

Advances in AI for Cloud, Edge, and Mobile Computing Applications



Shrikaant Kulkarni
P. William
Editors



CRC Press
Taylor & Francis Group

APPLE ACADEMIC PRESS

Advances in AI for Cloud, Edge, and Mobile Computing Applications



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Advances in AI for Cloud, Edge, and Mobile Computing Applications

Shrikaant Kulkarni, PhD

P. William, PhD

Editors



First edition published 2026

Apple Academic Press Inc.
1265 Goldenrod Circle, NE,
Palm Bay, FL 32905 USA

760 Laurentian Drive, Unit 19,
Burlington, ON L7N 0A4, Canada

CRC Press

2385 NW Executive Center Drive,
Suite 320, Boca Raton FL 33431

4 Park Square, Milton Park,
Abingdon, Oxon, OX14 4RN UK

© 2026 by Apple Academic Press, Inc.

Apple Academic Press exclusively co-publishes with CRC Press, an imprint of Taylor & Francis Group, LLC

Reasonable efforts have been made to publish reliable data and information, but the authors, editors, and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors, editors, and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged, please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, access www.copyright.com or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. For works that are not available on CCC please contact mpkbookspermissions@tandf.co.uk

Trademark notice: Product or corporate names may be trademarks or registered trademarks and are used only for identification and explanation without intent to infringe.

For Product Safety Concerns and Information please contact our EU representative GPSR@taylorandfrancis.com Taylor & Francis Verlag GmbH, Kaufingerstraße 24, 80331 München, Germany

Library and Archives Canada Cataloguing in Publication

Title: Advances in AI for cloud, edge, and mobile computing applications / Shrikaant Kulkarni, PhD, P. William, PhD, editors.

Names: Kulkarni, Shrikaant, editor | William, P., editor.

Description: First edition. | Includes bibliographical references and index.

Identifiers: Canadiana (print) 20250137232 | Canadiana (ebook) 20250137240 | ISBN 9781779643735 (hardcover) | ISBN 9781779520685 (softcopy) | ISBN 9781003637707 (ebook)

Subjects: LCSH: Computational intelligence. | LCSH: Artificial intelligence.

Classification: LCC Q342 .A38 2025 | DDC 006.3—dc23

Library of Congress Cataloging-in-Publication Data

CIP data on file with US Library of Congress

ISBN: 978-1-77964-373-5 (hbk)

ISBN: 978-1-77952-068-5 (pbk)

ISBN: 978-1-00363-770-7 (ebk)

About the Editors

Shrikaant Kulkarni, PhD

Adjunct Professor, Faculty of Business, Victorian Institute of Technology, Melbourne, Australia; Adjunct Professor, Center of Research Outcome and Impact, Chitkara University, Punjab, India

Shrikaant Kulkarni, PhD, is currently an Adjunct Professor in the Faculty of Business at the Victoria Institute of Technology, Melbourne, Australia, as well as an Adjunct Professor at the Center of Research Outcome and Impact at Chitkara University, Punjab, India. Dr. Kulkarni has been a senior academician and researcher for over four decades. He has delivered invited lectures and conducted sessions at national and international conferences and faculty development programs. He has guided many major and minor projects in engineering chemistry, green chemistry, nanotechnology, analytical chemistry, catalysis, chemical engineering materials, industrial organization, management, and AI. He has published over 100 research papers in national and international journals and conferences of repute. He has authored more than 50 book chapters. He has edited 25 books published by international publishers. Another 10 books are in the offing. He authored four textbooks in engineering chemistry. Dr. Kulkarni possesses MSc, MPhil, and PhD degrees in Chemistry, apart from master's degrees in Economics, Business Management, and Political Science. He has expertise in materials science, green chemistry and engineering, analytical chemistry, green nanoscience, nanotechnology, AI, and computational intelligence. He has expertise in wastewater treatment, green and analytical chemistry, and advanced areas in chemical engineering and materials science. He has published two patents and applied for two more.

P. William, PhD

Dean, Research & Development and Assistant Professor in the Department of Information Technology, Sanjivani College of Engineering, SPPU, Pune, India

P. William, PhD, is working as Dean of Research & Development and Assistant Professor in the Department of Information Technology at Sanjivani College of Engineering, SPPU, Pune, India. He has published many papers in Scopus-indexed journals and IEEE conferences. His field of research includes natural language processing, artificial intelligence, deep learning, machine learning, soft computing, cyber security, and cloud computing. He has been associated with numerous multinational companies, including IBM, TCS, etc., and educational groups. He is a focused and hardworking professional with experience in consulting in research, innovation, and development. He has served as session chair and keynote speaker at multiple SCOPUS-indexed international conferences and has been appointed as a reviewer in reputed SCOPUS/WOS-indexed journals. He is a member of the Institute of Electrical and Electronics Engineers (IEEE), the Quality Circle Forum of India (QCFI), and various other professional bodies. He received his Bachelor of Engineering in Computer Science and Engineering and Master of Technology in Computer Science and Engineering from CSVTU, Bhilai, India. He completed his PhD in the Department of Computer Science and Engineering in the School of Engineering & Information Technology, MATS University, Raipur, India.

Contents

Contributors..... ix

Preface xv

Introduction..... xvii

PART I: Computational Intelligence: Edge AI Services..... 1

1. Power Trading Framework of Cloud-Edge Computing in an Artificial Intelligence Bazaar 3
Sreenivas Mekala, Ananda Ravuri, K. Suresh Kumar,
Giovanna Gutiérrez Gayoso, and Sandeep Gupta

2. Anomaly Detection Using Artificial Intelligence in a Cloud Fuzzy Environment..... 25
M. K. Sharma, Vijaykumar, Bibin K. Jose, R. P. Ambilwade, and R. Revathi

3. Holistic Resource Management Based on AI and Sustainable Cloud Computing 51
Veena Prasad Vemuri, Baireddy Ravinder Reddy, Vineela Lagadapati,
Noor Firdoos Jahan, and Gowri Bharath Kumar

PART II: Computational Intelligence: Edge AI Security and Privacy 71

4. Application of Artificial Intelligence for Enhancement of Privacy and Security in Smart Environment 73
Ananda Ravuri, Khushboo Malik, Konda Hari Krishna, Abhijeet Das,
Ripon Bhattacharjee, Joshuva Arockia Dhanraj

5. Application of Artificial Intelligence in Wireless Sensor Network Through Quality of Services Enhancement..... 95
Tamirat Tagesse Takore, Syed Jahangir Badashah, Mohit Tiwari, A. Prabhu,
Venkanna Mood, and P. William

6. Recent Advancements of Artificial Intelligence in Edge Computing 117
Chetan Shelke, Rathnakar Achary, Sreenivas Mekala,
K. K. Ramachandran, and Shaileshkumar Wagh

PART III: Computational Intelligence: Edge Computing and AI applications 141

7. Future Innovations Using Edge Computing Infrastructure in IoT Applications 143
Neha Tyagi, Jitender Mittal, Dler Salih Hasan, Sunil Bajaja, and Shashidhar Sonnad

8. Understanding the Growth of Edge Computing with Artificial Intelligence	167
Chandrshekhhar Goswami, K. Suresh Kumar, R. Kaviyaraj, Prateek Srivastava, and R. Udhayakumar	
9. Perspectives of Machine Learning in the Convergence of Artificial Intelligence and Edge Computing	189
Abu Sarwar Zamani, R. M. Jagadish, Binod Kumar, Md. Atheeq Sultan Ghor, and Suhas A. Bhyratae	
10. Applications, Enabling Technologies, Vision of Edge Artificial Intelligence for 6G	213
Tamirat Tagesse Takore, Konda Hari Krishna, K. Suresh Kumar, R. Ganesh Kumar, Amit Jain, and T. Keerthika	
PART IV: Computational Intelligence: IoT Systems	239
11. AI Products in E-Commerce and Commercial Bank	241
M. Nithya, Piyush Mehta, Veena Prasad Vemuri, Meeta Joshi, and Tripti Tiwari	
12. Sustainable Green Human Resource Management Using Artificial Intelligence	265
Ashutosh Singh, Nakshatresh Kaushik, Meeta Joshi, A. Prabhu, and Rupesh Roshan Singh	
13. Natural Resource Management and Artificial Intelligence for Sustainable Development	285
Ankur Kulshrestha, Diksha Srivastava, M. Angulakshmi, Sachin Tripathi, and M. Z. M. Noman	
14. A Sustainable Development Goal Perspective Using Business Models and Artificial Intelligence	305
Alpana Sharma, Vidhika Tiwari, Rajesh Deb Barman, Kousik Boro, Ganesh P. Shukla, and P. William	
15. Different Requirements in Quality of Service Using an Adaptive Network Algorithm.....	323
J. Jayaganesh, K. Suresh Kumar, Konda Hari Krishna, Mohit Tiwari, R. Vettriselvan, and Chetan Shelke	
Index.....	345

Contributors

Rathnakar Achary

Associate Professor, Alliance College of Engineering and Design, Alliance University, Bangalore, Karnataka, India

R. P. Ambilwade

Assistant Professor, Computer Science, National Defense Academy, Pune, Maharashtra, India

M. Angulakshmi

Associate Professor, School of Computer Science Engineering and Information Systems, Vellore Institute of Technology, Vellore, Tamil Nadu, India

Syed Jahangir Badashah

Professor, ECE Department, Sreenidhi Institute of Science and Technology, Hyderabad, Telangana, India

Sunil Bajeja

Associate Professor, Faculty of Computer Applications, Marwadi University, Rajkot, Gujarat, India

Rajesh Deb Barman

Assistant Professor, Department of Commerce, Bodoland University, Kokrajhar, Assam, India

Ripon Bhattacharjee

Assistant Professor, National Law University, Tripura, India

Suhas A. Bhyratae

Associate Professor, Computer Science and Engineering, Sahyadri College of Engineering and Management, Mangalore, Karnataka, India

Kousik Boro

Research Scholar, Department of Commerce, Bodoland University, Kokrajhar, Assam, India

Abhijeet Das

Research Scholar, Department of Civil Engineering, C.V. Raman Global University (CGU), Bhubaneswar, Odisha, India

Joshuva Arockia Dhanraj

Center for Automation and Robotics (ANRO), Department of Mechatronics Engineering, Hindustan Institute of Technology and Science, Padur, Chennai, Tamil Nadu, India

Giovanna Gutiérrez Gayoso

Universidad Andina del Cusco, Universidad Andina del Cusco, Peru

Md. Atheeq Sultan Ghori

Associate Professor, Computer Science and Engineering Department, Telangana University, Nizamabad, Telangana, India

Chandrshekhhar Goswami

Associate Professor, Department of CSE, School of Computing, MIT ADT University, Pune, Maharashtra, India

Sandeep Gupta

Associate Professor, Department of Electrical Engineering, Graphic Era (Deemed to be University), Dehradun, Uttarakhand, India

Dler Salih Hasan

Lecturer, Computer Science and Information Technology, College of Science/University of Salahaddin, Erbil, Iraq

R. M. Jagadish

Associate Professor, Department of CSE (DS), Ballari Institute of Technology and Management, Ballari, Karnataka, India

Noor Firdoos Jahan

Professor, RV Institute of Management, Bangalore, Karnataka, India

Amit Jain

Professor, Computer Science and Engineering Department, OP Jindal University, Raigarh, Chhattisgarh, India

J. Jayaganesh

Assistant Professor, Department of Computer Science, Government Arts and Science College, Perumbakkam, Chennai, Tamil Nadu, India

Bibin K. Jose

Associate Professor, PG and Research Department of Mathematics, Sanatana Dharma College, Alappuzha, Kerala, India

Meeta Joshi

Professor, Faculty of Management Studies, Marwadi University, Gujarat, India

Nakshatresh Kaushik

Associate Professor, Management, Lloyd Business School, Greater Noida, Uttar Pradesh, India

R. Kaviyaraj

Assistant Professor, Computational Intelligence, RM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India

T. Keerthika

Associate Professor, Department of IT, Sri Krishna College of Engineering and Technology, Coimbatore, India

Konda Hari Krishna

Associate Professor, Department of CSE, School of Computing, Mohan Babu University, Sree Sainath Nagar, Rangampeta, Tirupati, Andhra Pradesh, India

Ankur Kulshrestha

Assistant Professor, Department of Mechanical and Aerospace Engineering, NIMS University, Jaipur, Rajasthan, India

Binod Kumar

Professor, MCA, JSPM Rajarshi Shahu College of Engineering, Pune, Maharashtra, India

Gowri Bharath Kumar

Assistant Professor, Computer Science and Engineering, Chalapathi Institute of Engineering and Technology, Lam, Guntur, Andhra Pradesh, India

K. Suresh Kumar

Associate Professor, MBA Department, Panimalar Engineering College, Varadarajapuram, Poonamallee, Chennai, Tamil Nadu, India

R. Ganesh Kumar

Associate Professor, Department of Computer Science and Engineering, CHRIST (Deemed to be University), School of Engineering and Technology, Kengeri Campus, Kanminike, Bangalore, Karnataka, India

Vineela Lagadapati

Assistant Professor, MBA, St. Martins Engineering College, Hyderabad, Telangana, India

Khushboo Malik

Assistant Professor, School of Law, CHRIST (Deemed to be University), Delhi–NCR Campus, India

Piyush Mehta

Associate Professor, United World School of Business, Karnavati University Uvarsad, Gandhinagar, Gujarat, India

Sreenivas Mekala

Associate Professor, Information Technology, Sreenidhi Institute of Science and Technology, Hyderabad, Telangana, India

Jitender Mittal

Research Scholar, Computer Science and Engineering, Amity University, Noida, Uttar Pradesh, India

Venkanna Mood

Associate Professor, ECE, St. Martin's Engineering College, Secunderabad, Telangana, India

M. Nithya

Computer Science and Engineering (5 Year Integrated), Sri Sairam Engineering College, Chennai, Tamil Nadu, India

M. Z. M. Nomani

Professor, Faculty of Law, Aligarh Muslim University, Aligarh, Uttar Pradesh, India

A. Prabhu

Assistant Professor, Electrical and Electronics Engineering, K. Ramakrishnan College of Engineering, Trichy, Tamil Nadu, India

K. K. Ramachandran

Director/Professor, Management/Commerce/International Business, Dr. G. R. D. College of Science, Coimbatore, Tamil Nadu

Ananda Ravuri

Senior Software Engineer, Intel Corporation, Hillsboro, Oregon, USA

Baireddy Ravinder Reddy

Assistant Professor, IT, Sreenidhi Institute of Science and Technology, Hyderabad, Telangana, India

R. Revathi

Assistant Professor, Computer Science, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India

Alpana Sharma

Assistant Professor, Commerce, Bhilai Mahila Mahavidyalaya, Bhilai, Chhattisgarh, India

M. K. Sharma

Professor, Department of Mathematics, Chaudhary Charan Singh University, Meerut, Uttar Pradesh, India

Chetan Shelke

Associate Professor, Alliance College of Engineering and Design, Alliance University, Bangalore, Karnataka, India

Ganesh P. Shukla

Assistant Professor, Department of Industrial and Production Engineering, Guru Ghasidas Vishwavidyalaya (A Central University), Bilaspur, Chhattisgarh, India

Ashutosh Singh

Assistant Professor, Institute of Business Management, GLA University, Mathura, Uttar Pradesh, India

Rupesh Roshan Singh

Mittal School of Business, Lovely Professional University, Punjab, India

Shashidhar Sonnad

Professor, Department of Electronics and Communication Engineering, Sharnbasva University, Kalaburagi, Karnataka, India

Diksha Srivastava

Assistant Professor, Department of Biotechnology and Chemical Engineering, Manipal University, Jaipur, Rajasthan, India

Prateek Srivastava

Department of Information Technology, School of Engineering and Technology (UIET), CSJM University, Kanpur, Uttar Pradesh, India

Tamirat Tagesse Takore

Assistant Professor, Department of Electrical and Computer Engineering, Wachemo University, India

Mohit Tiwari

Assistant Professor, Department of Computer Science and Engineering, Bharati Vidyapeeth's College of Engineering, Paschim Vihar, Delhi, India

Tripti Tiwari

Assistant Professor, Department of Management Studies, Bharati Vidyapeeth (Deemed to be University), Institute of Management and Research, Delhi, India

Vidhika Tiwari

Associate Professor, Department of Mechanical Engineering, DPGITM, Gurugram, Haryana, India

Sachin Tripathi

Symbiosis Law School, Symbiosis International (Deemed) University (SIU), WATHODA, Nagpur, Maharashtra, India

Neha Tyagi

Associate Professor, Computer Science and Engineering, Amity University, Noida, Uttar Pradesh, India

R. Udhayakumar

Professor of English, VelTech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, Tamil Nadu, India

Veena Prasad Vemuri

I/C Principal, Commerce/Management, NKES College of Arts Commerce and Science,
Mumbai University, India

R. Vettriselvan

Associate Professor, AMET Business School, Academy of Maritime Education and Training (AMET
University), Chennai, Tamil Nadu, India

Vijaykumar

Assistant Professor, Department of Agricultural Engineering, SRIET, CCS University, Meerut,
Uttar Pradesh, India

Shaileshkumar Wagh

Head and Assistant Professor, Department of Geography, MGSMs, Dadasaheb Dr. Suresh G. Patil
College, Chopda, Jalgaon, Maharashtra, India

P. William

Dean, Research and Development, Department of Information Technology, Sanjivani College of
Engineering, SPPU, Pune, Maharashtra, India

Abu Sarwar Zamani

Assistant Professor, Computer Science, Prince Sattam Bin Abdulaziz Alkharj, Riyadh, Saudi Arabia



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Preface

The book presents applications of advances in AI that have made inroads in almost all walks of life, and these applications in diverse areas are in great demand. The intent and content of the book will bring about a paradigm shift in the focus and complexion of not only the world of work but also the development and advancement of tools and technologies for the cause of humanity and the sustenance of human activities. These emerging technologies have made inroads in numerous applications by providing customized solutions. Such AI-powered initiatives are evolving constantly and need to go a long way in meeting the requirements of the world in terms of its outreach, agility, stability, affordability, robustness, infrastructural facilities, etc. These state-of-the-art technologies have a comparative advantage in terms of abysmally low latency, fast response, low bandwidth cost, and improved resilience. This is facilitated by data sources and IoT devices in the close neighborhood of the consumers, unlike cloud computing, wherein there is a globalized facility for data storage, analysis, and insight, while in edge computing, the IoTs are at the network edge near users. This book covers chapters that shed light and enlighten us on how we can use edge, cloud, and mobile computing as technologies of today and complement one another for extensive and widespread application areas by leveraging the promising features of advances in AI. The furtherance in IoTs, networking avenues, parallel computing, 5G, and sound infrastructure in general for machine learning has made it possible to employ edge, cloud, and mobile computing techniques in diverse areas and different walks of life.

This book sheds light on various aspects of computing in terms of trends and advances in edge, cloud, and mobile computing, integrating it with advanced AI, services, privacy and security concerns, algorithms, architectures, reliable AI-powered IoTs, advances in networking, and applications in different frontier areas.

The book is partitioned into three parts. Part I is devoted to Advanced Cloud AI services. Part II is dedicated to Advanced Edge AI services and the problems confronted during the application and execution of computing services. Part III deals with Cloud AI, Edge AI, and Mobile AI

Computing Applications. Part I contains five chapters. Chapter 1 dwells upon the provision of a power trading framework of cloud-edge computing in an artificial intelligence bazaar. Chapter 2 delves into anomaly detection using artificial intelligence in a cloud-fuzzy environment. Chapter 3 sheds light on how holistic resource management based on AI and sustainable cloud computing is done. Chapter 4 deliberates on the application of Artificial Intelligence for the enhancement of privacy and security in a smart environment. Chapter 5 is aimed at discussing the application of Artificial Intelligence in Wireless Sensor Networks through Quality of Services enhancement.

Part II of the book has five chapters. Chapter 6 deals with recent advancements of Artificial Intelligence in Edge Computing. Chapter 7 foresees future innovations in using Edge Computing Infrastructure for IoT applications. Chapter 8 aims to understand the growth of Edge Computing with Artificial Intelligence. Chapter 9 gives a perspective on the role of Machine Learning in the convergence of Artificial Intelligence and Edge Computing. Chapter 10 explains the applications, enabling technologies, and vision of Edge Artificial Intelligence for 6G. Part III is embodied with five topics. Chapter 11 discusses how AI products serve the areas of E-Commerce and the commercial banking sector. Chapter 12 deliberates on how sustainable human resources are managed by using artificial intelligence. Chapter 13 discusses the achievement of sustainable development goals by integrating natural resource management and AI. Chapter 14 deliberates on the use of business models in tandem with AI for the cause of sustainable development. Chapter 15 deals with different requirements for Quality of Service using an Adaptive Network Algorithm.

This volume will set a tone for academicians and researchers who will be guided not only in advancing their knowledge but also in becoming acquainted with the advances in Cloud, Edge, and Mobile computing in congruence with advances in AI, thereby enabling the development of sound technological solutions for the betterment of humanity in a sustainable manner.

Introduction

With the rapidly evolving technology landscape, the fusion of advanced AI services with cloud, edge, and mobile computing is catalyzing a monumental shift in the capabilities and applications of these platforms. This editorial explores the transformative impact of advanced AI across the computing continuum, highlighting its role in reshaping industries, enhancing user experiences, and ushering in a new era of intelligent computing.

In the dynamic landscape of computing, the convergence of advanced AI services with cloud, edge, and mobile platforms is orchestrating a revolution. This editorial delves into the transformative potential and far-reaching implications of infusing artificial intelligence across cloud, edge, and mobile computing applications, reshaping the way we interact with technology and unlocking unprecedented possibilities for innovation.

In the shifting landscape of computing, the integration of advanced AI services with cloud, edge, and mobile platforms marks a revolutionary chapter in technology. This editorial explores the transformative impact of advanced AI across these domains, unraveling the applications, benefits, and challenges that accompany the fusion of artificial intelligence with cloud, edge, and mobile computing.

In the ever-evolving landscape of technology, the convergence of advanced AI services with cloud, edge, and mobile computing is forging a new era of innovation. This editorial embarks on a journey to unravel the transformative impact of advanced AI across these computing domains, exploring their applications, benefits, and the synergy that is reshaping the way we interact with information and technology.

The landscape of technology and the convergence of advanced AI services with cloud, edge, and mobile computing is reshaping the way we interact with digital ecosystems. This editorial explores the transformative impact of advanced AI across these computing domains, unraveling the intricacies of its applications, benefits, and the revolutionary potential it holds for the future.

➤ **The Intelligent Cloud:**

1. **Data-Driven Insights:** Advanced AI services in the cloud are unlocking unparalleled capabilities in data analysis and interpretation. Machine learning algorithms, fueled by vast datasets, empower businesses to extract meaningful insights, predict trends, and make informed decisions. The cloud becomes a reservoir of computational power, offering scalable and intelligent solutions that transcend traditional storage and processing capabilities.
2. **Agile Decision-Making:** Cloud services enriched with Advanced AI introduce agility to decision-making processes. Organizations can respond dynamically to changing market conditions, optimize resource allocation, and gain a competitive edge. The cloud is capable of processing, and analyzing data in real time enhancing adaptability, and setting the stage for more responsive and data-driven business strategies.

➤ **The Edge of Intelligence:**

1. **Real-Time Processing:** The marriage of Advanced AI with edge computing brings intelligence closer to data sources. This decentralized approach enables real-time data processing, reducing latency and ensuring immediate responses. Edge devices equipped with AI capabilities can make informed decisions locally, making them indispensable in applications where low latency is critical, such as autonomous vehicles and industrial automation.
2. **Optimizing Bandwidth:** Edge computing, powered by AI algorithms, optimizes bandwidth usage by processing data locally. This exacerbates the data transmission efficiency and contains load on centralized cloud servers. The synergy between edge and AI technologies is reshaping the landscape of IoT, enabling a more responsive and interconnected network of devices.

➤ **The Intelligent Mobile Experience:**

1. **Intuitive Automation:** Mobile computing has evolved from being a simple utility to a hub of intelligent automation.

Advanced AI services in mobile applications provide users with intuitive experiences, from predictive text and voice recognition to smart assistants. This adaptability enhances user interactions, making mobile devices indispensable in our daily lives.

2. **Enhanced Security:** AI is fortifying the security landscape of mobile applications. Behavioral biometrics, facial recognition, and AI-driven threat detection contribute to robust security measures. The fusion of AI with mobile computing ensures not only enhanced user privacy but also safeguards against emerging cybersecurity threats.

➤ **Intelligent Cloud Computing:**

- 1 **Data Insights and Predictive Analytics:** Advanced AI services in the cloud are amplifying the capabilities of traditional computing. Machine learning algorithms process colossal datasets, extracting meaningful insights and empowering businesses with predictive analytics. From forecasting market trends to optimizing supply chains, cloud-based AI is a catalyst for data-driven decision-making.
2. **Scalability and Customization:** The marriage of cloud computing and advanced AI enables unparalleled scalability. Businesses can effortlessly scale their computational resources based on demand, while AI algorithms customize user experiences. This adaptability fosters agility, allowing organizations to respond dynamically to evolving requirements.

➤ **Empowering Edge Computing:**

- 1 **Decision-Making in Real-Time:** Edge computing, bolstered by advanced AI, facilitates real-time decision-making by processing data closer to its source. This is particularly crucial in scenarios where latency is a constraint, such as autonomous vehicles and industrial automation. Edge AI minimizes delays, enhancing responsiveness and efficiency.
2. **Optimized Resource Utilization:** The fusion of edge computing and AI optimizes resource utilization. Devices at the edge can analyze and filter data locally, reducing the need

for continuous communication with centralized servers. This not only enhances efficiency but also conserves bandwidth, making edge devices more self-reliant.

➤ **Redefining Mobile Computing:**

1. **Intelligent Automation:** Mobile computing experiences are being revolutionized by advanced AI services. From predictive text and voice recognition to personalized recommendations, AI algorithms enhance user interactions, making mobile devices more intuitive and user-friendly.
2. **Enhanced Security Measures:** AI contributes significantly to bolstering security in mobile applications. Facial recognition, behavioral biometrics, and AI-driven threat detection enhance device security, ensuring that sensitive information is safeguarded. Mobile devices are evolving into secure, intelligent companions.

➤ **Applications Transforming Industries:**

1. **Healthcare Revolution:** The integration of AI services across cloud, edge, and mobile computing is reshaping healthcare. From AI-driven diagnostics to personalized treatment plans, the intelligent processing of health data is enhancing patient care. Mobile applications with AI features enable remote monitoring and proactive health management.
2. **Smart Manufacturing and IoT Integration:** Industries are witnessing a paradigm shift with AI services optimizing manufacturing processes. Edge computing, coupled with AI algorithms, facilitates predictive maintenance, quality control, and supply chain efficiency. The Internet of Things (IoT) ecosystem is thriving with interconnected devices making real-time decisions, transforming manufacturing into a smarter and more responsive landscape.
3. **Healthcare Revolution:** The integration of AI services in cloud, edge, and mobile computing is reshaping healthcare. From remote patient monitoring to AI-driven diagnostics, these technologies enhance patient care, streamline operations, and contribute to the evolution of personalized medicine.

4. **Industry 4.0 in Action:** Manufacturing is witnessing a revolution with AI services optimizing processes. Edge computing, paired with AI algorithms, facilitates predictive maintenance, quality control, and supply chain efficiency. The Industrial Internet of Things (IIoT) is flourishing, driving Industry 4.0 transformations.

➤ **Cloud Computing Elevated:**

1. **Data Insights Beyond Imagination:** Advanced AI in cloud computing isn't merely a technological leap; it's a paradigm shift. Cloud services are no longer confined to data storage; they have evolved into hubs of intelligent data processing. Machine learning algorithms, fueled by vast datasets, empower businesses with unparalleled insights, predictive analytics, and strategic decision-making capabilities.
2. **Agility Redefined:** The integration of Advanced AI services in the cloud enhances organizational agility. Businesses can dynamically adapt to evolving market trends and customer preferences. Whether it's scaling operations or optimizing resources, the fusion of AI with cloud computing provides unprecedented flexibility, fostering a new era of responsive and adaptive enterprises.

➤ **Mobile Computing Redefined:**

1. **Intelligence in Every Hand:** The infusion of Advanced AI in mobile computing is transforming smartphones into intelligent companions. From voice recognition to natural language processing, mobile devices are becoming adept at understanding and anticipating user needs. Personalized experiences, intelligent automation, and enhanced security are becoming standard features.
2. **Empowering Applications:** AI in mobile applications extends beyond personal assistants. From healthcare apps offering diagnostics to productivity tools with predictive features, mobile applications are leveraging AI to enhance user experiences. This trend is not just about convenience;

it's about tailoring applications to individual preferences and needs.

➤ **Challenges and Ethical Considerations:**

1. **Ethical AI Development:** As AI services become increasingly sophisticated, ethical considerations surrounding their development and deployment become paramount. Ensuring fairness, transparency, and unbiased decision-making is imperative to build trust among users and stakeholders.
2. **Security in the Age of AI:** The integration of AI in cloud, edge, and mobile computing introduces new security challenges. Protecting AI models, and data privacy, and guarding against adversarial attacks demand robust security measures to maintain the integrity and trustworthiness of AI-powered applications.
3. **Ethical AI Development:** As AI services become more sophisticated, ethical considerations become imperative. Ensuring fairness, transparency, and unbiased decision-making are critical to building trust among users and stakeholders.
4. **Security in the Age of AI:** The integration of AI in cloud, edge, and mobile computing introduces new security challenges. Protecting AI models, and data privacy, and guarding against adversarial attacks demand robust security measures to maintain the integrity and trustworthiness of AI-powered applications.
5. **Ethical AI Development:** The ascent of AI services demands ethical considerations. Fairness, transparency, and accountability must be embedded in AI systems to foster trust. Striking the balance between AI innovation and ethics ensures responsible innovation and mitigates unintended consequences.
6. **Security Imperatives:** AI integration introduces new security challenges. Safeguarding AI models, protecting sensitive data, and defending against adversarial attacks are critical aspects. A robust security framework is essential to instill confidence in AI-powered applications.

7. **Ethical AI Integration:** As AI services become increasingly sophisticated, ethical considerations become paramount. Ensuring fairness, transparency, and unbiased decision-making are essential to building trust among users and stakeholders.
 8. **Security in the Age of AI:** The integration of AI introduces new security challenges. Protecting AI models, ensuring data privacy, and guarding against adversarial attacks demand robust security measures to maintain the integrity and trustworthiness of AI-powered applications.
 9. **Ethical AI Deployment:** As Advanced AI becomes increasingly integrated into cloud, edge, and mobile computing, addressing ethical considerations is paramount. Striking a balance between innovation and ethical responsibility involves ensuring transparency, fairness, and unbiased decision-making in AI models.
 10. **Data Privacy and Security:** The proliferation of AI services necessitates robust measures to protect data privacy and security. Strengthening encryption, implementing secure protocols, and fostering a culture of responsible data handling are critical to maintaining trust in AI-driven applications.
- **Cloud Computing: Powering Intelligent Scalability:**
1. **AI-Driven Data Analytics:** Advanced AI is injecting unparalleled intelligence into cloud computing. The ability to process vast datasets and derive meaningful insights is not just enhancing data analytics but is fundamentally altering how businesses make informed decisions. The cloud, powered by AI algorithms, becomes a hub for predictive analytics, trend forecasting, and data-driven strategies.
 2. **Scalable Machine Learning:** Cloud services are evolving beyond conventional storage and processing capabilities. With Advanced AI, machine learning algorithms become scalable, enabling businesses to seamlessly integrate intelligent decision-making processes into their operations. The cloud becomes a playground for experimentation and innovation, fostering a culture of continuous improvement.

- 3. Personalized Cloud Experiences:** Advanced AI brings personalization to the forefront of cloud computing. From tailored user interfaces to dynamic content recommendations, the cloud adapts to individual preferences. This personal touch enhances user experiences, making cloud applications more intuitive, engaging, and responsive.

➤ **Edge Computing: Real-Time Intelligence at the Fringe:**

- 1. Real-Time Decision-Making:** Edge computing, fueled by Advanced AI, empowers devices to make decisions closer to the data source. This is a game-changer in scenarios where low latency is critical, such as autonomous vehicles, smart cities, and industrial automation. Real-time processing at the edge enhances responsiveness and reduces dependence on centralized systems.
- 2. Optimized Resource Utilization:** The fusion of edge computing and AI optimizes resource utilization. Devices at the edge process, and filter data locally, transmitting relevant data to the cloud. This not only reduces bandwidth usage but also enhances efficiency, making edge devices more autonomous and responsive.
- 3. Edge AI for IoT Revolution:** The Internet of Things (IoT) benefits significantly from the marriage of Edge AI. Smart devices equipped with AI capabilities can analyze and act on data locally, reducing dependence on centralized cloud resources. Edge AI is at the forefront of revolutionizing IoT applications, from smart homes to industrial IoT.

➤ **Mobile Computing: Intelligence in Your Palm:**

- 1. Smart Automation:** Advanced AI services are transforming mobile computing by introducing smart automation. From predictive text and voice recognition to virtual assistants, mobile devices become intuitive and adaptive, simplifying tasks for users and enhancing overall user experiences.
- 2. Enhanced Security:** AI contributes to bolstering security measures in mobile applications. Behavioral biometrics,

facial recognition, and AI-driven threat detection enhance device security. Mobile devices equipped with AI capabilities offer a secure and seamless digital experience.

3. **Personalized Recommendations:** Mobile applications leverage AI algorithms to provide personalized experiences. Whether it's content suggestions, tailored recommendations, or personalized fitness plans, AI enhances user engagement by delivering experiences that align with individual preferences.

➤ **The Power of Advanced AI in the Cloud:**

1. **Intelligent Data Processing:** Advanced AI services in the cloud are elevating data processing to unprecedented levels. Machine learning algorithms, fueled by vast datasets, empower businesses to derive actionable insights, predict trends, and make informed decisions. The cloud becomes a powerhouse for computation, enabling scalability and agility in data analytics.
2. **Predictive Analytics and Machine Learning:** The integration of AI in the cloud facilitates predictive analytics, ensuring businesses predict trends, and optimize strategies. Machine learning models deployed on cloud platforms continuously learn from data, providing valuable predictions and recommendations. This enhances decision-making processes across diverse industries.
3. **Personalized User Experiences:** Cloud-based applications, enriched by advanced AI, offer personalized user experiences. From content recommendations to tailored services, the cloud leverages machine learning algorithms to understand and anticipate user preferences. This personalization fosters higher user engagement and satisfaction.
4. **IoT Empowerment:** Edge computing, empowered by Advanced AI, is a game-changer for the Internet of Things (IoT). Smart devices equipped with AI capabilities can process data locally, reducing dependence on centralized cloud resources. This results in faster and more efficient decision-making within the IoT ecosystem.

➤ **AI's Influence on Mobile Computing:**

- 1 **Intelligent Automation:** Mobile computing experiences are undergoing a paradigm shift with the infusion of Advanced AI. Smartphones equipped with AI-driven features offer intelligent automation, such as predictive typing, voice recognition, and contextual recommendations, making devices more intuitive and user-friendly.
2. **Enhanced Security Measures:** AI contributes to robust security in mobile applications. From biometric authentication to behavior analysis, AI algorithms fortify mobile devices against potential threats. This enhances data security and safeguards user privacy, ensuring a trustworthy mobile computing environment.
3. **Augmented Reality and Immersive Experiences:** AI services on mobile devices enhance augmented reality (AR) experiences. Computer vision and machine learning enable real-time recognition and interaction with the surrounding environment, delivering immersive AR applications. This extends to gaming, navigation, and interactive content creation.

➤ **Conclusion:**

The fusion of advanced AI services with cloud, edge, and mobile computing is propelling us into a new era of intelligent computing. As these technologies evolve, addressing challenges and ethical considerations will be pivotal. Responsible development practices, collaboration between industry stakeholders, and ongoing research will shape the future landscape where AI-driven applications seamlessly enhance our daily lives, transforming how we work, connect, and innovate. The journey into this intelligent era promises not just efficiency and convenience but a fundamental shift in the way we perceive and interact with the digital world.

The convergence of advanced AI services with cloud, edge, and mobile computing is propelling us into an era where intelligence is seamlessly integrated into every facet of our digital experiences. From redefining how businesses operate to enhancing

the capabilities of everyday devices, advanced AI is leaving an indelible mark on the technological landscape. As we navigate this intelligent horizon, it is of vital importance to observe ethical considerations, ensure robust security measures, and foster collaborative efforts between technology experts, businesses, and regulators. The fusion of advanced AI with computing platforms holds the key to unlocking unprecedented possibilities, creating a future where intelligence is not just a feature but an integral fabric of our digital existence.

The amalgamation of advanced AI services with cloud, edge, and mobile computing signifies a pinnacle in intelligent computing. The synergistic collaboration across these platforms is not merely a technological evolution but a paradigm shift in harnessing computing power. As we navigate the challenges and ethical considerations, the promise of a future where AI augments human capabilities, optimizes processes, and transforms industries is becoming a tangible reality. The fusion of advanced AI services with cloud, edge, and mobile computing is propelling us into an era where intelligent, adaptive computing reshapes the very fabric of our digital experiences. The journey has just begun, and the possibilities are limitless as we stride into a future shaped by the profound impact of AI-infused computing.

The fusion of advanced AI services with cloud, edge, and mobile computing is a transformative force, reshaping industries, redefining user experiences, and propelling us into a future where intelligent computing is the norm. As we embrace the opportunities presented by these technologies, it is imperative to navigate ethical considerations and security challenges. By fostering collaboration between technology experts, businesses, and regulators, we can harness the full potential of advanced AI, creating a future where intelligent computing elevates our capabilities and reshapes the way we live and work. The horizon is intelligent, and the journey has just begun.

The fusion of advanced AI services with cloud, edge, and mobile computing is not just a technological evolution; it's a paradigm shift. As these technologies continue to interweave, the innovation potential becomes boundless. Navigating ethical considerations

and addressing security challenges will be crucial, but the promise of a more intelligent and responsive digital ecosystem is within reach. The future, shaped by the seamless integration of advanced AI across cloud, edge, and mobile computing, holds the key to unlocking unprecedented possibilities and transforming the way we live, work, and interact with the digital world.

PART I

**Computational Intelligence:
Edge AI Services**



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

CHAPTER 1

Power Trading Framework of Cloud-Edge Computing in an Artificial Intelligence Bazaar

SREENIVAS MEKALA,¹ ANANDA RAVURI,² K. SURESH KUMAR,³
GIOVANNA GUTIÉRREZ GAYOSO,⁴ and SANDEEP GUPTA⁵

¹*Information Technology, Sreenidhi Institute of Science and Technology, Hyderabad, Telangana, India*

²*Intel Corporation Hillsboro, Oregon, USA*

³*MBA Department, Panimalar Engineering College, Varadarajapuram, Poonamallee, Chennai, Tamil Nadu, India*

⁴*Universidad Andina del Cusco, Universidad Andina del Cusco, Peru*

⁵*Department of Electrical Engineering, Graphic Era (Deemed to be University), Dehradun, Uttarakhand, India*

ABSTRACT

In the present world, technologies are spurred by the rapid expansion of Internet of Things and the flourishing progress in deep learning technology; we are witnessing a continuous proliferation of applications of Artificial Intelligence. Simultaneously, advancements in established computing paradigms, namely edge computing and cloud computing, offers range of computing solution to meet the ever-growing demands for widespread AI services. However, several complex issues persist within these computing frameworks that include the inefficient utilization of computing resources, self-interested behavior in power trading computing mechanisms as well

Advances in AI for Cloud, Edge, and Mobile Computing Applications.

Shrikaant Kulkarni and P. William (Eds.)

© 2025 Apple Academic Press, Inc. Co-published with CRC Press (Taylor & Francis)

as inefficiency in management of AI services. To address these challenges, we introduce a blockchain-based framework for trading computing power, which we refer to as AI-Bazaar. Multiple roles are assumed with AI consumers in AI bazaar and have the flexibility for both services of AI and mining using blockchain technology. Consequently, the study formulates the problem of trading computing resources as a Stackelberg game. Leveraging the WoLF principle (Win or Learn), we have devised a PB-MARL algorithm to seek equilibrium within AI-Bazaar, ensuring equitable profits for both AI consumers and CPPs. We have conducted numerical simulations to illustrate the effectiveness of the proposed framework.

1.1 INTRODUCTION

In the present world, technologies are spurred by the rapid expansion of Internet of Things and the flourishing progress in deep learning technology; we have witnessed a surge in the emergence of artificial intelligence (AI) applications [1–3]. Concurrently, advancements in established computing paradigms, namely cloud computing and edge computing, have been able to meet the growing demands for number of applications of AI. These advances offer robustness in computing power, high bandwidth, low latency thereby enabling the services of AI to be available ubiquitously [4–6]. Consequently, the widespread availability of services of AI has that gives rise to a new computing energy trading framework that necessitates seamless efficient and connectivity computing power trading [7–10].

In this new computing framework, computing nodes are distributed extensively, and computing power resources exhibit heterogeneity. The edge nodes, in particular, grapple with constraints stemming from limited resources of computing, which currently cannot adequately support computationally intensive services of AI.

Meanwhile, cloud nodes offer a solution that are profitable for sharing power trading for computing in between the edge nodes, specifically the consumers of AI [11, 12]. To unlock the potential of different computing nodes and facilitate the exchange of computing resources, there is an urgent need to establish a trusted framework for trading and sharing of computing power. Within this framework, the computing-power provider (CPP) consolidates computing resources from various nodes and makes them available to computing nodes constrained of resources that enable wide range of services of AI that meets the operational needs.

However, there are significant challenges within this computing-power trading and sharing framework. For instance:

The resources of computing power within the computing power provider are not completely optimized [13]. While others may be working normally or even be idle, some processing of nodes that assigned for consumers of AI that do not have enough resources to fulfill their progressively complicated computational requirements. This situation results in underutilized computing power and exacerbates the disparity in computing resource allocation.

The energy trading lacks an equitable profit-sharing mechanism. Mostly, the CPP tends to prioritize its self-interest and focuses solely on its own gains. As the number and diversity of consumers of AI continue to grow, the power trading computing framework is evolving into a new era characterized by economy of free market. Consequently, it becomes imperative to address the self-centeredness and monopolistic tendencies of the CPP and establish a profit-sharing mechanism that ensures fairness.

1.2 LITERATURE REVIEW

Presently, the quality management for services that is inefficient in AI. Alongside the proliferation of AI services, there has been a constant and rapid increase in computing demands. However, due to the computing environment constrained of resources, the existing framework falls short in meeting the precise, personalized, and requirement of high-quality service.

In essence, these aforementioned challenges have intensified the pressing necessity For development of power trading framework that is fully optimized, fair in profit distribution, and operationally efficient to support widespread AI services.

Significantly, within the computing power trading framework, the CPP offers paid resource services catering to compute-intensive AI applications. Blockchain technology, known for its immutability, accessibility, and tamper-proof nature, offers several advantages for trading computing resources. These benefits include recording the distributed computing power, facilitating contract compliance in computing transactions, and establishing pricing mechanisms without the need for trusted intermediaries [14, 15]. In particular, with advancements in blockchain and consensus mechanism has become well-suited for AI applications. For

instance, ongoing research explores consensus protocols based on proof of deep leaning (PoDL), PoL, PoQ, and more [16–19]. The protocols aim to connect the entire learning process in a decentralized blockchain manner to build a streamlined intelligent system and replace the conventional Proof of Work hashing challenge with assignments for training the model. These protocols help resource-constrained computing nodes learn more effectively while addressing the problem of wasted computing power. In order to improve the quality of AI services, it becomes sense to incorporate blockchain-based consensus protocols for proofs of valuable work into the computing-power trade framework.

In our earlier research [20], we created a power trading networking paradigm for ubiquitous AI by taking into account the networking, CPP, and AI consumer perspectives. This framework, however, lacked a thorough trading system for processing power. In this paper, we describe the AI-Bazaar computer trading framework, which leverages the blockchain's proof of useful work consensus process, known as the PoL consensus protocol. AI-bazaar is designed to address the three issues previously mentioned.

Specifically, Consumers of AI assumed with multiple roles that they have the freedom for contribution of computing energy that they have provided from CPP for services of AI and blockchain mining, effectively resolving the first as well as third challenges. Additionally, formulated the problem of computing trading using Stackelberg game within AI-Bazaar to ensure a fair distribution of interests, thereby addressing the second issue.

This work contributes in the following ways:

1. The study proposed power treating framework for ubiquitous services of AI based on blockchain, featuring a multi-role capability.
2. To achieve a balance in profits among consumers of AI and CPP. The study formulates the computing trading problem using Stackelberg game for single leader multiple followers. We also provide proof of the existence and uniqueness of the AI bazaar equilibrium and Nash Equilibrium at each stage of Stackelberg game.
3. Utilizing the Principle of win or learn fast, the study designs PB-MARL algorithm for assisting AI consumers and CPP for finding profits on equitable ratio. The findings of the experiment confirm the effectiveness of proposed power trading framework.

1.3 METHODOLOGY

1.3.1 FRAMEWORK DISTRIBUTION

The research study introduces AI-Bazaar, a power trading framework characterized by its multi-role capability allowing consumers to freely choose to be miners, AI service providers, and features balanced with profit ensuring satisfactory computing profits for all parties involved in services of AI.

As illustrated in Figure 1.1, AI-Bazaar is dedicated to achieving a balance in computing power allocation, linking dispersed computing resources, incentivizing computing nodes to share their resources, and accommodating the diverse customized needs of consumers of AI. Within AI-Bazaar, there are two primary entities: AI consumers and CPP. Specifically, resource-limited AI consumers leverage blockchain to request computing power from CPP to meet their specific requirements. To provide a more detailed overview, the framework of AI Bazaar comprises of different layers.

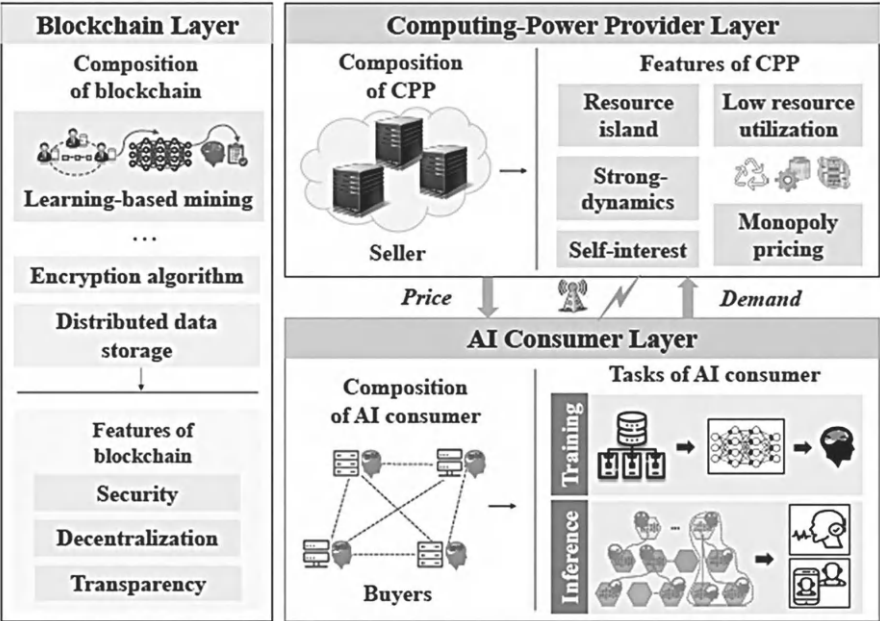


FIGURE 1.1 Proposed framework of AI bazaar.

There is a community in consumers of AI in AI Bazaar that are, represented by “i” belonging to the set N , as depicted in the layers for consumers of AI in Figure 1.1. These AI consumers primarily engage in two categories of tasks:

1. **AI Training:** AI consumers require substantial computational resources to facilitate AI training services for augmented reality, speech recognition, transportation control, and natural language processing. These services are powered by neural networks (NNs), federated learning (FL) algorithms, reinforcement learning (RL), and comparable technologies.
2. **AI Inference:** AI consumers also need computing power for AI inference. This involves processing tasks using the knowledge acquired during training to manage real-world applications. AI inference retains the ability to learn and can be used to quickly respond to fresh datasets.

To fulfill these AI demands, acquiring sufficient bandwidth is also essential within AI-Bazaar. However, this bandwidth provisioning is typically handled by the network operator and falls outside the scope of the framework for power trading. Now, let’s introduce the computing-power provider layer:

1.3.2 POWER PROVIDER LAYER

With the rapid growth of the IoT and the emergence of novel computing frameworks, the distribution of power has shifted from cloud’s centralized nodes to network nodes [21–23]. In general, computing nodes within these frameworks are widely dispersed, and the power computing resources they offer exhibit heterogeneity. The role of the CPP is to consolidate computing resources from these distributed nodes through resource pooling [24, 25]. This consolidation provides a cost-effective solution to address the resource limitations of edge devices, which are the AI consumers [26, 27].

Consequently, establishing a trustworthy framework of power trading is crucial to enable reliable resource sharing between CPP and AI consumers. Blockchain technology, with its openness, cryptographic security, and

decentralized nature, offers a range of solutions pertinent to computing, leading us to the introduction of the blockchain layer.

Blockchain incorporates a range of technologies, encompassing consensus mechanisms, encryption algorithms, distributed data storage, and more [28, 29]. These technologies collectively bestow upon blockchain the attributes of decentralization, transparency, security, and other characteristics [30]. These qualities are pivotal in establishing a secure and trustworthy framework of power trading [31, 32]. The PoL-based consensus mechanism is specifically used in this study as a response to the enormous processing power waste brought on by work-based protocols. Instead of producing random hashes, PoL treats the training of NNs as a meaningful computational process, thereby acting as a learning-based mining service [33]. With the help of this innovative method, AI users can play numerous roles inside the framework of power trade, such as being a miner or an AI service provider [20].

The multi-functionality of the NNs in AI-Bazaar is also clear from Figure 1.1. They can be seen in the learning-based blockchain as pieces of a puzzle on the one hand. On the other hand, they act as the deep learning (DL) algorithms' underlying support. Therefore, this multipurpose Neutral network strengthens the AI services for the consumers.

1.3.3 PROBLEM FORMULATION

The problem formulation for the given study is given below.

1.3.3.1 DESCRIPTION OF AI MODEL

The AI-Bazaar model is depicted in Figure 1.2. It's worth noting that various members with different computing power resources in the framework employ distinct strategies, which can be categorized into the procurement strategies adopted by AI consumers and the pricing strategies implemented by the CPP. In AI-Bazaar, the CPP primarily focuses on providing computing power whereas consumers of AI Bazaar submit requests to rent power trading from the CPP for the execution of diverse tasks and the operation of blockchain applications for AI services.



FIGURE 1.2 AI bazaar model.

Assuming the presence of CPP and NAI consumers for computing-power, we model the computing trade dynamics between the consumers of AI and CPP as the multiple followers of Stackelberg game. All AI consumers will initially pay the same unit price for processing power under the CPP. Each AI consumer then develops their own buying plan to obtain processing resources to handle a variety of activities. This computing power's standard unit price is denoted as "p" and falls within the range $[p_{\min}, p_{\max}]$, where $p_{\min}$ represents the minimum unit price, and $p_{\max}$ represents the maximum unit price for computing power.

Within AI-Bazaar, AI consumers are involved in both AI Inference and AI training tasks, leading to the dual roles. When the puzzle for mining is solved successfully, the tasks for training AI consumers are expedited. In such instances, the AI consumer assumes the role of a miner, vying for the privilege of bookkeeping within the blockchain and earning the associated block reward. The AI consumer can then act as an AI service provider, meeting each client's specific needs, if the CPP gives them the training results they need for inference. The ability to switch between various jobs is another advantage that AI consumers have. For example, they can divide their computing resources between meeting business demands and participating in blockchain mining.

In this study, we employ a consensus protocol rooted in learning, specifically, the PoL [17]. According to PoL, a neural network with a function for small loss can effectively supply AI services of high quality by treating the local training process as a puzzle [35]. The participant who successfully solves this puzzle becomes the winner and is rewarded with the issuance of a block [36, 37]. Once a consensus is reached, other nodes in the network adjust their NN parameters according to transaction of winner [38]. To execute the process of PoL among AI consumers, we employ a standardized dataset consisting solely of normal data. The model is only allowed to go live if its performance on this standardized dataset meets the required criteria, which also acts as a safeguard against poisoning attacks.

Let's assume that consumer "i" decides to allocate a portion of their computing power, denoted as β_i , for mining, which we refer to as the ratio for role playing. Consequently, the fraction of the computing power of AI consumers dedicated for mining within the entire blockchain network is represented as α_i , calculated as follows:

$$\alpha_i = \beta_i F_i / \sum_{j=1 \text{ to } N} \beta_j F_j \quad (1.1)$$

Similar to the concept of proof of work (PoW) [30], The likelihood of completing the learning-based challenge is taken into account as a random variable with a Poisson distribution. In a similar vein, the likelihood that miner "i" will solve this puzzle and reach consensus is indicated by the symbol ρ_i and can be represented as:

$$\rho_i = \alpha_i e^{\hat{\epsilon}} (-\epsilon T p_{-i}) \quad (1.2)$$

where ϵ is a constant parameter connected to the allocated training time (TTrain), with the result that $\hat{\epsilon} = 1/TTrain$. We make the assumption that each block has an equal number of transactions, indicated by the symbol $B_i = B$, without losing generality. Additionally, we use miner "i" 's trained NNs' loss function value as the evaluation indicator, indicated by the symbol i .

Upon successfully solving the learning-based puzzle, the winning miner broadcasts their solution over the entire network, while other nodes verify its accuracy and reach a consensus. The award goes to the first miner who successfully acquires the bookkeeping right. Miners can earn rewards through two avenues: one from successful mining, denoted as R , and the other from performance rewards, denoted as R_p , which is calculated is the result of the block size (B) and the performance reward factor $()$, or $R_p = B$. As a result, AI consumers get profited through mining as given below:

$$Um\_i = (R + \lambda B) \alpha i e^{\hat{}} (-\varepsilon \tau \delta B) \quad (1.3)$$

1.3.4 PB-MARL ALGORITHM

We use a two-stage Stackelberg game to describe the exchange of computing power among computing-power members. The following definitions provide the profit functions for each stage:

The profit made by the CPP in the AI-Bazaar is calculated by deducting the service cost from the payments made by Consumers of AI, presuming members for computing and power trading have quasi-linear utilities [39]. Based on the needs put forth by AI users, the CPP develops its computing-power pricing strategy in the first step within the strategy space $P = p \mid p_{\min} \leq p \leq p_{\max}$, with the goal of maximizing its profit U_{cpp} . This can be expressed as:

$$P1: \text{maximize } p \sum (p - C) F_i, I = 1 \text{ to } N \text{ subject to } p \in [p_{\min}, p_{\max}] \quad (1.4)$$

Here, C represents the price of electricity per computational unit. Self-interested CPPs are encouraged to participate in AI-Bazaar and use idle computer capacity for AI services by this revenue-incentive mechanism.

Additionally, the incentives from blockchain, mining risk costs, and payments for renting computing power from the CPP all contribute to the profit AI consumers make in AI-Bazaar. Based on the unit price of computing established in the first step, AI customers choose how much computational power to acquire within the purchase strategy space $F = F_i \mid i \in N: F_{\min} \leq F_i \leq F_{\max}$. A multi-role strategy can be used by consumer I to increase their personal profit. U_i , which is expressed as:

$$P2: \text{maximize } F_i F_i ((1 - \beta_i) C_b - u_i p) + m_i U_m \mid \text{subject to } F_i \in [F_{\min}, F_{\max}] \quad (1.5)$$

Here, U_m represents m_i is the parameter for effect of token expressing the monetary equivalent of mining, and u_i is a weight factor that indicates the significance of the payment cost in the overall profit for the i^{th} AI consumer. C_b stands for the profit per computer power unit produced by meeting business needs.

To summarize, the goals of both CPP and AI consumers are to maximize profits [40–43]. In the analysis that follows, we'll look at the equilibrium characteristics and trade-off factors relating to both parties' profits.

