT (A pre-trained ResNet50 architecture was used as the feature extractor)	SE and mean keratometry	Myopia/hyperopia = model trained on circular optic disc OCT images (Accuracy = 74.7%; Cohen's kappa = 0.49; Matthew's correlation = 0.49) AUC for high myopia 0.956 AUC for hyperopia 0.869
<i>Ocular appearance images</i>							
Yang, Yahan et al. 2020 [28]	China	6–18	Data from the Myopia AI Program for deep learning + prospective clinical trial	2350 eyes (processed by 7050 pictures)	DLA to learn the features of ocular appearance images Comparison of DLA with ophthalmologists	Myopia (SE ≤ -0.5 D) Noncycloplegic refractive error measures	Validation test AUC = 0.927 (95% CI, 0.858–0.961) Clinical trial Sensitivity [DLA: 84.00% (95% CI, 73.50–94.50%) vs ophthalmologists: 64.00% (95% CI, 48.00–72.00%)] Specificity [DLA: 74.00% (95% CI, 61.40–86.60%) vs ophthalmologists: 53.33% (95% CI, 30.00–66.00%)]

AI Artificial intelligence, DLA Deep learning algorithm, SE Spherical equivalent, OCT Optical coherence tomography, AUC Area under the receiver operating characteristic curve, CI Confidence intervals, MMD Myopic macular degeneration, MAE Mean absolute error, ResNet Residual network

tools to detect URE, new studies have emerged trying to incorporate deep learning models into portable and smart-phone based devices to improve accuracy of photorefraction (Table 25.3).

Chun et al. [37] used a deep learning model based on eccentric photorefraction photographs and were able to identify different categories of refractive error with accuracies ranging from 75.0% for $\geq +5.0$ D to 83.3% for > -0.5 D and $< +0.5$ D. The overall accuracy of the model was 81.6%, surpassing the accuracy of 75% achieved by Linde et al. [60]. In the Linde et al. [60] study, the authors used pupillary red reflex images taken by a handheld fundus camera and evaluated the accuracy of a deep learning model to predict refractive error. The model accuracy was good in predicting the spherical component, but not excellent. The authors highlight some important limitations, such as the exclusion of several photographs due to low quality and age of participants equal or above 16 years, limiting the applicability of the results to younger children.

Xu et al. [61] applied AI into multiple eccentric photorefraction images (Table 25.3). The results of the study were better (MAE = 0.18 D to 0.07 D) than previous studies, and higher than the Varadarajan et al. [29] study that used retinal photography (MAE = 0.56–0.91 D). The authors used a combination of a convolutional neural network and a recurrent neural network and found higher accuracy of refractive error detection compared with Chun et al. [37] and Linde et al. [60] that also used eccentric photorefraction images. The generalization of the model may be limited by the sample size and uneven distribution of age and race. Thus, further studies are important to replicate those results before large-scale vision screening with those models are implemented.

Traditional Measures of Refractive Error and Its Progression

The prediction of refractive error and its progression may be challenging, particularly due to need to use cycloplegic refraction to measure the

refractive error in an accurate manner. In general, to diagnose URE, measures should be done with the gold standard measure, cycloplegic eye drops to avoid overestimations of myopia or underestimations of hyperopia [62, 63]. The use of cycloplegic drops is a time-consuming procedure than can take up to 1 h or 2 h with temporary side effects (up to 72 h), such as photophobia, blurred vision and glare perception. Nevertheless, non-cycloplegic refraction can induce errors of 1 D, especially in younger children [64].

Long-term population-based cohort studies often have large volumes of data that are difficult to manage with traditional database systems. The use of big data with machine learning models is becoming popular to manage those large volumes of data. Machine learning, a subfield of AI enables a machine to learn and improve from experience [65]. Lin et al. [66] applied big data and machine learning technology to predict the onset of high myopia in Chinese children (Table 25.4). The authors used real world clinical refraction data and found good performance AUC to predict the onset of high myopia by 18 years of age, over 3 years, and decreased performance over longer periods of time.

By using support vector machine Yang et al. [67] found very high accuracy (93%) of a model containing several input variables to predict myopia at grade 6 in students (Table 25.4). Although the results are encouraging, according to Foo et al. and Zhang et al. further research is necessary, such as interpopulation validation, before these models can be generalised and applied to clinical practice [68, 69]. In another study, Li et al. [70] used machine learning algorithms to determine risk factors for myopia progression in The Anyang Childhood Eye Study after 5 years. Uncorrected distance visual acuity and SE were the factors with higher weight in the prediction. While parental myopia had a decrease in weight as children got older.

Rampat et al. [27] used AI tools, such as novel polynomial decomposition basis, to predict subjective refraction from wavefront aberrometry data (an objective method for

Table 25.3 Studies using artificial intelligence-based algorithms and photorefractive images

Study	City and Country	Age (years)	Design	Sample size	Methods	Outcome measure	Performance of the best model
Chun et al. 2020 [37]	Seoul, Republic of Korea	6 months to 8 years	Cross-sectional	305 images	Smartphone eccentric photorefractive images according to spherical values as measured by cycloplegic refraction (CNN)	Refractive error (sphere)	DLA accuracy = 81.6% ≤ −5.0 D = 80.0% > −5.0 D and ≤ −3.0 D = 77.8% > −3.0 D and ≤ −0.5 D = 82.0% > −0.5 D and < +0.5 D = 83.3% ≥ +0.5 D and < +3.0 D 82.8% ≥ +3.0 D and < +5.0 D = 79.3% ≥ +5.0 D = 75.0%
Xu et al. 2022 [61]	Suzhou, China	18–56	Cross-sectional	Training set 3907 groups of images Validation set 1167 groups of images Test set 1000 groups of images	Multiple eccentric photorefractive images (CNN and RNN)	SE (sphere and cylinder)	MAE: Accuracy Sphere = 0.17 D; 89.5% Cylinder = 0.07 D; 96.7% SE = 0.18 D; 89.4% High myopia AUC 0.9942 [95% CI], 99.34–99.50%, accuracy of 97.13% (95% CI) 96.93–97.34%, specificity of 97.19% (95% CI) 96.95–97.43%, and sensitivity of 96.97% (95% CI) 96.63–97.31%
Linde et al. 2023 [60]	Choithram, India and Dargaville, New Zealand	≥ 16 and 20–60	Pooled analysis of two prospective cohort studies	Training set 288 IR and 528 colour images Validation set 143 IR images and colour images	IR and colour images of pupillary crescents were extracted from eccentric photorefractive images—smart-phone based (CNN)	Refractive error (sphere and cylinder)	Accuracy (infrared images and a multiclass classifier) Sphere = 75% Cylinder = 50%

IR Infrared, CNN Convolutional neural network, RNN Recurrent neural network, DLA Deep learning algorithm, SE Spherical equivalent, MAE Mean absolute error values, AUC Area under the receiver operating characteristic curve, CI Confidence intervals

Table 25.4 Studies on the prediction of refractive error and its progression using artificial intelligence-based algorithms

Study	City and Country	Age (years)	Design/data source	Sample size	Outcome measure (output)	Input variables	Best model	Performance of the best model
Lin et al. 2018 [66]	China	6–20	Real-world clinical refraction data from 8 ophthalmic centres	129,242 individuals	SE and onset of high myopia ($SE \leq -6.0$ D) up to 10 years in the future	Age, SE and annual progression rate	Random forest machine learning	High myopia development by 18 years of age, over 3 years: AUC ranged from 0.940 to 0.985
Yang, Xu et al. 2020 [67]	Henan province, China	Students from grade 1–6	School-based	3112 children	Myopia at grade 6	Family history, sex, indoor and outdoor activities, AL and K	Support vector machine	AUC = 0.98 Accuracy = 93% Sensitivity = 94% Specificity = 94%
Rampat et al. 2020 [27]	Paris, France	Mean $\pm$ SD 36.03 $\pm$ 11.24	Hospital-based and cross-sectional	Training set 3729 eyes Validation 350 eyes	Subjective refraction: power vectors (non-cycloplegic)	Wavefront aberrometry, LD/HD polynomial	Gradient boosted trees (XGBoost) algorithms	MAE M: 0.301 $\pm$ 0.252 D J0: 0.120 $\pm$ 0.094 D J45: 0.094 $\pm$ 0.084 D
Hernandez et al. 2022 [71]	India	15–70	Hospital-based	Training set 1244 samples Testing set 518 samples	Manifest refraction by wavefront aberrometry: power vectors (M, J0, and J45)	Age, sex, Zernike coefficients up to 5th order, and pupil related metrics provided by the autorefractor	Random forest regressor, gradient boosting regressor, extreme gradient boosting regressor and ASB (average of the 3 models)	ASB performed better MAE decreased by M: 42.5% (0.27 D) J0: 29.4% (0.06 D) J45: 41.7% (0.05 D) Accuracy (reduction in the 95% limits of agreement) M: 0.63 D J0: 0.14D J45: $\pm$ 0.08 D
Tang, Tao et al. 2020 [74]	Peking, China	6–18	Retrospective study and cross-sectional datasets	1011 myopic children	AL and physiological elongation of AL	Age, sex, CCT, cycloplegic SE, K-mean and WTW	Robust linear regression model	AL: $R^2 = 0.87$; RMSE = 0.38 Estimated physiological AL elongation: 0.003–0.116 mm/year

(continued)

Table 25.4 (continued)

Study	City and Country	Age (years)	Design/data source	Sample size	Outcome measure (output)	Input variables	Best model	Performance of the best model
Li et al. 2022 [70]	Beijing, China	6–9	School-based cohort study	2740 children	Myopia progression after 5 years (cycloplegic autorefraction)	AL, near work, time outdoors, living and nutritional habits, wearing spectacles, and parental myopia	Random forest algorithm	Accuracy of 80% for an absolute error of 0.50 D Weight (mean, range) UDVA = 28%, 22–39% SE = 20%, 7–28% AL = 13%, 10–14%
Barraza-Bernal et al. 2023 [72]	Wuhan, China	5–16	Community-based cross-sectional data	Cross-sectional (n = 12,780) Longitudinal (n = 226) Validation (n = 81)	Myopia onset	Age, sex and sphere	Support vector regression + Gaussian process regression	$R^2 = 0.77$; RMSE = 0.97 D Bias = −0.05 D Limits of agreement 0.85 D (95% CI, −0.91–0.80 D)
Zhu et al. 2023 [73]	Huzhou, China	6–9	Retrospective	179 children Training sets 125 Test set 54	SE and AL at grade 6	Age, sex, grade 1–6, SE and AL	OMP and MLP for SE, KR, MLP and ET	SE: R^2 Grade 5: 0.8997 Grade 4: 0.7839 Grade 3: 0.7177 AL: R^2 Grade 5: 0.7546 Grade 4: 0.5456 Grade 3: 0.8755

SE Spherical equivalent, AL Axial length, CCT Central corneal thickness, K keratometry, WTW White-to-white corneal diameter, R Coefficient of determination, CI Confidence Intervals, RMSE Root mean squared error, LD/HD Low degree/high degree, MLP Multilayer perceptron algorithm, ET Extra Tree algorithm, OMP Orthogonal matching pursuit, KR Kernel ridge, R^2 Coefficient of determination, UDVA Uncorrected distance visual acuity, ABS Custom assembly model, MAE Mean absolute error

measuring refraction). The results showed good accuracy and precision with great potential in future to clinical practice aiding in the prediction of refraction in an objective manner (Table 25.4). Limitations of the study include the use of non-cycloplegic refraction. Similarly, Hernandez et al. [71] assessed the performance of machine learning models to predict subjective refraction from wavefront aberrometry data and demographic factors. MAE values were higher than the ones achieved by the AI models in the Rampat et al. [27] study. The results of both studies support the use of machine learning approaches as additions to the software to improve accuracy of autorefractors, although further studies are necessary to validate the results.

In a previous study cross-sectional population based data of 12,780 Chinese children and longitudinal data of 226 children were used to develop a machine learning-based algorithm to assess the risk of myopia onset (Table 25.4) [72]. The authors found that support vector regression and Gaussian process regression resulted in the best performing algorithm with a Pearson correlation coefficient between prediction and measured data of 0.77 (bias of -0.05 D) and the limits of agreement of 0.85 D (95% CI: -0.91 – 0.80 D). The choice of final algorithm was also based on a parsimonious choice of the number of variables to input, resulting in the inclusion of 3 variables (age, sphere, and sex). Although, the bias obtained from the study was small and probably not clinically significant, the limits of agreement are still high and further research is necessary to improve the algorithm. Apart from refractive error, AL has emerged as an important end point of clinical trials. Thus, published studies have also tried to predict AL as an outcome. Zhu et al. [73] used machine learning models to predict SE and AL at grade 6 (Table 25.4). The authors found that data from single years does not predict well future SE and AL for greater span predictions. The performance of the models varied, with the

orthogonal matching pursuit performing better for SE prediction and the Kernel ridge and the multilayer perceptron algorithm performing better for AL predictions.

New studies are emerging on the hypothesis that abnormal growth due to myopia should be differentiated from physiological elongation. Tang et al. [74] conducted a study using a machine learning-based model to estimate physiological elongation of AL in a sample of Chinese school-aged myopic children. The authors found that the robust linear regression model had a good performance in the predicting of physiological AL elongation (Table 25.4). However, the sample size was limited and more data on hyperopic and emmetropic children is necessary. Further studies should be implemented to improve the algorithm. Nevertheless, the algorithm shows promising results and in future may be used to follow-up children under orthokeratology treatment or other myopia control treatments.

New studies are emerging on the prediction of SE without cycloplegia in school-aged children using machine learning algorithms. By using big data acquisition and machine learning techniques, Du et al. [75] have shown that AI models have a promising role in predicting the difference before and after cycloplegia (R^2 : 0.889–0.927; MSE 0.250–0.380; MAE 0.372–0.436). Nevertheless, there is still room for improvement as the AUC to detect a change >0.50 D was 0.822, with a specificity of 76.2% and a sensitivity of 74.4%. Further studies are necessary to increase applicability to clinical practice.

The current challenges of developing an algorithm for the prediction of SE and SE progression, include lack of appropriate data for training and validation and the necessity of using the less possible number of inputs with an acceptable diagnostic performance. However, SE progression depends on multiple parameters such as parental myopia and environmental factors, such as time spent outdoors and near-work.

Deep Learning Assisted Technology for Treatment and Follow-Up

Most studies of applications of AI in URE have been developed to assist eye care providers in disease screening, diagnosis, and prognosis. Recent studies are also focusing on potential applications of AI for treatment, especially for myopic patients. The available myopia control therapies include low dose atropine, orthokeratology, anti-myopic glasses and contact lenses, red-light therapy, and prevention with environmental intervention [76–78].

New studies with AI tools are emerging in the orthokeratology field to help to fit the lenses, to identify the treatment zone and to evaluate the performance of the treatment [39, 79]. The selection of the best orthokeratology lenses is characterized by challenges such as the choice of the return zone depth and landing zone angle. A previous retrospective study developed a new machine learning algorithm to aid in the prescription of orthokeratology lenses in Chinese adolescents with myopia [79]. The authors found that R^2 values varied from 0.69 to 0.97. Similarly, Fan et al. [39] found that machine learning models provide efficient methods to estimate the alignment curve and in orthokeratology lens fitting for vision shaping treatment (R^2 varied from 0.73 to 0.91). Misplacement of the treatment zone leads to poor outcomes, such as declines in visual acuity, diplopia, and glare. Automated deep learning algorithms and corneal topography maps were used by two previous studies to identify this zone. Tang et al. [80] found an accuracy intersection over union of 0.90 ± 0.06 (range, 0.60–0.97). Similarly, Zhang et al. [81] developed a deep learning algorithm to determine the treatment zone and peripheral steepened zone following orthokeratology treatment. The authors used corneal topography maps and a deep neural network providing an automatic tool (to replace manual annotation) with good accuracy to determine the two treatment zones. Wu et al. [82] analysed the role of 19 factors on intraocular pressure (IOP) in myopic children treated with topical atropine.

The authors found that XGBoost was the best predictive model and baseline IOP was the most accurate predictive factor for final IOP. Although, the results of three studies showed good accuracy, further studies are necessary to validate the results.

Further avenues of AI can be explored as it would be highly valuable to have a tool to predict individual responses to different myopia control modalities as well as criteria for success. We anticipate that future models will be able to use traditional measures of refraction, AL, and retinal features to predict response to treatment and improve treatment approaches to reduce the risk of high myopia development.

Zhang et al. [69] proposed an AI-integrated telemedicine platform for the management of myopia. Firstly, the suggested workflow includes screening of myopia with portable devices coupled with AI technology. At a second stage risk stratifying based on AI tools and clinical data, images and genomics would be achieved with the aim of developing personalized treatment plans and environmental counselling (e.g. wearable devices with objective measures). In a third stage threatening conditions such as MMD would be monitored with imaging at home or in community-based centres, triaging patients with pathological myopia to tertiary centres. In Fig. 25.7, we proposed a 3-step approach to URE that may be implemented in future for screening, detection, follow-up, and treatment of URE and associated complications.

Challenges to Implementation into Clinical Practice

Limitations of current prevalence, risk factors, and screening studies are related with the need for cycloplegia. Thus, if AI-based methods could replace the present cycloplegia-related methods, they could revolutionize future studies and improve our understanding of refractive errors and their risk factors in all regions of the world. Machine learning methods seem to be

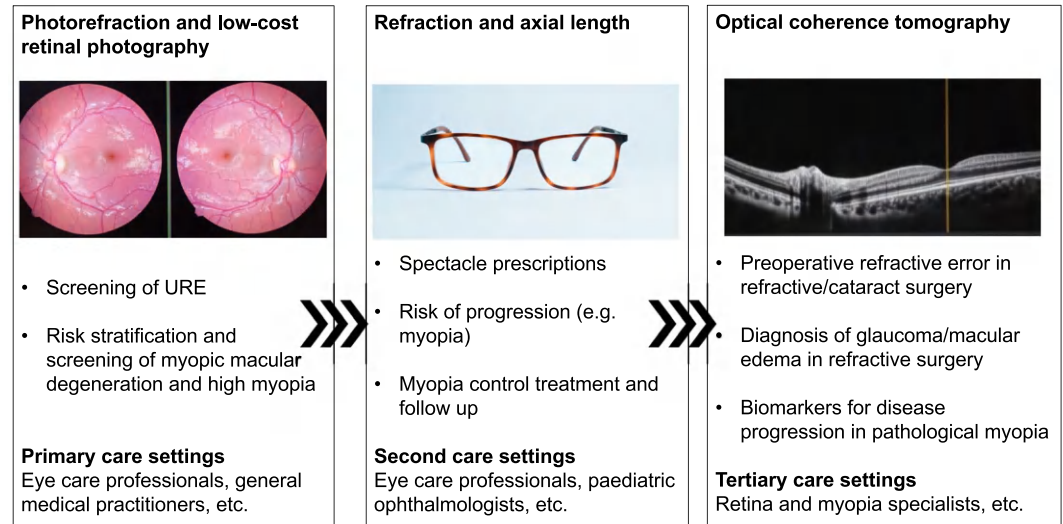


Fig. 25.7 Examples of future potential applications of AI to primary, second and tertiary care settings. URE, uncorrected refractive errors

efficient and accurate tools that in future can be implemented into clinical practice replacing some of the traditional methods. Currently, there are several challenges to implementation of AI technology into clinical practice for the management of refractive errors. Although, the accuracy of AI models was reported to be high in several research and clinical studies, which is one of the most important steps that supports successful introduction in clinical practice, other aspects of implementation and effective integration into the clinical workflow should be considered to achieve higher cost–benefit and better outcomes for the patients. For example, the choice of input variables for the models is important as myopia clinics around the world are overloaded with patients and the number of input variables should be clinical accessible and easy to collect. Real-world deployment may lead to decreases in model accuracy and strategies to overcome those limitations should be implemented. For example, cloud-based systems seem to improve the review process in the screening of diabetic retinopathy [83].

Additional limitations are similar to predictions of other health outcomes, such as mismatches between training and testing datasets that can induce decrease in performance of deep

learning models, especially in multi-ethnic scenarios. There are racial and ethnic differences in URE. Thus, further studies applying deep learning with populations of different races are necessary. Diagnostic performance of deep learning algorithms based on imaging decreases with low image resolution.

According to Zhang and Zou [84] more studies are necessary to compare the results due to differences in data and testing sets. Thus, it is necessary to establish a multicentre dataset and conduct several studies with updated algorithms to establish a benchmark for URE. For example, a previous study in Singapore have provided a proof of concept of a blockchain enabled AI platform for use in deep learning studies [35]. The China Alliance of Research in High Myopia (CHARM) is a new consortium that plans to involve over 100,000 Chinese patients to identify and quantify biomarkers of high myopia using AI [85]. The Myopia Risk Calculator Consortium (MRC) is employing classical machine learning methodologies with cutting-edge AI innovations to detect the onset myopia and risk of myopia progression. The MRC Consortium is strategically expanding its investigative purview to include transfer learning methodologies, seeking to augment the

precision and adaptability of our models to a variety of population datasets, which will entail the generation of specialized synthetic data. In future, the MRC will work towards data sharing and improvement of the diagnostic performance of the current models to deploy the best models for clinical practice.

Conclusions

Several AI based models have shown to be effective in detecting URE, providing information on spectacle prescription, risk of pathological myopia, identification of myopic lesions and follow-up data in myopia control treatment. Current gold standard measures are being challenged by the new AI methods that seem to be less time-consuming and, in some cases, superior to retinal specialists. Nevertheless, these new technologies will need further refinement and limitations need to be properly addressed with more research before introduction into daily clinical practice. Continuous exploration to further refine AI technologies is necessary to improve its efficacy.

New areas of further development where AI can be helpful are the understanding of the natural history of the URE, especially myopia, and determine the most efficient and sensitive biomarkers for disease progression in pathological myopia to use as outcome measures in future treatment trials. AI may also be used to collect better and valid data on outdoor activity, near work time and other children's activities. Additionally new avenues can be explored by using deep learning and computer vision technology to large-scale myopia screening using photos and regular cameras, such as smartphones.

References

1. Pascolini D, Mariotti SP. Global estimates of visual impairment: 2010. *Br J Ophthalmol*. 2011;96(5):614–8. <https://doi.org/10.1136/bjophthalmol-2011-300539>.
2. Steinmetz JD, Bourne RRA, Briant PS, et al. Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study. *Lancet Glob Health*. 2021;9(2):e144–60. [https://doi.org/10.1016/S2214-109X\(20\)30489-7](https://doi.org/10.1016/S2214-109X(20)30489-7).
3. Naidoo KS, Leasher J, Bourne RR, et al. Global vision impairment and blindness due to uncorrected refractive error, 1990–2010. *Optom Vis Sci*. 2016;93(3):227–34. <https://doi.org/10.1097/OPX.0000000000000796>.
4. Chen J, Yang C, Zheng W, et al. Global, regional, and national epidemiology of visual impairment in working-age individuals, 1990–2019. *JAMA Ophthalmol*. Published online December 7, 2023. <https://doi.org/10.1001/jamaophthalmol.2023.5617>.
5. World Health Organization. Report of the 2030 targets on effective coverage of eye care; 2022. <http://apps.who.int/>.
6. Fu Z, Hong H, Su Z, Lou B, Pan CW, Liu H. Global prevalence of amblyopia and disease burden projections through 2040: a systematic review and meta-analysis. *Br J Ophthalmol*. 2020;104(8):1164–70. <https://doi.org/10.1136/bjophthalmol-2019-314759>.
7. Killeen OJ, Niziol LM, Elam AR, et al. Visual impairment from uncorrected refractive error among participants in a novel program to improve eye care access among low-income adults in Michigan. *Ophthalmology*. Published online September 2023. <https://doi.org/10.1016/j.ophtha.2023.09.025>.
8. Lou L, Yao C, Jin Y, Perez V, Ye J. Global patterns in health burden of uncorrected refractive error. *Invest Ophthalmol Vis Sci*. 2016;57(14):6271. <https://doi.org/10.1167/iovs.16-20242>.
9. Smith TST, Frick KD, Holden BA, Fricke TR, Naidoo KS. Potential lost productivity resulting from the global burden of uncorrected refractive error. *Bull World Health Organ*. 2009;87(6):431–7. <https://doi.org/10.2471/blt.08.055673>.
10. Bullimore MA, Brennan NA. The underestimated role of myopia in uncorrectable visual impairment in the United States. *Sci Rep*. 2023;13(1):15283. <https://doi.org/10.1038/s41598-023-42108-y>.
11. Naidoo KS, Fricke TR, Frick KD, et al. Potential lost productivity resulting from the global burden of Myopia. *Ophthalmology*. 2019;126(3):338–46. <https://doi.org/10.1016/j.ophtha.2018.10.029>.
12. Yu M, Hu Y, Han M, et al. Global risk factor analysis of myopia onset in children: a systematic review and meta-analysis. *PLoS ONE*. 2023;18(9):e0291470. <https://doi.org/10.1371/journal.pone.0291470>.
13. Hashemi H, Khabazkhoob M, Fayaz M, et al. Refractive errors and their associated factors in school-children: a structural equation modeling. *Ophthalmic Epidemiol*. Published online March 25, 2022. p. 1–11. <https://doi.org/10.1080/09286586.2022.2048397>.
14. Plotnikov D, Williams C, Atan D, et al. Effect of education on myopia: evidence from the United Kingdom ROSLA 1972 Reform. *Invest Ophthalmol Vis Sci*. 2020;61(11):7. <https://doi.org/10.1167/iovs.61.11.7>.

15. Mountjoy E, Davies NM, Plotnikov D, et al. Education and myopia: assessing the direction of causality by mendelian randomisation. *BMJ*. 2018;361:k2022. <https://doi.org/10.1136/bmj.k2022>.
16. Nickels S, Hopf S, Pfeiffer N, Schuster AK. Myopia is associated with education: Results from NHANES 1999–2008. Tsai DC, ed. *PLoS One*. 2019;14(1):e0211196. <https://doi.org/10.1371/journal.pone.0211196>.
17. Clark R, Kneepkens SCM, Plotnikov D, et al. Time spent outdoors partly accounts for the effect of education on Myopia. *Invest Ophthalmol Vis Sci*. 2023;64(14):38. <https://doi.org/10.1167/iovs.64.14.38>.
18. World Council of Optometry. World Council of Optometry Resolution: The Standard of Care For Myopia Management by Optometrists.; 2021. Accessed 7 Dec 2023. <https://worldcouncilofoptometry.info/resolution-the-standard-of-care-for-myopia-management-by-optometrists/>.
19. Chen KS, Au Eong JTW, Au Eong KG. Changing paradigm in the management of childhood Myopia. *Eye*. Published online November 15, 2023. <https://doi.org/10.1038/s41433-023-02831-2>
20. Saxena R, Dhiman R, Gupta V, et al. Prevention and management of childhood progressive myopia: National consensus guidelines. *Indian J Ophthalmol*. 2023;71(7):2873–81. https://doi.org/10.4103/IJO.IJO_387_23.
21. Foo LL, Tan CS, Noel B, et al. Factors influencing myopia stabilisation in young myopic adult Singaporeans. *Br J Ophthalmol*. Published online August 27, 2023:bjo-2023-323680. <https://doi.org/10.1136/bjo-2023-323680>.
22. Ting DSW, Cheung CYL, Lim G, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA*. 2017;318(22):2211. <https://doi.org/10.1001/jama.2017.18152>.
23. Burlina PM, Joshi N, Pekala M, Pacheco KD, Freund DE, Bressler NM. Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks. *JAMA Ophthalmol*. 2017;135(11):1170. <https://doi.org/10.1001/jamaophthalmol.2017.3782>.
24. Asaoka R, Murata H, Iwase A, Araie M. Detecting preperimetric glaucoma with standard automated perimetry using a deep learning classifier. *Ophthalmology*. 2016;123(9):1974–80. <https://doi.org/10.1016/j.ophtha.2016.05.029>.
25. Fang X, Deshmukh M, Chee ML, et al. Deep learning algorithms for automatic detection of pterygium using anterior segment photographs from slit-lamp and hand-held cameras. *Br J Ophthalmol*. 2022;106(12):1642–7. <https://doi.org/10.1136/bjophthalmol-2021-318866>.
26. Tham YC, Anees A, Zhang L, et al. Referral for disease-related visual impairment using retinal photograph-based deep learning: a proof-of-concept, model development study. *Lancet Digit Health*. 2021;3(1):e29–40. [https://doi.org/10.1016/S2589-7500\(20\)30271-5](https://doi.org/10.1016/S2589-7500(20)30271-5).
27. Rampat R, Debellemanière G, Malet J, Gatineau D. Using artificial intelligence and novel polynomials to predict subjective refraction. *Sci Rep*. 2020;10(1):8565. <https://doi.org/10.1038/s41598-020-65417-y>.
28. Yang Y, Li R, Lin D, et al. Automatic identification of myopia based on ocular appearance images using deep learning. *Ann Transl Med*. 2020;8(11):705–705. <https://doi.org/10.21037/atm.2019.12.39>.
29. Varadarajan AV, Poplin R, Blumer K, et al. Deep learning for predicting refractive error from retinal fundus images. *Invest Ophthalmol Vis Sci*. 2018;59(7):2861. <https://doi.org/10.1167/iovs.18-23887>.
30. Ma Y, He X, Zou H, Lu L, Qu X, Zhu J. Myopia screening: combining visual acuity and non-cycloplegic autorefraction. *Optom Vis Sci*. 2013;90(12):1479–85. <https://doi.org/10.1097/OPX.0000000000000095>.
31. Langeslag-Smith MA, Vandal AC, Briane V, Thompson B, Anstice NS. Preschool children's vision screening in New Zealand: a retrospective evaluation of referral accuracy. *BMJ Open*. 2015;5(11):e009207. <https://doi.org/10.1136/bmjopen-2015-009207>.
32. Ying GS, Kulp MT, Maguire M, et al. Sensitivity of screening tests for detecting vision in preschoolers-targeted vision disorders when specificity is 94%. *Optom Vis Sci*. 2005;82(5):432–8. <https://doi.org/10.1097/01.opx.0000162660.14378.30>.
33. Yang D, Li M, Li W, et al. Prediction of refractive error based on ultrawide field images with deep learning models in Myopia patients. *Front Med (Lausanne)*. 2022;9. <https://doi.org/10.3389/fmed.2022.834281>.
34. Shi Z, Wang T, Huang Z, Xie F, Song G. A method for the automatic detection of myopia in Optos fundus images based on deep learning. *Int J Numer Method Biomed Eng*. 2021;37(6). <https://doi.org/10.1002/cnm.3460>.
35. Tan TE, Anees A, Chen C, et al. Retinal photograph-based deep learning algorithms for myopia and a blockchain platform to facilitate artificial intelligence medical research: a retrospective multicohort study. *Lancet Digit Health*. 2021;3(5):e317–29. [https://doi.org/10.1016/S2589-7500\(21\)00055-8](https://doi.org/10.1016/S2589-7500(21)00055-8).
36. Zou H, Shi S, Yang X, et al. Identification of ocular refraction based on deep learning algorithm as a novel retinoscopy method. *Biomed Eng Online*. 2022;21(1):87. <https://doi.org/10.1186/s12938-022-01057-9>.
37. Chun J, Kim Y, Shin KY, et al. Deep learning-based prediction of refractive error using photorefractive images captured by a smartphone: model development and validation study. *JMIR Med Inform*. 2020;8(5):e16225. <https://doi.org/10.2196/16225>.

38. Yoo TK, Ryu IH, Kim JK, Lee IS. Deep learning for predicting uncorrected refractive error using posterior segment optical coherence tomography images. *Eye*. 2022;36(10):1959–65. <https://doi.org/10.1038/s41433-021-01795-5>.
39. Fan Y, Yu Z, Tang T, et al. Machine learning algorithm improves accuracy of ortho-K lens fitting in vision shaping treatment. *Cont Lens Anterior Eye*. 2022;45(3):101474. <https://doi.org/10.1016/j.clae.2021.101474>.
40. Guo C, Ye Y, Yuan Y, et al. Development and validation of a novel nomogram for predicting the occurrence of myopia in schoolchildren: a prospective cohort study. *Am J Ophthalmol*. 2022;242:96–106. <https://doi.org/10.1016/j.ajo.2022.05.027>.
41. Chen S, Li L, He L, et al. Establishment and validation of a risk assessment model for myopia among Chinese primary school students during the COVID-19 pandemic: a lasso regression approach. *Heliyon*. 2023;9(10):e20638. <https://doi.org/10.1016/j.heliyon.2023.e20638>.
42. Sogawa T, Tabuchi H, Nagasato D, et al. Accuracy of a deep convolutional neural network in the detection of myopic macular diseases using swept-source optical coherence tomography. *PLoS ONE*. 2020;15(4):e0227240. <https://doi.org/10.1371/journal.pone.0227240>.
43. Li Y, Feng W, Zhao X, et al. Development and validation of a deep learning system to screen vision-threatening conditions in high myopia using optical coherence tomography images. *Br J Ophthalmol*. 2022;106(5):633–9. <https://doi.org/10.1136/bjophthalmol-2020-317825>.
44. Alzubaidi L, Zhang J, Humaidi AJ, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data*. 2021;8(1):53. <https://doi.org/10.1186/s40537-021-00444-8>.
45. Jain R, Yoo TK, Ryu IH, Song J, Kolte N, Nariani A. Deep transfer learning for ethnically distinct populations: prediction of refractive error using optical coherence tomography. *Ophthalmol Ther*. Published online November 13, 2023. <https://doi.org/10.1007/s40123-023-00842-6>.
46. Du R, Xie S, Fang Y, et al. Deep learning approach for automated detection of myopic maculopathy and pathologic Myopia in fundus images. *Ophthalmol Retina*. 2021;5(12):1235–44. <https://doi.org/10.1016/j.oret.2021.02.006>.
47. Hemelings R, Elen B, Blaschko MB, Jacob J, Stalmans I, De Boever P. Pathological myopia classification with simultaneous lesion segmentation using deep learning. *Comput Methods Programs Biomed*. 2021;199:105920. <https://doi.org/10.1016/j.cmpb.2020.105920>.
48. Foo LL, Lim GYS, Lanca C, et al. Deep learning system to predict the 5-year risk of high myopia using fundus imaging in children. *NPJ Digit Med*. 2023;6(1):10. <https://doi.org/10.1038/s41746-023-00752-8>.
49. Dong L, Hu XY, Yan YN et al. Deep learning-based estimation of axial length and subfoveal choroidal thickness from color fundus photographs. *Front Cell Dev Biol*. 2021;9. <https://doi.org/10.3389/fcell.2021.653692>.
50. Jeong Y, Lee B, Han JH, Oh J. Ocular axial length prediction based on visual interpretation of retinal fundus images via deep neural network. *IEEE J Sel Top Quantum Electron*. 2021;27(4):1–7. <https://doi.org/10.1109/JSTQE.2020.3038845>.
51. Oh R, Lee EK, Bae K, Park UC, Yu HG, Yoon CK. Deep learning-based prediction of axial length using ultra-widefield fundus photography. *Korean J Ophthalmol*. 2023;37(2):95–104. <https://doi.org/10.3341/kjo.2022.0059>.
52. Sung MS, Heo H, Piao H, Guo Y, Park SW. Parapapillary atrophy and changes in the optic nerve head and posterior pole in high myopia. *Sci Rep*. 2020;10(1):4607. <https://doi.org/10.1038/s41598-020-61485-2>.
53. Sung MS, Heo MY, Heo H, Park SW. Bruch's membrane opening enlargement and its implication on the myopic optic nerve head. *Sci Rep*. 2019;9(1):19564. <https://doi.org/10.1038/s41598-019-55926-w>.
54. Zhang Q, Xu L, Wei W Bin, Wang YX, Jonas JB. Size and shape of bruch's membrane opening in relationship to axial length, gamma zone, and macular Bruch's membrane defects. *Invest Ophthalmol Vis Sci*. 2019;60(7):2591. <https://doi.org/10.1167/iovs.19-27331>.
55. Öner V, Taş M, Türkücü FM, Alakuş MF, İşcan Y, Yazıcı AT. Evaluation of peripapillary retinal nerve fiber layer thickness of myopic and hyperopic patients: a controlled study by stratus optical coherence tomography. *Curr Eye Res*. 2013;38(1):102–7. <https://doi.org/10.3109/02713683.2012.715714>.
56. Huang G, Wen Y, Qian B, Bi L, Chen T, Sheng B. Attention-based multi-scale feature fusion network for myopia grading using optical coherence tomography images. *Vis Comput*. Published online December 22, 2023. <https://doi.org/10.1007/s00371-023-03189-y>.
57. Wang Y, Du R, Xie S, et al. Machine learning models for predicting long-term visual acuity in highly myopic eyes. *JAMA Ophthalmol*. Published online October 26, 2023. <https://doi.org/10.1001/jamaophthalmol.2023.4786>.
58. Blade PJ, Candy TR. Validation of the PowerRefractor for measuring human infant refraction. *Optom Vis Sci*. 2006;83(6):346–53. <https://doi.org/10.1097/01.opx.0000221402.35099.fb>.
59. Sravani NG, Nilagiri VK, Bharadwaj SR. Photorefractive estimates of refractive power varies with the ethnic origin of human eyes. *Sci Rep*. 2015;5(1):7976. <https://doi.org/10.1038/srep07976>.
60. Linde G, Chalakkal R, Zhou L, et al. Automatic refractive error estimation using deep learning-based analysis of red reflex images. *Diagnostics*. 2023;13(17):2810. <https://doi.org/10.3390/diagnostics13172810>.

61. Xu D, Ding S, Zheng T, et al. Deep learning for predicting refractive error from multiple photorefractive images. *Biomed Eng Online*. 2022;21(1):55. <https://doi.org/10.1186/s12938-022-01025-3>.
62. Wilson LB, Melia M, Kraker RT, et al. Accuracy of autorefraction in children. *Ophthalmology*. 2020;127(9):1259–67. <https://doi.org/10.1016/j.ophtha.2020.03.004>.
63. Morgan IG, Iribarren R, Fotouhi A, Grzybowski A. Cycloplegic refraction is the gold standard for epidemiological studies. *Acta Ophthalmol*. 2015;93(6):581–5. <https://doi.org/10.1111/aos.12642>.
64. Guo X, Shakarchi AF, Block SS, Friedman DS, Repka MX, Collins ME. Noncycloplegic compared with cycloplegic refraction in a Chicago school-aged population. *Ophthalmology*. 2022;129(7):813–20. <https://doi.org/10.1016/j.ophtha.2022.02.027>.
65. Gandomi AH, Chen F, Abualigah L. Machine learning technologies for big data analytics. *Electronics (Basel)*. 2022;11(3):421. <https://doi.org/10.3390/electronics11030421>.
66. Lin H, Long E, Ding X, et al. Prediction of myopia development among Chinese school-aged children using refraction data from electronic medical records: a retrospective, multicentre machine learning study. *PLoS Med*. 2018;15(11):e1002674. <https://doi.org/10.1371/journal.pmed.1002674>.
67. Yang X, Chen G, Qian Y, et al. Prediction of Myopia in adolescents through machine learning methods. *Int J Environ Res Public Health*. 2020;17(2):463. <https://doi.org/10.3390/ijerph17020463>.
68. Foo LL, Ang M, Wong CW, et al. Is artificial intelligence a solution to the myopia pandemic? *Br J Ophthalmol*. 2021;105(6):741–4. <https://doi.org/10.1136/bjophthalmol-2021-319129>.
69. Zhang C, Zhao J, Zhu Z, et al. Applications of artificial intelligence in myopia: current and future directions. *Front Med (Lausanne)*. 2022;9. <https://doi.org/10.3389/fmed.2022.840498>.
70. Li SM, Ren MY, Gan J, et al. Machine learning to determine risk factors for Myopia progression in primary school children: the anyang childhood eye study. *Ophthalmol Ther*. 2022;11(2):573–85. <https://doi.org/10.1007/s40123-021-00450-2>.
71. Hernández CS, Gil A, Casares I, et al. Prediction of manifest refraction using machine learning ensemble models on wavefront aberrometry data. *J Optom*. 2022;15:S22–31. <https://doi.org/10.1016/j.optom.2022.03.001>.
72. Barraza-Bernal MJ, Ohlendorf A, Sanz Diez P, et al. Prediction of refractive error and its progression: a machine learning-based algorithm. *BMJ Open Ophthalmol*. 2023;8(1):e001298. <https://doi.org/10.1136/bmjophth-2023-001298>.
73. Zhu S, Zhan H, Yan Z, et al. Prediction of spherical equivalent refraction and axial length in children based on machine learning. *Indian J Ophthalmol*. 2023;71(5):2115–31. https://doi.org/10.4103/IJO.IJO_2989_22.
74. Tang T, Yu Z, Xu Q, et al. A machine learning-based algorithm used to estimate the physiological elongation of ocular axial length in myopic children. *Eye and Vision*. 2020;7(1):50. <https://doi.org/10.1186/s40662-020-00214-2>.
75. Du B, Wang Q, Luo Y, et al. Prediction of spherical equivalent difference before and after cycloplegia in school-age children with machine learning algorithms. *Front Public Health*. 2023;11. <https://doi.org/10.3389/fpubh.2023.1096330>.
76. Lawrenson JG, Shah R, Huntjens B, et al. Interventions for myopia control in children: a living systematic review and network meta-analysis. *Cochrane Database Syst Rev*. 2023;2023(2). <https://doi.org/10.1002/14651858.CD014758.pub2>.
77. Logan NS, Bullimore MA. Optical interventions for myopia control. *Eye*. Published online September 22, 2023. <https://doi.org/10.1038/s41433-023-02723-5>.
78. Agyekum S, Chan PP, Adjei PE, et al. Cost-effectiveness analysis of myopia progression interventions in children. *JAMA Netw Open*. 2023;6(11):e2340986. <https://doi.org/10.1001/jamanetworkopen.2023.40986>.
79. Fan Y, Yu Z, Peng Z, et al. Machine learning based strategy surpasses the traditional method for selecting the first trial Lens parameters for corneal refractive therapy in Chinese adolescents with myopia. *Cont Lens Anterior Eye*. 2021;44(3):101330. <https://doi.org/10.1016/j.clae.2020.05.001>.
80. Tang Y, Chen Z, Wang W, et al. A deep learning-based framework for accurate evaluation of corneal treatment zone after orthokeratology. *Transl Vis Sci Technol*. 2021;10(14):21. <https://doi.org/10.1167/tvst.10.14.21>.
81. Zhang M, Guo Y, Zhou C, et al. Deep neural network with self-attention based automated determination system for treatment zone and peripheral steepened zone in Orthokeratology for adolescent myopia. *Contact Lens and Anterior Eye*. Published online November 2023:102081. <https://doi.org/10.1016/j.clae.2023.102081>.
82. Wu TE, Chen HA, Jhou MJ, Chen YN, Chang TJ, Lu CJ. Evaluating the effect of topical atropine use for myopia control on intraocular pressure by using machine learning. *J Clin Med*. 2020;10(1):111. <https://doi.org/10.3390/jcm10010111>.
83. Blair JPM, Rodriguez JN, Lasagni Vitar RM, et al. Development of LuxIA, a cloud-based ai diabetic retinopathy screening tool using a single color fundus image. *Transl Vis Sci Technol*. 2023;12(11):38. <https://doi.org/10.1167/tvst.12.11.38>.
84. Zhang J, Zou H. Artificial intelligence technology for myopia challenges: a review. *Front Cell Dev Biol*. 2023;11. <https://doi.org/10.3389/fcell.2023.1124005>.
85. He HL, Liu YX, Song H, Xu TZ, Wong TY, Jin ZB. Initiation of China Alliance of Research in High Myopia (CHARM): protocol for an AI-based multimodal high myopia research biobank. *BMJ Open*. 2023;13(12):e076418. <https://doi.org/10.1136/bmjopen-2023-076418>.

Artificial Intelligence and Prediction of Eye Diseases

26

Tahereh NaseriBooriabadi ,
Mohammad Hassan Emamian ,
and Andrzej Grzybowski 

Abstract

The early detection and prediction of eye diseases play a crucial role in preventing permanent visual impairments and vision loss. Given the significant impact of visual impairments on people's quality of life, there is a growing need to enhance the speed and accuracy of disease diagnosis and improve the performance of diagnostic tools. The use of machine learning (ML) models, deep learning, and artificial intelligence (AI) to expedite the prediction of the onset and progression of ocular diseases has sparked a significant revolution in the field of ophthalmology. This

chapter highlights the significance of predicting eye diseases and visual impairments. It explores the use of AI in predicting major diseases of the eye such as diabetic retinopathy (DR), glaucoma, and age-related macular degeneration (AMD), cataract, myopia, and keratoconus. In the next section, we will delve into the application of AI for the analysis of eye images to predict systemic diseases. Subsequently, we will investigate the potential of AI in the realm of ophthalmology.

Keywords

Artificial Intelligence · Prediction · Progression · Screening · Diabetic retinopathy · Glaucoma · Age-related macular degeneration · Cataract · Myopia · Keratoconus · Dataset · Machine learning · Deep learning

T. NaseriBooriabadi
Department of Health Information Technology,
School of Allied Medical Sciences, Shahroud
University of Medical Sciences, Shahroud, Iran
e-mail: taherehnaseribooriabadi@gmail.com

M. H. Emamian
Ophthalmic Epidemiology Research Center,
Shahroud University of Medical Sciences,
Shahroud, Iran
e-mail: emamian@shmu.ac.ir

A. Grzybowski (✉)
Department of Ophthalmology, University
of Warmia and Mazury, Olsztyn, Poland
e-mail: ae.grzybowski@gmail.com

Institute for Research in Ophthalmology,
Foundation for Ophthalmology Development,
Poznan, Poland

Prediction of Eye Diseases and Visual Impairments

The impact of vision on human life is significant, as it plays a crucial role in various dimensions and stages of life. The lack of vision can make daily personal and social activities challenging [1]. In 2020, the global labor force

population reported approximately 18.1 million blind individuals and 142.6 million people with moderate to severe visual impairments [2]. The economic ramifications of blindness extend to countries, families, and the quality of life of affected individuals [3]. Notably, the economic burden due to blindness primarily comprises care costs in nursing homes (\$41.8 billion in the US), other medical care expenses (\$30.9 billion in the US), and reduced labor force participation (\$16.2 billion in the US) [4].

Visual impairments not only result in diverse costs but also impact various dimensions of life, particularly the quality of life of affected individuals [1]. Furthermore, the main causes of blindness among individuals aged 50 and above in 2020 were attributed to cataract (15.02 million cases), glaucoma (3.6 million cases), uncorrected refractive errors (2.3 million cases), age-related macular degeneration (AMD) (8.8 million cases), and diabetic retinopathy (DR) (0.86 million cases) [5]. These conditions entail permanent visual impairments and can lead to complete blindness [1]. As several risk factors associated with these conditions are preventable, early prediction and diagnosis are imperative for effective treatment and the prevention of blindness [5].

Advantages of Using AI in Eye Diseases

By analyzing patient data, artificial intelligence (AI) demonstrates the capacity to anticipate disease progression and execute healthcare interventions tailored to individual patient needs in an automated manner. Furthermore, AI facilitates the automated screening of diseases, the prioritization of high-risk cases, and the prompt and precise identification of diseases, thereby contributing to the enhancement of eye care [6]. Global eye health signifies the attainment of the highest level of vision health and optimal functioning of the eye, leading to the overall improvement of general well-being and quality of life. The escalating expansion of

digital health and AI applications offers a transformative opportunity for reshaping eye health, encompassing the facilitation and amplification of access to healthcare and the support of data-driven clinical decision-making [7]. The utilization of AI models holds the potential to revolutionize the processes of detecting and predicting retinal diseases, resulting in a substantial reduction of the societal and economic repercussions associated with blindness on a global scale [8]. AI possesses significant potential in improving patients' accessibility to screening and clinical diagnosis services as well as in reducing healthcare costs. This is particularly pertinent in scenarios characterized by higher disease risk or in settings where resources are constrained [9]. For instance, previous research underscores the effectiveness of AI in diagnosing DR and highlights its utility in low- and middle-income countries that face resource limitations, as AI is capable of conducting proper and precise screening without necessitating expert human resources [10].

Automated screening and diagnosis, while conserving time and minimizing the likelihood of misdiagnosis, adjunctively serve to alleviate the burden on healthcare professionals and mitigate financial costs [11]. The progression in the computational capacity of machines, machine learning (ML) algorithms, and access to extensive volumes of medical data derived from electronic health records and wearable health devices [12]; has elevated the diagnostic precision of imaging tools through AI technologies [13]. Ophthalmology, reliant on imaging data and objective criteria, necessitates comprehensive image interpretation for the diagnosis of eye diseases. Consequently, a notable body of cutting-edge AI research has driven advancements in the diagnosis of eye diseases [11].

Integration of AI in ophthalmology presents an opportunity to achieve favorable treatment outcomes and increases the precision of eye surgeries [14]. It is anticipated that the utilization of AI and deep learning algorithms will revolutionize image analysis methods and techniques. This will facilitate significant advancements

in the timely diagnosis and accurate prognosis estimation of ocular diseases [15]. The cost-effectiveness and affordability of incorporating AI into ophthalmology [16], coupled with the growing popularity of digital health and AI applications in recent years [7], highlight the potential of telemedicine services, portable handheld devices, and AI-based automatic algorithms [17] to alleviate the burden of eye diseases through screening programs.

AI in Predicting Eye Diseases

AI models are shown to be able to diagnose many common eye diseases such as DR, AMD, glaucoma, and cataract [18]. Researchers utilized various learning and ML methods for categorizing eye diseases [19]. ML algorithms using feature extraction demonstrated high sensitivity and relatively low specificity in DR diagnosis [19]. ML is still utilized in certain tasks, but higher sensitivity and specificity can be achieved by employing deep learning (DL) techniques [19]. Both ML and DL have their advantages and limitations, and the selection between the two depends on the nature of the problem and the availability of data [20]. The effectiveness of using ML algorithms in predicting DR has been confirmed [21]. The implementation of a data-driven approach in DR management, improved resource allocation, better design of essential screening programs, and the integration of clinical data and molecular profiles have been associated with increased accuracy in predicting DR [22]. In a study utilizing the deep convolutional neural network, researchers developed a self-diagnostic model named EYENET to predict strabismus, exophthalmos, cataract, glaucoma, and uveitis, with the potential to reduce workload of physicians [23].

We have categorized the applications of AI algorithms in ophthalmology into two groups:

1. The use of AI algorithms in major diseases of the eye
2. Application of AI algorithms in systemic diseases based on retinal image analysis.

In this chapter, the use of AI for predicting major diseases of the eye (including DR, glaucoma, AMD, cataracts, myopia, and keratoconus) will be described. Additionally, the examination of the prediction of systemic diseases based on retinal image analysis is presented.

AI Application in the Major Eye Diseases

Diabetic Retinopathy

Diabetic Retinopathy is a serious complication of diabetes and is one of the main causes of blindness. Patients with DR face medical expenses, reduced quality of life, disabilities, decreased productivity, higher insurance costs, and lost income due to early retirement or premature death [3]. The prevention of blindness can be achieved by delaying the onset of DR in individuals with diabetes and by halting the progression of the disease in affected patients [24]. The accurate prediction of the progression of DR stands as a paramount challenge for ophthalmologists involved in the treatment of affected patients. Therefore, timely diagnosis of DR is pivotal to prevent irreversible disease advancement [25]. Compared to humans, ML algorithms can detect DR faster and more accurately [26]. In traditional screenings, ophthalmologists mainly rely on fundus images, which can be intricate and susceptible to errors including torsion [26]. Usman et al. found that previous researchers have used a combination of different risk factors to improve the performance of DR prediction. They also reported that traditional models relied on manual statistical analysis and feature engineering, which were less robust and precise compared to ML models [25].

It was found that ML is well suited for predicting the progression of diabetic DR by leveraging its ability to discover complex hidden patterns from large medical clinical records and recognize the intricate relationships between them over time [25]. The introduction of automated diagnosis for DR presents a potential alternative to conventional screening methods,

offering a means to conserve time, resources, and effort for ophthalmologists. Given the escalating diabetic patient population, the development of reliable DR diagnosis systems becomes essential [27].

A new hybrid deep learning model by combining several distinct deep neural network architectures was proposed to diagnose and classify specific lesions associated with DR, which could provide more accurate and interpretable insights. The diverse and high-quality dataset of retina images used in this model improved the power of the model and the generalizability of its results [28]. In another study, the Bilayered Neural Network (BNN) was employed to analyze retinal images for detecting DR with a high accuracy of 98.5%, surpassing previous studies. The BNN model outperformed other models in terms of predictive performance, highlighting its effectiveness in diagnosing DR from retinal fundus images [29]. Utilizing DeepDR Plus, Omer successfully predicted the progression of retinopathy within 5 years based solely on fundus images [30]. Implementing this system in the hospital resulted in an increase in the average screening interval from 12 months to approximately 32 months. This led to reduced screening needs, cost savings, and improved patient care. Additionally, the system effectively tailored screening intervals based on individual patient risk, consequently aiding in the prevention of blindness caused by late detection of vision-threatening retinopathy. Furthermore, this system empowered doctors to devise personalized treatment plans based on each patient's risk of disease progression, ultimately contributing to the prevention of retinopathy progression [30]. Dai et al. introduced a new system designed to improve screening efficiency and reduce errors in existing models. The proposed model's high accuracy and reliability can lead to the creation of automatic screening tools for timely DR diagnosis [31]. Furthermore, the streamlined diagnosis process may enhance screening accessibility and affordability, especially in regions with a shortage of skilled ophthalmologists [31]. Usman et al. found that some prediction models use only retinal images as predictors in DR

progression risk. They also reported that some ML based predictive models consider a combination of ocular symptoms, retinal images, socio-demographic factors, and clinical data from electronic medical records, and molecular and genetic profiles [25]. They also highlighted that risk prediction models demonstrate improved performance when retinal images and clinical data are utilized together [25]. Given the diversity of methods and predictors used across different studies, there is no agreement on the specific predictors necessary to create accurate prediction models for DR progression. The optimal predictors and the type of prediction to be used depend on the specific ML algorithm and the model complexity [25]. Figure 26.1 presents features, predictive models, and outputs used in studies on AI in predicting diabetic retinopathy (DR). Table 26.1 summarizes the recent articles on AI in predicting DR.

Glaucoma

After cataracts, glaucoma is the second most common cause of blindness globally. The blindness that occurs as a result of glaucoma is usually permanent. Projections suggest that by 2040, approximately 111.8 million individuals will experience the impact of this disease [36]. Identifying glaucoma in its early stages helps avoid the need for complex and costly procedures, leading to reduced expenses for individuals and globally. Glaucoma patients often require long-term monitoring, and accurately predicting the prognosis of this disease is a critical component of its treatment [37]. However, most public glaucoma datasets are primarily cross-sectional and focus on optic disc segmentation and diagnosis. With the advancement of AI technology, deep learning models utilizing longitudinal data sets can accurately predict a patient's visual field and anticipate the progression of the disease [38]. The application of AI in the diagnosis and prediction of glaucoma is rapidly evolving. By integrating AI algorithms, conventional functional and structural tests for glaucoma diagnosis can be strengthened [39].

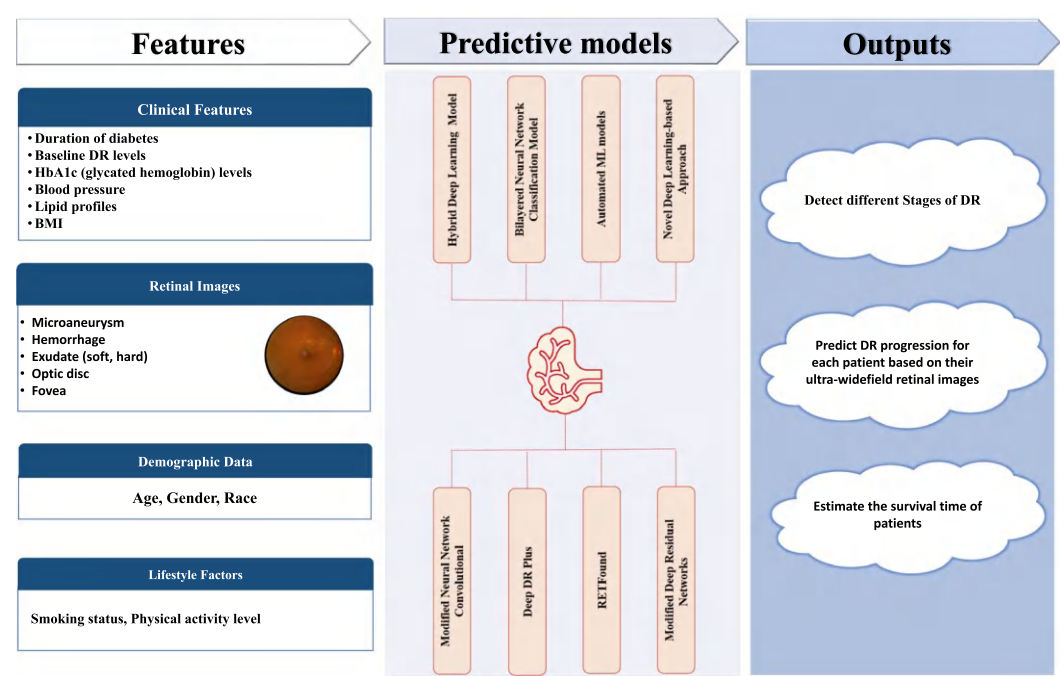


Fig. 26.1 Features, predictive models, and outputs used in studies on AI in predicting diabetic retinopathy (DR)

While AI research in glaucoma has encountered limitations such as the lack of multimodal evaluation and long-term natural progression, significant progress has been achieved in this field. Multiple studies confirm the successful use of AI in diagnosing glaucoma by analyzing structural changes in the fundus [40–44] and OCT images [45–47]. To improve the assessment of glaucoma, screening, and predicting its progression, AI can be used to analyze the features of the cup, optic disc, optic nerve head in the fundus as well as the retinal layers in OCT images, and patterns of visual field impairment [48]. It was demonstrated that a deep learning algorithm exhibited a high level of performance in the diagnosis of referable glaucomatous optic neuropathy (GON). The reported area under the curve (AUC) values ranged from 0.855 to 0.945, indicating the algorithm’s strong predictive capabilities. Furthermore, the study has identified crucial features associated with referable GON, which can significantly contribute to the accurate predictions made by the algorithm [49].

Another study demonstrated that deep learning algorithms enhance our understanding of the underlying mechanisms of glaucoma by identifying both clinical and subclinical symptoms associated with the disease [50]. Additionally, these algorithms can predict the risk of glaucoma by analyzing color fundus photographs, particularly in high-risk populations. It was postulated that integrating deep learning algorithms into routine screenings might facilitate timely detection and treatment of patients [50]. Yusefi et al. developed an AI-powered dashboard designed to assist clinicians in identifying glaucomatous visual impairment and monitoring its progression [51]. The performance validation of this dashboard, using independent reference datasets, demonstrated its impressive ability to detect possible non-progression with 94% accuracy and possible progression of glaucomatous lesions with 77% accuracy [51]. Thakhur et al. highlighted the potential of AI to process and learn from large-scale datasets, uncovering subtle disease patterns. They reported that this capability holds promise for significantly

Table 26.1 A summary of studies on AI in predicting diabetic retinopathy (DR)

Authors, publication year	Model	Dataset	Features	Feature extraction	Input	Classifier	Output	Performance (internal)	Performance (external)
Jabbar et al. 2024 [28]	Hybrid Deep Learning Model	Kaggle EyePACS1	Retinal lesions, Microaneurysms, Hemorrhages, Exudates	Transfer learning models (GoogleNet and ResNet-16)	Retinal fundus images of patients with DR	RSVM, DT, NB, RF	Detect different DR Stages	Accuracy: 94% Precision: 97% Recall: 89% F1 score: 96%	–
Omer 2024 [29]	Bilayered Neural Network classification model	Kaggle Dataset (IDRID)	Optic Disc, Fovea, Hemorrhages, Microaneurysms	vSLAM	Retinal fundus images	Bilayered Neural Network	Detect different Stages of DR	ROC curve: 0.985, Accuracy: 98.5%	–
Silva et al. 2024 [32]	Automated ML models	Ultra-widefield retinal images of patients with diabetes, along with information on DR progression	Presence of microaneurysms, hemorrhages, exudates, and other retinal abnormalities associated with DR	CNN	Ultra-widefield retinal images of patients with DR	CNN	Predict DR progression for each patient based on their ultra-widefield retinal images	Mild NPDR Sensitivity: 63%, Specificity: 63%, PPV 63%, NPV 63%, Accuracy 63.2% Moderate NPDR Sensitivity: 75%, Specificity: 75%, PPV 75%, NPV 75%, Accuracy 75% Mild NPDR Sensitivity: 72%, Specificity: 63%, PPV 26%, NPV 92%, Accuracy 64.3% Moderate NPDR Sensitivity: 80%, Specificity: 72%, PPV 45%, NPV 22%, Accuracy 73.8%	
Singh et al. 2024 [31]	A novel deep learning-based approach for detecting DR using transfer learning	A dataset of Indian DR images	Microaneurysms, soft exudates, hard exudates, and hemorrhages	VGG16 and VGG19	Fundus images from the IDRiD dataset	LR, RF, SVM, NN, XBoost, AdaBoost	Detect DR	Accuracy 90.4% AUC 0.985 F1-score 90.3% Precision 90.7% Recall 90.4%	–
Dai et al. 2024 [30]	DeepDR Plus	DR Progression Study (DRPS) cohort	Retinal fundus images such as age, gender, smoking status, duration of diabetes, baseline DR level, body mass index, glycated hemoglobin, blood pressure, and lipid profiles	ResNet-50 soft-attention mechanism	Baseline retinal fundus images and clinical metadata	ResNet-50	Predict the risk of DR progression	For years 1–5 AUC 0.822–0.896 C-indexes 0.754–0.846 BS 0.153–0.241	AUC: 0.738–0.886

(continued)

Table 26.1 (continued)

Authors, publication year	Model	Dataset	Features	Feature extraction	Input	Classifier	Output	Performance (internal)	Performance (external)
Zhou et al. 2023 [33]	RETFound, a self-supervised learning model	Public datasets including MEH, AlzEye, Diabetic retinopathy, Glaucoma, and Multiclass disease	Retinal images used for training the model	SSL-Retinal model on retinal images via SSL from scratch	Retinal images for training and testing the model	The model is fine-tuned for specific tasks like ocular disease diagnostic classification and prognosis	Predict conversion of fellow eye to wet-AMD in 1 year	RETFound achieved an AUROC of 0.943 on the Kaggle APTOS-2019 dataset, 0.822 on the IDRID dataset, and 0.884 on the MESSIDOR-2 dataset for diabetic retinopathy classification, significantly outperforming SL-ImageNet	RETFound fine-tuned on the Kaggle APTOS-2019 dataset achieved an AUROC of 0.822 on the IDRID dataset and 0.738 on the MESSIDOR-2 dataset, showing significantly higher performance compared to other models
Bajwa et al. 2023 [34]	A modified Convolutional Neural Network (CNN)	Fundus images collected from the Sindh Institute of Ophthalmology and Visual Sciences	The characteristics extracted from the fundus images, are then used as inputs to the modified CNN model	The CNN model	The pre-processed fundus images	The modified CNN model	Predict different stages of DR	Accuracy 93.72% Sensitivity: 97.3% Specificity: 92.9% AUC ^{male patient} : 0.975 AUC ^{female patient} : 0.951 AUC ^{all patients} : 0.969	Accuracy 93.4% Sensitivity 97.42% Specificity: 89.45% Clinical Expert Accuracy: 93.30% Sensitivity: 97.30% Specificity: 92.90%
Sambyal et al. 2023 [35]	Modified deep residual networks for both binary (healthy vs. non-healthy) and multi-stage classification of DR images	MESSIDOR	Retinal pathologies associated with DR, such as hemorrhages, microaneurysms, soft exudates, and hard exudates	Modified residual network models	Fundus images	Modified ResNet18, ResNet34, and ResNet50	Classify labels (the severity stage of DR)	The modified ResNet models Accuracy: 99.47% for binary classification (ResNet18, ResNet34) Accuracy 99.87% for binary classification (ResNet50) Accuracy: 99.37%	–

enhancing the accuracy and effectiveness of glaucoma diagnosis in the future [52].

Suganthé et al. showed that the CNN algorithm has better accuracy by effectively analyzing features such as the cup-to-disc ratio, which is critical for glaucoma diagnosis [53]. These algorithms, with their deep learning capabilities, can identify subtle patterns and changes in eye images and help diagnose glaucoma in time. Using large data sets, these algorithms can learn complex relationships within the data and make more accurate and reliable predictions compared to traditional methods [53]. Subbarayadu et al. revealed that logistic regression, among various machine learning models, achieved an impressive 95% accuracy in classifying glaucoma cases [54].

Qidwai et al. developed an innovative system tailored to suggest the optimal glaucoma surgery approach for individual patients, emphasizing minimal invasiveness. The recommendations were based on a comprehensive assessment of key factors, including patient age, visual acuity, intraocular pressure, and visual field measurements. Remarkably, this system achieved an accuracy of 91%, along with sensitivity of 80% and specificity of 90% [55].

Wang et al. explored the utility of deep and machine learning models in leveraging structured data from electronic health records to predict glaucoma progression and surgical risk. Notably, when dealing with complex, high-dimensional data, decision tree models and deep learning architectures outperformed traditional regression models. External evaluation confirmed that the proposed model exhibited relatively robust performance [56].

In the field of glaucoma management, predictive models enhanced by AI have emerged as powerful tools for understanding patient risk profiles and predicting disease progression rates. [57]. Kamal et al. demonstrated that the use of explainable AI and interpretable ML techniques enables the analysis of predictive models for glaucoma. They showed this approach allows for a better understanding of the identified risk factors, thereby proving valuable in treatment planning [58].

AI holds significant promise for transforming glaucoma management by enhancing diagnostic accuracy, enabling timely diagnoses, and tailoring therapeutic interventions. A comprehensive review of existing research underscores the importance of further validation studies for AI models and systems, ensuring their optimal integration at the bedside. Given the ongoing advancements in AI technology and the availability of large datasets, the prediction of glaucoma progression appears to be increasingly promising. Figure 26.2 presents features, predictive models, and outputs used in studies on AI in predicting glaucoma. Table 26.2 summarizes the recent articles on AI in predicting glaucoma.

Age-Related Macular Degeneration (AMD)

AMD stands as a primary contributor to visual impairment, and projections suggest that the prevalence of this condition will impact approximately 288 million individuals by the year 2040 [62]. The use of AI can predict AMD by analyzing retinal medical images such as OCT or fundus imaging. AI algorithms are capable of detecting subtle retinal changes that may indicate AMD or its progression. These algorithms can also more accurately predict the risk of AMD by analyzing genetic data, lifestyle factors, and medical records [63]. This information enables physicians to identify individuals who are at risk of developing AMD and take timely action to prevent blindness [63]. AI-based analyses show promise in predicting the progression of AMD and assessing the response to treatment [64]. Corradetti et al. reported that AI algorithms were developed to predict the progression of moderate AMD towards macular neovascularization or geographic atrophy [65]. AI has emerged as a potent tool in the realm of AMD management, offering the potential to forecast disease progression, pinpoint biomarkers, and facilitate therapeutic decision-making [66]. AI algorithms can accurately predict the best corrected visual acuity, which is key to assessing AMD progression [67]. AI-based models have

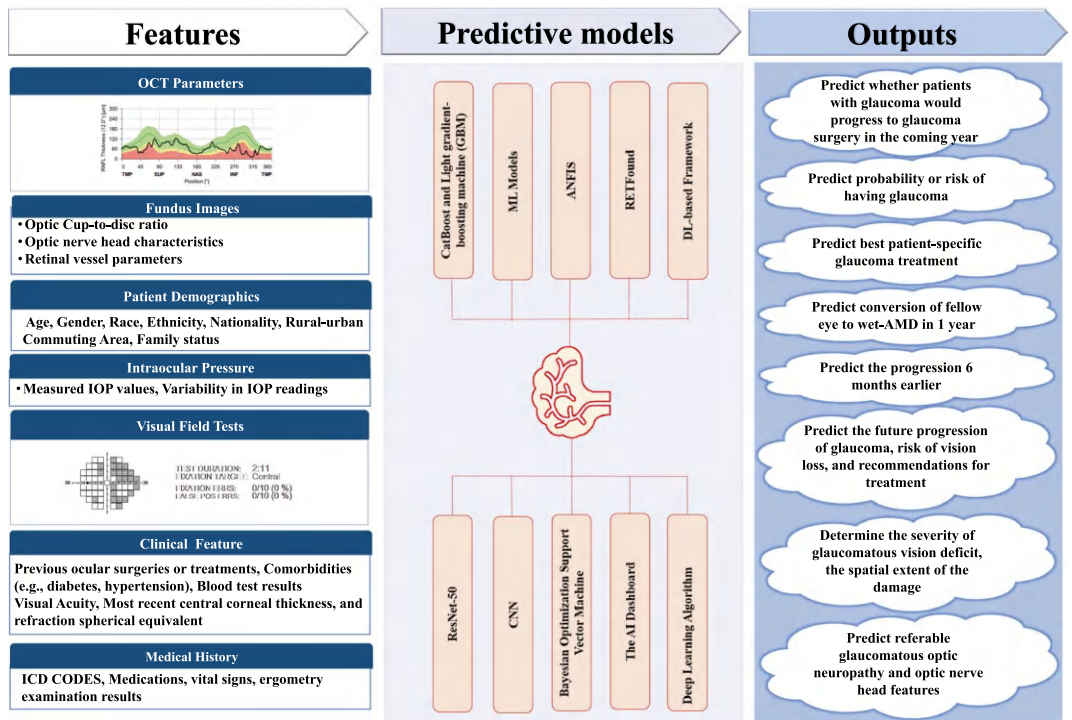


Fig. 26.2 Features, predictive models, and outputs used in studies on AI in predicting glaucoma

been developed and validated for the screening of AMD, as well as for predicting the progression of both dry and wet AMD at specific time intervals [68]. Dong et al. findings revealed that AI demonstrates the capability to detect AMD in color fundus photographs. The findings of this study underscore the utility of automated tools employing AI for the diagnosis of AMD [69]. Crincoli et al. specifically focused on the diagnosis and treatment of AMD, recognizing the potential of decision-making systems based on AI, GANs, and deep learning models in predicting treatment outcomes, evaluating visual acuity, and classifying disease patterns. The findings of this review confirm that artificial intelligence has a significant impact on increasing the accuracy and reliability of medical evaluations in ophthalmology [70]. Leng et al.'s findings suggest that CNNs are the most widely accepted deep learning algorithms for diagnosing AMD. Specifically, ResNets, a type of CNN, are highly effective in accurately diagnosing this condition.

The types of AMD and the layers of the network are both crucial factors in the model training process. Appropriate network layers can significantly enhance the reliability of the model [71]. As new diagnostic methods continue to emerge, more datasets will be generated to train deep learning models in the future. This is expected to greatly benefit fundus screening programs, improve long-term medical treatments, and alleviate the workload of physicians [71]. Recently RETFound which stands for Retinal Foundation outperforms other models in identifying serious eye conditions by using self-supervised learning on a large number of fundus photos [33]. RETFound improves model accuracy and reduces the annotation burden on experts, expanding the use of retinal imaging in clinical AI applications. For example, it demonstrated the strongest performance (AUROC of 0.862) in predicting the likelihood of the fellow eye developing wet-AMD within a year, surpassing other methods [33]. Figure 26.3 presents features,

Table 26.2 A summary of studies on AI in predicting glaucoma

Authors, publication year	Model	Dataset	Features	Feature extraction	Input	Classifier	Output	Performance (Internal)	Performance (External)
Wang et al. 2024 [56]	ML models [1] penalized LR (lasso, ridge, and elastic net); [2] tree-based models (RF, gradient boosted machines, and XGBoost) [3] deep learning models	The Sight Outcomes Research Collaborative (SOURCE)	Demographics (age, sex, race), ocular measurements (intraocular pressure, optic nerve cup-to-disc ratio), medical history (diagnosis codes, medications), and laboratory results	Not explicitly mention	Demographics: age, sex, race Ocular measurements: intraocular pressure, optic nerve cup-to-disc ratio Medical history: diagnosis codes, medication prescriptions Laboratory results	LR, RF, and XBoost	Predict whether patients with glaucoma would progress to surgery in the coming year	AUROC: 0.735–0.771	AUROC: 0.706–0.754
Landau Prat et al. 2024 [59]	Multiple predictive models for glaucoma using ML techniques	195,969 visits of 37,692 adults who had undergone comprehensive ophthalmic and systemic evaluations	Visual acuity, intraocular pressure, detailed anterior and posterior segment evaluation, demographic data, vital signs, detailed ergometry examination results, blood test results, and comorbidities (hypertension, diabetes mellitus)	CatBoost	Intraocular pressure, cup-to-disc ratio, visual field defects Systemic features: Age, sex, race/ethnicity, body mass index, diabetes status, hypertension status	CatBoost and Light gradient-boosting machine (GBM)	Predict probability or risk of having glaucoma	Prediction Models Based on Ocular and Systemic Features AUC ROC: 0.84 Prediction Models Based on Systemic Features Only AUC ROC: 0.72 Prediction Model Based on Ocular Features Only AUC ROC: 0.86	–

(continued)

Table 26.2 (continued)

Authors, publication year	Model	Dataset	Features	Feature extraction	Input	Classifier	Output	Performance (Internal)	Performance (External)
Qidwai et al. 2023 [55]	AI-based prediction system called iMIGS	The dataset consists of a retrospective case series of 372 patients who underwent minimally invasive glaucoma surgery (MIGS) from July 2016 to May 2020 at Queen Victoria Hospital, NHS Foundation Trust, UK	Age, visual acuity, intraocular Pressure (IOP), visual field, number of medicinal drops being taken daily Approximate Type of Glaucoma	Clinical measurements were collected during a baseline visit prior to surgery, including detailed patient information such as age, gender, type of glaucoma, and prior interventions. The significant variables were selected for the predictive model	Age, visual acuity, intraocular pressure (IOP), visual field, number of medicinal drops being taken daily, approximate type of Glaucoma	The ANFIS model acts as a classifier by utilizing fuzzy logic rules and neural network learning to predict the best MIGS treatment option based on the input features	Predict best patient-specific glaucoma treatment	Accuracy: 91%, Sensitivity: 80%, Specificity: 90%	–
Zhou et al. 2023 [33]	RETFound, a self-supervised learning model	Public datasets including MEH, AlzEye, Diabetic retinopathy, Glaucoma, and Multiclass disease	Retinal images used for training the model	SSL-Retinal trains the model on retinal images via SSL from scratch	Retinal images for training and testing the model	The model is fine-tuned for specific tasks like ocular disease diagnostic classification and prognosis	Predict conversion of fellow eye to wet-AMD in 1 year	–	–

(continued)

Table 26.2 (continued)

Authors, publication year	Model	Dataset	Features	Feature extraction	Input	Classifier	Output	Performance (Internal)	Performance (External)
Hussain et al. 2023 [60]	DL-based framework	The longitudinal dataset used for training comprised measurements from 105 glaucomatous eyes	AGE, GENDER, BCVA, REFR, CCT, AXL, RNFL, IOP and VF MD	Visual field test parameters were extracted directly from the test results OCT features were extracted using image processing techniques on the OCT scans Clinical features were extracted from patient medical records	OCT + Baseline, VF + Baseline, OCT + VF + Baseline at M3, OCT + VF + Baseline at M3 + Future Synthetic OCT, OCT + VF + Baseline at M6 + Future Synthetic OCT	Deep neural network DL (CNN + LSTM network) + GAN	Predict progression 6 months earlier	AUC 0.83, for predicting the progression 6 months earlier The use of synthetic future images enabled the model to accurately predict the vision loss even earlier (9 months earlier) with an AUC of 0.81, compared to using only structural (AUC 0.68) or only functional measures (AUC 0.72)	–

(continued)

Table 26.2 (continued)

Authors, publication year	Model	Dataset	Features	Feature extraction	Input	Classifier	Output	Performance (Internal)	Performance (External)
Huang et al. 2023 [38]	ResNet-50, residual learning framework	Longitudinal visual field tests and fundus images from glaucoma patients	Visual field features (mean deviation, pattern standard deviation, and other standard visual field indices), Fundus image features (optic nerve head parameters, retinal nerve fiber layer thickness, and other morphological features)	Not explicitly mention	CFPs of the eye at the baseline visit	–	Predict disease progression, risk of vision loss, or recommendations for treatment	The AUC for different progression criteria ranges from 0.71 to 0.80, indicating good predictive ability	–
Helen and Gokila 2023 [23]	Convolutional Neural Network (CNN)	11,929 eye images covering 11 different eye diseases, including DR, glaucoma, cataracts, and others	The raw pixel values of the eye images	The convolutional and pooling layers of the CNN model	The eye images	The final layers of the CNN model	Predict eye disease class label for the input eye image	Accuracy 91.4% Precision 0.70 Recall 0.72 F1 score 0.71	–
Eswari and Balamurali 2021[61]	Bayesian optimization support vector machine	Medical records and clinical data from diabetic patients, Ocular parameters, demographic information, and clinical history	Demographic data (age, gender) Clinical history (duration of diabetes, blood pressure, glucose, Hypertension, Hemoglobin, Diabetes Mellitus status, red blood cells), cup-to-disc ratio)	Not explicitly mention	Visual acuity, Blood pressure, Blood sugar, Diabetes duration, age, diabetes mellitus, Hypertension, CDR, family history	BOSVM, LSVM, LGSVM	Predict whether the diabetic patient is likely to develop glaucoma or not	Accuracy 96.6%, AUC 0.83 (training set), Accuracy BOSVM 97.4% (testing set) SE 0.943 SP 0.951	–

(continued)

Table 26.2 (continued)

Authors, publication year	Model	Dataset	Features	Feature extraction	Input	Classifier	Output	Performance (Internal)	Performance (External)
Yousefi et al. 2020 [51]	The AI dashboard	Out of 31,591 visual fields (VFs) on 8,077 subjects, 13,231 VFs from the most recent visit of each patient	–	–	The VF data from the most recent visit of glaucomatous and non-glaucomatous patients	An AI-enabled glaucoma dashboard that visualizes the severity and extent of functional loss and characteristic patterns of VF loss in patients with glaucoma	Determine the severity of glaucomatous vision deficit, the spatial extent of the damage as well as a means for monitoring the disease progression	For detecting likely non progression Specificity 94%, For detecting likely progression, Sensitivity 77%	–
Phene et al. 2019 [49]	Deep learning algorithm	The retrospective dataset consists of 86,618 fundus images	Vertical CDR, Notch, RNFL defect, Circumlinear vessels, Laminar dot, Disc hemorrhage, Nasalization emerging, Rim comparison, Nasalization directed, Rim comparison, b PPA	Expert graders	Vertical Cup-to-Disc Ratio, Rim width (Inferior, Superior, Temporal), Notch, Laminar dot sign, Nasalization (Emerging, Directed), Baring of circumlineal vessels, Disc hemorrhage, b PPA, RNFL defect	Deep learning Algorithm	Predict referable glaucomatous optic neuropathy and optic nerve head features from color fundus images	–	Dataset A (validated by glaucoma specialists): AUC of 0.945 Dataset B (diabetic tele-retinal screening program): AUC of 0.855 Dataset C (glaucoma clinic): AUC of 0.881

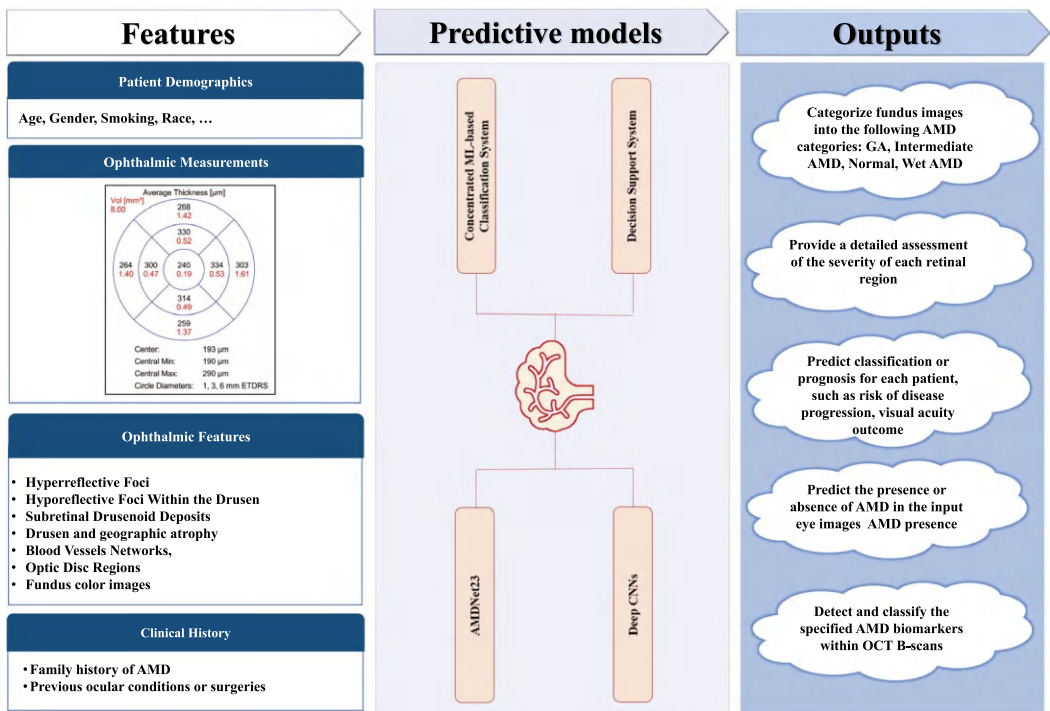


Fig. 26.3 Features, predictive models, and outputs used in studies on AI in predicting age-related macular degeneration (AMD)

predictive models, and outputs used in studies on AI in predicting age-related macular degeneration (AMD). Table 26.3 summarizes the recent articles on AI in predicting AMD.

Cataract

In 2020, an estimated 15.2 million individuals aged 50 years and older experienced vision loss as a result of cataracts on a global scale, while 78.8 million people were affected by moderate to severe visual impairments caused by cataracts [76]. Despite the increased availability of cataract surgery in the last two decades, cataracts continue to be the leading cause of blindness worldwide. There is a growing emphasis on comprehensive eye care, including preventive measures, early detection, surgical referral, and post-operative rehabilitation for cataract patients. However, there remains global inequality in addressing these needs [77]. The

increasing advancements in AI technologies, alongside the digitization of patient data and the persisting healthcare needs, have positioned AI as a pivotal tool in the future of cataract management [78]. The results of a systematic review have underscored the primary AI research in cataracts encompass the areas of screening and diagnosis, refinement of intraocular lens power calculation for cataract surgery, classification of different stages of cataract surgery, prediction of operation times to enhance operating room efficiency, and prediction of the risk of posterior capsule opacification (PCO) [79]. Tognetto et al. reported that, AI can be also utilized for operation scheduling, providing intraoperative support, and managing postoperative complications [80]. Singh et al. proposed a novel approach for early-stage cataract diagnosis using a VGG16 model based on a convolutional neural network [81]. Their results indicated that the model achieved the lowest loss rate of 0.1651 at epoch 20, demonstrating notable performance and error

Table 26.3 A summary of studies on AI in age-related macular degeneration (AMD)

Authors, publication year	Model	Dataset	Feature	Feature Extraction	Input	Classifier	Output	Performance (Internal)	Performance (External)
Abd ElKhalek et al. 2024 [72]	Concentrated ML-based classification system	More than 864 retinal images, including AMD	Age, sex, and their specific AMD category, Fundus color images	GLCM GLRLM	Fundus Images	LGBM, XGB, HGB, AdaBoost, RF, MLP, DT, LR, SVM, KNN, Top 2, Top 3, Top 4, Top 5, Top 6, Top 7, Top 8, Top 9, Top 10	Categorize fundus images into the following AMD categories: GA, Intermediate AMD, Normal, Wet AMD Provide a detailed assessment of the severity of each retinal region Aid physicians in ongoing treatment and follow-up for AMD patients	Accuracy: 96.85% Sensitivity: 93.72% Specificity: 97.89% Precision: 93.86%, F1 score: 93.72% ROC curve: 95.85% Balanced Accuracy: 95.81% Weighted sum: 95.38%	–
Borkhade and Akashe. 2024 [73]	Decision support system	The dataset were acquired from the Maghribia data set and BinRushed Dataset. For collection of real time images, we have taken approval from Advance Eye Hospital and Institute, Mumbai, India and Central Research laboratory, Datta Meghe Institute of Medical sciences, Wardha, India, to collect the images	Blood Vessels Networks, Optic Disc Regions	Mean, variance, and entropy, variance, mean, entropy, bright pixel count, and INST ratio	Retinal Images (Blood Vessels Networks, Optic Disc Regions)	Linear SVM, Fine Gaussian SVM, Quadratic SVM	Predict classification or prognosis for each patient, such as risk of disease progression, visual acuity outcome	Linear SVM Classification Accuracy (Resized images 512 * 512): 96% Classification Accuracy (Resized images 256 * 256): 92%	–

(continued)

Table 26.3 (continued)

Authors, publication year	Model	Dataset	Feature	Feature Extraction	Input	Classifier	Output	Performance (Internal)	Performance (External)
Ali et al. 2024 [74]	Hybrid CNN-LSTM deep learning approach (AMNet23)	A diverse dataset of 2000 experimental fundus images	Preprocessed retinal images for age-related macular degeneration	CNN-LSTM	Retinal images	CNN-LSTM	Predict the presence or absence of AMD in the input eye images AMD presence	AMD detection accuracy: 96.50%	–
Saha et al. 2019 [75]	Deep CNNs	Spectral-domain (SD)-OCT images of 153 patients who were diagnosed with early or intermediate AMD in at least one eye at the Doheny Eye Centers between 2010 and 2014	Hyperreflective foci, hyporeflective foci within the drusen, and other early AMD biomarkers	CNNs	Approximately 20,000 OCT B-scans from 153 Patients The Biomarkers of Interest Included: Hyperreflective Foci Hyporeflective Foci Within the Drusen Subretinal Drusenoid Deposits	Fine tuning deep CNN	Detect and classify the specified AMD biomarkers within OCT B-scans	Identify the presence early AMD pathologies Accuracy: 86–89% Identify SDD Accuracy: 80–86% Identify IHRF Accuracy: 809% Identify hRF Accuracy: 88%	–

(continued)

reduction capabilities. This study underscored the potential of deep learning techniques, such as the VGG16 model, in revolutionizing the diagnosis of eye diseases, alongside the application of advanced computational methods in medical diagnosis. The diagnostic accuracy of this technique was reported as 96.1% based on Kaggle’s cataract and eye diseases dataset [81]. Tognetto et al. study revealed that diagnostic tools leveraging AI exhibited diagnostic accuracy comparable to that of experienced ophthalmologists [80]. Notably, in certain scenarios—such as cataract diagnosis—AI outperformed human experts. As technology continues to advance, clinical decision support systems have emerged in the form of smartphone applications and AI-based imaging tools. These applications hold promise for enhancing ophthalmic care across various domains, including triage, diagnosis, and ongoing monitoring [80]. Figure 26.4 presents features, predictive models, and outputs used in

studies on AI in predicting cataract. Table 26.4 summarizes the recent articles on AI in predicting cataracts.

Myopia

Myopia emerges as a prominent global health concern, with projections indicating that by 2050, approximately half of the world’s population will be affected by this condition [86]. Once myopia sets in during adulthood, it becomes irreversible. Therefore, parents, healthcare professionals, and policymakers are increasingly focused on monitoring its progression in children to prevent it from advancing to severe or pathological stages. Anticipating the advancement of myopia can lay the groundwork for refining clinical practices, shaping health policies, and tailoring personalized interventions to effectively manage myopia in school-aged

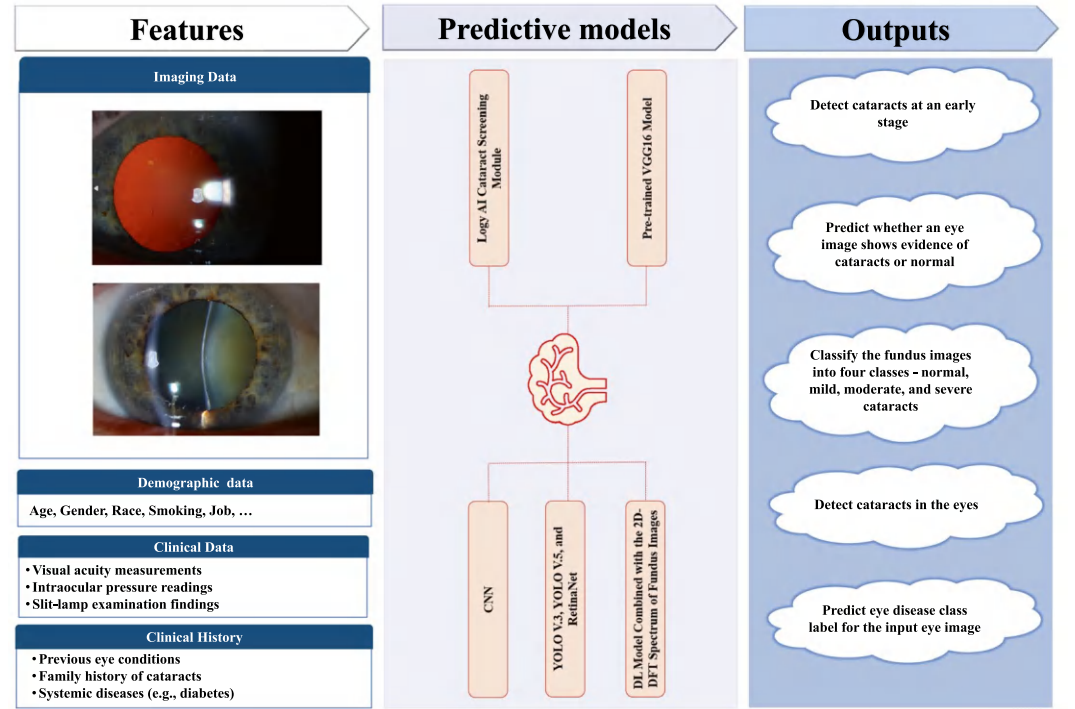


Fig. 26.4 Features, predictive models, and outputs used in studies on AI in predicting cataract

Table 26.4 A summary of studies on AI in predicting cataract

Authors, publication year	Model	Dataset	Feature	Feature extraction	Input	Classifier	Output	Performance (Internal)	Performance (External)
VM et al. 2024 [82]	Logy AI cataract screening module	794 eye images	Smartphone images captured by patients, age, gender	CNN	Smartphone images captured by patients, age, gender	Dense layer	Detect cataract at an early stage	Accuracy cataract detection: 90.08% Sensitivity: 90.38% Specificity: 89.87% F1 score: 87.98% PPV: 85.71% NPV: 93.29% AUC: 0.8946 Accuracy Immature Cataracts: 88.02% Accuracy mature cataract: 97.16% Accuracy normal category:90.08%	
Ueno et al. 2024 [83]	You Only Look Once V.3 (YOLO V.3), YOLO V.5, and RetinaNet	6106 images used to develop the AI model after excluding duplicates and poor-quality images	Not explicitly detailed, but they involve the characteristics of various corneal diseases and cataracts as captured in anterior segment photographs	YOLO V.3 and V.5 use Darknet-53 and its improved version for feature extraction, while RetinaNet uses ResNet-101	Anterior segment photographs captured using slit-lamp microscopy and smartphone cameras	YOLO V.3, YOLO V.5, and RetinaNet algorithms	Classify 36 eye diseases into 9 categories, along with a predictive score for each category	AUC ROC: 0.992 Sensitivity 95% CI: 0.429 (0.099 to 0.816) Specificity 95% CI:1.000 (0.971 to 1.000) Urgent: 1.00 Semi-urgent: 0.867 Routine: 0.853 Observation: 1.00	

(continued)

Table 26.4 (continued)

Authors, publication year	Model	Dataset	Feature	Feature extraction	Input	Classifier	Output	Performance (Internal)	Performance (External)
Singh et al. 2023 [81]	A pre-trained VGG16 Model	Images of eyes with both normal instances and those with cataracts obtained from Kaggle	Images of eyes	VGG16	Preprocessed eye images downsized to a standard size such as 224×224 pixels	CNN	Predict whether an eye image shows evidence of cataracts or normal	Accuracy 96.10% Precision 100% Recall 93% F1 score 96.40%	–
Yadav and Yadav 2023[84]	DL model combined with the 2D- DFT spectrum of fundus images	Randomly selected retinal images (n = 1835) from various open-source databases and images obtained from the Internet	Tiny blood vessels in fundus images	CNN model	2D DFT spectrograms of fundus images	Softmax classifier	Classify the fundus images into four classes - normal, mild, moderate, and severe cataracts	Accuracy: 93.10% Precision 93.09% Sensitivity: 93.13% Specificity: 97.71% F1 score: 93.08%	–
Chakraborty and Jana 2023 [85]	Convolutional Neural Network Model	1) a structured ophthalmic dataset of 5,000 patients with age, and color fundus images from both eyes 2) the retinal fundus images of the cataract eyes and normal eyes	Pixel values of the images, patterns in the images related to cataract symptoms	Convolutional layers in the CNN	Digital images of eye conditions (perhaps grayscale or RGB images)	CNN model	Detect cataract in the eyes	Accuracy overall: 97.47% Precision: 96.64% Recall: 98.29% F1 score: 97.46%	
Helen and Gokila 2023 [23]	Convolutional Neural Network (CNN)	11,929 eye images covering 11 different eye diseases, including DR, glaucoma, cataracts, and others	The raw pixel values of the eye images	The convolutional and pooling layers of the CNN model	The eye images	The final layers of the CNN model	Predict eye disease class label for the input eye image	Accuracy: 92.1%, precision: 80%, Recall: 71 %, F1 score 78%	

children [87]. Hashemi et al. showed that changes in lens power and axial length occur during the period of emmetropization [88].

Factors such as gender, age of onset, severity of myopia at onset, prolonged near work, less time outdoors, size of myopia in environment, and parental myopia are related to myopia progression [89]. These findings aid in identifying at-risk children, enabling them to receive appropriate treatment and lifestyle counseling [89]. It is essential to assess and monitor patients with mild to moderate myopia clinically. This process involves gathering data and records that include myopia-related risk factors such as near work time, gender, outdoor activity time, and genetic and racial factors, as well as best corrected visual acuity, refraction, axial length, and other eye measurements (e.g. intraocular pressure and ocular surface conditions). Analyzing and interpreting this data is complex, compounded by the limited of reliable risk prediction models for monitoring and prediction of myopia progression. Additionally, due to the large volume and complexity of the data, manual analysis is challenging [90].

Conventional machine learning methods possess the capability to process large amounts of data in a nonlinear manner and extract numerous potential predictor variables, making them suitable for myopia data analysis [91]. Yang et al. developed a risk prediction model using SVM, which showed high accuracy (93%) in predicting myopia in the future [92]. Li et al. focused on elementary school students over a 5-year period and found that uncorrected farsightedness, spherical equivalent, axial length, flat keratometry, gender, and parental myopia were significant in predicting myopia progression. They underscored the necessity of annual monitoring for children with myopia, emphasizing the dominance of eye-related factors over environmental and genetic factors [93]. Lin et al. employed a ML model to forecast the onset and progression of myopia up to 8 years into the future. The researchers observed that a substantial volume of data enhanced the performance of ML, and that the choice of the target variable could be tailored to clinical needs, such as the

utilization of axial length rather than refraction examinations for children using orthokeratology lenses [94]. Furthermore, Tang et al. demonstrated that a robust regression model achieved high accuracy in predicting axial length growth and stressed the importance of leveraging data from electronic medical records to enhance prediction models. These findings advocate for integrating behavioral data of children with myopia with clinical data to facilitate intelligent interventions for the prevention and management of myopia [95]. The integration of wearable devices equipped with AI algorithms in Mrochen et al. study aimed to investigate the behavioral patterns and vision habits of children with myopia has yielded promising results. The findings suggest that the Vivid monitor has the potential to facilitate the design and implementation of more effective interventions for managing and treating myopia in children [96].

Prior to the widespread adoption of AI in myopia management, it is imperative to prioritize the optimization of AI models to enhance their interpretability, human-machine interactions, generalization ability, as well as their validity and efficacy. This necessitates a concerted effort to refine clinical standards, incorporate large-scale clinical datasets, and establish a standardized evaluation framework for AI models in clinical settings. Moreover, the development and reinforcement of pertinent regulations are crucial for comprehensive oversight of the practical applications of AI in myopia [87].

Artiemjew et al. developed a myopia risk calculator as a pioneering force in the evolving landscape of healthcare. They evaluated the performance of a myopia risk calculation model using a robust methodology, combining Monte Carlo simulations and cross-validation techniques. The results demonstrate that using a combination of classifiers and regression models, specifically employing lasso regression techniques, resulted in optimal myopia risk prediction. Further investigations by integrating medical images and synthetic data through advanced Monte Carlo simulations are deemed necessary [97]. Furthermore, the utilization of novel technologies in addressing the unresolved

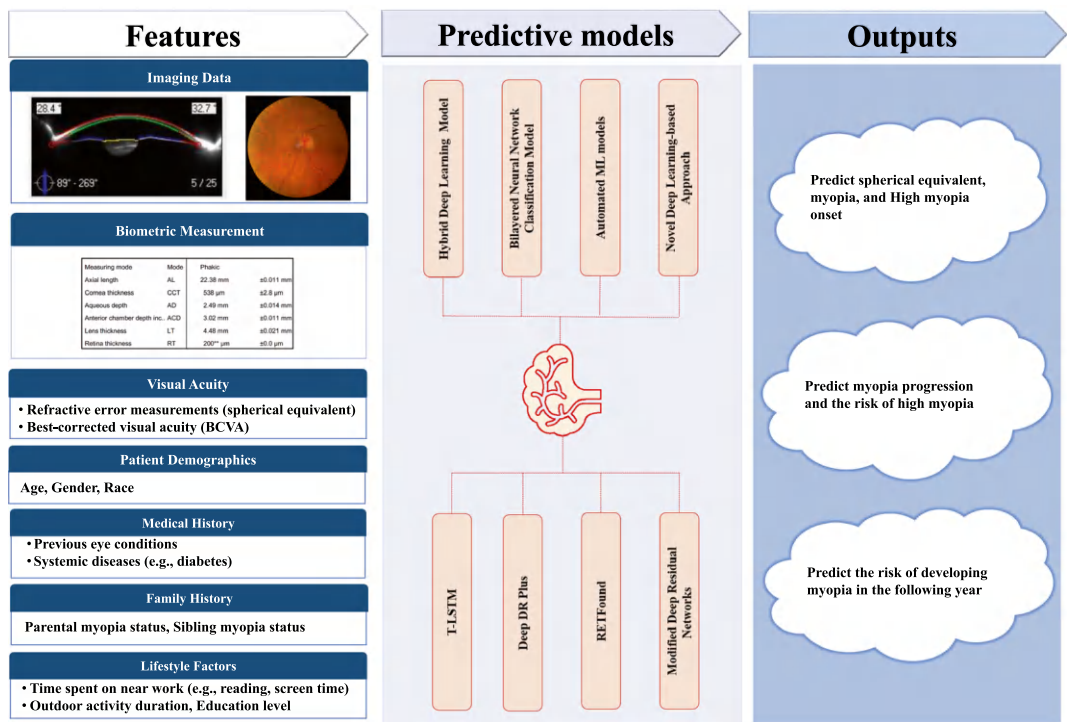


Fig. 26.5 Features, predictive models, and outputs used in studies on AI in predicting myopia

challenges associated with myopia management is advocated [87]. The paramount influence of AI technology in the analysis of medical images and data, particularly in addressing the complexities of eye diseases is also emphasized. Figure 26.5 presents features, predictive models, and outputs used in studies on AI in predicting myopia. Table 26.5 summarizes the recent articles on AI in predicting Myopia.