TABLE 1.1 PB-MARL Algorithm

- Step 1: Initialization
- Step 1: S_{cpp}^t state is observed
- Step 2: Probability policy π_{cpp} for selected action P^*
- Step 3: State S_{cpp}^{t+1} and R_{cpp} is observed for next reward
- Step 4: Q_{cpp} is updated (S_{cpp}^t, P^t)
- Step 5: Average policy $\bar{\pi}_{cpp}(S_{cpp}^t, P)$ is updated
- Step 6: Current strategy is updated $\pi_{cpp}(S_{cpp}^t, P)$
- Step 7: End

We have developed the proposed algorithm to acquire optimal policies within a dynamic environment involving multiple agents. This algorithm serves as an extension of Q-learning and is effective in mitigating the issues of severe oscillations often encountered in single-agent reinforcement learning. In practice, the proposed algorithm using WoLF principle.

The RL framework empowers participants to acquire the best strategy through a process of trial and error, utilizing feedback generated by their experiences and actions within an interactive environment. In this study, we represent the states of AI consumers and CPP as $s_u^t = p^t$ and $s_{cpp}^t = [F_t - 1 \ i]$ $i \in \mathbb{N}$, respectively. Here, F_t denotes the action of AI consumers in terms of submitted computing-power purchase, and p^t represents the price policy of power trading determined by the CPP. The action space for Consumers of AI is denoted as A_u , while A_{CPP} represents the action space for the CPP.

For Consumers of AI, upon observing the CPP's actions regarding computing-power unit pricing in time slot t , each AI consumer independently determines its own action $F_t \ i$ based on its current state. In a similar fashion, following equation (5), we define the MDP for Consumers of AI. Here, the state space is represented as $s_u^t = p^t$, the action space is defined as $F_t \in A_u$, and the reward function is expressed as:

$$R_u = F_t ((1 - \beta_i)C_b - u_{ipt}) + m_i(R + \lambda B)\alpha_i e^{-(\epsilon \tau \delta_i B)}. \quad (1.6)$$

1.4 RESULTS AND DISCUSSION

We have developed the PB-MARL algorithm to learn optimal policies in a dynamic environment with multiple agents. The extension of Q-learning

is this algorithm [44] and offers a solution to prevent the severe oscillatory problem that can be encountered in single-agent RL scenarios. Specifically, the PB-MARL algorithm dynamically adjusts the learning parameters δwincpp and $\delta\text{losecpp}$ to update the learning rate δcpp , following the principles of WoLF.

Algorithm 1 provides a detailed breakdown of the PB-MARL algorithm for the CPP within the multi-agent system. In this algorithm, αcpp (in the range of $(0, 1)$) where the learning rate is represented, while γcpp (in the range of $(0, 1)$) represents the discount rate. The initialization of certain parameters initiates the procedure. The Q-function for the unit price of power trading in cloud and edge computing is defined as $Q\text{cpp}$ (stcpp , pt) by taking note of the pertinent state, action, and reward. The Q-table is updated in a manner similar to Q-learning in accordance with the guideline given in step 4. The average policy cppis is then used to assess the current strategy's 'win' or 'failure.' This policy is updated in step 5, with $C(\text{stcpp})$ representing the count of state occurrences.

The CPP can then revise the price policy it has already provided, interacting with the agents and the environment to increase the overall reward.

The proposed power trading framework was implemented using Python 3.7.

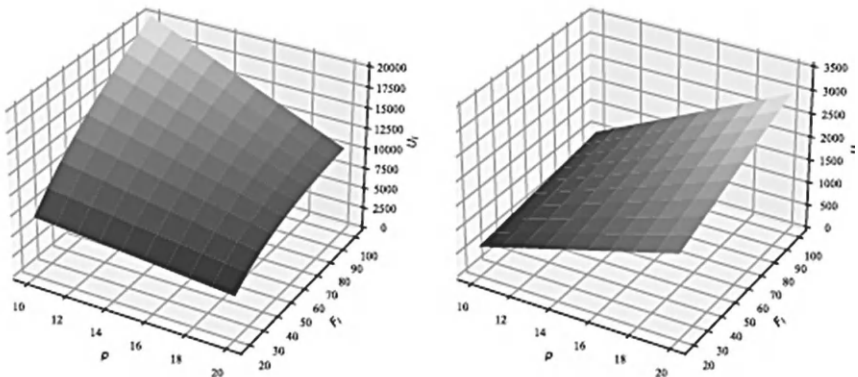


FIGURE 1.3 Profit model for AI bazaar and CPP.

In order to find the equilibrium of AI Bazaar in a non-stationary environment with several participants, we first employed the PB-MARL algorithm. The performance for convergence of three algorithms is shown

in Figure 1.3. These algorithms are PGA-APP (which only considers local rewards when choosing actions), MiniMax Q-learning (which substitutes a “MiniMax” operator determined through the solution of a linear program in standard algorithm of Q-learning), and PB-MARL. The ideal answer according to theory is represented by the red line. The PGA-APP algorithm’s low performance and inability to converge make this clear. On the other hand, both the PB-MARL and MiniMax learning algorithms show convergence to the equilibrium of AI Bazaar, with PB-MARL significantly outperforming MiniMax Q-learning due to the WoLF principle [45, 46].

We plotted the computing-power members’ profit models to get a more in-depth insight of AI-Bazaar. Each member of a unilateral model has its own optimal behavior, regardless of the tactics used by other computing-power members [47, 48]. Consequently, problem P2 focuses on optimizing the consumer’s profit, as depicted in the left part, while to increase the profit for the CPP, solve problem P1.

Here, we evaluated AI-Bazaar’s performance in profit-balancing. Specifically, we used the chart for representation of profit model for consumers of AI in AI Bazaar for indicating profit through graphical representation. The blue chart also shows the revenue generated by applying the PB-MARL method to solve the Stackelberg game [49, 50]. Both sides earn from the PB-MARL algorithm’s successful balancing of the profit between consumers and the CPP. “Furthermore, we compared the profits of computing-power members across four algorithms, namely PB-MARL, PGA-APP, MiniMax Q-Learning and IGA-WOLF.” The proposed algorithm closely approaches the equilibrium of AI Bazaar when compared to the other baseline algorithms.

In our proposal, consumers of AI have flexibility for taking on multiple roles simultaneously. It’s worth considering how the ratio of roles played by an AI consumer impacts their profits [51, 52]. As depicted in Figure 1.4, we analyze the earnings of a single consumer under various block reward conditions.

In Figure 1.4(a), we observe that with a smaller ratio on role-playing, where mining using blockchain technology tends to be more lucrative, the profit of consumers of AI tends to increase. However, as the ratio of role playing (β_i) increases, the profit decreases in consumers of AI. This decrease is attributed to the heightened risk of revenue loss when dedicating a significant portion of computing power to mining. Additionally,

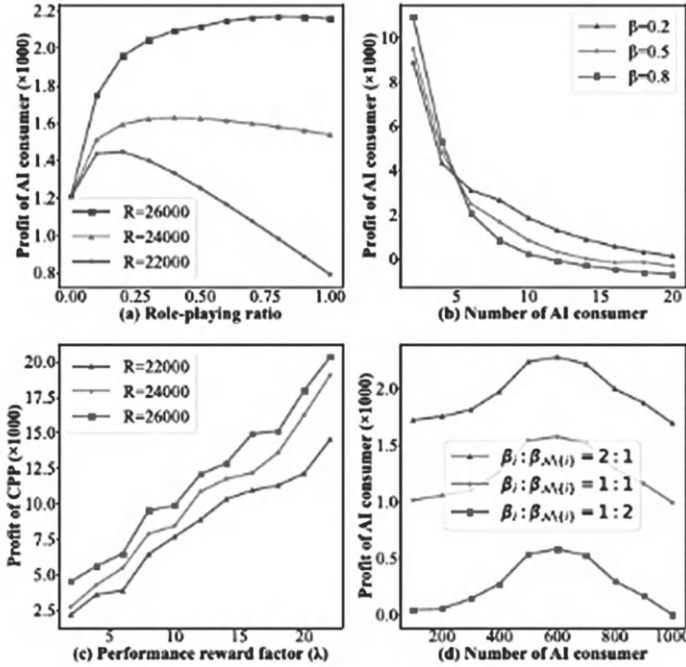


FIGURE 1.4 Comparison of blockchain factors.

lower block rewards result in reduced profits for Consumers of AI, making them more susceptible to the risks associated with a high role-playing ratio.

In Figure 1.4(b), we compare the profits of consumers of AI in various roles as per the number of consumers of AI. For smaller-scale AI consumer groups, an individual consumer can easily obtain mining rewards when they prefer to act as miners. In this scenario, a larger β_i leads to greater rewards from blockchain mining for the consumer.

As more AI users compete for processing power on AI-Bazaar, their chances of successfully solving the PoL puzzle and their exposure to mining risks decrease [53]. Consequently, with a higher number of Consumers of AI, an increase in the role-playing ratio (β_i) leads to diminishing returns for consumers from blockchain activities. Beyond a certain point, it becomes less profitable compared to consumers who prefer mining with a lower computing power allocation. To address this situation involving a large user base, one potential solution is to introduce higher rewards to incentivize participation of more users in blockchain mining.

The performance reward factor's (PRF) influence is depicted in Figure 1.4(c). An increase in the CPP's profit corresponds to a higher, as shown in Figure 1.4(c). This happens because more AI users are motivated by self-interest to buy computing resources when performance rewards are larger. Additionally, a bigger block reward results in greater financial gains for the CPP. This is so because a fixed block reward encourages increased demand for processing power from AI users, which raises the CPP's overall benefits.

Furthermore, as the number of Consumers of AI increases, the utility of individual Consumers of AI may decrease [54]. This occurs because the competition for computer resources increases as more Consumers of AI join AI-Bazaar; it becomes harder for them to complete PoL problems and earn incentives, leading to worse performance. However, as more AI users buy processing resources to support their work, the CPP's usefulness might rise.

In our suggested framework, Consumers of AI are capable of playing numerous roles at once. It is important to think about how the profit of an AI customer is impacted by the role-playing ratio. Figure 1.4 contrasts a single consumer's profit under various block rewards. Figure 1.4(a) demonstrates that the profit of the AI consumer exhibits an upward trend for a low role-playing ratio, where blockchain mining is more lucrative. However, as β_i (the role-playing ratio) rises, the profit of the AI consumer falls. This is due to the larger risk of income loss when mining receives the lion's share of processing power [55]. Additionally, when Consumers of AI select a higher role-playing ratio, lesser block rewards result in lower revenues for them. With more Consumers of AI entering AI-Bazaar, they are less likely to successfully solve PoL puzzles, thus taking on more mining risks. Consequently, with a larger β_i , an AI consumer's returns from the blockchain gradually diminish, eventually becoming less profitable compared to another consumer who prefers mining with less computing power.

To address this situation, especially in a large-scale user environment, we can incentivize more users to participate in blockchain mining by offering higher rewards.

There is no denying the surge in new technologies currently emerging, with blockchain standing out as a decentralized ledger system gaining significant attention for its applicability across various domains [56, 57]. Since its beginnings in the 1920s, blockchain has continued to be

a potentially game-changing technological advancement that has the ability to revolutionize collaboration, automate payments, track markets, and record transactions [58]. It might do away with the requirement for centralized authorities to monitor and certify multiparty agreements and transactions [59]. All mining nodes inside the blockchain process each transaction through cryptographic hashing and verification [60], creating immutable, secure, and time stamped records accessible to all involved parties [61].

Making decisions based on data and enabling robots to learn from it is what AI does, is another rapidly growing field, projected to reach a market value of up to USD 13 trillion by 2030 [62].

While various innovations strive to bolster the security of information in smart homes, there are still vulnerabilities to address. The consensus mechanism in blockchain ensures the permanent recording of all transactions [64], making it practically impossible to execute control attacks on data being sent or stored via a single point of entry. Instead, attackers would need to compromise the majority of network nodes to succeed [65].

The idea of decentralized AI, which combines blockchain and AI technologies, has advanced in recent years [66]. With this method, trustworthy, properly tagged data may be executed and securely stored on the blockchain in a distributed, intermediary-free manner [67]. Blockchains are now recognized as a trusted platform for preserving such data, while AI excels at processing vast datasets [68]. Smart contracts can be programmed into blockchains, allowing reliable third parties to monitor user sharing and data access [69]. Through exposure to autonomous systems, machines, and various scenarios, AI can adapt and learn, leading to precise and reliable outcomes for decision making universally accepted by all mining nodes on the blockchain [70, 71].

This setup instills trust and endorsement from all stakeholders. AI processes on blockchains facilitate decentralized decision-making to enhance trust and security in decision outcomes and data sharing across countless contribution of independent operations, organize as well as vote for future decisions [72].

The fusion of AI with blockchain technology has produced a number of useful benefits [73]. Blockchain guarantees the safekeeping of patient records in the medical industry, where AI can extract valuable insights from this data, aiding medical practitioners. This combined approach has proven valuable in addressing challenges [74]. BurstIQ, a blockchain-based

firm that offers data solutions for the healthcare industry, is a significant example. It provides a health wallet that manages patient information using blockchain technology, AI, and big data, enabling medical professionals to access it as needed [75]. The integration of these technologies is revolutionizing the financial services sector by enhancing mutual trust and transaction speeds [34].

1.5 CONCLUISON

To conclude the research study AI bazaar has been developed using blockchain technology for a power trading framework with the use of wide range of services of AI. Within the Framework of AI bazaar, AI consumers are afforded the versatility to switch between roles as Blockchain miners and service providers of AI. They are also granted the freedom to allocate the computing power they rent from the CPP in accordance with their personalized requirements. This arrangement gives rise to Stackelberg game at two-stage involving the AI consumers and CPP. Additionally, the study has devised PB-MARL algorithm, leveraging the WoLF principle, to identify the equilibrium within AI-Bazaar. This equilibrium enhances the provision of services based on learning within computing environments constrained of resources. Our findings of the experiment underscore the ability of both CPP and AI consumers to make decisions of satisfactory level and reap the benefits. The study aims to further refine the practical implementations of this framework, alongside the development of smart contracts for ensuring the transparency and security of the transactional processes.

KEYWORDS

- **power trading**
- **cloud-computing**
- **edge-computing**
- **artificial intelligence**

REFERENCES

1. Zhang, Z., Xiao, Y., Ma, Z., Xiao, M., Ding, Z., Lei, X., Karagiannidis, G. K., & Fan, P. (2019). 6G wireless networks: Vision, requirements, architecture, and key technologies. *IEEE Vehicular Technology Magazine*, 14(3), 28–41.
2. Li, E., Zeng, L., Zhou, Z., & Chen, X. (2020). Edge AI: On-demand accelerating deep neural network inference via edge computing. *IEEE Transactions on Wireless Communications*, 19(1), 447–457.
3. Wang, J., Jiang, C., Zhang, H., Ren, Y., Chen, K., & Hanzo, L. (2019). Thirty years of machine learning: The road to pareto-optimal next-generation wireless networks. *CoRR*, abs/1902.01946.
4. Wang, X., Han, Y., Wang, C., Zhao, Q., Chen, X., & Chen, M. (2019). In-edge AI: Intelligentizing mobile edge computing, caching, and communication by federated learning. *IEEE Network*, 33(5), 156–165.
5. Shih, Y., Chung, W., Pang, A., Chiu, T., & Wei, H. (2017). Enabling low-latency applications in fog-radio access networks. *IEEE Network*, 31(1), 52–58.
6. Gong, Y., Yao, H., Wang, J., Li, M., & Guo, S. (2022). Edge intelligence-driven joint offloading and resource allocation for future 6G industrial Internet of Things. *IEEE Transactions on Network Science and Engineering*.
7. Shao, G., Chen, X., Zeng, X., & Wang, L. (2020). Deep learning hierarchical representation from heterogeneous flow-level communication data. *IEEE Transactions on Information Forensics and Security*, 15, 1525–1540.
8. Zhou, Z., Chen, X., Li, E., Zeng, L., Luo, K., & Zhang, J. (2019). Edge Intelligence: Paving the last mile of artificial intelligence with edge computing. *Proceedings of the IEEE*, 107(8), 1738–1762.
9. Wang, X., Han, Y., Leung, V. C. M., Niyato, D., Yan, X., & Chen, X. (2020). Convergence of edge computing and deep learning: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, 22(2), 869–904.
10. Wang, F., Yao, H., Zhang, Q., Wang, J., Gao, R., Guo, D., & Guizani, M. (2021). Dynamic distributed multi-path aided load balancing for optical data center networks. *IEEE Transactions on Network and Service Management*.
11. Mai, T., Yao, H., Zhang, N., Xu, L., Guizani, M., & Guo, S. (2021). Cloud mining pool aided blockchain-enabled Internet of Things: An evolutionary game approach. *IEEE Transactions on Cloud Computing*.
12. Li, F., Yao, H., Du, J., Jiang, C., Han, Z., & Liu, Y. (2022). Auction design for edge computation offloading in SDN-based ultra dense networks. *IEEE Transactions on Mobile Computing*, 21(5), 1580–1595.
13. Ren, J., Yu, G., He, Y., & Li, G. Y. (2019). Collaborative cloud and edge computing for latency minimization. *IEEE Transactions on Vehicular Technology*, 68(5), 5031–5044.
14. Xiong, Z., Zhang, Y., Niyato, D., Wang, P., & Han, Z. (2018). When mobile blockchain meets edge computing. *IEEE Communications Magazine*, 56(8), 33–39.
15. Li, J., Liu, T., Niyato, D., Wang, P., Li, J., & Han, Z. (2021). Contract-theoretic pricing for security deposits in sharded blockchain with Internet of Things (IoT). *IEEE Internet of Things Journal*, 8(12), 10052–10070.

16. Liu, Y., Yu, F. R., Li, X., Ji, H., & Leung, V. C. M. (2020). Blockchain and machine learning for communications and networking systems. *IEEE Communications Surveys & Tutorials*, 22(2), 1392–1431.
17. Qiu, C., Yao, H., Wang, X., Zhang, N., Yu, F. R., & Niyato, D. (2020). AI-Chain: Blockchain energized edge intelligence for beyond 5G networks. *IEEE Network*, 34(6), 62–69.
18. Lu, Y., Huang, X., Dai, Y., Maharjan, S., & Zhang, Y. (2020). Blockchain and federated learning for privacy-preserved data sharing in industrial IoT. *IEEE Transactions on Industrial Informatics*, 16(6), 4177–4186.
19. Chenli, C., Li, B., Shi, Y., & Jung, T. (2019). Energy-recycling blockchain with proof-of-deep-learning. In *IEEE International Conference on Blockchain and Cryptocurrency* (pp. 19–23).
20. Wang, X., Ren, X., Qiu, C., Cao, Y., Taleb, T., & Leung, V. C. M. (2021). Net-in-AI: A computing-power networking framework with adaptability, flexibility, and profitability for ubiquitous AI. *IEEE Network*, 35(1), 280–288.
21. Chalapathi, G. S. S., Chamola, V., Vaish, A., & Buyya, R. (2021). Industrial Internet of Things (IIoT) applications of edge and fog computing: A review and future directions. In *Fog/Edge Computing for Security, Privacy, and Applications* (pp. 293–325). Springer.
22. Kumar, A., Sharma, S., Goyal, N., Gupta, S. K., Kumari, S., & Kumar, S. (2022). Energy-efficient fog computing in Internet of Things based on routing protocol for low-power and lossy network with Contiki. *International Journal of Communication Systems*, 35(4), e5049.
23. Zhang, C. (2020). Design and application of fog computing and Internet of Things service platform for smart city. *Future Generation Computer Systems*, 112, 630–640.
24. Mei, J., Wei, M., Sun, Y., Li, J., Zhou, G., & Zhang, X. (2023). The architecture of computing power network towards federated learning: Paradigms and perspectives. In *2023 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)* (pp. 1–6). IEEE.
25. Zhao, Q., Wei, M., Xie, Y., & Lei, B. (2023). Communication and computing integration: Characteristic of computing power network towards 6G. In *2023 IEEE/CIC International Conference on Communications in China (ICCC Workshops)* (pp. 1–6). IEEE.
26. Ren, X., Qiu, C., Wang, X., Han, Z., Xu, K., Yao, H., & Guo, S. (2022). AI-Bazaar: A cloud-edge computing power trading framework for ubiquitous AI services. *IEEE Transactions on Cloud Computing*.
27. Zaman, K., Zhaoyun, S., Shah, B., Hussain, A., Hussain, T., Khan, U. S., & Sarra, B. (2023). Efficient power management optimization based on whale optimization algorithm and enhanced differential evolution. *Alexandria Engineering Journal*, 79, 652–670.
28. Yadav, A. S., Singh, N., & Kushwaha, D. S. (2023). Evolution of blockchain and consensus mechanisms & its real-world applications. *Multimedia Tools and Applications*, 1–46.
29. Wenhua, Z., Qamar, F., Abdali, T. A. N., Hassan, R., Jafri, S. T. A., & Nguyen, Q. N. (2023). Blockchain technology: Security issues, healthcare applications, challenges and future trends. *Electronics*, 12(3), 546.

30. Mukherjee, P., & Pradhan, C. (2021). Blockchain 1.0 to blockchain 4.0—The evolutionary transformation of blockchain technology. In *Blockchain Technology: Applications and Challenges* (pp. 29–49). Cham: Springer International Publishing.
31. Li, C., Yuan, Y., & Wang, F. Y. (2022). A novel framework for data trading markets based on blockchain-enabled federated learning. In *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)* (pp. 3392–3397). IEEE.
32. Zhang, P., Ding, S., & Zhao, Q. (2023). Exploiting blockchain to make AI trustworthy: A software development lifecycle view. *ACM Computing Surveys*.
33. Qiu, C., Wang, X., Yao, H., Du, J., Yu, F. R., & Guo, S. (2020). Networking integrated cloud–edge–end in IoT: A blockchain-assisted collective Q-learning approach. *IEEE Internet of Things Journal*, 8(16), 12694–12704.
34. Soleymani, F., & Paquet, E. (2020). Financial portfolio optimization with online deep reinforcement learning and restricted stacked autoencoder—DeepBreath. *Expert Systems with Applications*, 156, 113456.
35. Wang, X., Ren, X., Qiu, C., Xiong, Z., Yao, H., & Leung, V. C. (2022). Integrating edge intelligence and blockchain: What, why, and how. *IEEE Communications Surveys & Tutorials*.
36. Liu, Y., Lan, Y., Li, B., Miao, C., & Tian, Z. (2021). Proof of Learning (PoLe): Empowering neural network training with consensus building on blockchains. *Computer Networks*, 201, 108594.
37. Singh, S., Kumar, A., & Kathuria, M. (2022). Understanding the public, private and consortium consensus algorithms in blockchain technology. *International Journal of Blockchains and Cryptocurrencies*, 3(3), 269–288.
38. Qiu, C., Aujla, G. S., Jiang, J., Wen, W., & Zhang, P. (2022). Rendering secure and trustworthy edge intelligence in 5G-enabled IIoT using proof of learning consensus protocol. *IEEE Transactions on Industrial Informatics*, 19(1), 900–909.
39. Bahreini, T., Badri, H., & Grosu, D. (2018). An envy-free auction mechanism for resource allocation in edge computing systems. In *2018 IEEE/ACM Symposium on Edge Computing* (pp. 313–322).
40. Król, M., Mastorakis, S., Oran, D., & Kutscher, D. (2019). Compute first networking: Distributed computing meets ICN. In *Proceedings of the 6th ACM Conference on Information-Centric Networking* (pp. 67–77).
41. Tang, X., Cao, C., Wang, Y., Zhang, S., Liu, Y., Li, M., & He, T. (2021). Computing power network: The architecture of convergence of computing and networking towards 6G requirement. *China Communications*, 18(2), 175–185.
42. Lee, J., Mtibaa, A., & Mastorakis, S. (2019). A case for compute reuse in future edge systems: An empirical study. In *2019 IEEE Globecom Workshops* (pp. 1–6). IEEE.
43. Chang, Z., Guo, W., Guo, X., Zhou, Z., & Ristaniemi, T. (2020). Incentive mechanism for edge-computing-based blockchain. *IEEE Transactions on Industrial Informatics*, 16(11), 7105–7114.
44. Jouffe, L. (1998). Fuzzy inference system learning by reinforcement methods. *IEEE Transactions on Systems, Man, and Cybernetics: Part C*, 28(3), 338–355.
45. Yang, H., Xiong, Z., Zhao, J., Niyato, D., Wu, Q., Poor, H. V., & Tornatore, M. (2020). Intelligent reflecting surface assisted anti-jamming communications: A fast reinforcement learning approach. *IEEE Transactions on Wireless Communications*, 20(3), 1963–1974.

46. Yao, H., Mai, T., Wang, J., Ji, Z., Jiang, C., & Qian, Y. (2019). Resource trading in blockchain-based industrial Internet of Things. *IEEE Transactions on Industrial Informatics*, 15(6), 3602–3609.
47. Ren, M., Guo, H., & Wang, Z. (2022). Mitigation of block withholding attack based on zero-determinant strategy. *PeerJ Computer Science*, 8, e997.
48. Gupta, D., Bhatt, S., Bhatt, P., Gupta, M., & Tosun, A. S. (2022). Game theory-based privacy preserving approach for collaborative deep learning in IoT. In *Deep Learning for Security and Privacy Preservation in IoT* (pp. 127–149). Springer Singapore.
49. Zhou, M., Wan, Z., Wang, H., Wen, M., Wu, R., Wen, Y., & Zhang, W. (2023). MALib: A parallel framework for population-based multi-agent reinforcement learning. *Journal of Machine Learning Research*, 24(150), 1–12.
50. Rapetswa, K., & Cheng, L. (2023). Towards a multi-agent reinforcement learning approach for joint sensing and sharing in cognitive radio networks. *Intelligent and Converged Networks*, 4(1), 50–75.
51. Wamba-Taguimdje, S. L., Fosso Wamba, S., Kala Kamdjoug, J. R., & Tchatchouang Wanko, C. E. (2020). Influence of artificial intelligence (AI) on firm performance: The business value of AI-based transformation projects. *Business Process Management Journal*, 26(7), 1893–1924.
52. Fan, H., Han, B., & Gao, W. (2022). (Im)balanced customer-oriented behaviors and AI chatbots' efficiency–flexibility performance: The moderating role of customers' rational choices. *Journal of Retailing and Consumer Services*, 66, 102937.
53. Chen, C., Wu, Z., Lai, Y., Ou, W., Liao, T., & Zheng, Z. (2023). *Challenges and Remedies to Privacy and Security in AIGC: Exploring the Potential of Privacy Computing, Blockchain, and Beyond*. arXiv preprint arXiv:2306.00419.
54. Ahmad, T., Zhang, D., Huang, C., Zhang, H., Dai, N., Song, Y., & Chen, H. (2021). Artificial intelligence in sustainable energy industry: Status quo, challenges, and opportunities. *Journal of Cleaner Production*, 289, 125834.
55. Chen, H., Chen, Y., Xiong, Z., Han, M., He, Z., Liu, B., & Ma, Z. (2023). Prevention method of block withholding attack based on miners' mining behavior in blockchain. *Applied Intelligence*, 53(9), 9878–9896.
56. Baynham-Herd, Z. (2017). Enlist blockchain to boost conservation. *Nature*, 548, 523.
57. Maxmen, A. (2018). AI researchers embrace Bitcoin technology to share medical data. *Nature*, 555, 293–295.
58. Nakamoto, S. (2009). *Bitcoin: A Peer-to-Peer Electronic Cash System*. <https://bitcoin.org/bitcoin.pdf> (accessed on 17 October 2024).
59. Taherdoost, H. (2022). An overview of trends in information systems: Emerging technologies that transform the information technology industry. *Cloud Computing and Data Science*, 4, 1–16.
60. Moosavi, N., & Taherdoost, H. (2022). Blockchain and Internet of Things (IoT): A disruptive integration. In *Proceedings of the 2nd International Conference on Emerging Technologies and Intelligent Systems (ICETIS 2022)* (pp. 1–10). Springer.
61. Swan, M. (2015). *Blockchain: Blueprint for a New Economy*. O'Reilly Media, Inc.
62. Pandl, K. D., Thiebes, S., Schmidt-Kraepelin, M., & Sunyaev, A. (2020). On the convergence of artificial intelligence and distributed ledger technology: A scoping review and future research agenda. *IEEE Access*, 8, 57075–57095.

63. Lin, J., Shen, Z., & Miao, C. (2017). Using blockchain technology to build trust in sharing LoRaWAN IoT. In *Proceedings of the 2nd International Conference on Crowd Science and Engineering* (pp. 38–43). Beijing, China.
64. Dai, Y., Xu, D., Maharjan, S., Chen, Z., He, Q., & Zhang, Y. (2019). Blockchain and deep reinforcement learning empowered intelligent 5G beyond. *IEEE Network*, 33(3), 10–17.
65. Salimitari, M., Chatterjee, M., Yuksel, M., & Pasiliao, E. (2017). Profit maximization for bitcoin pool mining: A prospect theoretic approach. In *Proceedings of the 2017 IEEE 3rd International Conference on Collaboration and Internet Computing (CIC)* (pp. 267–274). San Jose, CA, USA.
66. Singh, S. K., Rathore, S., & Park, J. H. (2020). Blockiotintelligence: A blockchain-enabled intelligent IoT architecture with artificial intelligence. *Future Generation Computer Systems*, 110, 721–743.
67. Dinh, T. N., & Thai, M. T. (2018). AI and blockchain: A disruptive integration. *Computer*, 51(9), 48–53.
68. Taherdoost, H. (2022). A critical review of blockchain acceptance models—Blockchain technology adoption frameworks and applications. *Computers*, 11(2), 24.
69. Wood, G. (2014). Ethereum: A secure decentralized generalized transaction ledger. *Ethereum Project Yellow Paper*, 151, 1–32.
70. Kumar, A., Abhishek, K., Nerurkar, P., Ghalib, M. R., Shankar, A., & Cheng, X. (2022). Secure smart contracts for cloud-based manufacturing using Ethereum blockchain. *Transactions on Emerging Telecommunications Technologies*, 33, e4129.
71. Li, D., Deng, L., Cai, Z., & Souri, A. (2022). Blockchain as a service models in the Internet of Things management: Systematic review. *Transactions on Emerging Telecommunications Technologies*, 33, e4139.
72. Wang, S., Yuan, Y., Wang, X., Li, J., Qin, R., & Wang, F. Y. (2018). An overview of smart contract: Architecture, applications, and future trends. In *Proceedings of the 2018 IEEE Intelligent Vehicles Symposium (IV)* (pp. 108–113). Changshu, China.
73. Makarius, E. E., Mukherjee, D., Fox, J. D., & Fox, A. K. (2020). Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization. *Journal of Business Research*, 120, 262–273.
74. Fusco, A., Dicuonzo, G., Dell’Atti, V., & Tatullo, M. (2020). Blockchain in healthcare: Insights on COVID-19. *International Journal of Environmental Research and Public Health*, 17(19), 7167.
75. Daley, S. (2020). *Tastier Coffee, Hurricane Prediction and Fighting the Opioid Crisis: 31 Ways Blockchain and AI Make a Powerful Pair*. BuiltIn. Retrieved from: <https://builtin.com/artificial-intelligence/blockchain-ai-examples> (accessed on 17 October 2024).

CHAPTER 2

Anomaly Detection Using Artificial Intelligence in a Cloud Fuzzy Environment

M. K. SHARMA,¹ VIJAYKUMAR,² BIBIN K. JOSE,³ R. P. AMBILWADE,⁴ and R. REVATHI⁵

¹Department of Mathematics, Chaudhary Charan Singh University, Meerut, Uttar Pradesh, India

²Department of Agricultural Engineering, SCRIET, CCS University, Meerut, Uttar Pradesh, India

³PG and Research Department of Mathematics, Sanatana Dharma College, Alappuzha, Kerala, India

⁴Computer Science, National Defense Academy, Pune, Maharashtra, India

⁵Computer Science, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India

ABSTRACT

Anomaly detection is vital for ensuring the quality of service in cloud infrastructures. However, the cloud environment poses challenges due to its complex and inconspicuous anomalies, high variability, and the absence of anomaly labels. Existing methods struggle to effectively address these issues. In this research, we introduce a novel approach for anomaly detection in the cloud fuzzy environment using CAT-IADEF.

Three convolutional neural networks are used in our method to collect sequence information, and adversarial training is used to increase the model's robustness while exaggerating minute anomalies. To increase accuracy, we use the Peaks Over Threshold (POT) approach to dynamically decide the detection threshold in real-time. We effectively explain anomalies by looking at anomaly score vectors in addition. Experimental findings on seven different types of open datasets are collected and other cutting-edge methods are outperformed by CAT-IADEF. In comparison to the baseline method, it performs better. Additionally, we include our framework for explaining anomalies into a number of models, demonstrating its usefulness on experimental datasets.

2.1 INTRODUCTION

The idea of “cloud computing” has substantially changed over time, going from a promising combination of data centre and virtualization technologies to a streamlined method for offering computer services to end consumers. It provides powerful computing capabilities that let users access and utilize cloud resources in a variety of ways, swiftly adapting to their specific requirements [1]. However, as the volume of data and the scope of network applications have dramatically expanded, the deployment of cloud platforms has also increased. This expansion results in the need for cloud computing systems to be highly reliable [2]. This reliability requirement extends to the capability of these systems to detect anomalies within time series data and provide insights into the potential causes or circumstances behind these anomalies.

Initially, many researchers focused on anomaly detection using univariate time series that focused on analyzing metrics of system like I/O request counts, memory usage, and throughput. However, as cloud systems have grown in complexity, monitoring data has evolved into multiple time series. For example, when monitoring CPU abnormalities, it becomes necessary to collect data on various CPU-related metrics, such as CPU user space utilization, CPU idle rates, and CPU time spent waiting for output and input. This complexity, along with the growing dataset sizes, has made multivariate time series for anomaly detection a significant area of research interest [3]. Anomaly detection is the process of spotting unusual actions that deviate from the usual behavior patterns

[4]. This solution safeguards your system in real-time by identifying and preventing occurrences that have the potential to lead to substantial financial losses, data breaches, and other detrimental incidents [5].

Traditional anomaly detection techniques, which relied on static thresholds, identified data points as outliers when they exceeded predetermined threshold values [6]. However, as data scales have increased, obtaining labeled data for training has become challenging, leading to issues like insufficient time series data and a lack of annotations [7]. Moreover, addressing anomaly detection that is proposed as unsupervised anomaly detection, in which the training process is ensured that have become key priorities [8]. To facilitate further research, this paper introduces the concept of anomaly explanation, which focuses on identifying dimensions where anomalies are most likely to occur. We present an adversarial training method for anomaly explanation, starting with the extraction of key sequence features using convolutional neural networks (CNN) [9]. Subsequently, we construct an adversarial training framework comprising three CNNs in conjunction with the dynamic threshold selection method known as POT [10].

The need to identify security across numerous applications, including parking area monitoring, spotting illegal activity, detection of abandoned object, identification of accidents, detection of crimes, depends heavily on the ability to notice anomalies in surveillance videos [11]. Nevertheless, humans find it difficult and time-consuming to manually discover anomaly. This is mainly because security systems create such enormous amounts of data that human analysis is difficult.

The demand for automated systems that can recognize video anomalies has increased dramatically in recent years. The systems involve CCTV scene driven by alarms, identification of people by biometrics, automatic traffic infraction detection, and video-based aberrant behavior identification [1]. Automated systems significantly alleviate the burden on human resources and time, making anomaly detection in surveillance videos more efficient and cost-effective.

The identification and tracking of anomalies in recorded video footage pose an evolving research challenge in the field of surveillance [12]. Many academic researchers have proposed various methods, employing techniques like image processing and machine learning algorithms [13].

Multiple purposes are served through experienced substantial growth due to utilization of sovereign camera in private and public places.

Detecting anomalies using diverse scenarios that of course various safety measures as essential part in video footage of surveillance camera. However, identification of anomaly using surveillance footage is done manually can be a tough and labor-intensive task [14]. To address this, researchers have introduced automated systems for video anomaly detection, significantly reducing the need for time and human labor [15]. Notwithstanding these developments, there is still room for enhancing the scalability, reliability and accuracy of video surveillance systems.

An innovative technology called the ‘Surveillance Video Anomaly Detection system’ is intended to automatically spot suspicious or anomalous behavior in video surveillance material without requiring human intervention [16]. The system achieves this by scrutinizing video frames, recognizing deviations from normal patterns of activity or movement [17]. It employs machine learning algorithms and other techniques to assess pixel positions in the video frame during an event [18].

In the past, anomaly detection techniques mostly concentrated on locating items that deviated from expected paths. In the video frame different types of items are available; these algorithms must be improved before they can be used effectively in video surveillance [19]. As a result, there are now two main methods for detecting video anomalies [20]. The first method involves calculating the reconstruction error of upcoming frames in order to determine the error magnitude [21]. To do this, large discrepancies between expected future frames and actual frames are flagged. The second method involves making future frame predictions based on past frames and giving frames that significantly differ from the projected frame a high anomaly score [22].

Due to recent advancements in emergence of new models and capabilities of new hardware and intelligent learning techniques has gained popularity in anomaly detection [23] in cloud fuzzy environment. However, their application also introduces several challenges. One significant challenge lies in the generation of large-scale data, necessitating substantial process of computational power. The demand for high computational power also presents a resource-intensive hurdle. The demand for high computational power also presents a resource-intensive hurdle [24].

In addition, we incorporate an anomaly explanation module that relies on the calculation of an attribution matrix for anomaly for every data point.

The present study makes several significant contributions:

- Addressing the challenge of identifying “subtle” anomalies and enhancing the robustness of models employed adversarial training for amplification of reconstruction errors and thereby improving both anomaly detection accuracy and sensitivity.
- To enhance the efficiency of computation through exploration of static threshold, we Effective integration of dynamic threshold POT using selection technology, resulting in improved detection performance.
- We introduce a comprehensive anomaly explanation framework to support post-detection anomaly handling. This framework involves calculating the number of anomalies based on = threshold matrix and anomaly score at each data point, enabling a partial explanation of the detected anomalies.

2.2 METHODS

This section includes the methodology part.

In the realm of artificial intelligence, various subsets are tailored for diverse applications and use cases [25]. This study primarily emphasizes using Machine Learning. Machine Learning is a potent technology applicable to anomaly detection, although the specific approach varies significantly depending on the problem at hand. The effectiveness of an ML algorithm can vary based on factors such as feature selection or the weighting assigned to each feature within the dataset [26]. Even when the same model is employed on identical datasets, variations may occur. If a model possesses too few features that are only occasionally effective, it may become prone to over fitting [27]. To gain a deeper understanding of ML techniques and their application to available data, it is valuable to review and compare existing solutions. Machine Learning can be broadly categorized Semi-supervised Learning, Unsupervised Learning and Supervised Learning [28].

2.2.1 SUPERVISED LEARNING

In supervised learning, knowledge is extracted from pre-existing labeled datasets, often referred to as the “training set [29].” These datasets allow the model to compare its predicted outputs to known labels. While a high-quality training set is essential for creating an effective

model, the training process itself plays a pivotal role in producing a reliable predictor [30]. SL involves the development of a classifier model through training, which can subsequently predict either discrete or continuous outputs. Typically, the performance of a supervised learning model, such as accuracy, is assessed before predictions are made. Classification and regression techniques are commonly employed in SL tasks [30, 31]. In classification, data is grouped into distinct categories, and the probability of test samples falling into each category can be calculated [32, 33]. The highest probability for every category can be selected for prediction. In regression, input factors are used to forecast continuous responses, such as changes in temperature or variations in electricity demand [34].

Numerous works have been conducted in SL, including the detection of malicious attacks based on temporal and spatial information extracted from consecutive video frames. Various techniques like support vector machines [35], convolutional neural network [36], hidden markov models [37], gaussian regression (GR) [38], long short-term memory [39] and multiple instance learning [40] have been employed, each anomaly detection has its own merits and demerits.

2.2.2 UNSUPERVISED LEARNING

Unsupervised learning groups data by uncovering hidden patterns or inherent structures without relying on predetermined output variables [41]. Unlike supervised learning, it operates independently, requiring no labeled input data or training. Consequently, it can offer more quantifiable performance metrics. Although some researchers use pre-existing labeled data to verify UL results, this is not always feasible in practice. UL is often used for dimensionality reduction to identify and remove redundant data, thereby reducing noise [42]. Clustering techniques allow samples to belong to multiple clusters or just one, addressing problems such as market research and object identification [43].

An example of UL in anomaly detection is the use of OCSVM [44] (one-class support vector machines) combined with deep feature extraction techniques employing pre-trained CNN on surveillance footage [45]. Various UL techniques, including generative adversarial networks and principal component analysis, have been employed.

2.2.3 SEMI-SUPERVISED LEARNING

Semi-supervised learning (SSL) is a machine learning approach that harnesses both labeled and unlabeled data to construct a classifier [46]. This is particularly advantageous when only a labeled data with limited data set is available. SSL employs a training process akin to supervised learning for creation of prediction model using small labeled data sets. This predictor then categorizes unlabeled samples and assigns confidence scores to pseudo-labeled samples. SSL relies on assumptions like smoothness and clustering before training unlabeled examples. It assumes that unlabeled data can be confidently labeled during the prediction process [47, 49]. SSL models, such as the anomaly detection (AE) model, are instrumental in using both unlabelled and labeled data for detection of anomalies within a dataset [50].

2.2.4 COMPARATIVE PERFORMANCE

The use of supervised learning techniques for anomaly detection has various benefits, including the capacity to identify particular types of anomalies and the exact identification of abnormalities using labeled data [51]. These methods are very beneficial for security and monitoring applications. However, they may be sensitive to environmental changes and require a significant amount of labeled data, which affects their accuracy [52].

The use of unsupervised learning techniques for anomaly detection has several benefits, including the capacity to discover abnormalities instantly and the lack of a need for labeled data [53]. They are able to spot anomalous patterns and categorize those that deviate from the usual [54]. They can, however, be susceptible to environmental changes and may not detect certain sorts of anomalies [55].

2.2.5 PROBLEM STATEMENT

The sequence of arrange values with specific statistical indicators for the particular phenomena arranged in the chronological order over the time is referred as time series. Time series of single variable, denoted as $T = \{x_1, x_2, \dots, x_t\}$, And the collected data is recorded over the time for the entire series that comprise of single variable. In the multivariate series the data point can include the vector representation of x_t .

The importance of the research is given below:

- Concerning anomaly detection, we start by normalizing the minimum and the maximum values of time series (T). The *W* of window series is used as the training input for generation of output *Y*. Initially, L_t is considered as the loss of reconstruction. Then, the study calculates the S_t as anomaly score and threshold is determined as *Threshold* based on the peculiarity score for the previous window. Regular data points are assumed to have good reconstructions, while anomalies have reconstruction errors compared to the threshold, resulting in labels $y \in (0, 1)$.
- With regard to anomaly explanation, we compute the reconstruction loss for each data point for each time series *T*. This data allows us to forecast a label sequence, $Y = \{y_1, y_2 \dots y_t\}$, where each $y_t \in (0, 1)$. This sequence allows us to identify the dimension with the most occurrences of ones, providing insight into the anomaly's characteristics.

2.2.6 PROPOSED FRAMEWORK OF CAT-IADEF

“The proposed framework for CAT-IADEF (Convolutional adversarial training based integrated anomaly detection with explanation framework) includes window sequence *W*, which was created by processing the time series data *T*, is shown on the left side of the diagram to start.” A convolutional auto-encoder receives this window sequence as input, producing a group of latent variables known as *Z*. The middle of the diagram shows two sets of auto-encoders, *ED1* and *ED2* (each comprising its own encoder and decoder), are trained adversarially. This adversarial training phase aims to enhance the ability of the *ED2* model to detect subtle anomalies and generate the reconstructed data from both auto-encoders.

During this adversarial training, the reconstruction loss *L* is computed for both the anomaly score S_i is computed by weighted averaging over sets of auto-encoders. A multivariate time series anomaly can be easily found using this method. The Anomaly detection model in cloud fuzzy environment is located on the diagram's right side. It records the label matrix are for the corresponding anomaly score at each data point of cloud using the anomaly score as input. This module decides whether each data point is anomalous or not by comparing the anomaly score to a Zero or 1

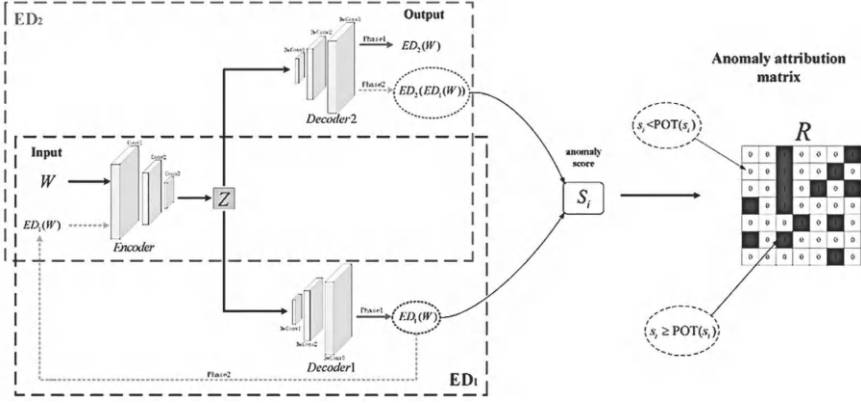


FIGURE 2.1 Proposed framework of CAT-IADEF.

to the value of threshold. Then the columns of the ‘matrix R’ are added up, as well as the highest sum of the column denotes the dimension with the greatest number of ones. The dimension where anomalies are most likely to exist is this one.

2.2.7 TRAINING USING CNN

Techniques for adversarial training in the GAN approach have shown promising results in the area of time series data anomaly detection. Figure 2.1 shows how we integrated two CNN decoders into the CAT-IADEF model in this context. The concurrent process is performed at dual stage for training the entire model.

As depicted in Figure 2.1, *ED1* encompasses *Encoder* and *Decoder1*, while *ED2* comprises *Encoder* and *Decoder2*. Initially, both *ED1* and *ED2* are tasked with reconstructing the input window sequence *W*, followed by adversarial training. *ED1* aims to deceive *ED2*, while *ED2* strives to ascertain the authenticity of the data. Here are the specifics of the two-stage training:

Phase 1: Minimizing Reconstruction. L2 norm is applied during Phase 1:

$$L1 = \|W - ED1(W)\|_2 \quad (2.1)$$

$$L2 = \|W - ED2(W)\|_2 \quad (2.2.)$$

Minimization of loss is main objective of the study:

$$\min \|W - ED1(W)\|_2 \quad (2.3)$$

$$\min \|W - ED2(W)\|_2 \quad (2.4)$$

Phase 2: In the second phase, *ED2* is trained to distinguish actual data. Meanwhile, *ED1* is trained to deceive *ED2*, and *ED2* reconstructs the data generated by *ED1* and quantifies the reconstruction error relative to the actual data *W*. Thus, the reconstruction loss in this phase is expressed as:

$$L3 = \|W - ED2(ED1(W))\|_2 \quad (2.5)$$

At this point, *ED1* aims for minimization of difference between two outputs of *W* and *ED2*. In contrast, *ED2* seeks to maximize the difference $\|W - ED2(ED1(W))\|_2$ to discern the data *ED1(W)* generated by *ED1* from the genuine data *W*. Hence, the training objectives at this stage are as follows:

$$\min_{ED1} \max_{ED2} \|W - ED2(ED1(W))\|_2 \quad (2.6)$$

ED1 strives to minimize this error, while *ED2* seeks to maximize it, effectively amplifying it. Consequently, we modify the reconstruction loss by introducing a sign:

$$L*1 = +\|W - ED2(ED1(W))\|_2 \quad (2.7)$$

$$L*2 = -\|W - ED2(ED1(W))\|_2 \quad (2.8)$$

2.2.8 DUAL-STAGE TRAINING

The functions for loss of reconstruction are identified in the study for *ED1* and *ED2* for combining the various stages and formulation of training the objectives for *ED1* and *ED2* throughout the entire training process:

$$LED1 = 1n\|W - ED1(W)\|_2 + (1-1n)\|W - ED2(ED1(W))\|_2 \quad (2.9)$$

$$LED2 = 1n\|W - ED2(W)\|_2 - (1-1n)\|W - ED2(ED1(W))\|_2 \quad (2.10)$$

To assess the effectiveness of CAT-IADEF, we conducted five distinct experiments, outlined as follows:

2.2.8.1 COMPARISON OF PERFORMANCE

The research study compares the performance with baseline method and the proposed approach. The results of these experiments are explained in the “anomaly detection” section in Section 4.4.

2.2.8.2 ANOMALY EXPLANATION

We evaluated the anomaly explanation capabilities of CAT-IADEF. The results of these experiments are explained in the results and discussion part under anomaly explanation section.

2.2.8.3 VALIDATION OF THE STUDY

We validated the anomaly explanation framework on various models. The results of these experiments can be found in the “anomaly explanation” section.

2.2.9 ANALYZING THE ABLATION

We analyzed the impact of individual model components on overall model performance.

- Sensitivity Analysis of Parameters: We examined how different parameters influenced the model’s performance.
- Experimental Environments: We selected the following techniques for performance comparison:
 - Dimensionality reduction-based PCA,
 - Density-based LOF,
 - Machine learning- and statistics-based COPOD, and
 - Classification-based OCSVM.

These four methods are considered classical approaches. Additionally, we included recent deep learning methods with strong performance, such as TranAD, USAD, Omni Anomaly, and MAD_GAN.

- Datasets: We utilized four different datasets related to waste treatment. Comprehensive information about these datasets is available. From each dataset, we carefully selected a number of data subsets for training and model evaluation in our experiments.

2.2.9.1 *SMAP*

Remote sensing data are included in the soil data gathered by satellites [56]. We updated a two-layer palmer model and included passive and active satellite derived for taking readings.

2.2.9.2 *MSL*

The mars science laboratory rover collected data for MSL, which is similar to the SMAP datasets [56].

2.2.9.3 *SWAT*

The safe water treatment dataset is derived from actual industrial water treatment facilities used to produce filtered water. It includes information from 4 days of aberrant function and 7 days of normal operation [57].

2.2.9.4 *SKAB*

This dataset is essential for analyzing anomaly detection techniques. More than 30 datasets and Python modules for algorithm evaluation are now included in the benchmark. Each dataset is an accumulation of the multi-variate time series which is placed on them from the sensors [58].

Information on each dataset's instance count, training-to-testing ratio, feature count, and percentage of anomalies is shown in Table 2.1. These four accessible datasets, with different training/testing methodologies, are frequently used in anomaly detection. Partitions, distinct numbers of features, and diverse anomaly ratios. Our approach's effectiveness and robustness are validated using these datasets, encompassing different anomaly proportions, training/test splits, and feature counts [59].

The evaluation metrics for anomaly detection is used to evaluate model performance; we employ common metrics such as precision, accuracy [60].

Calculation framework is utilized for independence matrix for direct identification of serial numbers that belongs to two dimensions with most of the anomalies in point of cloud fuzzy environment. This approach enables relevant members for focusing on observation of specific dimensions that works with multivariate time series data. It also simplifies data review for researchers to a certain extent.

2.3 RESULTS AND DISCUSSION

While deep learning methods are generally recognized as superior, Machine learning algorithm for anomaly detection that outperforms using the experimental data set. Then it showcases CAT-IADEF’s overall performance, 4 classical methods of machine learning are compared for anomaly detection in cloud fuzzy environment. In these comparative experiments, the adjustment for dynamic threshold is applied fairly to all the machine learning methods. The average value is obtained using the set of experiments which is conducted for several times.

TABLE 2.1 Characteristics of Dataset

Dataset	Instances	Testing	Training	Anomaly
MSL	132,046	56	44.2	11
SMAP	562,800	76	24.0	13.1
SKAB	46,806	80	20.1	35
SWaT	946,719	48	52.5	12

To assess various models, we initiated our analysis with four machine learning methods. The performance of PCA, specifically the number of latent variables it yields, is not reliable, resulting in subpar results across the seven datasets. COPOD, on the other hand, excels as it doesn’t require parameter adjustments or sample distance calculations, and its computational overhead is minimal. However, COPOD and OCSVM demonstrate strong performance on the T1 dataset but struggle on the remaining datasets. This can be attributed to the challenges posed by non-stationary and imbalanced time series data in the face of changing scenes. While LOF’s performance on other datasets isn’t remarkable, it achieves the best F1 score on the dataset T1, primarily since there is a distribution of local time series of subset is conducive to LOF’s k-fields calculation.

In the realm of deep learning, USAD’s reliance on a basic AE model and linear transformations limits its impact, resulting in marginal effectiveness. In contrast, CAT-IADEF employs a CNN network and considers inter-feature relationships, enhancing its anomaly detection capabilities. MAN_GAN and TranAD utilize an attention mechanism for anomaly detection. TranAD incorporates positional encoding within the transformer structure, which performs exceptionally well on the dataset using SWaT. Omni anomaly, while not standing out across most datasets, particularly

TABLE 2.2 Comparison of Performance

Methods	SMAP				MSL				SWaT			
	A	P	R	F1	A	P	R	F1	A	P	R	F1
PCA	0.890	0.110	0.999	0.198	0.934	0.031	0.999	0.060	0.858	0.394	0.696	0.503
LOF	0.794	0.691	0.999	0.817	0.931	0.410	0.999	0.581	0.813	0.978	0.696	0.813
COPOD	0.794	0.319	0.999	0.484	0.936	0.108	0.999	0.195	0.894	0.503	0.731	0.596
OCSVM	0.791	0.366	0.999	0.536	0.934	0.108	0.999	0.195	0.844	0.214	0.862	0.343
MAD_GAN	0.980	0.870	0.999	0.919	0.954	0.401	0.999	0.571	0.956	0.952	0.696	0.803
Omni Anomaly	0.982	0.859	0.999	0.924	0.965	0.441	0.999	0.612	0.959	0.976	0.969	0.812
USAD	0.980	0.849	0.999	0.918	0.969	0.488	0.999	0.648	0.959	0.989	0.688	0.811
TranAD	0.980	0.857	0.999	0.918	0.906	0.246	0.999	0.395	0.960	0.997	0.688	0.814
CAT-IADEF	0.981	0.857	0.999	0.923	0.979	0.596	0.999	0.747	0.958	0.971	0.696	0.811

except for P1, processes input sequentially, retaining essential information and reconstructing all inputs, regarding the characteristics of input data. This approach hinders its ability to detect anomalies that closely resemble normal trends. CAT-IADEF’s adversarial training, however, empowers it to amplify and discern “subtle” anomalies, even when other models might consider them routine data in Table 2.2.

2.3.1 ANALYSIS OF ABLATION

To assess the impact of individual model components on the overall performance of model, crucial element one by one as well as evaluates the absence by influencing the model’s performance using the experimental dataset.

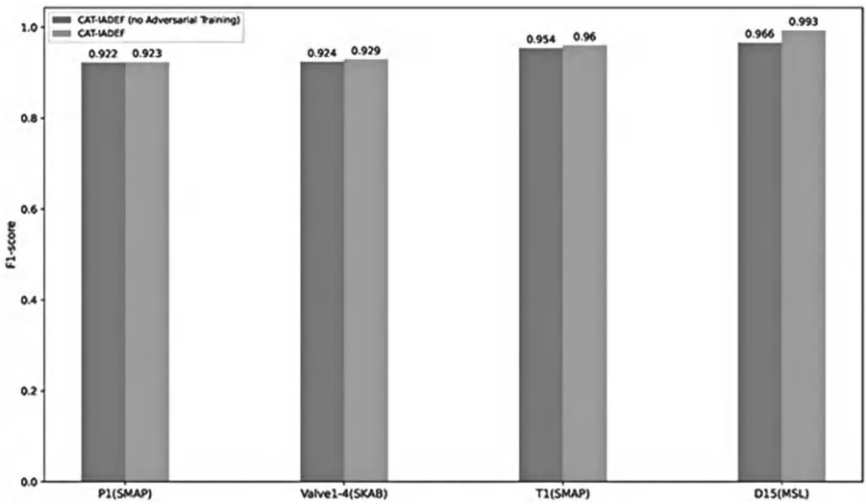


FIGURE 2.2 CAT-IADEF performance for adversarial training.

As depicted in Figure 2.2, The performance of the proposed model without and with adversarial training is exhibited for the decline across the four different types of data sets that ranging from the Maximum drop of 2.7% and the minimum drop is 0.11% the performance analysis shows the adversarial training in the proposed model which plays a crucial role in enabling the model for detection of anomalies through amplification of various errors.