Keratoconus

The prevalence of keratoconus ranges from 0.2 to 4.8 individuals per 100,000 population [105]. Early detection of this condition facilitates the monitoring of disease progression and enables timely interventions to prevent its advancement [106].

The application of AI in the screening and early diagnosis of corneal diseases has shown promising results [105]. Notably, the use of AI algorithms in keratoconus screening was first

proposed in 1997 by Smolek et al., who utilized a neural network to analyze corneal topographic image and discovered that the neural network can differentiate between keratoconus and sub-clinical keratoconus [107]. Subsequent studies have demonstrated the effectiveness of AI, particularly deep learning models, in accurately identifying and classifying keratoconus based on corneal topography images. These models exhibited high sensitivity and specificity, with the ResNet model achieving a very high accuracy in keratoconus classification (AUC: 0.995) [108]. Mohammadpour et al. compared the performance of AI-based classifiers and clinical experts in distinguishing normal corneas and ectasia states. They found that AI-based classifiers, particularly Sirius Phoenix, exhibited the highest agreement with clinical diagnosis in the timely identification of keratoconus [109]. Furthermore, compared to other classifiers, Phoenix demonstrated the highest accuracy in distinguishing keratoconus from subclinical keratoconus. The researchers emphasized

Table 26.5 A summary of studies on AI in predicting myopia

Authors, publication year	Model	Dataset	Feature	Feature extraction	Input	Classifier	Output	Performance	External validation
Zhao et al. 2024 [98]	RF+ XGBoost	Electronic medical records of pediatric and adolescent patients who underwent cycloplegic refraction measurements at the Eye and Ear, Nose, and Throat Hospital of Fudan University between January 2005 and December 2019. Included 88,250 individuals with 408,255 refraction records	Age at baseline and spherical equivalent (SE) at various intervals	The features were extracted from the electronic medical records	Age and SE	RF and XGBoost	Predict SE, myopia, and High myopia onset	Myopia onset prediction AUC: 0.845 - 0.953 over 15 years High myopia onset AUC (0.807-0.997 +13 years, with the 14th year at 0.765)	-
Li et al. 2024 [99]	A ML approach to predict myopia progression and the risk of developing high myopia in children	15,512 elementary school students from 17 cities in Hubei Province, China	Visual acuity, cycloplegic autorefraction, height and weight measurements, and information collected through questionnaires (e.g., parental myopia, education level, eye-related habits)	Features were directly collected	The various measurements and questionnaire responses	XGB, KNeighbors, Decision Tree, LR, GussianNB	Predict myopia progression and the risk of high myopia	AUC _{SG} : 0.96 AUC _{KNeighbors} : 0.93 AUC _{Decision Tree} : 0.92 AUC _{LR} : 0.85 AUC _{GaussianNB} : 0.8	-
Li et al. 2024 [100]	Two ML models (a multivariate linear regression model to predict the progression of myopia, and a logistic regression model to predict the risk of high myopia	Data gathered from five independent cohorts, encompassing 612,530 medical records from 227,543 patients ranging from infants to young adults	Age at baseline, spherical equivalent (SE) at baseline, and the time interval between baseline and follow-ups	The features were extracted directly from the medical records	612,530 medical records from five independent cohorts, covering patients ranging from infants to young adults	The multivariate linear regression algorithm for predicting the progression of myopia and the logistic regression algorithm for predicting the risk of high myopia	Predict progression of myopia and the risk of high myopia	R2 0.964 MAE 0.119D AUC 0.99 R2 S ₁ : 0.950 MAE S ₁ : 0.119D R2 S ₂ : 0.950 MAE S ₂ : 0.121D R2 S ₃ : 0.99 MAE S ₃ : 0.806 MAE S ₄ : -0.066D AUC S ₄ : 0.96 In the Beijing Children's Eye Study R ² : 0.749 MAE: 0.178D AUC: 0.99	

(continued)

Table 26.5 (continued)

Authors, publication year	Model	Dataset	Feature	Feature extraction	Input	Classifier	Output	Performance	External validation
Foo et al. 2023 [101]	A modular deep learning system (DLS) based on single-time point objective data and fundus imaging algorithms. Three distinct algorithms: Image Model, Clinical Model, Mixed Model Image+Clinical)	A school-based cohort in Singapore, comprising 998 children aged 6–12 years old at baseline: The primary dataset included 7456 baseline fundus images of 1878 eyes An independent test dataset consisted of 821 baseline fundus images of 189 eyes, along with clinical data	Fundus images and clinical data (age, gender, race, parental myopia, and baseline spherical equivalent	DenseNet-121 deep neural network models, RF. Features extracted from image data, were combined with the clinical data-based scores to derive an overall mixed-model risk score	- Image model (fundus images) - Clinical model (age, gender, race, parental myopia, and baseline spherical equivalent) Mixed Model (Image+Clinical): fundus images and clinical data	-	Predict high myopia development in school-going children potentially serves as a clinical decision-support tool for early intervention	Image model Fundus photo AUC: 0.93 Accuracy: 84% Sensitivity: 92% Specificity: 84%, Race, Gender AUC: 0.95 Accuracy: 86%, Sensitivity: 85%, Specificity: 86%, Clinical Model Baseline SE, AUC: 0.90(0.84–0.96) 0.80 (0.80–0.80) 0.77 (0.77–0.77) 0.80 (0.80–0.80) Baseline SE, Age, Race, Gender AUC: 0.97 Accuracy: 87% Sensitivity: 85% Specificity: 87% Baseline SE, Age, Race, Gender 1-year SE progression AUC: 0.98 Accuracy: 92% Sensitivity: 92% Mixed Model Fundus photo, Baseline SE AUC: 0.97, Accuracy: 82%, Sensitivity: 100%, Specificity: 80%, Fundus photo, Baseline SE, Age, Race, Gender AUC: 0.98, Accuracy: 82%, Sensitivity: 100%, Specificity: 80%, Fundus photo, Baseline SE, Age, Race, Gender, 1-year SE progression AUC: 0.95, Accuracy: 88%, Sensitivity: 94%, Specificity: 88%	Image model Fundus photo AUC: 0.93 Accuracy: 86% Sensitivity: 88% Specificity: 86% Fundus photo+Age+Race+Gender AUC: 0.91 Accuracy: 78% Sensitivity: 88% Specificity: 77%, Clinical Model Baseline SE AUC: 0.93, Accuracy: 65%, Sensitivity: 100%, Specificity: 61%, Baseline SE, Age, Race, Gender AUC: 0.94, Accuracy: 74%, Sensitivity: 100%, Specificity: 72% Baseline SE, Age, Race, Gender, 1-year SE progression AUC: 0.97, Accuracy: 84%, Sensitivity: 100%, Specificity: 82%, Mixed Model Fundus photo, Baseline SE AUC: 0.97, Accuracy: 82%, Sensitivity: 100%, Specificity: 80%, Fundus photo, Baseline SE, Age, Race, Gender AUC: 0.98, Accuracy: 82%, Sensitivity: 100%, Specificity: 80%, Fundus photo, Baseline SE, Age, Race, Gender, 1-year SE progression AUC: 0.95, Accuracy: 88%, Sensitivity: 94%, Specificity: 88%

(continued)

Table 26.5 (continued)

Authors, publication year	Model	Dataset	Feature	Feature extraction	Input	Classifier	Output	Performance	External validation
Huang et al. 2023 [102]	Time-Aware Long Short-Term Memory (T-LSTM)	Examined uncorrected visual acuity, sphere, astigmatism, axis, corneal curvature, and axial length of 75,172 eyes from 37,586 children and adolescents aged 6–20 years in Chengdu, China	Uncorrected visual acuity, sphere, astigmatism, axis, corneal curvature, and axial length	T-LSTM	Historical vision records (variable-length time series data)	Extracted features from historical vision records	Quantitatively predict spherical equivalent (SE) within two and a half years	The mean absolute prediction error for spherical equivalent on the testing set was 0.140 ± 0.103 diopters. This error exhibited variation across different lengths of historical records and prediction durations, spanning from 0.050 ± 0.040 diopters to 0.168 ± 0.187 diopters	
Chen et al. 2023 [103]	Logistic regression models	Data from the Guangzhou Twins Eye Study (GTES), Guangzhou Outdoor Activity Longitudinal Study (GOALS), and Singapore Cohort Study of the Risk factors for Myopia (SCORM) were used	Age, sex, cycloplegic refraction, ocular biometry (axial length, corneal curvature radius, anterior chamber depth), uncorrected visual acuity, and parental myopia history	Not explicitly detailed, but involves the collection of ocular biometry and cycloplegic refraction data	Age, sex, cycloplegic refraction, ocular biometry (axial length, corneal curvature radius, anterior chamber depth), uncorrected visual acuity, and parental myopia history	LR	Predict risk of developing myopia in the following year	Accuracy 92%, AUC 0.78–0.81, C-statistic 95% CI Age, sex, SE, myopic parent 0.91 0.87–0.94, Age, sex, AL, CR, ACD, UCVA, myopic parent 0.81 0.73–0.88, Age, sex, AL/CR, ACD, UCVA, myopic parent 0.78 0.71–0.85	Accuracy 93% AUC 0.76–0.80, C-statistic 95% CI Age, sex, SE, myopic parent 0.92 0.92–0.92 Age, sex, AL, CR, ACD, UCVA, myopic parent 0.80 0.79–0.80 Age, sex, AL/CR, ACD, UCVA, myopic parent 0.76 0.75–0.76
Li et al. 2022 [93]	RF	2740 children from grade 1 to grade 6 were examined annually over a 5-year period	Uncorrected distance visual acuity (UDVA), spherical equivalent, axial length, flat keratometry reading (K1), gender, and parental myopia	Questionnaires	Longitudinal, cycloplegic autorefraction data	RF	Predict myopia progression	Accuracy >80% for myopia progression	-
Wong et al. 2021 [104]	Multivariable Logistic Regression	-	-	-	Non-cycloplegic refraction, ocular biometry, accommodation measurements, and other factors such as gender, age, parental myopia, near-work time, and outdoor time	LR	Binary classification of myopia (spherical equivalent ≤ 0.5 diopter)	Predicting myopia onset AUC _{SEA} : 0.66 AUC _{SE} : 0.64 AUC _{AL} : 0.62 AUC _{Age, gender, parental myopia, SE, AL and PRA} : (0.74) AUC _{Age, gender, Parental Myopia, SE, and AL} : (0.70) AUC _{Age, Gender, Parental Myopia, And SE} : (0.65)	-

that while AI classifiers can assist in diagnosis, they cannot substitute the expertise of clinical experts, particularly prior to refractive surgery. Additionally, advancements in AI have the potential to facilitate more precise and timely diagnosis of keratoconus and support less experienced doctors [109]. Maeda et al. detailed the application of AI and neural networks in diagnosing and evaluating seven types of video keratography: normal, astigmatism (with-the-rule), keratoconus (in three stages), keratectomy post-photorefractive, and post-keratoplasty. The findings showed that for each category, accuracy and specificity exceeded 90%, while sensitivity ranged from 44 to 100% [110]. The diagnostic systems employed automated decision trees, support vector machines, and various neural network architectures. Data inputs were derived from diverse corneal imaging devices [110]. While it may take time to integrate AI techniques into corneal imaging devices, their use at the bedside is gaining popularity. Most reviewed studies indicate a strong ability to differentiate between normal cases and keratoconus, with a relatively lower ability for subclinical keratoconus [111].

A new diagnostic model named CorNet, which utilizes CORVIS ST raw data for keratoconus detection was developed [112]. The study demonstrated the model's effectiveness in learning temporal and spatial features independently from raw data, with an AUC of 0.971. This research supports the use of biomechanical data for keratoconus diagnosis, potentially enhancing early detection and treatment outcomes in ophthalmology [112].

A systematic review and meta-analysis on the application of AI algorithms using topography, tomography, biomechanical, and optical deviation data, as well as evaluating corneal layers, revealed the importance of identifying diagnostic limitations for timely keratoconus diagnosis. Integrating data from Placido and Scheimpflug systems enhances machine learning and deep learning models' performance. Researchers emphasized on the need for multicenter studies to build stronger AI models with larger datasets [113]. They indicated that future solutions for

timely ectatic cornea detection involve developing AI methods and corneal imaging technologies, but improving diagnostic quality based on AI models is crucial [113]. The potential of ChatGPT to revolutionize the ophthalmology, spans across diverse areas, including patients, healthcare professionals/systems, research, and education/training [114]. For instance, Delsoz et al. assessed the accuracy of ChatGPT-4 and ChatGPT-3.5 in initial diagnosis of different corneal disorders and compared the model performance with three cornea specialists [115]. ChatGPT-4.0 demonstrated significantly higher accuracy in diagnosing patients with various corneal conditions compared to ChatGPT-3.5. The diagnostic agreement between ChatGPT-4.0 and ChatGPT-3.5 for corneal conditions was 65%. However, when comparing ChatGPT-4.0 with three cornea specialists, the agreement levels were 85, 80, and 75% respectively [115]. Figure 26.6 presents features, predictive models, and outputs used in studies on AI in predicting keratoconus. Table 26.6 summarizes the recent articles on AI in predicting keratoconus.

Prediction of Systemic Diseases Based on Eye Data

For decades, researchers have been looking for retinal changes to identify ocular biomarkers for systemic diseases [122]. Chronic and non-communicable diseases present significant barriers to achieving a long and healthy life. Timely diagnosis can facilitate prevention, monitoring, and treatment, leading to improved patient outcomes. Therefore, efforts to enhance the accessibility, simplicity, and cost-effectiveness of screening techniques are crucial. The correlation between eye biomarkers and systemic diseases, as well as oculomics health, offers promising opportunities for diagnosing systemic diseases due to the relatively low-cost, quick, and non-invasive nature of eye techniques [123]. Oculomics involves the use of big data, AI, and eye imaging to detect retinal biomarkers of systemic disease [124]. The accessible yet complex visual image datasets

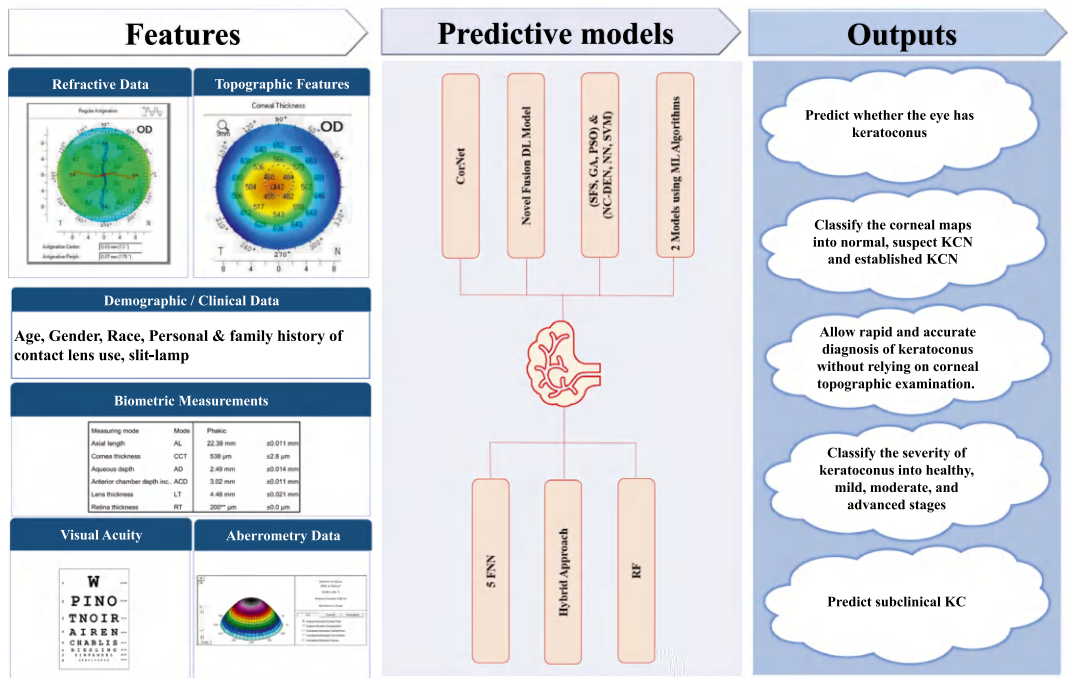


Fig. 26.6 Features, predictive models, and outputs used in studies on AI in predicting keratoconus

can be effectively analyzed using AI, offering a non-invasive method for predicting systemic diseases by assessing risk factors, retinal characteristics, and biomarkers. Most current algorithms for diagnosing systemic diseases focus on retinal imaging, while other modalities such as OCT, and OCT angiography hold great promise for further research [122]. Peng et al. found that previous studies have demonstrated that AI-driven analysis of eye images can effectively predict various diseases and systemic disease risk factors, highlighting its potential as a powerful tool for screening systemic diseases [125]. Recent research has demonstrated the use of AI to analyze eye images for the prediction of various cardiovascular diseases. These include atherosclerosis of the carotid artery [126, 127], congenital heart disease [128], serious adverse cardiovascular events [129], coronary artery calcium score [127, 130], coronary artery disease [131], ischemic stroke [132], and high blood pressure [133, 134].

Moreover, AI has proven to be effective in analyzing eye images to predict peripheral

neuropathy caused by diabetes [135], and diabetes [136–139]. Chang et al. developed a deep learning model to predict arteriosclerosis by examining retinal fundus images. They discovered that these images also have the potential to predict atherosclerosis of carotid arteries, and that this prediction could enhance traditional risk classification scores such as the Framingham risk score for cardiovascular mortality [126]. Furthermore, the study by Poplin et al. demonstrated that retinal fundus image analysis using deep learning alone can predict multiple cardiovascular risk factors, including age, gender, and systolic blood pressure. Since these risk factors are essential components of many cardiovascular risk calculators, it is evident that this model has the capability to directly predict cardiovascular risk [129]. Son et al. study also revealed identifiable patterns in retinal vessels and fovea for predicting coronary artery calcium score [127]. Prakash et al. emphasized the potential value of analyzing retinal images in identifying cardiovascular diseases, as the researchers found that a low

Table 26.6 A summary of studies on AI in predicting keratoconus

Authors, publication year	Model	Dataset	Feature	Feature Extraction	Input	Classifier	Output	Performance internal	External validation
Zhang et al. 2024 [112]	CorNet (autonomously learns comprehensive temporal and spatial features from raw Corvis ST data)	1786 Corvis ST raw data (1112 normal eyes and 674 keratoconus (KC) eyes)	The raw biomechanical data obtained from Corvis ST These data points included corneal surface elevation measurements during deformation	Four ResNet-inspired convolutional structures	Raw Corvis ST data Each raw data sample includes anterior and posterior corneal surface elevation during air-puff-induced dynamic deformation	End-to-end CorNet model	Predicts whether the eye has keratoconus	AUC: 0.971 Sensitivity: 92.49% Specificity: 91.54%	–
Al-Timemy et al. 2023 [116]	Novel fusion deep learning (DL) method based on Xception and InceptionResNetV2 architectures	Two datasets were used, one collected in Egypt and another in Iraq, comprising images from normal, suspect, and established keratoconus (KCN) cases	Sagittal (Sag), corneal pachymetry (CorPac), and elevation front (ElevF) maps	Fine-tuned on the corneal map data	3 * 299 * 299 sized images from the corneal maps	SVM, DT	Classify the corneal maps into normal, suspect KCN and established KCN	AUC: 0.99, Accuracy>97%	AUC 0.91–0.92 Accuracy 88–92%
Smily Yeya Jothi et al. 2023 [117]	(SFS, GA, PSO) and (NC-DEN, NN, SVM)	Taken from Devottam Gaurav et al. study (Review of Machine Learning Approach for Drug Development Process)	446 features (not mentioned)	SFS, GA, PSO	-	NC-DEN, NN, SVM	Identify the disease at an early stage	Average Accuracy: 95% Sensitivity: 95% Specificity: 94%	–

(continued)

Table 26.6 (continued)

Authors, publication year	Model	Dataset	Feature	Feature Extraction	Input	Classifier	Output	Performance internal	External validation
Tan et al. 2022 [118]	5-FNN	–	Age, gender, visual acuity, objective and manifest refraction, slit-lamp microscopy, noncontact intraocular pressure, funduscopy, corneal morphology, and corneal biomechanics	–	Age, gender, visual acuity, objective and manifest refraction, slit-lamp microscopy, noncontact intraocular pressure, funduscopy, corneal morphology, and corneal biomechanics	–	Allow rapid and accurate diagnosis of keratoconus without relying on corneal topographic examination	Accuracy: 99.6% Sensitivity: 99.3% Specificity: 100% Precision: 100%	Accuracy: 98.7% Sensitivity: 97.4% Specificity: 100% Precision: 100%
Alyasser et al. 2022 [119]	A hybrid approach (ML+DL)	5430 eyes at different stages of KCN severity from the Department of Ophthalmology and Visual Sciences, Paulista Medical School, Federal University of São Paulo, São Paulo, Brazil 1531 eyes from the Department of Ophthalmology, Jichi Medical University, Tochigi, Japan	Topography, tomography, pachymetry, anterior chamber angle, volume, height, corneal and lens densitometry, and other ocular indices	Not explicitly mentioned	Topography, tomography, pachymetry, anterior chamber angle, volume, height, corneal and lens densitometry, and other ocular indices	FPA and k-means algorithms	Detect keratoconus or normal eyes	Accuracy > 95% Sensitivity > 90% Specificity > 95%	Accuracy > 95% Sensitivity > 90% Specificity > 95%

(continued)

Table 26.6 (continued)

Authors, publication year	Model	Dataset	Feature	Feature Extraction	Input	Classifier	Output	Performance internal	External validation
Herber et al. 2021 [120]	Two models using ML algorithms (LDA and RF), with the RF model being optimized into the DK1	318 keratoconic eyes and 116 healthy eyes	DCR and corneal thickness-related parameters from Corvis ST tonometry	Recursive feature elimination for RF and stepwise Wilks-Lambda for LDA	The selected DCR and pachymetric parameters from the Corvis ST device	LDA, RF	Classification of the severity of keratoconus into healthy, mild, moderate, and advanced stages	LDA healthy eyes Sensitivity 82%, Specificity 97%, Mild KC eyes Sensitivity 73%, Specificity 81% Moderate KC Sensitivity 62%, Specificity 83% Advanced KC Sensitivity 68%, Specificity 95% RF model: Healthy eyes Sensitivity: 91% Specificity: 94% Mild KC Sensitivity: 80% Specificity: 90% Moderate KC Sensitivity: 63% Specificity: 87% Advanced KC Sensitivity: 72% Specificity: 95% Overall Accuracy LDA: 71% Overall Accuracy RF: 78% KC detection (all subgroups of KC severity in both models) Accuracy 93%	–

(continued)

Table 26.6 (continued)

Authors, publication year	Model	Dataset	Feature	Feature Extraction	Input	Classifier	Output	Performance internal	External validation
Castro-Luna et al. 2021 [121]	RF		Age, BSCVA, Sph eq, KMAX, CCT, IoP, HOA RMS, corneal biomechanical data		Age, BSCVA, Sph eq, KMAX, CCT, IoP, HOA RMS, corneal biomechanical data	RF	Predict subclinical KC	RF Specificity: 93% Sensitivity: 86% Overall accuracy rate: 89% DT Sensitivity: 50% Specificity: 93% Balanced accuracy: 71%	

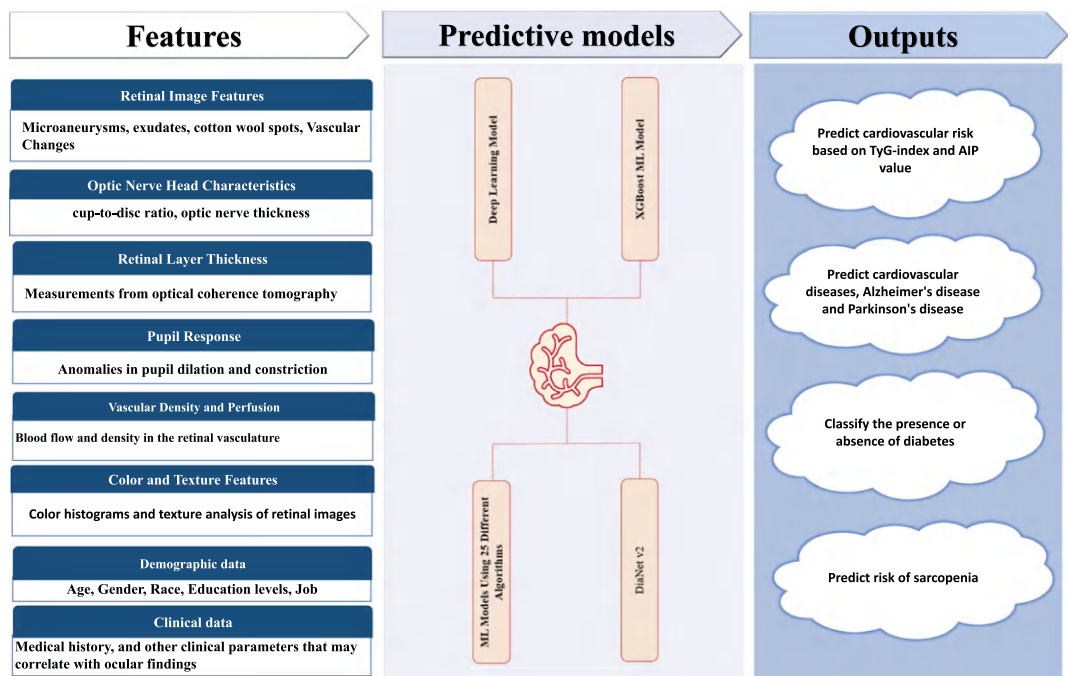


Fig. 26.7 Features, predictive models, and outputs used in studies on AI in predicting systemic diseases

arteriolar to venular diameter ratio (AVR) is valid evidence of an increased risk of cardiovascular diseases [140]. RETFound has the capability to accurately detect various diseases from retinal images. Its ability to predict systemic diseases, such as ischemic stroke, myocardial infarction, heart failure, and a neurodegenerative disease like Parkinson’s disease over a 3-year period, is highlighted in a recent study. By adapting RETFound to various challenging detection and prediction tasks, it has demonstrated superior performance compared to existing models, showcasing its versatility and broad disease detection capabilities [33]. Figure 26.7 presents features, predictive models, and outputs used in studies on AI in predicting systemic diseases. Table 26.7 summarizes the recent articles on AI in predicting systemic diseases.

Conclusion and Future Horizons

The meta-analysis of research evidence revealed that AI has sparked a revolution in ophthalmology by achieving high diagnostic accuracy in the analysis of eye images [143]. The compelling performance of several retrospective DR screening validation studies utilizing deep learning for automated DR screening based on retinal color images has sparked considerable enthusiasm for the integration of AI in healthcare, thereby expanding its scope beyond ophthalmology [19]. The assessment of 21 AI algorithms utilized in automatic DR screening indicated differing levels of performance [17]. Given the availability of reliable and high-quality algorithms, it is essential to conduct external validation of these AI algorithms under real-world conditions

Table 26.7 A summary of studies on AI in predicting systemic diseases on ophthalmic data

Authors, publication year	Model	Dataset	Feature	Feature extraction	Input	Classifier	Output	Performance	External validation
Zhang et al. 2024 [124]	ML models using 25 different algorithms	Korean National Health and Nutrition Examination Survey	Demographic, anthropometric, and ophthalmological measurements	AutoFeat for automated feature engineering and Spearman correlation analysis for dimensionality reduction	Clinical questionnaires and ophthalmic measurements	Gradient Boosting, LGBM, Random Forest, SVM	Predict cardiovascular risk based on TyG-index and AIP value	TyG-index cut-offs at 8.0, 8.75, and 8.93 AUC: 0.812, 0.873, and 0.911, For the atherogenic index of plasma (AIP) at 0.34, AUC 0.849, For the 0.318 cut-offs, AUCs: 0.844 Notably, gender-based variations were observed for certain thresholds. For TyG-index at AUC male:8.0 AUC female:8.75	TyG-index cut-offs at 8.0, 8.75, and 8.93 AUCs: 0.809, 0.863, and 0.901) For the atherogenic index of plasma (AIP) at 0.34, AUC 0.849, For the 0.318 cut-offs, AUC: 0.842 For the 0.318 cut-offs, AUC: 0.836 Notably, gender-based variations were observed for certain thresholds For TyG-index at AUC male:8.0 AUC female:8.75
Prakash et al. 2024[140]	Deep learning model	Retinal images of patients with cardiovascular disease	Abnormalities in retinal images	UNET (U-shaped convolutional neural network)	Abnormalities in retinal images	ResNet-50 and VGG-16	Predict cardiovascular disease based on retinal abnormalities	Overall accuracy VGG16 97.46% Overall accuracy ResNet50 95.32%	–

(continued)

Table 26.7 (continued)

Authors, publication year	Model	Dataset	Feature	Feature extraction	Input	Classifier	Output	Performance	External validation
Al-Absi et al. 2024 [141]	DiaNet v2 (DL model)	15,011 retinal images from the Qatar Biobank (QBB) and Hamad Medical Corporation (HMC)	Retinal images, focusing on areas such as the optic disc, macula, and intermediate zone	VGG-11 network	Pre-processed retinal images	DenseNet-121, ResNet-50, EfficientNet, VGG-11, and MobileNet_v2	Classify the presence or absence of diabetes	Accuracy > 92% Sensitivity > 93% Specificity > 91%	–
Kim et al. 2022 [142]	XGBoost ML model	Korean National Health and Nutrition Examination Survey	Ophthalmological examinations and demographic factors	SHapley Additive exPlanations	Ocular measurements and demographic details	XGBoost technique	Prediction risk of sarcopenia	AUC _{male} : 0.746 AUC _{female} : 0.762	AUC _{male} : 0.751 AUC _{female} : 0.785

before their clinical implementation [17]. A study by Thakhur et al. revealed some limitations associated with available datasets, the absence of a standardized definition of disease progression, variations in methodologies, and diverse forecasting approaches. Consequently, the implementation of AI in glaucoma management remains a complex and evolving endeavor [52].

Several concerns surround the use of AI in ophthalmology, including safety issues, cybersecurity, ownership of health data and AI algorithms, black box problems, and the risk of exacerbating healthcare inequality. These concerns can be mitigated by collecting high-quality data from the target population, implementing stronger security protocols, and evaluating algorithms under real-world conditions. Although researchers are eager to develop and assess AI models for predicting and identifying eye diseases, not all algorithms have obtained regulatory approval to become commercial products [144]. It is crucial to validate AI algorithms prior to their use in medical settings. In Europe, medical AI algorithms and software must comply with European medical device laws and obtain proper certification to be used in the European market. These certifications are classified into four levels: I, IIa, IIb, and III, based on the perceived risk level. Class I certification involves a self-certification process without external oversight or verification [145]. AI, particularly deep learning, has exhibited impressive capabilities in automating the analysis and interpretation of retinal images. It has demonstrated effectiveness in diagnosing and quantifying various retinal pathologies, including DR, AMD, glaucoma, and retinal vascular disorders. AI can serve as an intelligent assistant in disease screening, diagnosis, and monitoring. However, there are considerable challenges related to data availability and diversity that need to be addressed to ensure the validity and generalizability of AI models. The practical application of AI-based retinal analysis at the point of care, along with the development of regulatory frameworks for the deployment of intelligent technology, are essential for widespread adoption. The successful implementation

of AI in retinal imaging depends on close collaboration between ophthalmologists and computer specialists for design and validation, as well as the responsible use of these technologies [146]. Despite the extensive studies on the application of AI in retina analysis, the detection of structures, lesions, and disease diagnosis from fundus photographs and OCT images have primarily involved deep learning algorithms. The focus remains on striving to reach clinical specialist-level diagnostic performance [147]. In a comprehensive study, researchers underscored the valuable role of image analysis in training AI algorithms for diagnosing eye diseases. They emphasized the significant impact of AI technology on the clinical landscape for both ophthalmologists and patients. However, the study also outlined several limitations, including concerns about the validity of educational models and the underutilization of highly sensitive and specific AI tools at the bedside. Furthermore, the researchers noted the predominant focus on well-known diseases, neglecting the lesser-known eye conditions. They also highlighted the absence of a fully developed diagnostic system [148]. The approach of researchers has shifted from developing AI algorithms and analyzing abnormal physiological eye structures to focusing on researching eye disease diagnosis systems. Overcoming existing limitations requires gathering more international ophthalmic data for algorithm training and testing. In today's dynamic world, AI systems must be capable of continuous learning to maintain adaptability. Incorporating techniques such as gradient-based learning, modular neural networks, and meta learning can enable AI systems to continuously learn throughout their life cycle. These methods can enhance learning efficiency and facilitate the transfer of AI knowledge to related tasks [146]. Through advancements in AI, machine learning, deep learning, and big medical data, it is now possible to diagnose a range of disorders by analyzing eye images, including endocrine, cardiovascular, neurological, renal, and hematology conditions. However, there is still much to learn about the relationship between the eye and other diseases, necessitating ongoing improvements in

AI algorithms and a comprehensive understanding of physiological and pathological mechanisms [149]. The use of AI may improve certain eye care indicators such as increased referral for screening and adherence to referral for DR. However, this has not been conclusively proven yet. The implementation of AI in the prediction and screening of DR presents operational challenges in terms of workflow, technical aspects, ethical considerations, and acceptability among employees, patients, and health economics. The development of AI for DR screening from retinal color images has been accompanied by the rise of deep learning, which has increased the accuracy of diagnosis in both internal and external evaluations across diverse populations [145]. Integrating AI topics, particularly AI generalities and its impact on ophthalmology, into the basic and clinical sciences curriculum of the American Academy of Ophthalmology is an important initiative. Support from reputable organizations for conducting specialized training courses and fellowships on the clinical implementation of AI and decision support systems is also crucial. These efforts represent effective measures for advancing AI in ophthalmology. One notable accomplishment of AI is its highly accurate interpretation of retinal images, surpassing sometimes human experts in the identification of eye and non-eye disorders. This capability holds promise for practical and clinical applications in the future. Currently, the abundance of medical data has laid the groundwork for the integration of AI, deep learning, and instant analysis. This convergence is poised to bring about a profound transformation in healthcare by enhancing accuracy, speed, and workflow efficiency, while simultaneously reducing costs, errors, and ensuring universal access to healthcare [150].

References

1. World Health Organization. Blindness and vision impairment. 2023. Accessed 15 Dec 2023. Available from: <https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment>.
2. Marques AP, Ramke J, Cairns J, Butt T, Zhang JH, Muirhead D, et al. Global economic productivity losses from vision impairment and blindness. *eClinicalMedicine*. 2021;35:100852. <https://doi.org/10.1016/j.eclim.2021.100852>.
3. Chakravarthy SR, Mugambi D, Keshavjee K. AI can improve the economics of blindness prevention in Canada. *Stud Health Technol Inform*. 2024;312:82–6. <https://doi.org/10.3233/shti231317>.
4. Rein DB, Wittenborn JS, Zhang P, Sublett F, Lamuda PA, Lundeen EA, et al. The economic burden of vision loss and blindness in the United States. *Ophthalmology*. 2022;129(4):369–78. <https://doi.org/10.1016/j.ophtha.2021.09.010>.
5. Steinmetz JD, Bourne RRA, Briant PS, Flaxman SR, Taylor HRB, Jonas JB, et al. Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study. *Lancet Glob Health*. 2021;9(2):e144–60. [https://doi.org/10.1016/S2214-109X\(20\)30489-7](https://doi.org/10.1016/S2214-109X(20)30489-7).
6. Honavar SG. Eye of the AI storm: exploring the impact of AI tools in ophthalmology. *Indian J Ophthalmol*. 2023;71(6):2328–40. https://doi.org/10.4103/ijo.ijo_1478_23.
7. Tan TF, Thirunavukarasu AJ, Jin L, Lim J, Poh S, Teo ZL, et al. Artificial intelligence and digital health in global eye health: opportunities and challenges. *Lancet Glob Health*. 2023;11(9):e1432–43. [https://doi.org/10.1016/S2214-109X\(23\)00323-6](https://doi.org/10.1016/S2214-109X(23)00323-6).
8. Hassan B, Raja H, Hassan T, Akram MU, Raja H, Abd-alrazaq AA, et al. A comprehensive review of artificial intelligence models for screening major retinal diseases. *Artif Intell Rev*. 2024;57(5):111. <https://doi.org/10.1007/s10462-024-10736-z>.
9. Jiang X, Xie M, Ma L, Dong L, Li D. International publication trends in the application of artificial intelligence in ophthalmology research: an updated bibliometric analysis. *Ann Transl Med*. 2023;11(5):219. <https://doi.org/10.21037/atm-22-3773>.
10. Charles RC, Justus R, Jennifer RE, Iris G, David M, Matthew JB, et al. Artificial intelligence for diabetic retinopathy in low-income and middle-income countries: a scoping review. *BMJ Open Diabetes Res Care*. 2023;11(4):e003424. <https://doi.org/10.1136/bmjdr-2023-003424>.
11. Bali A, Mansotra V. Analysis of deep learning techniques for prediction of eye diseases: a systematic review. *Arch Comput Methods Eng*. 2024;31(1):487–520. <https://doi.org/10.1007/s11831-023-09989-8>.
12. Ahuja AS, Wagner IV, Dorairaj S, Checo L, Hulzen RT. Artificial intelligence in ophthalmology: a multidisciplinary approach. *Integr Med Res*. 2022;11(4):100888. <https://doi.org/10.1016/j.imr.2022.100888>.
13. Liu C, Jiao D, Liu Z. Artificial intelligence (AI)-aided disease prediction. *Bio Integration*. 2020;1(3):130–6. <https://doi.org/10.15212/bioi-2020-0017>

14. Srivastava O, Tennant M, Grewal P, Rubin U, Seamone M. Artificial intelligence and machine learning in ophthalmology: A review. *Indian J Ophthalmol*. 2023;71(1):11–7. https://doi.org/10.4103/ijo.IJO_1569_22.
15. Anton N, Doroftei B, Curteanu S, Catălin L, Ilie OD, Tărcoveanu F, et al. Comprehensive review on the use of artificial intelligence in ophthalmology and future research directions. *Diagnostics (Basel)*. 2022;13(1):100. <https://doi.org/10.3390/diagnostics13010100>.
16. Wu H, Jin K, Yip CC, Koh V, Ye J. A systematic review of economic evaluation of artificial intelligence-based screening for eye diseases: from possibility to reality. *Surv Ophthalmol*. 2024. <https://doi.org/10.1016/j.survophthal.2024.03.008>.
17. Kubin AM, Huhtinen P, Ohtonen P, Keskitalo A, Wirkkala J, Hautala N. Comparison of 21 artificial intelligence algorithms in automated diabetic retinopathy screening using a handheld fundus camera. *Ann Med*. 2024;56(1):2352018. <https://doi.org/10.1080/07853890.2024.2352018>.
18. Du XL, Li WB, Hu BJ. Application of artificial intelligence in ophthalmology. *Int J Ophthalmol*. 2018;11(9):1555–61. <https://doi.org/10.18240/ijo.2018.09.21>.
19. Grzybowski A, Singhanetr P, Nanegrungsunk O, Ruamviboonsuk P. Artificial intelligence for diabetic retinopathy screening using color retinal photographs: from development to deployment. *Ophthalmol Ther*. 2023;12(3):1419–37. <https://doi.org/10.1007/s40123-023-00691-3>.
20. Das D, Biswas SK, Bandyopadhyay S. A critical review on the diagnosis of diabetic retinopathy using machine learning and deep learning. *Multimedia Tools Appl*. 2022;81(18):25613–55. <https://doi.org/10.1007/s11042-022-12642-4>.
21. Tsao H-Y, Chan P-Y, Su EC-Y. Predicting diabetic retinopathy and identifying interpretable biomedical features using machine learning algorithms. *BMC Bioinform*. 2018;19(9):283. <https://doi.org/10.1186/s12859-018-2277-0>.
22. Al-Sari N, Kutuzova S, Suvitaival T, Henriksen P, Pociot F, Rossing P, et al. Precision diagnostic approach to predict 5-year risk for microvascular complications in type 1 diabetes. *EBioMedicine*. 2022;80:104032. <https://doi.org/10.1016/j.ebiom.2022.104032>.
23. Helen D, Gokila S. EYENET: an eye disease detection system using convolutional neural network. In: 2023 2nd International Conference on Edge Computing and Applications (ICECAA). 2023 Jul 19; IEEE. p. 839–42. <https://doi.org/10.1109/ICECAA58104.2023.10212139>.
24. Wong TY, Sabanayagam C. Strategies to tackle the global burden of diabetic retinopathy: from epidemiology to artificial intelligence. *Ophthalmologica*. 2020;243(1):9–20. <https://doi.org/10.1159/000502387>.
25. Usman TM, Saheed YK, Nsang A, Ajibesin A, Rakshit S. A systematic literature review of machine learning based risk prediction models for diabetic retinopathy progression. *Artif Intell Med*. 2023;143. <https://doi.org/10.1016/j.artmed.2023.102617>.
26. Zhu C-Z, Hu R, Zou B-J, Zhao R-C, Chen C-L, Xiao Y-L. Automatic diabetic retinopathy screening via cascaded framework based on image- and lesion-level features fusion. *J Comput Sci Technol*. 2019;34(6):1307–18. <https://doi.org/10.1007/s11390-019-1977-x>.
27. Vij R, Arora S. A systematic review on deep learning techniques for diabetic retinopathy segmentation and detection using ocular imaging modalities. *Wire Pers Commun*. 2024;134(2):1153–229. <https://doi.org/10.1007/s11277-024-10968-w>.
28. Jabbar A, Liaqat HB, Akram A, Sana MU, Azpíroz ID, Diez IDLT, et al. A lesion-based diabetic retinopathy detection through hybrid deep learning model. *IEEE Access*. 2024;12:40019–36. <https://doi.org/10.1109/ACCESS.2024.3373467>.
29. Omer HK. Diabetic retinopathy detection using Bilayered Neural Network classification model with resubstitution validation. *MethodsX*. 2024;12:102705. <https://doi.org/10.1016/j.mex.2024.102705>.
30. Dai L, Sheng B, Chen T, Wu Q, Liu R, Cai C, et al. A deep learning system for predicting time to progression of diabetic retinopathy. *Nat Med*. 2024;30(2):584–94. <https://doi.org/10.1038/s41591-023-02702-z>.
31. Singh D, Dobhal DC. A deep learning-based transfer learning approach fine-tuned for detecting diabetic retinopathy. *Procedia Comput Sci*. 2024;233:444–53. <https://doi.org/10.1016/j.procs.2024.03.234>.
32. Silva PS, Zhang D, Jacoba CMP, Fickweiler W, Lewis D, Leitmeyer J, et al. Automated machine learning for predicting diabetic retinopathy progression from ultra-widefield retinal images. *JAMA Ophthalmol*. 2024;142(3):171–7. <https://doi.org/10.1001/jamaophthalmol.2023.6318>.
33. Zhou Y, Chia MA, Wagner SK, Ayhan MS, Williamson DJ, Struyven RR, et al. A foundation model for generalizable disease detection from retinal images. *Nature*. 2023;622(7981):156–63. <https://doi.org/10.1038/s41586-023-06555-x>.
34. Bajwa A, Nosheen N, Talpur KI, Akram S. A prospective study on diabetic retinopathy detection based on modify convolutional neural network using fundus images at Sindh Institute of Ophthalmology & Visual Sciences. *Diagnostics (Basel)*. 2023;13(3). <https://doi.org/10.3390/diagnostics13030393>.
35. Sambyal N, Saini P, Syal R, Gupta V. Modified residual networks for severity stage classification of diabetic retinopathy. *Evol Syst*. 2023;14(1):17–35. <https://doi.org/10.1007/s12530-022-09427-3>.
36. Allison K, Patel D, Alabi O. Epidemiology of glaucoma: the past, present, and predictions for the

- future. *Cureus*. 2020;12(11):e11686. <https://doi.org/10.7759/cureus.11686>.
37. Schuman JS, De Los Angeles Ramos Cadena M, McGee R, Al-Aswad LA, Medeiros FA. A case for the use of artificial intelligence in glaucoma assessment. *Ophthalmol Glaucoma*. 2022;5(3):e3–e13. <https://doi.org/10.1016/j.ogla.2021.12.003>.
 38. Huang X, Kong X, Shen Z, Ouyang J, Li Y, Jin K, et al. GRAPE: a multi-modal dataset of longitudinal follow-up visual field and fundus images for glaucoma management. *Sci Data*. 2023;10(1):520. <https://doi.org/10.1038/s41597-023-02424-4>.
 39. Linyu Z, Li T, Min X, Guofan C. The application of artificial intelligence in glaucoma diagnosis and prediction. *Front Cell Dev Biol*. 2023;11:1173094. <https://doi.org/10.3389/fcell.2023.1173094>.
 40. Ting DSW, Cheung CY, Lim G, Tan GSW, Quang ND, Gan A, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA*. 2017;318(22):2211–23. <https://doi.org/10.1001/jama.2017.18152>.
 41. Liu H, Li L, Wormstone IM, Qiao C, Zhang C, Liu P, et al. Development and validation of a deep learning system to detect glaucomatous optic neuropathy using fundus photographs. *JAMA Ophthalmol*. 2019;137(12):1353–60. <https://doi.org/10.1001/jamaophthalmol.2019.3501>.
 42. Medeiros FA, Jammal AA, Thompson AC. From machine to machine: an OCT-trained deep learning algorithm for objective quantification of glaucomatous damage in fundus photographs. *Ophthalmology*. 2019;126(4):513–21. <https://doi.org/10.1016/j.ophtha.2018.12.033>.
 43. Chang J, Lee J, Ha A, Han YS, Bak E, Choi S, et al. Explaining the rationale of deep learning glaucoma decisions with adversarial examples. *Ophthalmology*. 2021;128(1):78–88. <https://doi.org/10.1016/j.ophtha.2020.06.036>.
 44. Xu Y, Hu M, Liu H, Yang H, Wang H, Lu S, et al. A hierarchical deep learning approach with transparency and interpretability based on small samples for glaucoma diagnosis. *NPJ Digit Med*. 2021;4(1):48. <https://doi.org/10.1038/s41746-021-00417-4>.
 45. Fu H, Baskaran M, Xu Y, Lin S, Wong DWK, Liu J, et al. A deep learning system for automated angle-closure detection in anterior segment optical coherence tomography images. *Am J Ophthalmol*. 2019;203:37–45. <https://doi.org/10.1016/j.ajo.2019.02.028>.
 46. Thompson AC, Jammal AA, Berchuck SI, Mariottoni EB, Medeiros FA. Assessment of a segmentation-free deep learning algorithm for diagnosing glaucoma from optical coherence tomography scans. *JAMA Ophthalmol*. 2020;138(4):333–9. <https://doi.org/10.1001/jamaophthalmol.2019.5983>.
 47. Sun S, Ha A, Kim YK, Yoo BW, Kim HC, Park KH. Dual-input convolutional neural network for glaucoma diagnosis using spectral-domain optical coherence tomography. *Br J Ophthalmol*. 2021;105(11):1555–60. <https://doi.org/10.1136/bjophthalmol-2020-316274>.
 48. Huang X, Islam MR, Akter S, Ahmed F, Kazami E, Serhan HA, et al. Artificial intelligence in glaucoma: opportunities, challenges, and future directions. *Biomed Eng Online*. 2023;22(1):126. <https://doi.org/10.1186/s12938-023-01187-8>.
 49. Phene S, Dunn RC, Hammel N, Liu Y, Krause J, Kitade N, et al. Deep learning and glaucoma specialists: the relative importance of optic disc features to predict glaucoma referral in fundus photographs. *Ophthalmology*. 2019;126(12):1627–39. <https://doi.org/10.1016/j.ophtha.2019.07.024>.
 50. Thakur A, Goldbaum M, Yousefi S. Predicting glaucoma before onset using deep learning. *Ophthalmol Glaucoma*. 2020;3(4):262–8. <https://doi.org/10.1016/j.ogla.2020.04.012>.
 51. Yousefi S, Elze T, Pasquale LR, Saedi O, Wang M, Shen LQ, et al. Monitoring glaucomatous functional loss using an artificial intelligence-enabled dashboard. *Ophthalmology*. 2020;127(9):1170–8. <https://doi.org/10.1016/j.ophtha.2020.03.008>.
 52. Thakur S, Dinh LL, Lavanya R, Quek TC, Liu Y, Cheng CY. Use of artificial intelligence in forecasting glaucoma progression. *Taiwan J Ophthalmol*. 2023;13(2):168–83. <https://doi.org/10.4103/tjo.tjo-d-23-00022>.
 53. Suganthe RC, Geetha M, Latha RS, Preethi S, Rishika M, Sandhiya D, editors. Glaucoma prediction using supervised learning. In: 2023 International Conference on Computer Communication and Informatics (ICCCI), vol. 23. IEEE; 2023. p. 1–6. <https://doi.org/10.1109/ICCCI56745.2023.10128405>.
 54. Subbarayudu Y, Reddy GV, Kumar MR, Naik MA, Nagini RV, Singh B. An efficient novel approach for glaucoma classification on retinal fundus images through machine learning paradigm. In: MATEC Web of Conferences, vol. 392. EDP Sciences; 2024. p. 01108. <https://doi.org/10.1051/mateconf/202439201108>.
 55. Qidwai U, Sivapalan T, Ratnarajan G. IMIGS: an innovative AI based prediction system for selecting the best patient-specific glaucoma treatment. *MethodsX*. 2023;10:102209. <https://doi.org/10.1016/j.mex.2023.102209>.
 56. Wang SY, Ravindranath R, Stein JD, Amin S, Edwards PA, Srikumaran D, et al. Prediction models for glaucoma in a multicenter electronic health records consortium: the sight outcomes research collaborative. *Ophthalmol Sci*. 2024;4(3):100445. <https://doi.org/10.1016/j.xops.2023.100445>.
 57. Yoon BW, Lim S-H, Shin JH, Lee J-W, Lee Y, Seo JH. Analysis of oral microbiome in glaucoma patients using machine learning prediction models. *J Oral Microbiol*. 2021;13(1):1962125. <https://doi.org/10.1080/20002297.2021.1962125>.

58. Kamal MS, Dey N, Chowdhury L, Hasan SI, Santosh K. Explainable AI for glaucoma prediction analysis to understand risk factors in treatment planning. *IEEE Trans Instrum Meas*. 2022; 71:1–9. <https://doi.org/10.1109/TIM.2022.3171613>.
59. Landau Prat D, Kapelushnik N, Arazi M, Zloto O, Leshno A, Klang E, et al. Glaucoma prediction models based on ocular and systemic findings. *Ophthalmic Res*. 2024;67(1):29–38. <https://doi.org/10.1159/000535879>.
60. Hussain S, Chua J, Wong D, Lo J, Kadziauskiene A, Asoklis R, et al. Predicting glaucoma progression using deep learning framework guided by generative algorithm. *Sci Rep*. 2023;13(1):19960. <https://doi.org/10.1038/s41598-023-46253-2>.
61. Eswari MS, Balamurali S. An intelligent machine learning support system for glaucoma prediction among diabetic patients. In: 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE). IEEE; 2021. p. 447–49. <https://doi.org/10.1109/icacite51222.2021.9404635>.
62. Wong WL, Su X, Li X, Cheung CMG, Klein R, Cheng C-Y, et al. Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. *Lancet Glob Health*. 2014;2(2):e106–16. [https://doi.org/10.1016/S2214-109X\(13\)70145-1](https://doi.org/10.1016/S2214-109X(13)70145-1).
63. Miranda M, Santos-Oliveira J, Mendonça AM, Sousa V, Melo T, Carneiro Â. Human versus artificial intelligence: validation of a deep learning model for retinal layer and fluid segmentation in optical coherence tomography images from patients with age-related macular degeneration. *Diagnostics*. 2024;14(10):975. <https://doi.org/10.3390/diagnostics14100975>.
64. Haseltine W, Hazel K. Artificial Intelligence to Manage the AMD Burden. *Retinal Physician*. 2024; 21(January/February):E1–E3. Available from: <https://www.retinalphysician.com/issues/2024/januaryfebruary/artificial-intelligence-to-manage-the-amd-burden/>.
65. Giulia C, Srinias S. Artificial intelligence in AMD imaging. *Retina Today*; 2022. Available from: <https://retinatoday.com/articles/2022-may-june/artificial-intelligence-in-amd-imaging>.
66. Yvette W. Artificial intelligence: reshaping AMD management. *miVision education*; 2024. Available from: <https://www.mieducation.com/pages/artificial-intelligence-reshaping-amd-management>.
67. Schmidt-Erfurth U, Bogunovic H, Sadeghipour A, Schlegl T, Langs G, Gerendas BS, et al. Machine learning to analyze the prognostic value of current imaging biomarkers in neovascular age-related macular degeneration. *Ophthalmol Retina*. 2018;2(1):24–30. <https://doi.org/10.1016/j.oret.2017.03.015>.
68. Bhuiyan A, Wong TY, Ting DSW, Govindaiah A, Souied EH, Smith RT. Artificial intelligence to stratify severity of age-related macular degeneration (AMD) and predict risk of progression to late AMD. *Transl Vis Sci Technol*. 2020;9(2):25. <https://doi.org/10.1167/tvst.9.2.25>.
69. Dong L, Yang Q, Zhang RH, Wei WB. Artificial intelligence for the detection of age-related macular degeneration in color fundus photographs: a systematic review and meta-analysis. *eClinicalMedicine*. 2021;35:100875. <https://doi.org/10.1016/j.eclinm.2021.100875>.
70. Crincoli E, Sacconi R, Querques L, Querques G. Artificial intelligence in age-related macular degeneration: state of the art and recent updates. *BMC Ophthalmol*. 2024;24(1):121. <https://doi.org/10.1186/s12886-024-03381-1>.
71. Leng X, Shi R, Wu Y, Zhu S, Cai X, Lu X, et al. Deep learning for detection of age-related macular degeneration: A systematic review and meta-analysis of diagnostic test accuracy studies. *PLoS ONE*. 2023;18(4):e0284060. <https://doi.org/10.1371/journal.pone.0284060>.
72. Abd El-Khalek AA, Balaha HM, Alghamdi NS, Ghazal M, Khalil AT, Abo-Elsoud MEA, et al. A concentrated machine learning-based classification system for age-related macular degeneration (AMD) diagnosis using fundus images. *Sci Rep*. 2024;14(1):2434. <https://doi.org/10.1038/s41598-024-52131-2>.
73. Borkhade G, Akashe S. Machine learning classifier to analyze prognostic value for current imaging biomarker in Age Macular Degeneration. In: 2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), vol. 2. IEEE; 2024. p. 1–6. <https://doi.org/10.1109/IATMSI60426.2024.10503004>.
74. Ali MA, Hossain MS, Hossain MK, Sikder SS, Khushbu SA, Islam M. AMDNet23: Hybrid CNN-LSTM deep learning approach with enhanced pre-processing for age-related macular degeneration (AMD) detection. *Intell Syst Appl*. 2024;21:200334. <https://doi.org/10.1016/j.iswa.2024.200334>.
75. Saha S, Nassisi M, Wang M, Lindenberg S, Kanagasigam Y, Sadda S, et al. Automated detection and classification of early AMD biomarkers using deep learning. *Scientific Reports*. 2019;9(1):10990. <https://doi.org/10.1038/s41598-019-47390-3>.
76. Pesudovs K, Lansingh VC, Kempen JH, Steinmetz JD, Briant PS, Varma R, et al. Cataract-related blindness and vision impairment in 2020 and trends over time in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study. *Invest Ophthalmol Vis Sci*. 2020;62(8):3523.
77. Cicinelli MV, Buchan JC, Nicholson M, Varadaraj V, Khanna RC. Cataracts. *The Lancet*. 2023;401(10374):377–89. [https://doi.org/10.1016/S0140-6736\(22\)01839-6](https://doi.org/10.1016/S0140-6736(22)01839-6).
78. Lindegger DJ, Wawrzynski J, Saleh GM. Evolution and applications of artificial intelligence to cataract

- surgery. *Ophthalmol Sci.* 2022;2(3):100164. <https://doi.org/10.1016/j.xops.2022.100164>.
79. Gutierrez L, Lim JS, Foo LL, Ng WY, Yip M, Lim GYS, et al. Application of artificial intelligence in cataract management: current and future directions. *Eye and Vis.* 2022;9(1):3. <https://doi.org/10.1186/s40662-021-00273-z>.
 80. Tognetto D, Giglio R, Vinciguerra AL, Milan S, Rejdak R, Rejdak M, et al. Artificial intelligence applications and cataract management: a systematic review. *Surv Ophthalmol.* 2022;67(3):817–29. <https://doi.org/10.1016/j.survophthal.2021.09.004>.
 81. Singh G, Guleria K, Sharma S, editors. A pre-trained VGG16 model for cataract eye disease prediction. In: 2023 3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON). IEEE; 2023. p. 1–6. <https://doi.org/10.1109/SMARTGENCON60755.2023.10442647>.
 82. VM MA, Tiwari N, Khobragade V, Pareek M, Panchbhavi A, Ghosh P, et al. Clinical validation of artificial intelligence-based cataract screening solution with smartphone images (Logy AI cataract screening module). *Int J Advances Med.* 2024;11(2):71. <https://doi.org/10.18203/2349-3933.ijam20240007>.
 83. Ueno Y, Oda M, Yamaguchi T, Fukuoka H, Nejima R, Kitaguchi Y, et al. Deep learning model for extensive smartphone-based diagnosis and triage of cataracts and multiple corneal diseases. *Br J Ophthalmol.* 2024;bjo-2023-324488. <https://doi.org/10.1136/bjo-2023-324488>.
 84. Yadav S, Yadav JKPS. Automatic cataract severity detection and grading using deep learning. *J Sensors.* 2023;2023:2973836. <https://doi.org/10.1155/2023/2973836>.
 85. Chakraborty S, Jana S, editors. Early prediction of cataract using convolutional neural network. In: 2023 IEEE Devices for Integrated Circuit (DevIC). IEEE; 2023. p. 446–50. <https://doi.org/10.1109/DevIC57758.2023.10135056>.
 86. Holden BA, Fricke TR, Wilson DA, Jong M, Naidoo KS, Sankaridurg P, et al. Global prevalence of Myopia and high myopia and temporal trends from 2000 through 2050. *Ophthalmology.* 2016;123(5):1036–42. <https://doi.org/10.1016/j.optha.2016.01.006>.
 87. Zhang C, Zhao J, Zhu Z, Li Y, Li K, Wang Y, et al. Applications of artificial intelligence in Myopia: current and future directions. *Fron Med.* 2022;9:840498. <https://doi.org/10.3389/fmed.2022.840498>.
 88. Hashemi H, Khabazkhoob M, Azizi E, Iribarren R, Lanca C, Grzybowski A, et al. Longitudinal changes in crystalline lens thickness and power in children aged 6–12 years old. *Eye.* 2024;38(7):1283–9. <https://doi.org/10.1038/s41433-023-02882-5>.
 89. Lanca C, Emamian MH, Wong YL, Hashemi H, Khabazkhoob M, Grzybowski A, et al. Three-year change in refractive error and its risk factors: results from the Shahroud School Children Eye Cohort Study. *Eye (Lond).* 2023;37(8):1625–32. <https://doi.org/10.1038/s41433-022-02219-8>.
 90. Zhang J, Zou H. Artificial intelligence technology for myopia challenges: a review. *Front Cell Dev Biol.* 2023;11:1124005. <https://doi.org/10.3389/fcell.2023.1124005>.
 91. Obermeyer Z, Emanuel EJ. Predicting the future—big data, machine learning and clinical medicine. *N Engl J Med.* 2016;375(13):1216–9. <https://doi.org/10.1056/NEJMp1606181>.
 92. Yang X, Chen G, Qian Y, Wang Y, Zhai Y, Fan D, et al. Prediction of Myopia in adolescents through machine learning methods. *Int J Environ Res Public Health.* 2020;17(2):463. <https://doi.org/10.3390/ijerph17020463>.
 93. Li S-M, Ren M-Y, Gan J, Zhang S-G, Kang M-T, Li H, et al. Machine learning to determine risk factors for Myopia progression in primary school children: the anyang childhood eye study. *Ophthalmol Ther.* 2022;11(2):573–85. <https://doi.org/10.1007/s40123-021-00450-2>.
 94. Lin H, Long E, Ding X, Diao H, Chen Z, Liu R, et al. Prediction of myopia development among Chinese school-aged children using refraction data from electronic medical records: a retrospective, multicentre machine learning study. *PLoS Med.* 2018;15(11):e1002674. <https://doi.org/10.1371/journal.pmed.1002674>.
 95. Tang T, Yu Z, Xu Q, Peng Z, Fan Y, Wang K, et al. A machine learning-based algorithm used to estimate the physiological elongation of ocular axial length in myopic children. *Eye Vis.* 2020;7(1):50. <https://doi.org/10.1186/s40662-020-00214-2>.
 96. Mrochen M, Zakharov P, Tabakcı BN, Tanrıverdi C, Kılıç A, Flitcroft DI. Visual lifestyle of myopic children assessed with AI-powered wearable monitoring. *Invest Ophthalmol Vis Sci.* 2020;61(7):82.
 97. Artiemjew P, Cybulski R, Emamian MH, Grzybowski A, Jankowski A, Lanca C, et al., editors. Predicting children's Myopia risk: a Monte Carlo approach to compare the performance of machine learning models. In: International Conference on Agents and Artificial Intelligence; 2024: SCITEPRESS Science and Technology Publications. <https://doi.org/10.5220/0012435500003636>.
 98. Zhao J, Yu Y, Li Y, Li F, Zhang Z, Jian W, et al. Development and validation of predictive models for myopia onset and progression using extensive 15-year refractive data in children and adolescents. *J Transl Med.* 2024;22(1):289. <https://doi.org/10.1186/s12967-024-05075-0>.
 99. Li W, Tu Y, Zhou L, Ma R, Li Y, Hu D, et al. Study of myopia progression and risk factors in Hubei

- children aged 7–10 years using machine learning: a longitudinal cohort. *BMC Ophthalmol.* 2024;24(1):93. <https://doi.org/10.1186/s12886-024-03331-x>.
100. Li J, Zeng S, Li Z, Xu J, Sun Z, Zhao J, et al. Accurate prediction of myopic progression and high myopia by machine learning. *Precis Clin Med.* 2024;7(1). <https://doi.org/10.1093/pcmedi/pbae005>.
 101. Foo LL, Lim GYS, Lanca C, Wong CW, Hoang QV, Zhang XJ, et al. Deep learning system to predict the 5-year risk of high myopia using fundus imaging in children. *NPJ Digit Med.* 2023;6(1):10. <https://doi.org/10.1038/s41746-023-00752-8>.
 102. Huang J, Ma W, Li R, Zhao N, Zhou T. Myopia prediction for children and adolescents via time-aware deep learning. *Sci Rep.* 2023;13(1):5430. <https://doi.org/10.1038/s41598-023-32367-0>.
 103. Chen Y, Tan C, Foo LL, He S, Zhang J, Bulloch G, et al. Development and validation of a model to predict who will develop Myopia in the following year as a criterion to define Premyopia. *Asia Pac J Ophthalmol (Phila).* 2023;12(1):38–43. <https://doi.org/10.1097/apo.0000000000000591>.
 104. Wong YL, Yuan Y, Su B, Tufail S, Ding Y, Ye Y, et al. Prediction of myopia onset with refractive error measured using non-cycloplegic subjective refraction: the WEProm Study. *BMJ Open Ophthalmol.* 2021;6(1):e000628. <https://doi.org/10.1136/bmjophth-2020-000628>.
 105. Santodomingo-Rubido J, Carracedo G, Suzuki A, Villa-Collar C, Vincent SJ, Wolffsohn JS. Keratoconus: an updated review. *Cont Lens Anterior Eye.* 2022;45(3). <https://doi.org/10.1016/j.clae.2021.101559>.
 106. Larkin DFP, Chowdhury K, Burr JM, Raynor M, Edwards M, Tuft SJ, et al. Effect of corneal cross-linking versus standard care on keratoconus progression in young patients: the KERALINK randomized controlled trial. *Ophthalmology.* 2021;128(11):1516–26. <https://doi.org/10.1016/j.opht.2021.04.019>.
 107. Smolek MK, Klyce SD. Current keratoconus detection methods compared with a neural network approach. *Invest Ophthalmol Vis Sci.* 1997;38(11):2290–9.
 108. Kuo BI, Chang WY, Liao TS, Liu FY, Liu HY, Chu HS, et al. Keratoconus screening based on deep learning approach of corneal topography. *Transl Vis Sci Technol.* 2020;9(2):53. <https://doi.org/10.1167/tvst.9.2.53>.
 109. Mohammadpour M, Heidari Z, Hashemi H, Yaseri M, Fotouhi A. Comparison of artificial intelligence-based machine learning classifiers for early detection of keratoconus. *Eur J Ophthalmol.* 2022;32(3):1352–60. <https://doi.org/10.1177/112067>.
 110. Maeda N, Klyce SD, Smolek MK. Neural network classification of corneal topography. Preliminary demonstration. *Invest Ophthalmol Vis Sci.* 1995;36(7):1327–35.
 111. Niazi S, Jiménez-García M, Findl O, Gatziofufas Z, Doroodgar F, Shahriari MH, et al. Keratoconus diagnosis: from fundamentals to artificial intelligence: a systematic narrative review. *Diagnostics (Basel).* 2023;13(16). <https://doi.org/10.3390/diagnostics13162715>.
 112. Zhang P, Yang L, Mao Y, Zhang X, Cheng J, Miao Y, et al. CorNet: Autonomous feature learning in raw Corvis ST data for keratoconus diagnosis via residual CNN approach. *Comput Biol Med.* 2024;172: 108286. <https://doi.org/10.1016/j.combiomed.2024.108286>.
 113. Hashemi H, Doroodgar F, Niazi S, Khabazkhoob M, Heidari Z. Comparison of different corneal imaging modalities using artificial intelligence for diagnosis of keratoconus: a systematic review and meta-analysis. *Graefes Arch Clin Exp Ophthalmol.* 2024;262(4):1017–39. <https://doi.org/10.1007/s00417-023-06154-6>.
 114. Ting DSJ, Tan TF, Ting DSW. ChatGPT in ophthalmology: the dawn of a new era? *Eye.* 2024;38(1):4–7. <https://doi.org/10.1038/s41433-023-02619-4>.
 115. Delsoz M, Madadi Y, Raja H, et al. Performance of ChatGPT in diagnosis of corneal eye diseases. *Cornea.* 2024;43(5):664–70. <https://doi.org/10.1097/ico.0000000000003492>.
 116. Al-Timemy AH, Alzubaidi L, Mosa ZM, Abdelmotaal H, Ghaeb NH, Lavric A, et al. A deep feature fusion of improved suspected keratoconus detection with deep learning. *Diagnostics.* 2023;13(10):1689. <https://doi.org/10.3390/diagnostics13101689>.
 117. Smily Yeya Jothi E, Anitha J, Hemanth J. Keratoconus classification using feature selection and machine learning approach. In: Jabbar MA, Ortiz-Rodríguez F, Tiwari S, Siarry P, editors. *Applied machine learning and data analytics. AMLDA 2022. Communications in Computer and Information Science*, vol. 1818. Cham: Springer; 2023. p. 39–51. https://doi.org/10.1007/978-3-031-34222-6_4.
 118. Tan Z, Chen X, Li K, Liu Y, Cao H, Li J, et al. Artificial intelligence-based diagnostic model for detecting keratoconus using videos of corneal force deformation. *Transl Vis Sci Technol.* 2022;11(9):32. <https://doi.org/10.1167/tvst.11.9.32>.
 119. Alyasseri ZAA, Al-Timemy AH, Abasi AK, Lavric A, Mohammed HJ, Takahashi H, et al. A hybrid artificial intelligence model for detecting keratoconus. *Appl Sci.* 2022;12(24):12979.
 120. Herber R, Pillunat LE, Raiskup F. Development of a classification system based on corneal biomechanical properties using artificial intelligence predicting keratoconus severity. *Eye Vis.* 2021;8(1):21. <https://doi.org/10.1186/s40662-021-00244-4>.

121. Castro-Luna G, Jiménez-Rodríguez D, Castaño-Fernández AB, Pérez-Rueda A. Diagnosis of sub-clinical keratoconus based on machine learning techniques. *J Clin Med*. 2021;10(18):4281. <https://doi.org/10.3390/jcm10184281>.
122. Ran AR, Hui HYH, Cheung CY, Wong TY. Chapter 12—Artificial intelligence in ophthalmology III: systemic disease prediction. In: Krittanawong C, editor. *Artificial Intelligence in Clinical Practice*. Cambridge: Academic Press; 2024. p. 119–25. <https://doi.org/10.1016/B978-0-443-15688-5.00041-3>.
123. Patterson EJ, Bounds AD, Wagner SK, Kadri-Langford R, Taylor R, Daly D. Oculomics: a crusade against the four horsemen of chronic disease. *Ophthalmol Ther*. 2024;13:1427–51. <https://doi.org/10.1007/s40123-024-00942-x>.
124. Zhang Y, Li S, Wu W, Zhao Y, Han J, Tong C, et al. Machine-learning-based models to predict cardiovascular risk using oculomics and clinic variables in KHNANES. *BioData Mining*. 2024;17(1):12. <https://doi.org/10.1186/s13040-024-00363-3>.
125. Peng Q, Tseng R, Tham YC, Cheng CY, Rim TH. Detection of systemic diseases from ocular images using artificial intelligence: a systematic review. *Asia Pac J Ophthalmol (Phila)*. 2022;11(2):126–39. <https://doi.org/10.1097/apo.0000000000000515>.
126. Chang J, Ko A, Park SM, Choi S, Kim K, Kim SM, et al. Association of cardiovascular mortality and deep learning-funduscopy atherosclerosis score derived from retinal fundus images. *Am J Ophthalmol*. 2020;217:121–30. <https://doi.org/10.1016/j.ajo.2020.03.027>.
127. Son J, Shin JY, Chun EJ, Jung KH, Park KH, Park SJ. Predicting high coronary artery calcium score from retinal fundus images with deep learning algorithms. *Transl Vis Sci Technol*. 2020;9(6):28. <https://doi.org/10.1167/tvst.9.2.28>.
128. Ng WWY, Liang H, Peng Q, Zhong C, Dong X, Huang Z, et al. An automatic framework for perioperative risks classification from retinal images of complex congenital heart disease patients. *Inter J Mach Learn Cybern*. 2022;13(2):471–83. <https://doi.org/10.1007/s13042-021-01419-0>.
129. Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, et al. Predicting cardiovascular risk factors in retinal fundus photographs using deep learning. *Nat Biomed Eng*. 2018;2(3):158–64. <https://doi.org/10.1038/s41551-018-0195-0>.
130. Rim TH, Lee CJ, Tham Y-C, Cheung N, Yu M, Lee G, et al. Deep-learning-based cardiovascular risk stratification using coronary artery calcium scores predicted from retinal photographs. *Lancet Digital Health*. 2021;3(5):e306–16. [https://doi.org/10.1016/S2589-7500\(21\)00043-1](https://doi.org/10.1016/S2589-7500(21)00043-1).
131. Zhong P, Li Z, Lin Y, Peng Q, Huang M, Jiang L, et al. Retinal microvasculature impairments in patients with coronary artery disease: an optical coherence tomography angiography study. *Acta Ophthalmol*. 100(2):225–233. <https://doi.org/10.1111/aos.14806>.
132. Lim G, Lim ZW, Xu D, Ting DSW, Wong TY, Lee ML, et al. Feature isolation for hypothesis testing in retinal imaging: an ischemic stroke prediction case study. *Proc AAAI Conf Artif Intel*. 2019;33(01):9510–5. <https://doi.org/10.1609/aaai.v33i01.33019510>.
133. Dai G, He W, Xu L, Pazo EE, Lin T, Liu S, et al. Exploring the effect of hypertension on retinal microvasculature using deep learning on the East Asian population. *PLoS ONE*. 2020;15(3):e0230111. <https://doi.org/10.1371/journal.pone.0230111>.
134. Zhang L, Yuan M, An Z, Zhao X, Wu H, Li H, et al. Prediction of hypertension, hyperglycemia and dyslipidemia from retinal fundus photographs via deep learning: A cross-sectional study of chronic diseases in central China. *PLoS ONE*. 2020;15(5):e0233166. <https://doi.org/10.1371/journal.pone.0233166>.
135. Benson J, Estrada T, Burge M, Soliz P. Diabetic peripheral neuropathy risk assessment using digital fundus photographs and machine learning. *Annu Int Conf IEEE Eng Med Biol Soc*. 2020;2020:1988–91. <https://doi.org/10.1109/EMBC44109.2020.9175982>.
136. Aslam TM, Hoyle DC, Puri V, Bento G. Differentiation of diabetic status using statistical and machine learning techniques on optical coherence tomography angiography images. *Transl Vis Sci Technol*. 2020;9(4):2. <https://doi.org/10.1167/tvst.9.4.2>.
137. Govindaswamy N, Ratra D, Dalan D, Doralli S, Tirumalai AA, Nagarajan R, et al. Vascular changes precede tomographic changes in diabetic eyes without retinopathy and improve artificial intelligence diagnostics. *J Biophotonics*. 2020;13(9):e202000107. <https://doi.org/10.1002/jbio.202000107>.
138. Islam T, Imran A, Arefeen A, Hasan M, Shahnaz C. Source and camera independent ophthalmic disease recognition from fundus image using neural network. In: 2019 IEEE international conference on signal processing, information, communication & systems (SPICSCON). 2019. p. 59–63. <https://doi.org/10.1109/spicscon48833.2019.9065162>.
139. Zhang K, Liu X, Xu J, Yuan J, Cai W, Chen T, et al. Deep-learning models for the detection and incidence prediction of chronic kidney disease and type 2 diabetes from retinal fundus images. *Nat Biomed Eng*. 2021;5(6):533–45. <https://doi.org/10.1038/s41551-021-00745-6>.
140. Prakash YP, Siri B, Jabez J. Identifying the abnormalities in retinal images towards the prediction of cardiovascular disease using deep learning. In: 2024 international conference on wireless communications signal processing and networking (WiSPNET). IEEE; 2024. p. 1–6. <https://doi.org/10.1109/WiSPNET61464.2024.10532859>.

141. Al-Absi HRH, Pai A, Naeem U, Mohamed FK, Arya S, Sbeit RA, et al. DiaNet v2 deep learning based method for diabetes diagnosis using retinal images. *Sci Rep.* 2024;14(1):1595. <https://doi.org/10.1038/s41598-023-49677-y>.
142. Kim BR, Yoo TK, Kim HK, Ryu IH, Kim JK, Lee IS, et al. Oculomics for sarcopenia prediction: a machine learning approach toward predictive, preventive, and personalized medicine. *Epma J.* 2022;13(3):367–82. <https://doi.org/10.1007/s13167-022-00292-3>.
143. Alhazimi A, Almarek F. Impact of artificial intelligence on diagnosing eye diseases—a meta-analysis. *Adv Hum Biol.* 2024;14(1):42–54. https://doi.org/10.4103/aihb.aihb_124_23.
144. Veritti D, Rubinato L, Sarao V, De Nardin A, Foresti GL, Lanzetta P. Behind the mask: a critical perspective on the ethical, moral, and legal implications of AI in ophthalmology. *Graefes Arch Clin Exp Ophthalmol.* 2024;262(3):975–82. <https://doi.org/10.1007/s00417-023-06245-4>.
145. Grzybowski A, Brona P. Approval and certification of ophthalmic AI devices in the European Union. *Ophthalmol Ther.* 2023;12(2):633–8. <https://doi.org/10.1007/s40123-023-00652-w>.
146. Shao Y, Jie Y, Liu ZG. Guidelines for the application of artificial intelligence in the diagnosis of anterior segment diseases. *Int J Ophthalmol.* 2023;16(9):1373–85. <https://doi.org/10.18240/ijo.2023.09.03>.
147. Yang J, Wu S, Dai R, Yu W, Chen Y. Publication trends of artificial intelligence in retina in 10 years: where do we stand? *Front Med.* 2022;9. <https://doi.org/10.3389/fmed.2022.1001673>
148. Zhao J, Lu Y, Zhu S, Li K, Jiang Q, Yang W. Systematic bibliometric and visualized analysis of research hotspots and trends on the application of artificial intelligence in ophthalmic disease diagnosis. *Front Pharmacol.* 2022;13. <https://doi.org/10.3389/fphar.2022.930520>.
149. Omar MO, Ali MJA, Qabillie SE, Haji AI, Takriti MB, Atif AH, et al. Beyond vision: potential role of ai-enabled ocular scans in the prediction of aging and systemic disorders. *Siriraj Med J.* 2024;76(2):106–15. <https://doi.org/10.33192/smj.v76i2.266303>.
150. Grzybowski A. Artificial intelligence in ophthalmology 2023 Post congress report. In: International online conference in ophthalmology 2023, 23 June 2023. Held by Ministry of Education and Science Republic of Poland. Accessed 1 May 2021. Available from: https://aiinophthalmology.com/wp-content/uploads/2023/09/Report-AI-2023_ENG.pdf.

Leveraging Artificial Intelligence (AI) to Enhance Nursing Care for Patients Undergoing Eye Surgery

27

Morteza Shamsizadeh

Abstract

In today's healthcare landscape, the fusion of technology and patient care is reshaping our understanding of medical possibilities. Eye surgeries known for their delicate procedures and precision, highlight this evolution. This chapter delves into the pivotal role of Artificial Intelligence (AI) in bolstering nursing care for eye surgeries. AI's tools offer a pathway to improved patient outcomes and an updated perspective on improving nursing care. This examination emphasizes the collaborative potential of technology and hands-on care, charting a promising future for those dedicated to advancing healthcare. Nurses and AI can work together to optimize patient outcomes in eye surgeries by complementing each other's strengths and addressing each other's limitations. Nurses have the human touch, clinical judgment, and ethical values that are essential for providing holistic and compassionate care to patients. AI has the computational power, data-driven insights, and automation capabilities that can enhance the accuracy, efficiency, and consistency of care delivery.

Keywords

Artificial intelligence · Machine learning · Nursing care · Delirium · Nursing diagnosis · Eye surgeries

Background

AI's Role in Nursing Care of Patients Undergoing Eye Surgery: A Brief Overview of Healthcare Technology Evolution

The evolving role of technology in healthcare has ushered in transformative changes, revolutionizing patient care and clinical practices. With a specific focus on artificial intelligence (AI), the landscape of nursing care for eye surgeries has seen significant advancements. The integration of AI technologies has augmented the capabilities of healthcare professionals, providing tailored solutions to enhance the quality and precision of care in the intricate field of ophthalmic surgery [1–4].

Traditionally, healthcare has been characterized by manual processes, but the advent of technology, particularly AI, has paved the way for more efficient, accurate, and personalized care. AI systems, driven by machine learning algorithms, can analyze vast amounts of data,

M. Shamsizadeh (✉)
School of Nursing, University of Pittsburgh,
Pittsburgh, PA, USA
e-mail: Mos179@pitt.edu

offering insights and predictions that aid healthcare providers in decision-making. This capability is particularly beneficial in the intricate and sensitive realm of eye surgeries where precision and individualized care are paramount [5, 6].

In the context of nursing care for eye surgeries, AI contributes to both preoperative and postoperative phases. It plays a pivotal role in patient risk assessment, helping nurses identify potential complications and tailor care plans to individual patient needs [7–9]. AI-driven platforms facilitate personalized patient education, delivering information tailored to specific eye surgeries and ensuring patients are well-informed and prepared. Communication enhancement is another crucial aspect where AI technologies shine. They enable seamless coordination among healthcare providers involved in preoperative care, ensuring that the entire team is well-informed and aligned in their approach. This collaborative environment enhances the overall efficiency and effectiveness of nursing care [7, 10, 11].

As technology continues to evolve, the integration of AI in nursing care promises ongoing improvements. The precision, data-driven insights, and enhanced communication facilitated by AI empower nurses to deliver a higher standard of care. This synergy between human expertise and technological innovation underscores the potential for continued advancements in patient outcomes and the overall quality of healthcare in the field of ophthalmic surgery [12, 13].

AI is not meant to replace nurses, but rather to support and enhance their expertise and skills [14]. By using AI, nurses can deliver better outcomes and experiences for patients undergoing eye surgeries. Some of the ways that AI can help nurses are:

- AI can analyze large amounts of patient data and identify patterns, risks, and outcomes that humans might miss [5],
- AI can optimize medication dosages, prevent adverse events, and reduce hospital readmissions by using predictive analytics and machine learning algorithms [15],

- AI can improve patient education, satisfaction, and adherence by using chatbots (ChatGPT, Claude, Bing, Bard, etc.) virtual assistants, and wearable devices [16, 17],
- AI can augment the human experience and trust between nurses and patients by providing empathetic, timely, and accurate communication [18].

The Role of Nursing in Eye Surgery Care

Delving into the intricate domain of eye surgeries, nurses assume pivotal roles in the orchestration of preoperative and postoperative care, navigating a landscape characterized by precision and sensitivity. Their responsibilities, ranging from meticulous assessments to vigilant monitoring, form the backbone of a patient-centric care continuum, ensuring both the safety and well-being of individuals undergoing these transformative procedures [19, 20].

Nurses take on the crucial responsibility of conducting thorough assessments, meticulously gathering patient histories, and identifying potential risk factors. They collaborate with the healthcare team to develop tailored care plans, ensuring that each intervention aligns with the individual needs of the patient. Education becomes a focal point as nurses adeptly communicate preoperative instructions, easing anxieties and fostering a sense of preparedness among patients scheduled for eye surgeries [21].

In the perioperative period, nurses facilitate a seamless transition from preoperative preparations to the operating room. Their responsibilities extend to maintaining open lines of communication with the surgical team, orchestrating the logistics, and ensuring that the patient is mentally and physically ready for the impending procedure. This intermediary role positions nurses as integral members of the surgical care team, contributing to the overall success of the eye surgery [22].

The postoperative phase marks a critical juncture where nurses vigilantly monitor

patients for immediate complications. Their responsibilities encompass the management of pain, administration of prescribed medications, and continuous assessment of vital signs. In this intermediate phase, nurses act as liaisons between the patient and the broader health-care team, providing real-time updates on the patient's condition and promptly addressing any emerging concerns [21, 23].

Throughout both preoperative and postoperative phases, nurses embody a commitment to patient-centered care. Their intermediate role involves not only executing clinical interventions but also offering empathetic support, addressing patient concerns, and fostering a healing environment. By recognizing the interplay between the physical and emotional aspects of care, nurses contribute to a holistic patient experience, elevating the standard of nursing care in the context of eye surgeries [11, 24].

Overall, the nursing role in vision improvement is the provision of comprehensive and holistic care for patients with eye conditions or impairments, as well as the promotion of eye health and prevention of vision loss in the general population. Nurses can contribute to vision improvement by:

- Assessing, diagnosing, and treating eye problems, such as infections, injuries, allergies, and dryness, using appropriate medications, devices, and procedures [25].
- Educating and counseling patients and their families about eye diseases, such as cataracts, glaucoma, diabetic retinopathy, and macular degeneration, and their management, such as surgery, medication, and lifestyle changes [21].
- Collaborating and communicating with other health professionals, such as ophthalmologists, optometrists, and pharmacists, to provide coordinated and integrated care for patients with complex eye conditions [26].
- Advocating and participating in eye health promotion and vision loss prevention activities, such as screening, immunization, smoking cessation, and environmental protection, in various settings, such as schools, workplaces, and communities [27].

- Engaging and empowering patients and their families to participate in their own eye care and vision improvement, by providing information, support, and resources [28].

Nurses play a vital role in eye surgery care, as they improve the quality of life and well-being of patients with eye problems, and reduce the impact of vision loss on individuals, families, and society. Their complex responsibilities are essential for the smooth delivery of preoperative and postoperative care, and their skill in balancing precision and compassion makes them key contributors to the best patient outcomes and quality of care.

Exploration of How AI Technologies Enhance Nursing Care

The integration of Artificial Intelligence (AI) technologies, notably machine learning, marks a paradigm shift in the landscape of nursing care. In the specialized field of eye surgeries, where precision and personalized interventions are paramount, the exploration of AI's potential becomes particularly significant.

Machine Learning Advancements

Machine learning, a subset of AI, empowers nursing care with the ability to analyze vast datasets and discern intricate patterns. In the preoperative phase, machine learning algorithms contribute to patient risk assessment by identifying potential complications based on individual health histories. This data-driven approach enables nurses to tailor preoperative care plans, optimizing interventions for enhanced patient safety and personalized attention. Machine learning is a type of artificial intelligence that can help nurses pre- and post-surgery by analyzing data and providing insights, predictions, and recommendations [29–31]. Here are some examples of how machine learning can help nurse's pre and post surgeries:

- Machine learning can help nurses screen and diagnose eye conditions, select the best surgical approach, and monitor postoperative recovery by analyzing data from patient records, images, and sensors [32, 33],
- Machine learning can help nurses optimize medication dosages, prevent adverse events, and reduce hospital readmissions by using predictive analytics and machine learning algorithms [10, 34],
- Machine learning can help nurses improve patient education, satisfaction, and adherence by using chatbots, virtual assistants, and wearable devices [33],
- Machine learning can help nurses evaluate their surgical skills and performance by using video analysis, motion tracking, and feedback systems [33],

Machine learning is not meant to replace nurses, but rather to support and enhance their expertise and skills. By using machine learning, nurses can deliver better outcomes and experiences for patients undergoing eye surgeries.

Importance of Optimizing Preoperative and Postoperative Care for Patients Undergoing Eye Surgeries

Optimizing preoperative and postoperative care for patients undergoing eye surgeries is of paramount importance in ensuring successful outcomes and preserving visual health. The delicate nature of the eye, coupled with the intricacies of surgical procedures, necessitates a meticulous approach to patient care. From the initial assessment to the postoperative recovery phase, attentive and personalized nursing care is essential to mitigate risks, enhance patient comfort, and maximize the chances of a seamless recuperation [35, 36].

Preoperative Care Significance

Preoperative care is the preparation and management of a patient before surgery. It aims to ensure that the patient is in the best possible

condition for the procedure, and to reduce the risk of complications and adverse outcomes. AI can play a significant role in enhancing preoperative care by supporting nurses in various aspects, such as:

- **Diagnosis:** AI can assist nurses as part of an eye care team in screening and diagnosing eye conditions that may necessitate surgery, including cataracts, glaucoma, diabetic retinopathy, and more. AI can analyze images and data from various sources, such as fundus cameras, optical coherence tomography, and electronic health records, and provide accurate and timely diagnosis [37].
- **Treatment:** AI can assist nurses as part of an eye care team in selecting and optimizing the best treatment options for eye surgery patients, including intraocular lens (IOL) power calculation, surgical procedure planning, and postoperative care. AI can use data from previous cases, patient preferences, and clinical guidelines to recommend the most suitable IOL type, power, and position, as well as the optimal surgical technique, anesthesia, and medication [38].
- **Education and engagement:** AI can help nurses to enhance their knowledge and skills regarding eye surgery, as well as to improve patient education and engagement. AI can provide personalized and adaptive learning platforms, virtual health assistants, and gamified interventions to deliver relevant and engaging information and feedback to nurses and patients [39].

In the realm of eye surgeries, preoperative care serves as the foundation for a patient's journey toward improved vision. Thorough preoperative assessments, including risk evaluations and comprehensive health histories, enable healthcare providers to tailor interventions and anticipate potential challenges. This proactive approach not only enhances the safety of the surgical process but also ensures that patients are well-prepared and informed, contributing to a smoother transition into the operating room.

Postoperative Care

Equally important is the focus on post-surgery care as the recovery process following eye surgeries requires thorough care and continuous monitoring. The immediate postoperative period is characterized by the vulnerability of the eye, and nursing interventions during this phase significantly influence the overall success of the surgery. Attending to patients' comfort, managing pain effectively, and monitoring for potential complications are essential elements of postoperative care that directly impact the recovery trajectory. By leveraging AI, nurses can improve their postoperative care performance and productivity, and ultimately improve patient safety and satisfaction [40–42]. AI can play a crucial role in enhancing postoperative care by supporting nurses in various aspects, such as:

- **Monitoring:** AI can help nurses to monitor the vital signs, wound healing, pain level, and recovery progress of eye surgery patients, using sensors, wearables, and smart devices. AI can alert nurses to any abnormal or critical changes, and provide timely and appropriate interventions [10, 43].
- **Evaluation:** AI can help nurses to evaluate the outcomes and satisfaction of eye surgery patients, using data analysis, natural language processing, and machine learning. AI can measure the effectiveness and quality of the surgical procedure, the postoperative care, and the patient education and engagement [10, 31].
- **Innovation:** AI can help nurses to innovate and improve their postoperative care practices, using evidence-based research, best practices, and feedback. AI can provide nurses with continuous learning opportunities, clinical decision support, and personalized recommendations [44].
- **Analyze the patient's data,** such as age, medical history, surgical procedure, and vital signs, to generate a list of possible nursing diagnoses, such as risk of infection, impaired vision, and anxiety [25].
- **Rank the nursing diagnoses** by their priority and severity, and suggest the most appropriate ones for the patient, such as impaired vision and anxiety [10, 45].
- **Provide the nurse with a care plan template,** which includes the nursing diagnoses, the expected outcomes, the interventions, and the evaluation criteria [10].
- **Help the nurse to customize the care plan** according to the patient's specific needs and preferences, such as providing eye drops, glasses, and education materials, as well as offering emotional support and reassurance [46].
- **Monitor the patient's progress and response** to the care plan, and alert the nurse to any changes or complications, such as signs of infection, bleeding, or increased pain [47].
- **Update the care plan accordingly,** and provide the nurse with feedback and suggestions for improvement [48].

In essence, the importance of optimizing pre-operative and postoperative care in the context of eye surgeries extends beyond the technical aspects of the procedures. It embodies a commitment to comprehensive, patient-centric care that considers both the delicate nature of the eye and the unique needs of individuals undergoing these transformative surgeries.

Transforming Patients Education: The Power of Immersive Virtual Reality (VR) Simulations

Immersive Virtual Reality (VR) technology has ushered in a transformative era in healthcare education, offering patients an unprecedented level of engagement and understanding through experiential simulations [49]. One of the most compelling applications of VR in healthcare is the creation of tailored educational experiences that allow patients to visually immerse themselves in virtual environments replicating specific surgical procedures. Unlike traditional educational materials, such as pamphlets or

videos, VR simulations provide a dynamic and interactive platform that can be customized to align with the individual patient's case, enhancing relevance and engagement [50, 51].

These VR educational experiences are meticulously designed to simulate the surgical journey, from pre-operative preparations to the intricacies of the surgical procedure itself. Patients can navigate through a virtual representation of the operating room, interact with virtual instruments, and observe the step-by-step progression of the surgical process. This immersive exposure empowers patients with a comprehensive understanding of what to expect during the actual surgery, demystifying the procedure and alleviating anxiety or apprehensions they may have [46, 52, 53].

Furthermore, the tailored nature of these VR simulations enables healthcare providers to address specific concerns or questions unique to each patient's case. By experiencing a realistic representation of the surgical procedure in a virtual environment, patients can engage more actively in discussions with their healthcare team, ask informed questions, and make well-informed decisions about their care [54]. Additionally, healthcare providers can utilize VR simulations as a valuable educational tool to enhance communication, foster collaboration, and ensure that patients and their families are well-prepared and confident as they approach the actual surgical experience [55, 56].

In conclusion, immersive VR educational experiences represent a groundbreaking advancement in patient education and engagement within the healthcare landscape. By providing patients with tailored, experiential simulations of surgical procedures, VR technology enhances transparency, understanding, and empowerment, ultimately contributing to improved patient outcomes and satisfaction. As the integration of VR technology continues to evolve, its potential to revolutionize patient education and transform the healthcare experience remains profoundly promising.

Personalized Patient Education: Enhancing Healthcare with AI-Driven Platforms

Personalized patient education is the process of providing tailored and relevant information and guidance to patients according to their individual characteristics, needs, and preferences. Personalized patient education can help patients understand their health conditions, treatment options, and self-care practices, as well as improve their health literacy, adherence, and outcomes [57]. However, traditional methods of patient education, such as printed materials, verbal instructions, or generic online resources, may have limitations in effectiveness, accessibility, and scalability, especially in the context of increasing patient diversity and complexity [58, 59].

Artificial intelligence (AI)-driven platforms are emerging technologies that can enhance personalized patient education by leveraging data analytics, natural language processing, and machine learning. AI-driven platforms can collect and analyze various data sources, such as electronic health records, medical images, laboratory tests, genomics, and wearable devices, to generate personalized and evidence-based content for patients. AI-driven platforms can also interact with patients through natural language, such as voice or text, and provide feedback, reminders, and reinforcement. AI-driven platforms can also adapt to patients' learning styles, progress, and feedback, and provide continuous and timely updates [5, 60, 61].

Several studies have demonstrated the potential and feasibility of using AI-driven platforms for personalized patient education in various clinical scenarios, such as chronic diseases, surgery, and mental health. For example [5, 62]. AI-driven platforms for personalized patient education have several advantages over conventional methods, such as greater effectiveness, accessibility, and scalability. However, there are also some challenges and limitations that

need to be addressed, such as data quality and privacy, algorithm interpretability and transparency, ethical and legal implications, and user acceptance and trust [63]. Therefore, AI-driven platforms for personalized patient education require rigorous validation, evaluation, and integration with clinical practice to ensure their safety, effectiveness, and value [5, 64].

Communication Enhancement with AI

Communication is a vital aspect of healthcare delivery, as it enables the exchange of information, coordination of care, and collaboration among healthcare providers. However, traditional methods of communication, such as phone calls, emails, or paper-based records, may have limitations in timeliness, accuracy, and accessibility, especially in complex and dynamic situations. AI-powered communication tools are emerging technologies that can enhance communication and data sharing among healthcare providers by leveraging natural language processing, speech recognition, and machine learning. One of the examples of successful communication strategies leveraging AI for improved preoperative care is the use of AI-powered tools to capture and document patient-provider conversations during clinical encounters. This can help nurses and surgeons to review and share the relevant information and instructions before surgery, as well as to reduce the documentation burden and errors. For instance, Nuance Dragon Ambient eXperience is an AI-powered tool that automatically generates structured and accurate clinical notes from the patient-provider dialogue [65]. These AI-powered communication tools can improve the quality, efficiency, and safety of preoperative care, as well as the satisfaction and engagement of healthcare providers and patients.

AI holds significant promise for enhancing communication between patients and healthcare providers [5]. Hospitals, for example, are deploying chatbots to oversee patient health remotely and respond to their inquiries. This innovation can potentially enhance

patient health by helping them navigate intricate medication regimens, ensuring that clinicians receive regular updates on patients' conditions, and allowing for early intervention at the first signs of trouble [66, 67]. Additionally, AI can facilitate more effective conversations between doctors and patients by offering insights into patients' physical and mental health, identifying high-risk scenarios when providers are experiencing stress, and reducing unnecessary tests [68, 69]. During medical/nursing visits, AI enhances patient-provider interactions by using ambient listening tools to transcribe clinical encounters, translating these discussions into structured notes for electronic health records. This process also offers opportunities to automate follow-ups and manage next steps, such as making referrals and scheduling future appointments [70].

AI-powered communication tools have the potential to improve the quality, efficiency, and safety of healthcare delivery, as well as the satisfaction and engagement of healthcare providers. However, there are also some challenges and limitations that need to be addressed, such as data quality and privacy, algorithm validity and reliability, user acceptance and trust, and ethical and legal implications. Therefore, AI-powered communication tools require rigorous evaluation, validation, and integration with clinical practice to ensure their safety, effectiveness, and value. In addition to the complexities surrounding the implementation of AI tools in healthcare, there is the critical concern of interoperability. Healthcare organizations venturing into the adoption of AI technologies must navigate the challenge of creating seamlessly integrated systems. These systems are essential for facilitating effective communication not only between departments within the organization but also across various patient care units. Achieving interoperability ensures that vital patient information flows efficiently and securely between different components of the healthcare ecosystem, ultimately enhancing the quality of care delivered. One framework commonly utilized to gauge the level of electronic medical record adoption and, by extension, interoperability

within healthcare organizations is the HIMSS (Healthcare Information and Management Systems Society) Electronic Medical Record Adoption Model (EMRAM). HIMSS EMRAM provides a structured approach for assessing the maturity and functionality of electronic medical record systems, serving as a benchmark for organizations striving to optimize their technological infrastructure for the delivery of healthcare services.

Integrating the Nursing Process with AI

The nursing process is a systematic, patient-centered approach to delivering individualized care based on the unique needs and responses of the patient. It is a framework that guides nurses in providing holistic, evidence-based care across various healthcare settings [71, 72]. The nursing process consists of five interconnected stages:

1. **Assessment:** This initial phase involves collecting comprehensive data about the patient's health status, including physical, psychological, social, cultural, and environmental factors. Through observation, interviews, and examination, nurses gather information that will inform subsequent stages of the process.
2. **Diagnosis:** Based on the assessment data, nurses identify actual or potential health problems and formulate nursing diagnoses. These diagnoses are distinct from medical diagnoses and focus on the patient's responses to health conditions or life processes.
3. **Planning:** In this stage, nurses collaborate with patients, families, and other healthcare team members to develop individualized care plans. Goals and outcomes are established, interventions are selected, and priorities are set based on the identified nursing diagnoses.
4. **Implementation:** This stage involves executing the planned interventions outlined in the care plan. Nurses provide direct care, coordinate with other healthcare professionals, and

facilitate patient education and empowerment. Implementation encompasses a range of activities aimed at promoting, maintaining, or restoring health and well-being.

5. **Evaluation:** The final stage of the nursing process involves assessing the effectiveness of the care provided and evaluating the achievement of goals and outcomes. Nurses collect data, compare actual outcomes with expected outcomes, identify any variances, and modify the care plan as needed to optimize patient outcomes.

The nursing process is dynamic and cyclical, allowing for ongoing assessment, planning, intervention, and evaluation as patients' needs and responses change over time. It emphasizes a proactive, patient-centered approach to care that prioritizes individualized interventions, evidence-based practice, and continuous quality improvement.

Nursing Diagnosis

The nursing diagnosis is a critical component of the nursing process. A nursing diagnosis, as defined by the North American Nursing Diagnosis Association (NANDA) International, is a clinical judgment regarding an individual, family, or community's response to actual or potential health issues or life processes [73]. This diagnosis is formulated based on data gathered during the nursing assessment and serves as the basis for developing the care plan. It directs the selection of nursing interventions aimed at achieving outcomes for which the nurse is responsible. Essentially, the nursing diagnosis underpins the nursing process, and the nursing process provides the framework for the nursing diagnosis [74, 75]. The nursing diagnosis is crucial in guiding the choice of nursing interventions to attain the desired patient outcomes. Unlike medical diagnoses, which identify diseases or conditions, nursing diagnoses focus on the patient's response to these conditions, providing a framework for holistic and individualized care. Through a systematic assessment,

nurses identify specific patient problems or risks, enabling them to formulate appropriate interventions that address not only the physical but also the emotional, psychological, and social aspects of care [71, 76]. We requested that ChatGPT provide us with nursing diagnoses for patients undergoing eye surgeries [77]. By doing so, nurses can utilize AI platforms to guide their practice and ultimately seek consultations with experienced professionals. Here's the list of nursing diagnoses:

Pre-operative Phase:

- Risk for injury related to impaired vision.
- Anxiety related to uncertainty about surgery.
- Risk for infection related to pre-operative preparations.
- Altered comfort related to anticipation of surgery.
- Risk for dehydration related to pre-operative fasting.
- Risk for impaired skin integrity related to pre-operative preparations.
- Disturbed sleep pattern related to pre-operative anxiety.
- Risk for aspiration related to pre-operative medications.