The studies related to dynamic thresholding for the removal of component related to dynamic threshold from the proposed model as well as the model is trained for the anomaly detection using the data sets in the experiment. Figure 2.3 illustrates the positive impact on POT dynamic threshold selection method for the proposed model. The improvement or noticeable and observed in data set D15 where the performance increases from 0.7–0.99 representing a remarkable enhancement of 36%. Conversely the improvement is very smaller and occurs in the data set P1 with performance increasing from 0.86–0.93 representing and boosting of 8.5%.

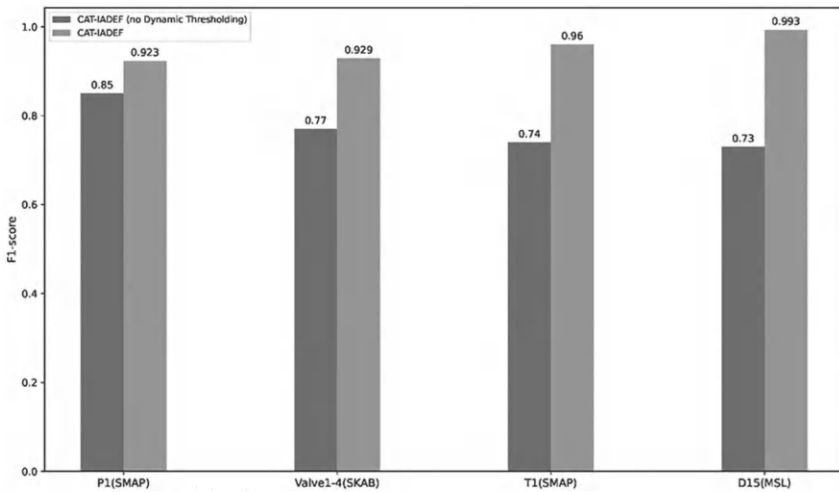


FIGURE 2.3 CAT-IADEF for dynamic thresholding.

2.3.2 ANALYSIS OF SENSITIVITY

In case of loss a function, where the researcher can employ alpha and beta for performing the weighted average using *Decoder1* as well as *Decoder2*, with $\alpha + \beta$ equaling 1. When β is set to 0 and α is set to 1, the loss function exclusively. Considering the loss of reconstruction of decoder one. Conversely, when β is set to 1 and α is set to 0, the model solely considers for the loss of reconstruction in *Decoder2* on the generated data. As In figure the experiments are conductor using data sets for anomaly detection in cloud environment, exploring seven different cases to demonstrate the impact of parameter pairs on model performance.

TABLE 2.3 Comparison of Dataset

α	β	R	P	F1
0	1	0.99	0.87	0.9312
0.1	0.9	0.99	0.87	0.9322
0.3	0.7	0.99	0.87	0.9302
0.5	0.5	0.99	0.85	0.9176
0.7	0.3	0.99	0.85	0.9186
0.9	0.1	0.99	0.85	0.9186
1	0	0.99	0.85	0.9167

As indicated in Table 2.3, when the sum of α and β ranges from 0–1, it becomes evident that as α approaches 1 and β approaches 0 (meaning a greater emphasis on accurate data), the F1 score tends to decrease, reaching its maximum value. Conversely, when α approaches 0 and β approaches 1 (indicating a higher focus on the reconstruction loss of generated data, essentially anomalies resembling standard data in simulated data), F1 takes on higher values.

In practical scenarios, training data may not always conform to the normal distribution, and examining the sensitivity of α and β allows the model to be more adaptable to varying experimental conditions and data characteristics.

In today’s landscape, there is a rapid proliferation of cloud environments across various domains, replacing the use of physical hardware-based machines with virtual machines to execute applications [61, 62]. This shift enhances service availability and concurrently reduces costs. Open Stack, an open-source cloud platform [63, 64], is experiencing a surge in adoption within both academic and industrial settings [65]. It offers accessible cloud services for research and production purposes [66].

However, one of the challenges encountered in cloud environments is the timely detection and prediction of anomalies before they manifest [67, 68]. In conventional approaches, anomalies are identified through manual monitoring, involving the tracking of threshold levels and system heartbeat [69, 70]. Current research efforts are focused on leveraging machine learning techniques to proactively detect anomalies before their occurrence [71].

Edge and cloud computing, cyber security, and artificial intelligence must be combined with a well-known technology for digitalizing industrial applications and sectors in order to increase its effectiveness, Sustainability and dependability [72–74]. Even though, the difficulty of assembling various data from various sensors, as well as securely managing and monitoring of large database, has shown up as a critical issue within the context of IIoT. This study advocates building a integrated smart IIoT infrastructure that are successfully handle these concerns in order to address these problems. This proposed infrastructure uses AI to provide security, gather different types of data, enable intercontinental connection, and execute real-time anomaly detection.

The transfer of energy data to edge IIoT servers is made possible by the design and construction of intelligent data gathering devices. The transport layer security protocol and hash encoding credentials are used to strengthen security on these servers. Additionally, a message queuing telemetry transport protocol is used to safeguard data transfer between these servers. Large amounts of data can be managed using edge and cloud databases. An algorithm using a collection of isolation forest models is created and put into use on edge and cloud servers to detect anomalies in specific electrical equipment in real-time [75]. MLS data consists of a vast number of points, making it computationally expensive and inefficient to analyze each element and its surroundings individually for feature extraction. To enhance the efficiency of the extraction and detection process for point cloud that can be obtained from the previous steps which are structured as well as examined for pillar arrangement in 3D vertical points. The cloud points are then subdivided into a 2-dimensional grid consisting of m columns and n rows [76].

His rapid growth in the development of network applications has resulted in a significant volume of network data being generated from various services catering to large user populations. Securing both network applications and internet-connected devices has consistently garnered the attention of researchers. Numerous studies have proposed solutions for intrusion detection systems and intrusion prevention systems. However, there is an urgent need to fine-tune contemporary data mining and statistical techniques to effectively address the

challenges posed by the expanding landscape of internet applications, including issues related to bandwidth management, network intrusion detection, and scalability. An intrusion detection system that leverages a combination of prediction and learning mechanisms in order to enhance the accuracy of anomaly detection within a network intrusion context [48].

2.4 CONCLUSION

We introduce a novel model for explaining anomalies using adversarial convolution techniques, designed for performing Time series explanation using the dataset for anomaly detection in cloud fuzzy environment. In contrast to traditional linear approaches, our convolutional structure allows for parameter sharing, significantly reducing the network's parameter count and effectively preventing over fitting. By employing adversarial training, our model amplifies subtle anomalies, enhancing its ability to distinguish them from normal data. Moreover, the sensitivity adjustments during adversarial training make our model applicable across a wider range of environments. Additionally, we leverage the POT technique to optimize anomaly detection, focusing on modeling the extreme values above a threshold. For anomaly explanation, we employ a straightforward loss function calculation method. Our model's performance is evaluated on different types of data sets available in public, and the results validate its effectiveness. Notably, CAT-IADEF outperforms all other models in our comparisons, establishing it as an excellent choice for anomaly detection in cloud fuzzy environment.

Looking ahead, we plan to enhance the model's capabilities by incorporating multi-head attention mechanism as a form of attention mechanism to enable the model to focus on specific unusual sequence points. We also aim to introduce the residual connections using convolutional neural networks for expedition of model training. The findings intend to investigate the cost and environmental impact of deploying the model in real-world scenarios, striving to minimize energy consumption and environmental pollution.

KEYWORDS

- **anomaly detection**
- **artificial intelligence**
- **cloud fuzzy environment**
- **convolutional neural network**

REFERENCES

1. Mohammadian, V., Navimipour, N. J., Hosseinzadeh, M., & Darwesh, A. (2021). Fault-tolerant load balancing in cloud computing: A systematic literature review. *IEEE Access*, 10, 12714–12731.
2. Nedelkoski, S., Bogatinovski, J., Mandapati, A. K., Becker, S., Cardoso, J., & Kao, O. (2020). Multi-source distributed system data for AI-powered analytics. In *Service-Oriented and Cloud Computing: 8th IFIP WG 2.14 European Conference, ESOC 2020, Heraklion, Crete, Greece, September 28–30, 2020, Proceedings* (pp. 161–176). Springer International Publishing.
3. Chalapathy, R., & Chawla, S. (2019). *Deep Learning for Anomaly Detection: A Survey*. arXiv preprint arXiv:1901.03407.
4. Mihailescu, M. I., Nita, S. L., Rogobete, M., & Marascu, V. (2023). Unveiling threats: Leveraging user behavior analysis for enhanced cybersecurity. In *2023 15th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)* (pp. 01–06). IEEE.
5. Chithanuru, V., & Ramaiah, M. (2023). An anomaly detection on blockchain infrastructure using artificial intelligence techniques: Challenges and future directions—A review. *Concurrency and Computation: Practice and Experience*, e7724.
6. Lu, T., Wang, L., & Zhao, X. (2023). Review of anomaly detection algorithms for data streams. *Applied Sciences*, 13(10), 6353.
7. Ren, H., Xu, B., Wang, Y., Yi, C., Huang, C., Kou, X., & Zhang, Q. (2019). Time-series anomaly detection service at Microsoft. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 3009–3017).
8. Audibert, J., Michiardi, P., Guyard, F., Marti, S., & Zuluaga, M. A. (2020). USAD: Unsupervised anomaly detection on multivariate time series. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 3395–3404).
9. Wang, Q., Zhang, X., Wang, X., & Cao, Z. (2021). Log sequence anomaly detection method based on contrastive adversarial training and dual feature extraction. *Entropy*, 24(1), 69.

10. Wen, P., Yang, Z., Wu, L., Qi, S., & Chen, J. (2022). An unsupervised convolutional adversarial anomaly detection model for IoT data infrastructure. In *2022 IEEE 24th International Conference on High Performance Computing & Communications; 8th International Conference on Data Science & Systems; 20th International Conference on Smart City; 8th International Conference on Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys)* (pp. 1336–1341). IEEE.
11. Liu, K., Zhu, M., Fu, H., Ma, H., & Chua, T. S. (2020). Enhancing anomaly detection in surveillance videos with transfer learning from action recognition. In *Proceedings of the 28th ACM International Conference on Multimedia* (pp. 4664–4668).
12. Singh, V., Singh, S., & Gupta, P. (2020). Real-time anomaly recognition through CCTV using neural networks. *Procedia Computer Science*, 173, 254–263.
13. Hou, Y., Li, Q., Zhang, C., Lu, G., Ye, Z., Chen, Y., & Cao, D. (2021). The state-of-the-art review on applications of intrusive sensing, image processing techniques, and machine learning methods in pavement monitoring and analysis. *Engineering*, 7(6), 845–856.
14. Irfanullah, Hussain, T., Iqbal, A., Yang, B., & Hussain, A. (2022). Real-time violence detection in surveillance videos using convolutional neural networks. *Multimedia Tools and Applications*, 81(26), 38151–38173.
15. Defard, T., Setkov, A., Loesch, A., & Audigier, R. (2021). PaDiM: A patch distribution modeling framework for anomaly detection and localization. In *International Conference on Pattern Recognition* (pp. 475–489). Cham: Springer International Publishing.
16. Patrikar, D. R., & Parate, M. R. (2022). Anomaly detection using edge computing in video surveillance system. *International Journal of Multimedia Information Retrieval*, 11(2), 85–110.
17. Rezaee, K., Rezakhani, S. M., Khosravi, M. R., & Moghimi, M. K. (2021). A survey on deep learning-based real-time crowd anomaly detection for secure distributed video surveillance. *Personal and Ubiquitous Computing*, 1–17.
18. Berroukham, A., Housni, K., Lahraichi, M., & Boulfrifi, I. (2023). Deep learning-based methods for anomaly detection in video surveillance: A review. *Bulletin of Electrical Engineering and Informatics*, 12(1), 314–327.
19. Abdusalomov, A., Baratov, N., Kutlimuratov, A., & Whangbo, T. K. (2021). An improvement of the fire detection and classification method using YOLOv3 for surveillance systems. *Sensors*, 21(19), 6519.
20. Zhu, Y., & Newsam, S. (2019). *Motion-Aware Feature for Improved Video Anomaly Detection*. arXiv preprint arXiv:1907.10211.
21. Zhong, Y., Chen, X., Jiang, J., & Ren, F. (2022). A cascade reconstruction model with generalization ability evaluation for anomaly detection in videos. *Pattern Recognition*, 122, 108336.
22. Luo, W., Liu, W., Lian, D., & Gao, S. (2021). Future frame prediction network for video anomaly detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11), 7505–7520.
23. Al-Amri, R., Murugesan, R. K., Man, M., Abdulateef, A. F., Al-Sharafi, M. A., & Alkahtani, A. A. (2021). A review of machine learning and deep learning techniques for anomaly detection in IoT data. *Applied Sciences*, 11(12), 5320.

24. Du, H., Zhang, R., Niyato, D., Kang, J., Xiong, Z., Kim, D. I., & Poor, H. V. (2023). Exploring collaborative distributed diffusion-based AI-generated content (AIGC) in wireless networks. *IEEE Network*, (99), 1–8.
25. Odyurt, U., Pimentel, A. D., & Alonso, I. G. (2022). Improving the robustness of industrial Cyber-Physical Systems through machine learning-based performance anomaly identification. *Journal of Systems Architecture*, 131, 102716.
26. Sarvari, S., Sani, N. F. M., Hanapi, Z. M., & Abdullah, M. T. (2020). An efficient anomaly intrusion detection method with feature selection and evolutionary neural network. *IEEE Access*, 8, 70651–70663.
27. Ullah, I., & Mahmoud, Q. H. (2022). Design and development of RNN anomaly detection model for IoT networks. *IEEE Access*, 10, 62722–62750.
28. Dong, Y., Chen, K., & Ma, Z. (2023). *Comparative Study on Semi-Supervised Learning Applied for Anomaly Detection in Hydraulic Condition Monitoring System*. *arXiv preprint arXiv:2306.02709*.
29. Chillara, S., Kavitha, A. S., Neginhal, S. A., Haldia, S., & Vidyullatha, K. S. (2019). Music genre classification using machine learning algorithms: A comparison. *International Research Journal of Engineering and Technology*, 6(5), 851–858.
30. Jayatilake, S. M. D. A. C., & Ganegoda, G. U. (2021). Involvement of machine learning tools in healthcare decision making. *Journal of Healthcare Engineering*, 2021, 1–14.
31. El Mrabet, M. A., El Makkaoui, K., & Faize, A. (2021). Supervised machine learning: A survey. In *2021 4th International Conference on Advanced Communication Technologies and Networking (CommNet)* (pp. 1–10). IEEE.
32. Wang, S., Balarezo, J. F., Kandeeppan, S., Al-Hourani, A., Chavez, K. G., & Rubinstein, B. (2021). Machine learning in network anomaly detection: A survey. *IEEE Access*, 9, 152379–152396.
33. Meng, L., McWilliams, B., Jarosinski, W., Park, H. Y., Jung, Y. G., Lee, J., & Zhang, J. (2020). Machine learning in additive manufacturing: A review. *JOM*, 72(7), 2363–2377.
34. Nasteski, V. (2017). An overview of the supervised machine learning methods. *Horizons. B*, 4, 51–62.
35. Dhiman, H. S., Deb, D., Muyeen, S. M., & Kamwa, I. (2021). Wind turbine gearbox anomaly detection based on adaptive threshold and twin support vector machines. *IEEE Transactions on Energy Conversion*, 36(4), 3462–3469.
36. Gao, J., Song, X., Wen, Q., Wang, P., Sun, L., & Xu, H. (2020). *RobustTAD: Robust Time Series Anomaly Detection via Decomposition and Convolutional Neural Networks*. *arXiv preprint arXiv:2002.09545*.
37. Leon-Lopez, K. M., Mouret, F., Arguello, H., & Tournet, J. Y. (2021). Anomaly detection and classification in multispectral time series based on hidden Markov models. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–11.
38. Fan, J., Zhang, Q., Zhu, J., Zhang, M., Yang, Z., & Cao, H. (2020). Robust deep auto-encoding Gaussian process regression for unsupervised anomaly detection. *Neurocomputing*, 376, 180–190.
39. Tran, K. P., Du Nguyen, H., & Thomassey, S. (2019). Anomaly detection using long short term memory networks and its applications in supply chain management. *IFAC-PapersOnLine*, 52(13), 2408–2412.

40. Feng, J. C., Hong, F. T., & Zheng, W. S. (2021). MIST: Multiple instance self-training framework for video anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 14009–14018).
41. Shu, X., & Ye, Y. (2023). Knowledge discovery: Methods from data mining and machine learning. *Social Science Research*, 110, 102817.
42. Shi, G., Huang, H., & Wang, L. (2019). Unsupervised dimensionality reduction for hyperspectral imagery via local geometric structure feature learning. *IEEE Geoscience and Remote Sensing Letters*, 17(8), 1425–1429.
43. Huang, G., Song, S., Gupta, J. N., & Wu, C. (2014). Semi-supervised and unsupervised extreme learning machines. *IEEE Transactions on Cybernetics*, 44(12), 2405–2417.
44. Natalino, C., Udalcovs, A., Wosinska, L., Ozolins, O., & Furdek, M. (2021). Spectrum anomaly detection for optical network monitoring using deep unsupervised learning. *IEEE Communications Letters*, 25(5), 1583–1586.
45. Wu, C., Shao, S., Tunc, C., & Hariri, S. (2020). Video anomaly detection using pre-trained deep convolutional neural nets and context mining. In *2020 IEEE/ACS 17th International Conference on Computer Systems and Applications (AICCSA)* (pp. 1–8). IEEE.
46. Hacoheh, G., & Weinshall, D. (2023). *Pruning the Unlabeled Data to Improve Semi-Supervised Learning*. *arXiv preprint arXiv:2308.14058*.
47. Tang, H., & Jia, K. (2022). Towards discovering the effectiveness of moderately confident samples for semi-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 14658–14667).
48. Imran, J., Jamil, F., & Kim, D. (2021). An ensemble of prediction and learning mechanism for improving accuracy of anomaly detection in network intrusion environments. *Sustainability*, 13(18), 10057.
49. Carlini, N. (2021). Poisoning the unlabeled dataset of semi-supervised learning. In *30th USENIX Security Symposium (USENIX Security 21)* (pp. 1577–1592).
50. Van Engelen, J. E., & Hoos, H. H. (2020). A survey on semi-supervised learning. *Machine Learning*, 109(2), 373–440.
51. Cui, M., Wang, J., & Yue, M. (2019). Machine learning-based anomaly detection for load forecasting under cyberattacks. *IEEE Transactions on Smart Grid*, 10(5), 5724–5734.
52. Şengönül, E., Samet, R., Abu Al-Haija, Q., Alqahtani, A., Alturki, B., & Alsulami, A. A. (2023). An analysis of artificial intelligence techniques in surveillance video anomaly detection: A comprehensive survey. *Applied Sciences*, 13(8), 4956.
53. Apostol, I., Preda, M., Nila, C., & Bica, I. (2021). IoT botnet anomaly detection using unsupervised deep learning. *Electronics*, 10(16), 1876.
54. Fahad, L. G., & Tahir, S. F. (2021). Activity recognition and anomaly detection in smart homes. *Neurocomputing*, 423, 362–372.
55. Nayak, R., Pati, U. C., & Das, S. K. (2021). A comprehensive review on deep learning-based methods for video anomaly detection. *Image and Vision Computing*, 106, 104078.
56. Hundman, K., Constantinou, V., Laporte, C., Colwell, I., & Soderstrom, T. (2018). Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 387–395).

57. Mathur, A. P., & Tippenhauer, N. O. (2016). SWaT: A water treatment testbed for research and training on ICS security. In *2016 International Workshop on Cyber-Physical Systems for Smart Water Networks (CySWater)* (pp. 31–36). IEEE.
58. Katser, I. D., & Kozitsin, V. O. (2020). Skoltech anomaly benchmark (SKAB). Available online: <https://www.kaggle.com/dsv/1693952> (accessed on 17 October 2024).
59. Wen, P., Yang, Z., Wu, L., Qi, S., Chen, J., & Chen, P. (2022). A novel convolutional adversarial framework for multivariate time series anomaly detection and explanation in cloud environment. *Applied Sciences*, 12(20), 10390.
60. Girish, L., & Rao, S. K. (2023). Anomaly detection in cloud environment using artificial intelligence techniques. *Computing*, 105(3), 675–688.
61. Zobaed, S. M., & Salehi, M. A. (2023). *Confidential Computing Across Edge-to-Cloud for Machine Learning: A Survey Study*. arXiv preprint arXiv:2307.16447.
62. Snehi, J., Snehi, M., Prasad, D., Simaiya, S., Kansal, I., & Baggan, V. (2022). SDN-based cloud combining edge computing for IoT infrastructure. In *Software Defined Networks: Architecture and Applications* (pp. 497–540).
63. Hiran, K. K., Doshi, R., Fagbola, T., & Mahrishi, M. (2019). *Cloud Computing: Master the Concepts, Architecture and Applications with Real-World Examples and Case Studies*. Bpb Publications.
64. Corradi, A., Fanelli, M., & Foschini, L. (2014). VM consolidation: A real case based on OpenStack cloud. *Future Generation Computer Systems*, 32, 118–127.
65. Airaj, M. (2017). Enable cloud DevOps approach for industry and higher education. *Concurrency and Computation: Practice and Experience*, 29(5), e3937.
66. Malburg, L., Klein, P., & Bergmann, R. (2020). *Using Semantic Web Services for AI-Based Research in Industry 4.0*. arXiv preprint arXiv:2007.03580.
67. Gokhroo, M. K., Govil, M. C., & Pilli, E. S. (2017). Detecting and mitigating faults in cloud computing environment. In *2017 3rd International Conference on Computational Intelligence & Communication Technology (CICT)* (pp. 1–9). IEEE.
68. Bello, S. A., Oyedele, L. O., Akinade, O. O., Bilal, M., Delgado, J. M. D., Akanbi, L. A., & Owolabi, H. A. (2021). Cloud computing in construction industry: Use cases, benefits and challenges. *Automation in Construction*, 122, 103441.
69. Shwetha, R., & Kirubanand, V. B. (2021). Remote monitoring of heart patients using robotic process automation (RPA). In *ITM Web of Conferences* (Vol. 37, p. 01002). EDP Sciences.
70. Imran, Iqbal, N., Ahmad, S., & Kim, D. H. (2021). Health monitoring system for elderly patients using intelligent task mapping mechanism in closed loop healthcare environment. *Symmetry*, 13(2), 357.
71. Bin Mofidul, R., Alam, M. M., Rahman, M. H., & Jang, Y. M. (2022). Real-time energy data acquisition, anomaly detection, and monitoring system: Implementation of a secured, robust, and integrated global IIoT infrastructure with edge and cloud AI. *Sensors*, 22(22), 8980.
72. Jan, Z., Ahamed, F., Mayer, W., Patel, N., Grossmann, G., Stumptner, M., & Kuusk, A. (2022). Artificial intelligence for industry 4.0: Systematic review of applications, challenges, and opportunities. *Expert Systems with Applications*, 119456.
73. Brecher, C., Buchsbaum, M., & Storms, S. (2019). Control from the cloud: Edge computing, services and digital shadow for automation technologies. In *2019*

- International Conference on Robotics and Automation (ICRA)* (pp. 9327–9333). IEEE.
74. Gleißner, M., Dotzler, J., Hartig, J., Aßmuth, A., Bulitta, C., & Hamm, S. (2021). IT security of cloud services and IoT devices in healthcare. *CLOUD COMPUTING* 2021, 10.
75. Rodríguez-Cuenca, B., García-Cortés, S., Ordóñez, C., & Alonso, M. C. (2015). Automatic detection and classification of pole-like objects in urban point cloud data using an anomaly detection algorithm. *Remote Sensing*, 7(10), 12680–12703.
76. Wróblewski, A., Wodecki, J., Trybała, P., & Zimroz, R. (2022). A method for large underground structures geometry evaluation based on multivariate parameterization and multidimensional analysis of point cloud data. *Energies*, 15(17), 6302.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

CHAPTER 3

Holistic Resource Management Based on AI and Sustainable Cloud Computing

VEENA PRASAD VEMURI,¹ BAIREDDY RAVINDER REDDY,²
VINEELA LAGADAPATI,³ NOOR FIRDOOS JAHAN,⁴ and
GOWRI BHARATH KUMAR⁵

¹*Commerce/Management, NKES College of Arts Commerce and Science, Mumbai University, Maharashtra, India*

²*IT, Sreenidhi Institute of Science and Technology, Hyderabad, Telangana, India*

³*St. Martins Engineering College, Hyderabad, Telangana, India*

⁴*RV Institute of Management, Bangalore, Karnataka, India*

⁵*Computer Science and Engineering, Chalapathi Institute of Engineering and Technology, Lam, Guntur, Andhra Pradesh, India*

ABSTRACT

The widespread demand for hosting application services in the cloud has been fueled by the deployment of cloud data centers (CDCs) on a global scale. Furthermore, modern apps' resource needs have sharply increased, especially in businesses that use a lot of data. Cloud providers give top importance to reducing the energy usage of servers within cloud computing networks in order to reduce costs related to operation and enhance service sustainability by concentrating applications onto less working servers. Cloud services are commonly duplicated among servers as a result of the providers' need for high levels of availability and

Advances in AI for Cloud, Edge, and Mobile Computing Applications.

Shrikaant Kulkarni and P. William (Eds.)

© 2025 Apple Academic Press, Inc. Co-published with CRC Press (Taylor & Francis)

dependability; it increases resource overhead and server energy utilization as well. The management of cloud resources must find equilibrium among service consolidation and replication to increase server availability and dependability while lowering energy consumption. These two goals might struggle. The artificial intelligence-based cuckoo optimization-based energy-reliability conscious resource scheduling approach (CRUZE) is a resource management solution for sustainable cloud computing. Utilizing the CloudSim toolset, we assess CRUZE's performance against the most advanced systems currently available.

3.1 INTRODUCTION

For facilitating and conveying programming answers for the overwhelming majority modern applications, cloud computing has secured itself as a trustworthy, savvy, and versatile computing administration choice [1]. To take advantage of the on-demand, flexible, and adaptable asset provisioning, which reduces capital expenditures and construction upkeep costs [2], many businesses have switched to CDCs. Most cloud providers give a broad assortment of sending decisions, allowing clients to fit resources for their necessities. Nonetheless, the extension of artificial intelligence and Internet of Things (IoT) applications in Industry 4.0 [3] has expanded the prerequisite for cloud assets all the more for the most part. Cloud service organizations like Amazon, Microsoft, and Google, specifically, have seen an expansion in the quantity of cloud server farms and a huge reliance on them [4].

For these cloud data centers to operate efficiently, a significant quantity of energy is needed. In particular, effective cloud resource management is required to cut energy usage, lower costs, and lessen carbon footprints. A significant portion of the energy used to meet this demand, which is often substantial, is used for cooling [2]. Nearly as much energy as the computer nodes themselves may be used by a CDC's cooling infrastructure [5]. It is a costly endeavor since major public cloud providers must make investments in expansive cooling infrastructures [6]. The creation of comprehensive energy-aware resource management models that take cooling and computing costs into account has been regarded as a significant open topic [6]. The deficiency of research emphasizes the requirement for workload planning for CDCs that takes energy, warm, and cooling costs into account as optimizing objectives [7, 8].

3.1.1 PROBLEM STATEMENT

It seeks to provide resource management for sustainable cloud computing due to the intricate relationships between energy usage, building upkeep, and the cooling system [9]. Another issue is the coordination of planning choices for grouped tasks that all the while ponder cooling power and dealing with power. While lowering energy usage, it's also important to consider how to get rid of hot or hazardous materials that might hinder framework function [10]. Due to the need to fulfill non-fixed requirements, cooling systems, and as a result, host power and temperature assessments, change dynamically in this way [11]. Likewise, there may be connections between a datacenter's operational activities. It is typical to assign jobs to a cloud environment, create service level agreements also known as SLAs for each activity, and regularly employ many uncontested initiatives for each task [12]. Then, rather than being evaluated for every project, difficult quick reaction time and SLA encroachment would be analyzed at the particular work level. Furthermore, the CPU, RAM, storage, and network resource capacities of the host systems in hybrid public-private clouds vary. These problems make scheduling much more difficult since now the system representation must additionally account for host heterogeneity and task interdependence across tasks.

3.1.2 MOTIVATION TO STUDY

In this examination, we fundamentally expand our previous work [13, 14] by introducing an all encompassing asset the board procedure for energy-limiting distributed computing utilizing computerized reasoning. The trouble of this issue might be settled utilizing heuristic procedures like the transformative calculation. The cuckoo optimization (CO) method was suggested for usage in this situation because it surpasses Particle Swarm Optimization also known as PSO and ACO also known as Ant Colony Optimization in terms of accuracy, speed, and integration when used to resolve batch process and task scheduling challenges [15, 16]. The fundamental goal of this study is to foster a Cuckoo improvement based powerful asset usage method called CRUZE for comprehensive administration of all assets (servers, organizations, capacity, cooling frameworks), upgrade energy viability, abbreviate carbon impression, and

lessening administration disappointment in equipment, administration, and so on. This study expands upon the conceptual frameworks suggested in [18–20]. Utilizing cloud assets; our technique bunches workloads into cluster in view of quality of service (QoS) qualities, and afterward utilizes a CRUZE to coordinate the assets to such an extent that the errand might be executed.

3.1.3 OBJECTIVE OF STUDY

1. Application consolidation onto fewer active servers to decrease server energy usage in cloud computing systems can save operational costs and boost service sustainability.
2. Cloud service providers must be provided with high levels of security in order to prevent the frequent duplication of cloud services across servers, which increase server energy consumption and overhead.

3.2 LITERATURE REVIEW

Since cloud users' needs are always changing, the management of cloud resources holistically is a challenging issue [21]. The majority of existing research [22, 23] exclusively considers servers' energy management, ignoring other components that consume a substantial amount of energy. This section compares the currently used resource management methods with our suggested strategy.

3.2.1 ENERGY – CONSCIOUS CLOUD RESOURCE PLANNING

Some research has proposed an EARS, or Energy-Aware Resource Scheduling, technique to execute workloads in virtual cloud environments [24]. Positions are seen as simply homogeneous by the EARS approach, which models the power and dissatisfaction profiles for CDC and performs them utilizing an event-based cloud test framework. The energy expenses of cloud server farms may be reduced with this strategy, and position fulfillment rates can be increased. In a similar vein, researchers [25] proposed a VM scheduling (VMS) method to increase customer satisfaction by reducing

server energy usage and determining how energy utilization affects SLA violation rates. For the management of IT infrastructure, researchers [26] presented a holistic approach (HA) to cut down on execution costs and energy consumption. Holistic workload scaling (HWS) is a strategy that devised to allow simultaneous scaling of resources vertically and horizontally while reducing latency without taking into account energy usage of cloud services [27]. A holistic energy optimization framework (HEOF) that depends on high figuring capacities was presented in examinations to at the same time lessen the warm effect and the expense of cooling [28]. They characterize energy utilization as an errand center task and booking issue. A server-based cloud-empowered design (SCA) was proposed by creators to expand the energy-effectiveness of a few equipment components, including computer processors, memory, and capacity [29]. Furthermore, a contextual investigation of a video following application is utilized to test SCA's presentation, and the consequences of the trials exhibit that SCA performs better as far as memory use. In a virtualization-based cloud datacenter, scientists introduced an all-encompassing Model for Asset The board HMRM to bring down the energy utilization of different assets including memory, stockpiling, and systems administration [30]. A comprehensive methodology for cloud administration provisioning (HACSP) was put out by certain examinations to address expected and unexpected changes in asset needs progressively while limiting energy costs [31]. As indicated by specialists, energy productive server farm the executives (EEDCM) approach was created under accessibility constraints [32]. They likewise examined the meaning of accessibility and planned a slope climbing calculation to stay away from disappointment zones.

3.2.2 SCHEDULING CLOUD RESOURCES WITH RELIABILITY IN MIND

Scientists [33] have suggested a method for improving cloud service reliability that addresses the double-dealing of organizational and capacity assets. The virtual machine's running state is stored using this approach, which makes use of service check pointing. Hub disappointment indicator is likewise made to utilize accessible organization assets. A united scattering based multi-cloud storage framework was made by studies to give a reliable, secure, and reasonable cloud administration

[34]. A few specialists recommended a Multi-Objective Asset Booking strategy to further develop the execution speed and reliability of cloud administrations [35]. A compromise between execution speed and reliability is likewise recognized by the creators for the successful execution of HPC (High Performance Computing) workload. To offer an issue the executives system, scientists introduced a Versatile and In the nick of time booking method utilizing spot and on-request examples [36]. Focuses on the Heterogeneous Spot Events based Auto-scaling approach [37] that is suggested for usage while running web applications under a trustworthy management. Researchers recommended a replication-based express the board method that actually remains conscious of diverse state fortifications on a few VMs while executing certified world and imitation streaming apps [38]. While booking assets for responsibility execution, concentrates on introduced the Powerful Bunching Association Title Approach to diminish issue decrease in task disappointment [39]. Researchers [40] suggested a Cloud Data Center Resource Optimization strategy. It makes a reasonable model to expand the exhibition measurements steadfastness and energy while booking assets.

3.2.3 SCHEDULING CLOUD RESOURCES WITH ENERGY AWARENESS USING CUCKOO OPTIMIZATION

A few studies suggested a GTCO or Group Technology based architecture and Cuckoo Streamlining approach to effectively organize VMs to cloud duties. When virtual machines are operational, GTCO reduces energy costs and outperforms cooperative effort-based asset planning [41]. To organize circulating figure tasks and increase energy productivity for the execution of homogenous exercises, experts promoted a Cuckoo Optimization based Task Scheduling, or COTC, technique [42]. Focuses on recommending the employment of a Cuckoo Optimization-based Resource Management (CORM) device for work wishing to enhance load change and reduce energy needs [43]. CORM beats cooperative effort-based asset planning for terms of energy proficiency while utilizing cloud assets. Algorithm based Task Scheduling the wellness esteem (execution season) of CSATS was just estimated by the creators with different probabilities to distinguish the best cloud asset for executing applications. A cuckoo search meta-heuristic (CSMH) approach was put out by certain examinations to

limit the energy utilization of cloud responsibilities [45]. Using CloudSim [46], the effectiveness of COTC [47] and CSMH [48] has been assessed.

3.3 METHODOLOGY

We provide and discuss each of the system's major components since our suggested method functions inside a comprehensive CDC [49]. Our method takes into account every level of cloud service provisioning.

3.3.1 SOFTWARE AS A SERVICE (SAAS)

The workload manager receives requests or user sessions from this layer, which manages entering user workloads (interactive or batch type). Contingent upon the meaning of the workload and any QoS necessities, for example, cutoff times and monetary limitations, the workload management keeps up with the responsibility line in a specific request.

3.3.2 PAAS OR PLATFORM AS A SERVICE

This layer establishes a controller to coordinate all of the system's activities. Three objectives guide the controller's efficient scheduling of the supplied cloud resources: i) Increase cloud datacenter stability and sustainability; ii) Reduce energy usage; and iii) Maximize resource use. The following five sub modules make up a controller (middleware) as well: VM resource supervisor, Fault supervisor, energy supervisor, cooling supervisor, and workload supervisor.

3.3.3 INFRASTRUCTURE AS A SERVICE (IAAS)

Information about cloud infrastructure, including VMs and CDCs, is included in the IaaS layer. The virtualization layer also permits task balancing via VM movement. The temperature changes of various VMs running on various centers are estimated, watched, and constrained by a proactive temperature-mindful scheduler. Power management units (PMUs) power and manage various components in cloud data centers.

By saving the current conditions of virtual machines, dynamic irregular access memory (Measure) gives a check-pointing system [50]. An uninterruptible power supply (UPS) is utilized to keep the lights on when the essential power sources fizzle. A chiller an energy administrator directs the age of energy from both sustainable and non-inexhaustible sources. Environmentally friendly power sources (sunlight based and wind) are progressively vigorously underlined in maintainable CDCs [14, 51].

3.3.4 MODEL DESIGN

- **Energy Model:** According to many studies, resource use and energy consumption are inversely correlated [1, 17, 54, 55, 57, 58].
- **Model for Reliability:** The capacity to offer the appropriate service is what makes a cloud service reliable [2, 9]. Model for capacity planning: Studies show that the memory, disk, and network usage at a specific time t describe the capacity model [61].
- **Model for Temperature:** To develop a temperature model for calculating datacenter temperature, we combined the computer room air conditioning (CRAC) model with the RC warm model [62–65]. Warm obstruction and intensity limit of the host are indicated by the letters R and C, individually.
- **Model for Renewable Energy:** The Energy Reuse Effectiveness as referred as ERE is calculated using a renewable energy model [66].
- **Model for Using Waste Heat:** The distribution proportion (RR) is resolved utilizing the locale warming administration-based waste heat utilization model [67].
- **Model of cooling management:** The efficacy or efficiency of datacenter cooling systems (DCS) is assessed using a water-based cooling management model [68].

3.3.5 PROVIDING OF RESOURCES

As shown in Table 3.1, a resource provisioning technique called Q-aware is utilized to distribute resources for clustered workload in accordance with the needs of various clusters.

TABLE 3.1 Workload Clustering Using K-Means

Cluster	Name of Cluster	Workloads
C1	Compute	Computer technology and performance assessment
C2	Storage	Services for online backup and storage
C3	Communication	Applications for the Internet, mobile computing services, and important websites
C4	Administration	Development and testing of software and projects, productivity tools with a graphical focus, financial services at the forefront, and software for projects and online transaction processing

3.3.6 ALGORITHM FOR RESOURCE ALLOCATION BASED ON CUCKOO OPTIMIZATION

Our suggested scheduling approach aims to reduce overall cloud energy use while enhancing system dependability. Consolidating these two objectives is in many cases seen as a compromise on the grounds that, albeit moving VMs to less dynamic server’s decreases framework energy utilization, server disappointment might adversely affect various VMs and brings down framework trustworthiness. Conflictingly, expanding the quantity of VM duplicates improves framework steadiness yet additionally brings about higher energy consumptions since additional servers are kept dynamic and required for handling. To convey financially savvy cloud benefits, a compromise between energy use and reliability is important to relieve this effect. In addition, the dependability of the system component is impacted by the frequent switching on and off of the server. Power tweak decreases the trustworthiness of server parts like memory and capacity. In order to decrease power usage with as little influence as possible on the dependability of cloud services, new energy-aware resource management strategies are required [69, 70].

The Cuckoo Optimization or CO algorithm is a resource scheduling method that is created to execute user workload while taking dependability and energy consumption into consideration. An objective function aims to complete all n tasks while simultaneously minimizing system energy use and increasing server dependability. The CO estimation, which takes on cuckoo qualities and the technique associated with laying eggs, is motivated by the cuckoo bird’s presence [60]. For tackling optimization issues like bunch process scheduling and occupation planning, the CO calculation

has been displayed to perform better compared to PSO and ACO with regards to exactness, speed, and assembly [56]. The CO algorithm is capable of both local and global search [52, 53]. It is recognized that the planning and execution of jobs on suitable cloud assets is a NP-complete issue, requiring the advancement of inventive asset booking calculations with the most significant levels of reliability and manageability of cloud administrations [51]. Because of its capacity to naturally plan assets for responsibility execution, generally straightforward reconciliation with customary enhancement strategies, and simplicity of change in a unique cloud climate, we have picked the CO calculation for booking provisioned assets. The CO-based resource scheduling algorithm is as follows:

Algorithm 1

Cuckoo optimization based resource scheduling algorithm.

```

Input: Set of Provisioned Resources
        Set of Workloads
Output: Execute Workloads
1. Start
2. Initialize Resource Set  $[R_1, R_2, \dots, R_{q_{wr}}]$ 
3. Evaluate Resource Utilization ( $R_U$ ) using Eq. (24) for every resource to determine its Profit
4. Rank the Resources ( $R_{U1} \geq R_{U2} \geq \dots \geq R_{Un}$ ) based on  $R_U$ 
5. Select Best Resource with maximum  $R_U$ 
6. Create instances of that resource  $R_{q_{wr}} = [I_1, I_2, \dots, I_k]$ 
7. Evaluate ELR for each resource using Eq. (33)
8. Each resource generates instances inside their corresponding ELR
9. Evaluate the Fitness Value ( $F$ ) for all instances using Eq. (23)
10. Choose instance which has high  $F$ 
11. Execute Workloads using selected instance of resource
12. if (All Workloads Executed Successfully == FALSE) then
13.   While do
14.     Continue Execution
15.     Monitor Performance
16.     if (Instances Required  $\geq$  Instances Provided) then
17.       While do
18.         Add New Stable Instance
19.         Calculate  $E_{consumption}$  and  $R_{Service}$ 
20.         if ( $E_{consumption} < TV_E$ ) then
21.           if ( $R_{Service} > TV_R$ ) then
22.             break
23.           else
24.             Declare Current Instance is Unstable
25.             Remove Unstable Instance
26.             continue
27.           else
28.             continue
29.         else
30.           continue
31.       else Stop

```

3.4 RESULT AND DISCUSSION

All of the studies used one of the four workloads listed in Table 3.2.

TABLE 3.2 Cloud Workload Details

Workload	Cluster	Explanation
Testing of performance	C1(Compute)	Cruze processes and changes a 713 MB JPG image file to a PNG. It is a workload (in the form of a Cloudlet) to convert one JPG picture into one PNG file.
Backup data and storage	C2 (Storage)	To store 5 TB of data and make a backup of the data is considered to be work.
Website	C3 (Communication)	Workload is defined as the volume of visitors to a university website during the admissions period.
Testing and development of software	C4 (Administration)	Workload refers to the design and testing of agri-info software that forecasts agricultural production.

The proposed technique for holistic resource management is assessed in light of various boundaries, including scope quantification, energy management, and waste heat utilization, energy from sustainable sources, virtualization, and cooling management. The 12 hour time frame for the temporal assessments includes 2500 workload submissions. The following various test scenarios are used to assess CRUZE’s performance:

- **Planning for Capacity and Renewable Energy:** Memory, disk, and network usage have all been considered as exhibition boundaries for scope quantification, and these metrics are expressed as rates (%). Figure 3.1 (a) shows how much memory is used when running the workloads. CRUZE uses less memory while processing the same number of jobs. Memory use in CRUZE is higher than it is in HMRM, CSRE, and CSMH by 23.68%, 24.35%, and 24.81%, respectively.

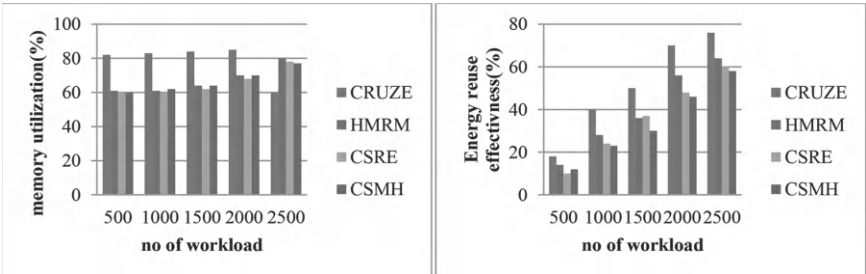


FIGURE 3.1 (a) Memory utilization (b) Energy reuses effectiveness.

As far as energy reuse for sustainable power, we surveyed CRUZE’s viability. The level of all out energy used by IT hardware that is utilized for cooling, lighting, and IT gadgets is known as energy reuse proficiency. Figure 3.1(b) displays the amount of renewable energy that was utilized again when performing various workloads.

- **Energy Management:** We examined CRUZE’s performance in regard to usage of energy in order to compute the value of usage of energy, which is expressed in kilowatt hours (kWh). Energy use fluctuates as workloads grow. According to Figure 3.2(a), CRUZE uses an average of 16.35%, 17.61%, and 19.12% less energy than other workloads. This is because CRUZE uses less energy and executes clustered workloads rather than discrete workloads, which reduces network traffic and switch count.

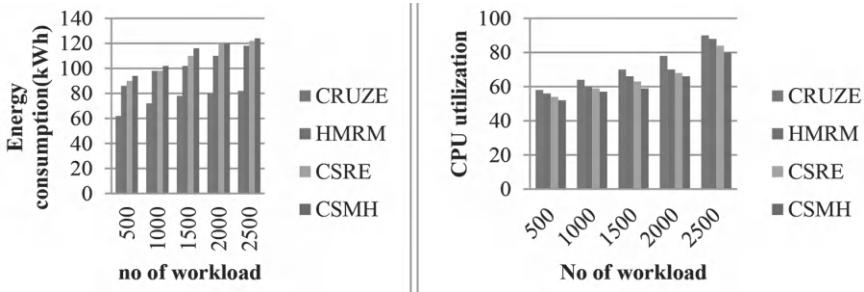


FIGURE 3.2 (a) Energy consumption (b) CPU utilization.

- **Virtualization:** As far as central processor use and VM Co-Area Cost, we surveyed CRUZE’s virtualization productivity. The change in computer chip use for the given workloads is displayed in Fig 3.2 (b). Since CRUZE employs provisioning methods to find the best resources, its average CPU consumption is greater than that of HMRM, CSRE, and CSMH, respectively, by 10.12%, 13.45%, and 14.69%.
- **Management of Cooling:** By measuring Cruze’s DCS Effectiveness, we evaluated its efficiency. The quantity of heat that can be removed from a system for each unit of energy needed to keep the CDC in a cool condition is measured by the DCS effectiveness. The variance in DCS Productivity for CRUZE, HMRM, CSRE,

and CSMH in connection to the accomplishment of various responsibilities is depicted in Figure 3.3. Because Cruze has a centralized heating management module for efficient cooling management, its average DCS Efficiency is greater than that of HMRM, CSRE, and Cruze, respectively, by 8.98%, 9.23%, and 10.56%.

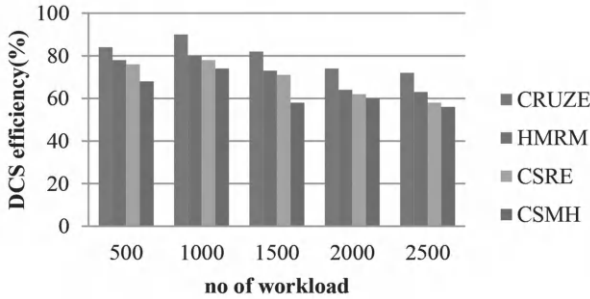


FIGURE 3.3 DCS efficiency.

- **Fault management and Application QoS:** In terms of fault tolerance, we examined CRUZE's reliability and fault detection rate. The ratio of identified hardware, software, and network problems or failures to all other faults or failures in the system is known as the fault detection rate.

Execution cost, execution time, and execution security have all been taken into account when determining an application's QoS. Execution cost, as per studies, is the aggregate sum of cash that may be spent on really running the program for 60 minutes. It is measured in Cloud Dollars Figure 3.4 shows the fault management and QoS application

- **Trade-Offs Between Several Performance Metrics:** Figure 3.5 illustrates the trade-off between energy consumption, reliability, and CPU utilization when CRUZE is used to carry out tasks. The value of CPU use and dependability are falling with rising energy consumption, whereas cloud service reliability is growing as CPU utilization rises. It is abundantly obvious that reliability and CPU use are negatively correlated with energy consumption, while reliability is correlated with CPU utilization. For these results, 35 resources and 2500 workloads were taken into account.

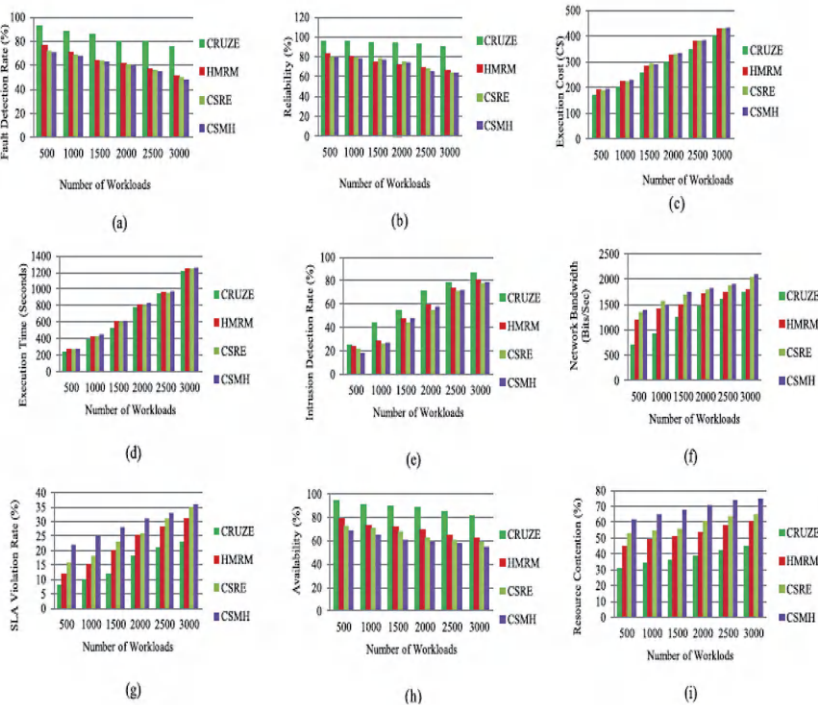


FIGURE 3.4 Algorithm comparisons are available for a number of metrics, including the rate of fault detection, cost, time, rate of intrusion detection, network bandwidth, availability, rate of SLA violations, and rate of resource contention.

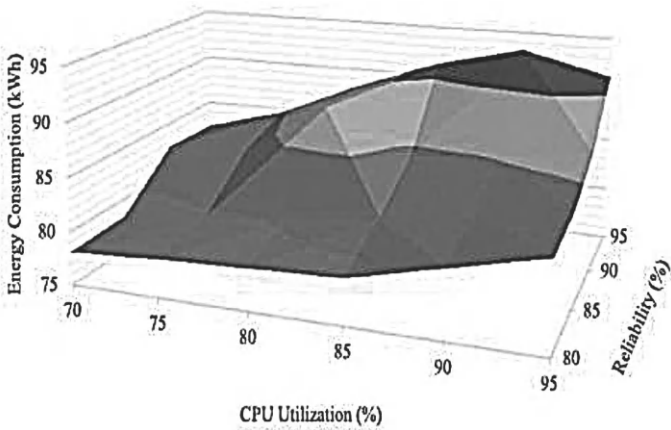


FIGURE 3.5 Energy use, dependability, and CPU use are all traded off.

3.6 CONCLUSION

We introduced the CO estimation based resource arranging strategy named CRUZE to increment energy productivity and decrease carbon impressions in cloud service provider while keeping up with cloud organization dependability. Dynamically managing failures is the foundation of this strategy. Furthermore, CRUZE orchestrates arranged assets for the execution of assorted jobs and changes the assets at runtime as per the QoS necessities of responsibilities, which might forestall or relieve asset under- and over-usage. Initial findings indicate that CRUZE reduces latency by 7.23%, execution costs by 14.64%, execution time by 11.15%, and energy reuse effectiveness by 19.65%. Additionally, CRUZE improves DCS Efficiency by 10.65%, reliability by 16.21%, intrusion detection rates by 19.36%, CPU utilization by 14.59%, memory utilization by 24.81%, disk utilization by 18.20%, network utilization by 11.21%, and memory utilization by 25.81%. At long last, the compromise between asset use, energy utilization, and trustworthiness for the execution of responsibilities is introduced.

KEYWORDS

- **artificial intelligence**
- **resource management**
- **CRUZE**
- **cloud computing**
- **cuckoo optimization**

REFERENCES

1. Bharany, S., Sharma, S., Khalaf, O. I., Abdulsahib, G. M., Al Humaimedy, A. S., Aldhyani, T. H., & Alkahtani, H. (2022). A systematic survey on energy-efficient techniques in sustainable cloud computing. *Sustainability*, 14(10), 6256.
2. Shuja, J., Gani, A., Shamshirband, S., Ahmad, R. W., & Bilal, K. (2016). Sustainable cloud data centers: A survey of enabling techniques and technologies. *Renewable and Sustainable Energy Reviews*, 62, 195–214.

3. Chun, K. W., Kim, H., & Lee, K. (2019). A study on research trends of technologies for Industry 4.0: 3D printing, artificial intelligence, big data, cloud computing, and internet of things. In *Advanced Multimedia and Ubiquitous Engineering: MUE/FutureTech 2018* (pp. 397–403). Springer Singapore.
4. Wankhede, P., Talati, M., & Chinchamatpure, R. (2020). Comparative study of cloud platforms-Microsoft Azure, Google Cloud Platform, and Amazon EC2. *J. Res. Eng. Appl. Sci.*, 5(02), 60–64.
5. Pakbaznia, E., & Pedram, M. (2009). Minimizing data center cooling and server power costs. In *Proceedings of the 2009 ACM/IEEE International Symposium on Low Power Electronics and Design* (pp. 145–150).
6. Gill, S. S., Garraghan, P., Stankovski, V., Casale, G., Thulasiram, R. K., Ghosh, S. K., & Buyya, R. (2019). Holistic resource management for sustainable and reliable cloud computing: An innovative solution to global challenge. *Journal of Systems and Software*, 155, 104–129.
7. Chang, C. W., Liu, C. Y., & Yang, C. Y. (2017). Energy-efficient heterogeneous resource management for wireless monitoring systems. *Journal of Systems and Software*, 131, 168–180.
8. Taherizadeh, S., Jones, A. C., Taylor, I., Zhao, Z., & Stankovski, V. (2018). Monitoring self-adaptive applications within edge computing frameworks: A state-of-the-art review. *Journal of Systems and Software*, 136, 19–38.
9. Gill, S. S., & Buyya, R. (2018). A taxonomy and future directions for sustainable cloud computing: 360° view. *ACM Computing Surveys (CSUR)*, 51(5), 1–33.
10. Akbari, A., Khonsari, A., & Ghoreyshi, S. M. (2020). Thermal-aware virtual machine allocation for heterogeneous cloud data centers. *Energies*, 13(11), 2880.
11. Chaudhry, M. T., Ling, T. C., Manzoor, A., Hussain, S. A., & Kim, J. (2015). Thermal-aware scheduling in green data centers. *ACM Computing Surveys (CSUR)*, 47(3), 1–48.
12. Mastelic, T., Oleksiak, A., Claussen, H., Brandic, I., Pierson, J. M., & Vasilakos, A. V. (2014). Cloud computing: Survey on energy efficiency. *ACM Computing Surveys (CSUR)*, 47(2), 1–36.
13. Li, X., Jiang, X., Garraghan, P., & Wu, Z. (2018). Holistic energy and failure aware workload scheduling in cloud datacenters. *Future Generation Computer Systems*, 78, 887–900.
14. Guitart, J. (2017). Toward sustainable data centers: A comprehensive energy management strategy. *Computing*, 99(6), 597–615.
15. Rajabioun, R. (2011). Cuckoo optimization algorithm. *Applied Soft Computing*, 11(8), 5508–5518.
16. Yang, X. S. (2014). Swarm intelligence-based algorithms: A critical analysis. *Evolutionary Intelligence*, 7, 17–28.
17. Li, X., Garraghan, P., Jiang, X., Wu, Z., & Xu, J. (2017). Holistic virtual machine scheduling in cloud datacenters towards minimizing total energy. *IEEE Transactions on Parallel and Distributed Systems*, 29(6), 1317–1331.
18. Buyya, R., & Gill, S. S. (2018). *Sustainable Cloud Computing: Foundations and Future Directions*. arXiv preprint arXiv:1805.01765.
19. Shahidinejad, A., Ghobaei-Arani, M., & Masdari, M. (2021). Resource provisioning using workload clustering in cloud computing environment: A hybrid approach. *Cluster Computing*, 24(1), 319–342.