Intra-operative Phase:

- Risk for injury related to positioning on the operating table.
- Risk for infection related to surgical site exposure.
- Acute pain related to surgical intervention.
- Risk for aspiration related to anesthesia drugs administration.
- Altered sensory perception related to anesthesia effects.
- Risk for bleeding related to surgical procedure.
- Risk for hypothermia related to operating room environment.
- Risk for impaired skin integrity related to surgical draping.

- Risk for impaired tissue integrity related to surgical manipulation.

Post-operative Phase:

- Risk for infection related to surgical wound.
- Acute pain related to surgical site.
- Risk for impaired vision related to surgical outcomes.
- Anxiety related to post-operative recovery.
- Risk for falls related to post-operative dizziness.
- Altered comfort related to post-operative discomfort.
- Risk for dehydration related to post-operative nausea.
- Risk for constipation related to post-operative medications.
- Impaired mobility related to post-operative restrictions.
- Disturbed sleep pattern related to post-operative pain.
- Risk for aspiration related to post-operative nausea and vomiting.
- Risk for impaired gas exchange related to post-operative respiratory complications.
- Risk for impaired renal function related to post-operative medications.
- Altered nutrition: less than body requirements related to post-operative dietary restrictions.
- Risk for impaired social interaction related to hospitalization.
- Risk for impaired cognition related to post-operative medications.
- Risk for impaired physical activity related to recovery.
- Risk for impaired mobility related to visual impairment.
- Altered body image related to surgical changes.
- Risk for impaired self-care related to recovery.
- Risk for impaired home maintenance related to recovery.
- Risk for impaired sleep patterns related to post-operative discomfort.

- Risk for impaired swallowing related to postoperative pain, nausea and vomiting.
- Risk for impaired communication related to vision changes.
- Risk for impaired cardiovascular function related to anesthesia effects.
- Risk for impaired liver function related to anesthesia drugs metabolism.
- Risk for impaired social interaction related to recovery restrictions.
- Risk for impaired environmental interpretation syndrome related to vision changes.
- Risk for impaired physical activity related to recovery restrictions.
- Risk for impaired verbal communication related to vision changes.
- Risk for impaired renal function related to post-operative medications.
- Risk for impaired comfort related to eye bandaging.
- Risk for impaired self-care related to recovery.

These nursing diagnoses encompass major concerns for patients undergoing eye surgeries, but they should be tailored to individual patient assessments and clinical contexts.

Delirium and Eye Surgeries

Delirium is a common and serious postoperative complication of eye surgeries, especially in older patients and those with pre-existing cognitive impairment [78–80]. Delirium is characterized by acute and fluctuating disturbances in attention, awareness, and cognition, which can impair patient recovery and increase the risk of adverse outcomes. The exact causes of delirium are not fully understood, but some possible factors include surgery-related stress, anesthesia, inflammation, infection, pain, medication, dehydration, and sleep deprivation [81–83]. The European Society of Anesthesiologists has delivered dedicated guidelines that reported the need for routine monitoring using validated scales [84]. Therefore, it is important to understand the risk factors and preventive measures for

postoperative delirium in order to enhance nursing care for eye surgeries. AI-based platforms can help in identifying patients who are at high risk of developing postoperative delirium and in providing personalized care to prevent its occurrence; for example, machine learning-based prediction models showed excellent performance in predicting delirium [85]. Moreover, an AI-based approach provides a hopeful method for early detection and intervention, which has the potential to transform care strategies for this susceptible group with delirium [86]. To prevent and manage delirium, nurses should assess patients' baseline cognitive function and delirium risk before surgery, and implement perioperative interventions such as optimizing hydration, nutrition, oxygenation, analgesia, and sedation, as well as providing orientation, stimulation, and mobilization [87–90]. Nurses should also monitor patients for signs and symptoms of delirium, and promptly report and treat any episodes using non-pharmacological and pharmacological approaches [91, 92]. According to a study, nurses are more likely to use AI to prevent and manage delirium in hospitals. The study provides insights into the user acceptance of a machine learning-based application supporting delirium management in hospitals [93]. By recognizing and addressing delirium, nurses can improve the quality and safety of care for patients undergoing eye surgeries. AI can contribute to the prevention and management of delirium. Here are some possible examples:

- AI can help detect early signs of delirium and predict which patients are at high risk of developing delirium during their intensive care unit (ICU) stay. For instance, AI algorithms can analyze electronic health record data and identify patients who are likely to develop delirium based on their age, medical history, vital signs, and medications [94].
- AI can help monitor the vital signs, neurological status, and cognitive function of patients who are at risk of or have delirium, using sensors, wearables, and smart devices. AI can alert nurses and doctors to any abnormal or critical changes, and provide timely

and appropriate interventions. For example, researchers have used AI and rapid-response EEGs to detect delirium in critically ill older adults, and provide feedback and suggestions for improvement [95].

- AI can help provide personalized and evidence-based care for patients who have delirium, using data analysis, natural language processing, and machine learning. AI can measure the effectiveness and quality of the delirium management, such as the use of medications, non-pharmacological interventions, and patient education. AI can also help optimize the care plan according to the patient's specific needs and preferences, and update the plan accordingly [96].
- AI can help prevent delirium by addressing the modifiable risk factors, such as pain, infection, dehydration, sleep deprivation, and sensory impairment. AI can also help promote the protective factors, such as cognitive stimulation, physical activity, and social interaction. For instance, researchers have developed a novel AI-based software that can prevent postoperative delirium by providing early detection and management of the condition [97–99].

Also, AI can play a pivotal role in:

1. **Risk prediction and early identification** with data analysis and continuous monitoring,
2. **Personalized care plans** with machine learning algorithms and adaptive interventions,
3. **Medication management** with algorithmic decision support and monitoring medication effects,
4. **Environmental optimization** with sensory monitoring and sleep pattern analysis,
5. **Predictive analytics for complication prevention** with data patterns and continuous observation,
6. **Rehabilitation and cognitive training** with virtual rehabilitation programs.

We should be mindful that the integration of AI into the healthcare ecosystem for delirium

prevention requires a collaborative effort among healthcare providers, data scientists, and technology experts. By leveraging the capabilities of AI, healthcare teams can move toward a more proactive, personalized, and data-driven approach to decrease the incidence and impact of delirium in patients, particularly in the vulnerable postoperative period.

Conclusions

The strategic integration of Artificial Intelligence (AI) into nursing care for patients undergoing eye surgeries heralds a paradigm shift in healthcare practices. Central to this transformation is the utilization of AI-powered communication tools, fostering seamless coordination among healthcare providers and enhancing interdisciplinary collaboration. Machine learning algorithms play a pivotal role in personalized patient education, tailoring information to individual needs and preferences. The strategic use of AI extends to risk assessment and stratification, empowering nurses to proactively address potential complications with precision and individualized attention. Looking ahead, emerging technologies like virtual reality, natural language processing, and sentiment analysis present promising avenues to revolutionize patient education and emotional support. In summary, the strategic deployment of AI in nursing care for eye surgeries not only optimizes efficiency but also elevates patient experiences, epitomizing a visionary approach that aligns technology with the holistic needs of individuals undergoing ocular interventions.

Acknowledgements I extend my heartfelt appreciation to Dr. Seaman and Dr. Alexander, both esteemed professors at the Pittsburgh School of Nursing, for their invaluable contributions to the revision of this book chapter. Their expertise, insights, and meticulous attention to detail have significantly enriched the quality and depth of the content. Also, I would like to acknowledge the contribution of ChatGPT (<https://chat.openai.com/>) in enriching the writing and final editing of this book chapter.

References

- Gutierrez L, Lim JS, Foo LL, Ng WY, Yip M, Lim GYS, et al. Application of artificial intelligence in cataract management: current and future directions. *Eye Vis.* 2022;9(1):3.
- Secinaro S, Calandra D, Secinaro A, Muthurangu V, Biancone P. The role of artificial intelligence in healthcare: a structured literature review. *BMC Med Inform Decis Mak.* 2021;21:1–23.
- Grzybowski A. Artificial intelligence in ophthalmology: promises, hazards and challenges. Springer; 2021.
- Tan TF, Thirunavukarasu AJ, Jin L, Lim J, Poh S, Teo ZL, et al. Artificial intelligence and digital health in global eye health: opportunities and challenges. *Lancet Glob Health.* 2023;11(9):e1432–43.
- Alowais SA, Alghamdi SS, Alsuehaby N, Alqahtani T, Alshaya AI, Almohareb SN, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Med Educ.* 2023;23(1):689.
- Väänänen A, Haataja K, Vehviläinen-Julkunen K, Toivanen P. AI in healthcare: a narrative review. *F1000Research.* 2021;10:6.
- Seibert K, Domhoff D, Bruch D, Schulte-Althoff M, Fürstenau D, Biessmann F, et al. Application scenarios for artificial intelligence in nursing care: rapid review. *J Med Internet Res.* 2021;23(11):e26522.
- Solanki SL, Pandrowala S, Nayak A, Bhandare M, Ambulkar RP, Shrikhande SV. Artificial intelligence in perioperative management of major gastrointestinal surgeries. *World J Gastroenterol.* 2021;27(21):2758.
- Robert N. How artificial intelligence is changing nursing. *Nurs Manage.* 2019;50(9):30.
- Ng ZQP, Ling LYJ, Chew HSJ, Lau Y. The role of artificial intelligence in enhancing clinical nursing care: a scoping review. *J Nurs Manag.* 2022;30(8):3654–74.
- Hohenstein J, Kizilcec RF, DiFranzo D, Aghajari Z, Mieczkowski H, Levy K, et al. Artificial intelligence in communication impacts language and social relationships. *Sci Rep.* 2023;13(1):5487.
- Maddox TM, Rumsfeld JS, Payne PR. Questions for artificial intelligence in health care. *JAMA.* 2019;321(1):31–2.
- Zahlan A, Ranjan RP, Hayes D. Artificial intelligence innovation in healthcare: literature review, exploratory analysis, and future research. *Technol Soc.* 2023;102321.
- Clancy TR. Artificial intelligence and nursing: the future is now. *JONA: J Nurs Admin.* 2020;50(3):125–7.
- Levivien C, Cavagna P, Grah A, Buronfosse A, Courseau R, Bézie Y, et al. Assessment of a hybrid decision support system using machine learning with artificial intelligence to safely rule out prescriptions from medication review in daily practice. *Int J Clin Pharm.* 2022;44(2):459–65.
- Dave M, Patel N. Artificial intelligence in healthcare and education. *Br Dent J.* 2023;234(10):761–4.
- Charow R, Jeyakumar T, Younus S, Dolatabadi E, Salhia M, Al-Mouaswas D, et al. Artificial intelligence education programs for health care professionals: scoping review. *JMIR Med Educ.* 2021;7(4):e31043.
- Guzman AL, Lewis SC. Artificial intelligence and communication: a human-machine communication research agenda. *New Media Soc.* 2020;22(1):70–86.
- Nespolo RG, Yi D, Cole E, Valikodath N, Luciano C, Leiderman YI. Evaluation of artificial intelligence-based intraoperative guidance tools for phacoemulsification cataract surgery. *JAMA Ophthalmol.* 2022;140(2):170–7.
- Xu H, Chen J, Tang T, Yang S. Cultivation path of compound talents in ophthalmic diagnosis, treatment, and nursing based on artificial intelligence. *J Clin Nurs Res.* 2022;6(5):106–11.
- Hinkle JL, Cheever KH. Brunner and Suddarth's textbook of medical-surgical nursing. Wolters Kluwer India Pvt Ltd; 2022.
- DeWit SC, Stromberg H, Dallred C. Medical-surgical nursing: concepts & practice. Elsevier Health Sciences; 2016.
- Lesin M, Sundov ZD, Jukic M, Puljak L. Postoperative pain in complex ophthalmic surgical procedures: comparing practice with guidelines. *Pain Med.* 2014;15(6):1036–42.
- de Oliveira Serra MAA, da Silva Filho FF, de Albuquerque AdO, dos Santos CAA, de Carvalho Junior JF, Silva RdA. Nursing care in the immediate postoperative period: a cross sectional study. *Online Braz J Nurs.* 2015;14(2):161–7.
- Fouad H, Hassanein AS, Soliman AM, Al-Feel H. Analyzing patient health information based on IoT sensor with AI for improving patient assistance in the future direction. *Measurement.* 2020;159:107757.
- Mahani MA, Farokhzadian J, Bahramnezhad F, Nematollahi M. Nurses' perceptions of critically ill patients' eye care: a qualitative study in Iran. *BMC Nurs.* 2023;22(1):56.
- Momeni Mehrjardi Z, Mirzaei S, Gohari M, Hafezieh A, Nasiriani K. Effect of training eye care clinical guideline for ICU patients on clinical competence of eye care in nurses. *Crit Care Res Pract.* 2021;1–6.
- Perry JP, Tullo AB. Care of the ophthalmic patient: a guide for nurses and health professionals. Springer; 2013.
- Jeong GH. Artificial intelligence, machine learning, and deep learning in women's health nursing. *Korean J Women Health Nurs.* 2020;26(1):5–9.
- Brom H, Carthon JMB, Ikeaba U, Chittams J. Leveraging electronic health records and machine learning to tailor nursing care for patients at high risk for readmissions. *J Nurs Care Qual.* 2020;35(1):27.
- Zabihi MR, Rashtiani S, Mashayekhi Y, Amirinia F, Gholamkar V, Kor S, et al. Applications of machine learning for hemodialysis nursing cares based on a

- machine learning algorithm. *J Nurs Rep Clin Pract*. 2023;1(1):4–9.
32. Egert M, Steward JE, Sundaram CP. Machine learning and artificial intelligence in surgical fields. *Indian J Surg Oncol*. 2020;11:573–7.
 33. Booth RG, Strudwick G, McBride S, O'Connor S, López ALS. How the nursing profession should adapt for a digital future. *bmj*. 2021;373.
 34. Kwon JY, Karim ME, Topaz M, Currie LM. Nurses “seeing forest for the trees” in the age of machine learning: using nursing knowledge to improve relevance and performance. *CIN: Comput Inform Nurs*. 2019;37(4):203–12.
 35. Chuang J, Shih KC, Chan TC, Wan KH, Jhanji V, Tong L. Preoperative optimization of ocular surface disease before cataract surgery. *J Cataract Refract Surg*. 2017;43(12):1596–607.
 36. Ramirez DA, Brodie FL, Rose-Nussbaumer J, Ramanathan S. Anxiety in patients undergoing cataract surgery: a pre-and postoperative comparison. *Clin Ophthalmol*. 2017;1979–86.
 37. King CR, Shambe A, Abraham J. Potential uses of AI for perioperative nursing handoffs: a qualitative study. *JAMIA Open*. 2023;6(1):o0ad015.
 38. Ronquillo CE, Peltonen LM, Pruinelli L, Chu CH, Bakken S, Beduschi A, et al. Artificial intelligence in nursing: priorities and opportunities from an international invitational think-tank of the Nursing and Artificial Intelligence Leadership Collaborative. *J Adv Nurs*. 2021;77(9):3707–17.
 39. Abuzaid MM, Elshami W, Fadden SM. Integration of artificial intelligence into nursing practice. *Heal Technol*. 2022;12(6):1109–15.
 40. Hopkins BS, Mazmudar A, Driscoll C, Svet M, Goergen J, Kelsten M, et al. Using artificial intelligence (AI) to predict postoperative surgical site infection: a retrospective cohort of 4046 posterior spinal fusions. *Clin Neurol Neurosurg*. 2020;192:105718.
 41. Bian Y, Xiang Y, Tong B, Feng B, Weng X. Artificial intelligence-assisted system in postoperative follow-up of orthopedic patients: exploratory quantitative and qualitative study. *J Med Internet Res*. 2020;22(5):e16896.
 42. Mofatteh M. Neurosurgery and artificial intelligence. *AIMS Neurosci*. 2021;8(4):477.
 43. Da Costa CA, Pasluosta CF, Eskofier B, Da Silva DB, da Rosa RR. Internet of health things: toward intelligent vital signs monitoring in hospital wards. *Artif Intell Med*. 2018;89:61–9.
 44. Ng FYC, Thirunavukarasu AJ, Cheng H, Tan TF, Gutierrez L, Lan Y et al. Artificial intelligence education: an evidence-based medicine approach for consumers, translators, and developers. *Cell Rep Med*. 2023;4(10).
 45. O'Connor S, Yan Y, Thilo FJ, Felzmann H, Dowding D, Lee JJ. Artificial intelligence in nursing and midwifery: a systematic review. *J Clin Nurs*. 2023;32(13–14):2951–68.
 46. Bakshi SK, Lin SR, Ting DSW, Chiang MF, Chodosh J. The era of artificial intelligence and virtual reality: transforming surgical education in ophthalmology. *Br J Ophthalmol*. 2020.
 47. Swearingen PL, Wright J. All-in-one nursing care planning resource-e-book: medical-surgical, pediatric, maternity, and psychiatric-mental health. Elsevier Health Sciences; 2019.
 48. Woodnutt S, Allen C, Snowden J, Flynn M, Hall S, Libberton P, et al. Could artificial intelligence write mental health nursing care plans? *J Psychiatr Mental Health Nurs*. 2023.
 49. Saab MM, Landers M, Murphy D, O'Mahony B, Cooke E, O'Driscoll M, et al. Nursing students' views of using virtual reality in healthcare: a qualitative study. *J Clin Nurs*. 2022;31(9–10):1228–42.
 50. Pensieri C, Pennacchini M. Virtual reality in medicine. Handbook on 3D3C platforms: applications and tools for three dimensional systems for community, creation and commerce. Springer; 2016. p. 353–401.
 51. Chen F-Q, Leng Y-F, Ge J-F, Wang D-W, Li C, Chen B, et al. Effectiveness of virtual reality in nursing education: meta-analysis. *J Med Internet Res*. 2020;22(9):e18290.
 52. Choi J, Thompson CE, Choi J, Waddill CB, Choi S. Effectiveness of immersive virtual reality in nursing education: systematic review. *Nurse Educ*. 2022;47(3):E57–61.
 53. Saab MM, Hegarty J, Murphy D, Landers M. Incorporating virtual reality in nurse education: a qualitative study of nursing students' perspectives. *Nurse Educ Today*. 2021;105:105045.
 54. Rizzetto F, Bernareggi A, Rantas S, Vanzulli A, Vertemati M. Immersive virtual reality in surgery and medical education: diving into the future. *Am J Surg*. 2020;220(4):856–7.
 55. Bryant L, Brunner M, Hemsley B. A review of virtual reality technologies in the field of communication disability: implications for practice and research. *Disabil Rehabil Assist Technol*. 2019;15.
 56. Shorey S, Ang E, Ng ED, Yap J, Lau LST, Chui CK. Communication skills training using virtual reality: a descriptive qualitative study. *Nurse Educ Today*. 2020;94:104592.
 57. Doupi P, van der Lei J. Design and implementation considerations for a personalized patient education system in burn care. *Int J Med Informatics*. 2005;74(2–4):151–7.
 58. Hiligsmann M, Ronda G, van der Weijden T, Boonen A. The development of a personalized patient education tool for decision making for postmenopausal women with osteoporosis. *Osteoporos Int*. 2016;27:2489–96.
 59. Bastable SB. Essentials of patient education. Jones & Bartlett Learning; 2016.
 60. Talyshinskii A, Naik N, Hameed BZ, Juliebø-Jones P, Somani BK. Potential of AI-driven chatbots in urology: revolutionizing patient care through artificial intelligence. *Curr Urol Rep*. 2023;1–10.

61. Sarella PNK, Mangam VT. AI-driven natural language processing in healthcare: transforming patient-provider communication. *Ind J Pharm Pract.* 2024;17(1).
62. Aggarwal A, Tam CC, Wu D, Li X, Qiao S. Artificial intelligence-based chatbots for promoting health behavioral changes: systematic review. *J Med Internet Res.* 2023;25:e40789.
63. Keskinbora KH. Medical ethics considerations on artificial intelligence. *J Clin Neurosci.* 2019;64:277–82.
64. McGreevey JD, Hanson CW, Koppel R. Clinical, legal, and ethical aspects of artificial intelligence–assisted conversational agents in health care. *JAMA.* 2020;324(6):552–3.
65. Maheshwari K, Ruetzler K, Saugel B. Perioperative intelligence: applications of artificial intelligence in perioperative medicine. *J Clin Monit Comput.* 2020;34:625–8.
66. Esmaeilzadeh P, Mirzaei T, Dharanikota S. Patients' perceptions toward human–artificial intelligence interaction in health care: experimental study. *J Med Internet Res.* 2021;23(11):e25856.
67. Lin SY, Mahoney MR, Sinsky CA. Ten ways artificial intelligence will transform primary care. *J Gen Intern Med.* 2019;34:1626–30.
68. Butow P, Hoque E. Using artificial intelligence to analyse and teach communication in healthcare. *The Breast.* 2020;50:49–55.
69. Derevianko A, Pizzoli SFM, Pesapane F, Rotili A, Monzani D, Grasso R, et al. The use of Artificial Intelligence (AI) in the radiology field: what is the state of doctor-patient communication in cancer diagnosis? *Cancers.* 2023;15(2):470.
70. Scheder-Bieschin J, Blümke B, de Buijzer E, Echterdiek F, Nacsa J, Ondresik M, et al. Improving emergency department patient-doctor conversation through an artificial intelligence symptom taking tool: an action-oriented design pilot study. *medRxiv.* 2020:2020.11.13.20230953.
71. Seaback WW. Nursing process: concepts and applications. 2012.
72. Alfaro-LeFevre R. Applying nursing process: the foundation for clinical reasoning. Lippincott Williams & Wilkins; 2012.
73. Gordon M. Manual of nursing diagnosis. Jones & Bartlett Publishers; 2014.
74. Locsin RC, Purnell M. A contemporary nursing process: the (un) bearable weight of knowing in nursing. Springer Publishing Company; 2009.
75. Löfgren U, Wälivaara B-M, Strömbäck U, Lindberg B. The nursing process: a supportive model for nursing students' learning during clinical education—a qualitative study. *Nurse Educ Pract.* 2023;72:103747.
76. Ackley BJ, Ladwig GB. Nursing diagnosis handbook-e-book: an evidence-based guide to planning care. Elsevier Health Sciences; 2010.
77. ChatGPT O. optimizing language models for dialogue. OpenAI. Retrieved 01/01/2024 from <https://chat.openai.com/>.
78. Aya AG, Pouchain P-H, Thomas H, Ripart J, Cuvillon P. Incidence of postoperative delirium in elderly ambulatory patients: a prospective evaluation using the FAM-CAM instrument. *J Clin Anesth.* 2019;53:35–8.
79. Rengel KF, Pandharipande PP, Hughes CG. Postoperative delirium. *La Presse Médicale.* 2018;47(4):e53–64.
80. Raats JW, Steunenbergh SL, de Lange DC, van der Laan L. Risk factors of post-operative delirium after elective vascular surgery in the elderly: a systematic review. *Int J Surg.* 2016;35:1–6.
81. Vincent J-L, Moore FA, Bellomo R, Marini JJ. Textbook of critical care. Elsevier Health Sciences; 2022.
82. Williams ST, Dhesi JK, Partridge JS. Distress in delirium: causes, assessment and management. *Eur Geriatric Med.* 2020;11:63–70.
83. Wilson JE, Mart MF, Cunningham C, Shehabi Y, Girard TD, MacLulich AM, et al. Delirium. *Nat Rev Dis Primers.* 2020;6(1):90.
84. Aldecoa C, Bettelli G, Bilotta F, Sanders RD, Audisio R, Borozdina A, et al. European Society of Anaesthesiology evidence-based and consensus-based guideline on postoperative delirium. *Eur J Anaesthesiol* | EJA. 2017;34(4):192–214.
85. Xie Q, Wang X, Pei J, Wu Y, Guo Q, Su Y, et al. Machine learning–based prediction models for delirium: a systematic review and meta-analysis. *J Am Med Direct Assoc.* 2022;23(10):1655–68. e6.
86. Nowakowska K, Sakellarios A, Kaźmierski J, Fotiadis DI, Pezoulas VC. AI-enhanced predictive modeling for identifying depression and delirium in cardiovascular patients scheduled for cardiac surgery. *Diagnostics.* 2023;14(1):67.
87. Vlisides P, Avidan M. Recent advances in preventing and managing postoperative delirium. *F1000Research.* 2019;8.
88. Emme C. “It should not be that difficult to manage a condition that is so frequent”: a qualitative study on hospital nurses' experience of delirium guidelines. *J Clin Nurs.* 2020;29(15–16):2849–62.
89. Khan A, PhD OB, Oh-Park M, PhD NAF, PhD MS, Oldham M. Preventing delirium takes a village: systematic review and meta-analysis of delirium preventive models of care. *J Hosp Med.* 2019;14(9):558–64.
90. Chua S, Wrigley S, Hair C, Sahathevan R. Prediction of delirium using data mining: a systematic review. *J Clin Neurosci.* 2021;91:288–98.
91. Rivosecchi RM, Smithburger PL, Svec S, Campbell S, Kane-Gill SL. Nonpharmacological interventions to prevent delirium: an evidence-based systematic review. *Crit Care Nurse.* 2015;35(1):39–49.
92. Faught DD. Delirium: the nurse's role in prevention, diagnosis, and treatment. *Medsurg Nurs.* 2014;23(5).

93. Zhao H, You J, Peng Y, Feng Y. Machine learning algorithm using electronic chart-derived data to predict delirium after elderly hip fracture surgeries: a retrospective case-control study. *Frontiers in Surgery*. 2021;8:634629.
94. Stollings JL, Kotfis K, Chanques G, Pun BT, Pandharipande PP, Ely EW. Delirium in critical illness: clinical manifestations, outcomes, and management. *Intensive Care Med*. 2021;47(10):1089–103.
95. Mulkey MA, Huang H, Albanese T, Kim S, Yang B. Supervised deep learning with vision transformer predicts delirium using limited lead EEG. *Sci Rep*. 2023;13(1):7890.
96. Bhattacharyya A, Sheikhalishahi S, Torbic H, Yeung W, Wang T, Birst J, et al. Delirium prediction in the ICU: designing a screening tool for preventive interventions. *JAMIA open*. 2022;5(2):ooac048.
97. Leinert C, Fotteler M, Kocar TD, Dallmeier D, Kestler HA, Wolf D, et al. Supporting SURgery with GERiatric Co-Management and AI (SURGE-Ahead): a study protocol for the development of a digital geriatrician. *PLoS ONE*. 2023;18(6):e0287230.
98. Yoon H-K, Yang H-L, Jung C-W, Lee H-C. Artificial intelligence in perioperative medicine: a narrative review. *Korean J Anesthesiol*. 2022;75(3):202–15.
99. Bellini V, Valente M, Bertorelli G, Pifferi B, Craca M, Mordonini M, et al. Machine learning in perioperative medicine: a systematic review. *J Anesth Analgesia Crit Care*. 2022;2(1):2.

Exploring the Potential of ChatGPT in Ophthalmology: A Vision for Future Healthcare

28

Anfisa Ayalon, Lauren M. Wasser,
Andrew M. Williams, and José-Alain Sahel

Abstract

Generative artificial intelligence (AI) is transforming many industries, including the field of medicine. ChatGPT, an AI chatbot powered by a large language model, has rapidly gained popularity since its launch. This chapter explores ChatGPT's diverse applications in ophthalmology, including enhancing clinical decision-making and tasks, improving patient and medical education, and assisting in research. The authors tested this technology by writing the Results section of this chapter, subsequently reflecting on the user experience and identifying potential challenges of the technology.

Keywords

Artificial intelligence · ChatGPT ·
Ophthalmology · AI challenges · Medical
chatbot

Anfisa Ayalon and Lauren M. Wasser are contributed equally to this work.

A. Ayalon (✉) · L. M. Wasser · A. M. Williams ·
J.-A. Sahel
Department of Ophthalmology, University of Pittsburgh,
The UPMC Vision Institute, Pittsburgh, PA, USA
e-mail: doctor.anfi@gmail.com

Introduction

ChatGPT is an artificial intelligence (AI) chatbot built on large language models that uses natural language processing to engage in human-like conversations with users (OpenAI; San Francisco, CA) [1]. The initial ChatGPT-1 model, released in 2018, was not publicly available and comprised 117 million parameters [2]. By 2020, the ChatGPT-3 model boasted 175 billion parameters by 2020 [2]. This technology demonstrated continuous rapid improvement in functionality, speed, and the quality of its output (Fig. 28.1) [2]. On November 30th, 2022, it was launched for public use, and within two months it had reached 100 million users, making it one of the most rapidly growing consumer applications to date [3]. The rising usage of ChatGPT on Google Trends illustrates its growing interest (Fig. 28.2). Since its release, the application of ChatGPT has been explored across various medical disciplines [4–6]. The culmination of its development in medicine in 2023 was the announcement of the integration of GPT-4 into America's largest healthcare software company, Epic Systems (Verona, WI) [7]. This marked the first system-wide incorporation of ChatGPT into the electronic health record, signifying an important milestone in integrating AI in medicine [7]. In this chapter, we explore the application of ChatGPT in ophthalmology and examine both the user experience and the potential challenges of this technology.

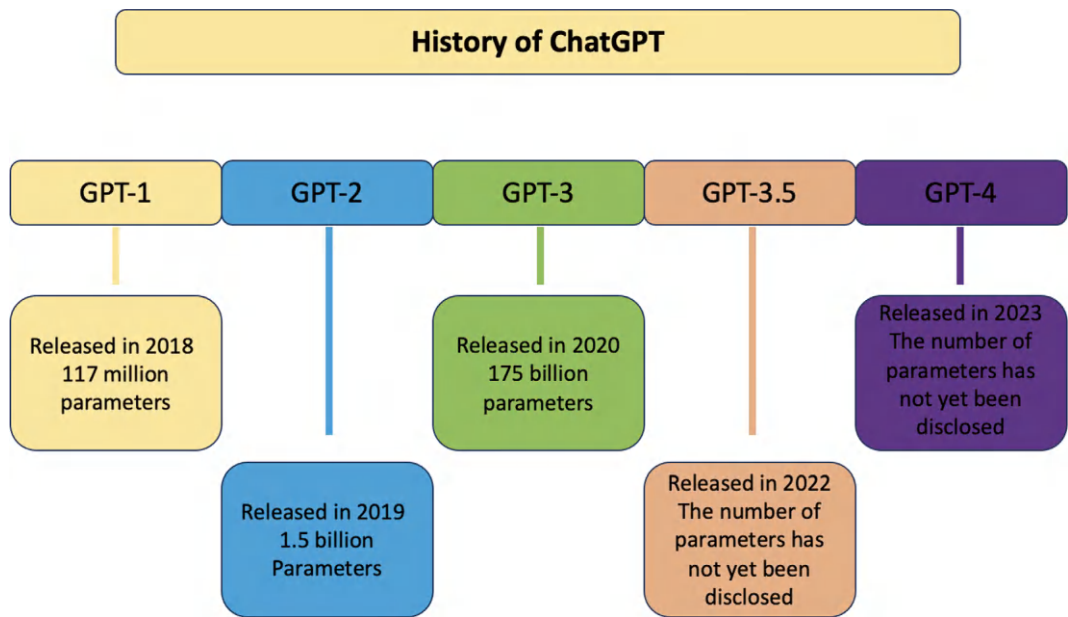


Fig. 28.1 History of ChatGPT. ChatGPT has significantly improved in its functioning, marked by a substantial increase in its number of parameters

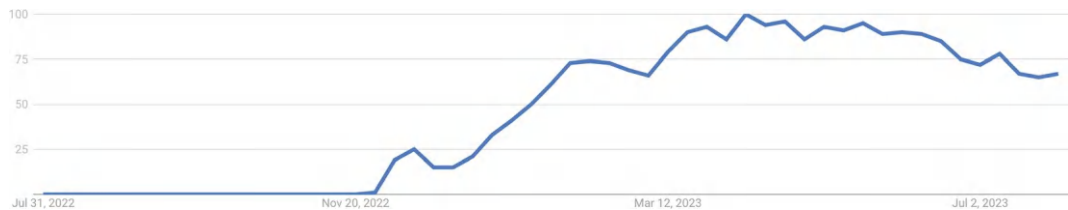


Fig. 28.2 Google Trends analysis of the search term ‘ChatGPT’ over 12 months worldwide. Google Trends depicts global interest in ChatGPT over time since its release in November 2022, with peaks showing increased queries

Materials and Methods

The Results section of the chapter was generated by the NLP-based technology, ChatGPT (OpenAI; San Francisco, CA, March 14 Version), in response to the following prompt provided by the authors: [Act as a medical researcher writing an article about ChatGPT and ophthalmology]. ChatGPT utilized the provided information to generate responses, which were then reviewed by the authors for accuracy. This assessment was conducted to evaluate the efficiency and effectiveness of ChatGPT in content generation and to assess its applicability in ophthalmology academic writing. The Introduction

and Discussion sections of this chapter were written by the authors. These sections explore the application of ChatGPT in the field of ophthalmology, potential challenges, key considerations related to integrating this technology into ophthalmology practices and provide a critical analysis of the authors’ experience in utilizing ChatGPT.

Results

In recent years, the interplay between artificial intelligence (AI) and medicine has radically transformed the landscape of patient care,

diagnostics, and research methodologies. One such AI model, ChatGPT, has begun to make significant inroads into the realm of ophthalmology. Developed by OpenAI, ChatGPT, rooted in the GPT-4 architecture, serves as a compelling example of the potential that conversational AI models hold for eye care professionals and patients alike.

Patient education is fundamental in ophthalmology. Whether discussing the intricacies of an ocular disease, post-operative care, or the importance of regular check-ups, eye care professionals must ensure that their patients understand, remember, and act upon the information given. Enter ChatGPT. Its conversational design makes it an ideal tool to simplify complex terminologies and concepts related to the eye, tailor information to the patient's level of understanding, and reinforce key instructions. Additionally, the model's multilingual capabilities cater to diverse patient populations, mitigating language barriers.

Beyond patient education, AI models like ChatGPT could provide invaluable support in clinical decision-making. As ophthalmologists are inundated with vast amounts of patient data from imaging techniques like Optical Coherence Tomography, fundus photography, and visual field tests, an intelligent system can assist in the rapid analysis and interpretation of these results. By "conversing" with ChatGPT, doctors can quickly get insights, differential diagnoses, and up-to-date research information, bolstering their confidence in clinical decisions. Ophthalmology, like any medical specialty, is ever evolving. New studies, novel treatments, and evolving guidelines emerge regularly. For practitioners, staying current can be a challenge. ChatGPT, with its expansive database, offers an immediate reference tool, enabling doctors to query the latest research, compare clinical outcomes, and even discuss complex cases with an AI-backed "colleague." While it doesn't replace peer consultations, it supplements them, especially in situations where immediate feedback is crucial.

In regions with limited access to ophthalmologic care, AI tools can help bridge the gap. Telemedicine has already shown potential in this

space but combining it with AI-driven conversational platforms can further amplify its impact. Patients can discuss symptoms, get preliminary advice, or be guided through basic eye tests, all while interfacing with ChatGPT. Though it cannot replace a physical exam or the nuanced diagnosis of a skilled ophthalmologist, it can serve as an interim solution, ensuring patients in remote areas aren't left in the dark, both figuratively and literally.

Discussion

In ophthalmology, ChatGPT has the potential to be useful in various settings, including medical education, clinical decision-making and assistance with clinic tasks, patient education, and research [8–14].

Augmenting Clinic Decision Making and Clinic Tasks

In the clinical setting, ChatGPT has already demonstrated superiority over preexisting AI technologies in assisting with ophthalmological diagnoses [15]. In 90% of cases it achieved a correct diagnosis and provided accurate differential diagnoses in 100% of cases, outperforming a well-known diagnostic support tool—Isabel Pro Differential Diagnosis Generator [15]. Integrating ChatGPT into electronic health records systems such as Epic can enhance the physician experience in several ways. Firstly, it can significantly reduce the time spent on writing outpatient clinic and operative notes, currently, a time-consuming process [16]. Physicians can input a short prompt, and ChatGPT, trained on specific data, can autonomously complete sentences and paragraphs. Secondly, this technology can contribute to real-time diagnosis support, offering specific guidance or tips to physicians when collecting patient history and attempting to establish a diagnosis [13]. Thirdly, ChatGPT can recommend treatment options, assist in prescribing ophthalmic drugs, and calculate dosages, while

also highlighting potential adverse events in patients with underlying conditions. Fourthly, ChatGPT can compile a summary of all the risk factors and patient parameters, facilitating the prediction of disease course and outcome. Other clinical applications have included assisting with clinic tasks such as prior authorizations and discharge summaries, providing quality responses to eye care questions, assisting in ophthalmic triage, and offering guidance for physician decision-making [8, 11–13, 16–21].

Ophthalmic Patient Education

Creating patient education materials is a costly process, as demonstrated by a study conducted across Ontario cancer centers, revealing an average annual cost of \$85,220 for pamphlets alone [22]. Patient consultations with medical staff are expected to incur even higher expenses. Despite the costs, patient education holds significant importance. The emergence of new AI technology presents an opportunity to optimize patient education, reduce expenses associated with material development and address general patient inquiries. A recent report indicated that postoperative instructions delivered by ChatGPT supplied adequate procedure-specific information to ophthalmology patients, covering common procedures including intravitreal injections and cataract surgery [18]. Another group conducted a study involving 112 questions from oculoplastic patients to assess ChatGPT's responses [19]. Their findings indicated that ChatGPT-3.5 provided comprehensive responses with a 5.1% error rate. In 10.5% of cases, responses by ChatGPT contained both correct and incorrect/outdated information [19]. While ChatGPT shows great potential for enhancing patient education and reducing costs, physician oversight is necessary to ensure the accuracy of its output [19]. Additionally, with the rapid-paced development of this technology, newer versions like ChatGPT-4, utilizing more updated data from the internet, may offer enhanced results and warrant further study.

Ophthalmic Medical Education

In ophthalmic medical education, ChatGPT holds promise as an accessible source of information, real-time supplemental learning, and in simulating virtual patient scenarios [23]. However, the real-world application of ChatGPT for these tasks remains unexplored. Several studies indicate that the earliest versions of ChatGPT demonstrated low accuracy in answering resident-level ophthalmology questions and were ineffective in assisting with preparation for ophthalmic examinations [24, 25]. In these studies, ChatGPT correctly answered only 56% of questions in the Ophthalmic Knowledge Assessment Program exam and 46% in the OphthoQuestions free trial for board certification examinations [24, 25]. A recent study on the performance of GPT-4 has recently demonstrated improved performance [26]. Ongoing research into how ChatGPT can enhance ophthalmic education is necessary to fully understand its potential.

Research

In ophthalmic research settings, ChatGPT can be widely implemented for a range of tasks. These include brainstorming ideas, assistance with programming, checking for data errors, and academic medical writing [27–29]. The open-sourced, approachable nature of this cutting-edge technology sparked our curiosity regarding its capabilities in ophthalmologic academic writing. We decided to test ChatGPT and see how it would fare in writing the Results section of this chapter. During the preparation of this book chapter, ChatGPT underwent multiple upgrades. Faced with the rapid pace at which ChatGPT models were being upgraded, we re-engaged with ChatGPT software for the Results section several times. Doing so, we realized that the upgraded versions were better equipped to handle more nuanced input and deliver higher-quality output. Consequently, we utilized GPT-4 to rewrite the final version of the Results. The generated text was concise, well-structured, and

articulately written. Readers may reflect on the quality of the ChatGPT-generated text in the Results section of this chapter.

Alternate Generative AI Platforms

Apart from ChatGPT (OpenAI), the market offers several advanced chatbots, including Bing Chat (Microsoft), Bard Gemini (Google), Grock (xAI), Llama (Meta), Claude (Anthropic AI), and Pi (Inflection AI) [30]. These chatbots exhibit varying levels and speeds of performance, along with differences in accessibility [30]. Currently, GPT-4 has demonstrated superior performance in handling common-sense multiple-choice questions compared to other chatbots [30]. While searching the literature in PubMed for chatbot performance in ophthalmology, it becomes apparent that a significant portion of the literature focuses on ChatGPT and its applications in ophthalmology. This suggests that the surge in popularity has led to more extensive testing of ChatGPT within the field of ophthalmology, compared to other chatbots.

Technology Limitations

While ChatGPT offers remarkable opportunities in ophthalmology, it also presents numerous challenges. Upon investigation, we learned that GPT-4, available through a paid subscription, is the only version with internet access and without information cutoffs. Meanwhile, the freely available GPT-3.5 completed its training phase in 2021 and is unable to access information after this time [31]. As a result, it is missing several years of critical knowledge, creating a discrepancy between the cutting-edge nature of this technology and the outdated content it can generate. When considering integrating ChatGPT in patient education, a lack of up-to-date information poses a significant limitation.

ChatGPT can produce misinterpreted output otherwise known as AI hallucinations [32]. When we asked the chatbot to provide references for the generated Results section, the older versions of GPT 3.5 created ersatz references (Fig. 28.3). GPT-4 was more forthcoming regarding its knowledge cutoff and provided

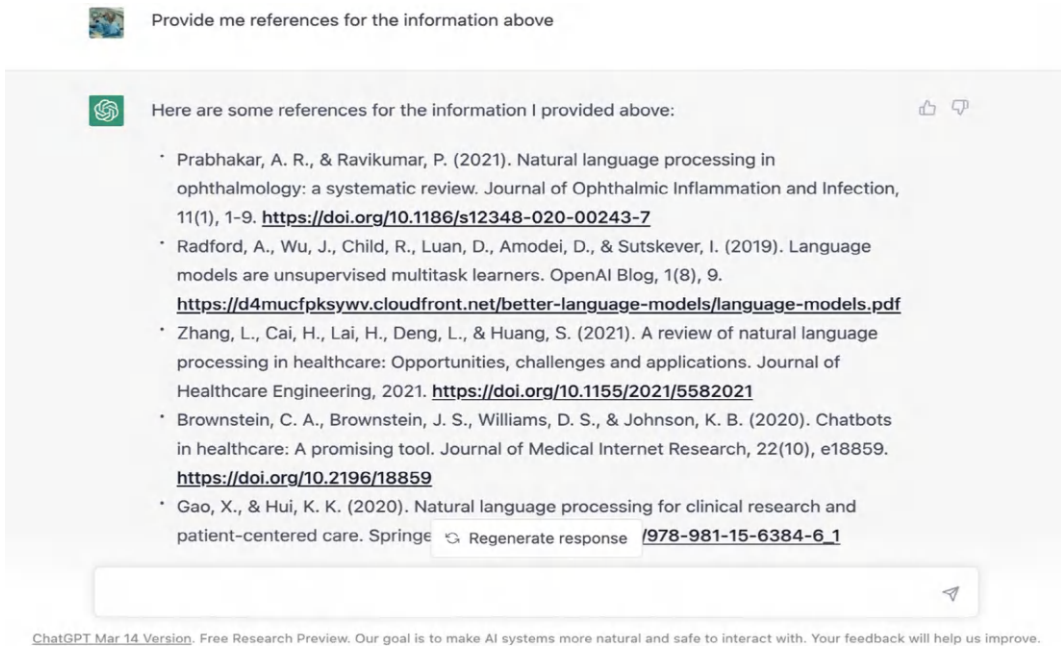


Fig. 28.3 Falsified references generated by ChatGPT (GPT3.5, March 14 Version, 2023)

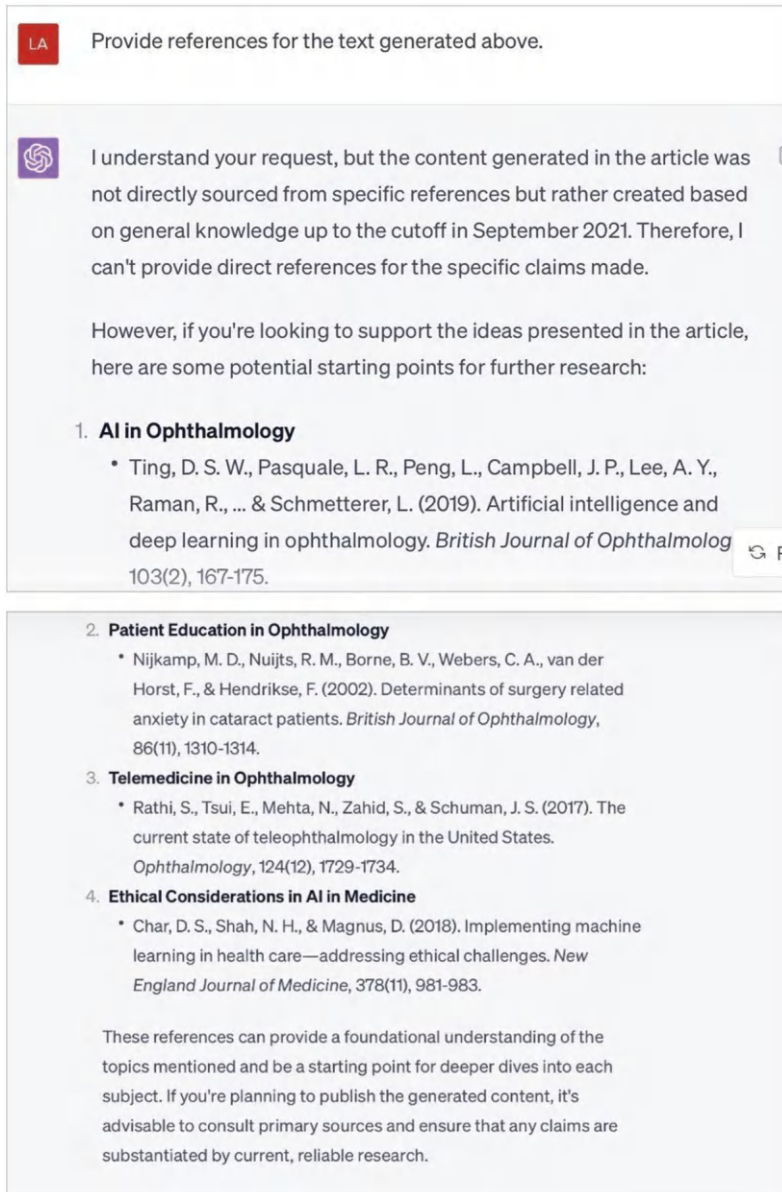


Fig. 28.4 Improved response of ChatGPT (GPT4, August 3 Version, 2023) when prompted to provide references

accurate references, however, they were not always directly related to the generated text (Fig. 28.4). This limitation should be considered during its integration in patient education and academic writing.

ChatGPT's responses rely on user input, which can lead to subjective outputs [33]. The model requires user caution. This becomes more

significant when users are unfamiliar with the subject matter, as they risk receiving misleading information and acting on potentially harmful medical advice. Effective prompt engineering is critical to elicit precise and dependable results. Establishing best practices for interacting with ChatGPT can ensure more consistent and accurate outcomes.

It is important to consider the ethical implications of ChatGPT during its integration into medicine and healthcare. Built on unsupervised datasets from the internet, ChatGPT may inadvertently amplify pre-existing biases [4]. Likewise, historical medical biases related to gender, race, or socioeconomic status, if present in the training data, may be perpetuated in the model's outputs. Proposals for mitigating these biases have included data transparency and human oversight for equitable and patient-centric outcomes [34]. Ophthalmology academic journals have already begun to address the ethical considerations and authorship guidelines for the publication of articles produced by this technology [35, 36]. The incorporation of ChatGPT in manuscript writing may necessitate updates to the review process of academic journals to ensure they remain current. Furthermore, considering ChatGPT's inability to cross-reference provided information, and the author's inability to manually reference the source of AI-generated text, is it sufficient to merely acknowledge its use in text creation? This difficulty in pinpointing the source of the output raises concerns about copyright issues and plagiarism. As AI's presence in academia becomes more mainstream, editorial teams and peer reviewers will need to establish clear guidelines on the acceptable use of AI in research and manuscript preparation.

ChatGPT technology, like many other machine learning-based tools, requires collection and storage of massive amounts of input data from users. This raises concerns regarding potential leakage of sensitive information and data utilization by third parties [37]. On April 25, 2023, OpenAI introduced a feature which allows the user to disable their chat history in ChatGPT [38]. Conversations with disabled chat histories are not utilized for model training or improvement and are not saved or visible in the history sidebar [38]. However, according to the OpenAI private policy, the company may disclose the user's information to the company's "affiliates, vendors, and service providers to process in accordance with our [OpenAI]

instructions; to law enforcement and other third parties for the legal reasons described above; to parties involved in transactions; and to other users and third parties you [the user] choose to share it with" [39]. In the medical field, where privacy concerns are especially sensitive, these policies may require more stringent handling. Therefore, legal measures and safeguards specifically tailored to protect medical data privacy should be considered when using ChatGPT technology in healthcare.

In summary, ChatGPT presents as an accessible tool with the promise to contribute to the field of ophthalmology. It is important to recognize its limitations, including outdated information and biases. Additionally, utilizing this technology for patient care introduces significant data privacy concerns. Despite these challenges, this technology is increasingly becoming embedded across various sectors, including medicine, and is here to stay. It is essential to find a balanced approach to integrating this technology, enhancing patient care and research, while also safeguarding our vulnerable patient population.

References

1. Haque MU, Dharmadasa I, Sworna ZT et al (2022) I think this is the most disruptive technology: exploring sentiments of ChatGPT early adopters using Twitter data. arXiv preprint [arXiv:2212.05856](https://arxiv.org/abs/2212.05856)
2. Ray PP. ChatGPT: a comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet Things Cyber-Physical Syst.* 2023;3:121–54. <https://doi.org/10.1016/j.iotcps.2023.04.003>.
3. Tate T, Doroudi S, Ritchie D, Xu Y (2023) Educational research and AI-generated writing: confronting the coming tsunami. <https://doi.org/10.35542/osf.io/4mec3>.
4. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: An overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Front Artif Intell.* 2023;6:1169595. <https://doi.org/10.3389/frai.2023.1169595>.
5. Howard A, Hope W, Gerada A. ChatGPT and antimicrobial advice: The end of the consulting infection doctor? *Lancet Infect Dis.* 2023;23(4):405–6. [https://doi.org/10.1016/S1473-3099\(23\)00113-5](https://doi.org/10.1016/S1473-3099(23)00113-5).

6. Patel SB, Lam K. ChatGPT: the future of discharge summaries? *Lancet Digit Health*. 2023;5(3):e107–8. [https://doi.org/10.1016/S2589-7500\(23\)00021-3](https://doi.org/10.1016/S2589-7500(23)00021-3).
7. Benj E (2018) Gpt-4 will hunt for trends in medical records thanks to microsoft and epic. Available at: <https://arstechnica.com/information-technology/2023/04/gpt-4-will-hunt-for-trends-in-medical-records-thanks-to-microsoft-and-epic/>. Accessed 18 Apr 2023
8. Antaki F, Touma S, Milad D, et al. Evaluating the performance of ChatGPT in ophthalmology: an analysis of its successes and shortcomings. *Ophthalmology Sci*. 2023;3(4):100324. <https://doi.org/10.1016/j.xops.2023.100324>.
9. Ting DSJ, Tan TF, Ting DSW (2023) ChatGPT in ophthalmology: the dawn of a new era? Eye. Preprint at <https://doi.org/10.1038/s41433-023-02619-4>
10. Panthier C, Gatineau D. Success of ChatGPT, an ai language model, in taking the french language version of the European board of ophthalmology examination: a novel approach to medical knowledge assessment. *J Fr Ophthalmol*. 2023;46(7):706–11. <https://doi.org/10.1016/j.jfo.2023.05.006>.
11. Waisberg E, Ong J, Zaman N, et al. Gpt-4 for triaging ophthalmic symptoms. *Eye*. 2023;37(18):3874–5. <https://doi.org/10.1038/s41433-023-02595-9>.
12. Lyons RJ, Arepalli SR, Fromal O (2023) Artificial intelligence chatbot performance in triage of ophthalmic conditions. *Can J Ophthalmol*. Preprint at <https://doi.org/10.1016/j.cjco.2023.07.016>
13. Ali MJ. ChatGPT and lacrimal drainage disorders: performance and scope of improvement. *Ophthalm Plast Reconstr Surg*. 2023;39(3):221–5. <https://doi.org/10.1097/IOP.0000000000002418>.
14. Momenaei B, Wakabayashi T, Shahlaee A, et al. Appropriateness and readability of ChatGPT -4 generated responses for surgical treatment of retinal diseases. *Ophthalmol Retina*. 2023;7(10):862–8. <https://doi.org/10.1016/j.oret.2023.05.022>.
15. Balas M, Ing EB. Conversational ai models for ophthalmic diagnosis: Comparison of ChatGPT and the Isabel pro differential diagnosis generator. *JFO Open Ophthalmology*. 2023;1:100005. <https://doi.org/10.1016/j.jfop.2023.100005>.
16. Singh S, Djalilian A, Ali MJ. ChatGPT and ophthalmology: exploring its potential with discharge summaries and operative notes. *Semin Ophthalmol*. 2023;38(5):503–7. <https://doi.org/10.1080/08820538.2023.2209166>.
17. Bernstein IA, Zhang YV, Govil D, et al. Comparison of ophthalmologist and large language model chatbot responses to online patient eye care questions. *JAMA Netw Open*. 2023;6(8):e2330320–e2330320. <https://doi.org/10.1001/jamanetworkopen.2023.30320>.
18. Nanji K, Yu CW, Wong TY, et al. Evaluation of post-operative ophthalmology patient instructions from ChatGPT and Google Search. *Can J Ophthalmol*. 2023;S0008–4182(23):00306–X. <https://doi.org/10.1016/j.cjco.2023.10.001>.
19. Al-Sharif EM, Penteadou RC, Dib El Jalbout N, et al. Evaluating the accuracy of ChatGPT and Google BARD in fielding oculoplastic patient queries: a comparative study on artificial versus human intelligence. *Ophthalmic Plast Reconstr Surg*. 2024. <https://doi.org/10.1097/IOP.0000000000002567>. doi:10.1097/IOP.0000000000002567.
20. Shemer A, Cohen M, Altarescu A, et al. Diagnostic capabilities of ChatGPT in ophthalmology. *Graefes Arch Clin Exp Ophthalmol*. 2024. <https://doi.org/10.1007/s00417-023-06363-z.6>.
21. Madadi Y, Delsoz M, Lao PA, et al. ChatGPT Assisting Diagnosis of Neuro-ophthalmology Diseases Based on Case Reports. 2023;128:1356. <https://doi.org/10.1101/2023.09.13.23295508>
22. Papadakis J, Samoil D, Giannopoulos E, et al. The cost of patient education materials development: opportunities to identify value and priorities. *J Cancer Educ*. 2020;37(3):834–42. <https://doi.org/10.1007/s13187-020-01893-0>.
23. Gurnani B, Kaur K. Leveraging ChatGPT for ophthalmic education: A critical appraisal. *Eur J Ophthalmol*. 2023. <https://doi.org/10.1177/11206721231215862>.
24. Teebagay S, Colwell L, Wood E, et al. ChatGPT in ophthalmology: a pilot study. *Invest Ophthalmol Vis Sci*. 2023;64(8):5376–5376.
25. Mihalache A, Popovic MM, Muni RH. Performance of an artificial intelligence chatbot in ophthalmic knowledge assessment. *JAMA Ophthalmol*. 2023;141(6):589–97. <https://doi.org/10.1001/jamaophthalmol.2023.1144>.
26. Teebagay S, Colwell L, Wood E, Yaghy A, Faustina M. Improved performance of ChatGPT-4 on the OKAP examination: a comparative study with ChatGPT-3.5. *J Acad Ophthalmol*. 2023;15(2):e184–e187. <https://doi.org/10.1055/s-0043-1774399>
27. Ahaley SS, Pandey A, Juneja SK, et al. ChatGPT in medical writing: a game-changer or a gimmick? *Perspect Clin Res*. 2023;10:4103. https://doi.org/10.4103/picr.picr_167_23.
28. Meyer JG, Urbanowicz RJ, Martin PC, et al. ChatGPT and large language models in academia: opportunities and challenges. *BioData Min*. 2023;16(1):20. <https://doi.org/10.1186/s13040-023-00339-9>.
29. Lingard L. Writing with ChatGPT: an illustration of its capacity, limitations & implications for academic writers. *Pers Med Educ*. 2023;12(1):261–70. <https://doi.org/10.5334/pme.1072>.
30. Team G, Anil R, Borgeaud S, et al. Gemini: a family of highly capable multimodal models; 2023. arXiv preprint arXiv:2312.11805. <https://doi.org/10.48550/arXiv.2312.11805>
31. OpenAI. ChatGPT; 2023. <https://help.openai.com/en/articles/6783457-what-is-chatgpt> Accessed 24 Jan 2023
32. Alkaissi H, McFarlane SI. Artificial hallucinations in ChatGPT: implications in scientific writing.

- Cureus. 2023;15(2):e35179. <https://doi.org/10.7759/cureus.35179>.
33. Jin JQ, Dobry AS. ChatGPT for healthcare providers and patients: practical implications within dermatology. *J Am Acad Dermatol*. 2023;89(4):870–1. <https://doi.org/10.1016/j.jaad.2023.05.081>.
34. Ferrara E. Should ChatGPT be biased? Challenges and risks of bias in large language models; 2023. arXiv preprint [arXiv:2304.03738](https://arxiv.org/abs/2304.03738)
35. Salimi A, Saheb H. Large language models in ophthalmology scientific writing: ethical considerations blurred lines or not at all? *Am J Ophthalmol*. 2023;254:177–81. <https://doi.org/10.1016/j.ajo.2023.06.004>.
36. Liebreinz M, Schleifer R, Buadze A, et al. Generating scholarly content with ChatGPT: ethical challenges for medical publishing. *Lancet Digit Health*. 2023;5(3):e105–6. [https://doi.org/10.1016/S2589-7500\(23\)00019-5](https://doi.org/10.1016/S2589-7500(23)00019-5).
37. Li J. Security implications of AI Chatbots in health care. *J Med Internet Res*. 2023;25:e47551. <https://doi.org/10.2196/47551>.
38. OpenAI. ChatGPT; 2024. <https://openai.com/blog/new-ways-to-manage-your-data-in-chatgpt#OpenAI>. Accessed 9 Jan 2024
39. OpenAI. ChatGPT; 2024. <https://openai.com/policies/privacy-policy>. Accessed 9 Jan 2024.

Analysis of International Publication Trends in Artificial Intelligence in Ophthalmology (2019–2023)

29

Kai Jin , Wenyue Shen, Lu Yuan,
Andrzej Grzybowski , and Juan Ye 

Abstract

Background: Over the past two decades, there has been a rapid expansion in literature concerning artificial intelligence (AI) in ophthalmology, showcasing its pivotal role in advancing this field. This analysis aims to offer a comprehensive longitudinal bibliometric assessment of AI-focused papers in ophthalmology. **Methods:** With timespan from January 1st, 2019 to December 31st, 2023, we retrieved publications related to artificial intelligence in ophthalmology from the Web

of Science Core Collection (WoSCC) database. Our analysis encompassed annual scientific output, leading authors, highly cited documents, productive countries, institutions, and journals. Additionally, we conducted a co-citation analysis of literature, examined the top co-occurring keywords, and performed a cluster analysis based on keyword co-occurrence. Data visualization was conducted using VOSviewer. **Results:** Totally, 3314 papers related to AI in ophthalmology were examined, showcasing a surge in output from 2019 to 2023, followed by a subsequent decline. Prevalent contributors included China and the USA, with prolific author Ting DSW standing out. Noteworthy journals such as “*Scientific Reports*” and “*Ophthalmology*” were prominent. Keywords analysis revealed a burgeoning interest in OCT parameters, AI techniques, and their practical applications in ophthalmology, signaling dynamic research trajectories within the evolutionary progress of this field. **Conclusion:** This analysis represented an in-depth overview of AI in ophthalmology research, aiding scientists in grasping the knowledge of frontier domains and potential implications of AI. Future research is expected to explore the link between ocular and systemic biomarkers, telemedicine, real-world studies, and the advancement and application of novel AI algorithms, like visual converters.

K. Jin · W. Shen · J. Ye (✉)
Eye Center, The Second Affiliated Hospital, School
of Medicine, Zhejiang University, Hangzhou,
Zhejiang, China
e-mail: yejuan@zju.edu.cn

K. Jin
e-mail: jinkai@zju.edu.cn

W. Shen
e-mail: wy_shen@zju.edu.cn

L. Yuan
Department of Ophthalmology, The Children’s
Hospital, National Clinical Research Center
for Child Health, Zhejiang University School
of Medicine, Hangzhou, China
e-mail: yuanlu@zju.edu.cn

A. Grzybowski
Institute for Research in Ophthalmology, Foundation
for Ophthalmology Development, Poznan, Poland
e-mail: ae.grzybowski@gmail.com

Keywords

Artificial intelligence · Ophthalmology ·
Publication trends · Data analysis ·
Machine learning applications

Introduction

Since its inception in the 1950s, AI has significantly contributed to auxiliary disease diagnosis, marking a rising sector [1]. The integration of artificial intelligence (AI) into healthcare holds immense promise for transforming medical diagnosis and therapeutic strategies [2]. Not only does it offer expedited and precise diagnosis, but also it minimizes medical errors and enhances overall efficiency to a certain degree. AI's advanced algorithms and learning capabilities are making it progressively pervasive across all healthcare domains due to its remarkable advantages [3]. The successful applications of AI has already been demonstrated in specific clinical areas such as ophthalmology [4], dermatology, radiology, and oncology [5, 6]. Nowadays, the real potential of AI extends to screening, diagnosis, categorization, and treatment guidance, fostering precision medicine [7].

Recently, numerous researchers in ophthalmology specializing in AI have utilized medical images to develop intricate deep learning models for comprehensive analyses. These models have demonstrated exceptional effectivity and accuracy in automatic screening and diagnosis of prevalent vision-threatening conditions, such as diabetic retinopathy (DR) [8], glaucoma [9], age-related macular degeneration (AMD) [10], cataracts, and various other anterior segment diseases [11]. With the advancement of deep learning technologies, researchers are currently exploring disorder detection beyond the conventional realm of early-stage [12] diagnosis, expanding into prognosis prediction [13]. Image-based medical assistive diagnostic systems significantly foster large-scale population screening, heightening clinical work efficiency, and devising innovative strategies to tackle medical resource scarcity. Furthermore, the

fusion of AI with telemedicine is emerging as another practical solution to address resource shortages [14].

The burgeoning interest in ophthalmic applications of AI and a plethora of related publications, pose challenges for researchers to keep abreast of the latest developments and research priorities in the field. Ongoing studies highlight the rapid development of AI in ophthalmology, which is still in its early stages. This bibliometric analysis extensively evaluates the AI literature in ophthalmology, outlines global research trends, hotspots, and visualizes publication patterns. This study is an invaluable resource for researchers seeking a comprehensive overview of the development of AI in ophthalmology.

Method

Search Strategy

We conducted a comprehensive search of the WoSCC database, a reputable scholarly repository established in 1985, to ensure that high-quality peer-reviewed journals were included in this study (Table 29.1). The citation indexes included the Science Citation Index Expanded (SCI-EXPANDED) and the Social Sciences Citation Index (SSCI). To retrieve relevant data, a set of AI-related keywords (encompassing terms such as ‘artificial intelligence’, ‘machine learning’, and ‘deep learning’) and a range of topics associated with ophthalmology (including ‘eye diseases’, ‘eye care’, ‘retinal diseases’, and ‘corneal diseases’) were applied. The search strategy employed was: TS (Topic)=((‘artificial intelligence’ OR ‘machine learning’ OR ‘deep learning’) AND (‘ophthalmology’ OR ‘eye diseases’ OR ‘retinal diseases’ OR ‘corneal diseases’)). The search timespan ranged from January 1, 2019, to December 31, 2023 (publication date), with the final search conducted on January 13, 2024. The WoSCC yielded a total of 5159 documents, of which 3314 documents underwent further analysis after the screening process illustrated in Fig. 29.1, following the exclusion of 1845 documents.”

Table 29.1 Summary of data source and selection

Research database	Web of Science Core Collection
Citation indexes	SCI-EXPANDED and SSCI
Search period	Jan 1st, 2019–Dec 31st, 2023
Document Types	Original Article
Language	English
Searching keywords	(“ophthalmology”) AND (“artificial intelligence”)
Data extraction	Export with full records and cited references in plain text format
Sample size	3314

SCI-EXPANDED, Science Citation Index Expanded; SSCI, Social Sciences Citation Index
Citation counts reflect all the paper obtained on January 13, 2024

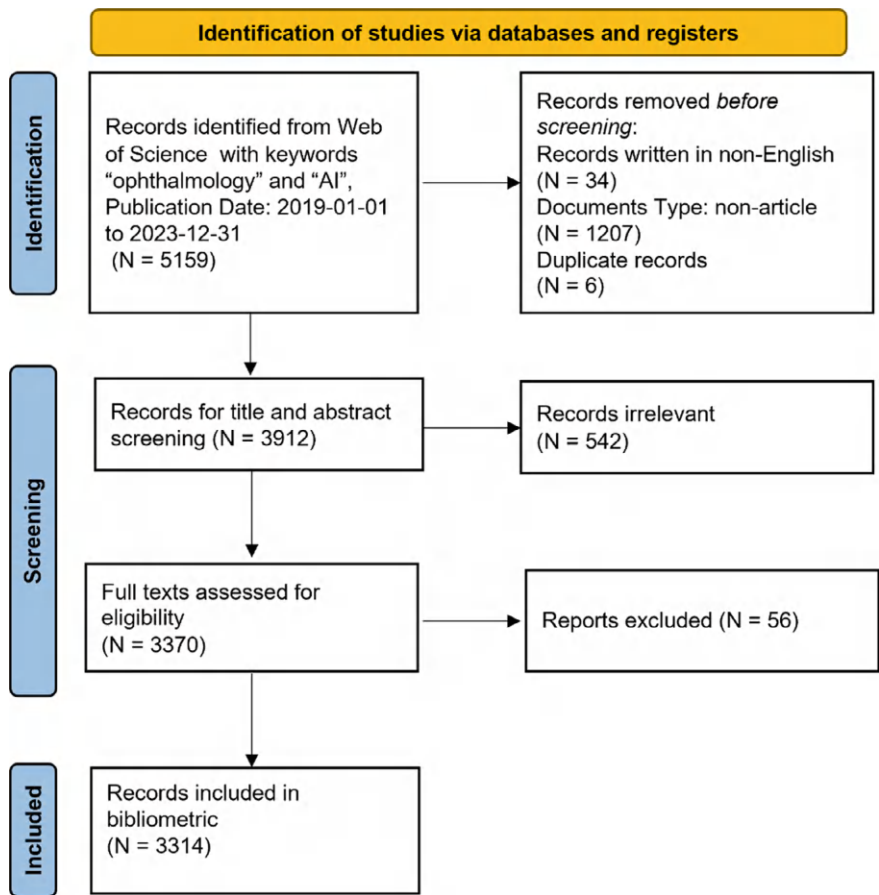


Fig. 29.1 Flowchart of search, screening, and bibliometric analysis

Data Collection

Following retrieval of pertinent documents from the WoSCC database, an analysis was conducted on the yearly scientific output, prominent authors in terms of productivity, highly-cited documents, leading countries in productivity, institutions with notable output, prolific journals, co-citation patterns within the literature, prevalent co-occurring keywords, and clusters formed by co-occurring keywords.

Data Analysis

The results from the WoSCC analysis and the citation reports were imported into Microsoft Excel and depicted graphically. VOSviewer 1.6.19 (<https://www.vosviewer.com/>) was utilized for quantitative research of bibliometrics and scientometrics. The network visualization in VOSviewer comprised items represented by circles with labels, where larger circles denoted higher item frequency, aiding in the identification of distinct clusters. Visual elements encompassed countries with more than 5 published articles, organizations with 22 or more publications, and co-occurring keywords (≥ 45 occurrences). These analyses facilitated the identification of pivotal research themes.

Results

A total of 5159 documents on AI applications in ophthalmology were searched from the WoSCC database. After screening, a total of 3314 documents related to AI in ophthalmology were utilized for further analysis (Fig. 29.1). 3314 documents were from 455 journals, 13,284 authors. Meanwhile, 3314 documents had a total of 62,807 references from 15,247 journals.

Overview of the Published Documents

An analysis was conducted on the multifaceted publication trends involving yearly scientific output, prominent authors, leading countries, highly productive organizations, and journals with significant publication yields.

Among the 3314 documents, the trend of yearly scientific output (Fig. 29.2) displayed a gradual increase from 2019 to a peak in 2022, followed by a decline in 2023. Concurrently, citations on the applications of AI in ophthalmology showed a continuous growth from 2019 to 2023.

A scrutiny of 13,284 authors identified the top 30 most prolific contributors, each with more than 23 documents, including Schmidt-Erfurth Ursula, Liu Jiang, Bogunovic Hrvoje, Zhao Yitian, Chen Xinjian, Lin Haotian, Ting Daniel S. W., Ye Juan, Yang Weihua, Chiang Michael F. and Lee Aaron Y. (Table 29.2). Schmidt-Erfurth Ursula had the highest number of publications with 31, accumulating 1007 citations at an average of 32.5 per publication.

The examining of countries revealed the top ten productive nations emerged, including China, USA, India, UK, South Korea, Australia, Germany, Saudi Arabia, Singapore, and Japan (Fig. 29.3). Network diagrams showcased the publication contributions of 60 countries, with China, the USA, and India as the most productive countries (Fig. 29.4). And China got the most citations per publication (Table 29.3).

The evaluation of 3914 organizations identified the top ten productive institutions, such as “Sun Yat-Sen University”, “Chinese Academy of Sciences”, “Capital Medical University”, “Moorfield Eye Hospital”, “Shanghai Jiao Tong University”, “Stanford University”, “Singapore National Eye Centre”, “Wenzhou Medical University”, “Medical University of Vienna”, and “National University of Singapore”.

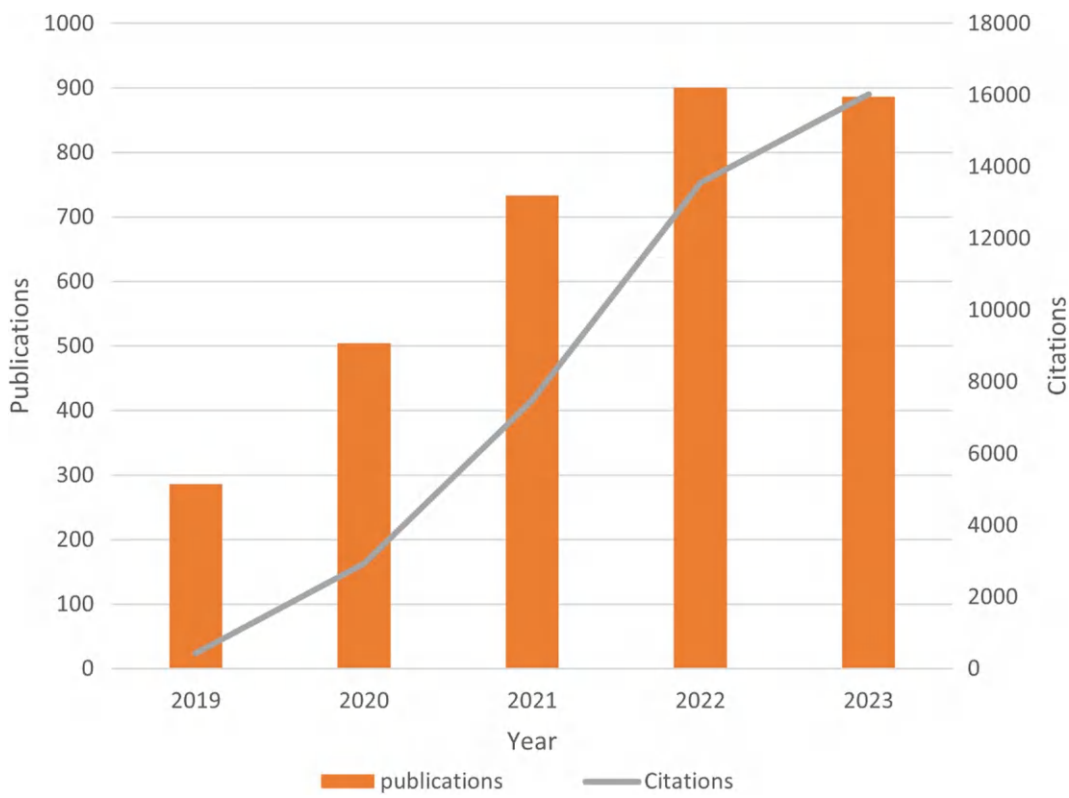


Fig. 29.2 Annual publication numbers and citations on applications of AI in ophthalmology from 2019 to 2023

Network diagrams displayed the publication contributions of 54 organizations (Fig. 29.5). The top ten high-yield publications in terms of volume and citations were analyzed from 455 journals. Noteworthy journals based on publication volume included “Scientific Reports”, “Translational Vision Science & Technology”, “IEEE Access”, “Biomedical Signal Processing and Control”, “Multimedia Tools and Applications”, “Diagnosis”, “Computer in Biology and Medicine”, “Biomedical Optics Express”, “British Journal of Ophthalmology”, “Frontiers in Medicine” (Fig. 29.6A). In terms of citations, significant journals included “Ophthalmology”, “IEEE Transactions on Medical Imaging”, “Investigative Ophthalmology & Visual Science” (Fig. 29.6B).

Top Cited Documents

Among 3314 documents, the top 20 most cited documents are listed in Table 29.4. The most cited document was the “CE-Net: Context Encoder Network for 2D Medical Image Segmentation”, which was published in 2019 by the journal of “IEEE Transactions on Medical Imaging”. The high-citation papers in medical AI span diverse journals, reflecting the field’s broad impact. They’re published across influential platforms like Medical Image Analysis, IEEE Transactions on Medical Imaging, Symmetry-Basel, Ophthalmology, British Journal of Ophthalmology, and New England and Journal of Medicine. These journals cover cutting-edge medical research, digital medicine,

Table 29.2 Top 20 authors on applications of AI in ophthalmology (2019–2023)

Rank	Author	Country	Publications	Citations	Average citations/ publications
1	Schmidt-Erfurth, Ursula	Austria	36	1073	29.8
2	Liu, Jiang	China	30	1514	50.5
3	Bogunovic, Hrvoje	Austria	28	511	18.3
4	Zhao, Yitian	China	28	1328	47.4
5	Lin, Haotian	China	26	378	14.5
6	Chen, Xinjian	China	25	268	10.7
7	Lee, Aaron Y	USA	25	404	16.2
8	Ting, Daniel S. W	Singapore	25	743	29.7
9	Ye, Juan	China	25	314	12.6
10	Chiang, Michael F	USA	24	423	17.6
11	Keane, Pearse A	UK	24	706	29.4
12	Yang, Weihua	China	24	193	8.0
13	Cheng, Ching-Yu	Singapore	23	389	16.9
14	Wong, Tien Yin	Singapore	23	662	28.8
15	Tabuchi, Hitoshi	Japan	22	313	14.2
16	Novo, Jorge	Spain	21	103	4.9
17	Chan, R. V. Paul	USA	20	417	20.9
18	Cheung, Carol Y	China	20	400	20.0
19	Elze, Tobias	USA	20	212	10.6
20	He, Mingguang	China	20	221	11.1

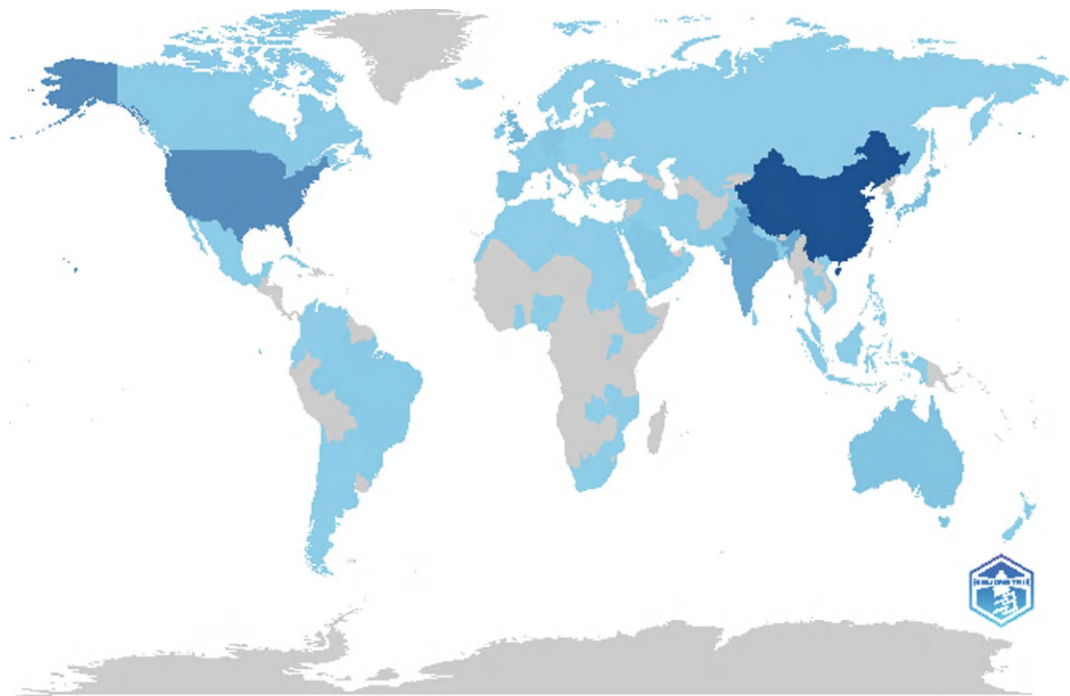


Fig. 29.3 Geographic distribution of publications on applications of AI in ophthalmology (2019–2023). The dark blue areas indicate the most publications and the light blue indicate the fewest publications. Grey areas means that no relevant data has been collected

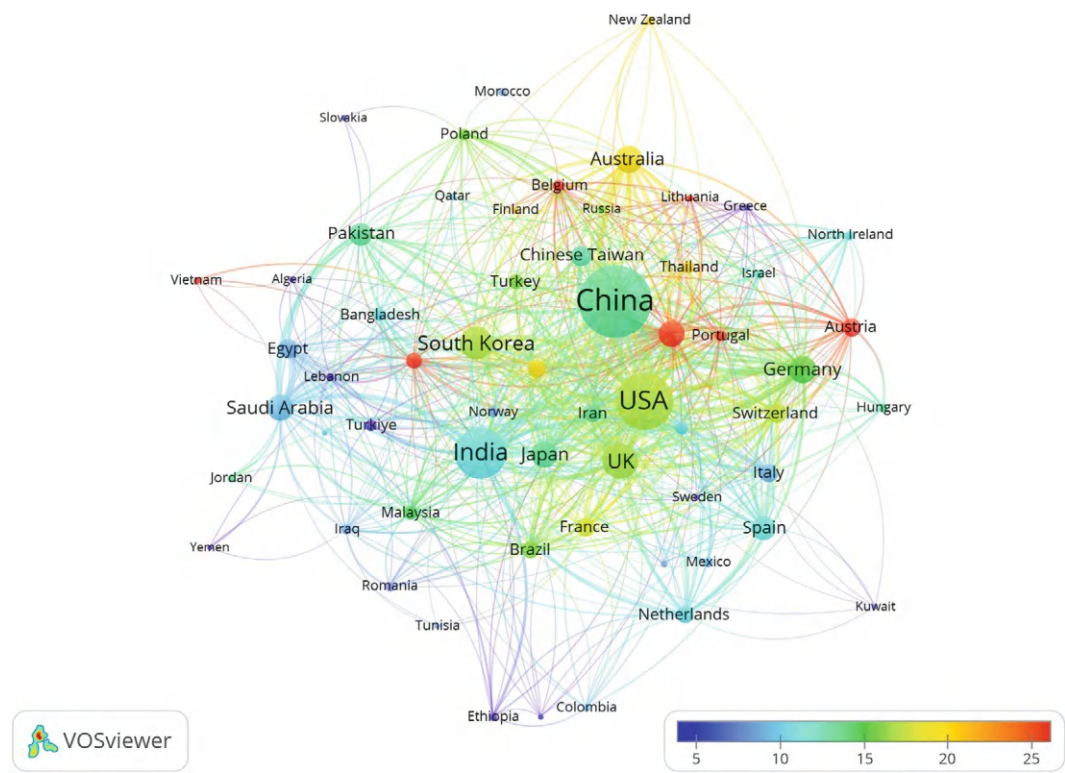


Fig. 29.4 Collaboration maps between highly contributed countries/regions on applications of AI in ophthalmology worldwide (2019–2023). Threshold=5 for the minimum number of publications of a country/region.

Circle size represents the number of publications. The colors indicate citations of the countries/regions

image analysis, and specialized areas like ophthalmology and diabetic retinopathy, showcasing the interdisciplinary nature and wide recognition of AI’s role in advancing healthcare. Additionally, platforms like IEEE Access, npj Digital Medicine, and others contribute to the dissemination of AI-driven innovations across various medical disciplines.

Keywords Analysis

The Fig. 29.7 shows the top 20 keywords with the strongest citation bursts from 2019 to 2023. The vertical axis lists the keywords, and the height of each bar represents the number of citations for that keyword in a particular year. The color gradient indicates the trend in citation volume. Presumably These keywords have

experienced a significant increase in citations in a relatively short period of time. By analyzing this chart, we can gain insights into the research topics that have recently garnered increased attention and those that may be on the rise in popularity within the scholarly community. It can also help to identify new research trends and topics that may shape the field of AI in ophthalmology.

By analyzing this chart, it is possible to gain insight into the research topics that have recently garnered increased attention and those that may be on the rise in popularity within the scholarly community. It can also help to identify new research trends and topics that may be shaping the field. Furthermore, the cluster analysis of the co-occurrence keywords (4 or more) was displayed by a network diagram, divided into three clusters, and displayed in four colors (Fig. 29.8).

Table 29.3 Top 30 countries on applications of AI in ophthalmology (2019–2023)

Rank	Country	Publications	Citations	Citations per publications
1	China	1171	15,649	13.3
2	USA	697	11,880	17.0
3	India	552	6284	11.4
4	UK	262	4349	16.6
5	South Korea	226	3813	16.9
6	Australia	149	2907	19.5
7	Germany	144	2224	15.4
8	Saudi Arabia	141	1334	9.5
9	Singapore	134	3557	26.5
10	Japan	132	1783	13.5
11	Spain	107	1263	11.8
12	Pakistan	98	1315	13.4
13	Chinese Taiwan	78	1011	13.0
14	Egypt	72	641	8.9
15	Switzerland	69	1191	17.3
16	Austria	68	1979	29.1
17	Canada	68	1332	19.6
18	Italy	62	562	9.1
19	France	60	1067	17.8
20	Iran	58	788	13.6
21	Netherlands	57	615	10.8
22	Brazil	55	875	15.9
23	U Arab Emirates	49	2101	42.9
24	Turkey	44	700	15.9
25	Malaysia	35	510	14.6
26	Bangladesh	32	348	10.9
27	Thailand	31	626	20.2
28	Denmark	30	322	10.7
29	Poland	27	424	15.7
30	Portugal	27	762	28.2

The red cluster: “artificial intelligence”, “deep learning”, “machine learning”, “accuracy”, “prediction”, “prevalence”.

The green cluster: “diabetic retinopathy”, “optic disc”, “deep neural network”, “retinal image”.

The blue cluster: “segmentation”, “feature extraction”, “task analysis”, “training”.

The yellow cluster: “oct”, “octa”, “quantification”, “macular edema”.

The red cluster comprised keywords emphasizing the contents of OCT-related parameters, such as “thickness” and “macular degeneration”. The green cluster comprised

keywords underlining the contents of the AI, such as “deep learning” and “convolutional neural network”. The blue cluster appeared in the fields in application of AI in ophthalmology, including keywords such as “classification” and “validation”.

Discussion

Contemporary academic research involves bibliometric analysis, an interdisciplinary scientific approach that employs mathematical and statistical methods to quantitatively assess various

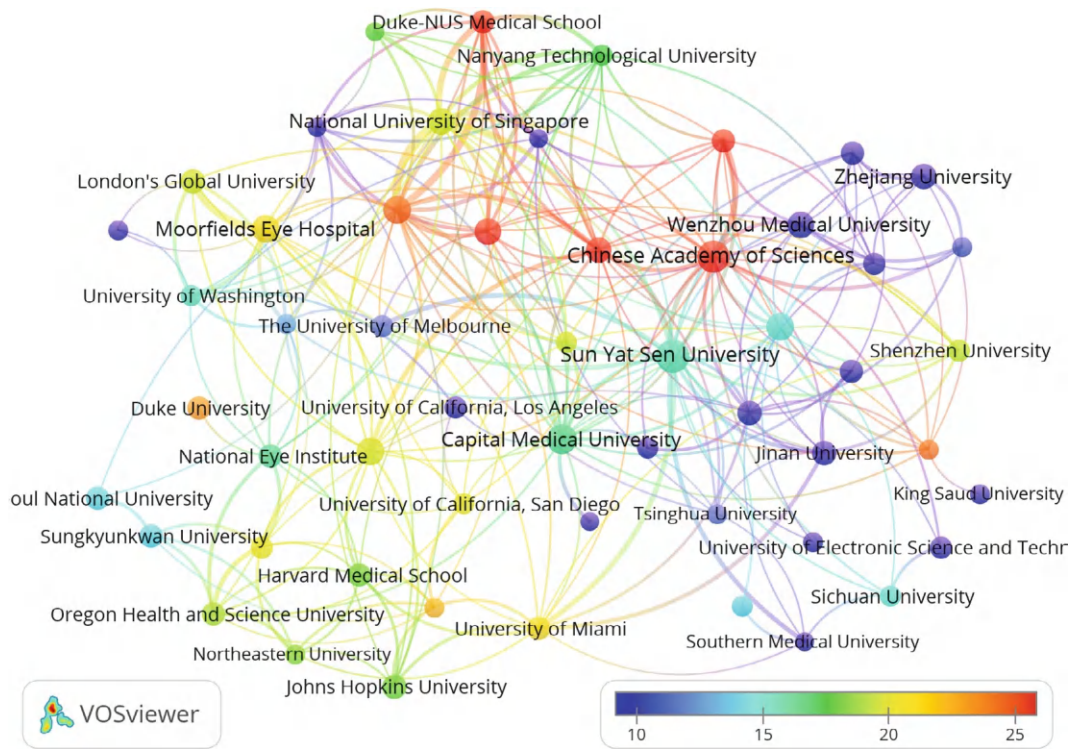


Fig. 29.5 Collaboration maps between highly contributed organizations on applications of AI in ophthalmology worldwide (2019–2023). Threshold = 22 for the

minimum number of publications of an organization. 1 of the organizations are not connected to each other, so they are omitted in the network

knowledge sources. This comprehensive knowledge system integrates mathematics, statistics, and linguistics while emphasizing quantification [15]. Presently, bibliometric analysis stands as a pivotal method for exploring research trends and cutting-edge developments. The visualization of relationships between documents/countries, regions/journals/authors can be achieved through a scientific knowledge map after bibliometric analysis [16]. As far as our understanding goes, comprehending the steps involved in bibliometric analysis and adhering to general guidelines is crucial to support scientists in their writing endeavors. Advance in biomedicine have facilitated the visualization of scientific knowledge using VOSviewer software which serves as a versatile tool for scientific analysis and mapping. In this study, we utilized VOSviewer to analyze literature related to AI in ophthalmology obtained from the WoSCC database. This

entailed evaluating publication trends, delineating primary document contributions, and outlining key thematic areas within the field.

Publication Trends of AI in Ophthalmology Researches

The rising significance and advancements of AI technology has spurred considerable interest in its application across diverse sectors, particularly in ophthalmology. This heightened demand for AI-driven solutions in eye care coincides with a substantial increase in research publications and citations, indicating the growing importance and influence of this field. This trend highlights the increasing recognition of AI's capacity to revolutionize ophthalmology and improve patient outcomes. It also signifies a burgeoning community of

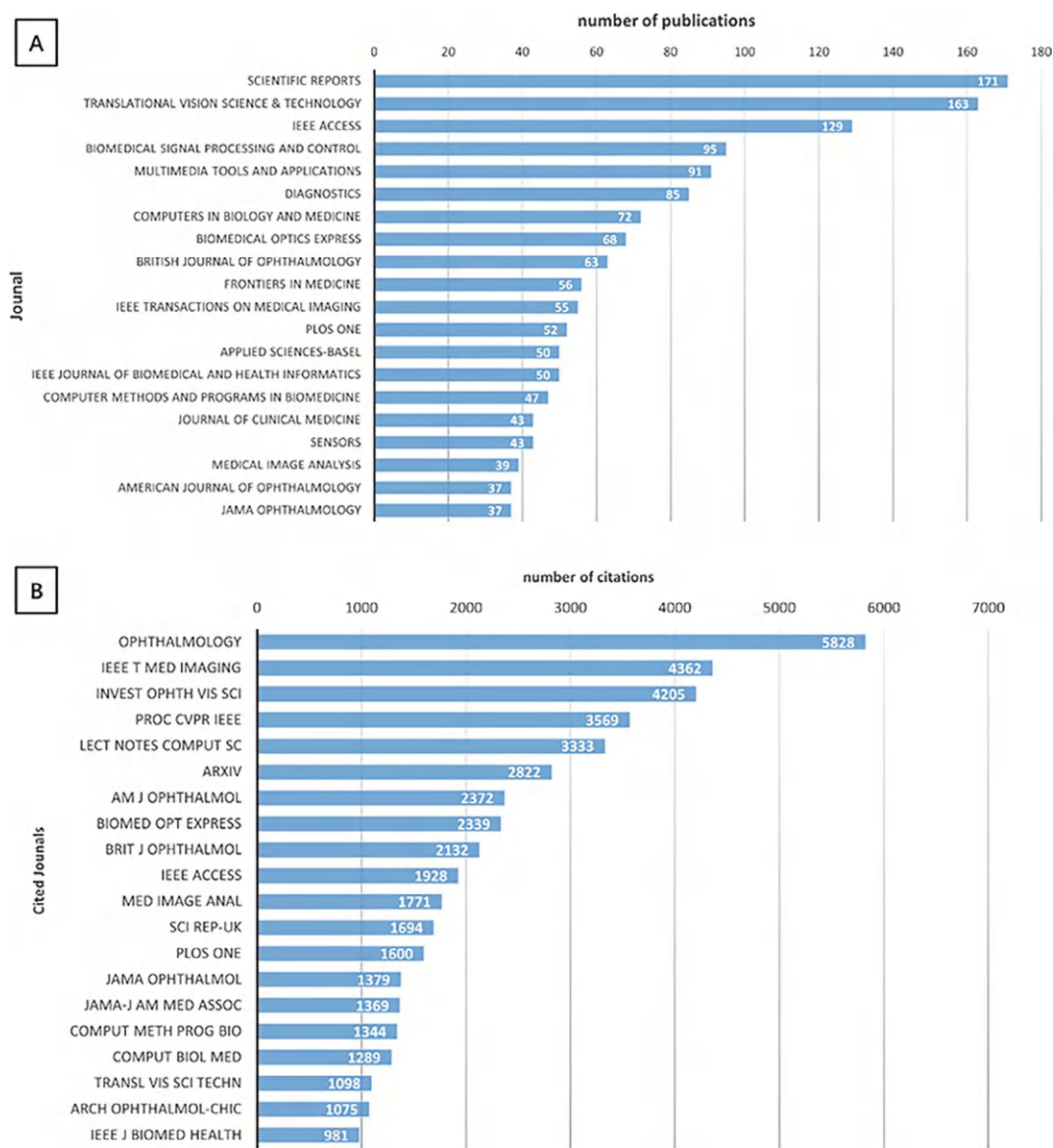


Fig. 29.6 (A) Most relevance journals (B) Most cited journals on applications of AI in ophthalmology (2019–2013)

researchers dedicated to exploring and advancing AI-based techniques and technologies to address challenges in this specialized field [17]. Consequently, keeping abreast of the latest AI developments and their practical applications in ophthalmology is imperative for both ophthalmologists and researchers to remain relevant and competitive in this rapidly evolving landscape [7].

Research Frontiers

AI-driven research in ophthalmology has garnered global interest. It was identified four distinct periods in healthcare development driven by corresponding technological advancements and applications (1990–1996, 1997–2003, 2004–2015, and 2016–2021), attributing these shifts to innovations such as the popularization

Table 29.4 Top citation publications of 3314 on applications of AI in ophthalmology (2019–2013)

Rank	Authors	Title	Journal	Publication year	Total citations	Average per year
1	Gu Zaiwang et al.	CE-Net: Context Encoder Network for 2D Medical Image Segmentation	<i>IEEE Transactions on Medical Imaging</i>	2019	990	165.0
2	Schlegl Thomas et al.	f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks	<i>Medical Image Analysis</i>	2019	507	84.5
3	Orlando Jose Ignacio et al.	REFUGE Challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs	<i>Medical Image Analysis</i>	2020	262	52.4
4	Li Xiaomeng et al.	Transformation-Consistent Self-Ensembling Model for Semisupervised Medical Image Segmentation	<i>IEEE Transactions on Neural Networks and Learning Systems</i>	2021	207	51.8
5	Qummar Shrish et al.	A Deep Learning Ensemble Approach for Diabetic Retinopathy Detection	<i>IEEE Access</i>	2019	197	32.8
6	Mateen Muhammad et al.	Fundus Image Classification Using VGG-19 Architecture with PCA and SVD	<i>Symmetry-Basel</i>	2019	191	31.8
7	Sayres Rory et al.	Using a Deep Learning Algorithm and Integrated Gradients Explanation to Assist Grading for Diabetic Retinopathy	<i>Ophthalmology</i>	2019	180	30.0
8	Li Tao et al.	Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening	<i>Information Sciences</i>	2019	166	27.7
9	Peng Yifan et al.	DeepSeeNet: A Deep Learning Model for Automated Classification of Patient-based Age-related Macular Degeneration Severity from Color Fundus Photographs	<i>Ophthalmology</i>	2019	163	27.2
10	Gadekallu Thippa Reddy et al.	Early Detection of Diabetic Retinopathy Using PCA-Firefly Based Deep Learning Model	<i>Electronics</i>	2020	160	32.0
11	Milea D. et al.	Artificial Intelligence to Detect Papilledema from Ocular Fundus Photographs	<i>New England and Journal of Medicine</i>	2020	148	29.6
12	Gadekallu Thippa Reddy et al.	Deep neural networks to predict diabetic retinopathy	<i>Journal of Ambient Intelligence and Humanized Computing</i>	2020	147	29.4

(continued)

Table 29.4 (continued)

Rank	Authors	Title	Journal	Publication year	Total citations	Average per year
13	Shanthi, T. et al.	Modified Alexnet architecture for classification of diabetic retinopathy images	<i>Computers & Electrical Engineering</i>	2019	141	23.5
14	Kwon Bum Chul et al.	RetainVis: Visual Analytics with Interpretable and Interactive Recurrent Neural Networks on Electronic Medical Records	<i>IEEE Transactions on Visualization and Computer Graphics</i>	2019	141	23.5
15	Liu Hanruo et al.	Development and Validation of a Deep Learning System to Detect Glaucomatous Optic Neuropathy Using Fundus Photographs	<i>JAMA Ophthalmology</i>	2019	140	23.3
16	Wang Shujun et al.	Patch-Based Output Space Adversarial Learning for Joint Optic Disc and Cup Segmentation	<i>IEEE Transactions on Medical Imaging</i>	2019	139	23.2
17	Bellemo Valentina et al.	Artificial intelligence using deep learning to screen for referable and vision-threatening diabetic retinopathy in Africa: a clinical validation study	<i>Lancet Digital Health</i>	2019	135	22.5
18	Yim Jason et al.	Predicting conversion to wet age-related macular degeneration using deep learning	<i>Nature Medicine</i>	2020	134	26.8
19	Asaoka Ryo et al.	Using Deep Learning and Transfer Learning to Accurately Diagnose Early-Onset Glaucoma From Macular Optical Coherence Tomography Images	<i>American Journal of Ophthalmology</i>	2019	127	21.2
20	Arcadu Filippo et al.	Deep learning algorithm predicts diabetic retinopathy progression in individual patients	<i>npj Digital Medicine</i>	2019	126	21.0

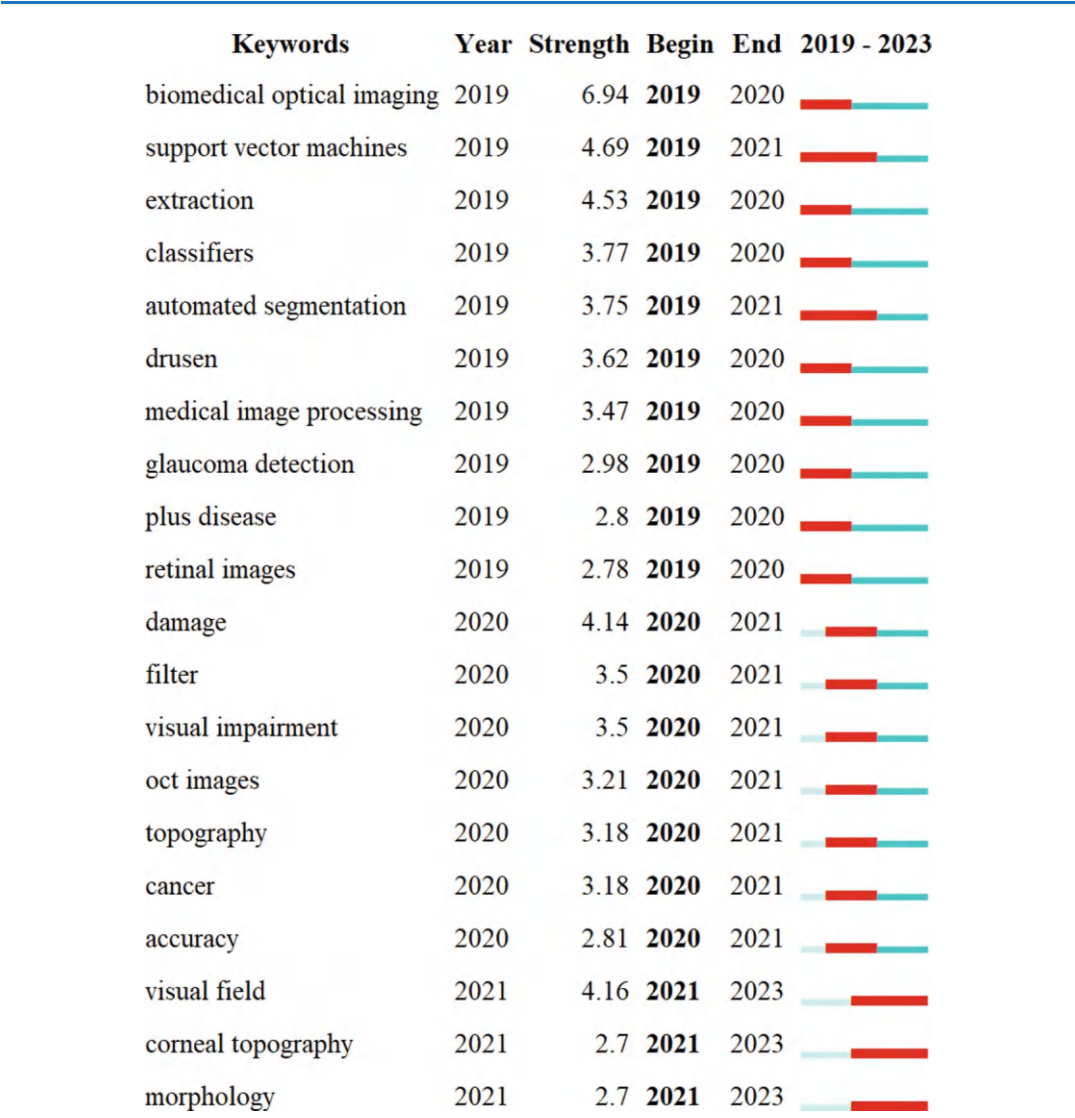


Fig. 29.7 Keywords with the strongest citation bursts on applications of AI in ophthalmology (2019–2023)

of the Internet, smartphone usage, the development of cloud computing, and advancements in algorithmic techniques, particularly deep learning [18]. In a former study, the USA, China, and the UK emerged as the three most productive countries, with researchers, particularly in the USA, collaborating with peers from neighboring countries [19]. Our findings spotlight the USA, China, and India as the leading nations in terms of citations and H-index scores within this domain. Undoubtedly, the USA spearheads this

research, consistently maintaining an indispensable role. Notably, affluent nations have been the prime impetus behind AI’s penetration into ophthalmology [20]. The USA’s developed status is in line with its standardized electronic systems and advanced infrastructure. Meanwhile, China’s AI research has seen a meteoric rise in the last four years, boasting the highest number of publications and ranking second in terms of citations and H-index.