20. Gill, S. S., & Buyya, R. (2018). Failure management for reliable cloud computing: A taxonomy, model, and future directions. *Computing in Science & Engineering*, 22(3), 52–63.
21. Singh, S., & Chana, I. (2016). A survey on resource scheduling in cloud computing: Issues and challenges. *Journal of Grid Computing*, 14, 217–264.
22. Singh, S., & Chana, I. (2015). Q-aware: Quality of service-based cloud resource provisioning. *Computers & Electrical Engineering*, 47, 138–160.
23. Natu, M., Ghosh, R. K., Shyamsundar, R. K., & Ranjan, R. (2016). Holistic performance monitoring of hybrid clouds: Complexities and future directions. *IEEE Cloud Computing*, 3(1), 72–81.
24. Li, M., Qin, C., & Lee, P. P. (2015). {CDStore}: Toward Reliable, Secure, and {Cost-Efficient} Cloud Storage via Convergent Dispersal. In *2015 USENIX Annual Technical Conference (USENIX ATC 15)* (pp. 111–124).
25. Qiu, Y., Jiang, C., Wang, Y., Ou, D., Li, Y., & Wan, J. (2019). Energy aware virtual machine scheduling in data centers. *Energies*, 12(4), 646.
26. Balis, B., Brzoza-Woch, R., Bubak, M., Kasztelnik, M., Kwolek, B., Nawrocki, P., & Zielinski, K. (2018). Holistic approach to management of IT infrastructure for environmental monitoring and decision support systems with urgent computing capabilities. *Future Generation Computer Systems*, 79, 128–143.
27. Pérez, J. F., Chen, L. Y., Villari, M., & Ranjan, R. (2018). Holistic workload scaling: A new approach to compute acceleration in the cloud. *IEEE Cloud Computing*, 5(1), 20–30.
28. Luo, C., Yang, L. T., Li, P., Xie, X., & Chao, H. C. (2015). A holistic energy optimization framework for cloud-assisted mobile computing. *IEEE Wireless Communications*, 22(3), 118–123.
29. Liu, B., Chen, Y., Blasch, E., Pham, K., Shen, D., & Chen, G. (2014). A holistic cloud-enabled robotics system for real-time video tracking application. In *Future Information Technology: FutureTech 2013* (pp. 455–468). Springer Berlin Heidelberg.
30. Guzek, M., Kliazovich, D., & Bouvry, P. (2013). A holistic model for resource representation in virtualized cloud computing data centers. In *2013 IEEE 5th International Conference on Cloud Computing Technology and Science (Vol. 1, pp. 590–598)*. IEEE.
31. Ferrer, A. J., Hernández, F., Tordsson, J., Elmroth, E., Ali-Eldin, A., Zsigri, C., & Sheridan, C. (2012). OPTIMIS: A holistic approach to cloud service provisioning. *Future Generation Computer Systems*, 28(1), 66–77.
32. Sitarum, D., Phalachandra, H. L., Gautham, S., Swathi, H. V., & Sagar, T. P. (2015). Energy efficient data center management under availability constraints. In *2015 Annual IEEE Systems Conference (SysCon) Proceedings* (pp. 377–381). IEEE.
33. Zhou, A., Wang, S., Zheng, Z., Hsu, C. H., Lyu, M. R., & Yang, F. (2014). On cloud service reliability enhancement with optimal resource usage. *IEEE Transactions on Cloud Computing*, 4(4), 452–466.
34. Singh, P., Agarwal, N., & Raman, B. (2018). Secure data deduplication using secret sharing schemes over cloud. *Future Generation Computer Systems*, 88, 156–167.
35. Azimzadeh, F., & Biabani, F. (2017). Multi-objective job scheduling algorithm in cloud computing based on reliability and time. In *2017 3rd International Conference on Web Research (ICWR)* (pp. 96–101). IEEE.

36. Poola, D., Ramamohanarao, K., & Buyya, R. (2016). Enhancing reliability of workflow execution using task replication and spot instances. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, 10(4), 1–21.
37. Qu, C., Calheiros, R. N., & Buyya, R. (2016). A reliable and cost-efficient auto-scaling system for web applications using heterogeneous spot instances. *Journal of Network and Computer Applications*, 65, 167–180.
38. Liu, X., Harwood, A., Karunasekera, S., Rubinstein, B., & Buyya, R. (2017). E-storm: Replication-based state management in distributed stream processing systems. In *2017 46th International Conference on Parallel Processing (ICPP)* (pp. 571–580). IEEE.
39. Abdulhamid, S. I. M., Abd Latiff, M. S., Madni, S. H. H., & Abdullahi, M. (2018). Fault tolerance aware scheduling technique for cloud computing environment using dynamic clustering algorithm. *Neural Computing and Applications*, 29, 279–293.
40. Zhang, P., Zhou, M., & Wang, X. (2020). An intelligent optimization method for optimal virtual machine allocation in cloud data centers. *IEEE Transactions on Automation Science and Engineering*, 17(4), 1725–1735.
41. Shahdi-Pashaki, S., Teymourian, E., Kayvanfar, V., Komaki, G. M., & Sajadi, A. (2015). Group technology-based model and cuckoo optimization algorithm for resource allocation in cloud computing. *IFAC-PapersOnLine*, 48(3), 1140–1145.
42. Sundarraj, R., & Vasudevan, V. (2017). An optimization algorithm for task scheduling in cloud computing based on multi-purpose cuckoo seek algorithm. In *Theoretical Computer Science and Discrete Mathematics: First International Conference, ICTCSDM 2016, Krishnankoil, India, December 19–21, 2016, Revised Selected Papers* (pp. 415–424). Springer International Publishing.
43. Abbasi, M. J., & Mohri, M. (2016). Scheduling tasks in the cloud computing environment with the effect of Cuckoo optimization algorithm. *SSRG International Journal of Computer Science and Engineering (SSRG-IJCSE)*, 3(8), 1–9.
44. Navimipour, N. J., & Milani, F. S. (2015). Task scheduling in the cloud computing based on the cuckoo search algorithm. *International Journal of Modeling and Optimization*, 5(1), 44–47.
45. Madni, S. H. H., Abd Latiff, M. S., & Abdulhamid, S. I. M. (2017). Optimal resource scheduling for IaaS cloud computing using cuckoo search algorithm. *Journal of Network and Computer Applications*, 77, 79–93.
46. Teixeira Sá, T., Calheiros, R. N., & Gomes, D. G. (2014). CloudReports: An extensible simulation tool for energy-aware cloud computing environments. In *Cloud Computing: Challenges, Limitations and R&D Solutions* (pp. 127–142). Springer International Publishing.
47. Agarwal, M., & Srivastava, G. M. S. (2018). A cuckoo search algorithm-based task scheduling in cloud computing. In *Advances in Computer and Computational Sciences: Proceedings of ICCCS 2016, Volume 2* (pp. 293–299). Springer Singapore.
48. Abdulhamid, S. I. M., Abd Latiff, M. S., Abdul—Salaam, G., & Madni, S. H. H. (2017). Optimal resource scheduling for IaaS cloud computing using cuckoo search algorithm. *International Journal of Grid and High-Performance Computing*, 9(2), 70–89.
49. Gill, S. S., & Buyya, R. (2019). Resource provisioning-based scheduling framework for execution of heterogeneous and clustered workloads in clouds: From fundamental to autonomic offering. *Journal of Grid Computing*, 17, 385–417.

50. Gill, S. S., Chana, I., Singh, M., & Buyya, R. (2019). RADAR: Self-configuring and self-healing in resource management for enhancing quality of cloud services. *Concurrency and Computation: Practice and Experience*, 31(1), e4834.
51. Li, X., Jiang, X., & He, Y. (2014). Virtual machine scheduling considering both computing and cooling energy. In *2014 IEEE International Conference on High Performance Computing and Communications, 2014 IEEE 6th International Symposium on Cyberspace Safety and Security, 2014 IEEE 11th International Conference on Embedded Software and Systems* (pp. 244–247). IEEE.
52. Yang, X. S., Deb, S., Zhao, Y. X., Fong, S., & He, X. (2018). Swarm intelligence: Past, present and future. *Soft Computing*, 22(18), 5923–5933.
53. Cuong-Le, T., Minh, H. L., Khatir, S., Wahab, M. A., Tran, M. T., & Mirjalili, S. (2021). A novel version of Cuckoo search algorithm for solving optimization problems. *Expert Systems with Applications*, 186, 115669.
54. Park, S., Park, S., & Park, Y. B. (2018). An architecture framework for orchestrating context-aware IT ecosystems: A case study for quantitative evaluation. *Sensors*, 18(2), 562.
55. Gill, S. S., & Buyya, R. (2018). SECURE: Self-protection approach in cloud resource management. *IEEE Cloud Computing*, 5(1), 60–72.
56. Deb, K. (2019). Constrained multi-objective evolutionary algorithm. In *Evolutionary and Swarm Intelligence Algorithms* (pp. 85–118). Springer, Cham.
57. Dewangan, B. K., Agarwal, A., Venkatadri, M., & Pasricha, A. (2019). Energy-aware autonomic resource scheduling framework for cloud. *International Journal of Mathematical, Engineering and Management Sciences*, 4(1), 41–54.
58. Möbius, C., Dargie, W., & Schill, A. (2013). Power consumption estimation models for processors, virtual machines, and servers. *IEEE Transactions on Parallel and Distributed Systems*, 25(6), 1600–1614.
59. Bilal, K., Khalid, O., Malik, S. U. R., Khan, M. U. S., Khan, S. U., & Zomaya, A. Y. (2016). Fault tolerance in the cloud. In *Encyclopedia of Cloud Computing* (pp. 291–300). Wiley.
60. Tsipianitis, A., & Tsompanakis, Y. (2020). Improved Cuckoo Search algorithmic variants for constrained nonlinear optimization. *Advances in Engineering Software*, 149, 102865.
61. Kouki, Y., & Ledoux, T. (2012). SLA-driven capacity planning for cloud applications. In *4th IEEE International Conference on Cloud Computing Technology and Science Proceedings* (pp. 135–140). IEEE.
62. Wang, L., Khan, S. U., & Dayal, J. (2012). Thermal aware workload placement with task-temperature profiles in a data center. *The Journal of Supercomputing*, 61(3), 780–803.
63. Zhang, S., & Chatha, K. S. (2007). Approximation algorithm for the temperature-aware scheduling problem. In *2007 IEEE/ACM International Conference on Computer-Aided Design* (pp. 281–288). IEEE.
64. Tang, Q., Gupta, S. K. S., & Varsamopoulos, G. (2008). Energy-efficient thermal-aware task scheduling for homogeneous high-performance computing data centers: A cyber-physical approach. *IEEE Transactions on Parallel and Distributed Systems*, 19(11), 1458–1472.

65. Lazic, N., Boutilier, C., Lu, T., Wong, E., Roy, B., Ryu, M. K., & Imwalle, G. (2018). Data center cooling using model-predictive control. In *Advances in Neural Information Processing Systems* (Vol. 31).
66. Tschudi, B., Vangeet, O., Cooley, J., & Azevedo, D. (2010). ERE: A metric for measuring the benefit of reuse energy from a data center. *White Paper*, 29, 82–83.
67. Karellas, S., & Braimakis, K. (2016). Energy–exergy analysis and economic investigation of a cogeneration and trigeneration ORC–VCC hybrid system utilizing biomass fuel and solar power. *Energy Conversion and Management*, 107, 103–113.
68. Liu, Z., Chen, Y., Bash, C., Wierman, A., Gmach, D., Wang, Z., & Hyser, C. (2012). Renewable and cooling aware workload management for sustainable data centers. In *Proceedings of the 12th ACM SIGMETRICS/PERFORMANCE Joint International Conference on Measurement and Modeling of Computer Systems* (pp. 175–186). ACM.
69. Sharma, Y., Javadi, B., Si, W., & Sun, D. (2016). Reliability and energy efficiency in cloud computing systems: Survey and taxonomy. *Journal of Network and Computer Applications*, 74, 66–85.
70. Garraghan, P., Moreno, I. S., Townend, P., & Xu, J. (2014). An analysis of failure-related energy waste in a large-scale cloud environment. *IEEE Transactions on Emerging Topics in Computing*, 2(2), 166–180.

PART II

**Computational Intelligence:
Edge AI Security and Privacy**



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

CHAPTER 4

Application of Artificial Intelligence for Enhancement of Privacy and Security in Smart Environment

ANANDA RAVURI,¹ KHUSHBOO MALIK,² KONDA HARI KRISHNA,³
ABHIJEET DAS,⁴ RIPON BHATTACHARJEE,⁵
JOSHUVA AROCKIA DHANRAJ⁶

¹*Intel Corporation, Hillsboro, Oregon, USA*

²*School of Law, CHRIST (Deemed to be University), Delhi–NCR Campus, India*

³*Department of CSE, School of Computing, Mohan Babu University, Tirupati, Andhra Pradesh, India*

⁴*Department of Civil Engineering, C.V. Raman Global University (CGU), Bhubaneswar, Odisha, India*

⁵*National Law University, Tripura, India*

⁶*Center for Automation and Robotics (ANRO), Department of Mechatronics Engineering, Hindustan Institute of Technology and Science, Padur, Chennai, Tamil Nadu, India*

ABSTRACT

With the advancement of urbanization, city public privacy and security are growing increasingly crucial. City public privacy and security is not just the basis of environmental growth, but also the primary assurance for the long-term security of inhabitants' existence. The idea of a smart

Advances in AI for Cloud, Edge, and Mobile Computing Applications.

Shrikaant Kulkarni and P. William (Eds.)

© 2025 Apple Academic Press, Inc. Co-published with CRC Press (Taylor & Francis)

environment is continually being explored in the present-day artificial intelligence (AI) setting. Urban public privacy and security have also introduced novel issues and difficulties. To objective of this paper is to employ AI innovation to create an effective public privacy and security data resources management platform in a smart environment. One of the main objectives of the study of AI was to allow systems to execute complicated activities that would ordinarily need human intellect. In this paper, the neural network (NN) data processing technique was used. As indication statistics, variables impacting urban public privacy and security were evaluated. The feed forward back-propagation NN (BPNN) was utilized in this study to anticipate indices statistics in real-time, which enabled the administration tasks of risk surveillance and initial alert of public privacy and security data indications to be accomplished. The outcomes indicated that the BPNN method had a mean prediction performance of roughly 89% for indication prediction, which was 16.1% points greater than the conventional NN method. The BPNN method's mean hazard alert accuracy pace was 90.3%, which was 16.5% greater than that of the typical NN method. This demonstrates that the BPNN method used in this study, which employs AI innovation, can perform advanced alerting of threats and administration of urban public privacy and security more effectively and precisely.

4.1 INTRODUCTION

The practice of building cities smartly has become more sophisticated in the past few decades. Additionally, more and more study is being done on smart environment. A smart city has been completely computerized, connected, and informative. It additionally utilizes the brains of people for independent thinking and reasoning [1–5].

Municipalities, healthcare, property management, transit, public utilities, and public privacy and security are among the more integrated, effective, and intelligent sectors in smart cities [6]. It offers citizens improved assistance for their personal and professional lives, as well as improving the environment in which businesses can grow. Additionally, it makes government administration and operations more effective [7–9]. The creation of a smart environment has seen some early success thanks to the “smart” idea’s widespread adoption. There has been an improvement in individual living circumstances.

Businesses have discovered business prospects from it as executives' understanding of managerial data has gotten better and more precise [8, 10–12].

On the fundamental question of a “smart environment,” nevertheless, researchers have differing opinions. Individuals describe and categorize “smart environment” in the context of technological advances and finances, according to studies on the topic. A smart environment should, from a technological perspective, employ rational and scholarly information and communication technologies (ICT) to fully use the network's potential for bridging the technological gap and promoting the sharing of data [14–18]. To make decision-making processes more empirical and reasonable in the administration of towns and the environment, the societal risk evaluation of participative smart environment development is applied. To make decision-making processes more empirical and reasonable in the administration of towns and the environment, the societal risk evaluation of participative smart environment development is applied [19]. At the same point, it is encouraged to distribute and use natural assets, societal wealth, and human resources scientifically [20]. A smart environment is defined from the standpoint of urban strategy as a structured and methodical expenditure on novel ICT for societal growth and security. A tangible way to support urbanization is through the data of a smart environment. Its main characteristic is the ability to emerge in a new manner and circumstance [21–23].

The connection between smart environments and ideas like smart cities, digital cities, ecological towns, and urban sustainability can be easily discovered, according to pertinent studies on smart environments [24–26]. A crucial connection in the adoption of information technology across a range of industries is the development of a smart environment. The primary distinction between two “smart environments” is that one's basic smart environment is based on the “intelligence of things.” The usage of smart assistance and innovations is a key component of a smart environment. Individuals come first is a key component of smart city planning [27–31]. An expansion of the digital globe is a computerized metropolis. Digital earth growth has been aided by the advancement of geographical data systems.

The risk to public privacy and security posed by smart environments, based on their present stage of development, is growing daily, which has

led to societal dissatisfaction and chaos. Citizens, developers, and executives are the end users as a smart city is developed. The development of smart environments is now being hampered by variables including the poor overall standard of citizens, a flawed talent framework, and managerial decision-making prejudice [32–36]. The peculiarities of large cities' corporate structures in which innovation and talent are in increasing supply. It is a complicated geographical location due to the large level of variables including demographic dimension, demographic density, societal movement, and complication [37]. There are several concerns associated with the present pace of urbanization in the globe from the viewpoint of societal growth and transformation. Urban security measures, for instance, fall short of ideal. Environmental privacy and security and urban security are not assured [38]. The growth of a smart environment is at a societal phase that poses a high danger from a geographical standpoint. To accomplish this, it is necessary to create a data resource management (DRM) framework that can comprehend urban public security and privacy at any moment and be utilized for risk evaluation and timely notification of elements that could damage urban public privacy and security in the future [39–41]. The administration capacities of city risk evaluation and timely alert can be efficiently realized by the smart environment public privacy and security administration platform built using a BPNN [42], according to this paper.

The novel aspect of this research is the development of an innovative DRM framework and the suggestion of a technique for data prediction employing the BPNN paradigm. This paper summarizes the drawbacks of conventional studies in use and employs retrospective reasoning to evaluate relevant sources. The NN framework is repeatedly trained in the empirical section of section 2, in the meantime. The evaluation findings of the BPNN paradigm and the conventional NN paradigm are contrasted in section 3's empirical section, and the impact is more pronounced.

4.2 LITERATURE REVIEW

The fundamental assurance for individuals to live and function in tranquility and happiness is urban public privacy and security. Certain academics have undertaken similar studies about the administration of urban public privacy and security [43]. Among them, Editorial E suggested a multiple-sensor imaging platform usage for urban public privacy and

security, namely a smart gadget subdivision for target recognition powered by AI, which could be utilized to trigger surveillance of targets [44]. To protect urban public privacy and security, Arbuckle deployed communication technologies to alert the public about impending fires and evacuation instructions [45].

Paul created an urban privacy and security communications network to improve city administration and increase public safety [46]. In the setting of a smart environment, Oliveira and Santos looked at the variables influencing the adoption of data release for security purposes [47]. The majority of the research listed above is conceptual, even though they have helped to some degree in the development of urban public privacy and security. There are not many distinct deployment and usage techniques. Thus, more study is required.

To do this, certain researchers suggested using AI technology to handle the data resources for public privacy and security in smart environments. To handle resources effectively in portable multiple-user radio frequency-optical diverse networks, Wu et al. [48] suggested an AI-based system.

From a conceptual viewpoint to a systematic standpoint, Lin and Zhao examined the latest developments in AI administration of resources, taking into account not just radio resource administration but also a variety of other kinds of materials [49]. The primary exploratory instruments of AI and machine learning (ML), according to Qi and Yao's proposal, might be included in hardware acceleration technologies and employed, among other things, for urban crisis security evacuation [50]. Although the expenses of these techniques are expensive, they are highly practicable and can translate concepts into particular uses. It needs more development to be a useful application. More advancement is thus necessary.

4.3 METHODOLOGY

4.3.1 URBAN PUBLIC PRIVACY AND SECURITY

Urban public privacy and security administration has grown in importance as urban public privacy and security administration has accelerated [51]. The management of societal security threats worldwide has received increasing emphasis as a result of the ongoing growth of smart environment economies, administration, and civilization. The risk management framework for urban public privacy and security worldwide is still in its

early stages [52]. They are a long way beyond the demands of financial inclusion, both in concept and in action. To construct a reasonably developed DRM framework, it is thus required to thoroughly investigate the risk supplies, risk control, and risk mitigation strategies of urban public privacy and security [53]. To set up a mix of identification, latency, and reaction operations, technological avoidance of public safety administration and avoidance is primarily dependent on security avoidance innovation and the growth of person-centered preventive activities.

Four broad groups of public risks might occur in urban areas: natural catastrophes, unintentional catastrophes, the propagation of the influenza virus, and terrorist acts. There are several gaps in urban public privacy and security, as demonstrated by several difficulties with urban security in the past decade, including bird flu, life-threatening acute respiratory diseases, and several terrorist acts [54]. According to an examination of societal hazards worldwide, there are currently issues with transforming institutions hazards, data stream and management hazards, environment, abuse, diversity, economic, and life hazards [56]. Nevertheless, based on the way that public security is now managed, there is still work that has to be done in terms of urgent medical care, and this approach to management is getting worse [57].

Analyzing the variables that influence the public safety of the city is required to set up a system of governance [58]. Metropolitan environmental safety, utilization of resources security, water grade, food, resource, and energy are particular regions where factor evaluation of urban public privacy and security must be conducted [59]. The self-control and self-discipline of metropolitan inhabitants are just as important to public privacy and security in the environment as official administration. Secondly, the term “safety” typically relates to the security of an individual’s everyday activities and also the security of populated areas, such as the security of fire, traffic, and other hazards [60]. Urban public privacy and security administration can be broken down into crisis public security administration, public healthcare crisis administration, medical crisis administration, and emergency administration of public facilities. In addition, the extent of administration of urban public privacy and security, fire and protection, inspection and administration, and healthcare can all be seen as major elements impacting urban public security [61].

After comprehending the factors that influence the state of urban security, ways to construct an effective urban privacy and security framework may be investigated. The urban privacy and security administration

platform described in this paper consists primarily of a crisis platform, an initial alert platform, and a safety educational framework. The deployment of these systems focuses primarily on manufacturing, food, building, fire, and natural catastrophe surveillance [61].

Some accomplishments have been made in worldwide urban crisis management in recent decades. Nonetheless, when confronted with novel duties and chances, there are nevertheless numerous issues since it does not recognize that the danger itself is a significant risk [63]. A vast number of large-scale, qualitative, and quantitative research have been conducted in the risk evaluation and evaluation of urban management organizations. There is nevertheless a significant void in the areas of smart surveillance and observation employing big data, cloud-based computation, detectors, and other data innovations. When a risk to public security administration arises, no matter how good and timely the response, it is bound to result in some damage to individual and material protection [64, 65]. As a result, the most efficient strategy for dealing with public security-related events is to build a security and risk-management framework that keeps public risks within individual authority.

4.3.2 DRM

The world's desire for smart environment design is currently increasing. Nevertheless, present privacy and safety-related DRM approaches are incapable of meeting the requirements of urban growth, particularly the study of smart environment technologies and dangers. Although certain near-ideal ideas and methodologies exist, they have not been thoroughly examined from the platform and other perspectives. The present study does not differentiate between the threats of urban data and the smart environment. Although there is a link between the two, the origins and presentations of each are distinct [66].

The public privacy and security administration of businesses continues to evolve in an empirical orientation with the advancement of technologies, science, and finance. Risk identification is conducted using a probability technique, and efficient action and administration are implemented.

In China, the management of public privacy and security has a century-long practice. The initial step involves the creation and use of the public privacy and security administration concept. Public privacy and security administration and novel initiatives are given top priority by the authorities.

Secondly, as Western urbanism accelerated in the early 20th era, the incorporation of public privacy and security administration concepts into businesses increased. A considerable amount of caution has been used in the public privacy and security administration of corporation finances and employees. Insurance is the primary tool for managing the privacy and security of the public. Public privacy and security administration has evolved into an academic discipline that is increasingly appreciated by professionals and academics [67, 68].

To achieve this, this paper examines the basis of public privacy and security statistics initially. The administration of resources is continued after that. The study has revealed that the primary sources of public privacy and security data for smart environments are climatic surveillance data, geographical study statistics, data from immigration inspections, data from managing wildfires, data from reconnaissance flights, and data from metropolitan floor surveillance networks. For this purpose, these data may be examined and handled in the administration of public privacy and security statistics. It is possible to transform the particular measurements and analyze the dangers that are included in them [69].

Numerous privacy and security administration demands are included in the DRM framework built in this paper, which is depicted in Figure 4.1. The illustration describes the structure, from data response to precise functioning and data processing. Realizing the significance of data is the ultimate objective. Among these, standardizing risk forecasting, early notification, and smart city monitoring and assessment processes are significant advancements in the world’s security risk administration [70].

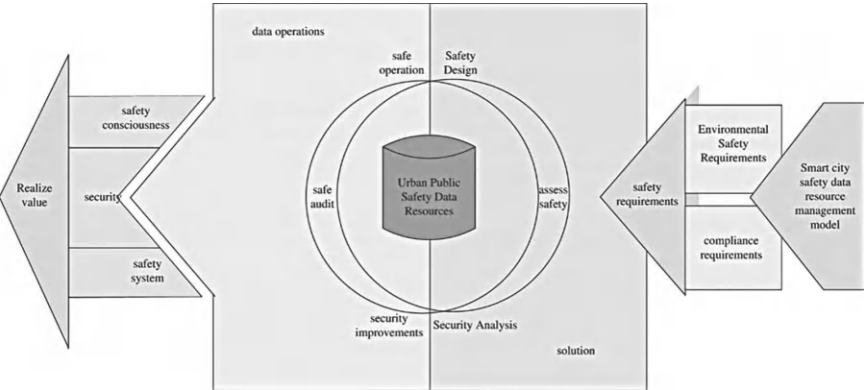


FIGURE 4.1 DRM.

Modern design, a networked financial system, and cutting-edge innovations come together to create smart environments, which promote development, household wealth, education, and improved societal privacy and security. Additionally, a smart environment would enhance data management structures, connected settlements, growth, labor force, revenue, and societal functions. Randomization has been demonstrated to be effective when a relationship to a dangerous ingredient is greater than a particular limit. Although a smart environment provides its citizens with extra benefits, its privacy and security have been jeopardized as a result of its dependence on data. By fusing smart layouts, networked populations, and cutting-edge technology, smart environments were intended to promote growth, employment, financial stability, training, and increasing societal security.

Smart environments employ data and modern innovations to make better choices that lead to advancement. Businesses can monitor issues to understand how market trends change and take action with the least expensive and most efficient solutions using the most accurate data in real time. Smart innovations typically create novel innovations that generate cash thanks to innovative attempts.

IoT solutions encourage collaboration with governmental agencies and other nonprofit organizations to effectively generate a single source of income. Conversely, smart environments could enable efficient resource and city-wide use while reducing damaging assaults. There are numerous uses for AI. The construction of “smart cities” generates vast volumes of data through smart devices and networks. As an outcome, gathering vast amounts of data could present a security risk that could end in breaches of confidentiality or criminal activity. Wireless sensor networks commonly offer dynamic data sharing for smart environment installations and research a multifunctional management strategy. The platforms depend on AI, which supports a framework that is secure, effective, and easily accessible.

Cyber risks and criminality are not obstacles to technical advancement or the prospect of utilizing ICT to its fullest potential. AI operational platforms could be used by Smart Environment services to monitor and display their worldwide findings in real time. Some Smart Environments may not be suitable for the cloud [71].

It needs to be modified in a cloud-based setting to satisfy specifications, and intelligence that is crucial for financial development must be added.

Data also needs to be described properly. The technological problem in digital societies is made worse by several factors, including unpredictable excessive power use, industrial attackers, and frontier exhaustion. To assess smart environments, a lot of data is needed [72].

Data Networks in the Cloud have employed solid-state discs to eliminate duplicate data and encrypted data transmission in their data points. Compared to physical network structures, cloud-based services typically offer more budgetary adaptability. A smart environment would be essential for enhancing the stability and prosperity of the worldwide economy.

Real-time monitoring of the city's geographical and meteorological factors is possible via the meteorological alert. The device would autonomously sound an alert, for instance, in the event of impending flooding, a typhoon, or a major earthquake. The system may also keep track of and interpret satellite signals during natural catastrophic events like landslides, fires, and floods and then sound an alert so that rescuers can arrive in time [73]. The environment above the city may be identified at any point during unmanned aerial automobile reconnaissance. To avoid unexpected hostile assaults or flight mishaps, it is possible to keep an eye out for foreign airplanes in the air at the same period. Certain troublesome neighborhoods may be monitored round-the-clock employing the city patrol robotic. Additionally, it can keep an eye on a few opportunistic spots, which helps the city's degree of public security be guaranteed more successfully.

Data classification, prediction, and other applications of BPNN are common. To identify and forewarn risk factors that can threaten public privacy and security in the future, this paper employs the BPNN method to anticipate the public security data of a smart environment.

4.3 RESULTS AND DISCUSSION

4.3.1 SYSTEM DEVELOPMENT PROCEDURES

Table 4.1 displays the development setting for the smart environment public privacy and security of the DRM platform developed in this paper. Table 4.1 shows that Python is being employed as the programming tool and that the Linux operating system is being employed. Python was selected since it is simple to grasp, compile, and completely functioning. The data supply for the trial training and referral is chosen to be the MySQL dataset.

TABLE 4.1 Context of system development.

	Development Context	Information's
D4	Dataset	MySQL
D3	Instrument	PyCharm
D2	Language	Python
D1	Computation platform	Linux XP

Additionally, this paper employs Python to keep track of local meteorological and traffic information to improve environment public privacy and security. The following are the traffic security surveillance digits:

```
import numpy as np
import pandas as pd
data_origin = pd.read_csv("180508.csv," header = None, encoding = 'GBK,'
names = [ "card,""date,""time," "station," "way," "money," "transaction. nature"],
dtype = dict(zip(['3,4,6],[ "category"]*3)))
data_origin.head()
data_stations = pd.read_csv("station.csv,"encoding = 'GBK')
data_stations.head()
```

The information from the primary traffic train stations in the metropolis may be collected using the above techniques, allowing for the realization of functions such as traffic data surveillance or disaster initial alerts.

Additionally, the following are the codes for surveillance of urban air conditions that were prepared for this paper:

```
N = 24
theta = np.linspace (0.0, 2*np.pi, N,endpoint = False)
radii = a24h
width = 2*np.pi/N*2
ax = plt. subplot (111, polar = True)
ax.set_ theta_ direction ( -1)
ax. set_ theta_ offset(np.pi/2.0)
ax.set_ xticks (np. Linspace (e, 2*np.pi, 24, endpoint = False))
ax. set_ xticklabels (range (24))
bars = ax.bar(theta, radii, width = width,bottom = 0.0)
r_ max = radii.max()
for r,bar in zip(radii, bars).
bar. set_ face color (plt. cm. jet (r**2/r max-*2))
bar. set_ alpha (0.5)
Plt.show()
```


This technique allows for the 24-hour surveillance of urban air quality. Additionally, it can provide initial notice when levels of urban smog, or PM_{2.5} particles, rise above those during routine events or natural catastrophes, enhancing urban public privacy and security.

4.3.2 EMPIRICAL FRAMEWORK OF SYSTEM EVALUATION

The BPNN system must be trained on statistics before being tested because it is employed in this paper for predicting statistics. This paper separates the data sets into training and evaluation sets to achieve this. In Table 4.2, the pertinent variable settings for the 2 groups are displayed. There are 50 training sites and 600 data sets in the training category 100 items and 20 test sites make up the evaluation set.

TABLE 4.2 Variable of System Evaluations

Set	Training	Evaluation
No. of sites	50	20
Sample book	600	100
Network	4 G	4 G
Power	40 W	30 W

Lastly, Table 4.3 of this paper’s risk evaluation indications is chosen, primarily focusing on the indications for urban data security, city environmental security, moral awareness among residents, and urban diplomacy security. The BPNN method is used to forecast these indicators’ readings. To fulfill the initial alert role, the system would sound an alert if the forecasted level were to potentially surpass the crucial threshold.

TABLE 4.3 Risk Evaluation Indications

Code	1st Level Indications	2nd Level Indications
F4	Diplomacy security	Diplomacy level
F3	Moral literacy	Moral consciousness and education degrees
F2	Environment security	Tracking the state of the environment The degree of traffic control on roads
F1	Data security	Level of communication facilities Security of data and encrypted

The training sets are split into two groups in this paper so that the BPNN method can have an acceptable level of training effectiveness 300 sets of data is taken every period. The outcomes of the two trainings are depicted in Figure 4.2, which compares the training data effectiveness with the conventional NN system. In the initial training, the accuracy of the conventional NN system is just 63.2% at the smallest level, 75% at the best, and only 68.4% on average, as shown in Figure 4.2(a). The BPNN system has the smallest accuracy (79.3%) and best accuracy (82.9%). The average is roughly 81% or 12.6% more than the conventional NN method. This, nevertheless, falls short of the evaluation’s requirements for accuracy. Figure 4.2(b) shows that, following the 2nd training, the BPNN system’s accuracy ranges from 83.3–93.8%. The mean score is 88.7%, which satisfies the standards of the fundamental evaluation outcomes. This demonstrates the way the BPNN system has a higher accuracy and a rapid rate of training.

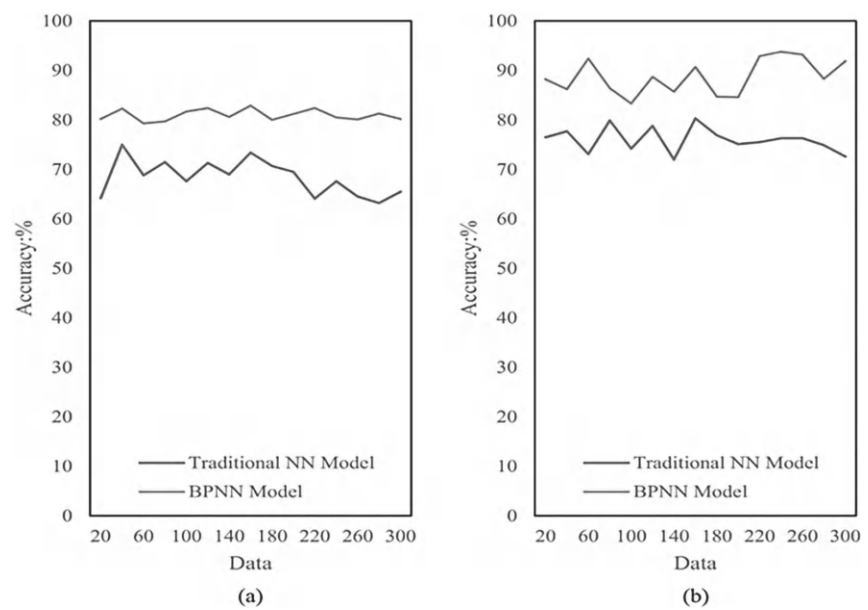


FIGURE 4.2 Outcomes of training data; (a) Initial training outcome; and (b) the 2nd training outcome.

It is then put through a thorough data evaluation process following the BPNN system has reached a reasonable level of accuracy. The model’s prediction outcomes are outlined by computing and forecasting each

indicator of the public privacy and security resource management platform. Figure 4.3 compares the findings with the classic NN approach in terms of predicting accuracy and the right proportion of risk alerts.

The accuracy of the BPNN framework for the prediction of indications is not below 85%, as can be observed in Figure 4.3(a). The average accuracy of the predictions is roughly 89%, and the greatest one is 93.6%. Just 79% and at its lowest, 66.3%, of the conventional NN system’s predictions are accurate. In comparison to the BPNN system, the mean accuracy is just 72.9%, a 16.1% decrease. The conventional NN system in Figure 4.3(b) has a risk alert accuracy that spans from 68.8–77.5%, with a mean of 73.8%.

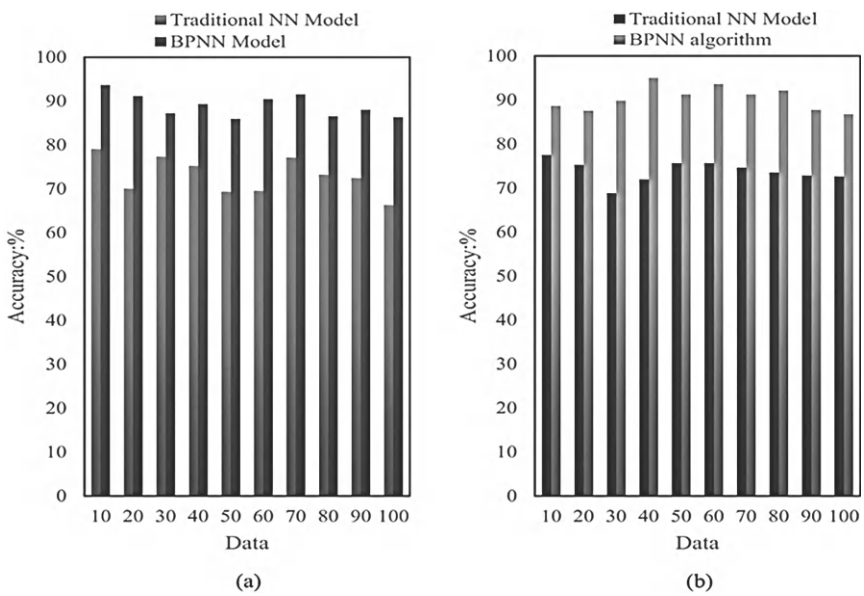


FIGURE 4.3 Risk evaluation capacity, (a) Indexes predicting accuracy and (b) risk alert accuracy percentage.

86.7–95%, with a mean of 90.3%, is the right proportion of risk alert for the BPNN system, which is 16.5% greater than that of the conventional NN system. Overall, this paper tests the data management platform using AI innovation’s BPNN system, and it produces positive outcomes. It demonstrates that the public privacy and security management platform for smart environments that was built in this paper has a particular level of consistency and dependability.

4.3.3 DEPLOYMENT AND CALCULATION FEASIBILITY OF OUTCOMES

The implementation of a smart environment platform that makes use of AI, IoT, and ML innovations is dependent on several variables, such as the accessibility and dependability of data, the system's adaptability, the needed computational strength and storing capability, the privacy and security consequences of the innovation, and the financial viability of the approach. The usage of IoT gadgets and detectors to gather real-time statistics from diverse sources can produce a huge volume of data that necessitates substantial computer resources to handle and evaluate. To train and optimally utilize the designs, ML methods need a lot of computing and storage supplies. However, they may be utilized to derive valuable information from this data. While this study paper addresses a few of these issues; others will need to be explored in subsequent studies. To manage massive batches of data and allow for real-time evaluation, the construction of a smart environment platform has been developed with scaling and adaptability at its core in this thorough smart environment assessment study. Furthermore, there are some difficulties with implementing adequate privacy and security safeguards that may be resolved in future studies to safeguard the data of city residents and avoid illegal accessibility.

4.3.4 FUTURE SUGGESTIONS FOR SMART STUDIES

For the future growth and application of deep learning (DL) in smart environments, the following suggestions may be useful:

Make DL methods more reliable and effective: Smart environments must successfully implement DL methods. Thus, there is a demand for ongoing study and design of more reliable and effective DL techniques that can manage huge quantities of data and offer precise predictions and observations.

1. **Privacy and Security Prioritization:** There is a possibility that private data will be made public as interconnected smart environments grow. The design and application of DL in smart environments must thus give great priority to privacy and security. Secure encrypted data may be used to do this, and only permitted individuals should have access to the data.

Collaboration with regional authorities and populations is essential to the performance of DL in smart environments. This is why it is crucial to involve regional authorities and populations in the growth and deployment of new technologies. This can make certain that innovation is developed to address the unique demands and issues of the surrounding area.

Create interfaces that are simple to use because DL innovation may be complicated and challenging for novices to comprehend. Hence, it's critical to provide easy-to-use interfaces that make it simple for local government employees and citizens to browse the data produced by DL methods [62].

2. **Prioritize Sustainability:** DL innovation may assist smart environments in improving resource allocation and minimizing trash. Nevertheless, it is critical to guarantee that the innovation is employed in a manner that stresses sustainability and minimizes environmental effects. This may be accomplished by employing renewable sources of energy and constructing smart environment systems with power effectiveness at their core [13].
3. **Considering the Legal Ramifications:** DL innovation has far-reaching societal and legal ramifications. As a result, it is critical to evaluate the legal implications of employing DL in smart environments and to guarantee that the innovation is employed in a manner that is appropriate and equitable for all citizens. This may be accomplished by encouraging openness and inclusiveness in the design and deployment processes.

4.4 CONCLUSION

Urbanization will quickly expand the number of people living in cities in the future. In some ways, the city's growth has been hampered by overpopulation. The city's economy has grown significantly as a result of the populace growth, which is a favorable growth. On the downside, the rise in the urban populace has inevitably increased the usage of energy, the strain on the environment, facilities, public assets, and so on, particularly in terms of public privacy and security, leading to increasing concealed hazards. The development of a more effective administration system is

therefore crucial in the current AI period. The initial stage in this study was to develop a pertinent concept of “smart environment,” which revealed that the term encompassed more profound implications than smart and “information-driven” cities. Additionally, this paper established a public privacy and security administration system and described the many public threats in smart environments. After that, the information’s source for public privacy and security was explained. By the data source, a DRM platform has been created. This paper evaluated the system’s capacity for risk evaluation using the BPNN paradigm to comprehend how the system’s administration impact would be affected. The findings indicated that the BPNN method had an accurate degree of threat alert that was 16.5% greater than the classic NN method. The data management platform in this paper was tested using the BPNN paradigm, and the outcomes were promising. However, there are nevertheless a lot of problems with this paper. Among these, public privacy and security management studies and the use of AI innovation both lack sufficient depth and maturity. So, there’s nevertheless opportunity for development.

KEYWORDS

- **artificial intelligence**
- **privacy**
- **security**
- **smart environment**
- **neural networks**

REFERENCES

1. Yin, C., Xiong, Z., Chen, H., Wang, J., Cooper, D., & David, B. (2015). A literature survey on smart cities. *Science China Information Sciences*, 58(10), 1–8.
2. Lai, C. S., Jia, Y., Dong, Z., Wang, D., Tao, Y., Lai, Q. H., Wong, R. T., Zobaa, A. F., Wu, R., & Lai, L. L. (2020). A review of technical standards for smart cities. *Clean Technologies*, 2(3), 290–310.
3. Kim, H., Choi, H., Kang, H., An, J., Yeom, S., & Hong, T. (2021). A systematic review of the smart energy conservation system: From smart homes to sustainable smart cities. *Renewable and Sustainable Energy Reviews*, 140, 110755.

4. Ghazal, T. M., Hasan, M. K., Alshurideh, M. T., Alzoubi, H. M., Ahmad, M., Akbar, S. S., Al Kurdi, B., & Akour, I. A. (2021). IoT for smart cities: Machine learning approaches in smart healthcare—A review. *Future Internet*, 13(8), 218.
5. Ahad, M. A., Paiva, S., Tripathi, G., & Feroz, N. (2020). Enabling technologies and sustainable smart cities. *Sustainable Cities and Society*, 61, 102301.
6. Javed, A. R., Shahzad, F., ur Rehman, S., Zikria, Y. B., Razzak, I., Jalil, Z., & Xu, G. (2022). Future smart cities: Requirements, emerging technologies, applications, challenges, and future aspects. *Cities*, 129, 103794.
7. Ismagilova, E., Hughes, L., Rana, N. P., & Dwivedi, Y. K. (2020). Security, privacy, and risks within smart cities: Literature review and development of a smart city interaction framework. *Information Systems Frontiers*.
8. Ghazal, T. M., Hasan, M. K., Alzoubi, H. M., Al Hammadi, M., Al-Dmour, N. A., Islam, S., Kamran, R., & Mago, B. (2022). Securing smart cities using blockchain technology. In *2022 1st International Conference on AI in Cybersecurity (ICAIC)* (pp. 1–4). IEEE.
9. Guo, Y. M., Huang, Z. L., Guo, J., Li, H., Guo, X. R., & Nkeli, M. J. (2019). Bibliometric analysis on smart cities research. *Sustainability*, 11(13), 3606.
10. Razmjoo, A., Gandomi, A. H., Pazhoohesh, M., Mirjalili, S., & Rezaei, M. (2022). The key role of clean energy and technology in smart city development. *Energy Strategy Reviews*, 44, 100943.
11. Heidari, A., Navimipour, N. J., & Unal, M. (2022). Applications of ML/DL in the management of smart cities and societies based on new trends in information technologies: A systematic literature review. *Sustainable Cities and Society*.
12. Singh, T., Solanki, A., Sharma, S. K., Nayyar, A., & Paul, A. (2022). A decade review on smart cities: Paradigms, challenges and opportunities. *IEEE Access*.
13. Ma, C. (2021). Smart city and cyber-security: Technologies used, leading challenges and future recommendations. *Energy Reports*, 7, 7999–8012.
14. Herath, H. M., & Mittal, M. (2022). Adoption of artificial intelligence in smart cities: A comprehensive review. *International Journal of Information Management Data Insights*, 2(1), 100076.
15. Mall, P. K., Narayan, V., Pramanik, S., Srivastava, S., Faiz, M., Sriramulu, S., & Kumar, M. N. (2023). FuzzyNet-based modelling smart traffic system in smart cities using deep learning models. In *Handbook of Research on Data-Driven Mathematical Modeling in Smart Cities* (pp. 76–95). IGI Global.
16. Blasi, S., Ganzaroli, A., & De Noni, I. (2022). Smartening sustainable development in cities: Strengthening the theoretical linkage between smart cities and SDGs. *Sustainable Cities and Society*, 80, 103793.
17. Dana, L. P., Salamzadeh, A., Hadizadeh, M., Heydari, G., & Shamsoddin, S. (2022). Urban entrepreneurship and sustainable businesses in smart cities: Exploring the role of digital technologies. *Sustainable Technology and Entrepreneurship*, 1(2), 100016.
18. Pandya, S., Srivastava, G., Jhaveri, R., Babu, M. R., Bhattacharya, S., Maddikunta, P. K., Mastorakis, S., Piran, M. J., & Gadekallu, T. R. (2023). Federated learning for smart cities: A comprehensive survey. *Sustainable Energy Technologies and Assessments*, 55, 102987.

19. Duygan, M., Fischer, M., Pärli, R., & Ingold, K. (2022). Where do smart cities grow? The spatial and socio-economic configurations of smart city development. *Sustainable Cities and Society*, 77, 103578.
20. van Twist, A., Ruijter, E., & Meijer, A. (2023). Smart cities & citizen discontent: A systematic review of the literature. *Government Information Quarterly*.
21. Javed, A. R., Ahmed, W., Pandya, S., Maddikunta, P. K., Alazab, M., & Gadekallu, T. R. (2023). A survey of explainable artificial intelligence for smart cities. *Electronics*, 12(4), 1020.
22. Zeng, S., Hu, Y., & Llopis-Albert, C. (2023). Stakeholder-inclusive multi-criteria development of smart cities. *Journal of Business Research*, 154, 113281.
23. Ullah, F., & Al-Turjman, F. (2023). A conceptual framework for blockchain smart contract adoption to manage real estate deals in smart cities. *Neural Computing and Applications*, 35(7), 5033–5054.
24. Kuo, Y. H., Leung, J. M., & Yan, Y. (2023). Public transport for smart cities: Recent innovations and future challenges. *European Journal of Operational Research*, 306(3), 1001–1026.
25. Alshamaila, Y., Papagiannidis, S., Alsawalqah, H., & Aljarah, I. (2023). Effective use of smart cities in crisis cases: A systematic review of the literature. *International Journal of Disaster Risk Reduction*, 80, 103521.
26. Savastano, M., Suci, M. C., Gorelova, I., & Stativă, G. A. (2023). How smart is mobility in smart cities? An analysis of citizens' value perceptions through ICT applications. *Cities*, 132, 104071.
27. Khanh, Q. V., Nguyen, V. H., Minh, Q. N., Van, A. D., Le Anh, N., & Chehri, A. (2023). An efficient edge computing management mechanism for sustainable smart cities. *Sustainable Computing: Informatics and Systems*, 38, 100867.
28. Silva, B. N., Khan, M., & Han, K. (2018). Towards sustainable smart cities: A review of trends, architectures, components, and open challenges in smart cities. *Sustainable Cities and Society*, 38, 697–713.
29. Al Nuaimi, E., Al Neyadi, H., Mohamed, N., & Al-Jaroodi, J. (2015). Applications of big data to smart cities. *Journal of Internet Services and Applications*, 6(1), 1–5.
30. Letaifa, S. B. (2015). How to strategize smart cities: Revealing the SMART model. *Journal of Business Research*, 68(7), 1414–1419.
31. Kumar, T. V., & Dahiya, B. (2017). Smart economy in smart cities. In *Smart Economy in Smart Cities* (pp. 3–76). Springer.
32. Yang, L., Elisa, N., & Eliot, N. (2019). Privacy and security aspects of e-government in smart cities. In *Smart Cities Cybersecurity and Privacy* (pp. 89–102). Elsevier.
33. Braun, T., Fung, B. C., Iqbal, F., & Shah, B. (2018). Security and privacy challenges in smart cities. *Sustainable Cities and Society*, 39, 499–507.
34. Chen, J., Ramanathan, L., & Alazab, M. (2021). Holistic big data integrated artificial intelligent modeling to improve privacy and security in data management of smart cities. *Microprocessors and Microsystems*, 81, 103722.
35. Moustaka, V., Theodosiou, Z., Vakali, A., Kounoudes, A., & Anthopoulos, L. G. (2019). Enhancing social networking in smart cities: Privacy and security borderlines. *Technological Forecasting and Social Change*, 142, 285–300.
36. Elmaghraby, A. S., & Losavio, M. M. (2014). Cyber security challenges in smart cities: Safety, security, and privacy. *Journal of Advanced Research*, 5(4), 491–497.

37. Al-Turjman, F., Zahmatkesh, H., & Shahroze, R. (2022). An overview of security and privacy in smart cities' IoT communications. *Transactions on Emerging Telecommunications Technologies*, 33(3), e3677.
38. Vattapparamban, E., Güvenç, I., Yurekli, A. I., Akkaya, K., & Uluagaç, S. (2016). Drones for smart cities: Issues in cybersecurity, privacy, and public safety. In *2016 International Wireless Communications and Mobile Computing Conference (IWCMC)* (pp. 216–221). IEEE.
39. Mousa, S. K., & Othman, M. (2020). The impact of green human resource management practices on sustainable performance in healthcare organizations: A conceptual framework. *Journal of Cleaner Production*, 243, 118595.
40. Hong, C. H., & Varghese, B. (2019). Resource management in fog/edge computing: A survey on architectures, infrastructure, and algorithms. *ACM Computing Surveys (CSUR)*, 52(5), 1–37.
41. Sinaeepourfard, A., García Almiñana, J., Masip Bruin, X., & Marín Tordera, E. (2017). Fog-to-Cloud (F2C) data management for smart cities. In *Proceedings of 2017 Future Technologies Conference (FTC): 29–30 November 2017, Vancouver, Canada* (pp. 162–172). The Science and Information (SAI) Organization.
42. Cilimkovic, M. (2015). Neural networks and backpropagation algorithm. *Institute of Technology Blanchardstown*, 15(1).
43. Zhuravleva, N. A., Cadge, K., Poliak, M., & Podhorska, I. (2019). Data privacy and security vulnerabilities of smart and sustainable urban space monitoring systems. *Contemporary Readings in Law and Social Justice*, 11(2), 56–62.
44. Editorial, B. (2021). Correction: Big data and development of smart city: System architecture and practical public safety example (vol. 17, no. 3, pp. 337–355, 2020). *Serbian Journal of Electrical Engineering*, 18(1), 137.
45. Arbuckle, J. (2019). Broadband deployment, public safety and defending local control. *Western City*, 95(6), 3–7.
46. Paul, S. (2017). Public safety communications systems: The future is smart. *Public Safety Communications*, 83(3), 14–15.
47. Oliveira, V., & Santos, G. D. (2019). Information technology acceptance in public safety in smart sustainable cities: A qualitative analysis. *Procedia Manufacturing*, 39, 1929–1936.
48. Wu, Z. Y., Ismail, M., Serpedin, E., & Wang, J. (2021). Artificial intelligence for smart resource management in multi-user mobile heterogeneous RF-light networks. *IEEE Wireless Communications*, 28(4), 152–158.
49. Lin, M., & Zhao, Y. (2020). Artificial intelligence-empowered resource management for future wireless communications: A survey. *China Communications*, 17(3), 58–77.
50. Qi, L., & Yao, K. (2021). Artificial intelligence enterprise human resource management system based on FPGA high-performance computer hardware. *Microprocessors and Microsystems*, 82, 103876.
51. Dooley, K. (2021). Direct passive participation: Aiming for accuracy and citizen safety in the era of big data and the smart city. *Smart Cities*, 4(1), 336–348.
52. Liaqat, M., Chang, V., Gani, A., Hamid, S. A., Toseef, M., & Shoaib, U. (2017). Federated cloud resource management: Review and discussion. *Journal of Network and Computer Applications*, 77, 87–105.

53. Garcia, A. A., & Dantec, C. (2018). Quotidian report: Grassroots data practices to address public safety. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW), 1–18.
54. Browne, K. (2018). Public safety and wildfire resilience. *Western City*, 94(10), 19–20, 31.
55. Zhao, Q., Han, S., Mao, Y., Leng, S., Min, G., Hu, J., et al. (2017). A Markovian analytical framework for public-safety video sharing by device-to-device communications. *Concurrency and Computation: Practice and Experience*, 29(16), 1–18.
56. Gutek, D. (2018). Improvement of safety in public transport using image server. *AUTOBUSY–Technika Eksploatacja Systemy Transportu*, 19(12), 82–85.
57. Zhao, Z. (2021). Community public safety evaluation system based on location information service architecture. *Mobile Information Systems*, 2021, 1–10.
58. Quinlan, A. (2020). Visions of public safety, justice, and healing: The making of the rape kit backlog in the United States. *Social & Legal Studies*, 29(2), 225–245.
59. Wan, S., Lu, J., & Fan, P. (2017). To smart city: Public safety network design for emergency. *IEEE Access*, 6, 1451–1460.
60. Valle, A. D., & Acciaio, E. (2017). SISMA Salt Lake City public safety building, Salt Lake City. *Costruzioni Metalliche*, 69(2), 45–58.
61. Leggio, W., & Snodgrass, A. (2017). Small town, big commitment: Public safety agencies in Nebraska City, Neb. worked together to solve their own. *JEMS: A Journal of Emergency Medical Services*, 42(1), 38–40.
62. Atitallah, S. B., Driss, M., Boulila, W., & Ghézala, H. B. (2020). Leveraging deep learning and IoT big data analytics to support smart cities development: Review and future directions. *Computer Science Review*, 38, 100303.
63. Kirkpatrick, L. O., & Breznau, N. (2016). The (non) politics of emergency political intervention: The racial geography of urban crisis management in Michigan. Available at SSRN 2754128.
64. Fasenfest, D. (2019). A neoliberal response to an urban crisis: Emergency management in Flint, MI. *Critical Sociology*, 45(1), 33–47.
65. Makui, A., Ashouri, F., & Barzinpour, F. (2019). Assignment of injuries and medical supplies in urban crisis management. *Journal of Applied Research on Industrial Engineering*, 6(3), 232–250.
66. Bill, R. (2017). The future of public safety communications. *Radio User*, 12(6), 12–14.
67. Kaplan-Dobbs, M., Kattan, J. A., & Tuazon, E. (2020). Increasing access to buprenorphine in safety-net primary care clinics: The New York City Buprenorphine Nurse Care Manager Initiative. *American Journal of Public Health*, 111(2), e1–e4.
68. Cottrill, J. (2017). City council divided on safety approach. *OHS Canada*, 33(3), 13.
69. Henkey, T. (2017). *Urban Emergency Management: Planning and Response for the 21st Century*. Butterworth-Heinemann.
70. Patel, S., Ranjbar, M., & Cummins, T. C. (2022). Safety and inner-city neighborhood change: Student and teacher perspectives. *Education and Urban Society*, 54(3), 227–248.
71. Allam, Z., & Dhunny, Z. A. (2019). On big data, artificial intelligence, and smart cities. *Cities*, 89, 80–91.

72. Su, Y., Gao, W., Guan, D., & Zuo, T. A. (2020). Achieving urban water security: A review of water management approaches from a technology perspective. *Water Resources Management*, 34, 4163–4179.
73. Molnar, A. (2015). The geo-historical legacies of urban security governance and the Vancouver 2010 Olympics. *The Geographical Journal*, 181(3), 235–241.