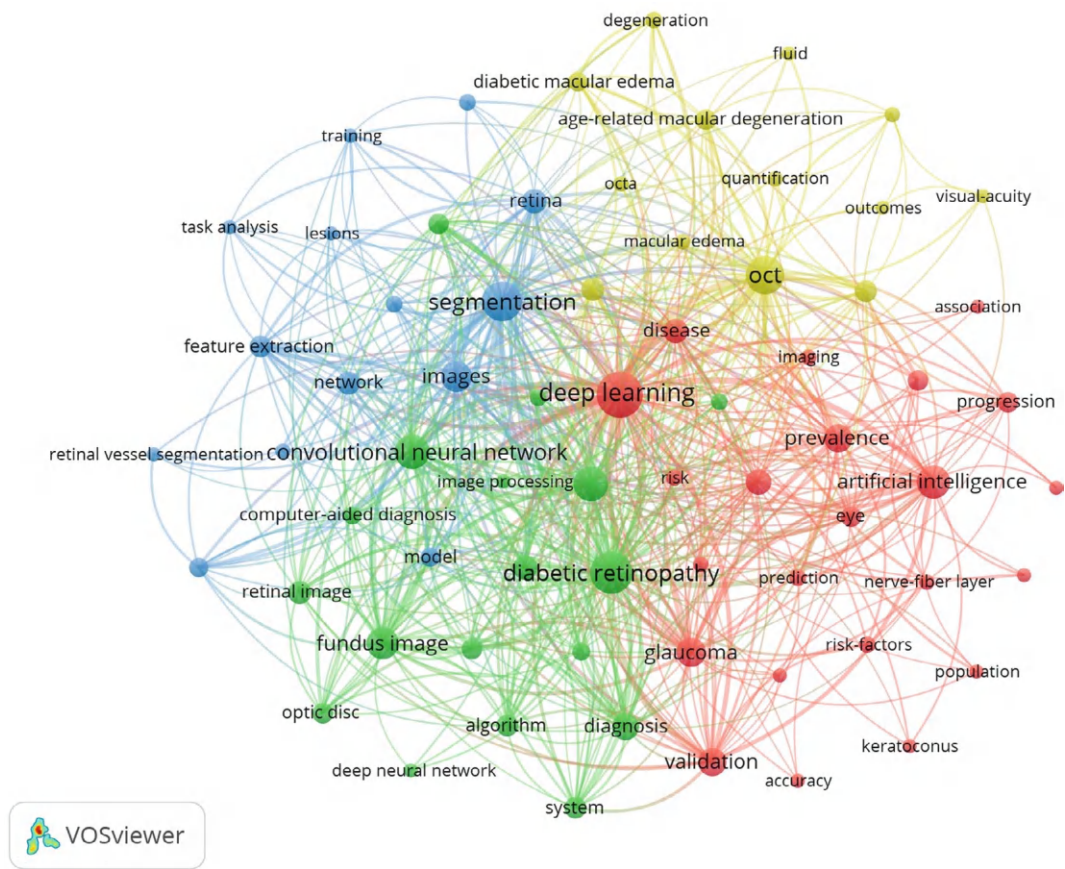


Fig. 29.8 The network of co-occurrence analysis of keywords on applications of AI in ophthalmology (2019–2023). Threshold = 45 for the minimum number of currencies of a keyword

Despite China’s prolific output, its citation count remains modest, signaling a need for enhanced the quality of research. This discrepancy might stem from delayed inception and uneven scientific resources. Moreover, incomplete electronic systems impede data colation difficult. However, populations are noteworthy in China and India allow for extensive data-driven AI studies as well as fueling their research growth.

Focuses of Research Activities

Prominent scholarly works received consistent citations. Table 29.4 delineates the top 20 AI-related papers in ophthalmology. Dominant

keywords in clustered research reveal deep learning, glaucoma, and retinal images as focal points in AI exploration. AI research predominantly targets prevalent ocular disorders that threaten vision. Retinal disease stands as a globally recognized primary cause of severe visual impairment and blindness. Projections suggest a staggering 288 million cases of AMD by 2040, with 600 million cases of diabetes, one third of which may develop DR, and an estimated increase of 111.8 million cases of glaucoma, imposing a substantial global disease burden [21–23]. Hence, scientists ardently embrace state-of-the-art technologies to enhance early disease identification and management of disease. The entry of AI ingress into ophthalmic therapy spans automatic screening,

diagnosis, classification, and prediction of disease progression.

Recent AI surge addresses anterior segment diseases and imaging. A burgeoning field involves infectious keratitis diagnosis and categorization [24]. Lin et al. present a robust AI platform for excellent cataract diagnosis [25]. Myopia and refractive surgery also witness AI applications in identifying vision-threatening conditions [26]. AI predictions for laser refractive surgery outcomes is underway, yet accurately predicting high myopia remains limited. Improvements in AI algorithms promise a burgeoning field.

Furthermore, digital advancements extend AI fundus image recognition beyond disease diagnosis and screening. The integration of systemic disorders with fundus images is emerging as a promising AI research focus. In chronic vascular disorders, there is a robust link between systemic diseases and fundus images. Sabanayagam et al. employed AI to detect chronic kidney disease from retinal images, enhancing community and primary healthcare diagnosis [27]. The work of Rim et al. predicted systemic biomarkers via DL on retinal images, opening new AI vistas in ophthalmology [28]. AI analyzes retinal images for coronary artery calcium and cardiovascular risk factors [29, 30], marking a pivotal juncture in systemic disease screening. AI is filling gaps in ophthalmologists' assessment of systemic disease.

AI-driven ophthalmic research matures, yet the digital era births emerging fields. Real-world-based studies captivate scholars and become a new research interest. Wu et al. pioneer cataract AI and multilevel referral model integration, exhibiting robust diagnosis [31]. The AI real expert (CARE) system trained on real-world data performs well in diverse retinal abnormalities [32]. Generative Adversary Networks (GANs) excel in image synthesis and data scarcity challenges, propelling ophthalmic research [33].

Conclusion

Our aim was to offer a comprehensive overview of AI research in ophthalmology. This analysis provided an extensive summary of ophthalmology-related AI studies spanning the last two decades. Through exhaustive searches and screenings, employing three search criteria over a 23-year period from 2000 to 2022, we believe our identification of ophthalmic AI research is comprehensive.

The analysis revealed distinct trends in AI-related ophthalmic research: (I) Dominance in ophthalmic AI research stemmed from high-income nations and those with substantial population bases. (II) Current focal points encompass chronic ocular ailments, notably retinopathy and glaucoma. (III) Future AI research trajectories in ophthalmology include exploring the correlation between ocular and systemic biomarkers, telemedicine, real-world investigations, and novel AI algorithms such as vision transformer models.

The integration of AI into ophthalmology is progressing rapidly, with prospective applications undergoing trials across various medical domains. Nonetheless, successful clinical implementation of this technology remains limited. Future AI research should prioritize bridging the gap between research findings and clinical practice.

References

1. Miller RA. Medical diagnostic decision support systems—past, present, and future: a threaded bibliography and brief commentary. *J Am Med Inform Assoc (JAMIA)*. 1994;1:8–27.
2. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nat Biomed Eng*. 2018;2:719–31.
3. Ting DSW, Pasquale LR, Peng L, Campbell JP, Lee AY, Raman R, et al. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol*. 2019;103:167–75.
4. Schmidt-Erfurth U, Sadeghipour A, Gerendas BS, Waldstein SM, Bogunović H. Artificial intelligence in retina. *Prog Retin Eye Res*. 2018;67:1–29.

5. Kann BH, Hosny A, Aerts H. Artificial intelligence for clinical oncology. *Cancer Cell*. 2021;39:916–27.
6. Hosny A, Parmar C, Quackenbush J. Artificial intelligence in radiology. *Nat Rev Cancer*. 2018;18:500–10.
7. Ting DSW, Peng L, Varadarajan AV, Keane PA, Burlina PM, Chiang MF, et al. Deep learning in ophthalmology: the technical and clinical considerations. *Prog Retin Eye Res*. 2019;72: 100759.
8. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316:2402–10.
9. Li Z, He Y, Keel S, Meng W, Chang RT, He M. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. *Ophthalmology*. 2018;125:1199–206.
10. Grassmann F, Mengelkamp J, Brandl C, Harsch S, Zimmermann ME, Linkohr B, et al. A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography. *Ophthalmology*. 2018;125:1410–20.
11. Li W, Yang Y, Zhang K. Dense anatomical annotation of slit-lamp images improves the performance of deep learning for the diagnosis of ophthalmic disorders. *Nat Biomed Eng*. 2020;4:767–77.
12. Bogunovic H, Montuoro A, Baratsits M, Karantonis MG, Waldstein SM, Schlanitz F, et al. Machine learning of the progression of intermediate age-related macular degeneration based on OCT imaging. *Investig Ophthalmol Vis Sci*. 2017;58:Bio141–Bio150.
13. Yim J, Chopra R. Predicting conversion to wet age-related macular degeneration using deep learning. *Nat Med*. 2020;26:892–9.
14. Rathi S, Tsui E, Mehta N, Zahid S, Schuman JS. The current state of teleophthalmology in the United States. *Ophthalmology*. 2017;124:1729–34.
15. Tian H, Chen J. A bibliometric analysis on global eHealth. *Digit Health*. 2022;8:20552076221091350.
16. Li B, Tan W, Wang Z, Zhou H, Zou J, Li Y, et al. Progress and prospects of gene therapy in ophthalmology from 2000 to 2022: a bibliometric analysis. *Heliyon*. 2023;9: e18228.
17. Ting DSW, Lin H, Ruamviboonsuk P, Wong TY, Sim DA. Artificial intelligence, the internet of things, and virtual clinics: ophthalmology at the digital translation forefront. *Lancet Digit Health*. 2020;2:e8–9.
18. Penteado BE, Fornazin M, Castro L. The evolution of artificial intelligence in medical informatics: a bibliometric analysis. Cham: Springer International Publishing; 2021. p. 121–33.
19. Boudry C, Al Hajj H, Arnould L, Mouriaux F. Analysis of international publication trends in artificial intelligence in ophthalmology. *Graefes Arch Clin Exp Ophthalmol*. 2022;260(5):1779–88.
20. De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med*. 2018;24(9):1342–50.
21. Tham YC, Li X, Wong TY, Quigley HA, Aung T, Cheng CY. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology*. 2014;121:2081–90.
22. Mitchell P, Liew G, Gopinath B, Wong TY. Age-related macular degeneration. *Lancet* (London, England). 2018;392:1147–59.
23. Raman R, Vasconcelos JC, Rajalakshmi R, Prevost AT, Ramasamy K, Mohan V, et al. Prevalence of diabetic retinopathy in India stratified by known and undiagnosed diabetes, urban-rural locations, and socioeconomic indices: results from the SMART India population-based cross-sectional screening study. *Lancet Glob Health*. 2022;10:e1764–73.
24. Wei Z, Wang S, Wang Z, Zhang Y, Chen K, Gong L, et al. Development and multi-center validation of machine learning model for early detection of fungal keratitis. *EBioMedicine*. 2023;88: 104438.
25. Lin H, Li R, Liu Z, Chen J, Yang Y, Chen H, et al. Diagnostic efficacy and therapeutic decision-making capacity of an artificial intelligence platform for childhood cataracts in eye clinics: a multicentre randomized controlled trial. *EClinicalMedicine*. 2019;9:52–9.
26. Tan TE, Anees A, Chen C, Li S, Xu X, Li Z, et al. Retinal photograph-based deep learning algorithms for myopia and a blockchain platform to facilitate artificial intelligence medical research: a retrospective multicohort study. *Lancet Digit Health*. 2021;3:e317–29.
27. Sabanayagam C, Xu D, Ting DSW, Nusinovici S, Banu R, Hamzah H, et al. A deep learning algorithm to detect chronic kidney disease from retinal photographs in community-based populations. *Lancet Digit Health*. 2020;2:e295–302.
28. Rim TH, Lee G, Kim Y, Tham YC, Lee CJ, Baik SJ, et al. Prediction of systemic biomarkers from retinal photographs: development and validation of deep-learning algorithms. *Lancet Digit Health*. 2020;2:e526–36.
29. Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng*. 2018;2:158–64.
30. Rim TH, Lee CJ, Tham YC, Cheung N, Yu M, Lee G, et al. Deep-learning-based cardiovascular risk stratification using coronary artery calcium scores predicted from retinal photographs. *Lancet Digit Health*. 2021;3:e306–16.
31. Wu X, Huang Y, Liu Z, Lai W, Long E, Zhang K, et al. Universal artificial intelligence platform

- for collaborative management of cataracts. *Br J Ophthalmol*. 2019;103(11):1553–60.
32. Lin D, Xiong J, Liu C, Zhao L, Li Z, Yu S, et al. Application of Comprehensive Artificial intelligence Retinal Expert (CARE) system: a national real-world evidence study. *Lancet Digit Health*. 2021;3:e486–95.
33. Lazaridis G, Lorenzi M, Ourselin S, Garway-Heath D. Improving statistical power of glaucoma clinical trials using an ensemble of cyclical generative adversarial networks. *Med Image Anal*. 2021;68: 101906.

Machine Learning Studies in Ocular Oncology

30

David E. Pelayes and Arun D. Singh

Abstract

This chapter explores the integration of artificial intelligence (AI) into the diagnosis and management of intraocular and conjunctival tumors, with a focus on uveal melanoma (UM) and small choroidal pigmented tumors (SCPT). It examines AI's ability to enhance early detection, risk assessment, and prognostication by leveraging advanced machine learning models and ophthalmic imaging data. The potential of AI for improving clinical decision-making and overcoming limitations in traditional diagnostic methods is highlighted, along with challenges and opportunities for future development.

Keywords

Artificial Intelligence · Small pigmented choroidal tumors · Small choroidal melanomas · Digital images · Learning machine

Introduction

In clinical oncology, the treatment of malignant tumors requires histologic confirmation to indicate treatment, but in intraocular tumors, such as uveal melanoma (UM), the treatment decision is based on clinical examination and complementary tests, specially ultrasound (US), which due to its low mean echogenicity allows, together with the clinical features, the confirmation of UM in the vast majority of cases [1, 2].

Unfortunately, this is not proven for small choroidal pigmented tumors (SCPT), which are define as those with a $a < 3$ mm thickness and a basal diameter < 10 mm (Fig. 30.1) in which the sensitivity and specificity of US is limited by the thickness of the tumors [3].

SPCT are treated only if their appearance or growth suggests malignancy. Several studies have identified risk factors for growth [4–9] and metastasis [8, 9]. These include tumor thickness greater than 2 mm, the presence of subretinal fluid, clinical symptoms, presence of orange pigment, and tumor margin within 3 mm of the optic disc margin, giving rise to the term: TFSOM (“To Find Small Ocular Melanoma”) [7–9] (Fig. 30.2).

The current treatment for PCPT (which includes atypical nevi and small melanomas), is periodic observation until growth [8, 9, 11], specially those without risk factors and those with foveolar location (since their treatment seriously

D. E. Pelayes (✉)
Buenos Aires University, Buenos Aires, Argentina
e-mail: davidpelayes@gmail.com

FASTA University, Mar del Plata, Argentina

Maimonides University, Buenos Aires, Argentina

A. D. Singh
Cleveland Clinic, Cole Eye Institute,
Cleveland, OH, USA



Fig. 30.1 SPCT with risk factors

compromises vision). But this conduct in cases of melanomas has the potential risk to generate metastasis, for that reason, observation is controversial [12, 13]. Moreover, this behavior is a contradiction in clinical oncology's practice, which considers earlier diagnosis and treatment as a mandatory first step to improve patient survival [14, 15].

The recent development of Artificial Intelligence (AI) in medicine and especially in the field of ophthalmology started with focus on: Ocular features of Diabetes, Glaucoma, Anterior Segment and Macular Degeneration among others has also a great potential in the field of ocular oncology for choroid melanocytic tumors, also being developed for tumors in other locations such as ocular surface tumors.

In a literature review, various articles are showing this aspect of AI. In this case, we will refer to the main ones published, the current

state of AI in ocular technology and the future potential of this tool.

Uveal Melanoma

Zabor et al. present a work to discuss the evolution of noninvasive diagnostic methods in the identification of choroidal nevus and determination of risk factors for malignant transformation as well as introducing the novel role that AI can play in the diagnostic process. They conclude that, AI applied to existing ophthalmic imaging systems could be used for both identification of choroidal nevus and as a tool to aid in earlier detection for transformation in malignant melanoma [16].

Singh AD et al. publish a study to develop a validated machine learning (ML) model to diagnose small choroidal melanoma. The training

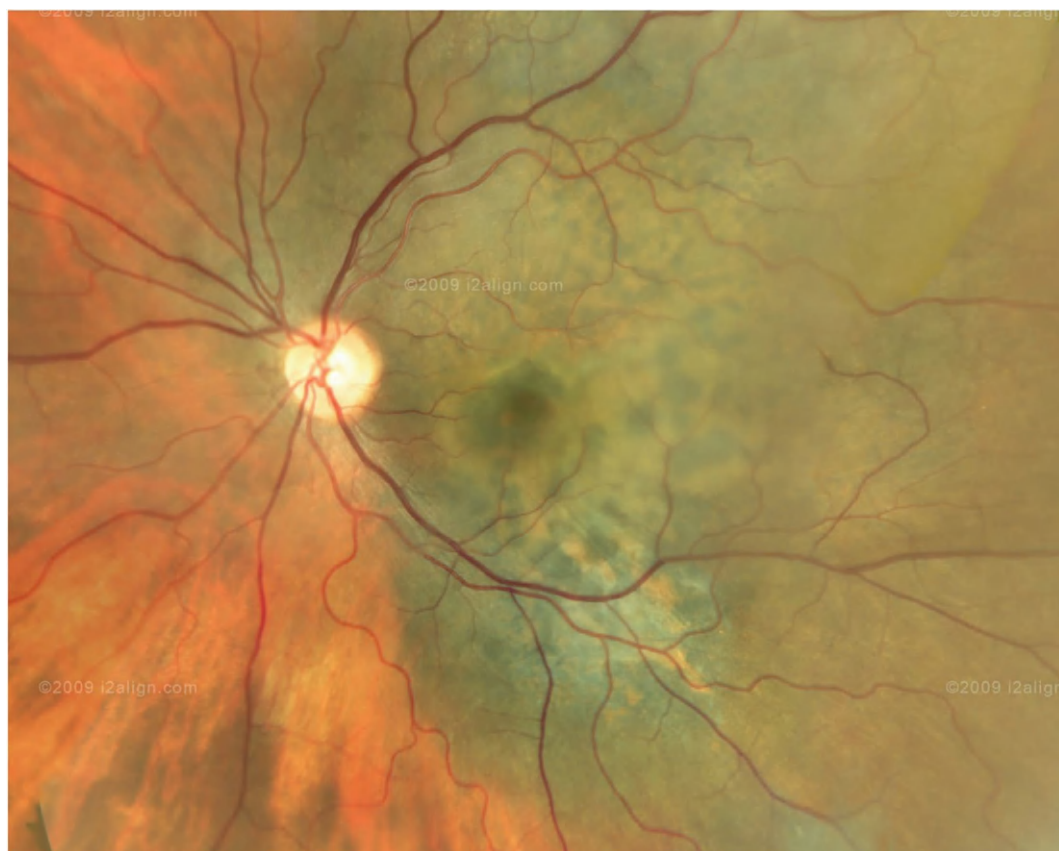


Fig. 30.2 SPCT after 13 months [10]

dataset included 123 patients diagnosed as small choroidal melanocytic tumor (5.0–16.0 mm in the largest basal diameter and 1.0 mm–2.5 mm in height; *Collaborative Ocular Melanoma Study criteria*). Those diagnosed as melanoma ($n=61$) had either documented growth or pathologic confirmation. Sixty-two patients with stable lesions classified as choroidal nevus were used as negative control group. The external validation dataset included 240 patients with small choroidal melanocytic tumor managed in a third party clinic, observed for malignant growth. The lasso logistic regression method was used to select variables to include in the final model and for the association of melanoma versus choroidal nevus.

Internal and external validation was performed to assess the model performance. The findings shows that a distance to optic

disc ≥ 3 mm and presence of drusen were associated with decreased odds of melanoma, whereas male versus female sex, increased height, subretinal fluid, and presence of orange pigment were associated with increased odds of choroidal melanoma. The area under the receiver operating characteristic “discrimination value” for this model was 0.880. The top four variables for inclusion in the model for internal validation (implying their importance as predictors of melanoma), were subretinal fluid, height, distance to the optic disc, and presence of orange pigment. When tested against the validation dataset, the prediction model could distinguish between choroidal nevus and melanoma with a high discrimination of 0.861. The final prediction model was converted into an online calculator to generate predicted probability of melanoma. They concluded that to minimize a

diagnostic uncertainty, a machine learning-based diagnostic prediction calculator can be readily applied for decision-making and counseling patients with small choroidal melanoma [17].

Lou et al. published another study about a an accurate machine-learning model to predict the 4-year risk of metastasis and death in UM, based on ocular ultrasound data. A total of 454 patients with UM were enrolled in this retrospective, single-center study. All patients were followed up for at least 4 years after the application of plaque brachytherapy. B-scan US was used to measure the basal diameters and thickness of tumors preoperatively and postoperatively. Random Forest (RF) algorithm was used to construct two prediction models: whether a patient will survive for more than 4 years and whether the tumor will develop metastasis within 4 years after treatment. The results of the predictive model achieved an area under the receiver operating characteristic curve (AUC) of 0.708 for predicting death using only a one-time follow-up record. Including the data from two additional follow-ups increased the AUC of the model to 0.883. We attained AUCs of 0.730 and 0.846 with data from one and three-time follow-up, respectively, for predicting metastasis. The model found that the amount of follow-up postoperative data significantly improved the prediction accuracy of death and metastasis. Furthermore, tumor treatment response was divided into four different patterns. The D(decrease)/S(stable) patterns were associated with a significantly better prognosis than I(increase)/O(other) patterns. The authors conclude that the study developed an RF model to predict the risk of metastasis and death from UM within 4 years based on US follow-up records following plaque brachytherapy, but with limitations that will require, validate this model with prospective datasets [18].

In other study by Chen et al., a ML model was used for outcome prediction of uveal melanoma patients in a chinese population. Here they analyzed the clinical characteristics and survival rate of 1553 patients diagnosed with UM over a period of 15 years. They also included factors

that were not included in previous study, such as the lesion largest basal diameter, thickness, tumor size, pigmentation, intraocular hemorrhage, ciliary body involvement, extraocular extension and TNM stage. They also constructed a comprehensive prognostic model to predict the risk of metastasis within 2 years and death after 2 years following treatment. They concluded that the largest basal diameter, size, stage, age, intraocular pressure and gender were associated with high risk of metastasis.

Donizy et al. publish a study with a ML model to identify a cost-effective prognostic tool using routine histologic and clinical variables for risk stratification. The authors identified very important prognostic parameters in a cohort of 164 enucleated primary uveal melanomas eyes from 164 patients without prior therapies. They validated the prognostic prediction of the top important parameters identified in the discovery cohort using 80 uveal melanomas from the Tumor Cancer Genome Atlas database with available gene expression prognostic signature (GEPS). The performance of three different survival analysis models (Cox proportional hazards (CPH), random survival forest (RSF), and survival gradient boosting (SGB)) was compared against GEPS using receiver operating curves (ROC). In all three selection methods, BAP1 status, nucleoli size, age, mitotic rate per 1mm^2 , and ciliary body infiltration were identified as significant overall survival (OS) predictors; and BAP1 status, nucleoli size, largest basal tumor diameter, tumor-infiltrating lymphocyte density, and tumor-associated macrophage density were identified as significant progression-free survival (PFS) predictors. ROC plots for the median survival time point showed that significant parameters in SGB studied model can predict OS better than GEPS. For PFS, SGB model performed similarly to GEPS. The time-dependent area under the curve (AUC) showed SGB model performing better than GEPS in predicting OS and metastatic risk.

This study also show that routine histologic and clinical variables are adequate for patient risk stratification in comparison with not readily accessible GEPS [19].

Conjunctival Tumors

Up to date, the studies in this specific type of tumors are more limited. One of the main reasons is that conjunctival melanomas are statistically less frequent generating less number of cases and images. Few-Shot Learning (FSL) is a system to obtain images in these cases. Based on prior knowledge of the lesion, it incorporates data and obtains results with a smaller number of images. Using the FSL, Yoo et al. [20] conducted a study composed of anonymized images of four types; conjunctival melanoma (136), nevus or melanosis (93), pterygium (75), and normal conjunctiva (94). Before training, two generative adversarial networks (GANs) were constructed to augment the dataset for low-shot learning. The collected data was randomly divided into training (70%), validation (10%), and test (20%) datasets. Moreover, 3D melanoma phantoms set using a smartphone were designed to build an external validation. The GoogleNet, InceptionV3, NASNet, ResNet50, and MobileNetV2 architectures were trained through transfer learning and validated using the test and the external validation datasets.

The DL model demonstrated a significant improvement in the classification accuracy of conjunctival lesions using synthetic images generated by the GAN models. MobileNetV2 with GAN-based augmentation displayed the highest accuracy (87.5%) in the four-class classification and 97.2% in the binary classification for the detection of conjunctival melanoma. It also showed an accuracy of 94.0% using 3D melanoma phantom images captured with the smartphone camera. In conclusion, this study described a low-shot deep learning model that can detect conjunctival melanomas using ocular surface images with a high accuracy.

Retinoblastoma

Retinoblastoma (RB) is the most common ocular malignant tumor worldwide [21]. It originates from immature retinal cells in children. In RB, magnetic resonance imaging (MRI) is routinely

used to confirm the diagnosis and determine the extent of the disease [22]. In current practice, images of patients with RB are qualitatively evaluated to support the diagnosis and obtain information on tumor extent and metastatic risk factors [23–25]. Related work in other tumors has shown that the quantitative evaluation of radiological features (i.e., radiomics) can provide additional information about tumor characteristics and harbor predictive and prognostic information [26]. For example, MRI-based radiomics models have been proposed for head and neck area [27, 28] and UM uveal melanoma [29].

Strijbis et al. conducted a retrospective study to evaluate the performance of multi-view convolutional neural networks (MV-CNNs) for automated eye and tumor segmentation on MRI in retinoblastoma patients. 40 retinoblastoma patients and 20 healthy control-eyes from 30 patients were included in a train/test (N=29 retinoblastoma-, 17 healthy-eyes) and independent validation (N=11 retinoblastoma-, 3 healthy-eyes) set. Imaging was done using 3.0 T Fast Imaging Employing Steady-state Acquisition (FIESTA), T2-weighted and contrast-enhanced T1-weighted sequences. Sclera, vitreous, lens, retinal detachment and tumor were manually delineated on FIESTA images to serve as a reference standard. Volumetric and spatial performance were assessed by calculating intra-class correlation (ICC) and dice similarity coefficient (DSC). Additionally, the effects of multi-scale, sequences and data augmentation were explored. Optimal performance was obtained by using a three-level pyramid MV-CNN with FIESTA, T2 and T1c sequences and data augmentation. Eye and tumor volumetric ICC was 0.997 and 0.996, respectively. Median [Interquartile range] DSC for eye, sclera, vitreous, lens, retinal detachment and tumor were 0.965 [0.950–0.975], 0.847 [0.782–0.893], 0.975 [0.930–0.986], 0.909 [0.847–0.951], 0.828 [0.458–0.962] and 0.914 [0.852–0.958], respectively. In conclusion, MV-CNN can be used to obtain accurate ocular structure and tumor segmentations in retinoblastoma.

Other retrospective observational study by Kaliki et al. [30], was designed to explore the

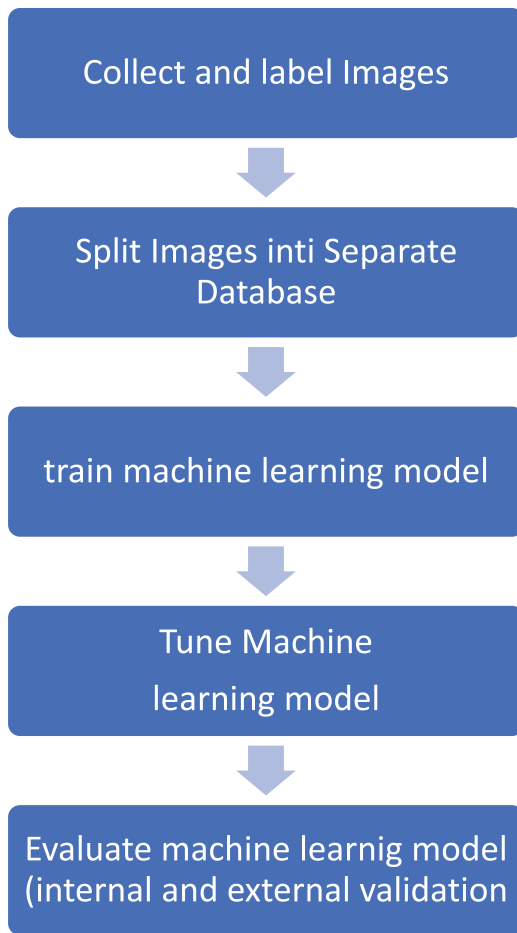
utility of AI and ML in the diagnosis and grouping of intraocular retinoblastomas (iRB). In 771 fundus images of 109 eyes, 181 images had no tumor and 590 images displayed iRB based on review by two independent ocular oncologists (with an interobserver variability of <1%). The sensitivity, specificity, positive predictive value, and negative predictive value of the trained AI model were 85%, 99%, 99.6%, and 67%, respectively. Of 109 eyes, the sensitivity, specificity, positive predictive value, and negative predictive value for detection of RB by AI model were 96%, 94%, 97%, and 91%, respectively. Of these, the eyes were normal (n=31) or belonged to group A (n=1), B (n=22), C (n=8), D (n=23), and E (n=24) RB based on review by two independent ocular oncologists (with an interobserver variability of 0%). The sensitivity, specificity, positive predictive value, and negative predictive value of the trained AI model were 100, 100, 100, and 100% for group A; 82%, 20, 21, 98%, 90%, and 96% for group B; 63, 99, 83, and 97% for group C; 78, 98, 90, and 94% for group D, and 92, 91, 73, and 98% for group E, respectively. The authors conclude that the AI model for iRB is highly sensitive in the detection of RB with high specificity for the classification of iRB.

Zhang et al. [31] published a study to apply a DL algorithm to reduce the burden of follow-up and offspring screening. The study includes RB from Beijing Tongren Hospital. Clinical-suspected and treated retinoblastoma patients were prospectively collected for validation. Images from the posterior pole and peripheral retina were collected, and reference standards were made according to the consensus of the multidisciplinary management team. A DL algorithm was trained to identify “normal fundus”, “stable retinoblastoma” in which specific treatment is not required, and “active retinoblastoma” in which specific treatment is required. The performance of each classifier included

sensitivity, specificity, accuracy, and cost-utility. A total of 36,623 images were included for developing the Deep Learning Assistant for Retinoblastoma Monitoring (DLA-RB) algorithm. In internal fivefold cross-validation, DLA-RB the model achieved an area under curve (AUC) of 0.998 (95% confidence interval [CI] 0.986–1.000) in distinguishing normal fundus and active retinoblastoma, and 0.940 (95% CI 0.851–0.996) in distinguishing stable and active retinoblastoma. They identify active retinoblastoma tumors from all clinical-suspected patients and active retinoblastoma from all treated retinoblastoma patients, the AUC of DLA-RB reached 0.991 (95% CI 0.970–1.000), and 0.962 (95% CI 0.915–1.000), respectively. In conclusion, DLA-RB achieved high accuracy and sensitivity in identifying active retinoblastoma from the normal and stable retinoblastoma fundus. This algorithm can be used to surveil the activity of retinoblastoma during follow-up and screen high-risk offspring. Compared with referral procedures to ophthalmologic centers, DLA-RB-based screening and surveillance is cost-effective and can be incorporated within telemedicine programs.

Conclusions

In general, a pool of data used for model training from individual sources and institutions are desirable and would increase the sensitivity and specificity of the AI based systems, but also, the generalizability of the model. With greater quantity, quality, and diversity of the collected data, potential discrepancies in the characteristics of the dataset would be avoided and could prevent the model overfitting on one specific characteristic of the training data. The figure shows an enhanced workflow of how to develop, train, validate, test and deploy a potential future AI model.



The figure shows an enhanced workflow of how to develop, train, validate, test and deploy a potential future AI model.

The ophthalmic oncology has AI as a tool with incredible potential for the diagnosis, follow up, indications of treatment and evolutions.

References

1. Shing AD, Elizalde J, Gonzalez-Candial M, Ramos-Esteban JC, Pelayes DE, Ortiz MA, et al. *Oncología Clínica Oftálmica*. Amsterdam, Netherlands: Elsevier; 2008. p. 280–4.
2. Chang M, Zimmerman LE, McLean IW. The persisting pseudomelanoma problem. *Arch Ophthalmol*. 1984;102:726–7.
3. Augsburger JJ, Corrêa ZM, Schneider S, Yassin RS, Robinson-Smith T, Ehya H, et al. Diagnostic transvitreal fine-needle aspiration biopsy of small melanocytic choroidal tumors in nevus versus melanoma category. *Trans Am Ophthalmol Soc*. 2002;100:225–32.
4. Butler P, Char DH, Zarbin M, Kroll S. Natural history of indeterminate pigmented choroidal tumors. *Ophthalmology*. 1994;101:710–7.
5. Singh AD, Mokashi AA, Bena JF, Jacques R, Rundle PA, Rennie IG. Small choroidal melanocytic lesions: features predictive of growth. *Ophthalmology*. 2006;113:1032–9.
6. Shields CL, Shields JA, Kiratli H, DePotter P, Cater JR. Risk factors for growth and metastasis of small choroidal melanocytic lesions. *Ophthalmology*. 1995;102:1351–61.
7. Shields CL, Shields JA. Clinical features of small choroidal melanoma. *Curr Opin Ophthalmol*. 2002;13:135–41.
8. Augsburger JJ, Schroeder RP, Territo C, Gamel JW, Shields JA. Clinical parameters predictive of enlargement of melanocytic choroidal lesions. *Br J Ophthalmol*. 1989;73:911–7.
9. Finger P, Kurli M, Shulman J. Treatment of small choroidal melanoma: an internet survey. *Int J Ophthalmol Vis Sci*. 2006;5:1–5.
10. Pelayes DE, Folgar AM, Chiaradia P, Zarate JO. Diagnostic algorithm in small pigmented choroid tumors (less than 3 mm thick). *Lat Am J Ophthalmol*. 2021;4:5.
11. Augsburger JJ, Correa ZM, Trichopoulos N. An alternative hypothesis for observed mortality rates due to metastasis after treatment of choroidal melanomas of different sizes. *Trans Am Ophthalmol Soc*. 2007;105:54–60.
12. Augsburger JJ. Is observation really appropriate for small choroidal melanomas. *Trans Am Ophthalmol Soc*. 1993;91:147–75.
13. Robertson DM. Changing concepts in the management of choroidal melanoma. *Am J Ophthalmol*. 2003;136:161–70.
14. Mims JL, Shields JA. Follow-up studies of suspicious choroidal nevi. *Ophthalmology*. 1978;85:929–43.
15. Shields CL, Materin MA, Teixeira L, Mashayekhi A, Ganguly A, Shields JA. Small choroidal melanoma with chromosome 3 monosomy on fine-needle aspiration biopsy. *Ophthalmology*. 2007;114:1919–24.
16. Zabor EC, Raval V, Luo S, Pelayes DE, Singh AD. A Prediction Model to Discriminate Small Choroidal Melanoma from Choroidal Nevus. *Ocul Oncol Pathol*. 2022 Feb;8(1):71–78. doi: 10.1159/000521541. Epub 2021 Dec 22. PMID: 35356604; PMCID: PMC8914269.
17. Zabor EC, Raval V, Luo S, Singh AD, Pelayes D. A prediction model to discriminate small & choroidal melanoma from choroidal nevus. *Ocul Oncol Pathol*. 2022;8:71–8.
18. Luo J, Chen Y, Yang Y, et al. Prognosis prediction of uveal melanoma after plaque brachytherapy based on ultrasound with machine learning. *Front Med (Lausanne)*. 2021;8: 777142.
19. Donizy P, Krzyzinski M, Markiewicz A, et al. Machine learning models demonstrate that

- clinicopathologic variables are comparable to gene expression prognostic signature in predicting survival in uveal melanoma. *Eur J Cancer*. 2022;174:251–60.
20. Yoo TK, Choi JY, Kim HK, et al. Adopting low-shot deep learning for the detection of conjunctival melanoma using ocular surface images. *Comput Methods Programs Biomed*. 2021;205: 106086.
 21. Kivelä T. The epidemiological challenge of the most frequent eye cancer: retinoblastoma, an issue of birth and death. *Br J Ophthalmol*. 2009;93(9):1129–31.
 22. de Graaf P, et al. Guidelines for imaging retinoblastoma: imaging principles and MRI standardization. *Pediatr Radiol*. 2012;42:2–14.
 23. De Jong MC, et al. Diagnostic accuracy of intraocular tumor size measured with MR imaging in the prediction of postlaminar optic nerve invasion and massive choroidal invasion of retinoblastoma. *Radiology*. 2016;279:817–26.
 24. Jansen RW, et al. MR imaging features of retinoblastoma: association with gene expression profiles. *Radiology*. 2018;288:506–15.
 25. Sirin S, et al. High-resolution MRI using orbit surface coils for the evaluation of metastatic risk factors in 143 children with retinoblastoma. *Neuroradiology*. 2015;57:805–14.
 26. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. 2016;278(2):563–77.
 27. Mes SW, et al. Outcome prediction of head and neck squamous cell carcinoma by MRI radiomic signatures. *Eur Radiol*. 2020;30:6311–21.
 28. Martens RM, et al. Predictive value of quantitative diffusion-weighted imaging and 18-F-FDG-PET in head and neck squamous cell carcinoma treated by (chemo) radiotherapy. *Eur J Radiol*. 2019;113:39–50.
 29. Su Y, et al. Value of MR-based radiomics in differentiating uveal melanoma from other intraocular masses in adults. *Eur J Radiol*. 2020;131: 109268.
 30. Kaliki S, Vempuluru VS, Ghose N, Patil G, Viriyala R, Dhara KK. Artificial intelligence and machine learning in ocular oncology: retinoblastoma. *Indian J Ophthalmol*. 2023;71(2):424–30.
 31. Zhang R, Dong L, Li R, Zhang K, Li Y, Zhao H, Shi J, Ge X, Xu X, Jiang L, Shi X, Zhang C, Zhou W, Xu L, Wu H, Li H, Yu C, Li J, Ma J, Wei W. Automatic retinoblastoma screening and surveillance using deep learning. *Br J Cancer*. 2023;129(3):466–74.

Application of Artificial Intelligence in Oculoplastics

31

Yilu Cai, Xuan Zhang, Jing Cao, Andrzej Grzybowski, Juan Ye, and Lixia Lou

Abstract

This chapter explores the application of artificial intelligence (AI) in oculoplastics, highlighting its integration into diagnosis, treatment planning, and prognostication for conditions affecting the eyelids, orbit, and lacrimal system. It outlines advancements in machine learning and deep learning technologies for analyzing medical images, measuring clinical parameters, and predicting outcomes with high accuracy. Emphasis is placed on AI's potential to improve efficiency, enhance diagnostic precision, and support surgical planning, offering transformative possibilities in oculoplastic practice.

Keywords

Artificial intelligence · Oculoplastics · Eyelid diseases · Orbital diseases · Lacrimal diseases

Introduction

Artificial intelligence (AI) refers to computer systems that can deal with tasks that typically require human intelligence. In the contemporary world, it is one of the most widely used computer technologies, involving, among other areas, computer science, mathematics, cognitive psychology, linguistics, philosophy, and neuroscience. AI has evolved at an incredible velocity since the first proposal of developing this technology in 1956 [1]. A significant subset of AI, named machine learning (ML), enables programs to learn and recognize data patterns and to make predictions or decisions according to data sets without the program being explicitly written [2]. Deep learning has been put forward based on ML, which analyzes data through multilayer neural networks and thus can learn more complex data sets and give more accurate conclusions [3].

AI technology progressed remarkably in the last two decades, which has been used in various medical applications, including ophthalmology [4–8]. Oculoplastics is a subspecialty of ophthalmology. It addresses abnormalities in the appearance of the eye, encompassing a range of conditions like eyelid diseases, orbital diseases, and lacrimal diseases. These diseases have a common denominator, that is the influence on the appearance of the periorbital area. Abnormal alterations in periocular appearance is

Y. Cai · X. Zhang · J. Cao · J. Ye · L. Lou (✉)
Eye Center, The Second Affiliated Hospital, School
of Medicine, Zhejiang University, Hangzhou, China
e-mail: louluxia110@zju.edu.cn

A. Grzybowski
Institute for Research in Ophthalmology, Foundation
for Ophthalmology Development, Poznan, Poland

detrimental to one's facial appearance [9], which have unneglectable effect on the patient's social lives, thus the significance of prompt diagnosis and treatment must not be disregarded.

The features of oculoplastic diseases are mostly significantly visual, and the diagnosis is usually based on image data. Thus, many research groups have explored the applications of AI in oculoplastics. Various works can already be done by AI, such as segmenting the relevant part [10, 11] and measuring clinical parameters [12, 13] from extraocular photographs, computed tomography (CT) images, and magnetic resonance images (MRI). Moreover, AI technology has been used to screen [14], diagnosis [15], and classify oculoplastic diseases [16, 17], while it also assists in treatment procedures such as surgical strategy planning in blepharoptosis [18]. Other significant applications include prediction [19] and assessment of postoperative outcomes [20]. The unique advantages of AI mainly lie in the accuracy and comprehensiveness of information processing, free from space restrictions [21], and labor cost saving [22].

In recent years, there has been a major surge in research focused on the application of artificial intelligence in oculoplastics. Here we summarize recent studies that concentrate on the applications of utilizing AI for diagnosis, treatment, and prognosis of oculoplastic diseases. We probe into their advantages, disadvantages and possibilities, aiming to present a comprehensive panorama of prospects for this research area.

Methodology

The scientific literatures were searched on Sep 30, 2023, using the PubMed and Web of Science. Relevant contributions were included. Search terms were used in specific combinations: "artificial intelligence", "deep learning", "machine learning", "neural network", "oculoplastics", "eyelid diseases", "orbital diseases", and "lacrimal diseases". We extracted relevant papers from the references of identified articles to expand the scope of our search. All reviewed documents

were accessible in full text and published in English. Irrelevant and duplicate articles were eliminated by examining the title and abstract. We excluded surveys, case reports, comments, guidelines, patent accounts, and editorials. Y.C. and X.Z. conducted paper screening, data extraction, and assessment. Extracted information included the author, publication year, research target, algorithm, training and targeting dataset, accuracy, and the area under the model's curve.

Application of AI Technology in Eyelid Diseases

Maintaining a natural contour of the eyelids is vital for facial aesthetics. Therefore, eyelid surgery is essential not just for the restoration of eyelid function, but also for addressing cosmetic issues. External photography of the eye can easily reveal major symptoms like inflammation, tumors, and dysfunction of the eyelid [23]. It enables the utilization of deep learning for analyzing eye/facial photographs, aiding in refining the appearance of eyelids. Advancements in facial recognition technology have set the groundwork for AI function in eyelid diseases [24]. Recent studies involving artificial intelligence applications in eyelid diseases are showed in Table 31.1.

AI-Assisted Segmentation and Measurement

Diagnosing eyelid diseases typically involves measuring periocular parameters, a task generally taken by specialized ophthalmologists. Considering the advantages of high efficiency and potential improvement in reliability and validity, Moriyama et al. first used AI technology for eyelid graphical information acquisition in 2006, who designed a system to output eyelid and iris structural parameters and track eye motion [25]. A ML algorithm proposed by Şimşek and Şirolu using photographs on which the face margin has been marked to output

Table 31.1 Recent studies involving artificial intelligence applications in eyelid diseases

Authors and citations	Research target	Dataset	AI algorithm	Accuracy	Area under the curve
Bahçeci Şimşek and Şirolu [26]	Palpebral distance, eye-opening area, and average eyebrow height acquisition	55 Full-face photographs	DLIB-ML algorithm	–	Palpebral distance, eye-opening area, and average eyebrow height acquisition
Van Brummen et al. [27]	Periocular information acquisition	418 photographs of periorbital areas	ResNet-50	–	Periocular information acquisition
Chen et al. [28]	MRD and levator muscle function measuring	822 eyes of 411 volunteers	–	–	MRD and levator muscle function measuring
Schulz et al. [21]	Capturing periocular morphology from videos	Consumer-grade video of 77 patients with FNP, 33 with ptosis, 33 with TED, and 65 controls	A customized program named ‘Video Analysis of the Eyelids.’	–	–
Lou et al. [56]	Measuring MRD automatically	132 eyes of 66 volunteers	–	–	–
Cao et al. [10]	Eyelid morphologic parameters measuring	2756 eyes of 1378 regular Asian participants	U-Net	–	–
Tabuchi et al. [14]	Screening ptosis via tablet device	1276 eyelid images taken by tablet device	MobileNetV2 CNNi	0.828	0.900
Hung et al. [15]	Identifying blepharoptosis from extraocular photographs accurately	500 unprocessed photos of blepharoptosis patients and controls	22 CNNs (best performance: DenseNet121 without pre-training)	DenseNet121:0.886	DenseNet121:0.95
Hung et al. [22]	Identifying blepharoptosis from extraocular photographs accurately	782 external ocular images	VGG-16 model	0.900	0.987

(continued)

Table 31.1 (continued)

Authors and citations	Research target	Dataset	AI algorithm	Accuracy	Area under the curve
Li et al. [17]	Distinguishing benign and malignant eyelid tumors	1417 digital camera images	DenseNet121(best)	0.818 (DenseNet121, external test set) 0.822 (DenseNet121, internal test set)	0.899 (DenseNet121, external test set) 0.955 (DenseNet121, internal test set)
Song et al. [18]	Decision-making between two ptosis surgery strategies	External ocular images of 152 eyes	Gradient-boosted decision tree	0.826	0.795
Qu et al. [20]	Providing 2D and 3D features of ptotic eyes	Extraocular photographs of 64 patients	Multichannel CNN	0.988	–
Lou et al. [12]	Standardized postoperative assessment	Extraocular photographs of 64 patients	U-Net (Attention R2U-Net)	–	–
Tan et al. [29]	Assessment of the surgical complexity of periorcular BCC patients	156 periocular BCC	Decision tree model	–	–
Sun et al. [19]	Predicting postoperative appearance before ptosis surgery	970 pairs of preoperative and postoperative ptosis images	Pix2Pix conditional GAN	–	–
Huang et al. [30]	Diagnose blepharoptosis and predict postoperative appearance	24,500 facial images	–	0.993	–

BBC, basal cell carcinoma; CNN, convolutional neural network; DenseNet, densely connected convolutional networks; FNP, facial nerve palsy; GAN, generative adversarial networks; MRD, marginal reflex distance; ResNet, residual network; R2U-Net, Recurrent Residual CNN-based U-Net; TED, thyroid eye disease

palpebral distance, eye-opening area, and average eyebrow height for subsequent comprehensive analysis of the parameters [26]. An open-source, fully automated dynamic-static periocular parameter extraction algorithm was proposed recently as an alternative to manual measurement, the output of which contains parameters of eyebrow and iris morphology [27]. A U-Net model was used to automatically and objectively segment eyelid morphologic features and to measure morphologic parameters [10].