CHAPTER 5

Application of Artificial Intelligence in Wireless Sensor Network Through Quality of Services Enhancement

TAMIRAT TAGESSE TAKORE,¹ SYED JAHANGIR BADASHAH,²
MOHIT TIWARI,³ A. PRABHU,⁴ VENKANNA MOOD,⁵ and
P. WILLIAM⁶

¹Department of Electrical and Computer Engineering, Wachemo University, India

²ECE Department, Sreenidhi Institute of Science and Technology, Hyderabad, Telangana, India

³Department of Computer Science and Engineering, Bharati Vidyapeeth's College of Engineering, Paschim Vihar, Delhi, India

⁴Electrical and Electronics Engineering, K. Ramakrishnan College of Engineering, Trichy, Tamil Nadu, India

⁵ECE, St. Martin's Engineering College, Secunderabad, Telangana, India

⁶Research and Development, Department of Information Technology, Sanjivani College of Engineering, SPPU, Pune, Maharashtra, India

ABSTRACT

Wireless sensor networks (WSNs) have evolved as a transformative technique with widespread applications, ranging from environmental monitoring to industrial automation. Ensuring quality of service (QoS) in

WSNs is paramount for reliable and efficient data collection and transmission. This research paper delves into the innovative integration of artificial intelligence (AI) techniques into WSNs to enhance QoS and elevate the performance of these networks. One key aspect of this research is the utilization of AI algorithms for dynamic routing and data aggregation. By incorporating machine learning models, WSNs can adapt their routing paths based on real-time conditions, mitigating congestion, and minimizing energy consumption. Additionally, data aggregation techniques using AI enable more accurate data fusion while preserving network resources. In conclusion, this research paper showcases the potential of Artificial Intelligence in revolutionizing WSNs by enhancing their QoS. Through dynamic routing, data aggregation, predictive maintenance, energy-efficient protocols, and security reinforcement, AI empowers WSNs to meet the demanding requirements of modern applications. The findings of this study provide valuable insights into the future of WSNs and their pivotal role in the internet of things (IoT) era, paving the way for more reliable and efficient deployments across diverse domains.

5.1 INTRODUCTION

Wireless Sensor Networks have become a key technology in the IoT era, where billions of devices are networked to enable smooth communication and data sharing. These networks are made up of countless tiny sensor nodes with limited resources that work together to collect data from their surroundings and send it to a centralized base station or sink node. WSNs are used in a variety of industries, such as manufacturing automation, smart cities, medical care, and environmental monitoring. The effectiveness of WSNs, nevertheless, heavily relies on the QoS they can provide, which includes elements like dependability, latency, energy efficiency, and scalability. The wireless sensor is primarily made up of a network system linked by different types of homes. It is currently primarily an evolving field, encompassing a variety of domestic and international domains with a rather high level of scientific excitement [9]. Embedded computer information is the world's most cutting-edge networking innovation. Processing technology can identify multiple monitoring methods using numerous sensors, and gather information and data, and then wirelessly communicate it to each user's hands [10]. There are various technical

configurations for wireless sensor networks. As an illustration, certain organizations exhibit traits like spontaneity, fault tolerance, speed, and small size [11]. The potential uses for wireless sensors in military and urban administration are likewise highly varied [12]. Diverse technical fields have greatly impacted transportation, education, healthcare finance, and other areas in current years as a result of the information technology industry's swift advancement. The use of information technology has significantly altered both manufacturing and daily living [13]. The way people live and work will alter as a result. The growth of China's industrialization and industrialization has been significantly aided by the advancement of information technology [14]. The two should be combined, according to the Chinese government [15]. The growth of the 17th Party's activities ought to encourage the quick expansion of business, and the improvement of fiscalization, information technology, and national resources [16]. It also fervently advocates for the in-depth inclusion of informatization in order to express ideas on speeding it. Various corporate information has various functions in the corporate strategy elements [17]. To motivate small enterprises to adopt information technology, new synergies must be created, and some corporate information services should also be highlighted [18]. Planning, accounting, tracking, and assessment are all essential components of the perfect cost control [19]. In order to govern contemporary corporate management and progressively generate a mature idea of cost-effectiveness, traditional corporate cost management methods frequently concentrate on cost control, employing technology and certain contents that can overlook operational expenses [20]. It has increasingly realized the value of investing in product development, moving away from the conventional approach of just concentrating on cost savings [21]. Additionally, the cost of acquiring and updating hardware and software for businesses can be greatly decreased by using periodically adjustable resources [22]. Enterprise performance management, also known as ERP, is the research of how to enhance organizational management techniques and procedures under financial and technical constraints so that investments can best serve the needs of the company [23].

The most important problem in WSN is QoS, which attracts a lot of attention. In a real-world setting, quality assurance is essential for the network's degree of efficiency, confidentiality, and integrity. Efficiency is based on the classification-based preference of the QoS variable. Network-oriented characteristics or application-oriented variables may be used to

determine the prioritization [24–26]. It is challenging to improve all QoS factors at once, for example, reducing latency leads to significant network energy usage [27]. Both variables are essential ones for the designing stage. In order to improve the performance, it is necessary to sustain a trade-off between the two factors. The throughput variable is stressed in the study [30, 31], and deadlines are essential parameters in harsh real-time applications [28, 29]. These examples show how the efficiency of the network might vary depending on the application. Throughput [32, 33], bandwidth [34], dependability [35–38], accessibility of resources [39, 40], adaptability [41, 42], and integrity [43, 44] are all improved by QoS, while end-to-end delay [45], jitter [46], different kinds of faults [47, 48], packet error ratio [49–51], packet loss ratio [52, 53], and usage of energy [54] are all decreased. Due to the intrinsic limitations of sensor nodes, like constrained processing power, memory, and energy resources, achieving excellent QoS in WSNs is a challenging issue. The conventional strategies for improving QoS in WSNs have frequently focused on static, rule-based methods that are unable to react to unforeseen, variable environmental situations. When it comes to managing and optimizing WSNs, AI has the possibility to be a game-changer as a transformative technology.

5.2 LITERATURE REVIEW

By facilitating intelligent decision-making, self-organization, and scalability in sensor networks, the combination of AI and WSNs results in a fundamental change. This study investigates the emerging subject of using AI methods to improve the QoS of WSNs, considering the various difficulties that these networks confront. We explore many facets of this implementation, demonstrating the numerous ways AI is used to enhance WSN efficiency.

While significant progress has been made in the integration of AI with WSNs to enhance QoS parameters, there exists a critical research gap that demands further exploration. The existing literature primarily focuses on AI technologies like ML, DL and reinforcement learning to optimize various aspects of WSNs, such as energy efficiency, data routing, and fault detection. However, a comprehensive investigation into the seamless integration of AI for holistic QoS enhancement in WSNs is still lacking.

One major research gap lies in the development of AI-driven solutions that can adapt dynamically to the changing and unpredictable conditions of wireless sensor environments. Real-world WSNs often

operate in dynamic and harsh environments, and existing AI models may struggle to maintain consistently high QoS levels under such conditions. Addressing this gap requires the development of AI algorithms and techniques specifically designed to handle the challenges unique to WSNs, such as limited energy resources, data transmission delays, and network scalability.

Furthermore, there is a need for research that evaluates the scalability and practicality of AI-enhanced WSNs in large-scale deployments. While AI algorithms have shown promise in small-scale experimental settings, their effectiveness and efficiency in real-world, large-scale WSNs remain uncertain. This research gap calls for empirical studies that assess the performance and scalability of AI-integrated WSNs under diverse scenarios and network sizes.

In summary, the research gap in the field of AI application in WSNs for QoS enhancement lies in the development of adaptive AI algorithms for dynamic WSN environments and empirical validation of AI's scalability and practicality in large-scale deployments. Closing this gap is essential to harness the full potential of AI in improving the QoS of wireless sensor networks, thereby enabling their broader adoption in critical applications.

5.3 RESEARCH METHODOLOGY

The research aims to investigate the application of AI techniques to enhance QoS in WSNs. This section outlines the research methodology that will be employed to achieve this goal.

The primary problem to be addressed is the optimization of QoS parameters such as latency, reliability, and energy efficiency in WSNs through the incorporation of AI algorithms. The specific objectives include developing AI models for optimizing QoS parameters and evaluating their performance in a WSN environment.

5.3.1 DATA COLLECTION

To collect data for experimentation, a WSN simulation environment will be set up using popular simulation tools like NS-3 or MATLAB. This will involve creating a virtual WSN with sensor nodes, a sink node, and communication models. Various scenarios will be designed to generate data, including varying network sizes, communication

patterns, and traffic loads. This data will be used for training and testing AI models. AI algorithms like ML, DL, and reinforcement learning will be considered for model development. Algorithms like neural networks, decision trees, and Q-learning may be employed to optimize QoS parameters. The data collected in the simulation environment will be divided into training and testing datasets. The training data will be used to train AI models to optimize QoS parameters. The trained AI models will be integrated into the WSN simulation environment to control and optimize the network's operation in real time. This will involve modifying the node behaviors based on AI-driven decisions. To evaluate the effectiveness of the AI models, performance metrics such as latency, packet delivery ratio, energy usage, and network lifetime will be measured and compared with traditional non-AI WSNs. A set of experiments will be designed to test the AI-enhanced WSN under various conditions, including different traffic loads, network topologies, and mobility patterns. The data collected during experiments will be analyzed to assess how AI algorithms impact QoS parameters. Statistical analysis and visualization techniques will be employed to draw meaningful conclusions. The results obtained from AI-enhanced WSNs will be compared with results from conventional WSNs to highlight the improvements in QoS parameters achieved through AI. The implications of the findings will be discussed, including the potential applications and limitations of AI in WSNs. The research will conclude by summarizing the key findings and contributions to the field of WSNs with AI-enhanced QoS. Suggestions for future research in this area will be presented, including exploring more advanced AI techniques and their potential impact on WSNs. This research methodology provides a structured approach to investigate the application of AI in enhancing QoS in WSNs. It encompasses data collection, AI model development, experimentation, and analysis, with the goal of improving the efficiency and reliability of wireless sensor networks.

5.4 RESULTS AND DISCUSSION

The simulation's goal is to trace the beginnings of WSN. The underlying setting used is NS-2, an object that employs techniques from a command language. Use the suggested architecture to execute all network simulations, and then write the necessary scripts and code.

TABLE 5.1 Simulation Variables

Variables	Value
Simulation Tool	Ns2
Data size	100 MB
Transferring Data	100
Network	WSN

The data from the WSN in Table 5.1 is flawed and requires analysis in a diverse setting completely. There is plenty of reading time in each setting. According on data research, the traffic is being utilized.

The arrival and cycle of data are referred to as idleness. When a node processes data, delay time, may be to blame; in this case, the gap between previous bits is the shortest. Average data provides default nodes, an average quantity of data, and raw node data to the toolbar. A path redirection delay and buffer delays are involved in the method of determining the transmission time in the interface path.

TABLE 5.2 Failure Measure Efficiency

No. of Data	SVM in Sec	KNN in Sec	RSVR in Sec
10	0.4	0.7	0.3
20	0.6	1.4	1.6
30	1.3	2.2	1.5
40	1.7	3.7	2.4
50	2.3	4.6	3.9

Table 5.2 displays the failure efficiency analysis of the method for the delay and buffer delays.

The transmission of data packet delay efficiency is shown in Figure 5.1. SVM achieved 2.3%, KNN 4.6%, and an identified 3.9% delay rate for RSVR.

The proportion among the transmission sequence numbers of the sent data is known as the safety efficiency packet transmission rate guarantees that data is transmitted quickly via the fixed link. Through the perforations, the high packet transfer rate is set.

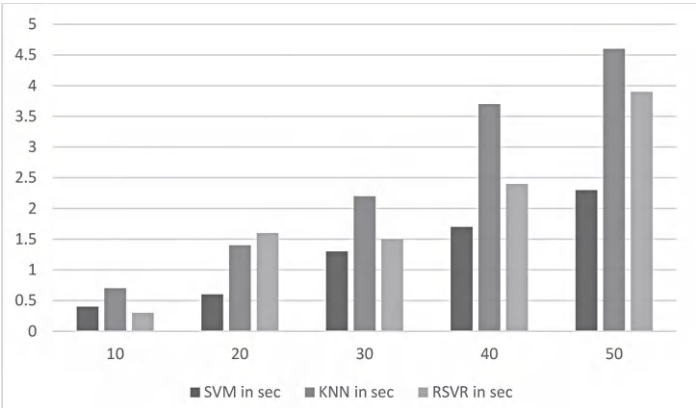


FIGURE 5.1 Efficiency of packet delivery delay.

TABLE 5.3 Security Performance

No. of Data	Security Performance in (%)		
	SVM	KNN	RSVR
10	27	38	49
20	56	59	68
30	58	69	76
40	68	81	90
50	76	91	92

The security efficiency for the three algorithms is shown in Table 5.3. Data security efficiency is represented graphically in Figure 5.2. According to the results, the SVM achieved 82% precision above 87% precision with KNN and 91% precision with RSVR.

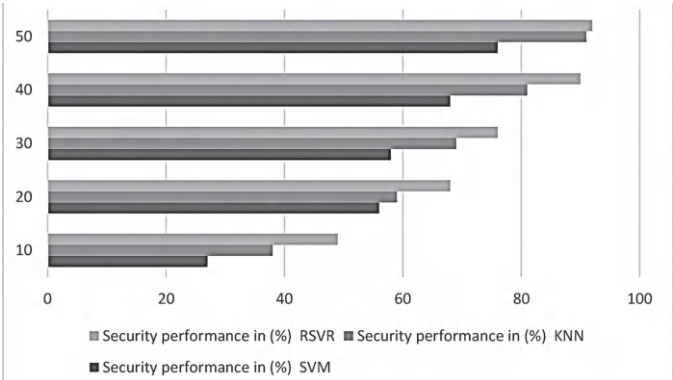


FIGURE 5.2 Algorithm efficiency.

According to the findings of the simulation, RSVR takes less time than KNN and SVM. The indexed data for temporal complexity are displayed in Table 5.4.

TABLE 5.4 Time Complexity

No of Data	SVM in sec	KNN in sec	RSVR in sec
10	23	20	9
20	17	31	11
30	29	17	19
40	36	27	34
50	27	25	18
60	48	39	28

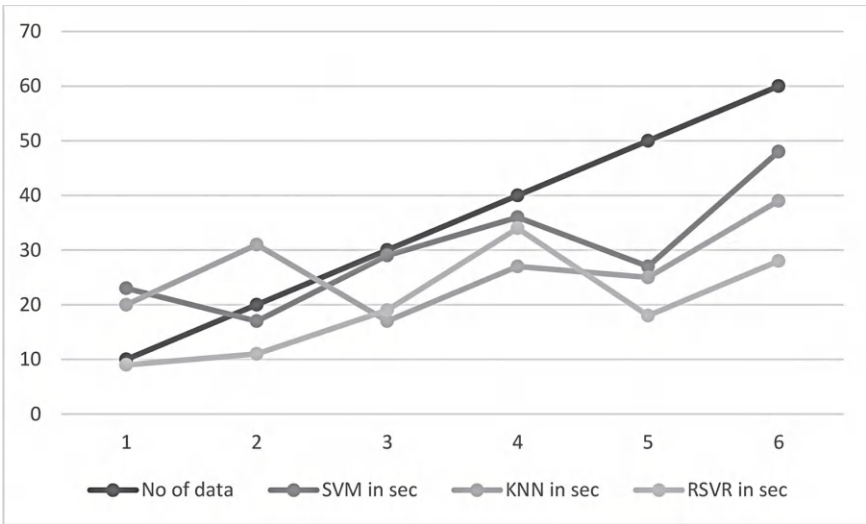


FIGURE 5.3 Graphical representation for time complexity.

The *time limit Analyzes* is shown in Figure 5.3. Based on time, the delay performance data count is determined. Regarding data size and complex time, the SVM estimates a 100 MB table of data transfer to be longer. KNN requires a minimum of 45 secs of data to be supplied, compared to 55 secs for SVM and 30 secs for RSVR. By using the RSVR, complex time is decreased.

The presented study focuses on tracking the origin of WSN in a complex environment, using the NS-2 simulation tool. The primary goal is to comprehensively process data from a wireless sensor network with a 100 MB data size. The research evaluates the performance of three different algorithms, namely support vector machine (SVM), k-nearest neighbors (KNN), and Recursive support vector regression (RSVR), in terms of packet delivery delay, security performance, and time complexity.

1. **Packet Delivery Delay:** The study's first significant finding is related to packet delivery delay. Fig 5.3 illustrates that SVM achieved the lowest delay rate at 4.4%, followed by RSVR at 4.8%, and KNN at 5.3%. This suggests that SVM and RSVR are more efficient in reducing packet delivery delays in the wireless sensor network. Low packet delivery delays are essential for real-time applications and critical data transmission.
2. **Security Performance:** Table 5.3 and Fig.5.4 delve into the security performance of the three algorithms. Security performance is crucial in wireless networks to ensure data confidentiality and integrity. SVM achieved an impressive 82% accuracy, while KNN and RSVR achieved 87% and 91% accuracy, respectively. These results indicate that RSVR outperforms SVM and KNN in terms of security performance. Higher security accuracy is crucial for safeguarding sensitive data in wireless sensor networks.
3. **Time Complexity:** Table 5.4 and Figure 5.3 provide insights into the time complexity of the algorithms. Time complexity is an important consideration as it directly impacts the computational resources required for data processing. RSVR consistently demonstrated the lowest time complexity across different data sizes. For instance, with 100 MB of data, RSVR took only 30 seconds, whereas SVM and KNN required 55 and 45 seconds, respectively. This suggests that RSVR is highly efficient in terms of computational time, making it an attractive choice for resource-constrained wireless sensor networks.
4. **Overall Discussion:** The study's findings highlight the importance of algorithm selection in wireless sensor network applications. In the context of packet delivery delay, SVM and RSVR exhibit superior performance, ensuring timely and efficient data

transmission. These results make them suitable for applications that require low-latency communication, such as IoT devices and real-time monitoring systems.

In terms of security performance, RSVR stands out with an impressive 91% accuracy, making it the preferred choice for applications that demand robust security measures. Ensuring data confidentiality and integrity is paramount in wireless networks, and RSVR's performance in this aspect is commendable [66].

Additionally, the study emphasizes the significance of time complexity, especially in resource-constrained environments. RSVR consistently demonstrates reduced time complexity, making it a practical choice for wireless sensor networks with limited computational resources.

In conclusion, the performance analysis of SVM, KNN, and RSVR algorithms in a wireless sensor network environment provides valuable insights for network designers and researchers [67–71]. The choice of algorithm should align with the specific requirements of the application, whether it prioritizes low packet delivery delays, high security accuracy, or efficient resource utilization. This study contributes to the ongoing efforts to optimize wireless sensor network performance, ultimately enhancing their usability in various domains.

5.4.1 CHALLENGES

Although there have been some difficulties incorporating IoT and WSN into Industry 4.0, intelligent systems may solve a variety of issues encountered by the industrial sector. Issues concerning security and data management have grown as a result of advances in IoT and WSN technology [1]. Manufacturers and businesses struggle to adequately control the growing amount of data being produced. AI algorithms have been used to handle Big Data and improve the intelligence of systems and gadgets. The data is processed using the algorithms throughout a variety of time frames. While corporations are generally unwilling to disclose their confidential data due to weak and inadequate organizational assistance for data in Industry 4.0, the data for education must be disseminated in a central repository. Industry 4.0 also has safety management difficulties [2].

1. **Big Data:** The extraordinary growth of Big Data technology and processing has resulted from the birth of several innovations and an upsurge in their use. Each device generates a significant amount of data. As big data volumes increase, advancements to big data packages run against constraints and need issues that must be resolved in order to effectively govern the volume of data used.
2. **Evolving to 6G:** The IoT and WSN face new challenges as a result of 6G. Implementing low-power, low-cost 6G devices present a significant barrier in terms of processing power [3]. For WSN and IoT, 6G also poses issues in terms of confidentiality and security.
3. **Updates:** Due to interoperability across standards, systems, and their elements, system elements could not be updated. As a result, in intelligent factories, networks are more susceptible to attack if even a small portion of a single system from a network is compromised.
4. **Environment:** WSN security poses a significant problem [4, 5]. Due to wireless data transfer, WSN nodes are not secure when set up in a risky setting. They are easily accessed by an attacker from anywhere in the world and can be controlled. Attacks on the internet can also alter how vulnerable sensor nodes are.

IoT devices are proliferating every day, bringing fresh problems for the reliability and extensibility of supply chain management systems [6, 7]. IoT device connectivity to the cloud or the Internet at the same time necessitates extensive access control, fault tolerance, data management, confidentiality, and security measures.

5. **Restricted Resources:** Additional issues in the WSN domain have an impact on the network's sensor nodes' energy consumption. Sensor nodes typically switch between sleeping and being active. In contrast to active mode, which has additional problems including increased energy usage, sleep mode is regarded as being "outside the network" [8]. They also passed away as a result of the increased energy usage. Sensor nodes often have low processing, memory, and power capacities. Another issue that prevents the incorporation of mobile sensor nodes with the Internet is sensor mobility [72].

5.5 CONCLUSION

In conclusion, the integration of AI into WSNs for the enhancement of QoS represents a significant advancement in the field of wireless communication and sensing. Through this research paper, we have explored the various ways in which AI can be harnessed to address the inherent challenges in WSNs and elevate their performance to new heights [55–60]. Our findings underscore the transformative potential of AI-driven solutions in optimizing QoS parameters such as reliability, latency, energy efficiency, and scalability in WSNs.

The deployment of AI algorithms in WSNs has demonstrated tangible benefits in terms of improving network management and resource allocation. By leveraging ML and DL technologies, WSNs can adapt to changing environmental conditions and network dynamics in real-time. This adaptability not only ensures robust and reliable data transmission but also extends the operational lifespan of sensor nodes, a critical concern in resource-constrained environments. One of the key takeaways from our research is the role of AI in enhancing fault tolerance and self-healing capabilities in WSNs. AI-based anomaly detection and fault prediction mechanisms enable WSNs to proactively identify and mitigate issues, reducing downtime and maintenance costs [61]. Additionally, AI-driven routing protocols optimize data routing paths, minimizing latency and congestion while maximizing the utilization of available network resources.

Energy efficiency remains a paramount concern in WSNs, especially in remote or inaccessible locations. Our study has highlighted the potential of AI algorithms to optimize power consumption by sensor nodes. AI-driven data compression, task scheduling, and adaptive duty cycling strategies ensure that energy resources are utilized judiciously, extending the network's longevity without compromising QoS [62, 63]. Furthermore, the scalability of WSNs, a fundamental requirement for applications ranging from smart cities to environmental monitoring, benefits greatly from AI-based solutions. Through dynamic node clustering, load balancing, and self-organization, AI enables WSNs to seamlessly scale to accommodate a growing number of sensors and adapt to fluctuations in data traffic.

In addition to technical advancements, our research underscores the economic implications of integrating AI into WSNs. By improving QoS parameters, AI-driven WSNs become more attractive for a broader range of applications, attracting greater investments, and fostering innovation.

Industries such as agriculture, healthcare, and industrial automation stand to benefit significantly from these advancements, leading to increased productivity and cost savings. However, it is essential to acknowledge the challenges and ethical considerations associated with the widespread adoption of AI in WSNs [64]. Privacy concerns, data security, and algorithmic biases must be addressed through rigorous regulations and best practices to ensure that AI-driven WSNs are not only technically superior but also trustworthy and responsible.

In conclusion, our research demonstrates that the fusion of AI and WSNs holds immense promise in enhancing Quality of Service across a spectrum of applications. As we move forward, continued collaboration between researchers, industry stakeholders, and policymakers will be crucial to unlocking the full potential of AI in WSNs [65]. By addressing technical, ethical, and regulatory challenges, we can usher in a new era of intelligent, efficient, and reliable wireless sensor networks that will play a pivotal role in shaping our interconnected world. The path ahead is undoubtedly exciting, and with the right approach, AI-driven WSNs have the potential to revolutionize the way we collect, process, and utilize data for the betterment of society.

Artificial Intelligence has emerged as a promising technology for improving the performance and capabilities of WSNs. However, while AI offers numerous advantages, it also comes with several limitations and challenges when applied to enhance the QoS in WSNs [73]. Understanding these limitations is crucial for developing effective solutions and strategies. Here, we discuss some of the key limitations associated with the application of AI in WSNs for QoS enhancement.

1. **Resource Constraints:** WSNs are typically deployed in resource-constrained environments where sensor nodes have limited processing power, memory, and energy resources. AI algorithms often require significant computational resources, which can be impractical for these resource-constrained nodes [74]. Implementing AI on such nodes can lead to excessive energy consumption and may not be feasible for long-term deployments.
2. **Complexity:** AI algorithms, especially deep learning models, tend to be complex and require extensive training and optimization. Deploying complex AI models in WSNs can be challenging due to the limited processing capabilities of sensor nodes. The

deployment and maintenance of AI models in such environments may require specialized expertise and substantial effort.

3. **Data Privacy and Security:** AI-driven WSNs generate and process a vast amount of data, often including sensitive information. Ensuring data privacy and security is a significant concern. AI models themselves can be vulnerable to attacks, and protecting the confidentiality and integrity of AI algorithms and their data inputs is challenging.
4. **Scalability:** The scalability of AI solutions in WSNs is another limitation. As the network size and complexity increase, managing and scaling AI models becomes more challenging. Ensuring that AI-driven enhancements work seamlessly in large-scale WSNs is a non-trivial task.
5. **Energy Efficiency:** AI algorithms can be power-hungry, which is particularly problematic in WSNs where energy conservation is a primary concern. AI-enhanced WSNs must strike a balance between improved QoS and energy efficiency. Running complex AI algorithms continuously can deplete node batteries quickly, leading to reduced network lifetime.
6. **Adaptation to Dynamic Environments:** WSNs often operate in dynamic and harsh environments where network conditions can change rapidly. AI models that are trained on static datasets may struggle to adapt to these dynamic conditions [75]. Ensuring that AI algorithms can continuously adapt and provide reliable QoS in changing environments is a challenge.
7. **Algorithm Robustness:** AI algorithms can be sensitive to variations in input data and noise. In WSNs, where sensor data may be noisy or inaccurate, ensuring the robustness of AI algorithms is essential. Developing AI models that can provide accurate results despite noisy data inputs is a non-trivial task.
8. **Overhead:** Implementing AI algorithms in WSNs introduces additional overhead, both in terms of computation and communication. This overhead can affect the overall network performance and may lead to increased latency, which is undesirable in applications requiring real-time data.

9. **Cost:** Developing and deploying AI solutions in WSNs can be cost-prohibitive. It may require investments in hardware, software, and expertise that are beyond the budget of many WSN deployments, particularly in resource-constrained settings.
10. **Interoperability:** AI-driven enhancements may require specific hardware or software components, which can lead to interoperability issues with existing WSN infrastructure. Ensuring seamless integration with legacy systems can be challenging.

In conclusion, while AI holds significant promise for enhancing the Quality of Service in wireless sensor networks, it is essential to recognize and address its limitations. Overcoming these limitations requires a multi-disciplinary approach that combines AI research with expertise in wireless networking, energy-efficient algorithms, and security measures to create practical and effective solutions for deploying AI in WSNs. Researchers and practitioners in this field must consider these limitations and generate approaches to mitigate them to realize the full potential of AI in enhancing WSN QoS.

KEYWORDS

- artificial intelligence (AI)
- deep learning (DL)
- energy efficiency
- machine learning (ML)
- quality of service (QoS)
- wireless sensor networks (WSNs)

REFERENCES

1. Mehta, R., Sahni, J., & Khanna, K. (2018). Internet of things: Vision, applications and challenges. *Procedia Computer Science*, 132, 1263–1269.
2. Forcina, A., & Falcone, D. (2021). The role of Industry 4.0 enabling technologies for safety management: A systematic literature review. *Procedia Computer Science*, 180, 436–445.

3. Sfar, A. R., Natalizio, E., Challal, Y., & Chtourou, Z. (2018). A roadmap for security challenges in the Internet of Things. *Digital Communications and Networks*, 4(2), 118–137.
4. Sicari, S., Rizzardi, A., Grieco, L. A., & Coen-Porisini, A. (2015). Security, privacy and trust in Internet of Things: The road ahead. *Computer Networks*, 76, 146–164.
5. Tan, Y., Yang, W., Yoshida, K., & Takakuwa, S. (2019). Application of IoT-aided simulation to manufacturing systems in cyber-physical systems. *Machines*, 7(1), 2.
6. Birkel, H., & Müller, J. M. (2021). Potentials of Industry 4.0 for supply chain management within the triple bottom line of sustainability: A systematic literature review. *Journal of Cleaner Production*, 289, 125612.
7. Mrugalska, B., & Ahmed, J. (2021). Organizational agility in Industry 4.0: A systematic literature review. *Sustainability*, 13(15), 8272.
8. Atif, S., Ahmed, S., Wasim, M., Zeb, B., Pervez, Z., & Quinn, L. (2021). Towards a conceptual development of Industry 4.0, servitisation, and circular economy: A systematic literature review. *Sustainability*, 13(11), 6501.
9. Chen, T. H., & Shih, W. K. (2010). A robust mutual authentication protocol for wireless sensor networks. *ETRI Journal*, 32(5), 704–712.
10. Das, M. L. (2009). Two-factor user authentication in wireless sensor networks. *IEEE Transactions on Wireless Communications*, 8(3), 1086–1090.
11. He, J., Xu, J., Zhu, X., Zhang, Y., Zhang, T., & Fu, W. (2014). Reputation-based secure sensor localization in wireless sensor networks. *The Scientific World Journal*, 2014, 2014.
12. Liu, D., Ning, P., Liu, A., Wang, C., & Du, W. K. (2008). Attack-resistant location estimation in wireless sensor networks. *ACM Transactions on Information and System Security (TISSEC)*, 11(4), 1–39.
13. Fortino, G., Di Fatta, G., Pathan, M., & Vasilakos, A. V. (2014). Cloud-assisted body area networks: State-of-the-art and future challenges. *Wireless Networks*, 20, 1925–1938.
14. Esfahani, A., Mantas, G., Maticsek, R., Saghezchi, F. B., Rodriguez, J., Bicaku, A., & Bastos, J. (2017). A lightweight authentication mechanism for M2M communications in industrial IoT environments. *IEEE Internet of Things Journal*, 6(1), 288–296.
15. Anthopoulos, L., Janssen, M., & Weerakkody, V. (2019). A Unified Smart City Model (USCM) for smart city conceptualization and benchmarking. In *Smart Cities and Smart Spaces: Concepts, Methodologies, Tools, and Applications* (pp. 247–264). IGI Global.
16. Arrington, K., & Murray, K. (2018). Architecture, design & engineering—Archiving digital assets: Past, present and future. In *Archiving Conference* (Vol. 15, pp. 139–142). Society for Imaging Science and Technology.
17. Popescu, A. D. (2020). Transforming capital markets by means of financial digital assets. *Revista de Științe Politice. Revue des Sciences Politiques*, (68), 108–119.
18. Zhang, J. F., Wu, Z. J., Feng, P. F., & Yu, D. W. (2011). Evaluation systems and methods of enterprise informatization and its application. *Expert Systems with Applications*, 38(7), 8938–8948.
19. Shi, J. J., & Halpin, D. W. (2003). Enterprise resource planning for construction business management. *Journal of Construction Engineering and Management*, 129(2), 214–221.

20. Mexas, M. P., Quelhas, O. L. G., & Costa, H. G. (2012). Prioritization of enterprise resource planning systems criteria: Focusing on construction industry. *International Journal of Production Economics*, 139(1), 340–350.
21. Gu, Y., Qi, L., & Wang, J. (2018). Breaking the monolith: Strategy, variety, and performance of enterprise information systems. *Journal of Systems Science and Systems Engineering*, 27, 727–770.
22. Mu, L., & Kwong, C. K. (2018). A multi-objective optimization model of component selection in enterprise information system integration. *Computers & Industrial Engineering*, 115, 278–289.
23. Amin, S. H., & Razmi, J. (2009). An integrated fuzzy model for supplier management: A case study of ISP selection and evaluation. *Expert Systems with Applications*, 36(4), 8639–8648.
24. Thakur, U. K., & Dethe, C. G. (2019). Optimizing network QoS using multichannel lifetime-aware aggregation-based routing protocol. In *Soft Computing and Signal Processing: Proceedings of ICSCSP 2018, Volume 2* (pp. 173–179). Springer Singapore.
25. Kumar, D. P., Amgoth, T., & Annavarapu, C. S. R. (2019). Machine learning algorithms for wireless sensor networks: A survey. *Information Fusion*, 49, 1–25.
26. Alsheikh, M. A., Lin, S., Niyato, D., & Tan, H. P. (2014). Machine learning in wireless sensor networks: Algorithms, strategies, and applications. *IEEE Communications Surveys & Tutorials*, 16(4), 1996–2018.
27. Vahabi, M., Ghazvini, M. F., Rasid, M. F. A., Azmir, R. S., & Abdullah, R. (2007). Trade-off between energy consumption and target delay for wireless sensor network. In *2007 IEEE International Conference on Telecommunications and Malaysia International Conference on Communications* (pp. 545–549). IEEE.
28. Tashtarian, F., Moghaddam, M. Y., Sohraby, K., & Effati, S. (2015). ODT: Optimal deadline-based trajectory for mobile sinks in WSN: A decision tree and dynamic programming approach. *Computer Networks*, 77, 128–143.
29. Martinez, A., Alfaro, F. J., Sánchez, J. L., & Duato, J. (2007). Deadline-based QoS algorithms for high-performance networks. In *2007 IEEE International Parallel and Distributed Processing Symposium* (pp. 1–9). IEEE.
30. Bodin, U., & Wolosz, K. (2015). Proportional throughput differentiation with cognitive load control on WSN channels. *EURASIP Journal on Wireless Communications and Networking*, 2015, 1–14.
31. Raha, A., Naskar, M. K., Avishek, P., Chakraborty, A., & Karmakar, A. (2013). A genetic algorithm inspired load balancing protocol for congestion control in wireless sensor networks using trust-based routing framework (GACCTR). *International Journal of Computer Network and Information Security*, 5(9), 9–16.
32. Soundari, A. G., & Jyothi, V. L. (2020). Energy-efficient machine learning technique for smart data collection in wireless sensor networks. *Circuits, Systems, and Signal Processing*, 39, 1089–1122.
33. Collotta, M., Pau, G., & Bobovich, A. V. (2017). A fuzzy data fusion solution to enhance the QoS and the energy consumption in wireless sensor networks. *Wireless Communications and Mobile Computing*, 2017, 1–14.
34. Alghamdi, T. A. (2021). Enhanced QoS routing protocol using maximum flow technique. *Computers & Electrical Engineering*, 89, 106950.

35. Sehgal, A., Agrawal, R., Bhardwaj, R., & Singh, K. K. (2020). Reliability analysis of wireless link for IoT applications under shadow-fading conditions. *Procedia Computer Science*, 167, 1515–1523.
36. Meng, T., Jing, X., Yan, Z., & Pedrycz, W. (2020). A survey on machine learning for data fusion. *Information Fusion*, 57, 115–129.
37. Sandhu, J. K., Verma, A. K., & Rana, P. S. (2018). A novel framework for reliable network prediction of small-scale wireless sensor networks (SSWSNs). *Fundamenta Informaticae*, 160(3), 303–341.
38. Dwivedi, R. K., Rai, A. K., & Kumar, R. (2020). Outlier detection in wireless sensor networks using machine learning techniques: A survey. In *2020 International Conference on Electrical and Electronics Engineering (ICE3)* (pp. 316–321). IEEE.
39. Zia, H., Harris, N., Merrett, G., & Rivers, M. (2015). Predicting discharge using a low complexity machine learning model. *Computers and Electronics in Agriculture*, 118, 350–360.
40. Gunduz, S., Arslan, B., & Demirci, M. (2015). A review of machine learning solutions to denial-of-services attacks in wireless sensor networks. In *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)* (pp. 150–155). IEEE.
41. Haseeb, K., Almogren, A., Ud Din, I., Islam, N., & Altameem, A. (2020). SASC: Secure and authentication-based sensor cloud architecture for intelligent Internet of Things. *Sensors*, 20(9), 2468.
42. Hemalatha, T., Ramesh, M. V., & Rangan, V. P. (2019). Effective and accelerated forewarning of landslides using wireless sensor networks and machine learning. *IEEE Sensors Journal*, 19(21), 9964–9975.
43. Elhoseny, M., Yuan, X., El-Minir, H. K., & Riad, A. M. (2016). An energy efficient encryption method for secure dynamic WSN. *Security and Communication Networks*, 9(13), 2024–2031.
44. Bhasker, L. (2014). Genetically derived secure cluster-based data aggregation in wireless sensor networks. *IET Information Security*, 8(1), 1–7.
45. Devi, V. S., Ravi, T., & Priya, S. B. (2020). Cluster based data aggregation scheme for latency and packet loss reduction in WSN. *Computer Communications*, 149, 36–43.
46. Sreedharan, P. S., & Pete, D. J. (2020). A fuzzy multicriteria decision-making-based CH selection and hybrid routing protocol for WSN. *International Journal of Communication Systems*, 33(15), e4536.
47. Syed, M., & Dubey, M. (2019). A novel adaptive neuro-fuzzy inference system-differential evolution (ANFIS-DE) assisted software fault-tolerance methodology in wireless sensor network (WSN). In *2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)* (pp. 736–741). IEEE.
48. Moridi, E., Haghparsat, M., Hosseinzadeh, M., & Jassbi, S. J. (2020). Fault management frameworks in wireless sensor networks: A survey. *Computer Communications*, 155, 205–226.
49. Poornima, I. G. A., & Paramasivan, B. (2020). Anomaly detection in wireless sensor network using machine learning algorithm. *Computer Communications*, 151, 331–337.

50. Phoemphon, S., So-In, C., & Niyato, D. T. (2018). A hybrid model using fuzzy logic and an extreme learning machine with vector particle swarm optimization for wireless sensor network localization. *Applied Soft Computing*, 65, 101–120.
51. Zhu, F., & Wei, J. (2017). Localization algorithm for large scale wireless sensor networks based on fast-SVM. *Wireless Personal Communications*, 95(3), 1859–1875.
52. Kamalesh, S., & Ganesh Kumar, P. (2017). Data aggregation in wireless sensor network using SVM-based failure detection and loss recovery. *Journal of Experimental & Theoretical Artificial Intelligence*, 29(1), 133–147.
53. Huang, T., Chen, H., Zhang, Z., & Cui, L. (2012). EasiPLED: Discriminating the causes of packet losses and errors in indoor WSNs. In *2012 IEEE Global Communications Conference (GLOBECOM)* (pp. 487–493). IEEE.
54. Moon, S. H., Park, S., & Han, S. J. (2017). Energy efficient data collection in sink-centric wireless sensor networks: A cluster-ring approach. *Computer Communications*, 101, 12–25.
55. Mekonnen, Y., Namuduri, S., Burton, L., Sarwat, A., & Bhansali, S. (2019). Machine learning techniques in wireless sensor network-based precision agriculture. *Journal of the Electrochemical Society*, 167(3), 037522.
56. Pundir, M., & Sandhu, J. K. (2021). A systematic review of quality of service in wireless sensor networks using machine learning: Recent trend and future vision. *Journal of Network and Computer Applications*, 188, 103084.
57. Rawat, P., & Chauhan, S. (2021). A survey on clustering protocols in wireless sensor networks: Taxonomy, comparison, and future scope. *Journal of Ambient Intelligence and Humanized Computing*, 1–47.
58. Shafique, K., Khawaja, B. A., Sabir, F., Qazi, S., & Mustaqim, M. (2020). Internet of things (IoT) for next-generation smart systems: A review of current challenges, future trends and prospects for emerging 5G-IoT scenarios. *IEEE Access*, 8, 23022–23040.
59. Romer, K., & Mattern, F. (2004). The design space of wireless sensor networks. *IEEE Wireless Communications*, 11(6), 54–61.
60. Shah, K., & Kumar, M. (2008). Resource management in wireless sensor networks using collective intelligence. In *2008 International Conference on Intelligent Sensors, Sensor Networks and Information Processing* (pp. 423–428). IEEE.
61. Ancillotti, E., Vallati, C., Bruno, R., & Mingozzi, E. (2017). A reinforcement learning-based link quality estimation strategy for RPL and its impact on topology management. *Computer Communications*, 112, 1–13.
62. de Souza, P. S. S., Rubin, F. P., Hohemberger, R., Ferreto, T. C., Lorenzon, A. F., Luizelli, M. C., & Rossi, F. D. (2020). Detecting abnormal sensors via machine learning: An IoT farming WSN-based architecture case study. *Measurement*, 164, 108042.
63. Gholipour, M., Haghighat, A. T., & Meybodi, M. R. (2017). Hop-by-Hop Congestion Avoidance in wireless sensor networks based on genetic support vector machine. *Neurocomputing*, 223, 63–76.
64. Han, Y., Bai, G., & Zhang, G. (2018). Transmission power control of wireless sensor networks based on optimal connectivity. *Computers & Electrical Engineering*, 70, 844–852.
65. ABID, A., KACHOURI, A., & MAHFOUDHI, A. (2014). Anomaly detection in WSN: Critical study with new vision. In *Proceedings of the International Conference*

- on Automation, Control, Engineering and Computer Science (ACECS'14) (pp. 37–46).
66. Alam, K. M., Kamruzzaman, J., Karmakar, G., & Murshed, M. (2012). Priority sensitive event detection in hybrid wireless sensor networks. In *2012 21st International Conference on Computer Communications and Networks (ICCCN)* (pp. 1–7). IEEE.
 67. Alshinina, R. A., & Elleithy, K. M. (2018). A highly accurate deep learning-based approach for developing wireless sensor network middleware. *IEEE Access*, 6, 29885–29898.
 68. Arbi, I. B., Derbel, F., & Strakosch, F. (2017). Forecasting methods to reduce energy consumption in WSN. In *2017 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)* (pp. 1–6). IEEE.
 69. binti Zubir, N. Z., Ramli, A. F., & Basarudin, H. (2017). Optimization of wireless sensor networks MAC protocols using machine learning: A survey. In *2017 International Conference on Engineering Technology and Technopreneurship (ICE2T)* (pp. 1–5). IEEE.
 70. Choudhary, A., Choudhary, S., & Mishra, A. (2019). Review on Fault Tolerance in Wireless Sensor Network. *International Journal of Computer Applications*, 182(45).
 71. Amri, S., Khelifi, F., Bradai, A., Rachedi, A., Kaddachi, M. L., & Atri, M. (2019). A new fuzzy logic-based node localization mechanism for wireless sensor networks. *Future Generation Computer Systems*, 93, 799–813.
 72. Khedr, A. M., & Omar, D. M. (2015). SEP-CS: Effective routing protocol for heterogeneous wireless sensor networks. *Ad Hoc Sensors and Wireless Networks*, 26(1–4), 211–234.
 73. Wang, F., & Liu, J. (2010). Networked wireless sensor data collection: Issues, challenges, and approaches. *IEEE Communications Surveys & Tutorials*, 13(4), 673–687.
 74. Zhou, Y., Fang, Y., & Zhang, Y. (2008). Securing wireless sensor networks: A survey. *IEEE Communications Surveys & Tutorials*, 10(3), 6–28.
 75. Othman, S. B., Bahattab, A. A., Trad, A., & Youssef, H. (2015). Confidentiality and integrity for data aggregation in WSN using homomorphic encryption. *Wireless Personal Communications*, 80, 867–889.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

CHAPTER 6

Recent Advancements of Artificial Intelligence in Edge Computing

CHETAN SHELKE,¹ RATHNAKAR ACHARY,¹ SREENIVAS MEKALA,²
K. K. RAMACHANDRAN,³ and SHAILESHKUMAR WAGH⁴

*¹Alliance College of Engineering and Design, Alliance University,
Bangalore, Karnataka, India*

*²Department of IT, Sreenidhi Institute of Technology and Science,
Hyderabad, Telangana, India*

*³Management/Commerce/International Business, Dr. G. R. D. College of
Science, Coimbatore, Tamil Nadu, India*

*⁴Department of Geography, MGSMs, Dadasaheb Dr. Suresh G. Patil
College, Chopda, Jalgaon, Maharashtra, India*

ABSTRACT

The internet of things (IoT) has grown significantly during the past several years. The IoT significantly aided the advancement of artificial intelligence (AI) by supplying sufficient data to train and infer models. The conventional cloud computing model may still have a difficult time autonomously managing the massive volumes of data that IoT generates and satisfying related practical requirements given this background and trend. In response, the academic and industrial communities have given edge computing (EC) a lot of attention. However, as EC research has continued to advance, academics have discovered that traditional (non-AI) methodologies have limitations when it comes to improving EC performance. As

EC researchers see how successfully AI is being used in other industries, they start to concentrate on the subject, particularly from the standpoint intelligent machines. Intelligent automation is a subfield that popularity over the past couple decades. In this essay, we'll first go through the formal definition of EC and the factors that have contributed to its rise as a popular computing model. We list the conventional approaches and draw attention to their shortcomings. In order to explore new research ideas in these two fields and to take use of the beneficial interactions between AI and EC, this article covers the study findings of utilizing AI to optimize EC and applying AI to other disciplines under the EC architecture.

6.1 INTRODUCTION

Due to the significant breakthroughs in the IoT and AI, 14th Industrial Revolution is changing how organizations update and modernize the manufacturing to implement digital factories that are smart [1–5]. In our daily lives, automatic devices and robotics are indispensable. Additionally, they are essential for manufacturers who utilize AI to complete clouds computing and integrating data analysis throughout their entire company, including the manufacturing facilities [6–9]. Since 2012, AI has developed quickly thanks to advancements in computing speed and power, which have sped up the design of several deep learning algorithms using graphics processing units (GPUs). AlexNet [10] was developed in 2012 to divide the deep neural network into two halves so that it could utilize the processing capacity of numerous GPUs. The ImageNet large scale visual recognition challenge (LSVRC) 2012 has an accuracy rate of roughly 84.6% using AlexNet. In 2014, VGGNet [11] increased accuracy to 92.7% thanks to the advantages of deeper neural network layers. GANs or generative adversarial networks [12] are introduced strong subtype of generative models for picture production challenges at the same time. In Besides the developments in the processing of images, there are several basic deep neural network systems in the disciplines of processing visual arts, artificial recognition of voices, processing natural languages as well as medical informatics techniques like DNN, CNN, and RNN have been employed [14–16].

AsIoT is deployed and used more intelligent devices, such as cell phones and detectors, are able to contact a network, producing a sizeable

volume of network data. There constrained computer capabilities and persistently demanding applications prevent them from achieving real-time and efficient execution, despite the fact that their processing power is growing extremely quickly. Cloud computing [17, 18] is able to handle very complicated computing operations and services to enable cloud-device collaboration. In the framework of cloud computing, consumers can trust the extraordinarily robust computing and storage capacities of a technology that is cloud-based facility to enhance the calculating performance of devices and space for storing and to expeditiously complete computationally demanding activities. The device-cloud cooperation mode does have significant drawbacks, though, namely large transmission delays and exceedingly high network bandwidth needs. Edge computing [19, 20] is suggested as a solution to the issues with using the cloud for data processing in order to provide users with the essential clerical services [21]. When compared to cloud technology, edge computing permitted specific user apps to be performed on edge servers that are close to intelligent devices, thereby reducing the amount of network bandwidth and data transmission latency needed for edge-cloud collaboration. Edge computing has another benefit over device-cloud computing in that it can ensure more successfully security of user data because long-distance data transports are eliminated. Therefore, utilizing edge computing to conclude various computer operations for intelligent machines has emerged as a significant development trend [22–24]. In this paper, these gadgets are referred to as edge gadgets.

The explosion of sensor technologies creates Industry 4.0's second cornerstone. The ability to respond to physical stimuli or chemical variations has been built into sensor devices, and the resulting impulse is transmitted for use in measurement and control systems [24–29]. Additionally, self-powered sensors with piezoelectric and tri-bioelectric bases are being researched as a way to further do away with the requirement for external power supplies for user interfaces and peripheral electronics. The use of piezoelectric and tri-bioelectric materials has advanced the development of wearable sensors due to this property [30–32]. With the help of AI data analytics and sensors on clothing, applications in medical care, monitoring the environment, human-machine interaction (HMI), and plant monitoring can record signals of muscles deformation, joint stretching, fluctuations in heartbeat frequency, and others (Figure 6.1).

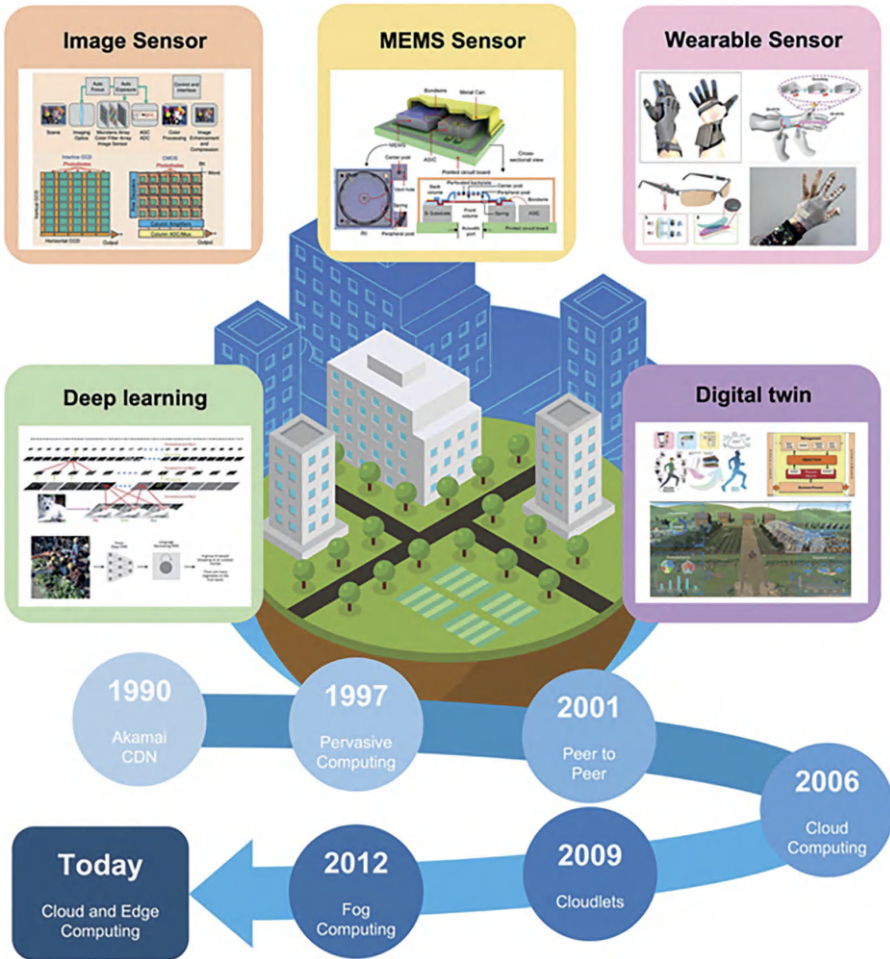


FIGURE 6.1 Advancement of cloud and edge computing, as well as the creation of sophisticated AI sensors.

All computation-intensive tasks from edge devices are typically offloaded to a processing edge server [33, 34] according to established scheduling algorithms for edge computing jobs. The wastage of materials used for processing and storing between cloud computing and edge computing facilities, however, may be the outcome [35–37]. Additionally, numerous devices may connect via an edge server simultaneously. The server may thus get overloaded with computational tasks, resulting in a backlog of work [38, 39]. Due to the longer completion times for all jobs

in the queue; the edge server's processing time for tasks now exceeds that of the edge devices. However, a lot of edge devices could remain inactive, wasting their computational power, and cloud data centers with lots of resources could be underused. By using task offloading, edge devices' computing obligations can be properly managed; in order to fix the issues mentioned above, incorporating an online data center, server farms, and edge devices. We should allocate providing them with computer tasks for on-demand processing and utilize the storage and computational capabilities of a remote data center, servers at the edge, and optimization targets, and system state. An edge computing infrastructure designed to handle greater AI workloads at and near the edge is a key component of an efficient edge AI paradigm. Enterprises may successfully utilize their data with industry-leading performance and limitless scalability when edge AI and storage solutions are combined [40–42].

The advantages of Edge AI are already being gained by numerous international businesses. Edge AI has applications across many industries, from enhancing assembly line production monitoring to operating autonomous vehicles. Additionally, as additional industrial uses for the technology continue to develop, the recent rollout of 5G infrastructure in several nations provides Edge AI an extra boost [45].

6.2 LITERATURE REVIEW

Edge computing with AI for enterprises has a number of benefits, including.

- Effective asset management and predictive maintenance.
- A minute or less per product for inspection.
- Reduction in field problems.
- Increased client satisfaction.
- Providing a significant Edge AI infrastructure.
- Enhancing city traffic control procedures.

6.2.1 TECHNOLOGIES ADVANCING CUTTING-EDGE AI

In the realm of edge AI, developments in expertise like machine learning, information science, and the Internet of Things are very important [46, 47]. The true issue, though, is in closely following the course of computer science advancements especially the software with advanced AI capabilities and hardware which can easily fit into an artificial intelligence and

machine learning environment. Thankfully, new hardware development is taking place in the domain of edge computing, which will free Edge AI from its current limitations. AIStorm, Esperanto Technologies, and Sima.ai are a few of the few companies creating microchips that can handle demanding AI workloads [48–50].

For \$15.3 billion, Intel purchased Mobileye, a Tel Aviv-based provider of vision-safety technologies, in August 2017. Recently, Kunlun AI chips of the second generation, an incredibly quick microprocessor regarding edge computing, began to produce in bulk by Chinese multinational technology juggernaut Baidu [51, 52].

Along with microchips, Microsoft, Amazon, and Intel, and Asus collaborated to create motherboards using the search engine's Edge TPU and Nvidia's Jetson Nano to enhance the capabilities of edge computing [53]. The first deep learning-capable video camera in the world, the Amazon AWS DeepLens, represents a significant leap in this approach to surveillance.

As opposed to analyzing data in a faraway cloud where it would be delayed by long-distance communication, edge technology analyzes data locally, it can respond to customer need instantly [54, 55]. The most well-known Edge AI applications, like autonomous automobiles, show how software can handle data and machine learning utilizing deep learning methods. The system can analyze data in a matter of milliseconds when used in conjunction with an autonomous vehicle, preventing accidents in real-time.

Edge can provide significant cost savings to your company due to the scalability of analytics and decreased latency in making critical decisions. In addition to saving time, the edge can conserve bandwidth by reducing the need for data transfer, which also improves the energy efficiency of the equipment. Edge AI usually offers the best value despite requiring local hardware and computing resources.

6.2.2 RECENT ADVANCEMENT

The effectiveness of deploying AI models at the edge has risen as a result of three recent developments:

- 1. Evolution of Neural Networks:** Now that neural networks and related AI technology have advanced, general machine learning is feasible. After properly training AI models, businesses are learning more about how to use them in production at the edge [56].

2. **Infrastructure Upgrades:** For AI to function at the edge, distributed processing power on a massive scale is necessary. As a result of recent developments in massively parallel GPUs, neural networks can now be utilized [57, 58].
3. **Adoption of IoT Devices:** The widespread adoption has been as an outcome of “Internet of Things,” an increase in information. Due to sudden ability to gather data from every aspect of a business, including industrial gauges, cutting-edge cameras, machines, and more, we now have the knowledge and resources required to deploy AI models at the edge [59, 60].

6.2.3 CHALLENGES OF EDGE AI

By accelerating operations and improving device energy efficiency, edge AI improves decision-making, safeguards data processing, greatly personalizes the user experience, and reduces expenses. Edge computing with AI offers several benefits. Regardless of what they are right now, they are all concentrated on enhancing processes and consumer satisfaction [61]. The three categories of smart parking management, traffic monitoring/prediction, and intelligent transportation management are reviewed and categorized in this section [62].

1. **Smart Transportation:** An immediate traffic control system is created using traffic conditions on roads monitoring, recognizing incidents for traffic safety in real-time and stopping disruptions that effect the availability of parking and traffic congestion. To accomplish this, the transportation sector uses IoT and AI technology to collect data about vehicles, drivers, and roadways [63, 64].
 - i. **Smart Parking Management (SPM):** In, the authors proposed a method for sharing bicycles based on computing at the edge that would cluster the initial data using a hybrid machine learning model (SOM-RT) and an autonomous mapping network [65]. Each station needed to have a certain number of bikes then calculated using each cluster as an RT. Experiments perform better in generalization and prediction accuracy than other methods [66].
 - ii. **Traffic Monitoring and Prediction (TMP):** An approach for cooperative improvement of smart transportation systems

was proposed by the authors of to deal with the issue of dynamic traffic changes for accurate traffic forecasting and for real-time detection of abnormal situations. Data collection from every intersection is made possible by putting up observation stations at various traffic intersections. To predict traffic patterns and forecast the network's entire traffic flow, the DBN-SVR technique is used. After the model was utilized to determine the traffic flow at a few selected intersections, the newest computer technology was used to process the data signals produced by the crossings [66, 67].

- **Intelligent Transportation Management (ITM):** The authors of this study used actual time distracted driving detection to an edge/cloud computing system based on a modified DCNN model and a VGG16 based model. In their article, Xu et al. describe a method for assessing driving behavior that uses vehicle edge-cloud architecture. The telemetric box of a moving vehicle transmits data to edge networks that show the driver and autopilot's actions. The edge networks return to the source the behavior ratings vehicles using the model for evaluating driving behavior created by the cloud server.

The model for evaluating driving behavior is developed using automobile data and built on a cloud server. After then, the edge networks receive the model on a regular basis for changes. Experimental findings indicate that durability and the effectiveness of the suggested technique [68, 69].

- **Vulnerable Security Feature:** Some computer specialists assert that edge computing decentralized structure enhances its security qualities. Locally gathered data, however, necessitates protection for more sites. An Edge AI infrastructure is more open to several types of cyberattacks as a result of the extra physical data points.
- **Limited Machine Learning Power:** On edge computing hardware platforms, machine learning demands more processing power. The processing speed in edge AI

infrastructure is constrained by the edge or IoT device speed. In order to increase accuracy and efficiency, huge, intricate Edge AI models often need to be reduced before being deployed to the Edge AI hardware [70, 71].

6.2.4 APPLICATION OF EDGE AI

- **Robotic Vehicles:** Edge AI-enabled autonomous vehicles are better equipped to identify road traffic features and navigate routes than humans because of their decision-making speed and accuracy. Without manual interference, it makes transit faster and safer.
- **Optical Automation Inspection:** Optical inspection by computers is a crucial part of manufacturing processes. With the use of a systematic Edge AI visual analysis, it is possible to identify defective components on a production line. Without requiring massive quantities of cloud-based data transmission, automated optical inspection provides incredibly accurate, ultrafast data processing [72, 73].

In addition to the use cases listed above, Edge AI could be quite helpful in emergency medical treatment, the improvement of industrial IoT security, and facial recognition technology. Every day, more use cases for Edge AI are discovered. By meeting everyone's personal and professional needs in the near future, edge AI will develop into a standard piece of day-to-day technology.

6.3 METHODOLOGY

Edge AI describes the installation of AI programs on real-world hardware. The phrase “edge AI” refers to the fact that, as opposed to being carried out centrally at a personal data center or internet-based computing facility, AI processing happens near the network's edge, proximity to the user and the location where the data is kept. The network's edge can be any place because the internet has a global reach. It may be a department shop, factory, hospital, or any of the technological advancements that surround us, including traffic lights, self-driving cars, and cell phones.

The “intelligence” of human cognition may now be used by computers and other devices wherever they are thanks to advancements in edge AI.

Intelligent AI-enabled systems acquire the ability to carry out the same duties in different conditions, similar to in actual life [74].

6.3.1 LOGIC BOARD SCENARIOS (COMPUTING SCENARIOS)

Edge computers and edge appliances constitute majority of the Internet of Things' frontier computing resources. When scheduling tasks, we also take cloud centers into account as a component of the whole system in order to benefit from them. Cloud data centers often contain a high number of powerful computing servers. Managing computing tasks requires effective use of the processor, storage spaces, the bandwidth, and other system resources [75]. Consequently, the distribution of computing assets, many computing scenarios are examined and summarized in this section. Edge gadgets, the servers, and scheduling are all available on the outer edge. Offloading computational tasks that edge sensor are capable of performing to servers along with other available or unoccupied edge devices could exist. They have resources for processing, but they can only provide a finite number of computational services for computing edge applications, storage, networks, and other things. They are far less resource-rich than edge servers, but they do possess resources for processing, storage, networks, and other things [76].

6.3.2 EDGE COMPUTING BASICS

Edge servers and devices make up the first scenario. In this case, an edge device has the choice of offloading a calculation to a server at the edge or carrying it out locally. The edge server performs it, which subsequently relays computation outcome to appropriate edge sensor. Using a cloud computing facility to offload work from devices is similar to this situation. It is the most straightforward case in edge computing. The locations for outsourcing computing tasks to edge servers have been determined. Additionally, the specific types of computing jobs executed are based on the capacities of the edge server and are fixed in nature. In this situation, additionally, there is no internet-based computing facility [76]. The result is, it is better suited for processing activities in relatively restricted environments that need less computation and strict latency

constraints. According to the aforementioned research, it is projected that the proposed scenario will achieve the Quality-of-Service levels. To gauge quality of service, use task completion time. We view all servers on the edge and devices at the edge as being the same. Keep in mind that the majority of the material is easily adaptable to heterogeneous devices and servers [77].

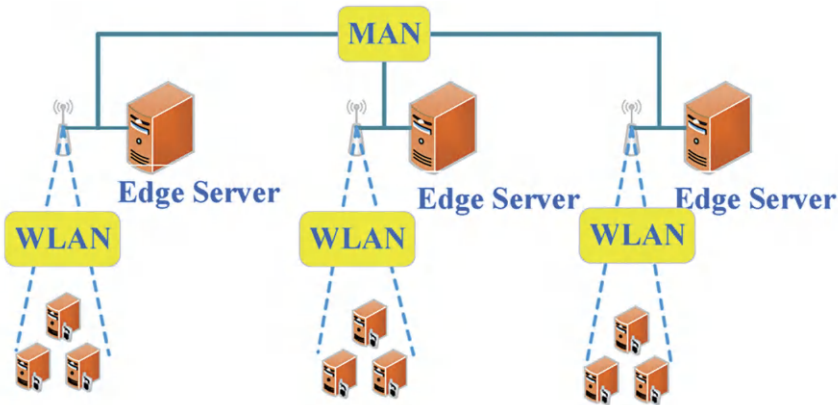


FIGURE 6.2 Standard LAN and MAN for edge computing.

Let t represent the amount of time it takes for a task that has been delegated to an edge server to complete, taking into account both the wait time before the task is processed as well as the latency associated with data transfer involving an edge server and an edge device. Let τ represent time it takes to execute single instruction and represent time it takes for an edge server to process a task. G represents all of the instructions needed to complete this project. The task's completion time can be calculated using the formula:

$$\tau = n(1 + P\tau)\tau + \Omega + \eta G$$

P is packet loss rate experienced during n packet transmissions involving an edge computer and an edge gadget, where n is total number of packets sent, including those sent in both directions involving an edge computer and an edge gadget. The quantity of transmission, interpreting, and delaying are all accounted for by the average packet latency involving an edge computer and an edge gadget.

6.3.3 SCHEDULER BASED EDGE COMPUTING

As opposed to the initial, it has an angle scheduling that carefully plan out jobs in this case; a computing work may be handled directly by an edge device or may be delegated to the outer edge processor. The edge processor logically distributes work to edge sensors and processors in accordance with the scheduling policies. The guidelines are created using information about all edge servers and devices' current processing, storage, job execution status, network state, and other factors. The scheduler then inputs the results of the calculation to the initial hardware. Fundamental characteristic of circumstance scheduler can reasonably schedule the computing jobs to multiple servers and edge devices, allowing for the successful completion of collaborative computing tasks across numerous devices at the edge and servers. Edge computing devices can be used to their fullest potential. There are a variety of computer jobs that can be processed as well as computing resources that are available at the edge. Versus a cloud computing facility, the edge's computation and storage capabilities are constrained. As illustrated in Figure 6.2, this architecture obviously works well for processing tasks requiring little computation and strict latency constraints [78].

This case suggests that the work could be assigned to an edge scheduler's designated edge server. The role of the target edge server completion time can be calculated using the same formula as in the first scenario:

$$\rho = n(1 + Pp)\rho_A + n(1 + Pp')\rho_B + \Omega + \eta G$$

Where Pp' is rate of data transpiration encountered with target edge computers and an edge scheduler and Pp is rate of data loss received during

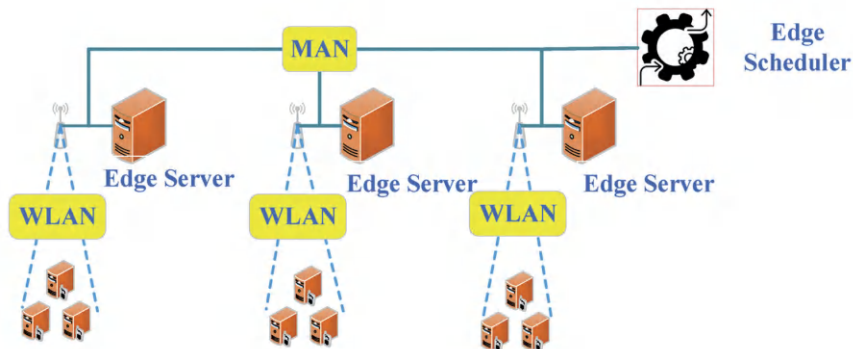


FIGURE 6.3 Edge computing with a scheduler.

the exchange of n data between a device at the edge and edge scheduler. The mean latency per packet with regards to an edge planner and an edge technology, ρ_A and ρ_B , which is the standard deviation of the latency per packet between an edge scheduler and a target edge server, respectively, indicate total amount of interpreting, queuing, and transportation delays n packets. If an edge device's task is straight towards the cutting-edge computer scheduled.

$$P\rho = P\tau, \tau = \rho_A, \rho_B = 0 \text{ and } \rho = \tau.$$

Let ' ρ ' stand for the time it takes to execute a single instruction, and let ' G ' represent the total number of instructions required to complete this task on an edge device. To figure out when a work scheduled for a target device will end, use the formula below:

$$\rho = n(1 + P\rho) \rho_A + n(1 + P\rho'') \rho_C + \Omega' + \eta' G'$$

During n packet transmissions involving a target edge technology and an edge scheduler, $P\rho''$ stands for packet loss rate. With a target edge technology and an edge scheduler, the average packet latency (ρ_C) is the sum of the transfer, interpreting, and holding delays for n packet. The amount of time until a task is processed at an edge device is Ω .

The primary difference between this scenario and the first one is an edge scheduling able to assign jobs to a sluggish server edge in accordance with network and edge servers' actual time conditions. It delivers more benefits than the first one if the job processing time makes up a considerable fraction of the total length [78].

6.3.4 EDGE-CLOUD COMPUTING

In this case, there isn't any edge scheduler who also contains edge servers, devices at the edge, and a cloud computing facility. Due to the ability of its cutting-edge tools to outsource work to its center for internet computing, it varies from first scenario. Based on the unique characteristics based on its computing activities and the users' QoS requirements, edge devices determine the appropriate offloading to a computing cloud facility or server at the edge. The primary difference between this scenario and the previous one is that it has fixed job types and offloading sites for edge computing. A cloud computing facility can be used to offload any jobs that need a lot of processing and are insensitive to delays on the edge. As

a result, as illustrated in Figure 6.3, only the sorts of edge environment resources in this architecture have an impact on how computing operations are processed. These architectural designs are used in [18].

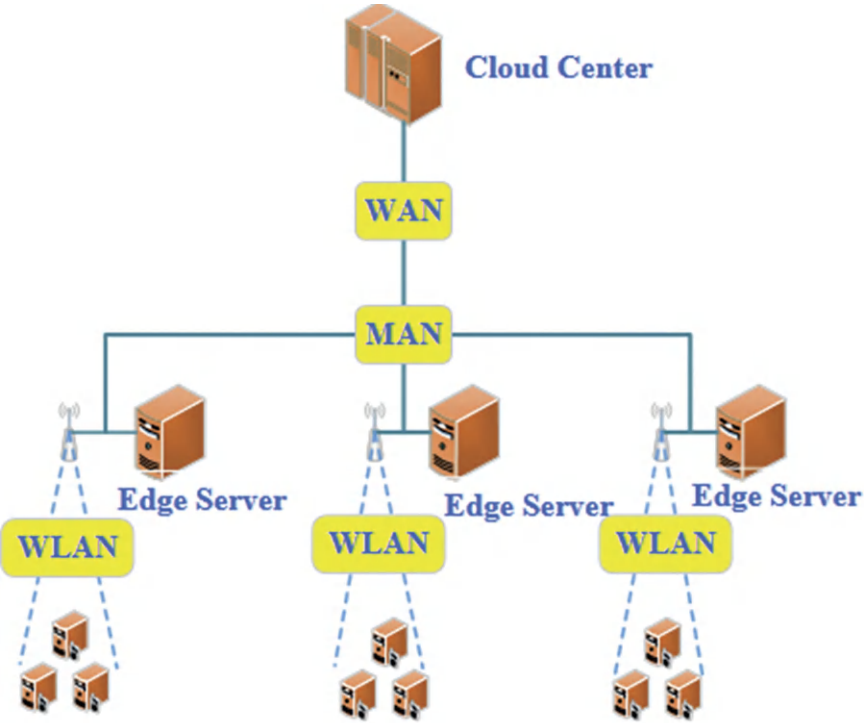


FIGURE 6.4 Edge-cloud computing structure.

In this case, an edge network or an internet-based server handles a task. The above situation is created with an estimated time of completion that depends on the task’s location.

6.3.5 CLOUD COMPUTING AT THE EDGE WITH SCHEDULERS

The 4th instance consists of a cloud-based data center, an edge scheduler, edge servers, and edge devices. In this case, an edge device has two options for carrying out computational tasks: locally or by offloading them to the scheduler at the edge. Third architectural style, in contrast, is the edge of the scheduler, which accepts all computing tasks that edge schedules and

offloading devices them to the appropriate computing entities for service performance in accordance with the available resources for computer power, storage capacity, network speed, and task parameters. It shares the same primary quality as the second one, namely that edge computing jobs’ offloading positions are ambiguous and their types are varied. Only certain sorts of edge environment resources have an impact on how computing jobs are processed, not the amount of computing. In Figure 6.4, its architecture is displayed.

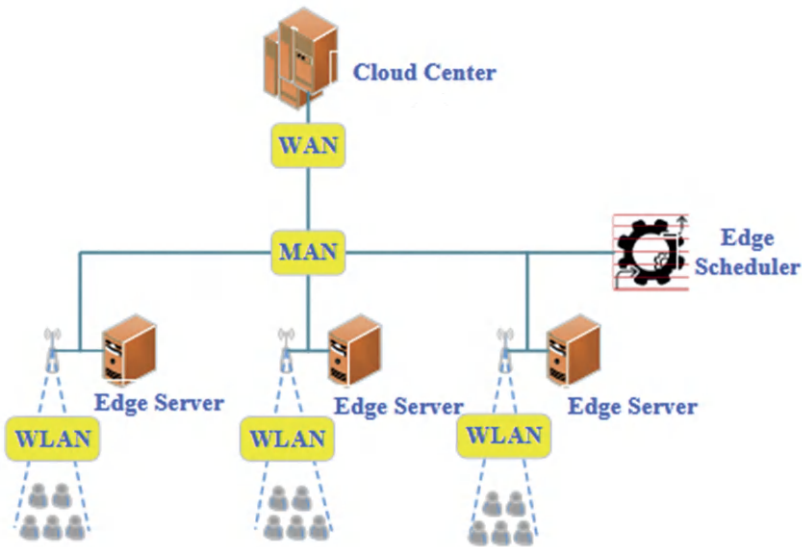


FIGURE 6.5 Edge-cloud computing with a scheduler.

In this case, these duties might be carried out by a cloud server, an unused edge server, a device, or both. An edge scheduler can assign them to the correct areas [18] (Figure 6.5).

6.4 RESULTS AND DISCUSSION

The computing activities are then examined to see if they can be precisely scheduled to the right node and accomplish the desired goals, such as using the least amount of time and energy to complete. We determine whether subtasks are interdependent and whether they may be divided based on

task attributes [19]. A computing task that is simple or well-integrated cannot be separated; it can only be completed all at once either partially offloaded at the edge of the server or at the edge equipment. This is special assessment criterion. If they can be divided into several segments depending on their code and/or data, they can be offloaded [81]. To sum up, depending on the computing task, there are three operational modes: local implementation, a portion of it being offloaded, and the entire thing. In light of the device's processing capacity, the status of the network, and the resource availability of a cloud computing facility, servers on the edge devices, and accurate offloading location of computing tasks needs to be carefully considered [82, 83].

- **Execution Locally:** A network of servers that store an edge device's resources and the state of the resources should be taken into consideration when deciding whether or not edge computing tasks are completed locally [15]. The computer task can only be carried out locally if the network bandwidth is insufficient to allow a job to be successfully uploaded, that is, if the staying network bandwidth is less than the memory bandwidth needed for the task's upload. Additionally, if the edge servers' computing resources are not available, which prevents the processing of computing duties in a timely manner, the tasks must be carried out locally. An edge server's load and the requirement regard to network activity are effectively reduced when the computational capabilities may satisfy the service requirements of an edge device.
- **Full Offloading:** To ascertain whether all activities involving edge computing transferred entirely to a scheduling system or edge computer. The present network, the capabilities of edge devices, the accessibility of the resources on edge servers and the results of system optimization must all be considered. If the accessible network bandwidth permits the efficient offloading of computational activities, local processing or offloading of the jobs is chosen, the servers at the edge or other devices on the edge are idle, and the results of local execution and full offloading to the edge servers are equivalent with respect to a scheduling goal. For instance, if the objective is to decrease the amount of time required to finish a task, it is critical to evaluate the execution

timings for local execution against offloading to an edges server/cloud computing center [17].

- **Partial Offloading:** Edge devices have just two options: Distributing an indivisible computing task as it relates to the edge scheduler, subsequently the appropriate server at the edge or idle edges device, locally or completely. Partial offloading benefits split table computer jobs. They should be treated as a scheduling unit when scheduling the distinct separate tasks, consideration should be given to the network, edge servers, and edge devices' resources [84]. If there is no interdependence between the subtasks of the split computing task, they may be allocated to several nodes and processed simultaneously in order to save energy and hasten job completion. However, this can have an adverse effect on the overall task's eventual processing impact. If there is any interdependence between the interdependent subtasks, they should be assigned to the same computational node for execution [80].

There are many different ways to divide tasks. Some studies examination of the application division problem of continuously updating division during application execution suggests a framework for application division in a dynamic mobile cloud environment [79]. Using their framework, they produce an online solution for the dynamic network connection to the cloud, which can drastically reduce the amount of time it takes for applications to finish.

6.5 CONCLUSION

The analysis and synthesis of computing scenarios, tasks, definition of optimization targets, and this study, scheduling strategies for computing tasks for an edge computing system are provided. The computer simulations for arranging work are categorized into 4 groups based on edge computing resources, and each group's composition and characteristics are carefully evaluated. Computing processes can be executed locally, partially offloaded, or fully offloaded depending on where they are performed. We then outline the optimization challenge to cut down on waiting time and energy use for offload computation using various queuing models. We also show the complexity of the solution. The majority of current studies' optimization objectives for edge computing job scheduling

approaches are to reduce delay, energy usage, or both. To handle such an optimization challenge, improved conventional work scheduling methods and some clever optimization algorithms can be applied. Due to the computational complexity of the examined optimization problems, most researchers favor using better heuristic algorithms or intelligent optimization algorithms over mathematical programming ones. The challenges and potential opportunities in the area of cooperative task scheduling in an edge computing paradigm are also covered in this work. This study should encourage additional investigation into a team approach to scheduling and using it for computing at the edge.

KEYWORDS

- **edge computing**
- **cloud computing**
- **CNN**
- **RNN**

REFERENCES

1. Shi, W., & Dustdar, S. (2016). The promise of edge computing. *Computer*, 49(5), 78–81.
2. Qin, M., Chen, L., Zhao, N., Chen, Y., Yu, F. R., & Wei, G. (2018). Power-constrained edge computing with maximum processing capacity for IoT networks. *IEEE Internet of Things Journal*, 6(3), 4330–4343.
3. Ghosh, A. M., & Grolinger, K. (2021). Edge-cloud computing for internet of things data analytics: Embedding intelligence in the edge with deep learning. *IEEE Transactions on Industrial Informatics*, 17(3), 2191–2200.
4. Zhou, P., Chen, W., Ji, S., Jiang, H., Yu, L., & Wu, D. (2019). Privacy-preserving online task allocation in edge-computing-enabled massive crowd sensing. *IEEE Internet of Things Journal*, 6(5), 7773–7787.
5. Jamwal, R., Agrawal, M., & Sharma, A. (2021). Application of artificial intelligence for multi-objective optimization of energy and quality of service in 5g mobile networks. *Applied Sciences*, 11(11), 5725.
6. Zhang, Y., Chen, H., & Chen, D. (2021). Edge computing for big data analytics: A survey. *Information Systems Frontiers*. <https://doi.org/10.1007/s10796-021-10153-5>.

7. Ajeev, B. H., Javaregowda, A., Ali, M., Modak, S., Patil, S., Khatua, S., Ramadoss, M., Kothavade, P. A., & Arulraj, A. K. (2020). Materials for sustainable development: An overview of advanced materials for energy and environment. *Advanced Materials Technologies*, 5(1).
8. Patel, M., Lee, W. S., & Peres, N. A. (2022). A review on the impact of advanced sensing technologies in food safety and quality management. *Sensors*, 22(22), 8558.
9. Lu, Y., Zeng, Z., Lu, H., & Lv, E. (2021). Internet of Things: A survey on security and privacy issues. *Sensors*, 21(21), 8369.
10. Ren, Z., Zhang, Z., Wei, J., Dong, B., & Lee, C. (2022). Data-driven edge computing: A review. *Nature Communications*, 13(1), 3859.
11. Mukhopadhyay, S. C., Tyagi, S. K. S., Suryadevara, N. K., Piuri, V., Scotti, F., & Zeadally, S. (2021). Sensors and actuators for the Internet of Things: A survey. *IEEE Sensors Journal*, 21(22), 24920.
12. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Communications of the ACM*, 63(11), 139.
13. Chen, Y., Zhao, X., & Jia, X. (2015). Remote sensing image classification with deep learning: A review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(8), 2381–2385.
14. Conrad, N., Verbeke, G., Molenberghs, G., Goetschalckx, L., Callender, T., Rahimi, K., Mason, J. C., McMurray, J. J. V., & Verbakel, J. Y. (2022). A framework for assessing quality in observational studies: A demonstration using the European Heart Journal. *European Heart Journal*, 43(29), 2862.
15. Tao, X., Ota, K., Dong, M., Qi, H., & Li, K. (2017). Performance guaranteed computation offloading for mobile-edge cloud computing. *IEEE Wireless Communications Letters*, 6(6), 774–777.
16. Wang, S., Zafer, M., & Leung, K. K. (2017). Online placement of multi-component applications in edge computing environments. *IEEE Access*, 5, 2514–2533.
17. Yang, L., Cao, J., Cheng, H., & Ji, Y. (2015). Multi-user computation partitioning for latency-sensitive mobile cloud applications. *IEEE Transactions on Computers*, 64(8), 2253–2266.
18. Cuervo, E., Balasubramanian, A., Cho, D., Wolman, A., Saroiu, S., Chandra, R., & Bahl, P. (2010). Maui: Making smartphones last longer with code offload. In *Proceedings of the MobiSys* (pp. 49–62). ACM.
19. Munoz, O., Pascual-Iserte, A., & Vidal, J. (2015). Optimization of radio and computational resources for energy efficiency in latency-constrained application offloading. *IEEE Transactions on Vehicular Technology*, 64(11), 4738–4755.
20. Lazar, A. (1983). The throughput time delay function of an M/M/1 queue (correspondence). *IEEE Transactions on Information Theory*, 29(6), 914–918.
21. Zhang, G., Zhang, W., Cao, Y., Li, D., & Wang, L. (2018). Energy-delay tradeoff for dynamic offloading in mobile-edge computing system with energy harvesting devices. *IEEE Transactions on Industrial Informatics*, 14(4), 4642–4655.
22. Chen, X. (2015). Decentralized computation offloading game for mobile cloud computing. *IEEE Transactions on Parallel and Distributed Systems*, 26(4), 974–983.

23. Chen, X., Jiao, L., Li, W., & Fu, X. (2016). Efficient multi-user computation offloading for mobile-edge cloud computing. *IEEE/ACM Transactions on Networking*, 24(5), 2795–2808.
24. Yuan, H., Bi, J., Zhou, M., Liu, Q., & Ammari, A. C. (2020). Bi-objective task scheduling for distributed green data center. *IEEE Transactions on Automation Science and Engineering*, 17(2), 770–780.
25. Fortino, G., Messina, F., Rosaci, D., & Sarne, G. M. L. (2020). ResIoT: An IoT social framework resilient to malicious activities. *IEEE/CAA Journal of Automatica Sinica*, 7(6), 1263–1278.
26. Wang, F. Y. (2020). Parallel intelligence: Belief and prescription for edge emergence and cloud convergence in cyber-physical systems. *IEEE Transactions on Computational Social Systems*, 7(6), 1105–1110.
27. Olivares-Rojas, J. C., Reyes-Archundia, E., Rodríguez-Maya, N. E., Gutiérrez-Gnecchi, J. A., Molina-Moreno, I., & Cerda-Jacobo, J. (2020). Machine learning model for the detection of electric energy fraud using an edge-fog computing architecture. In *Proceedings of the 2020 IEEE International Conference on Engineering Veracruz (ICEV)* (pp. 1–6). IEEE.
28. Utomo, D., & Hsiung, P. A. (2020). A multitiered solution for anomaly detection in edge computing for smart meters. *Sensors*, 20(18), 5159.
29. Zhang, Y., Ai, Q., Wang, H., Li, Z., & Zhou, X. (2020). Energy theft detection in an edge data center using threshold-based abnormality detector. *International Journal of Electrical Power & Energy Systems*, 121, 106162.
30. Ashraf, M. M., Waqas, M., Abbas, G., Baker, T., Abbas, Z. H., & Alasmay, H. (2022). FedDP: A privacy-protecting theft detection scheme in smart grids using federated learning. *Energies*, 15(17), 6241.
31. Guillén, M. A., Llanes, A., Imbernón, B., Martínez-España, R., Bueno-Crespo, A., Cano, J. C., & Cecilia, J. M. (2021). Performance evaluation of edge-computing platforms for the prediction of low temperatures in agriculture using deep learning. *The Journal of Supercomputing*, 77(2), 818–840.
32. Kaur, A., & Sood, S. K. (2020). Cloud-fog based framework for drought prediction and forecasting using artificial neural network and genetic algorithm. *Journal of Experimental & Theoretical Artificial Intelligence*, 32(2), 273–289.
33. Lee, K., Silva, B. N., & Han, K. (2020). Deep learning entrusted to fog nodes (DLEFN) based smart agriculture. *Applied Sciences*, 10(5), 1544.
34. Taneja, M., Byabazaire, J., Jalodia, N., Davy, A., Olariu, C., & Malone, P. (2020). Machine learning-based fog computing assisted data-driven approach for early lameness detection in dairy cattle. *Computers and Electronics in Agriculture*, 171, 105286.
35. Cordeiro, M., Markert, C., Araújo, S. S., Campos, N. G., Gondim, R. S., da Silva, T. L. C., & da Rocha, A. R. (2022). Towards smart farming: Fog-enabled intelligent irrigation system using deep neural networks. *Future Generation Computer Systems*, 129, 115–124.
36. Dahane, A., Benameur, R., & Kechar, B. (2022). An IoT low-cost smart farming for enhancing irrigation efficiency of smallholder's farmers. *Wireless Personal Communications*, 127(3), 3173–3210.

37. De Vita, F., Nocera, G., Bruneo, D., Tomaselli, V., Giacalone, D., & Das, S. K. (2021). Porting deep neural networks on the edge via dynamic K-means compression: A case study of plant disease detection. *Pervasive and Mobile Computing*, 75, 101437.
38. Gu, M., Li, K. C., Li, Z., Han, Q., & Fan, W. (2020). Recognition of crop diseases based on depthwise separable convolution in edge computing. *Sensors*, 20(14), 4091.
39. Zhang, R., & Li, X. (2021). Edge computing driven data sensing strategy in the entire crop lifecycle for smart agriculture. *Sensors*, 21(21), 7502.
40. Gupta, N., Khosravy, M., Patel, N., Dey, N., Gupta, S., Darbari, H., & Crespo, R. G. (2020). Economic data analytic AI technique on IoT edge devices for health monitoring of agriculture machines. *Applied Intelligence*, 50(11), 3990–4016.
41. Rajakumar, M. P., Ramya, J., & Maheswari, B. U. (2021). Health monitoring and fault prediction using a lightweight deep convolutional neural network optimized by Levy flight optimization algorithm. *Neural Computing and Applications*, 33(30), 12513–12534.
42. Umarale, D., Sodhani, S., Akhelikar, A., & Koshy, R. (2021). Attention detection of participants during digital learning sessions using edge computing. *SSRN Electronic Journal*, 575–584.
43. Blot, M., Picard, D., Cord, M., & Thome, N. (2016). *Gossip Training for Deep Learning*. *arXiv:1611.09726*.
44. Jin, P. H., Yuan, Q., Iandola, F., & Keutzer, K. (2016). *How to Scale Distributed Deep Learning?* *arXiv:1611.04581*.
45. Daily, J., Vishnu, A., Siegel, C., Warfel, T., & Amatya, V. (2018). *GossipGraD: Scalable Deep Learning Using Gossip Communication Based Asynchronous Gradient Descent*. *arXiv:1803.05880*.
46. Zhao, J., Mortier, R., Crowcroft, J., & Wang, L. (2018). Privacy-preserving machine learning based data analytics on edge devices. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 341–346).
47. Li, Y., Wang, T., Lin, K. K., & Liu, S. (2018). A network-centric hardware/algorithm co-design to accelerate distributed training of deep neural networks. In *Proceedings of the 51st Annual IEEE/ACM International Symposium on Microarchitecture (MICRO)* (pp. 175–188).
48. Wu, C. J., Zhang, J., Xu, Y., & Shafi, K. (2019). Machine learning at Facebook: Understanding inference at the edge. In *Proceedings of the IEEE International Symposium on High Performance Computer Architecture (HPCA)* (pp. 331–344).
49. Han, S., Pool, J., Tran, J., & Dally, W. (2015). Learning both weights and connections for efficient neural networks. In *Advances in Neural Information Processing Systems* (pp. 1135–1143).
50. Han, S., Mao, H., & Dally, W. J. (2015). *Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding*. *arXiv:1510.00149*.
51. Chen, Y. H., Emer, J., & Sze, V. (2016). Eyeriss: A spatial architecture for energy-efficient dataflow for convolutional neural networks. *ACM SIGARCH Computer Architecture News*, 44(3), 367–379.
52. Yang, T. J., Chen, Y. H., & Sze, V. (2017). Designing energy-efficient convolutional neural networks using energy-aware pruning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 5687–5695).

53. Reagen, B., Wentzlaff, D., & Mudge, T. (2016). Minerva: Enabling low-power, highly-accurate deep neural network accelerators. *ACM SIGARCH Computer Architecture News*, 44(3), 267–278.
54. Liu, S., Lin, Y., Zhou, Z., Nan, K., Liu, H., & Du, J. (2018). On-demand deep model compression for mobile devices: A usage-driven model selection framework. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services* (pp. 389–400).
55. Oh, Y. H., Yang, J., Lee, H., & Kim, S. (2018). A portable, automatic data quantizer for deep neural networks. In *Proceedings of the ACM PACT* (p. 17).
56. Lane, N. D., Georgiev, P., & Lin, M. (2016). DeepX: A software accelerator for low-power deep learning inference on mobile devices. In *Proceedings of the 15th International Conference on Information Processing in Sensor Networks* (p. 23).
57. Zeng, L., Li, E., Zhou, Z., & Chen, X. (2019). Boomerang: On-demand cooperative deep neural network inference for edge intelligence on industrial Internet of Things. *IEEE Network*.
58. Kang, Y., Ryu, J., Choi, H., & Kim, H. (2017). Neurosurgeon: Collaborative intelligence between the cloud and mobile edge. *ACM SIGPLAN Notices*, 52(4), 615–629.
59. Hu, W., Bao, D., Wang, D., & Liu, F. (2019). Dynamic adaptive DNN surgery for inference acceleration on the edge. In *Proceedings of the IEEE INFOCOM* (pp. 1–9).
60. Li, H., Hu, C., Jiang, J., Wang, Z., Wen, Y., & Zhu, W. (2018). JALAD: Joint Accuracy-and Latency-Aware Deep Structure Decoupling for Edge-Cloud Execution. *arXiv:1812.10027*.
61. Jeong, H. J., Lee, H. J., Shin, C. H., & Moon, S. M. (2018). IONN: Incremental offloading of neural network computations from mobile devices to edge servers. In *Proceedings of the ACM Symposium on Cloud Computing* (pp. 401–411).
62. Ko, J. H., Na, T., Amir, M. F., & Mukhopadhyay, S. (2018). Edge-Host Partitioning of Deep Neural Networks with Feature Space Encoding for Resource-Constrained Internet-of-Things platforms. *arXiv:1802.03835*.
63. Mao, J., Goh, H. C., Rahimian, M., & Raghavan, V. (2017). McDNN: A distributed mobile system with enhanced partition and deployment for large-scale DNNs. In *Proceedings of the 36th International Conference on Computer-Aided Design* (pp. 751–756).
64. Mao, J., Chen, X., Nixon, K. W., Krieger, C., & Chen, Y. (2017). MoDNN: Local distributed mobile computing system for deep neural networks. In *Proceedings of the Design, Automation & Test in Europe Conference & Exhibition (DATE)* (pp. 1396–1401).
65. Zhao, Z., Barijough, K. M., & Gerstlauer, A. (2018). Deep Things: Distributed adaptive deep learning inference on resource-constrained IoT edge clusters. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 37(11), 2348–2359.
66. Teerapittayanon, S., McDaniel, B., & Kung, H. (2016). Branchy Net: Fast inference via early exiting from deep neural networks. In *Proceedings of the 23rd International Conference on Pattern Recognition (ICPR)* (pp. 2464–2469).

67. Teerapittayanon, S., McDaniel, B., & Kung, H. T. (2017). Distributed deep neural networks over the cloud, the edge, and end devices. In *Proceedings of the 37th IEEE International Conference on Distributed Computing Systems (ICDCS)* (pp. 328–339).
68. Bolukbasi, T., Wang, J., Dekel, O., & Saligrama, V. (2017). *Adaptive Neural Networks for Efficient Inference*. arXiv:1702.07811.
69. Lo, Y. Y., Su, C. Y., Lee, C., & Chang, S. C. (2017). A dynamic deep neural network design for efficient workload allocation in edge computing. In *Proceedings of the IEEE International Conference on Computer Design (ICCD)* (pp. 273–280).
70. Leroux, S., Gagnon, S., & Beaulieu, C. (2017). The cascading neural network: Building the Internet of smart things. *Knowledge and Information Systems*, 52(3), 791–814.
71. Chen, T. Y. H., Ravindranath, L., Deng, S., Bahl, P., & Balakrishnan, H. (2015). Glimpse: Continuous, real-time object recognition on mobile devices. In *Proceedings of the ACM Conference on Embedded Networked Sensor Systems (SenSys)*.
72. Drolia, U., Guo, K., Tan, J., Gandhi, R., & Narasimhan, P. (2017). Cachier: Edge-caching for recognition applications. In *Proceedings of the IEEE International Conference on Distributed Computing Systems (ICDCS)* (pp. 276–286).
73. Drolia, U., Guo, K., & Narasimhan, P. (2017). Precog: Prefetching for image recognition applications at the edge. In *Proceedings of the ACM/IEEE Symposium on Edge Computing (SEC)* (p. 17).
74. Guo, P., Hu, B., Li, R., & Hu, W. (2018). Foggy Cache: Cross-device approximate computation reuse. In *Proceedings of the ACM MobiCom* (pp. 19–34).
75. Venugopal, S., Gazzetti, M., Gkoufas, Y., & Katrinis, K. (2018). Shadow puppets: Cloud-level accurate AI inference at the speed and economy of edge. In *Proceedings of the USENIX Workshop on Hot Topics in Edge Computing (HotEdge)*.
76. Kang, J., Emmons, F., Abuzaid, P., Bailis, P., & Zaharia, M. (2017). No scope: Optimizing neural network queries over video at scale. *Proceedings of the VLDB Endowment*, 10(11), 1586–1597.
77. Wang, J., Huang, K., Jiang, J., Zhao, X., & Zhang, J. (2018). Bandwidth-efficient live video analytics for drones via edge computing. In *Proceedings of the IEEE/ACM Symposium on Edge Computing (SEC)* (pp. 159–173).
78. Kao, Y. H., Krishnamachari, B., Ra, M. R., & Bai, F. (2015). Hermes: Latency optimal task assignment for resource-constrained mobile computing. In *Proceedings of the 2015 IEEE Conference on Computer Communications (INFOCOM)* (pp. 1894–1902).
79. Liu, J., Mao, Y., Zhang, J., & Letaief, K. B. (2016). Delay-optimal computation task scheduling for mobile-edge computing systems. In *Proceedings of the 2016 IEEE International Symposium on Information Theory (ISIT)* (pp. 1451–1455).
80. Carson, K., Thomason, J., Wolski, R., Krintz, C., & Mock, M. (2019). Mandrake: Implementing durability for edge clouds. In *Proceedings of the 2019 IEEE International Conference on Edge Computing (EDGE)* (pp. 95–101).
81. Zhao, J., Li, Q., Gong, Y., & Zhang, K. (2019). Computation offloading and resource allocation for cloud-assisted mobile edge computing in vehicular networks. *IEEE Transactions on Vehicular Technology*, 68(8), 7944–7956.
82. Lyu, X., Ni, W., Tian, H., Liu, R. P., Wang, X., Giannakis, G. B., & Paulraj, A. (2017). Optimal schedule of mobile edge computing for the Internet of Things using

- partial information. *IEEE Journal on Selected Areas in Communications*, 35(12), 2606–2615.
83. Mao, Y., Zhang, J., & Letaief, K. B. (2017). Joint task offloading scheduling and transmit power allocation for mobile-edge computing systems. In *Proceedings of the 2017 IEEE Wireless Communications and Networking Conference (WCNC)* (pp. 1–6).
84. Liu, L., Chang, Z., Guo, X., & Ristaniemi, T. (2017). Multi-objective optimization for computation offloading in mobile-edge computing. In *Proceedings of the 2017 IEEE Symposium on Computers and Communications (ISCC)* (pp. 832–837).

PART III

Computational Intelligence: Edge Computing and AI applications



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

CHAPTER 7

Future Innovations Using Edge Computing Infrastructure in IoT Applications

NEHA TYAGI,¹ JITENDER MITTAL,¹ DLER SALIH HASAN,²
SUNIL BAJEJA,³ and SHASHIDHAR SONNAD⁴

¹Computer Science and Engineering, Amity University, Noida, Uttar Pradesh, India

²Computer Science and Information Technology, College of Science/ University of Salahaddin, Erbil, Iraq

³Faculty of Computer Applications, Marwadi University, Rajkot, Gujarat, India

⁴Department of Electronics and Communication Engineering, Sharnbasva University, Kalaburagi, Karnataka, India

ABSTRACT

Although edge computing is a crucial component of creating 5G networks, there are not many edges computing-based applications or innovations in the nation. In the past several years, the Internet of Things (IoT) has seen an enormous increase in the volume of data produced by detectors, actuators, and other equipment. Presently, IoT data is handled in the cloud, primarily using computer power from far-off data centers. As a result, significant limitations in network bandwidth and delay in communication develop. This study promotes future innovation of edge computing in IoT applications that use sensor flows to enhance interactive applications.

By categorizing future innovations in terms of the shared cooperation of radio and computer systems, system optimization, and smart administration of resources, this study offers a comprehensive view of the present difficulties and efficient solutions. Following the processing of the human activity statistics, the actions of the associated surveillance items are identified, modeled, and modified using edge device dispersed computation. These systems are distinguished by their quick response time and minimal usage of energy. The outcomes of the study support. From a patent viewpoint, edge computation innovation is now far more effective at being recognized.

7.1 INTRODUCTION

The Internet is moving toward a future IoT that might connect a trillion or more gadgets. According to Ref. [1], over 50 billion gadgets will be connected to the Web by 2025.

Through the employment of numerous high-performance computers in the data point, cloud computing innovation gives customers access to virtually limitless computing capacity. It is among the essential techniques for processing and analyzing huge data [2]. Because of its enormous computing and storage capability, the current cloud-based computation system can manage the enormous volume of data coming from scattered IoT gadgets [3, 4].

Nevertheless, this is not appropriate for dispersed IoT networks or the real-time functioning of dependent-on-latency IoT applications due to its central method of functioning and the associated latency [5, 6]. Real-time reaction to collisions, errors, and catastrophes, for instance, is crucial in situations involving industrial manufacturing and functioning, and data transfer expenses are particularly pertinent in network data collection situations [7]. Edge computing innovation has been developed and vigorously encouraged as a result [8], and it is the foundation of edge information processing innovation. Edge computation is referred to as a dispersed accessible system that combines network-based storage, computation, and application fundamental features at the system's edge, near the origin of the items or data [9, 10]. Edge computation represents a fresh environmental paradigm. To enhance system assistance performance and accessible networking management skills, 5 different kinds of resources—network,

processing, storage, usage, and intelligence—are converged at the system's edge [11]. This facilitates the development of a network that is analogous to the wireless web.

It is critical to foster collaboration between cloud and edge computation to simplify the handling and evaluation of massive amounts of unorganized IoT data. Edge computing gadgets, in addition to offering computational power close to consumers, can undertake a variety of initial activities such as data classification and sorting, service-level compliance scoring, and variable assessments before engaging the central cloud [12, 13]. Additionally, cooperative processing concerning resource distribution, balance of load, and offloading duties involving edge and cloud computing can give substantial assistance to both latency-sensitive and compute-intensive operations in cellular IoT systems [14].

The majority of the gadgets will be positioned near the network's edge and may bring novel applications, altering numerous elements of conventional commercial manufacturing and daily lives. Apple timepieces, Oculus Rift protective gear [15, 16], Google Home [17, 18], Fitbit fitness monitors, and Google Glasses are among the products that have already emerged [19]. Edge IoT gadgets may include any type of detectors and chips with varying potential created by multiple producers and numerous uses may be developed to allow smart houses, medical care, transport, structures, and towns [20, 21]. Since the majority of the computing, preservation, and connecting facilities are contained in these broad, distributed data points that are controlled by the application service suppliers, it is extremely prevalent for these numerous edge gadgets to need to collaborate tightly with the application platforms [22, 23].

Professionals from both home and overseas have conducted numerous investigations on machine learning (ML) edge computation clone node identification [24, 25]. TangX and ChansonST suggested an optimum task downloading allocation technique in foreign nations depending on an investigation of fog and cloud combo. Through empirical modeling, it is feasible to conserve the communication spectrum and decrease network latency beneath the circumstance of service latency restriction by surrendering just a fraction of computer resources, so that the cloud's usage of energy may be kept to a low [26, 27].

Hence this study aims to investigate the future innovation of Edge computing in IoT applications based on a patent viewpoint.

7.2 LITERATURE REVIEW

This part examines the research literature on relevant studies conducted in the fields of IoT, IoT designs, and edge computing.

[28] Gave an in-depth examination of IoT, covering future innovations, IoT designs, and confidentiality and safety concerns. In the context of applications, they also investigated merging IoT with edge computing. They nevertheless failed to investigate IoT applications [29]. Present an in-depth look into IoT-enabling innovation, applications, and standards [30]. Analyzed IoT apps by prevalence and consumer kind across four categories: personal, smart surroundings, houses, and cars [31]. Investigated 20 IoT instances from three major IoT applications: smart areas, towns, and houses. They divided IoT applications into various groups: (i) settings depending on the areas and the kind of IoT usage; (ii) subjects: which emphasizes usage dimensions, that is, the number of individuals and gadgets; (iii) sensors: which emphasizes on needed sensor kinds in the IoT usage; (iv) network specifications, like system pace; and (v) other: for instance, application safety demands [32]. Discussed the concept of the IoT framework, assessed IoT-enabling innovation, and investigated prospective IoT applications [33]. Categorized IoT applications depending on approaches designed to assist difficulties in the IoT sector. The study conducted in [34] looked at IoT designs for domain-dependent applications and quickly highlighted IoT cloud platforms that allow data assessment [35]. Compared IoT gateway platforms to IoT needs, covering design demands [36]. Examined IoT uses for blockchain technology [37]. Provided cloud-of-things middleware and investigated different design approaches of middleware and application areas [38]. Developed IoT taxonomies, investigated IoT frameworks in the form of 3-layer, 5-, cloud, edge-based designs, and societal IoT, which emphasizes the connection between IoT devices and uses [39]. Conducted a review of IoT uses in the context of developing situations and investigated open study issues for IoT frameworks [40]. Researched the IoTs in the context of ideas and vision, uses, innovations, and study problems [41]. Examined IoT uses that advantage from edge computation and then developed a standard design for IoT, the framework of which depends on software-described networking (SDN) and edge computation.

7.3 METHODOLOGY

The 21st era is the age of the information economy. The ability of an organization to survive and thrive in the face of strong marketplace

competitiveness is more reliant on its intellectual and technical innovation capacity, as well as the successful administration and use of its intellectual property claims [42]. In the course of engaging in academic study and commercial tasks, businesses have created a huge number of patent successes [43]. Just by integrating patents and the marketplace setting and leveraging the outcomes of these patents will we be able to constantly drive the imaginative passion of partners while also improving the enterprise's technological competence and commercial competence [44].

Patents constitute intellectual property and are the primary supply of technological data, encompassing 90%–95% of the globe's academic and technology data. A patent is a popular indicator used to assess the output of corporate innovation operations. Patent data from diverse firms is generally straightforward to acquire and can more accurately represent enterprises' innovative efforts [45].

The patent portfolio concept employs patent actions and standards as assessment indices for company patents. Using the variety of patent uses as a sign of patent action in an enterprise's patent portfolio, it is considered that enterprises with a large number of patent uses are highly innovative, and their commercial potential is likewise stronger. The lot of patent use that an enterprise files represents not just the variety of patents it has, but also the level of action in its invention. The higher the enterprise's yearly patent uses, the more robust the enterprise's innovation efforts [46, 47].

7.3.1 SYSTEM FRAMEWORK OF EDGE COMPUTING

Edge computation may have greater depth and complicated features than previous versions. Its conceptual study frameworks and methodologies each have its own set of suitable ranges. The portion and the entirety are no longer united, and they fail to meet the concept of superposition of linear structures. Especially for other recognized structures, modeling and controlling them is hard. When the framework of the regulated item is fully unresolved structure study becomes more difficult. This additionally renders edge computation system framework detection and management an explosive subject in the present management area [48, 49]. In recent times, AI has been frequently deployed in the detection of black-box devices. Amongst them, the local finances have gotten a lot of interest. The illustrating of nonlinear structures has been revitalized by its exceptional edge computation capacities.

Finding that data interaction is inextricably linked to real-time information transfer, management of information and data gathering is not hard when studying information services. Loss of information or latency can be brought on by the incorrect choice of interaction mechanisms. Thus, the standard and channel for interaction that is chosen are crucial. For network information, the edge gadget itself must use the network component to reach the targeted webpage, collect the necessary information, and finish the data collection process [50]. Additional equipment is not required for data gathering because the edge gadgets linked to the cloud-based computation unit and the web conduct the transfer of information by wireless transmission [51]. The framework of the edge computing system is shown in Figure 7.1.

Regional users are served through edge computation. On the contrary, it may shorten processing times for services and increase productivity. On the contrary, it can lessen the need for networking and bandwidth and minimize structure overhead. Edge computation outperforms other computations in terms of response duration and service standard, and it satisfies the needs for minimal delay, strong dependability, and safety [52]. Edge computation also relieves the strain on the data center, lowers bandwidth needs, equalizes data processing, and increases total system performance as an add-on to other computations [53]. Edge computation has been increasingly popular in recent decades due to the IoTs' quick expansion in a variety of industries, including the Internet of Automobiles, wireless detectors and motors, smart houses, and software-described networking. Edge computation will eventually become increasingly prevalent in more disciplines and businesses, add to other computations, connect with it, and provide an appropriate technological assistance system for data processing in the IoT age [54].

Edge computation's real-time data assistance framework separates real-time data assistance into cloud usage assistance for edge computation. The primary task of real-time data assistance is handled by edge computation assistance. As the data recipient for edge computation assistance, cloud utilization assistance primarily offers web-based services. Database assistance is provided by two providers. Big data, cloud-based computation, and digitization have all benefited from the emergence of edge computation. It now spans a wide range of industries, including health-care, farming, geology, cosmology, and the IoTs. It has also advanced into the journalism and digital government industries. Enormous information's

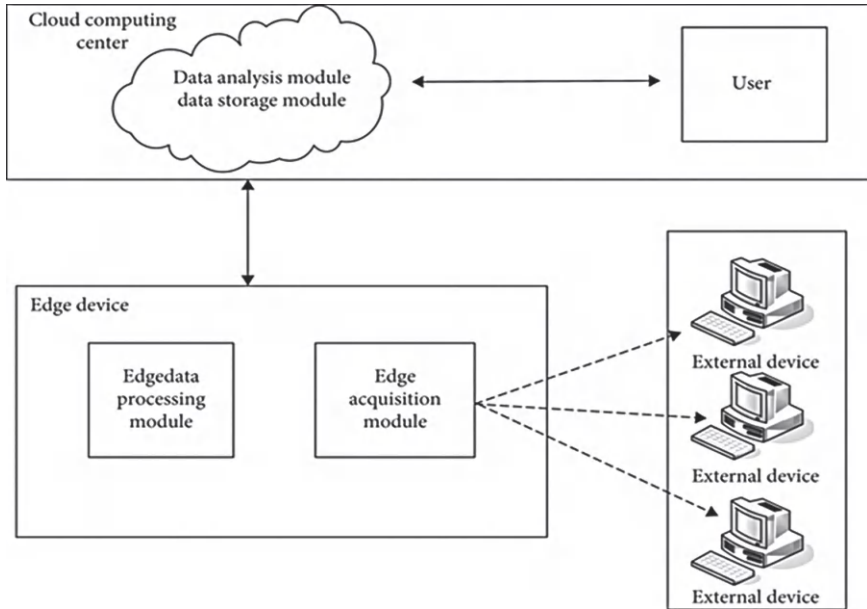


FIGURE 7.1 The framework of the edge computing system.

enormous worth opens up new prospects for progress in every industry [55]. Nevertheless, the production of enormous amounts of data also poses significant difficulties for data processing. It necessitates not only powerful computation and analytical skills but also a sizable storage area to house the data. The computational unit will likely experience too much strain as a result. Edge computation effectively addresses this issue.

7.3.2 DATA COLLECTION

Perform the data evaluation, gather data using techniques like AI and ML, reduce the divergence from the original data, and rectify the data to produce more precise data; provide a dispersed edge computation and cloud-based computation cooperative big data assessment and mining system to realize data creation and module-specific capability [56]; An integrated software framework often finishes the instructions and schedules, creates long-lasting steady interactions, and assures the proper functioning of the entire structure [57, 58]. The communication between the operational components is also quite straightforward.

A human activity data gathering, processing, and evaluation platform built on edge computation to recognize human actions in real-world settings. A human activity identification platform with outstanding performance, minimal energy usage, and quick reaction is achieved by employing dispersed computation of edge gadgets to alter these assumptions. A structure that may be directly modeled is acquired after data collecting and data preprocessing. Systems like classification and forecasts, cluster examination, and connection criteria may be built to assist system users in obtaining the significance contained by the mining aim and data type. It is vital to choose an appropriate microprocessor as the central component of the total system-edge computation to be able to fully embrace the aforementioned hardware's general layout and accomplish the related computation features [59].

The edge sensor receives the framework and then integrates the real-time data it receives to validate the fundamental framework. The structure will be restored to the cloud computation facility when the consumer or the edge unit it judges that the outcomes generated during the application of the framework do not correspond to the actual scenario. The cloud computation unit modifies the framework using the freshly obtained data and comments on outcomes, then transmits the updated system to the edge device for an additional evaluation. This process is repeated until the system's evaluation outcome satisfies the edge gadget's demands, at which point an outcome is produced. On edge systems, the design will be utilized routinely [60].

The framework's total operational design, which is composed of the data collection component, data preprocessing component, and data evaluation component, may then be acquired. This strategy of placing the assessment designs on the edge system can, on the one hand, lessen the load on the cloud's processing resources, and, on the contrary, for situations where latencies are a concern, the edge device's ability to conduct the data assessment job may successfully decrease the turnaround time for results. The platform's users receive timely presentations of the outcomes of the real-time assessment.

The edge device's pertinent sensors are employed to finish the real-time data gathering necessary for human body acknowledgment, after which the data is appropriately analyzed and delivered to the cloud. The cloud computation unit completes the modeling process and sends the framework back to the edge sensor using associated ML methods. The

learned model is used by the edge sensor to anticipate the posture activity of the presently tracking item [61]. When the framework's anticipated outcome fails to correspond to the existing state of affairs, it may alert the cloud-based computation unit of the precise outcome to the appropriate framework.

7.3.3 DATA SOURCE

Edge devices regulating numerous sensors will finish the gathering of tangible information in this framework. In the contexts of human action acknowledgment, devices that can be utilized as data sources for framework learning comprise quantitative data-gathering devices like gyroscopes, IMUs, and cardiac monitoring devices, as well as visual information catchers like thermal and infrared visual collection devices. In this case, IMU hardware is employed for gathering human activity data based on system specifications 1 and system attributes.

7.3.4 EQUIPMENT OF EDGE COMPUTATION

The present data processor component and the human activity identification component, which will be created later, are responsible for collecting and distributing gathered data, and an edge system that satisfies structure criteria is necessary to allow the execution of the aforementioned operations. The most commonly used and newly invented system is the most developed. Here is mostly an assortment of goods classified as systems, like the Raspberry Pi and Arduino. The Raspberry Pi B+, RPi2, and UNO versions of the Arduino are popular gadgets.

7.3.5 DATA COLLECTION OF HUMAN ACTIVITY

This study necessitates the tracked item perform the associated activities in line with specified directions for the action labels needed by the study, such as an array of everyday activities and certain actions beneath unique conditions, like Falling, through offering devices like as embedded mobile strength, sequential port, and Bluetooth connect to guarantee that the information gathered is more appropriate for model evaluation. When

the detector is switched on, the data collection begins to function autonomously and transmits the acquired tangible information to the sequential port at the rate specified.

7.3.6 DATA

Every data evaluation in this paper is done with SPSS19.0, the statistical evaluation is a two-sided assessment, relevance is described as 0.05, and $p < 0.05$ is regarded as relevant. The statistical findings are presented as mean (+/–) standard deviation (SD). When the examination data follows a normal pattern, the double T-test is employed to compare inside the category, while the autonomous sample T-test is employed to compare between the categories.

7.4 RESULTS AND DISCUSSION

7.4.1 PATENT MODIFICATION OF EDGE COMPUTATION

This study compiled data on the modifications in edge computation patents over the past few decades using literature reviews and data from the patent position, displayed in Table 7.1.

TABLE 7.1 Patents of Computing

	2016	2017	2018	2019	2020
Big data	12	17	13	9	10
Cloud computing	18	15	25	33	28
Fog computing	11	8	14	18	15
Edge computing	9	5	12	17	21

Table 7.1 shows that as time passes, future innovation continues to advance, and there is a rising number of edge computation patents. In 2016, the number of edge computation patents was roughly half that of cloud computing. However, after 5 years of expansion, edge computation patents have begun to look at cloud computing and have even begun to demonstrate a pattern outside cloud computing, indicating that edge computation has a lot of room to grow. In the experiment described in this paper, we must use equipment to gather data on human mobility. This

study initially evaluates the data collection capability of various pieces of equipment to guarantee the actual reliability of the data.