The widespread use of smartphones has created extra possibilities for AI applications in eyelid diseases. In a recent study, monocular multi-directional photographs captured with smartphones served as a dataset to construct an AI model for measuring marginal reflex distance (MRD) and evaluating levator muscle function [28]. The correlation coefficient between the model output and the actual measurements was approximately 0.9, indicating a strong association. Consumer-grade video, unlike static photographs, benefited from capturing the patient's natural periocular morphology, circumventing the interference of unconscious reflexes during the photography process. A software developed by Schulz et al. was capable to extract clinically meaningful eyelid measurements, such as MRD, blink lagophthalmos, and ocular surface area exposure, from consumer-grade videos of individuals with oculofacial disorders, including blepharoptosis, facial nerve palsy, and thyroid eye disease [21].

AI-Assisted Screening and Diagnosis

With decision-making capacity, AI could aid in the initial screening of eyelid diseases. The processing capabilities of graphic data have bolstered due to the advancement of deep learning, allowing the utilization of images from mobile devices with cameras as valuable data sources. Tabuchi et al. developed an iOS application using deep learning algorithms for the preliminary screening of blepharoptosis [14]. A significant concern of this application is that when

subjects photograph their face with their mobile phone, the camera angle may not accurately capture the open eyelid state, leading to reduced accuracy in diagnosing blepharoptosis.

In pursuit of performance in diagnosing blepharoptosis, Hung's team developed and trained 22 convolutional neural network (CNN) models to analyze unprocessed photo data. The best-performing DenseNet121 model achieved a diagnostic sensitivity of 90.1% and a diagnostic specificity of 82.4% [15]. The team developed a high-performing model for general practitioners to identify referable blepharoptosis using the VGG-16 CNN. The model demonstrated a sensitivity of 92%, specificity of 88%, and accuracy of 90%, outperforming non-ophthalmic plastic surgeons in testing [22].

Deep learning algorithms have also been utilized to distinguish between benign and malignant eyelid tumors using digital camera imaging [17]. Typically this task was carried out by experienced ophthalmologists. In this study, DenseNet121 once more emerged as the top-performing one among various algorithms, achieving a sensitivity of 96.1% and an area under the curve of 0.955.

AI-Assisted Treatment

AI technology has been used in surgical planning based on annotated data because of its powerful graphical analysis capabilities (Fig. 31.1). The decision-making model assists in selecting surgical plans for upper eyelid blepharoptosis surgery (either levator muscle resection or frontalis suspension) [18]. In this study, AI was merely used to analyze crucial eyelid parameters and offer recommendations, with manual annotations of these parameters. Notably, this study introduced measurements of periocular three-dimensional (3D) modeling for analysis, in addition to the measurements from two-dimensional (2D) images, providing a new application of deep learning. To enhance the outcome of eyelid surgery, a multichannel CNN was utilized to generate a 3D eye model, so that surgeons are accessible to specific features and can formulate different surgical plans [20].

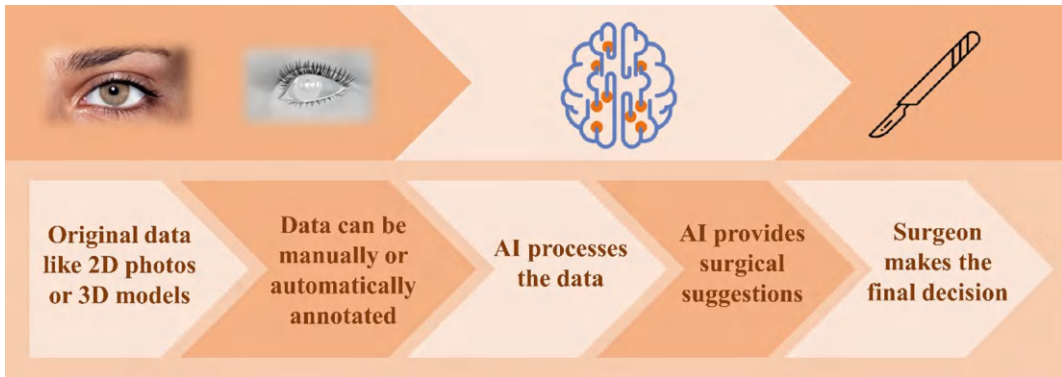


Fig. 31.1 The procedure of AI-assisted treatment

AI-Assisted Assessment of Treatment Outcomes

Treatment of eyelid diseases encompasses addressing both functionality and aesthetic aspects, with unavoidable subjectivity. Utilizing deep learning for preoperative prognostication and postoperative evaluation offers crucial information for both surgeons and patients. One common application involves integrating AI to measure preoperative and postoperative eyelid parameters, comparing them using statistical methods [26]. This aids in standardizing assessment and mitigating human errors [12]. A multi-channel CNN model was utilized by Qu et al. to generate a 3D eye model from 2D images before and after eye bag surgery. This measured alterations in 3D model features for postoperative assessment [20].

AI-Assisted Prediction

AI has been used totally as a predictor (Fig. 31.2). A decision tree model was created to forecast the complexity of excisional reconstruction surgery for patients with periocular basal cell carcinoma. This model utilized data on tumor size, surgical time, and risk evaluation collected at the first specialist appointment [29]. In addition, a fully automatic prediction system was established by Sun et al. with the assistance of a generative adversarial network. It was able to forecast postoperative appearance in cases of blepharoptosis, offering intuitive postoperative eye appearance solely based on preoperative extraocular photographs [19]. Furthermore, a comprehensive prediction system named the Global Attention-based Ophthalmic Image Measurement and Postoperative Appearance

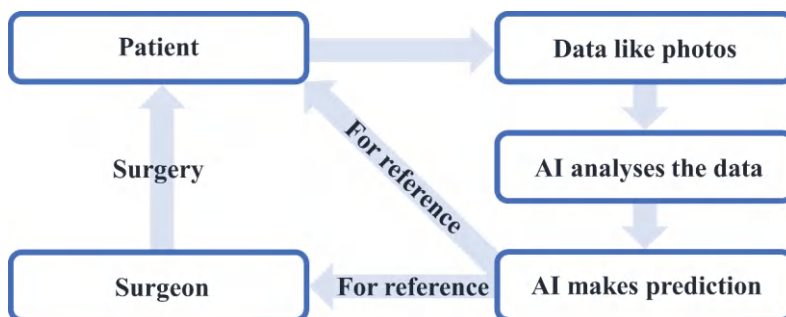


Fig. 31.2 The procedure of AI-assisted prediction

Prediction System was recently proposed [30]. It successfully overcame the challenges of different sampling models and morphological conditions, achieving ideal performance in measurement and postoperative prediction under various clinical situations.

Application of AI Technology in Orbital Diseases

Recent studies involving artificial intelligence applications in orbital diseases are showed in the Table 31.2.

AI-Assisted Segmentation and Measurement

CT and MRI imaging techniques are typically used to acquire crucial parameters of the orbit and adjacent soft tissues [31]. Due to the volumetric nature of orbital tissue, manual segmentation requires considerable time. Therefore, the significant advantages of automated segmentation algorithms became apparent. A multi-step automatic system proposed by Hamwood et al. was capable of segmenting the orbital area and locating the center point using two modified U-Net models [11]. The system performed slightly better at segmenting CT images than MRI images in their study since the bony orbits from CT images are more accurate. Pan et al. used U-Net++ to segment and extract periorbital features from a large-scale Chinese population's orbital CT imaging dataset [32]. The same methodology was utilized on a large-scale Mongolian population's orbital CT imaging dataset to assess periorbital characteristics by age group [33].

Apart from segmenting the bony orbit, deep learning can also automatically segment periorbital soft tissue [34], eyeball volume [35], cellulitis abscess lesions [36], and clinical target volume [37]. For thyroid eye disease, exophthalmos measurement is critical in assessing the severity of the disease, while current manual measurement is troublesome. Zhang et al. used a U-Net++ model to extract exophthalmos

distance from multi-directional CT images [38], with correlation coefficients of more than 0.95 between automatic and manual measurements.

AI-Assisted Screening and Diagnosis

Deep learning has diverse applications in screening orbital diseases. In 1998, Grus and his colleagues utilized a library of IgG autoantibodies from patients and healthy controls as training data, developing classification algorithms for endocrine eye diseases [39]. A recent developed diagnostic system is designed to classify thyroid eye disease by extracting signs based on extraocular photographs [16]. An integrated diagnostic model for thyroid eye disease was constructed by Karlin et al. using periocular photographs as training data, with performance being close to or even better than the performance of specialized ophthalmologists [40]. Lee et al. introduced a neural network utilizing CT images to evaluate the severity of Graves' ophthalmopathy, achieving a high area under the curve of 0.979 [41]. An automatic grading system for orbital burst fractures from CT images was presented by Li et al., based on a deep learning approach [42]. Shao et al. created an end-to-end diagnostic model for the automatic classification of benign and malignant orbital tumors from CT images, achieving an accuracy of 86.96% [43]. Deep learning has also been applied for identifying orbital cavernous venous malformations and nerve sheath tumors [44, 45], and determining periorbital area invaded by nasal/sinus tumor [46].

AI-Assisted Treatment

Deep learning can be applied in extracting disease status information and making treatment decisions. This provides valuable support for clinicians and may contribute to rehabilitation. AI measures the extent of enlarged extraocular muscles in Graves' ophthalmopathy treatment, aiding in addressing potential thyroid myopathy appropriately [47]. ML is also utilized to

Table 31.2 Recent studies involving artificial intelligence applications in orbital diseases

Authors and citations	Research target	Dataset	AI algorithm	Accuracy	Area under the curve
Hamwood et al. [11]	Orbital bony regions segmenting	443 slices of the training set and 363 pieces of the test set	Two CNNs in series and a search method	–	–
Pan et al. [32]	Extraction of Chinese periorbital features	Orbital CT images of 595 Chinese	U-Net++	–	–
Li et al. [33]	Classifying Mongolian periorbital features	Orbital CT images of 595 Mongolians	U-Net	–	–
Brown et al. [34]	Automatic segmentation of periorbital soft tissue	1018 MRI images	UNet-like CNN	–	–
Umapathy et al. [35]	Automatic quantification of eyeball volume	CT images of 80 patients	MRes-UNet	–	–
Jiang et al. [37]	Cellulitis abscess lesions segmentation	CT images of 67 patients	Context-aware CNN	–	–
Jiang et al. [37]	Extracting clinical target volume areas	CT images of 121 patients with GO	FCN-8 s network	–	–
Zhang et al. [38]	Automatically measuring exophthalmos distance	79 horizontal CT images and 99 sagittal CT images	U-Net++	–	–
Huang et al. [16]	Thyroid eye disease diagnosis and classification	Images of 3120 eyes	ResNet-50 U-Net	0.980(eye location) 0.930(sclera) 0.870(segmentation)	Over 0.850
Karlin et al. [40] Lee et al. [41]	Diagnosis of TAO Assessment of G.O. severity	Periocular photographs CT images of 288 cases	ResNet-18 A developed CNN	0.892 0.826(mild) 0.930(moderate to severe)	– 0.895(mild) 0.979(moderate to severe)
Li et al. [42]	Automatic grading for orbital burst fractures	CT scan data of 94 cases	Inception V3 CNN	0.87(patient grading)	0.9574
Shao et al. [43]	Automatically segmentation and classification of orbital tumor	64 patients (mean age: 51.69 years old)	ResNet-34	0.8696	0.8439 to 0.9546
Bi et al. [44]	Identification of orbital cavernous venous malformations and schwannoma	11,489 MRI images of cavernous hemangioma and 3478 of schwannoma	RCNN ResNet-101	0.911	0.954

(continued)

Table 31.2 (continued)

Authors and citations	Research target	Dataset	AI algorithm	Accuracy	Area under the curve
Nakagawa et al. [46]	Diagnosis of nasal/sinus tumor invasion	Orbital CT images of 168 lesions	Pre-trained CNN algorithm devoted to image classification	0.920	0.940
Hanai et al. [47]	Automatically measuring EEM	Orbital CT photographs of 371 participants	–	–	0.946
Hu et al. [48]	Assessment of the response of TAO patients to G.C. therapy	MRI images of 110 TAO patients	Logistic regression, decision tree, support vector machine classifiers	0.916(LR) 0.857(DT) 0.855(SVM)	0.875(LR) 0.781(DT) 0.844(SVM)
Wu et al. [50]	Assisting in the prediction of DON	CT images of 42 DON patients, 49 TAO patients without DON, and 87 controls	DMs-MAFM	0.96	–
Yoo et al. [52]	Predicting postoperative appearance before orbital decompression surgery	500 pairs of preoperative and postoperative extraocular images	Generative adversarial network	–	0.957

CNN, convolutional neural network; CT, computed tomography; DMs-MAFM, double multiscale and multi-attention fusion module; DON, dysthyroid optic neuropathy; D.T., decision tree; EEM, enlarged extraocular muscles; FCN, fully convolutional networks; G.C., glucocorticoid; GO, Graves’ Ophthalmopathy; L.R., logistic regression; MRes-UNet, 2D modified residual UNet; MRI, magnetic resonance imaging; RCNN, Regions with CNN features; ResNet, residual network; SVM, support vector machine; TAO, thyroid-associated ophthalmopathy

Table 31.3 Recent studies involving artificial intelligence applications in lacrimal diseases

Authors and citations	Research target	Dataset	AI algorithm	Accuracy	Area under the curve
Imamura et al. [53]	Classification of tear meniscus	117 anterior segment optical coherence tomography images	DenseNet169 (best performance)	–	0.778
Song et al. [54]	Preliminary screening of dacryocystitis	Demographic information and clinical diagnosis of 56 patients with dacryocystitis and 56 controls	A developed screening model	0.8571	–

evaluate the response of thyroid eye disease patients to glucocorticoid therapy. The model training relied on a dataset consisted of characteristic indicators obtained from T2-weighted imaging of extraocular muscles [48].

AI-Assisted Prognostication

AI was used to comprehensively assess and categorize clinical ocular manifestations, as well as predict the progression of thyroid eye disease [49]. For patients with thyroid eye disease, the incidence of the most severe progression, dysthyroid optic neuropathy, was approximately 4~8%. A decision-making system assisting in predicting dysthyroid optic neuropathy was established by Wu et al. which integrated attention mechanisms and neural learning networks [50]. Wang et al. developed a model to extract radiomic features from optic-nerve CT images, aiming to predict the effectiveness of glucocorticoid treatment [51]. A generative adversarial network was proposed to predict postoperative appearance after orbital decompression surgery [52]. A comparatively lower quality of synthesized images has been observed, but the similarity between predicted postoperative images and actual images was satisfying still. Considering the obvious postoperative appearance changes following orbital decompression surgery, image generation is valuable for surgical decision-making [52].

Application of AI Technology in Lacrimal Diseases

Nine single and 502 ensembled deep learning models proposed by Imamura et al. was able to distinguish between tear meniscus of lacrimal duct obstruction patients and normal subjects, utilizing anterior segment optical coherence tomography images [53]. DenseNet169 was the best-performing single model, with an area under the curve of 0.761. The best-performing mixed model incorporated VGG16, ResNet50, DenseNet121, DenseNet169, InceptionResNetV2, InceptionV3, and Xception, achieving an area under the curve of 0.824. An algorithm was established for the preliminary screening of dacryocystitis based on the parameters derived from ocular surface-related indices, offering a noninvasive and extensive screening approach [54]. Recent studies involving artificial intelligenceArtificial Intelligence (AI) applications in lacrimal diseases are showed in Table 31.3.

Discussion

In recent years, as the computer technology has rapidly advanced in the medical field [55], the importance of AI in oculoplastics has significantly grown. The widespread use of smartphones, which allows easy access to multi-angle photos or videos, is an opportunity for AI to

carry out parameter measurements and aid in diagnosing and treating eyelid diseases. Thyroid eye disease, particularly Graves' ophthalmopathy, is a focal point in the application of AI for orbital diseases. Various articles have explored AI applications in the glucocorticoid therapy of Graves' disease [41, 47, 51], given the variety and occasionally inconclusive nature of the clinical manifestations [41]. AI offers new solutions for parameter acquisition in clinical practice, addressing the time-consuming and inefficient aspects of manual procedures [38]. 3D eye models offer richer information compared to 2D eye models [18, 20]. The image generation capabilities of deep learning enable the prediction of postoperative appearance for patients undergoing oculoplastic surgery. This not only allows patients to gain a clearer understanding of appearance changes but also provides valuable surgical guidance for less experienced surgeons [19, 52].

However, applications of AI in oculoplastics are still facing some limitations currently. The accuracy of the supervised AI models relies on the quality of ground truth and human grading. Incorporating 3D eye models with AI models has the potential to address this issue and enhance measurement accuracy to some degree. Additionally, the limited accessibility of clinical data could impact the AI training. Many studies face the limitation in sample sizes due to data constraints. The more data are used for AI training, the better the AI's performance will be, suggesting the significance of large datasets. To overcome this limitation, possible solutions include the establishment of large-scale databases and the exploration of methods that fully utilize available data.

In the future, 3D modeling technology and image generation might be the hotspots in the field of AI in oculoplastics. 3D modeling amplifies the input dataset by providing parameters in a 3D scale. Involving 2D and 3D data together could increase the reproducibility and precision of AI diagnostic models [18]. Parameters extracted from 3D modeling avoids the problems

caused by incorrect camera angles during 2D image acquisition, mitigating the decline in AI diagnostic accuracy. Some researchers utilized a circular marker on patients' foreheads and calculated pixel/millimeter ratios, to extract eyelid morphological parameters from photographs. However, due to the fact that the forehead and the eyelid are typically on different coronal planes, it could introduce a minor error in pixel/millimeter ratios. While 3D modeling fundamentally eliminates this potential issue. 3D modeling offers more objective measurements on human face, considering the fact that face is not flat. Image generation provides prediction of postoperative appearance which is visible. Generated image with prediction of postoperative appearance could act as a reference for patients when deciding whether to undergo the surgery or at the recovery phases, promotes communication between clinicians and patients, and alleviates anxiety. It also holds promise as a tool for guiding less experienced surgeons, and assisting them in setting up their prediction style.

Conclusions

AI technology shows substantial potential for assisting in quantification, diagnosis, treatment, and prediction of oculoplastic diseases. Its efficiency in processing extensive data surpasses human capabilities, enabling more accurate evaluation and prediction. Thus, AI could support clinicians in screening, decision-making, and evaluating oculoplastic surgery outcomes. With the rapid evolution of AI algorithms, its role in oculoplastics is likely to broaden significantly in the future.

Acknowledgements This work was supported by the National Natural Science Foundation of China [No. 82000948 to L.L., No. U20A20386 to J.Y., No. 82330032 to J.Y.].

Declarations of interest None.

Abbreviations

AI	Artificial intelligence
CT	Computed tomography
CNN	Convolutional neural network
ML	Machine learning
MRD	Marginal reflex distance
MRI	Magnetic resonance imaging
2D	Two-dimensional
3D	Three-dimensional

References

- McCarthy J, Minsky ML, Shannon CE. A proposal for the Dartmouth summer research project on artificial intelligence—August 31, 1955. *AI Mag.* 2006;27(4):12–4. <https://doi.org/10.1609/aimag.v27i4.1904>.
- Nichols JA, Herbert Chan HW, Baker MAB. Machine learning: applications of artificial intelligence to imaging and diagnosis. *Biophys Rev.* 2019;11(1):111–8. <https://doi.org/10.1007/s12551-018-0449-9>.
- Dong S, Wang P, Abbas K. A survey on deep learning and its applications. *Comput Sci Rev.* 2021;40: 100379. <https://doi.org/10.1016/j.cosrev.2021.100379>.
- Jin K, Ye J. Artificial intelligence and deep learning in ophthalmology: current status and future perspectives. *Adv Ophthalmol Pract Res.* 2022;2(3): 100078. <https://doi.org/10.1016/j.aopr.2022.100078>.
- Grzybowski A, Singhanetr P, Nanegrungsunk O, et al. Artificial intelligence for diabetic retinopathy screening using color retinal photographs: from development to deployment. *Ophthalmol Ther.* 2023;12(3):1419–37. <https://doi.org/10.1007/s40123-023-00691-3>.
- Xu Z, Xu J, Shi C, et al. Artificial intelligence for anterior segment diseases: a review of potential developments and clinical applications. *Ophthalmol Ther.* 2023;12(3):1439–55. <https://doi.org/10.1007/s40123-023-00690-4>.
- Li H, Cao J, Grzybowski A, et al. Diagnosing systemic disorders with AI algorithms based on ocular images. *Healthcare (Basel).* 2023;11(12):1739. <https://doi.org/10.3390/healthcare11121739>.
- Grzybowski A. Artificial intelligence in Ophthalmology. Cham: Springer; 2021.
- Bhattacharjee K, Misra DK, Deori N. Updates on upper eyelid blepharoplasty. *Indian J Ophthalmol.* 2017;65(7):551–8. https://doi.org/10.4103/ijo.IJO_540_17.
- Cao J, Lou L, You K, et al. A novel automatic morphologic analysis of eyelids based on deep learning methods. *Curr Eye Res.* 2021;46(10):1495–502. <https://doi.org/10.1080/02713683.2021.1908569>.
- Hamwood J, Schmutz B, Collins MJ, et al. A deep learning method for automatic segmentation of the bony orbit in MRI and CT images. *Sci Rep.* 2021;11(1):13693. <https://doi.org/10.1038/s41598-021-93227-3>.
- Lou L, Cao J, Wang Y, et al. Deep learning-based image analysis for automated measurement of eyelid morphology before and after blepharoptosis surgery. *Ann Med.* 2021;53(1):2278–85. <https://doi.org/10.1080/07853890.2021.2009127>.
- Shao J, Huang X, Gao T, et al. Deep learning-based image analysis of eyelid morphology in thyroid-associated ophthalmopathy. *Quant Imaging Med Surg.* 2023;13(3):1592–604. <https://doi.org/10.21037/qims-22-551>.
- Tabuchi H, Nagasato D, Masumoto H, et al. Developing an iOS application that uses machine learning for the automated diagnosis of blepharoptosis. *Graefes Arch Clin Exp Ophthalmol.* 2022;260(4):1329–35. <https://doi.org/10.1007/s00417-021-05475-8>.
- Hung J, Perera C, Chen K, et al. A deep learning approach to identify blepharoptosis by convolutional neural networks. *Int J Med Inform.* 2021;148: 104402. <https://doi.org/10.1016/j.ijmedinf.2021.104402>.
- Huang X, Ju L, Li J, et al. An intelligent diagnostic system for thyroid-associated ophthalmopathy based on facial images. *Front Med (Lausanne).* 2022;9: 920716. <https://doi.org/10.3389/fmed.2022.920716>.
- Li Z, Qiang W, Chen H, et al. Artificial intelligence to detect malignant eyelid tumors from photographic images. *NPJ Digit Med.* 2022;5(1):23. <https://doi.org/10.1038/s41746-022-00571-3>.
- Song X, Tong W, Lei C, et al. A clinical decision model based on machine learning for ptosis. *BMC Ophthalmol.* 2021;21(1):169. <https://doi.org/10.1186/s12886-021-01923-5>.
- Sun Y, Huang X, Zhang Q, et al. A fully automatic postoperative appearance prediction system for blepharoptosis surgery with image-based deep learning. *Ophthalmol Sci.* 2022;2(3): 100169. <https://doi.org/10.1016/j.xops.2022.100169>.
- Qu Y, Lincoln B, Li S, et al. Effect of multi-channel convolutional neural network-based model on the repair and aesthetic effect of eye plastic surgery patients. *Comput Math Methods Med.* 2022;2022:5315146. <https://doi.org/10.1155/2022/5315146>.
- Schulz CB, Clarke H, Makuloluwe S, et al. Automated extraction of clinical measures from videos of oculofacial disorders using machine learning: feasibility, validity and reliability. *Eye (Lond).* 2023;37:1–7. <https://doi.org/10.1038/s41433-023-02424-z>.
- Hung J, Chen K, Perera C, et al. An outperforming artificial intelligence model to identify

- referable blepharoptosis for general practitioners. *J Pers Med.* 2022;12(2):283. <https://doi.org/10.3390/jpm12020283>.
23. Fukuda Y, Fujimura T, Moriwaki S, et al. A new method to evaluate lower eyelid sag using three-dimensional image analysis. *Int J Cosmet Sci.* 2005;27(5):283–90. <https://doi.org/10.1111/j.1467-2494.2005.00282.x>.
 24. Bao X, Sun Y, Zhan X, et al. Orbital and eyelid diseases: the next breakthrough in artificial intelligence? *Front Cell Dev Biol.* 2022;10:1069248. <https://doi.org/10.3389/fcell.2022.1069248>.
 25. Moriyama T, Kanade T, Xiao J, et al. Meticulously detailed eye region model and its application to analysis of facial images. *IEEE Trans Pattern Anal Mach Intell.* 2006;28(5):738–52. <https://doi.org/10.1109/TPAMI.2006.98>.
 26. Bahçeci Şimşek İ, Şirolu C. Analysis of surgical outcome after upper eyelid surgery by computer vision algorithm using face and facial landmark detection. *Graefes Arch Clin Exp Ophthalmol.* 2021;259(10):3119–25. <https://doi.org/10.1007/s00417-021-05219-8>.
 27. Van Brummen A, Owen JP, Spaide T, et al. PeriorbitAI: artificial intelligence automation of eyelid and periorbital measurements. *Am J Ophthalmol.* 2021;230:285–96. <https://doi.org/10.1016/j.ajo.2021.05.007>.
 28. Chen H, Tzeng S, Hsiao Y, et al. Smartphone-based artificial intelligence-assisted prediction for eyelid measurements: algorithm development and observational validation study. *JMIR Mhealth Uhealth.* 2021;9(10):e32444. <https://doi.org/10.2196/32444>.
 29. Tan E, Lin F, Sheck L, et al. A practical decision-tree model to predict complexity of reconstructive surgery after periocular basal cell carcinoma excision. *J Eur Acad Dermatol Venereol.* 2017;31(4):717–23. <https://doi.org/10.1111/jdv.14012>.
 30. Huang XR, Li Z, Lou LX, et al. GOMPS: global attention-based ophthalmic image measurement and postoperative appearance prediction system. *Expert Syst Appl.* 2023;232:120812. <https://doi.org/10.1016/j.eswa.2023.120812>.
 31. Weber AL, Sabates NR. Survey of CT and MR imaging of the orbit. *Eur J Radiol.* 1996;22(1):42–52. [https://doi.org/10.1016/0720-048X\(96\)00737-1](https://doi.org/10.1016/0720-048X(96)00737-1).
 32. Pan L, Chen K, Zheng Z, et al. Aging of Chinese bony orbit: automatic calculation based on UNet plus plus and connected component analysis. *Surg Radiol Anat.* 2022;44(5):749–58. <https://doi.org/10.1007/s00276-022-02933-8>.
 33. Li Z, Chen K, Yang J, et al. Deep learning-based CT radiomics for feature representation and analysis of aging characteristics of Asian bony orbit. *J Craniofac Surg.* 2022;33(1):312–8. <https://doi.org/10.1097/SCS.00000000000008198>.
 34. Brown RA, Fetco D, Fratila R, et al. Deep learning segmentation of orbital fat to calibrate conventional MRI for longitudinal studies. *Neuroimage.* 2020;208: 116442. <https://doi.org/10.1016/j.neuroimage.2019.116442>.
 35. Umapathy L, Winegar B, MacKinnon L, et al. Fully automated segmentation of globes for volume quantification in CT images of orbits using deep learning. *AJNR Am J Neuroradiol.* 2020;41(6):1061–9. <https://doi.org/10.3174/ajnr.A6538>.
 36. Fu R, Leader JK, Pradeep T, et al. Automated delineation of orbital abscess depicted on CT scan using deep learning. *Med Phys.* 2021;48(7):3721–9. <https://doi.org/10.1002/mp.14907>.
 37. Jiang J, Luo Y, Wang F, et al. Evaluation on auto-segmentation of the clinical target volume (CTV) for Graves' ophthalmopathy (GO) with a fully convolutional network (FCN) on CT images. *Curr Med Imaging.* 2021;17(3):404–9.
 38. Zhang Y, Rao J, Wu X, et al. Automatic measurement of exophthalmos based orbital CT images using deep learning. *Front Cell Dev Biol.* 2023;11:1135959. <https://doi.org/10.3389/fcell.2023.1135959>.
 39. Grus FH, Augustin AJ, Toth-Sagi K. Diagnostic classification of autoantibody repertoires in endocrine ophthalmopathy using an artificial neural network. *Ocul Immunol Inflamm.* 1998;6(1):43–50. <https://doi.org/10.1076/ocii.6.1.43.8082>.
 40. Karlin J, Gai L, LaPierre N, et al. Ensemble neural network model for detecting thyroid eye disease using external photographs. *Br J Ophthalmol.* 2022. <https://doi.org/10.1136/bjo-2022-321833>.
 41. Lee J, Seo W, Park J, et al. Neural network-based method for diagnosis and severity assessment of Graves' orbitopathy using orbital computed tomography. *Sci Rep.* 2022;12(1):12071. <https://doi.org/10.1038/s41598-022-16217-z>.
 42. Li L, Song X, Guo Y, et al. Deep convolutional neural networks for automatic detection of orbital blow-out fractures. *J Craniofac Surg.* 2020;31(2):400–3. <https://doi.org/10.1038/s41598-022-16217-z>.
 43. Shao J, Zhu J, Jin K, et al. End-to-end deep-learning-based diagnosis of benign and malignant orbital tumors on computed tomography images. *J Pers Med.* 2023;13(2):204. <https://doi.org/10.3390/jpm13020204>.
 44. Bi S, Chen R, Zhang K, et al. Differentiate cavernous hemangioma from schwannoma with artificial intelligence (AI). *Ann Transl Med.* 2020;8(11):710. <https://doi.org/10.21037/atm.2020.03.150>.
 45. Han Q, Du L, Mo Y, et al. Machine learning based non-enhanced CT radiomics for the identification of orbital cavernous venous malformations: an innovative tool. *J Craniofac Surg.* 2022;33(3):814–20. <https://doi.org/10.1097/SCS.00000000000008446>.
 46. Nakagawa J, Fujima N, Hirata K, et al. Utility of the deep learning technique for the diagnosis of orbital invasion on CT in patients with a nasal or sinonasal tumor. *Cancer Imaging.* 2022;22(1):52. <https://doi.org/10.1186/s40644-022-00492-0>.
 47. Hanai K, Tabuchi H, Nagasato D, et al. Automated detection of enlarged extraocular muscle in Graves'

- ophthalmopathy with computed tomography and deep neural network. *Sci Rep.* 2022;12(1):16036. <https://doi.org/10.1038/s41598-022-20279-4>.
48. Hu H, Chen L, Zhang J, et al. T2 -weighted MR imaging-derived radiomics for pretreatment determination of therapeutic response to glucocorticoid in patients with thyroid-associated ophthalmopathy: comparison with semiquantitative evaluation. *J Magn Reson Imaging.* 2022;56(3):862–72. <https://doi.org/10.1002/jmri.28088>.
 49. Salvi M, Dazzi D, Pellistri I, et al. Prediction of the progression of thyroid-associated ophthalmopathy at first, ophthalmologic examination: use of a neural network. *Thyroid.* 2002;12(3):233–6. <https://doi.org/10.1089/105072502753600197>.
 50. Wu C, Li S, Liu X, et al. DMs-MAFM+EfficientNet: a hybrid model for predicting dysthyroid optic neuropathy. *Med Biol Eng Comput.* 2022;60(11):3217–30. <https://doi.org/10.1007/s11517-022-02663-4>.
 51. Wang Y, Wang H, Li L, et al. Novel observational study protocol to develop a prediction model that identifies patients with Graves' ophthalmopathy insensitive to intravenous glucocorticoids pulse therapy. *BMJ Open.* 2021;11(12): e053173. <https://doi.org/10.1136/bmjopen-2021-053173>.
 52. Yoo TK, Choi JY, Kim HK. A generative adversarial network approach to predicting postoperative appearance after orbital decompression surgery for thyroid eye disease. *Comput Biol Med.* 2020;118: 103628. <https://doi.org/10.1016/j.compbiomed.2020.103628>.
 53. Imamura H, Tabuchi H, Nagasato D, et al. Automatic screening of tear meniscus from lacrimal duct obstructions using anterior segment optical coherence tomography images by deep learning. *Graefes Arch Clin Exp Ophthalmol.* 2021;259(6):1569–77. <https://doi.org/10.1007/s00417-021-05078-3>.
 54. Song X, Li L, Han F, et al. Noninvasive machine learning screening model for dacryocystitis based on ocular surface indicators. *J Craniofac Surg.* 2022;33(1):e23–8. <https://doi.org/10.1097/SCS.00000000000007863>.
 55. Rajesh A, Asaad M. Artificial intelligence and machine learning in surgery. *Am Surg.* 2023;89(1):9–10. <https://doi.org/10.1177/00031348221117024>.
 56. Lou L, Yang L, Ye X, et al. A novel approach for automated eyelid measurements in blepharoptosis using digital image analysis. *Curr Eye Res.* 2019;44(10):1075–9. <https://doi.org/10.1080/02713683.2019.1619779>.

Index

A

Age-Related Eye Disease Studies (AREDS), 124, 218–221

Age-related Macular Degeneration (AMD), 32, 37, 38, 121, 122, 125, 126, 217–221, 319, 324, 373, 374, 387, 388, 444

AREDS, 124

automated image analysis, 122

clinical features, 122, 123

deep learning, 126

imaging modalities, 122, 123

limitation, 131

Simplified Severity Scale, 124, 125

AI-based devices

- aims, 2, 3
- benefit for, 4, 5, 7
- challenges, 11–14
- continuous learning, 11
- cost-effectiveness, 14–16
- design factors, 14, 15
- developments, 2, 3
- DL models, 4
- doctor-patient relationship, 2, 3
- GANs, 9–11
- guidelines, 14
- human cognitive capacity, 3
- industrial revolution, 2
- life-science papers, 4
- medical specialties, 4
- optical coherence tomography images, 7, 8
- programs creation, 2
- resources, 7, 8
- risk factors, 4, 6
- safety, 7
- transfer learning, 8–10
- turing test, 2

AI-based risk stratification, 296

AI-DL models on external multi-ethnic testing datasets, 297

AI safety, 16, 78

Anaemia, 292

Area Under the Curve (AUC), 30

AREDS 9-step severity scale, 219

Artificial Intelligence (AI), 1–5, 7, 8, 11–16, 19–24, 28, 30–35, 37, 38, 46, 52–62, 69–79, 84, 93, 96, 97, 103–116, 122, 126, 129, 130, 138–140, 146, 147, 154, 155, 158, 159, 161–163, 173–181, 184–191, 196, 199, 201–203, 205–208, 223–229, 231, 232, 237–239, 243–249, 251–258, 264, 270, 281, 283, 284, 287–298, 300, 301, 305–315, 319–321, 323, 324, 326, 331, 333, 334, 338, 339, 343–345, 349–356, 359–364, 366–369, 373–378, 380–383, 386–388, 390, 391, 393–395, 398–400, 404, 405, 407, 408, 417–424, 426, 427, 433–437, 439, 443–457, 462, 466, 467, 469, 470, 473–476, 478, 479

Artificial intelligence in ophthalmology, 443

Artificial Neural Networks (ANN), 27

Artificial neurons, 27

Assisted treatment, 474

Assistive AI, 69, 72, 74

Augmented Intelligence, 69

Automatic Retinal Imaging and Analysis, 218–221

Autonomous systems, 52

B

Basic principles, 19

Bayes' theorem, 25

Big GANs, 220

Blood Pressure (BP), 291

Boundary equilibrium GANs, 220

C

CAC scores, 295

Cardiovascular disease, 114, 288, 299, 399, 404, 405

Cataract, 16, 33, 37, 38, 57–59, 105, 108, 195, 243–245, 247–249, 261, 265, 267, 268, 305, 310, 313, 314, 373–376, 385, 387, 390–392, 419, 420, 444, 457

Cataract detection, 244, 391

Cataract grading, 108

Cataract management, 109, 243, 248, 249, 387

- Cataract surgery, 28, 58, 109, 115, 244–246, 261–263, 265–270, 305, 306, 314, 387, 436
- ChatGPT, 21, 22, 28, 37, 48, 62, 146, 398, 418, 425, 427, 433–439
- Clinical decision-making, 116, 121, 127, 139, 146, 147, 206, 374, 433, 435
- Clinical development, 319, 321, 324, 326, 331
- Color Fundus Photograph (CFP), 122
- Conditional GANs, 220
- Conformité Européenne (CE) marking, 7
- Confusion matrix, 30
- Conventional Cox proportional hazard model, 296
- Conventional Neural Network (CNN), 27
- Coronary Artery Calcium (CAC), 295
- Coupling effects, 257
- COVID-19 pandemic, 4
- D**
- Data analysis, 4, 31, 115, 227, 321, 393, 421, 427
- Data bias, 97, 99
- Data labeling, 97, 99
- Dataset, 8, 12, 15, 21, 23–26, 28, 33, 34, 37, 41, 44–50, 53–55, 57–59, 61, 62, 83–85, 91, 96, 108, 111, 112, 115, 122, 124, 126–131, 140, 141, 143, 144, 147, 159–163, 173, 175, 178, 180–185, 187, 188, 190, 201, 205, 217–219, 224, 225, 228, 229, 231, 232, 244–246, 254, 267–270, 277, 282, 284, 291, 297, 298, 300, 310–312, 321, 323, 324, 331, 353–357, 360, 361, 364, 368, 369, 376–381, 383, 384, 386, 388–390, 392, 393, 396, 398, 400, 407, 419, 439, 463–466, 470, 473, 475, 478, 479
- DCNN technical refinements to synthetic images generation, 221
- Decision tree, 26
- Deep learning, 6, 20, 21, 27, 30–33, 37, 38, 40, 41, 43, 83, 91, 103, 106, 107, 111–113, 116, 121, 122, 126–131, 138–142, 144–146, 154, 158, 160, 163, 173, 182, 187, 190, 191, 199, 217, 226, 247, 257, 265, 268, 270, 273, 281, 287, 288, 319–321, 323, 326–330, 333, 339, 343, 353–355, 358–363, 367–369, 373–378, 380–382, 386, 389, 390, 394, 396, 398–400, 404, 405, 407, 408, 444, 450, 453, 454, 456, 465, 466, 469, 470, 473–475, 478, 479
- AMD, 126–128
- GA, 128, 129
- overview, 21
- retinal disease, 32
- RPD, 129, 130
- social aspects, 34
- technical aspects, 33
- Deep medicine, 2
- Deep-Learning (DL) models, 4
- Deep Neural Networks (DNN), 27, 28
- Delirium, 425–427
- Detection, 4, 25, 28, 32, 33, 35, 38, 54–60, 70, 74, 85, 89, 91, 97, 98, 106–113, 115, 121, 122, 128–131, 141, 161, 162, 172–174, 176, 177, 180, 182–185, 187, 189, 199, 200, 202, 203, 205, 206, 209, 223–229, 232, 253, 254, 267, 281, 282, 284, 288, 290, 294, 321, 334, 336, 339, 343, 349, 350, 352, 355, 362, 367, 373, 376, 377, 387, 389, 394, 398, 402, 404, 407, 426, 444, 453, 462, 465, 466
- Diabetic retinopathy, 7, 10, 14, 15, 32, 37, 38, 52, 61, 69–72, 74–77, 95, 104, 113, 123, 126, 130, 169–171, 174, 177, 182, 184, 185, 288, 299, 324, 330, 338, 352, 368, 373–376, 378, 379, 383, 419, 420, 444, 449, 450, 453, 454
- Digital Colour Fundus Photographs (CFP), 288
- AI-DL on, 289
- body composition factors, 289, 291
- chronic kidney disease, 294
- circulatory system, 291
- demographic and lifestyle factors, 290
- endocrine system, 294
- framework for AI-DL, 289
- framework for artificial intelligence, 289
- hematocrit, 292
- hematologic factors in external datasets, 292
- hemoglobin, 292
- imaging biomarker, 296
- lifestyle factors, 289
- longitudinal outcomes of systemic diseases, 298
- prediction of kidney disease, 294
- prediction of metabolic and endocrinological diseases, 292
- prediction of neurological diseases, 291
- predictive values and cross-sectional outcomes, 296
- red blood cell, 292
- renal system, 295
- risk stratification and delivery of tailored interventions, 297
- smoking status, 289
- systemic diseases, circulatory system, 293
- Discriminative network, 220
- DLS's classification accuracy, 220
- Dry eye disease, 333, 334, 339
- E**
- Experimental protocol, 94
- Eyelid diseases, 469, 470, 473, 474, 479
- Eye surgeries, 374, 417–421, 425–427
- F**
- Feedforward Neural Network (FFN), 27
- Forme frustre keratoconus, 223, 224
- Framingham Risk Score (FRS), 297
- Fundus Autofluorescence (FAF), 123
- G**
- Generative Adversarial Networks (GANs), 9–11, 220
- Generative AI, 19, 20, 22, 24, 33, 34, 106, 300, 437

Generative network, 220

Geographic Atrophy (GA), 55, 113, 122, 124, 125,
127–129, 131, 219, 320, 324, 380, 388, 400,
425

Glaucoma, 16, 32, 33, 37, 38, 56, 57, 60, 74, 76, 84,
104, 111, 112, 123, 126, 130, 137–147, 176,
195–199, 201, 203–209, 284, 319, 338, 352,
359, 373–377, 379, 380, 382, 383, 385, 386,
392, 407, 419, 420, 453, 454, 456, 457, 462

H

Hill-RBF formula, The, 311

Hoffer QST formula, The, 310

Holdout validation, 29

I

Iatrogenic Ectasia, 223, 224

ImageNet Large Scale Visual Recognition Challenge
(ILSVRC), 2

Implantable Collamer Lens (ICL), 255–257

Intraocular lens power calculation formula, 306

Introduction, 16, 70, 71, 74, 79, 121, 173, 176, 181, 208,
223, 229, 319, 368, 369, 375, 434

K

Kane formula, The, 309, 310, 314

Keras, 31

Keratoconus, 25, 33, 105, 106, 223–229, 231, 232, 235,
238, 252–254, 373, 375, 394, 398–402

Keratoconus suspects, 106, 223, 224, 253

K-Nearest Neighbor (KNN), 26

L

Lacrimal diseases, 469, 470, 478

Laplacian Pyramid GANs, 220

Large language models, 20–22, 37, 62, 139, 146, 196,
281, 283, 433

LASIK, 33, 107, 224, 228, 232, 251, 254

Leave-P-Out Cross Validation, 30

Linear regression, 24

Low-Shot Learning (LSL), 8

M

Machine Learning (ML), 1, 8, 12, 20, 21, 23, 24, 26,
27, 29–32, 34, 37–39, 46–48, 50, 72, 75,
83–85, 93, 94, 97–99, 105, 107, 109, 122, 126,
138–140, 145, 146, 159, 176, 189, 190, 199,
223, 225–229, 231, 244, 245, 252, 254, 255,
265, 273, 292, 296, 310–312, 314, 319, 323,
327, 329, 331, 333, 339, 343, 354, 360, 362,
364, 366–368, 373, 374, 380, 393, 398, 400,
407, 418–423, 426, 427, 439, 444, 450, 462,
464, 469, 470

ANN, 27

classification, 23

clustering, 23

decision tree, 26

DNN, 27, 28

glaucoma, 32

keratoconus, 33

K-means clustering, 26

KNN, 26

learning models, 24

logistic regression, 25

model development, 31

Naïve Bayes algorithm, 25

non-linear regression, 24, 25

overview, 21

parameters, 30

performance evaluation

confusion matrix, 30

K folds, 30

metrics, 29

ROC curve, 30

prediction, 23, 24

refractive surgery, 33

regression, 24

retinal disease, 32

social aspects, 34

software packages, 31, 32

SVM, 26

technical aspects, 33

Machine learning (ML) problems, 8

Major Adverse Cardiac Events (MACE), 297

MATLAB, 32

Mean Absolute Error (MAE), 290

Medicine, 1–3, 5, 12, 16, 23, 24, 99, 115, 122, 126, 143,
174, 217, 220, 223, 243, 281, 300, 433, 434,
439, 447, 449, 453, 454, 462

Metrics, 29

Model evaluation, 95, 276

Multilayer Perceptron (MLP), 27

Myopia, 33, 59, 76, 107, 109, 141, 184, 256, 306,

349–353, 355–357, 359–369, 373, 375, 390,
393–397, 457

N

Naïve Bayes algorithm, 25

Neovascular AMD, 59, 61, 122, 129, 130, 324, 328, 329

Nursing care, 417–420, 426, 427

Nursing diagnosis, 421, 424–426

O

Ocular imaging, 52, 107, 126, 288, 291

Oculoplastics, 262, 436, 469, 470, 478, 479

Ophthalmology, 1, 4, 5, 7, 11, 15, 16, 19, 32, 34, 37, 38,
42, 59, 61, 62, 84, 85, 87, 88, 90, 91, 93, 95,
97, 98, 103, 105, 109, 112–115, 122, 126, 138,
146, 147, 155, 161–163, 175, 181, 184, 187,
195, 196, 199, 206–208, 218, 244, 249, 257,
261, 262, 265, 273, 281–284, 288, 295, 300,

321, 331, 337, 352, 373–375, 379, 381, 398, 401, 404, 407, 408, 433–437, 439, 443–457, 462, 469

Optical Coherence Tomography (OCT), 4, 7, 52, 70, 104, 106, 123, 137, 142, 161, 189, 196, 221, 283, 288, 300, 319, 321, 333, 335, 339, 353, 358, 361, 420, 435, 454, 478

Optic neuropathies, 45, 56, 111, 112, 141, 184, 195, 283, 377, 386, 454, 477, 478

Orbital diseases, 469, 470, 475, 476, 479

Out-of-distribution, 98, 323, 324

OverFeat4 Deep Convolutional Neural Network (DCNN), 218

Overfitting, 29

P

PEARL-DGS formula, The, 310

Pediatric cataract, 109, 247, 248

Phacoemulsification, 246, 262, 263, 265, 267–270, 305, 306

Population-based studies, 288

Postoperative management, 258

Precision medicine, 255, 333

Predictability and generalisability of the AI-DL model, 297

Prediction, 1, 6, 16, 23, 24, 29, 31, 32, 37, 38, 41, 42, 44–47, 49, 51, 56, 57, 59–61, 84, 85, 91, 95, 98, 105, 107, 109, 113, 114, 127, 129–131, 139, 144–147, 163, 191, 196, 199, 203, 219, 220, 232, 245, 246, 249, 255–257, 265, 274, 276, 281, 289, 292, 294, 296, 297, 305, 306, 308–314, 329, 349, 353–356, 359, 361, 362, 364, 366, 368, 373–377, 380, 382, 383, 387, 393, 395, 397, 399, 404, 406, 408, 418, 419, 426, 436, 450, 457, 463, 464, 469, 474, 475, 477, 479

Progression, 32, 37, 54–57, 59–61, 76, 103, 105, 113, 124, 127–129, 131, 137, 138, 143, 144, 155–157, 160–162, 175–178, 190, 196, 199, 219–221, 228, 232, 235, 246, 328, 350, 351, 353, 362, 364–366, 368, 369, 373–378, 380, 381, 384–386, 388, 390, 393–397, 407, 422, 454, 457, 464, 478

Publication trends, 446, 451

Python, 31

R

Random Forest, 27

Receiver Operating Characteristic (ROC) Curve, 30

Recurrent Neural Network (RNN), 28

Refractive surgery, 16, 32, 33, 58, 105–107, 107, 108, 224, 226–229, 231, 232, 251–255, 257, 258, 310, 313, 335, 338, 398, 457

Reinforcement Learning (RL), 24, 28

ResNet 50 derived classifications, 219

Reticular Pseudodrusen (RPD), 55, 124, 129–131

Retinal imaging biomarkers, 295

Retinal microvascular abnormalities and systemic diseases, 288

Retinal vascular changes, 288

Retinopathy of prematurity, 32, 37, 58, 104, 105, 109, 110, 126, 153–155, 157, 162

Risk stratification, 129, 247, 296, 297, 349, 464

R language, 31

S

Screening, 5, 14, 15, 34, 37, 38, 52–54, 58–61, 72, 104–107, 116, 131, 142, 153–155, 157, 158, 161–163, 170–191, 197, 207, 223–225, 227, 229, 232, 247, 249, 252, 255, 258, 276, 290–292, 294, 297–299, 327, 328, 338, 345, 350, 352–354, 359, 360, 362, 367–369, 374–377, 381, 386, 387, 391, 394, 398, 399, 404, 407, 408, 419, 444–446, 453, 456, 457, 466, 470, 471, 473, 475, 478, 479

Self-attention GANs, 220

Self-Organization Maps (SOM), 28

Semi-structured data, 23

Semi-supervised learning, 24

Shallow medicine, 2, 3

Singapore I Vessel Assessment deep-learning system (SIVA-DLS), 295

SIVA-DLS, 296

SMILE, 251, 254–256

Software as a Medical Device (SaMD), 12, 14

Stratified K-fold cross validation, 30

Structured data, 23

Subretinal drusenoid deposits, 129, 130

Superior DL approach achieved an area-under-receiver-operating-characteristic-curve (AUC-ROC curve) value, 218

Supervised learning, 24

Support Vector Machine (SVM), 26

Surgical skill, 114, 115, 261–266, 269, 270, 420

Surgical training, 114, 207, 261, 270

Systematic COronary Risk Evaluation (SCORE) risk calculator, 297

Systemic biomarker, 288, 443, 457

Systemic disease, 16, 130, 287–289, 291–297, 300, 301, 373, 375, 398, 399, 404, 405, 457

T

TensorFlow, 31

Turing test, 1

U

Uncorrected refractive errors, 349, 354, 355, 368, 374

Underfitting, 29

Unstructured data, 23

Unsupervised learning, 24

US Food and Drug Administration (FDA) approval, 5, 7

V

Visual Impairment (VI), [154](#), [196](#), [247](#), [300](#), [349](#), [350](#),
[352](#), [359](#), [373](#), [374](#), [377](#), [380](#), [387](#), [425](#), [456](#)

W

Weka, [31](#)