The data gathered by RPi2 is the most solid among the 3 equipment, with tiny changes, broad memory, and a significant quantity of data being gathered. The data gathered by the UNO version of Arduino is the most accurate, and the variations are also just barely greater than RPi2, but its storage capacity is little, and the variety of data gathered is tiny. RPi B + has the biggest acquiring spectrum, and its optimum significance is also the biggest. After careful deliberation, we chose to use RPi2 hardware for collecting.

7.4.2 VARIOUS ALGORITHM IMPACTS

In Table 7.2, this study gathered the data using various techniques and contrasted how the results changed as they ran.

TABLE 7.2 Data Variation of Operation

	Accuracy	F-value	Operation Periods
MDD	0.936	0.923	43.5
Edge computing	0.984	0.964	18.6
RF	0.948	0.932	36.2
NB	0.946	0.944	35.2
DT	0.932	0.933	28.7

By analyzing the data, it can be demonstrated that edge computation performs better than other fundamental methods in terms of speed as well as precision, and recall. It combines learning in contrast to fragile classifiers like decision trees. The technique’s benefits are more plainly visible. Edge computation offers a higher degree of self-model normalization than other types of modeling, giving this sort of system a larger capacity for generality. With more data being generated, edge computing accuracy will become more apparent in practical applications.

This study allows multiple techniques to operate independently and tally usage to show the expense of executing different techniques. Various computer implementation outcomes may deviate because of variations in hardware setup, but there shouldn’t be much variation in the comparative execution times of various methods. Figure 7.2 displays the precise data.

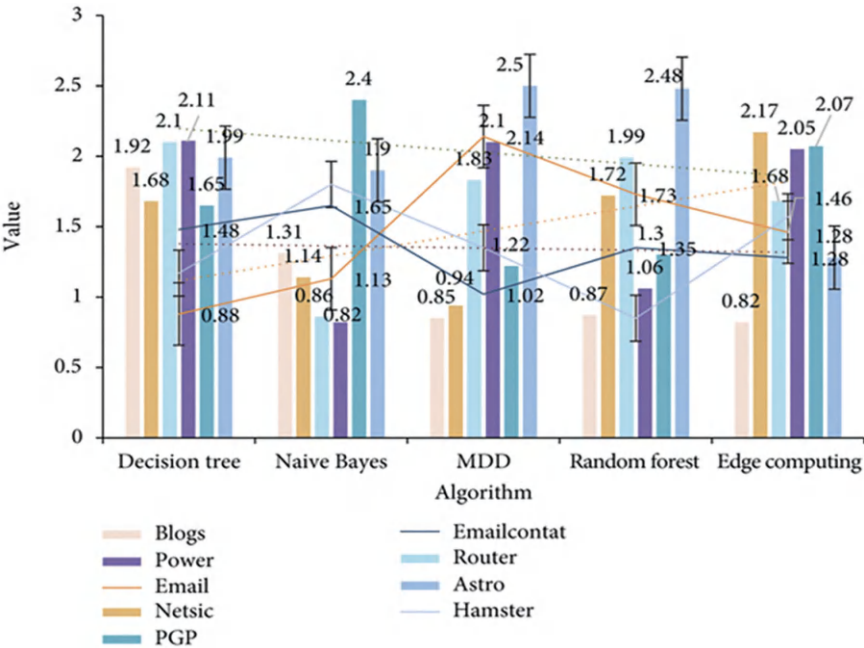


FIGURE 7.2 Computation duration of the method.

Figure 7.2 demonstrates that the benefits of edge-based computation in terms of processing cost are not immediately apparent. Two signs have costs that are higher compared to other computation techniques, whereas other signs have costs that are significantly less than other techniques. It is approximately 40% less than other techniques, especially in terms of Bloggers and Astro.

Figure 7.3 depicts the monotony of the rating of multiple techniques in the real system. As may be observed, the edge-based approach has clear benefits in this area. Its position monotony is greater than that of Naive Bayes (NB), MDD, and other methods, and its method's rating monotony is over 30% greater than that of other methods. Figure 7.4 depicts the computation duration of several methods per time unit.

Figure 7.4 shows that at the start of various methods, there is a slight variance in functioning performance between them, yet after conducting over six hundred instances, the performance between the methods starts to rise, and the rate of edge computation is contrasted to other methods.

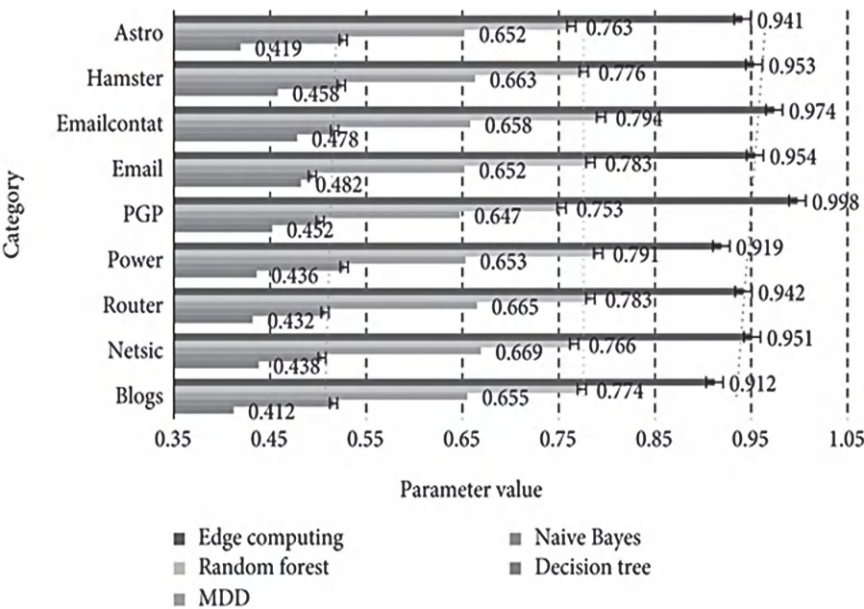


FIGURE 7.3 Monotonicity ranking of various methods.

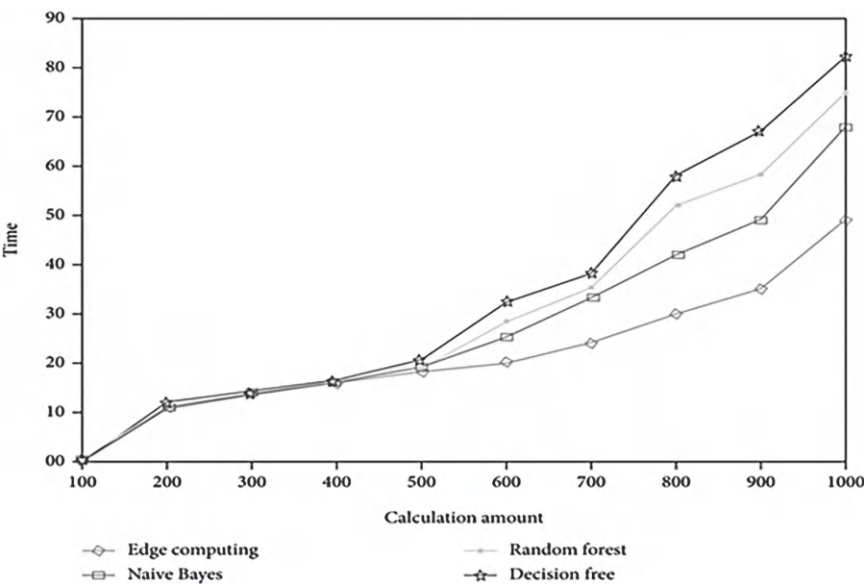


FIGURE 7.4 Duration of operation for various algorithm unit activities.

7.4.3 CLASSIFICATION OF FUTURE INNOVATIONS OF EDGE-IOT APPLICATIONS

The various QoS needs of IoT devices, which intrinsically combine mobile and computation assistance, demand modern interaction and computation innovations. In addition, statistical analysis and network intellect may serve critical roles in these networks to deal with unprecedented levels of data and unpredictable behavior of devices linked to them.

7.4.4 COORDINATED RESOURCE ADMINISTRATION

- 1. Communications:** Future wireless systems will need complex resource administration systems to increase spectrum and energy effectiveness, and also to prevent cross-tier disruption [62]. Additionally, the overhead of sharing data within the main networking may restrict total systems performance. To that aim, Cloud RAN is seen as a future innovation since it enables effective coordination among network parts under a cohesive regulator, and numerous radio-accessing innovations may be linked to optimize networking capability through this SDN-oriented centralized regulator. Furthermore, rather than outsourcing all work to edge servers, Device-to-Device (D2D) innovation can lessen server loads by allowing mobile applications to benefit from accessible computation resources in neighboring gadgets in a collaborative way [63].
- 2. Computation:** It is critical to develop a layered framework between edge and cloud systems to provide adaptable provision of computational resources with satisfying QoE like delay, power, and Central Processing Unit operations. The Fog-to-Cloud architecture [64] is an instance of cloud-fog coordinated operation combining resources from fog and cloud networking to leverage total computation resources. In future computation structures, any node, such as consumer electronics, could participate in computing processes. Among the execution that allows user gadgets to share their accessible computational power with external gadgets, or more, with the cloud and fog systems, is the idea of the user as a supplier. Furthermore, vehicle and drone-oriented fogs aid in offering on-demand computation assistance to nearby cellular consumers with minimal delay.

3. **Combined communication-computation:** Providing SBSs in future mobile systems with computational functionality will transform SBSs. Small-cell cloud-enhanced eNodeB is a real-world deployment of intelligent SBSs in an LTE-oriented architecture that combines mobile, cloud, and edge computation [65]. On the contrary, SBSs are linked to the wireless base through backhaul connections, while on the other side, SBSs are linked to portable handsets for program provision. In addition, utilizing SBSs' real-time connectivity to the main system supports SBSs in performing computation-related operations such as cache generation employing system-wide information. Both networking suppliers and cloud functions have to function together to optimize the exchange of backhaul lines, range, and server usage. Also, the interaction of cloud and edge computation can enable a wide range of IoT applications with varying QoS needs [66].

7.4.5 SMART EDGE-IOT APPLICATIONS

The massive volume of data gathered through IoT gadgets is a great resource of information sets that can aid in teaching a machine and learning different characteristics. This feature has allowed the use of ML approaches in a range of computation elements like data caches, task offloading, and gadget grouping to facilitate the decentralized functioning of dispersed IoT platforms with the help of a central cloud.

Content Cache in Edge Gadgets: The latest advancements in mobile designs allow networking components to have an in-depth understanding of all accessible resources. Furthermore, the presence of SBSs in the area of cellular users gives precise spatiotemporal data on preferred material, resulting in a more precise content cache and enhanced utilization of the restricted storage space.

Big Data Analysis with Edge Computation: The closeness of edge gadgets to cellular users facilitates the study and assessment of big data acquired from computing-hungry uses, particularly those containing signal processing operations via collection, sorting, and pre-processing [67]. Big data can be utilized to accurately anticipate famous content, store that content in mobile SBSs, and offer fast responses to end clients without placing additional strain on backhaul infrastructure [68, 69]. But for the

edge gadgets with limited resources and processing power to use them, the complicated data assessment methods that are often used for the cloud need to be reduced [70].

Deep Learning as a Method for Edge Computation: Making smart choices dependent on real-time training information is necessary to take advantage of the advantages offered by both edge and cloud-based computing systems. Learning-oriented methods employing both short- and long-term data appear potential in this context to deal with the diverse and flexible characteristics of portable IoT systems. For example, depending on the instant predictable circumstances, it is possible to determine whether to offload tasks with strict latency-sensitive needs. In contrast, making judgments for applications that have lesser latency-sensitive needs based on probabilistic long-term system effectiveness may be more effective [71]. To do this, a variety of ML techniques, such as neural networks (NN), support vector machines (SVM), k-means, and linear regression, can be used to make smart choices in a variety of network operating tasks, such as balance of load, fault administration, and adaptable allocating resources [72].

IoT Devices Grouping: IoT gadgets can be categorized and grouped due to specific criteria like operation and places rather than providing every IoT information to the edge system or central cloud. For example, house monitors may be analyzed by a centralized household manager, which may be fitted with ML capacity to identify any unexpected data (such as air pollution, sound, and doors accessed at uncommon daily times) and communicate this data to the house owner or competent officials. Furthermore, inside a neighborhood, data acquired from devices that encounter comparable circumstances, like position, environment, and climate evaluations, may be further analyzed and confirmed by the community manager. In this manner, redundant data transfer from countless gadgets may be prevented, lowering both connectivity and computation expenses. Furthermore, inter-cluster connectivity may be managed by employing other media like wireless [73].

7.4.6 ADAPTABLE ARCHITECTURE ADMINISTRATION

The unprecedented volume and diversity of linked items will inevitably complicate network administration and governance in the IoT age. To ensure adaptability, re-setup, and scaling, adaptable administration solutions must be implemented.

1. **SDN:** The implementation of a multilayer fog design necessitates adaptable package forward systems between numerous fog tiers, as well as from fog tiers to the cloud tier. SDN innovation aids in this procedure, particularly with the diversity and mobility properties of IoT gadgets (such as wearable gadgets). Furthermore, fog units may be incorporated into specific mobile base stations (BS) or shared by numerous BSs. In this case, an SDN-oriented mobile base that uses OpenFlow regulators and standards can improve transmission of data by reducing, mobility administration, authorization, and networking virtualization. Additionally, when a gadget travels between fog units, SDN assists in determining the ideal profitability based on end-to-end latency and mobile traffic and activates the Virtual Machine movement command based on the expense assessment [74].
2. **NFV:** By virtualizing multiple network services on common equipment, NFV innovation can provide greater architecture management, data exchange, and network safety. Furthermore, it minimizes the complication of network and IoT administration [75], and it serves a vital part in system performance improvements by adjusting radio, compute, and architecture assets based on the time of operation requirements [76]. Content distribution networks and the Platform as a Service approach, which enables customers to use a specific piece of software on shared servers, are a couple of instances of NFV use cases.
3. **Adaptable Radio:** New cognitive radio innovations are allowing mobile BSs to adjust their radio resources in response to requests, distribute the most effective frequency resource, and determine whether the network layer (such as macro, Pico, or femtocells) is most suited to fulfill a certain requirement.

7.5 CONCLUSION

The “big data” period has progressively swept through society in recent decades, and the emergence of cloud-based computation has increased the need for big data analysis and application capabilities. Because of its benefits, including reduced expenses for operation, flexible adaptability, and streamlined management and maintenance, cloud

computation has been extensively advocated. Firms relating to cloud computation have also grown quickly in the nation. To make dispersed IoT systems viable, edge computation and the central cloud together form a potent computation architecture. In this regard, this study offered a thorough overview of current study challenges and future innovations of edge computing, and it also introduced a distinctive structure for the optimization of important performance indicators in edge-IoT applications. The implementation of a data gathering and processing platform dependent on edge computation from a patent standpoint is the focus of this study. The original structure's issues with slow computation rate and excessive usage of energy are resolved by the cooperation and communication of edge sensors and cloud computation facilities. The system's various uses in various contexts are given the appropriate designs and execution.

The future IoT application necessitates considerable initiatives in the guidance of establishing self-organizing and self-resilient computation structures, as well as dealing with the diversity of IoT applications. Also, since interaction and computation are naturally incorporated in IoT frameworks, computation system optimizing involves taking into consideration the restrictions caused by mobile network restrictions, whether the restrictions are associated with inadequate assets, time-sensitive wireless connections, or additional problems like energy and expenses. Furthermore, combined communication-computation platforms with learning aspects, as well as adaptable architecture and administration of resources, must cohabit alongside dispersed IoT platforms.

KEYWORDS

- **big data analysis**
- **cloud computation**
- **edge computing**
- **internet of things**
- **resource optimization**
- **self-organizing systems**

REFERENCES

1. Ericsson Inc. (2010). *CEO to Shareholders: 50 Billion Connections 2020*. Retrieved from: <http://www.ericsson.com/thecompany/press/releases/2010/04/1403231> (accessed on 17 October 2024).
2. Sarkar, S., Chatterjee, S., & Misra, S. (2015). Assessment of the suitability of fog computing in the context of Internet of Things. *IEEE Transactions on Cloud Computing*, 6(1), 46–59.
3. Botta, A., De Donato, W., Persico, V., & Pescapé, A. (2016). Integration of cloud computing and Internet of Things: A survey. *Future Generation Computer Systems*, 56, 684–700.
4. Babu, S. M., Lakshmi, A. J., & Rao, B. T. (2015). A study on cloud-based Internet of Things: CloudIoT. In *2015 Global Conference on Communication Technologies (GCCT)* (pp. 60–65). IEEE.
5. Yang, Z., Zhou, Q., Lei, L., Zheng, K., & Xiang, W. (2016). An IoT-cloud-based wearable ECG monitoring system for smart healthcare. *Journal of Medical Systems*, 40, 1–1.
6. Hossain, M. S., & Muhammad, G. (2016). Cloud-assisted industrial Internet of Things (IoT)-enabled framework for health monitoring. *Computer Networks*, 101, 192–202.
7. Petrolo, R., Loscri, V., & Mitton, N. (2017). Towards a smart city based on cloud of things: A survey on the smart city vision and paradigms. *Transactions on Emerging Telecommunications Technologies*, 28(1), e2931.
8. Chen, N., Chen, Y., You, Y., Ling, H., Liang, P., & Zimmermann, R. (2016). Dynamic urban surveillance video stream processing using fog computing. In *2016 IEEE Second International Conference on Multimedia Big Data (BigMM)* (pp. 105–112). IEEE.
9. Dan, L. I., Xin, C., Chongwei, H., & Liangliang, J. I. (2015). Intelligent agriculture greenhouse environment monitoring system based on IoT technology. In *2015 International Conference on Intelligent Transportation, Big Data and Smart City* (pp. 487–490). IEEE.
10. Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. *IEEE Internet of Things Journal*, 3(5), 637–646.
11. Almajali, S., Salameh, H. B., Ayyash, M., & Elgala, H. (2018). A framework for efficient and secured mobility of IoT devices in mobile edge computing. In *2018 Third International Conference on Fog and Mobile Edge Computing (FMEC)* (pp. 58–62). IEEE.
12. Roman, R., Lopez, J., & Mambo, M. (2018). Mobile edge computing, fog et al.: A survey and analysis of security threats and challenges. *Future Generation Computer Systems*, 78, 680–698.
13. Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., & Ayyash, M. (2015). Internet of Things: A survey on enabling technologies, protocols, and applications. *IEEE Communications Surveys & Tutorials*, 17(4), 2347–2376.

14. Zhou, M., Wang, Y., Tian, Z., Lian, Y., Wang, Y., & Wang, B. (2018). Calibrated data simplification for energy-efficient location sensing in the Internet of Things. *IEEE Internet of Things Journal*, 6(4), 6125–6133.
15. Liu, Y., Xu, C., Zhan, Y., Liu, Z., Guan, J., & Zhang, H. (2017). Incentive mechanism for computation offloading using edge computing: A Stackelberg game approach. *Computer Networks*, 129, 399–409.
16. Lv, Z., Chen, D., Lou, R., & Wang, Q. (2021). Intelligent edge computing based on machine learning for smart city. *Future Generation Computer Systems*, 115, 90–99.
17. Catarinucci, L., De Donno, D., Mainetti, L., Palano, L., Patrono, L., Stefanizzi, M. L., & Tarricone, L. (2015). An IoT-aware architecture for smart healthcare systems. *IEEE Internet of Things Journal*, 2(6), 515–526.
18. Lee, I., & Lee, K. (2015). The Internet of Things (IoT): Applications, investments, and challenges for enterprises. *Business Horizons*, 58(4), 431–440.
19. Gumaei, A., Al-Rakhami, M., Hassan, M. M., Pace, P., Alai, G., Lin, K., & Fortino, G. (2021). Deep learning and blockchain with edge computing for 5G-enabled drone identification and flight mode detection. *IEEE Network*, 35(1), 94–100.
20. Olaniyan, R., Fadahunsi, O., Maheswaran, M., & Zhani, M. F. (2018). Opportunistic edge computing: Concepts, opportunities and research challenges. *Future Generation Computer Systems*, 89, 633–645.
21. Lv, Z., Qiao, L., Chen, D., Lou, R., Li, J., & Li, Y. (2020). Machine learning for proactive defense for critical infrastructure systems. *IEEE Communications Magazine*, 6(1).
22. Chang, W. J., Chen, L. B., Sie, C. Y., & Yang, C. H. (2020). An artificial intelligence edge computing-based assistive system for visually impaired pedestrian safety at zebra crossings. *IEEE Transactions on Consumer Electronics*, 67(1), 3–11.
23. Li, R., Ma, Q., Gong, J., Zhou, Z., & Chen, X. (2021). Age of processing: Age-driven status sampling and processing offloading for edge-computing-enabled real-time IoT applications. *IEEE Internet of Things Journal*, 8(19), 14471–14484.
24. Lv, Z., Han, Y., Singh, A. K., Manogaran, G., & Lv, H. (2020). Trustworthiness in industrial IoT systems based on artificial intelligence. *IEEE Transactions on Industrial Informatics*, 17(2), 1496–1504.
25. Liu, G., Chen, X., Zhou, R., Xu, S., Chen, Y. C., & Chen, G. (2021). Social learning discrete Particle Swarm Optimization based two-stage X-routing for IC design under Intelligent Edge Computing architecture. *Applied Soft Computing*, 104, 107215.
26. Moon, S., Park, J., & Lim, Y. (2021). Task migration based on reinforcement learning in vehicular edge computing. *Wireless Communications and Mobile Computing*, 2021, 1–0.
27. Kumar, S. M., & Majumder, D. (2018). Healthcare solution based on machine learning applications in IoT and edge computing. *International Journal of Pure and Applied Mathematics*, 119(16), 1473–1484.
28. Lin, J., Yu, W., Zhang, N., Yang, X., Zhang, H., & Zhao, W. (2017). A survey on Internet of Things: Architecture, enabling technologies, security and privacy, and applications. *IEEE Internet of Things Journal*, 4(5), 1125–1142.
29. Maier, M. V. (n.d.). The Internet of Things (IoT): What is the potential of Internet of Things applications for consumer marketing? (Bachelor's thesis, University of Twente).

30. Pflanzner, T., & Kertész, A. (2018). A taxonomy and survey of IoT cloud applications. *EAI Endorsed Transactions on Internet of Things*, 3(12), 1–14.
31. Asghari, P., Rahmani, A. M., & Javadi, H. H. (2019). Internet of Things applications: A systematic review. *Computer Networks*, 148, 241–261.
32. Ray, P. P. (2018). A survey on Internet of Things architectures. *Journal of King Saud University—Computer and Information Sciences*, 30(3), 291–319.
33. Razzaque, M. A., Milojevic-Jevric, M., Palade, A., & Clarke, S. (2015). Middleware for Internet of Things: A survey. *IEEE Internet of Things Journal*, 3(1), 70–95.
34. Lao, L., Li, Z., Hou, S., Xiao, B., Guo, S., & Yang, Y. (2020). A survey of IoT applications in blockchain systems: Architecture, consensus, and traffic modeling. *ACM Computing Surveys (CSUR)*, 53(1), 1–32.
35. Farahzadi, A., Shams, P., Rezazadeh, J., & Farahbakhsh, R. (2018). Middleware technologies for the cloud of things: A survey. *Digital Communications and Networks*, 4(3), 176–188.
36. Pallavi, S., & Smruti, R. S. (2017). Internet of Things: Architectures, protocols, and applications. *Journal of Electrical and Computer Engineering*.
37. Abbas, N., Zhang, Y., Taherkordi, A., & Skeie, T. (2017). Mobile edge computing: A survey. *IEEE Internet of Things Journal*, 5(1), 450–465.
38. Dastjerdi, A. V., Gupta, H., Calheiros, R. N., Ghosh, S. K., & Buyya, R. (2016). Fog computing: Principles, architectures, and applications. In *Internet of Things* (pp. 61–75). Morgan Kaufmann.
39. Xu, X., Fu, S., Qi, L., Zhang, X., Liu, Q., He, Q., & Li, S. (2018). An IoT-oriented data placement method with privacy preservation in a cloud environment. *Journal of Network and Computer Applications*, 124, 148–157.
40. Suganuma, T., Oide, T., Kitagami, S., Sugawara, K., & Shiratori, N. (2018). Multiagent-based flexible edge computing architecture for IoT. *IEEE Network*, 32(1), 16–23.
41. Wang, C., Liang, C., Yu, F. R., Chen, Q., & Tang, L. (2017). Computation offloading and resource allocation in wireless cellular networks with mobile edge computing. *IEEE Transactions on Wireless Communications*, 16(8), 4924–4938.
42. Lu, Y., Cuiying, S., Hongwei, C., & Hongwei, Y. (2020). Foreign body detection method for power transmission equipment based on edge computing and deep learning. *China Electric Power*, 53(6), 27–33.
43. Wu, W., An, S., Wu, C. H., Tsai, S. B., & Yang, K. (2020). An empirical study on green environmental system certification affects financing cost of high energy consumption enterprises taking metallurgical enterprises as an example. *Journal of Cleaner Production*, 244, 118848.
44. Zheng, L. (2020). Application of smart new energy: Solar electric bicycle power pile system based on edge computing. *Information Construction*, 263(8), 58–61.
45. Yueru, S., & Jun, L. (2020). Mobile edge computing offload strategy for multi-base station and multi-user scenarios. *Frontiers of Data and Computing Development*, 5(3), 130–140.
46. Shen, Y., & Chang, L. (2020). Visual analysis of enterprise electricity consumption behavior based on edge computing gateway. *Electrical Appliances and Energy Efficiency Management Technology*, 592(7), 89–94.

47. Haochen, H., Yutong, L., Wang, T., Zhaoming, Q., & Junwei, C. (2020). A control strategy for energy Internet edge computing system based on mixed stochastic H_2/H_1 method. *Proceedings of the Chinese Society of Electrical Engineering*, 656(21), 115–125.
48. Jianmin, Z., Yang, F., Wu, Z., Zhengkun, Z., & Yuwei, W. (2019). Multi-access edge computing (MEC) and key technologies. *Telecommunications Science*, 35(3), 160–169.
49. Ma, J., Jia, X., Shou, Y., & Yang, C. (2018). Industrial data collection based on edge computing. *Microcomputer Applications*, 37(4), 91–93.
50. Zhu, F., Lu, P., Jilong, L., Ke, L., & Bingtao, H. (2020). The surreal experience of home intelligence: Distributed real-time rendering at the edge. *Artificial Intelligence*, 18(5), 41–48.
51. Fajjari, I., Tobagi, F., & Takahashi, Y. (2018). Cloud edge computing in the IoT. *Annals of Telecommunications*, 73, 413–414.
52. Namasudra, S., & Roy, P. (2018). PpBAC: Popularity based access control model for cloud computing. *Journal of Organizational and End User Computing (JOEUC)*, 30(4), 14–31.
53. Li, X., & Wan, J. (2018). Proactive caching for edge computing-enabled industrial mobile wireless networks. *Future Generation Computer Systems*, 89, 89–97.
54. Lv, Z., & Qiao, L. (2020). Optimization of collaborative resource allocation for mobile edge computing. *Computer Communications*, 161, 19–27.
55. Gu, B., Chen, Y., Liao, H., Zhou, Z., & Zhang, D. (2018). A distributed and context-aware task assignment mechanism for collaborative mobile edge computing. *Sensors*, 18(8), 2423.
56. Sun, H., Zhou, F., & Hu, R. Q. (2019). Joint offloading and computation energy efficiency maximization in a mobile edge computing system. *IEEE Transactions on Vehicular Technology*, 68(3), 3052–3056.
57. Lv, Z., & Xiu, W. (2019). Interaction of edge-cloud computing based on SDN and NFV for next-generation IoT. *IEEE Internet of Things Journal*, 7(7), 5706–5712.
58. Ma, X., Liang, J., Liu, R., Ni, W., Li, Y., Li, R., Ma, W., & Qi, C. (2018). A survey on data storage and information discovery in the WSANs-based edge computing systems. *Sensors*, 18(2), 546.
59. Shrestha, B., & Lin, H. (2020). Data-centric edge computing to defend power grids against IoT-based attacks. *Computer*, 53(5), 35–43.
60. Wang, H., Zeng, M., Xiong, Z., & Yang, F. (2017). Finding main causes of elevator accidents via multi-dimensional association rule in an edge computing environment. *China Communications*, 14(11), 39–47.
61. Panagopoulos, A. D. (2015). *Handbook of Research on Next-Generation Mobile Communication Systems*. IGI Global.
62. Laghari, A. A., Wu, K., Laghari, R. A., Ali, M., & Khan, A. A. (2021). A review and state of the art of the Internet of Things (IoT). *Archives of Computational Methods in Engineering*, 1–9.
63. Masip-Bruin, X., Marín-Tordera, E., Tashakor, G., Jukan, A., & Ren, G. J. (2016). Foggy clouds and cloudy fogs: A real need for coordinated management of fog-to-cloud computing systems. *IEEE Wireless Communications*, 23(5), 120–128.

64. Malik, S. U., Akram, H., Gill, S. S., Pervaiz, H., & Malik, H. (2021). EFFORT: Energy efficient framework for offload communication in mobile cloud computing. *Software: Practice and Experience*, 51(9), 1896–1909.
65. El-Sayed, H., Sankar, S., Prasad, M., Puthal, D., Gupta, A., Mohanty, M., & Lin, C. T. (2017). Edge of things: The big picture on the integration of edge, IoT, and the cloud in a distributed computing environment. *IEEE Access*, 6, 1706–1717.
66. Nguyen, V., Khanh, T. T., Van Nam, P., Thu, N. T., Hong, C. S., & Huh, E. N. (2020). Towards flying mobile edge computing. In *2020 International Conference on Information Networking (ICOIN)* (pp. 723–725). IEEE.
67. Sharma, S. K., & Wang, X. (2017). Live data analytics with collaborative edge and cloud processing in wireless IoT networks. *IEEE Access*, 5, 4621–4635.
68. Zeydan, E., Bastug, E., Bennis, M., Kader, M. A., Karatepe, I. A., Er, A. S., & Debbah, M. (2016). Big data caching for networking: Moving from cloud to edge. *IEEE Communications Magazine*, 54(9), 36–42.
69. Sahni, Y., Cao, J., Zhang, S., & Yang, L. (2017). Edge mesh: A new paradigm to enable distributed intelligence in the Internet of Things. *IEEE Access*, 5, 16441–16458.
70. Mao, Y., Zhang, J., Song, S. H., & Letaief, K. B. (2017). Stochastic joint radio and computational resource management for multi-user mobile-edge computing systems. *IEEE Transactions on Wireless Communications*, 16(9), 5994–6009.
71. Mahdavinnejad, M. S., Rezvan, M., Barekatain, M., Adibi, P., Barnaghi, P., & Sheth, A. P. (2018). Machine learning for Internet of Things data analysis: A survey. *Digital Communications and Networks*, 4(3), 161–175.
72. Jindal, M., Gupta, J., & Bhushan, B. (2019). Machine learning methods for IoT and their future applications. In *2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)* (pp. 430–434). IEEE.
73. Shao, X., Yang, C., Chen, D., Zhao, N., & Yu, F. R. (2018). Dynamic IoT device clustering and energy management with hybrid NOMA systems. *IEEE Transactions on Industrial Informatics*, 14(10), 4622–4630.
74. Sun, X., & Ansari, N. (2016). EdgeIoT: Mobile edge computing for the Internet of Things. *IEEE Communications Magazine*, 54(12), 22–29.
75. Mazza, D., Tarchi, D., & Corazza, G. E. (2017). A unified urban mobile cloud computing offloading mechanism for smart cities. *IEEE Communications Magazine*, 55(3), 30–37.
76. Van Ling, F., Yannuzzi, M., Jain, A., Irons-Mclean, R., Lluch, O., Carrera, D., Pérez, J. L., Gutiérrez, A., Montero, D., Martí, J., & Masó, R. (2017). The unavoidable convergence of NFV, 5G, and fog: A model-driven approach to bridge cloud and edge. *IEEE Communications Magazine*, 55(8), 28–35.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

CHAPTER 8

Understanding the Growth of Edge Computing with Artificial Intelligence

CHANDRSHEKHAR GOSWAMI,¹ K. SURESH KUMAR,² R. KAVIYARAJ,³
PRATEEK SRIVASTAVA,⁴ and R. UDHAYAKUMAR⁵

¹Department of CSE, School of Computing, MIT ADT University, Pune, Maharashtra, India

²MBA Department, Panimalar Engineering College, Varadarajapuram, Poonamallee, Chennai, Tamil Nadu, India

³Computational Intelligence, RM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India

⁴Department of Information Technology, School of Engineering and Technology (UIET), CSJM University, Kanpur, Uttar Pradesh, India

⁵VelTech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, Tamil Nadu, India

ABSTRACT

The conventional cloud computing paradigm may run into several issues when attempting to handle the enormous amounts of data produced on its own and fulfill associated practical requirements. In response, the academic and industrial communities have paid significant focus to a novel computational paradigm known as edge computing. Nevertheless, scientists have discovered that conventional approaches have limitations in terms of improving the functioning of edge computing as the study of edge computing continues to go deeper. Edge computing scientists began

Advances in AI for Cloud, Edge, and Mobile Computing Applications.

Shrikaant Kulkarni and P. William (Eds.)

© 2025 Apple Academic Press, Inc. Co-published with CRC Press (Taylor & Francis)

to focus on artificial intelligence (AI) after witnessing its widespread implementation in an array of domains, particularly from the standpoint of machine learning (ML), a subset of AI that has grown in popularity over the past few decades. The current study proposes a platform for understanding the growth of edge computing using ML algorithms for software sensor development. This paper describes data system architecture and associated methods depending on statistics from a quality assurance department. Multiple challenges arose throughout development as a consequence of the constant growth of ML algorithms and the optimization of their variables. As an outcome, the most precise, data-driven, and current algorithms are constantly accessible to quality assurance experts with this approach.

8.1 INTRODUCTION

The cloud-based computing system has been extensively employed since the beginning and has significantly altered people's lifestyles [1]. Many prominent corporations, like Amazon, Google, and Microsoft, have created their cloud-based computing amenities (Google Cloud) [2]. With an extensive number of distantly situated systems, cloud computational can smartly offer consumers computation, preservation, and networking resources in real-time based on their requirements in terms of asset kind, size, and so forth [3]. In this instance, consumers can easily access this cloud assistance for a modest charge or for free [4].

The growth of the internet of things (IoT) has fueled the creation and implementation of an enormous variety of physical devices/sensors around the globe [5, 6]. These physical devices/sensors are capable of sensing the tangible atmosphere around them and converting those details into data. After these enormous quantities of data are transported to the cloud for computing or preservation, data users can browse cloud data based on their specific requirements and extract the statistics they require [7].

Nevertheless, with the continued growth and broad deployment of IoT, cloud computation has started to reveal an increasing number of issues [8]. For example, if data produced by worldwide terminal equipment is calculated and preserved in a centralized cloud, it will result in a slew of issues like minimal throughput, substantial delay, bandwidth delays, confidentiality of information, centralized weaknesses, and extra expenses (like transfer expenses, energy, preservation, and computation) [9, 10].

Numerous IoT usage instances, particularly the internet of vehicles (IoV), require substantial throughput and low delay needs for data processing, analysis, and outcome return [11].

To overcome the aforementioned difficulties of cloud computation, a novel computation model known as edge computing has gained extensive interest. To put it plainly, the fundamental tenet of edge computing architecture is to transfer the data processing, preservation, and computational tasks that the cloud initially demanded to the edge of the system close to terminal equipment [12, 13]. This aids in reducing data transfer periods and gadget response periods, freeing up system bandwidth, lowering data transfer costs, and achieving decentralization [14].

Artificial intelligence (AI) is a type of innovation that gives machines an appropriate degree of intellect so that they may carry out jobs just like people [15]. This research primarily concentrates on machine learning (ML), a newly prominent topic in AI, even though heuristic-oriented methods and data mining (DM) [16] have each performed an essential part in AI approaches to IoT in recent years. It is important to note that while DM and ML both employ large amounts of data; ML concentrates on simulating human learning whereas DM is intended to derive the laws from the data [17, 18]. ML is a top intellect than DM and symbolizes the route that AI will take in the years to come.

In the “big data age” ushered in by IoT, it is evident that ubiquitous AI usage, particularly ML, is now a must-follow pattern [19]. It’s important to note that this page concentrates on the newest generation of AI algorithms, such as deep learning (DL) and other techniques [20]. The system reliability and delay need of several of these usages are substantial, yet cloud computing frequently cannot meet these demands. Conversely, the novel edge computing architecture can fulfill these needs by implementing AI at the edge and allocating certain computation and preservation assets to edge equipment near the interface [21, 22]. Despite the advantages of edge computing, like decreased delay, increased safety, and greater confidentiality of information, the constrained computation and preservation capabilities of edge computing equipment have created novel issues [23]. A recent development in related studies involves the use of AI to improve edge computing and address issues that it encounters [24].

Software sensors calculate essential characteristics for complicated chemical procedures that are hard to quantify. The growth and implementation of software sensors in chemical manufacturing have been widespread over the

past 10 years [25, 26]. Nevertheless, no adequate option for their financial functioning and lifetime surveillance has been found, therefore the number of sensors is now limited [27]. The creation of an approach for system expense-, power-, and resource-effective functioning permits constant software updates for sensors [28, 29]. Multiple sensors are employed in chemical procedures for tracking crucial procedure parameters like the quality of the goods and procedure security. Samples undergoing examination are physically removed from the procedure and analyzed in labs. Sampling rates are frequently inadequate for operational control and surveillance [30, 31]. Systems developed on datasets with tiny and insufficient standard variances may produce disappointing results. As a result, the study and evaluation of fundamental data are required at the start of modeling [32].

The objective of the research is to understand the growth of edge computing with ML algorithms for the development of software sensors. The objective is to deliver an approach that fits the aforementioned standards while also supporting the qualification procedures indefinitely [33]. To that aim, this research established a quality-assurance framework that outlines the components needed to create such a system. Furthermore, this research has established a technique that enables the implementation of ML, and this research demonstrates its applicability. The technique provides an approach for a variety of labs [34]. In addition to the previously mentioned considerations, the growth intends to lessen the ecological influence of lab operations and to apply software sensors in a variety of manufacturing procedures. ML techniques have been created for computing essential material properties based on rapid, sustainable, and low-cost spectroscopy data. The ML techniques may acquire crucial sections of spectrum data that may predict both numerical and qualitative aspects [35, 36].

Procedure systems engineering (PSE) has been around for over 50 years in the chemical engineering field, with an emphasis on computing authority and the advancement of chemical procedures employing it to support more effective plant layout, functioning, and quality control for more profitable, ecologically conscious, and effective manufacturing [37]. A more affordable, ecologically friendly, and effective functioning of different procedures is made possible by the primary fields of IoT, cloud, edge computation, and ML. To accurately monitor the quality of numerous industrial procedures, ML techniques are currently being implemented [38, 39]. The effectiveness of conventional data designs, ML techniques, and other ML-related issues still need to be resolved [40],

and also the numerous ML approaches. Because of the complicated nature of chemical procedures, incorporating ML algorithms into constant or batch manufacturing procedures is difficult [41]. As a result, it is necessary to improve the integrated capabilities of commercial management platforms and ML procedures. The existing IoT and edge-computation instruments constantly facilitate functioning operations with ML designs, which is the foundation for attaining the major goals outlined above [42, 43]. Past data and practical knowledge must be used to upgrade the designs. Furthermore, ML designs, like machinery, require upkeep since the designs might crash or break over the period. Edge computing, which is conducted by IoT devices talking with the distant cloud, is critical in commercial digitization [44, 45]. Edge computing designs may be an excellent alternative for smart industries and smart towns to reduce latency. A gateway is used for interaction between the IoT and the edge.

8.2 LITERATURE REVIEW

Edge computing and AI are prominent study areas, and a few relevant studies have been released. The writers of Citation [46] concentrate on the drive and study effort of implementing AI algorithms at system edges. The most recent advances in ML in cellular edge computing are discussed in Article [47], which involves the creation of 5G networking in autonomous adaptable resource allocation, movement modeling, safety, and energy effectiveness. The research paper [48] evaluates the usage of DL in edge computing and concentrates on ways to employ DL to boost the growth of edge usage, such as smart multimedia, transport, city, and industrial. Reference [49] discusses many techniques of quick DL logic execution in a blend of end equipment, edge machines, and cloud, as well as techniques for instructing DL algorithms in numerous edge equipment. The article [50] covers in detail ways to build an edge computing design with interaction, computational authority, and power usage restrictions to obtain the most significant efficiency for DL training and logic. From the standpoint of methods and structures, [51] thoroughly summarizes the most recent techniques for overcoming interaction problems generated by AI logic and development at the system's edge.

Edge computation instruments serve a vital part in the preservation, onsite accessibility, and growth of ML systems [52]. The objective of edge

computation is to bring cloud facilities and assistance nearer to the entities that generate data [53]. Cloud computation enables easy, on-demand networking accessibility to a common collection of programmable computer facilities that may be easily implemented and delivered with minimum management [54, 55]. A collection of IoT systems that link various things and enable them to be handled, accessible, and mined by the information they produce and transmit with other equipment [56]. In a wider context, it provides a connection to the internet and processing capacity to things, equipment, sensors, or items that are not servers [57, 58]. IoT gadgets are also crucial for wireless signal identification and transfer. The functioning of contemporary businesses depends heavily on the various portals and equipment at the web's edge.

Prediction maintenance and quality assurance are two manufacturing facilities that are significantly more effective when manufacturing processes are digitalized [59].

In a fog-based context, cyber production places a high priority on observing the state and operation of information-driven devices. The main operational system in this paper is defined by Amazon Machine Image, and the interaction method used is called MT-Connect. MT-Connect are an open collection of principles that depends on common internet-related techniques [60, 61]. Manufacturers can utilize MT-Connect to keep an eye on the pace, humidity, emergency closure, and performance condition of their processes and machines in real-time [62]. Additionally, since this protocol is set up as a web-based application, every appliance that links to the equipment's network can readily access it [63]. Edge computations are frequently employed as computations for fog and clouds. The goal is to do functions on-site, provide predictions, and accelerate procedures. The development of smart manufacturing depends on recent hot themes like cloud, edge, and fog computation as well as the Internet of Things [64, 65]. Osmotic computing features allow for better coordination of computation, connectivity, preservation, data transport, and administration among cloud and IoT gadgets at edge computing levels [66, 67].

8.3 METHODOLOGY

It is critical to build established procedures architecture for the growth of edge computing with ML models to track quality assurance in a business setting. However, there are nevertheless several growths where this

is not the case. The ML approach data-mining paradigm was created in response to the increased need and current quality assurance for the algorithms. Data quality, system durability, and anticipated system functioning are all quality assurance necessities for ML. The method’s core is to describe hazards that could have a detrimental impact on application performance and the effectiveness of ML algorithms [68]. The deployment of the ML approach aids in the ongoing verification of algorithms during the prediction of features that substantially influence the quality of items (Figure 8.1).

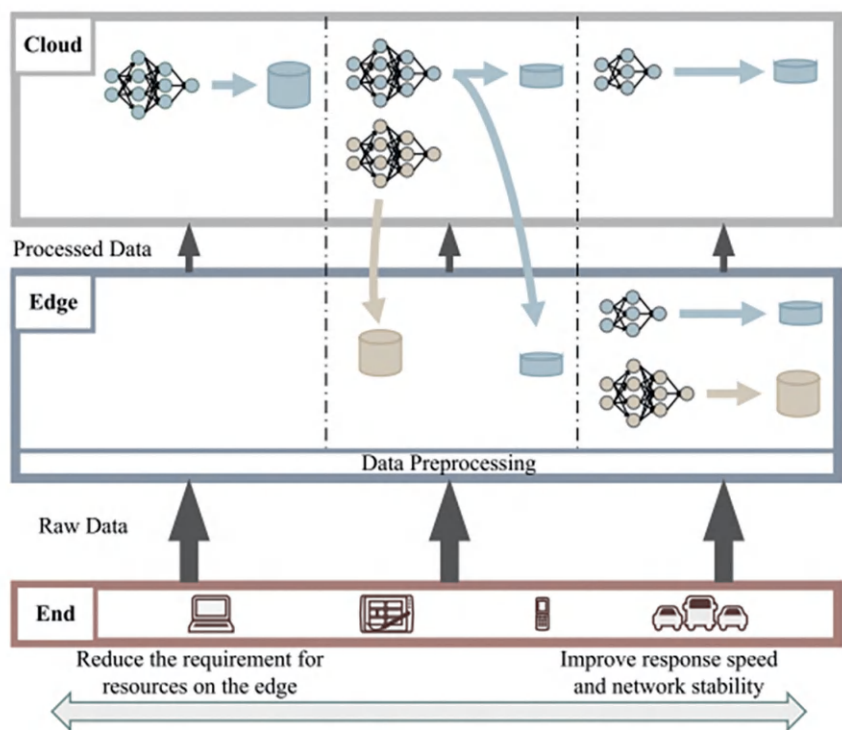


FIGURE 8.1 Hierarchy model of deploying AI in edge computing.

8.3.1 UNDERSTANDING EDGE COMPUTING WITH ML

Managing the quality of data and setting objectives for achievement are used in initiatives for the growth of edge computing with ML applications [69]. To determine if the designs generated are effective or not, the

requirements must be straightforward and quantifiable. These characteristics in this situation are the precision, dependability, and reproducibility of traditional laboratory tests. Aside from constant number monitoring, it is critical to communicate with the parties selected by the firm (such as chemical professionals, lab advancement scientists, and specialists). The ML paradigm advises assisting in the definition of constraints and application needs (robustness, adaptability) for commercial uses. The quality of the data and the numerical assessment of the data acquired are significant issues in the formation of ML algorithms.

Edge and cloud computation services are the exterior aspects of the design covered in this part. Cloud architecture solutions offer a great deal of processing capacity at an affordable price. Furthermore, virtual amenities are offered at a predefined hourly expense in these offerings, allowing us to spend the same amount for assistance as previously. The fact that numerous users may exchange credentials and capabilities at the same period is an enormous benefit. Cloud-based computation and MLOps make the growth, tracking, and future functioning of ML systems much easier [70]. This method is critical for the cloud-based storage of lab data as well as the collaborative management of data relevant to the production procedure, like humidity, stress, and quantitative evaluations. Data is transported from labs to manufacturing via different edge computing equipment and IoT. Employing the outputs of ML techniques and the state of the equipment engaged in manufacturing, the evaluation of the statistics thus acquired may offer timely assistance in the quality of the item. Additionally, in terms of data accessibility, data transfer, and modeling ought to operate smoothly. The appropriate design is depicted in the image as having two primary components (manufacturing and cloud) and 3 segments (lab, reports, and growth). The manufacturing procedure surveillance and action, lab gathering of data, and system running on-premises system describe the key aspects of setting while establishing the data pipes, algorithms growth, ML assistance, model surveillance, and reporting are all done web.

8.3.2 SECURED DATA COLLECTION AND OPERATION ON THE EDGE COMPUTING DEVICE

An important component of the growth initiative is making the produced designs accessible for manufacturing and validation even if anything goes bad between the cloud and the ground [71]. Troubleshooting also becomes

simpler if we have a problem with edge computing. Edge computing evaluation and expertise development happen near or at the origin of information and computing efficiency, out of centralized places, and towards the system's edges. Edge computation must emphasize that this approach does not depend on data facilities, but rather on ready-made systems produced in the cloud. Edge computation is a dispersed computation system that brings computation and data storage nearer together, decreasing response periods and minimizing possible range difficulties and concerns. As a consequence, it boosts the rapidity and effectiveness with which data is processed. This computer system is comparable to a cloud-based system, but it is more focused on apps. Edge computing analyses information from IoT gadgets at the system's edge and sends it to the cloud. In the method that this research devised, and also the assessment findings, lab data administration structure (LDAS) and ERP information should be accessible on the edge computing equipment. As a result, picking the best edge gadget on the market is critical. Numerous producers make a wide range of sharps, the specifications of which might vary greatly. A small personal computer (PC) is the edge gadget of this research choice. An important component of the study was that the equipment might be employed in severe field settings (temperatures ranging from 40°C–85°C), not just in the lab. The thoughtfully chosen edge computing device securely delivers the acquired data to the cloud and saves and operates the systems bundled following suitable instructions.

Using blockchain innovation to eliminate potential threats is one way. The solution enables secure data transport and ML model implementation to IoT and edge computing equipment [72]. Nevertheless, additional secured options exist in addition to or in conjunction with block-chains.

8.3.3 IMPLEMENTATION OF FRAMEWORK

The standard of the factor on which the designs may be formed is a crucial consideration in model creation. Where and the way a certain variable may be anticipated is also a crucial factor. For implementing software sensors for functional and possibly drilling smart sensors, designs for lab evaluations are thus helpful. The understanding of edge computing with ML applications objective in the initial stage of the ML approach is to comprehend that the measurement of nitrogen from functioning variables

and the assessment of quartz composition from drilling rock data have considerable possibilities for evaluating ML designs. Although the level of model inaccuracy and the pace were related to the computation, the success condition for nitrogen design calculation was defined by the repeatability value of the traditional assessment in the quartz design. In both situations, the created form satisfies the initial ML requirement. When analyzing the statistics in the 2nd stage, it was crucial to determine whether the conventional measurements in both situations can be supplanted by quick, non-destructive forms. These forms would then serve as a solid foundation for the setup of additional software sensors in the future. The stage of data understanding focuses on evaluation data and the faults that exist in our tests (reproducibility, repetition). The edge computing device, ERP, and LIMS were the appropriate data resources in each situation. To track the data dispersion from the 2 databases, this research employed specific train-test splits for both targeted factors throughout the modeling process. The dispersion of the application's training and testing samples was comparable. The designs were created using 10-fold cross-validation (10-cv) and deployed using PMML. Workers can use a visualization instrument, PC usage, internet browser, or perhaps mobile to monitor the outcomes and accuracy of installed systems.

The verification process should be used to review ever-newer readings from the edge device (Figure 8.2, Edge computing). ML upkeep determines if the sample is valuable enough to be included in the algorithm or not. Additionally, continually fresher datasets aid in monitoring the effectiveness of models that are presently in use (Reporting). The developed designs should be capable of handling these adjustments, so they are preserved and the data via usage. The Figure 8.2 illustrates the design growth phases for a material stream variable.

Exploration data assessment (EDA), pre-processing, outlier identification, train-test splits, and a specific method that takes into account the dispersion of the intended parameter are the primary components of the growth of edge computing with ML algorithms that outline the streamlined processes of data processing and modeling. The system's variables are then adjusted during the iterative phase, and low-error designs with the proper situations are then implemented. Every variable (nitrogen, quartz content and so on) in every group of lab datasets should be established independently for this procedure. Assessment outcomes from laboratories are frequently not directly applicable to interpreting or modeling. In all

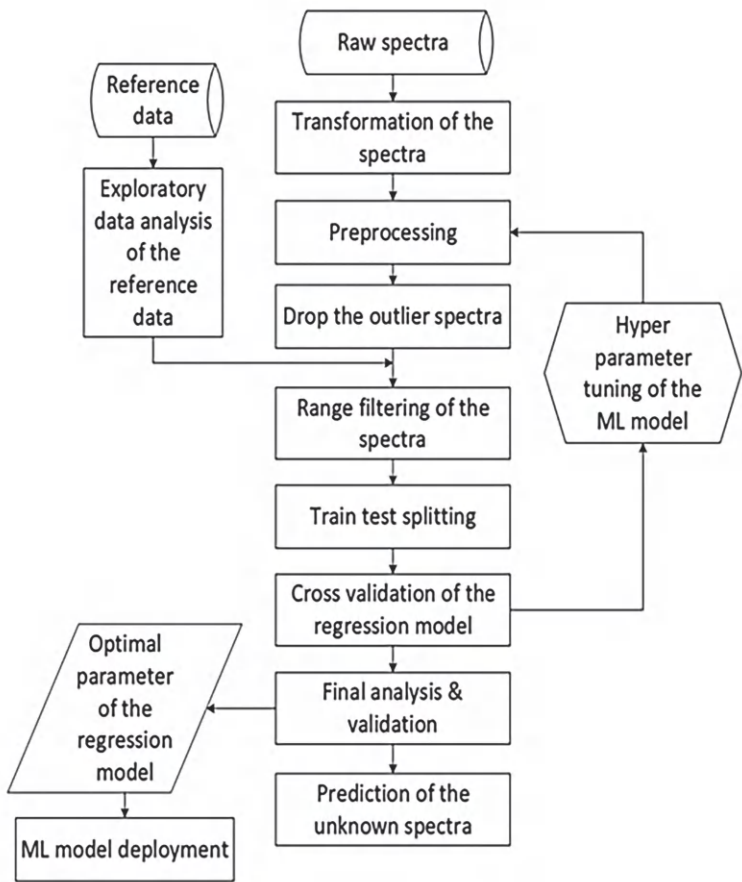


FIGURE 8.2 The primary phases in the growth of ML algorithms.

situations engaging the business peers, it should be connected to a certain calibration to understand edge computing with ML applications, or in numerous circumstances, certain quantitative methods should be utilized (Understanding edge computing with ML applications). According to the ML approach, this number is continually rising. As a result, it is important to investigate every variable’s contribution to the system’s layout.

Random forest (RF) is a tree-oriented method that blends the outcomes of many decision trees (DT) to produce the eventual result. The word “random” refers to the fact that this technique is a forest of arbitrarily produced DT [73]. The more straightforward DT approach was not used

since it has a substantial downside of causing over-matching, which may be mitigated when applying random forest regression (RFR). A further major benefit is that the RF technique is incredibly quick and durable when contrasted with other techniques.

Extreme gradient boosting (EGB) is a well-known gradient-boosted tree technique [74]. The technique's strategy seeks to forecast the intended target parameter precisely by merging values from smaller, weaker designs. The regularized objective functions are minimized by EGB, which blends a convex loss function (the disparity between anticipated and targeted outcomes) and a penalty word for system complication. Iteratively, the training is finished through the inclusion of novel trees that anticipate the remaining parts or faults of the prior trees, which are then integrated with the previous trees to form the final projection [75]. It includes bespoke optimization and also has a distinctive technique for building and pruning trees. It is a significant benefit because it allows for faster computation on massive databases.

8.4 RESULTS AND DISCUSSION

Throughout the real-time functioning of ML algorithms, it is critical to constantly assess the precision of the algorithms to evaluate when a specific paradigm is judged outstanding, favorable, or inadequate. When examining the designs, the three “R” scores of the conventional measures should be regarded [76]. In general, a framework is regarded as acceptable if the predicted accuracy of the novel datasets is within the reproduction threshold. Systems that surpass the repeatability level of traditional assessments are deemed unacceptable (Understanding edge computing with ML applications). Surveillance systems enable them to be graded depending on proportion excess. The surveillance structure reveals the frameworks with the most important shortcomings at the start of the position, indicating that assistance is required right away (Figure 8.1, The Design Deployment). Various measurements in conventional lab measuring procedures are analogous to system KPIs. We've been using circular assessments of different types to verify gadgets regularly for decades. As a result, when utilizing the ML method signs, the computations are extremely similar. The coefficient of correlation (R^2), root-mean-square deviation (RMSE), and relative percentage

deviations (RPD) are critical markers for monitoring ML algorithms. In this research, we examine 3 various methods to find the optimal one for estimating a particular variable. A special linear regression is augmented by a computation that may deal with non-linear situations. RF and EGB are the other 2 tree-oriented methods in use. The choice of a method for qualifying the data sets was a critical factor. Throughout the evaluations, the ML models should be stable and not dependent on the outlying dataset, and the methodologies of the competitive methods vary. The three methods should be optimized and evaluated for every intended factor, and the most effective of the three should be deployed on the edge instrument. The design kind could have altered throughout growth for deployed systems.

Because of its capacity to handle chaotic, co-linear, and imperfect databases, partial least squares (PLSR) are becoming more prominent as a multivariable data assessment technique. PLSR was augmented by a nonlinear iteration PLS (NIPLS) method that was supplemented by a nonlinear iteration assessment dependent on a recurrent estimation of the correlation matrix and gradient-oriented methods for calculating eigenvectors of the pertinent matrix.

Examining the sensitivity of the systems is an essential step in the construction of strong systems. Sensitivity assessments assess alterations in system sources as well as the distinct influences of every factor on results, and they offer data on the various consequences of every factor investigated. Furthermore, it is critical to generate a significant number of data sets, particularly unusual ones, to set up robust designs. For the latter step, severe datasets can be created using the layout of the study; these datasets contribute to the stability of the designs. We evaluated the accuracy of the designs for every lab characteristic for verification and testing databases during the growth of edge computing with ML. The systems were optimized so that KPIs did not vary considerably across training, verification, and evaluation databases, safeguarding the systems from excessive fitting.

The accuracy of the ML algorithms generated on the two evaluated characteristics is summarized in the subsequent two tables. It is crucial to highlight that the preprocessing of the spectrum before the three system kinds was identical for both attributes (Tables 8.1 and 8.2). The '10-cv' outcomes are expressed by the 'perf.' measure, which indicates the system's performance on data not utilized in the system's training.

TABLE 8.1 Outcomes of 10-cv and Performance of Nitrogen Content of ML Algorithms

	R2		RMSE		RPD	
	10-cv	perf	10-cv	perf	10-cv	perf
PLSR	0.999	0.975	0.01	0.03	57.7	6.3
RFR	0.972	0.929	0.089	0.084	5.9	3.7
EGB	0.999	0.747	0.005	0.11	31.6	1.9

TABLE 8.2 Outcomes of 10-cv and Performance of Quartz Content of ML Algorithms

	R2		RMSE		RPD	
	10-cv	perf	10-cv	perf	10-cv	perf
PLSR	0.90	0.73	2.03	2.40	3.16	1.93
RFR	0.62	0.93	1.43	4.67	1.62	4.01
EGB	0.91	0.87	1.96	4.66	3.40	2.77

Based on the findings in the above two tables, it is possible to derive that EGB is overfitted and operates poorly despite hyper-parameter adjustment. The PLSR system has a reasonable average performance, while the RFR is the best-tuned ML system of the three. These system findings demonstrate that this research may employ ML algorithms to examine the two main variables that comprised the research. By using the systems, this research can identify certain essential factors considerably quicker, allowing us to substantially lower the workload and reaction period of the lab. Additionally, with the help of industrial engineers, the setup of industrial software sensors may be addressed after evaluating the developed designs. In the instance of lube oils, the developer gives guidance on where to set up sensors; however, for upward wells, software sensors are permitted. The plant’s needed characteristics are intended to decrease total equipment efficiency during lubricant manufacturing and to locate the best source for upstream drilling. Garbage items are reduced with the assistance of the designs throughout the manufacturing of lube oils, and in the instance of quartz designs, we receive a more realistic image of the geographical deposits.

The designs are at present accessible by labs that are tracked via documentation and a web application. There are many limits for the dataset analyzed where the system reveals a distinction between the predicted value and the conventional assessment. The online application autonomously suggests the verification necessary for the ML system after 10 signs. Afterward, our data analyst peers evaluate the badly predicted

datasets and, if appropriate, understand and grow the edge computing model with ML.

Several additional samples can be measured swiftly using the sophisticated statistical algorithms of manufacturing and study facilities. The aforementioned design and growth of edge computing with ML algorithms may decrease lab tasks and enable tests with reduced safety and health administration hazards. Rather than difficult-to-measure, dangerous compounds, the success rate of ML algorithms implemented on edge-computing sensors grows for diverse qualifying properties and can fluctuate dramatically over a period. After deploying the system, it is critical to track and manage the edge component designs. Furthermore, the instrument is in charge of transmitting data in real-time and onsite accessibility to the most recent designs. The MOXA-8200 edge sensor was chosen for the investigation, and its setup and functioning caused significant challenges during growth. MOXA is an effective instrument for gathering data and maintaining certain designs, but adding more models leads to significant limits when utilizing the device. Since the demand for edge computing equipment is evolving rapidly, it is worthwhile to assess the devices utilized regularly. The instrument examined was obtained from a regional support firm, making it simple to verify that the most effective instrument is always available on-site. Employing several ML methods, this research described in this part may calculate hard-to-measure, troublesome variables. The designs created have the advantage of always having the proper ones accessible. Monitoring and updating designs is a difficult for studies and industrial labs, but edge computing ideas can help. They provide comprehensive transmission of data, layout, model creation, and implementation approaches. Nevertheless, the two methodologies constitute a significant potential as well as a significant challenge for the secure and uninterrupted delivery of commercial operations. As a result, it is critical to guarantee the systems' real-time accuracy and accessibility.

8.5 CONCLUSION

With the growth of the fourth industrial revolution and the potential provided by technological advancement, it is critical to bring research and development nearer to manufacturing, and sensors serve an important part in this. At the moment, software sensors are growing in popularity because they

can forecast important variables that are hard to monitor in manufacturing operations. Nevertheless, software sensors necessitate the creation of unique AI algorithms that should be constantly tracked, controlled, and upgraded. Edge computing is a highly intriguing novel computing model that has the potential to compensate for the flaws of present cloud computing, whereas AI is a tremendously prominent topic in academics and business. The technique detailed in the scholarly paper, as well as the case investigation offers a feasible approach for the use of software sensors. The implementation of ML algorithms in manufacturing includes several layered elements and procedures. ML is an organized procedure framework for ML software growth that emphasizes quality assurance to decrease possible hazards and guarantee the achievement of the ML program. The major components of applying the ML technique include model construction, ongoing cleansing of data, engineering features, model verification, performance surveillance, and dataset visualization. The continual construction of designs, sequential transmission of data, and on-site accessibility to the designs all require edge computing, which are additional crucial components of this methodology. The two laboratory tests included in the case investigation were conducted using ML algorithms that can be employed with software sensors. The architecture that is being given is also connected to a technique that incorporates edge computing components.

The future goal is to investigate further areas for study that mix the two, such as dispersed training and logic in the context of edge computing for implementing software sensors for different procedure divisions and using the architecture to further enhance production procedures.

KEYWORDS

- **AI algorithms**
- **edge computing**
- **fourth industrial revolution**
- **machine learning**
- **manufacturing optimization**
- **software sensors**

REFERENCES

1. Sharma, P., Kumari, R., & Aulakh, I. K. (2020). Task-aware energy-efficient framework for mobile cloud computing. In *ICT Systems and Sustainability: Proceedings of ICT4SD 2019, Volume 1* (pp. 387–395). Springer Singapore.
2. Moorthy, V., Venkataraman, R., & Rao, T. R. (2020). Security and privacy attacks during data communication in software-defined mobile clouds. *Computer Communications*, 153, 515–526.
3. Elgendy, I. A., & Yadav, R. (2022). Survey on mobile edge-cloud computing: A taxonomy on computation offloading approaches. In *Security and Privacy Preserving for IoT and 5G Networks: Techniques, Challenges, and New Directions* (pp. 117–158). Springer.
4. Gholivand, R., & Movahedi, Z. (2021). A Cloud-RAN based end-to-end computation offloading in Mobile Edge Computing. *Computer Communications*, 175, 193–204.
5. Shi, W., & Dustdar, S. (2016). The promise of edge computing. *Computer*, 49(5), 78–81.
6. Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. *IEEE Internet of Things Journal*, 3(5), 637–646.
7. Carvalho, G., Cabral, B., Pereira, V., & Bernardino, J. (2021). Edge computing: Current trends, research challenges, and future directions. *Computing*, 103(5), 993–1023.
8. Qin, M., Chen, L., Zhao, N., Chen, Y., Yu, F. R., & Wei, G. (2018). Power-constrained edge computing with maximum processing capacity for IoT networks. *IEEE Internet of Things Journal*, 6(3), 4330–4343.
9. Li, S., Zhai, D., Du, P., & Han, T. (2019). Energy-efficient task offloading, load balancing, and resource allocation in mobile edge computing enabled IoT networks. *Science China Information Sciences*, 62, 1–3.
10. Samie, F., Tsoutsouras, V., Bauer, L., Xydis, S., Soudris, D., & Henkel, J. (2016). Computation offloading and resource allocation for low-power IoT edge devices. In *2016 IEEE 3rd World Forum on Internet of Things (WF-IoT)* (pp. 7–12). IEEE.
11. Fu, Y., Yang, X., Yang, P., Wong, A. K., Shi, Z., Wang, H., & Quek, T. Q. (2021). Energy-efficient offloading and resource allocation for mobile edge computing enabled mission-critical internet-of-things systems. *EURASIP Journal on Wireless Communications and Networking*, 2021(1), 1–6.
12. Ghosh, A. M., & Grolinger, K. (2020). Edge-cloud computing for Internet of Things data analytics: Embedding intelligence in the edge with deep learning. *IEEE Transactions on Industrial Informatics*, 17(3), 2191–2200.
13. Wang, X., Han, Y., Leung, V. C., Niyato, D., Yan, X., & Chen, X. (2020). Convergence of edge computing and deep learning: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, 22(2), 869–904.
14. Li, H., Ota, K., & Dong, M. (2018). Learning IoT in edge: Deep learning for the Internet of Things with edge computing. *IEEE Network*, 32(1), 96–101.
15. Zhou, P., Chen, W., Ji, S., Jiang, H., Yu, L., & Wu, D. (2019). Privacy-preserving online task allocation in edge-computing-enabled massive crowdsensing. *IEEE Internet of Things Journal*, 6(5), 7773–7787.

16. Savaglio, C., Gerace, P., Di Fatta, G., & Fortino, G. (2019). Data mining at the IoT edge. In *2019 28th International Conference on Computer Communication and Networks (ICCCN)* (pp. 1–6). IEEE.
17. Xu, Z., Liu, W., Huang, J., Yang, C., Lu, J., & Tan, H. (2020). Artificial intelligence for securing IoT services in edge computing: A survey. *Security and Communication Networks*, 2020, 1–13.
18. Savaglio, C., & Fortino, G. (2021). A simulation-driven methodology for IoT data mining based on edge computing. *ACM Transactions on Internet Technology (TOIT)*, 21(2), 1–22.
19. Lei, L., Xu, H., Xiong, X., Zheng, K., Xiang, W., & Wang, X. (2019). Multiuser resource control with deep reinforcement learning in IoT edge computing. *IEEE Internet of Things Journal*, 6(6), 10119–10133.
20. Marjani, M., Nasaruddin, F., Gani, A., Karim, A., Hashem, I. A., Siddiqua, A., & Yaqoob, I. (2017). Big IoT data analytics: Architecture, opportunities, and open research challenges. *IEEE Access*, 5, 5247–5261.
21. Index, C. G. (2018). *Forecast and Methodology, 2016–2021 White Paper*. Cisco Global Cloud Index (2016–2021).
22. Ahmed, E., Yaqoob, I., Hashem, I. A., Khan, I., Ahmed, A. I., Imran, M., & Vasilakos, A. V. (2017). The role of big data analytics in the Internet of Things. *Computer Networks*, 129, 459–471.
23. Sasaki, Y. (2021). A survey on IoT big data analytic systems: Current and future. *IEEE Internet of Things Journal*, 9(2), 1024–1036.
24. Cai, H., Xu, B., Jiang, L., & Vasilakos, A. V. (2016). IoT-based big data storage systems in cloud computing: Perspectives and challenges. *IEEE Internet of Things Journal*, 4(1), 75–87.
25. Funes, E., Allouche, Y., Beltrán, G., Aguliera, M. P., & Jiménez, A. (2017). A predictive artificial neural network model as a simulator of the extra virgin olive oil elaboration process. *Journal of Near Infrared Spectroscopy*, 25(4), 278–285.
26. Souza, F. A., Araújo, R., & Mendes, J. (2016). Review of soft sensor methods for regression applications. *Chemometrics and Intelligent Laboratory Systems*, 152, 69–79.
27. Bidar, B., Sadeghi, J., Shahraki, F., & Khalilipour, M. M. (2017). Data-driven soft sensor approach for online quality prediction using state-dependent parameter models. *Chemometrics and Intelligent Laboratory Systems*, 162, 130–141.
28. Pattanayak, A. S., Pattnaik, B. S., Udgata, S. K., & Panda, A. K. (2020). Development of chemical oxygen demand (COD) soft sensor using edge intelligence. *IEEE Sensors Journal*, 20(24), 14892–14902.
29. Pattnaik, B. S., Pattanayak, A. S., Udgata, S. K., & Panda, A. K. (2021). Machine learning-based soft sensor model for BOD estimation using intelligence at the edge. *Complex & Intelligent Systems*, 7(2), 961–976.
30. Wang, Z., & Chiang, L. (2019). Monitoring chemical processes using a judicious fusion of multi-rate sensor data. *Sensors*, 19(10), 2240.
31. Asachi, M., & Camargo-Valero, M. A. (2023). Multi-sensor data fusion for monitoring of powdered and granule products: Current status and future perspectives. *Advanced Powder Technology*, 34(7), 104055.

32. Pereira Parente, A., de Souza Jr, M. B., Valdman, A., & Mattos Folly, R. O. (2019). Data augmentation applied to machine learning-based monitoring of a pulp and paper process. *Processes*, 7(12), 958.
33. Da Costa, N. L., Da Costa, M. S., & Barbosa, R. (2021). A review on the application of chemometrics and machine learning algorithms to evaluate beer authentication. *Food Analytical Methods*, 14, 136–155.
34. Koren, D., Lórinçz, L., Kovács, S., Kun-Farkas, G., & Vecseriné Hegyes, B. (2020). Comparison of supervised learning statistical methods for classifying commercial beers and identifying patterns. *Journal of Chemometrics*, 34(4), e3216.
35. Maione, C., Barbosa Jr, F., & Barbosa, R. M. (2019). Predicting the botanical and geographical origin of honey with multivariate data analysis and machine learning techniques: A review. *Computers and Electronics in Agriculture*, 157, 436–446.
36. Silva, A. J., Cortez, P., & Pilastri, A. (2020). Chemical laboratories 4.0: A two-stage machine learning system for predicting the arrival of samples. In *IFIP International Conference on Artificial Intelligence Applications and Innovations* (pp. 232–243). Cham: Springer International Publishing.
37. Reis, M. S., & Saraiva, P. M. (2021). Data-centric process systems engineering: A push towards PSE 4.0. *Computers & Chemical Engineering*, 155, 107529.
38. Reis, M. S., & Saraiva, P. M. (2022). Data-driven process system engineering—Contributions to its consolidation following the path laid down by George Stephanopoulos. *Computers & Chemical Engineering*, 159, 107675.
39. Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal*, 13, 8–17.
40. Mowbray, M., Vallerio, M., Perez-Galvan, C., Zhang, D., Chanona, A. D., & Navarro-Brull, F. J. (2022). Industrial data science: A review of machine learning applications for chemical and process industries. *Reaction Chemistry & Engineering*, 7(7), 1471–1509.
41. Studer, S., Bui, T. B., Drescher, C., Hanuschkin, A., Winkler, L., Peters, S., & Müller, K. R. (2021). Towards CRISP-ML(Q): A machine learning process model with quality assurance methodology. *Machine Learning and Knowledge Extraction*, 3(2), 392–413.
42. Lee, H. W., Bhatia, H., Park, S. Y., Kamga, M. H., Reimonn, T., Sha, S., Huang, Z., Galbraith, S., Liu, H., & Yoon, S. (2019). Process analytical technology and quality by design for animal cell culture. In *Cell Culture Engineering: Recombinant Protein Production* (pp. 365–390).
43. PremSankar, G., Di Francesco, M., & Taleb, T. (2018). Edge computing for the Internet of Things: A case study. *IEEE Internet of Things Journal*, 5(2), 1275–1284.
44. Sittón-Candanedo, I., Alonso, R. S., Corchado, J. M., Rodríguez-González, S., & Casado-Vara, R. (2019). A review of edge computing reference architectures and a new global edge proposal. *Future Generation Computer Systems*, 99, 278–294.
45. Sittón-Candanedo, I., Alonso, R. S., Rodríguez-González, S., García Coria, J. A., & De La Prieta, F. (2020). Edge computing architectures in industry 4.0: A general survey and comparison. In *14th International Conference on Soft Computing Models*

- in *Industrial and Environmental Applications (SOCO 2019)* (pp. 121–131). Springer International Publishing.
46. Zhou, Z., Chen, X., Li, E., Zeng, L., Luo, K., & Zhang, J. (2019). Edge intelligence: Paving the last mile of artificial intelligence with edge computing. *Proceedings of the IEEE*, 107(8), 1738–1762.
 47. McClellan, M., Cervelló-Pastor, C., & Sallent, S. (2020). Deep learning at the mobile edge: Opportunities for 5G networks. *Applied Sciences*, 10(14), 4735.
 48. Wang, F., Zhang, M., Wang, X., Ma, X., & Liu, J. (2020). Deep learning for edge computing applications: A state-of-the-art survey. *IEEE Access*, 8, 58322–58336.
 49. Wang, X., Han, Y., Leung, V. C., Niyato, D., Yan, X., & Chen, X. (2020). Convergence of edge computing and deep learning: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, 22(2), 869–904.
 50. Chen, J., & Ran, X. (2019). Deep learning with edge computing: A review. *Proceedings of the IEEE*, 107(8), 1655–1674.
 51. Shi, Y., Yang, K., Jiang, T., Zhang, J., & Letaief, K. B. (2020). Communication-efficient edge AI: Algorithms and systems. *IEEE Communications Surveys & Tutorials*, 22(4), 2167–2191.
 52. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., & Chou, R. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery*, 88, 105906.
 53. Escamilla-Ambrosio, P. J., Rodríguez-Mota, A., Aguirre-Anaya, E., Acosta-Bermejo, R., & Salinas-Rosales, M. (2018). Distributing computing in the Internet of Things: Cloud, fog, and edge computing overview. In *NEO 2016: Results of the Numerical and Evolutionary Optimization Workshop* (pp. 87–115). Springer International Publishing.
 54. Alippi, C., Fantacci, R., Marabissi, D., & Roveri, M. (2016). A cloud to the ground: The new frontier of intelligent and autonomous networks of things. *IEEE Communications Magazine*, 54(12), 14–20.
 55. Patel, P., Ali, M. I., & Sheth, A. (2017). On using the intelligent edge for IoT analytics. *IEEE Intelligent Systems*, 32(5), 64–69.
 56. Barik, R. K., Dubey, H., Misra, C., Borthakur, D., Constant, N., Sasane, S. A., Lenka, R. K., Mishra, B. S., Das, H., & Mankodiya, K. (2018). Fog assisted cloud computing in the era of big data and Internet of Things: Systems, architectures, and applications. In *Cloud Computing for Optimization: Foundations, Applications, and Challenges* (pp. 367–394).
 57. Boyes, H., Hallaq, B., Cunningham, J., & Watson, T. (2018). The industrial Internet of Things (IIoT): An analysis framework. *Computers in Industry*, 101, 1, 2.
 58. Serpanos, D., & Wolf, M. (2018). Industrial Internet of Things. In *Internet-of-Things (IoT) Systems: Architectures, Algorithms, Methodologies* (pp. 37–54).
 59. Yang, H., Kumara, S., Bukkapatnam, S. T., & Tsung, F. (2019). The Internet of Things for smart manufacturing: A review. *IJSE Transactions*, 51(11), 1190–1216.
 60. Chen, L., Zhou, S., & Xu, J. (2018). Computation peer offloading for energy-constrained mobile edge computing in small-cell networks. *IEEE/ACM Transactions on Networking*, 26(4), 1619–1632.

61. Ning, Z., Dong, P., Kong, X., & Xia, F. (2018). A cooperative partial computation offloading scheme for mobile edge computing enabled Internet of Things. *IEEE Internet of Things Journal*, 6(3), 4804–4814.
62. Du, J., Zhao, L., Feng, J., & Chu, X. (2017). Computation offloading and resource allocation in mixed fog/cloud computing systems with min-max fairness guarantee. *IEEE Transactions on Communications*, 66(4), 1594–1608.
63. Wang, Y., Tao, X., Zhang, X., Zhang, P., & Hou, Y. T. (2019). Cooperative task offloading in three-tier mobile computing networks: An ADMM framework. *IEEE Transactions on Vehicular Technology*, 68(3), 2763–2776.
64. Qiao, Y., Liu, Z., Lv, H., Li, M., Huang, Z., Li, Z., & Liu, W. (2019). An effective data privacy protection algorithm based on differential privacy in edge computing. *IEEE Access*, 7, 136203–136213.
65. Wu, D., Liu, S., Zhang, L., Terpenney, J., Gao, R. X., Kurfess, T., & Guzzo, J. A. (2017). A fog computing-based framework for process monitoring and prognosis in cyber-manufacturing. *Journal of Manufacturing Systems*, 43, 25–34.
66. Villari, M., Celesti, A., & Fazio, M. (2018). Towards osmotic computing: Looking at basic principles and technologies. In *Complex, Intelligent, and Software Intensive Systems: Proceedings of the 11th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS-2017)* (pp. 906–915). Springer International Publishing.
67. Neha, B., Panda, S. K., Sahu, P. K., Sahoo, K. S., & Gandomi, A. H. (2022). A systematic review on osmotic computing. *ACM Transactions on Internet of Things*, 3(2), 1–30.
68. Mäkinen, S., Skogström, H., Laaksonen, E., & Mikkonen, T. (2021). Who needs MLOps: What data scientists seek to accomplish and how can MLOps help? In *2021 IEEE/ACM 1st Workshop on AI Engineering-Software Engineering for AI (WAIN)* (pp. 109–112). IEEE.
69. Hussain, B., Du, Q., Zhang, S., Imran, A., & Imran, M. A. (2019). Mobile edge computing-based data-driven deep learning framework for anomaly detection. *IEEE Access*, 7, 137656–137667.
70. Dong, R., She, C., Hardjawana, W., Li, Y., & Vucetic, B. (2019). Deep learning for hybrid 5G services in mobile edge computing systems: Learn from a digital twin. *IEEE Transactions on Wireless Communications*, 18(10), 4692–4707.
71. Sivaganesan, D. D. (2019). Design and development of AI-enabled edge computing for intelligent IoT applications. *Journal of Trends in Computer Science and Smart Technology*, 1(2), 84–94.
72. Diro, A., Chilamkurti, N., Nguyen, V. D., & Heyne, W. (2021). A comprehensive study of anomaly detection schemes in IoT networks using machine learning algorithms. *Sensors*, 21(24), 8320.
73. Chencho, Li, J., Hao, H., Wang, R., & Li, L. (2021). Development and application of random forest technique for element level structural damage quantification. *Structural Control and Health Monitoring*, 28(3), e2678.
74. Sheridan, R. P., Wang, W. M., Liaw, A., Ma, J., & Gifford, E. M. (2016). Extreme gradient boosting as a method for quantitative structure–activity relationships. *Journal of Chemical Information and Modeling*, 56(12), 2353–2360.

75. Ching, P. M., Zou, X., Wu, D., So, R. H., & Chen, G. H. (2022). Development of a wide-range soft sensor for predicting wastewater BOD5 using an extreme gradient boosting (XGBoost) machine. *Environmental Research*, 210, 112953.
76. Plesser, H. E. (2018). Reproducibility vs. replicability: A brief history of a confused terminology. *Frontiers in Neuroinformatics*, 11, 76.

CHAPTER 9

Perspectives of Machine Learning in the Convergence of Artificial Intelligence and Edge Computing

ABU SARWAR ZAMANI,¹ R. M. JAGADISH,² BINOD KUMAR,³
MD. ATHEEQ SULTAN GHORI,⁴ and SUHAS A. BHYRATAE⁵

¹*Computer Science, Prince Sattam Bin Abdulaziz Alkharj, Riyadh, Saudi Arabia*

²*Department of CSE (DS), Ballari Institute of Technology and Management, Ballari, Karnataka, India*

³*MCA, JSPM Rajarshi Shahu College of Engineering, Pune, Maharashtra, India*

⁴*Computer Science and Engineering Department, Telangana University, Nizamabad, Telangana, India*

⁵*Computer Science and Engineering, Sahyadri College of Engineering and Management, Mangalore, Karnataka, India*

ABSTRACT

The widespread application of the internet of things (IoT) and the advancement of machine learning (ML) methods have sparked demand to converge artificial intelligence (AI) methods to the emerging field of edge computing. The ideal allocation of created tasks across the equipment in every tier (i.e., edge-cloud-fog) is among the difficulties faced by edge computation infrastructures. This research suggests employing

Reinforcement learning (RL) to resolve the task assignment issue (TAI) at the edge of the system and then this research suggests a new multi-layer expansion of RL methods that enables edge drivers to seek out a top-level driver with additional information to enhance functioning in complicated and unsure circumstances. The task assignment procedure is first developed while taking into account the trade-off between power usage and implementation duration. This research next provides a greedy approach as a foundation and uses the Pure-Edge-Sim simulation to build the multi-layer RL approach. Lastly, many simulations of every method with varying numbers of gadgets are reviewed to ensure adaptability. The simulation findings reveal that RL methods beat heuristic-oriented options, and the multi-layer strategy can increase functioning greatly in extreme gadget density conditions.

9.1 INTRODUCTION

Novel approaches for digital media and edge computing, specifically in urbanized and smart city settings [1–7], have emerged as a result of the improvements and developments in the convergence of AI and ML [8, 9], as well as the wide availability of portable equipment and online innovations. These factors also include an increasing emphasis on multimedia databases [10] and computing methods [11]. These frameworks are a result of several patterns coming together, such as (i) the widespread use of IoT gadgets and sensors in smart towns and urban regions [12]; (ii) the development of independent structures [13]; and (iii) the rising demand for multimedia equipment and computing process from incorporated recording devices and portable imaging equipment (such as drones) [14, 15]. The edge computing framework distributes the work to numerous equipment that are near the data supply as opposed to cloud processing techniques, which gather data and send it to a centralized system for data processing [16].

The cloud-based computing field has long been concerned about which highly demanded tasks edge computation might advance that cloud computation could not, and what the edge computation's "killer services" are. This has been true even throughout the initial phases of the technology's creation [17, 18]. Microsoft has been exploring what applications must be transferred from the cloud to the edge continuously since 2009

[19] to dispel any confusion. These applications range from real-time video analysis to voice order acknowledgment augmented and virtual reality and interactive cloud-based gaming [20].

In contrast, real-time video analysis is thought to be a game-changing use case for edge computation [21–23]. Real-time video analysis is a new application based on computer vision that constantly gathers high-definition films from security cameras and analyses them using high processing, significant bandwidth, substantial confidentiality, and minimal delay. Edge computation is the only practical strategy that can satisfy these stringent demands [24].

However, as described in [25] and [26], there are obstacles to all inventive progress. This is also true for IoT and portable innovations. Among them is the delay issue. Since some mobile uses, like virtual reality [27] and self-driving cars [28], must adhere to a specified delay standard, sending data to the cloud [29] is occasionally not an option. To solve this issue, the edge computation framework offers a dispersed data processing infrastructure close to the consumers, allowing delay-critical uses and lowering bandwidth consumption and cloud-based system congestion [30].

Nevertheless, a computing offloading paradigm that optimizes the finest feasible task allocation to computation equipment must be used to make effective use of the equipment in the edge-based computation design [31–33]. This research will focus on this issue, in which every edge equipment is given tasks with unique needs and must choose whether to carry out the computing itself or to send it to a cloud or another edge unit.

The technique to identify the most efficient allocation of tasks to equipment is known as the “task assignment issue (TAI),” and it is a combined optimization issue known as the operation of establishing where the calculation of every task occurs to minimize specific variables, like the above-mentioned delay and usage of energy, a major issue for a sustainable future [34] and an essential aspect of the green computation theory [35].

TAI may be expressed as a generalized assignment issue (GAI) [36, 37], which is an NP-hard issue. Among the classic methods for addressing GAP, novel approaches depending on dynamic software development ML [38] approaches, like as reinforcement learning (RL) and neural network (NN) RL (Deep RL), have arisen. Hence this research aims to investigate the ML-based RL convergence with AI in edge computing for task off loadings.

9.2 LITERATURE REVIEW

This study will discuss a few of the most significant techniques discovered through literature study in this part. Numerous approaches employ strategies or procedures that depend on heuristics that look for sub-optimal answers as it can be hard to identify ideal approaches to the TAI, particularly due to the prohibitive processing complication in IoT gadgets [39].

Greedy methods are prevalent methods that deliver suboptimal results at a cheap computing expense and, in some situations, yield results that are extremely near to ideal [40]. Provide a greedy strategy for task assignments utilizing a knapsack-oriented scheduling system (GKS). They present a system in their research that rates jobs based on their utility and weight and allocates the tasks with the greatest potential to occupy the knapsack. They test the method with i-Fog-Sim, among the most widely used for fog computation simulations.

Another method to address optimization issues is to use metaheuristic computations, the most common of which are Genetic Methods, which were motivated by the procedure of natural selection. Jia et al. [41] proposes and solves the collaborative multiple TAI for UAVs employing an altered genetic method. Through computational simulations, they find that their technique delivers computationally achievable and effective results.

Sen and Shen [42] suggest employing RL to resolve the TAI while accounting for timely usage of energy, and then offer a heuristic for addressing the issue and designing the reinforcement structure depending on the results of the suggested heuristic. They build the suggested RL structure as a model-free Q-learning technique and assess it in the i-Fog-Sim simulation. The simulation findings reveal that structure can cut task processing duration and energy usage by 10 to 20% when contrasted to other approaches.

9.3 METHODOLOGY

9.3.1 SYSTEM MODEL

Edge computation platforms are made up of a diverse set of equipment with distinct features and functions [16, 44–47]. Certain equipment has considerable computing capacity and acts as servers for processing tasks; whilst others with less computing capacity continuously produce tasks for

the apps they operate [48]. This results in a tiered design in which equipment is divided into tiers based on their function.

On the highest level of this design, a sequence of offloaded tasks seems, as certain lower-capacity equipment elects to transmit work to more efficient equipment for processing [49]. A task is defined as an indivisible component of computing produced by a specific usage that has unique features and limitations like maximal delay, data dimensions, and processing resources needed. The task assignment issue (TAI) is an essential part of these systems as it is required to identify the best feasible dispersion of tasks between equipment at each tier [50, 51].

An assignment issue is defined as a combinatorial optimization issue with a collection of drivers and an array of tasks [52]. Any task may be allocated to a driver to do, bearing various costs that differ based on the task-driver assignment. The assignment issue entails determining the appropriate task-driver assignment of every task and driver to satisfy particular restrictions while minimizing the expense of operation [53].

The suggested framework is composed of equipment that is divided into 3 levels based on their function. Figure 9.1 depicts the suggested offloading framework. The edge layer, which is nearest to the users, is made up of heterogeneous edge equipment that may have an intermittent link and a shifting location. This tier has the smallest delay and computing capability and is where tasks are produced from IoT edge equipment [54, 55]. Depending on the choice of the offloading technique, tasks may be processed regionally or dispatched to equipment in each of the three tiers.

The fog tier is the tier in the middle wherein fog systems are installed. Fog Systems are tiny data centers with interim processing capability that are situated between edge equipment and the cloud, resulting in intermediary delay [56].

The cloud tier, which has an exceptionally large processing power and a medium-high delay, is the layer on top. The cloud appears to be a single gadget, but in reality, it is an extensive collection of high-performance systems in a data center, thus its processing power may be extremely large [57, 58]. Nevertheless, because of the distance between them and the end customers, their delay is likewise considerable.

The tiers are linked together by both wireless and wired links. The fog and cloud layers are linked by cable, while the edge tier is linked through radio connections to the fog tier and other edge equipment. Every kind of link has different maximal bandwidth. Every edge equipment in the edge

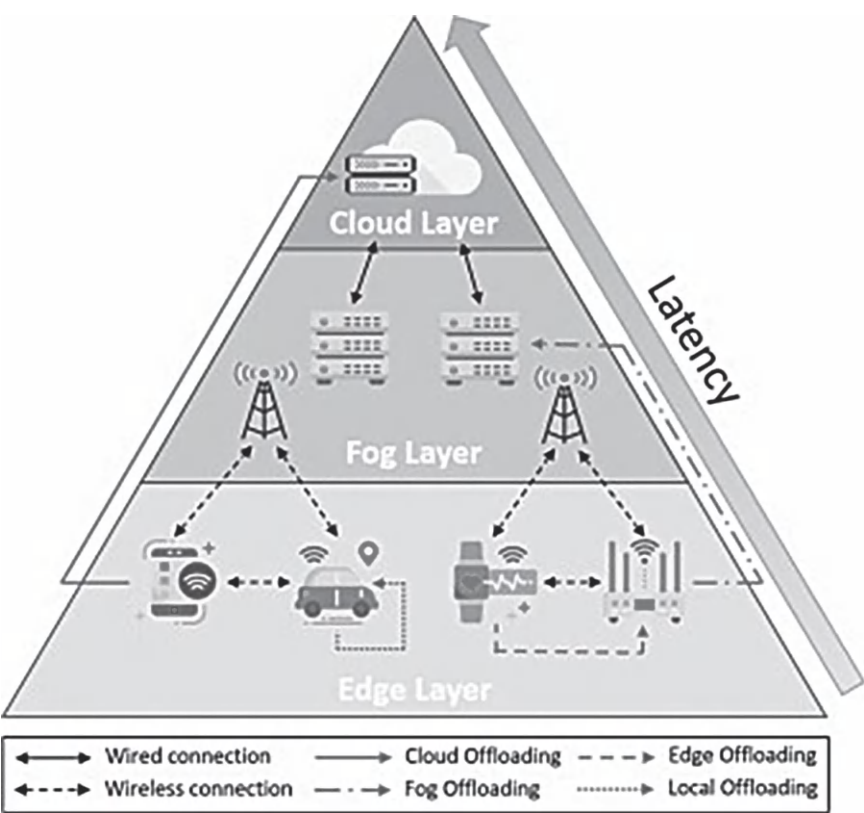


FIGURE 9.1 Edge computing framework.

tier conducts an offloading technique that determines where it offloads the tasks it produces based on the equipment’s regional data [59, 60]. A driver is a decision-making framework that leverages its view of the setting.

9.3.2 TASK OFFLOADING USING RL TECHNIQUE

This part provides a quick overview of the RL concept to help you better grasp our suggestion. After that, we’ll go over our single-layer RL-oriented strategy. We’ll go over the Markov decision procedure, RL, Q-learning, and the statistical definitions of quality, value, and compensation functioning in the subsections that follow. Lastly, this study will show an RL-oriented method for resolving the TAP, describing how the driver’s

condition is determined, how the compensation function is computed, and how the deployed method functions in general [61, 62].

Supervised learning is among the most popular ML techniques, and it makes use of input-output instances to forecast the results of novel data [63]. The only item that can be determined, nevertheless, in certain situations is the result of the activities that are made; data is only available as compensation for certain actions. The system is guided towards the optimum actions based on the utility of the input received after taking an activity. Instead of employing predefined combinations of inputs and outputs as in supervised learning, RL seeks to iteratively learn through trial and error the optimum action to perform in each circumstance based on a particular compensation operation [64].

The agent perceives (or states) the setting to which it is attached and has a range of potential actions to do to affect modification in that setting. Therefore, the agent chooses an activity to carry out from the activity collection after receiving as input the setting's present condition in every repetition. The driver will employ the action's result to determine whether the activity was the proper one. The action will alter the situation and create compensation.

9.3.3 TASK OFFLOADING USING MULTI-LAYER RL TECHNIQUE

The RL technique performance in the multi-driver TOI is enhanced in this part by extending the previous method to a multi-layer methodology.

Each device operates independently inside the technique described in part above, using regional data and global data that has been collected. Decisions are made as a result based on the agent's information and the current regional situation. Inadequate performance in challenging circumstances is brought about by the initial phases of the technique's incomplete information and biased perception of the context. To address this issue, this research suggests allowing equipment's RL driver to delegate the decision to offload to a top-level driver if it lacks sufficient data [65]. The decision will be made by a higher-level driver, who may be placed in the fog layer, based on its information and the overall status of the system. Both the regional and the higher-level drivers will pick up knowledge from the compensation received after carrying out the task from the offloading decision supplied to the questioning equipment.

In this improved design, an autonomous RL technique that can cooperate between layers is executed on both edge equipment and fog platforms [66, 67]. When an edge driver begins out without any information and conducts inquiries to study the decisions of the higher-level driver, the offloading question permits a passive exchange of information from fog drivers to the edge.

To integrate the novel offloading question activity into the list of viable activities for edge equipment and question administration in the fog system, this study expands the performance of the RL offloading technique from the previous part. Given that they each have a unique perspective on their surroundings, edge equipment, and fog systems behave differently when executing the identical technique [68, 69].

The edge equipment driver functions in a manner that is identical to the previous one, with the exception that it now can execute the offloading question. Because it is believed that a more powerful driver would make better decisions, rewards earned for actions inquired to a fog system are thought to have more significance [70].

Additionally, the Q-value of the query activity, which is employed to lessen the likelihood of selection while enhancing the agent's regional information, is penalized [71, 72]. On the contrary, the higher layer driver waits for offloading requests from other drivers before making an offloading choice and sending it back. It does this using its information and understanding of the present scenario as well as views of the real-time fog and cloud CPU applications [73]. As an outcome, when the offloading activity is completed, an incentive will be received that will help both the driver who made the inquiry and the higher-level driver learn more.

The same technique is used by every equipment as was already stated, but each one has its unique understanding of the state. This implies that all equipment will control a separate Q-Table that is regionally educated [74]. Each communication with equipment will be used by the higher-layer driver for updating their state globally.

9.4 RESULTS AND DISCUSSION

9.4.1 PERFORMANCE ASSESSMENT

In this part, this research contrasts the suggested method in a simulation edge computation context to the greedy and single-layer RL alternates.

This study will run a large number of simulations with every method to evaluate their efficacy employing an array of measures. The outcomes of the simulation can be found on our GitHub page [61].

The objective of the assessment is to collect sufficient information to contrast the offloading strategies equitably. As a result, this research will employ an edge computing simulation to investigate the actions of the methods as they transition from low to higher-density equipment situations. This study believes that a region is feasible for a Smart-Campus IoT situation [43], with a minimal count of gadgets of 10. On the contrary, a high-density situation with greater than 200 edge equipment is considered.

Every simulation will produce an array of measures that will be employed to evaluate the method's performance in the particular simulated situation. To avoid erroneous findings generated by the simulator's randomized element, measurements will be derived by averaging the outcomes of multiple simulations performed under identical circumstances.

9.4.2 STUDY CONFIGURATION

The assessment was carried out on an altered variant of the Pure-Edge-Sim edge computation simulator, and the source program for the addition is accessible on GitHub. To satisfy the needs, this study enhanced the simulator's capabilities by adding novel effectiveness indicators and RL methods. The simulator operates in a Java 1.8 context on a machine using an AMD Ryzen 9 5900X processor and 32 GB of DDR4 RAM.

The simulation edge computation circumstances comprise 3 tiers of equipment in an edge-cloud design, with the edge equipment randomly scattered regions. To test the adaptability of the suggested methods, the count of edge equipment for every simulated is raised by 10 until it reaches 200. To compute the average outcome, every simulation is run 10 instances for each setup, lasting 10 minutes. At the fog and edge levels, the link between equipment has a bandwidth of a hundred megabits per second, while the link to the cloud tier has a bandwidth of 20 megabits per second. The portable link's longest reach for edge equipment is 40 meters. Table 9.1 provides a summary of the Pure-Edge-Sim settings simulation variables.

TABLE 9.1 Simulation Pure-Edge-Sim Variables

Variables	Values
Average simulations	10 min
Period simulation	10 per setup
Area	200 × 200 m
Range of edge	40 m
Bandwidth of fog and edge	100 Mbps
Bandwidth of cloud	20 Mbps
Maximum no. of edge gadgets	200
Minimum no. of edge gadgets	10

Table 9.2 lists the qualities that the Pure-Edge-Sim simulator needs to define the components of the suggested 3-tier design. Sensors cannot execute regional offloading or accept duties from other equipment since all equipment, with have sufficient computational power to acquire and procedure tasks. Just sensors and mobile devices produce tasks for the network amongst edge equipment.

TABLE 9.2 Pure-Edge-Sim equipment features.

Equipment Kind	Edge Data center	Cloud Data center	Laptop	Smartphone	Raspberry	Sensor
Layer	Fog	Cloud	Edge	Edge	Edge	Edge
No. of devices	4	1	20%	30%	20%	40%
MIPS	1,000,000	2,000,000	110,000	25,000	16,000	–
CPU core	8	8	8	8	4	–
Generate tasks	–	–	No	Yes	No	Yes
Maximum consumption (Wh/s)	0.016	0.016	23.6	3.3	5.1	0.036
Idle consumption (Wh/s)	0.00015	0.00015	1.7	0.078	1.6	0.0011

Three simulated uses with distinct delay and processing needs are used to examine the actions of the offloading techniques. Table 9.3 lists the requirements of every program by Pure-Edge-Sim standards as well as the proportion of utilization on every system.

TABLE 9.3 Pure-Edge-Sim Application Variables

	Augmented Reality	Health Application	Data Processing
Task length (MI)	120,000	600,000	18,000
Application percentage	20	25	30
Maximum delay (s)	6	30	10
Generation rate (task/m)	45	35	45

9.4.3 MEASURES

This research describes the standard measures listed below for contrasting every method’s performance:

- **Task Performance Rate:** The portion of all tasks that complete their implementation. If the task’s execution period goes over its allotted period or the offloading procedure collapses, the task is not deemed to have been correctly completed. For the assessment procedure, this measure is among the most crucial.
- **Average Overall Duration:** The sum of the execution duration and the sending duration to the processing unit that is needed to effectively finish a task. This measure is particularly helpful for assessing how much delay every method causes.
- **Complete Average Overall Duration:** This measure uses the identical formula as Average Overall Duration, but also takes into account time lost on unsuccessfully completed tasks.
- **Tasks That Crashed Because of Delay:** The number of tasks that crashed since the maximum delay was reached during their execution.
- **Average CPU Consumption Per Equipment:** A device’s average CPU consumption. Helpful for calculating how much the equipment’s processing capabilities are being employed.
- **Average Energy Utilization per Equipment:** The typical amount of energy used by one gadget.

9.4.4 COMPARED TECHNIQUES

To assess the 3 methods presented in this work performance, this research built them in the simulation. Since the approaches’ computing costs are

modest in contrast to the large expense of processing the offloaded activities, this study assumes the tests that the CPU and power expenditures of the techniques are insignificant. The single-layer RL method, as well as the suggested multi-layer directed RL, will be evaluated in contrast to the greedy approach. Table 9.4 displays the variables that the greedy method’s heuristic took into consideration.

TABLE 9.4 Greedy Algorithm Variables

Greedy Algorithm Variables	Value
Closest gadget weighing	1.1
Cloud weighing	1.8
Fog system weighing	1.5
CPU trade-off	20

On the contrary, this research created 3 approaches depending on the RL approaches described in earlier parts that were put into practice. The initial one is the fundamental execution of an RL method, which executes regionally on every gadget without the aid of outside information and is referred to in the experiments as “Local RL.” The multi-layer RL method is implemented identically in the 2nd and 3rd approaches but with various starting circumstances. The “RL Multilayered Empty” kind begins every simulation with each of the gadgets’ Q-Tables (information) empty, whereas the “RL Multilayer” variant employs the Q-Table from earlier models on the fog systems with the identical setup to simulate how a network would behave if it had information at the beginning to enhance initial performance. Table 9.5 lists the variables that both techniques employed.

TABLE 9.5 RL Method Variables

RL Method Variables	Multi	Single
Studying ratio	0.6	0.6
Delay-energy-trade-off	0.003	0.003
Failure penalty	1000	1000
Discount factor	0.3	0.3
Average CPU refresh ratio	60	60
Query usage penalty	10	–
Query compensation	0.2	–
Initial query Q-value	10	–
Initial Q-value	200	200

The simulated outcomes that were run for every method and setting are shown in this part. All of the subfigures in Figure 9.2 indicate a measure that was specified to allow for comparing methods from an instance with 10 gadgets up to two hundred, and the result will be the average result of simulating 10 instances.

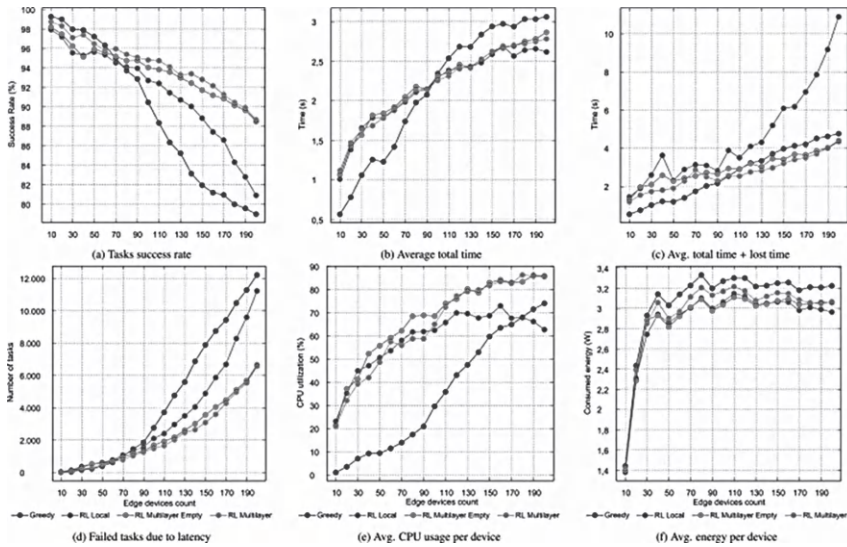


FIGURE 9.2 Outcome of simulation.

In practice, this has the greatest detrimental effects on the end-user, among the most important outcomes is the task implementation performance rate. According to every method and setup used to run the simulation, the performance rate is shown in Figure 9.2(a). This study may observe that the greedy technique performs better than the others in minimal equipment density situations up to a point of medium equipment density, or 70 equipment, at which point its performance rate begins to decline. While both multi-layer approaches are still capable of maintaining adequate performance under elevated equipment density circumstances, the performance of the greedy technique and the RL single-layer is drastically reduced.

This action results from the fact that there aren't many jobs in minimal equipment density circumstances, and most of them may be completed by fog and cloud-based systems without saturating, therefore the greedy

technique's heuristics perform more effectively than RL techniques in these situations. When a moderate density of equipment is attained, the suitable usage of assets becomes more important, and methods based on RL approaches may modify dynamically to maintain the highest performance rate feasible. Because optimum utilization of processing units is crucial in higher-density situations with an abundance of tasks, the greedy technique cannot obtain decent outcomes, and even the single-layer RL technique cannot enhance the outcome. Multi-layer RL approaches, on the other hand, have a high success rate because of the ability to delegate decision-making to higher-level drivers. The multi-layer RL technique, which begins with the Q-Table of the fog systems populated with the data discovered through earlier models, yields the highest performance rate as it enables the provision of helpful information to the equipment at an initial phase of studying.

Additionally, as demonstrated by the simulated instances in Figure 9.3, the multi-layer technique provides higher real-time efficiency and speed of convergence than the single-layer even in a state that begins with no prior understanding. At first, the single-layer approach performs poorly, but throughout the experiment, it gradually improves to its best state. Though the offloading of the query causes the performance rate to gradually increase on its own, the multi-layer technique rapidly surpasses the single-layer technique's highest performance rate. The optimal outcome is also immediately achieved and maintained throughout the experiment via the multi-layer technique, which builds on the Q-tables of the fog drivers learned from earlier experiments. On the contrary, because the greedy technique's heuristics lack any dynamic elements, they have an established and steady nature.

Figure 9.2(b) displays the average duration needed to finish a task for every technique and the count of equipment. Similar to performance rate, the greedy technique's average overall duration significantly alters performance depending on the number of equipment; whereas RL approaches gradually alter the average overall duration. As this statistic only takes into account tasks that have effectively concluded their implementation, it can be confusing to compare the 3 RL approaches because they all yield a comparable average overall duration.

Another measure is established that additionally takes into account the time lost owing to tasks that fail to be implemented correctly as a result of delay to better understand the performance of the procedures in terms

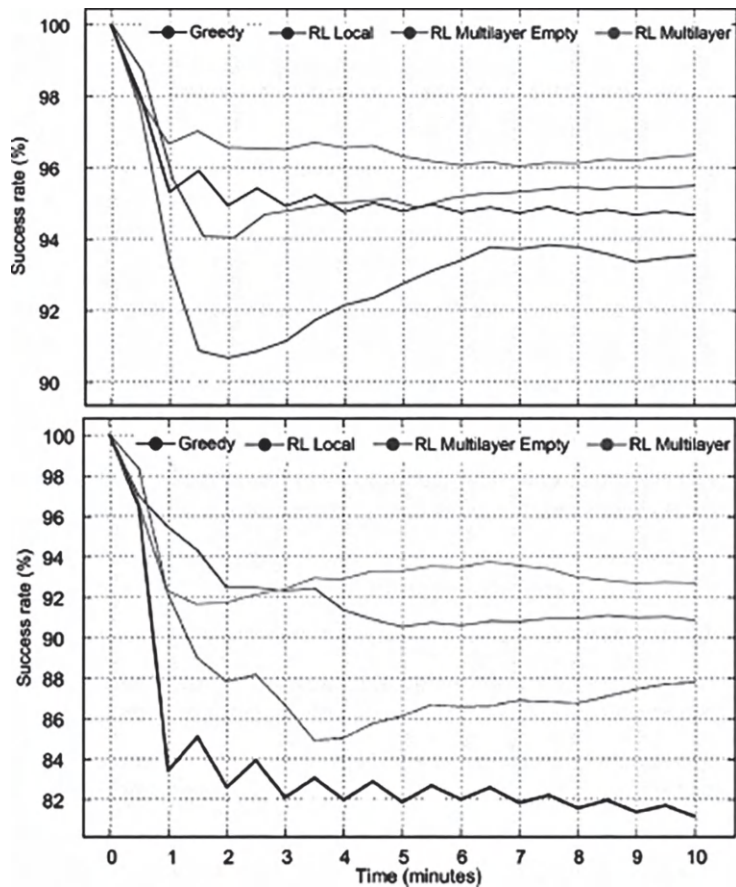


FIGURE 9.3 The development of task performance rates (on 70 and 170 equipment).

of the overall implementation duration. Figure 9.2(c) illustrates the actual effects of the technique’s decision to perform an inappropriate offloading. While the single-layer RL technique significantly lengthens the duration in excessive equipment density instances, the greedy technique’s actions behave similarly to Figure 9.2(b). This is because the influence of incorrect choices in complicated circumstances is extremely large (e.g., choosing to send tasks to the cloud system, saturating it, rather than dispersing them among the edge equipment). As may be observed in Figure 9.2(d), multi-layer RL approaches, on the other hand, eliminate the beginning ambiguity by delegating the decision, resulting in superior offloading decisions that decrease delay breakdowns.

The average CPU utilization per equipment, which shows the extent of resource utilization of the system's computing resources, is another pertinent statistic that may be evaluated. As resource utilization is maximized, suitable task dispersion among the equipment produces an elevated average CPU consumption per equipment. Minimal CPU consumption, on the other hand, shows that a few of the equipment are idle while a small number of them are being saturated by the technique. The 2 multi-layer RL approaches stand out among the others, and the greedy technique exhibits a low usage of computing power, as demonstrated in Figure 9.2(e), which depicts the simulation's output for this measure.

The average energy usage per equipment as determined from the experiments is shown in Figure 9.2(f), and similarly, the performance of techniques may be evaluated in the context of their energy usage, which is one of the elements of the optimization issue. The techniques that use RL have minimal energy usage, whereas the greedy techniques have the largest usage of energy. Due to the minimal power usage of all the gadgets in the situation, there is little variance between the strategies. The amazing aspect of this conclusion, however, is how the RL approaches can cut usage even when there is little room for enhancement while still greatly enhancing the outcome in the other measures.

Since application delay in the experiment situation is a crucial element that directly influences the rate of effective implementation, all of the experiments have been run using a trade-off factor-beta of 0.003, while power utilization has significance but has no impact on task performance rates. In other circumstances, though, where utilization of energy is critical, this study would be keen on boosting the beta variable. For instance, it would be especially crucial to correctly handle the utilization of energy to prevent exhausting battery life in an environment where the maximal allowable waiting periods for tasks are excessive and the edge equipment is driven by charges. This is true even if it results in longer task implementation periods.

In this instance, this action occurs since tasks have a limited maximal duration for waiting. As a result, any system saturation that makes a task wait to begin implementation will result in the task's due date being surpassed, at which point the job will be deemed unsuccessful. The outcome shows that as beta increases, tasks that are effectively completed must wait further because they are routed to systems that are more effective but also more congested. The outcomes also show that this study observes

the tendency for edge-level equipment to employ less CPU as beta rises. As a result, tasks are executed in the cloud and fog tiers, which are more energy-effective but incapable of managing an increased number of tasks, having a significant negative influence on task implementation duration.

This study may conclude that the greedy technique provides adequate performance in minimal and moderate equipment density situations after observing how the techniques performed in various simulation conditions. To preserve system performance, more sophisticated techniques should be used, nevertheless, as equipment density rises. The ideal simulation outcomes are achieved by RL techniques because they can adjust to complicated situations at an inexpensive computing expense. Additionally, the multi-layer technique differs from previous approaches since it performs well in the most crucial measures in challenging higher-density circumstances. This enhancement is the result of an improved offloading decision process that makes use of outside information, and it shows how well the multi-layer RL approach performs. As a result, RL techniques are effective tools for addressing the task assignment issue, and the suggestion is a practical and simple addition that can be made to any RL technique to boost output.

9.5 CONCLUSION

The TAI has been highlighted in this study as a crucial aspect of collaboration edge computation designs. There are several ways to solve TAI, as was demonstrated in the earlier part, but the ones depending on AI appear to have the most promise. The task offloading procedure in the suggested 3-layer cloud-edge-fog architecture is addressed using RL, an ML method.

To comprehend the effects of task dispersion and the restricted perspective of RL drivers and how they influence the operational action of the method in complicated scenarios, this research has explored various setups in this study. This research suggests a unique expansion of RL approaches to address these issues by enabling drivers to ask queries of a top-level driver who has more information and a wider perspective on the context.

In an altered edition of the Pure-Edge-Sim simulation, this research executed the suggestions along with a greedy alternate and ran multiple experiments to contrast the effectiveness of every method in various scenarios while adhering to established measurements, causing the

outcomes and models available for replication. The empirical findings demonstrated that, in comparison to the greedy and conventional RL methods, our suggested multilayer RL method outperformed them significantly in situations with an abundance of gadgets and tasks.

This concept uses a passive method of information transmission; yet, in future work, this research aims to investigate other methods, such as federation learning and information transmission, to create dispersed decision-making platforms for 5G networking and the Computation Continuum model. Applying neural networks to enhance the effectiveness of Q-Learning can help it overcome some drawbacks like a lack of stages and discrete elements. This is an exciting field for future studies.

KEYWORDS

- artificial intelligence
- machine learning
- reinforcement learning
- edge computing
- task offloading

REFERENCES

1. Russell, S., & Norvig, P. (2002). *Artificial Intelligence: A Modern Approach*. Prentice Hall.
2. Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255–260.
3. Grosky, W. I. (1994). Multimedia information systems. *IEEE Multimedia*, 1(1), 12–24.
4. Chew, L. W., Chia, W. C., Ang, L. M., & Seng, K. P. (2012). Low-memory video compression architecture using strip-based processing for implementation in wireless multimedia sensor networks. *International Journal of Sensor Networks*, 11(1), 33–47.
5. Silva, B. N., Khan, M., & Han, K. (2018). Towards sustainable smart cities: A review of trends, architectures, components, and open challenges in smart cities. *Sustainable Cities and Society*, 38, 697–713.
6. Ang, L. M., Seng, K. P., Zungeru, A. M., & Ijamaru, G. K. (2017). Big sensor data systems for smart cities. *IEEE Internet of Things Journal*, 4(5), 1259–1271.

7. Mehmood, Y., Ahmad, F., Yaqoob, I., Adnane, A., Imran, M., & Guizani, S. (2017). Internet-of-things-based smart cities: Recent advances and challenges. *IEEE Communications Magazine*, 55(9), 16–24.
8. Zhou, Z., Chen, X., Li, E., Zeng, L., Luo, K., & Zhang, J. (2019). Edge intelligence: Paving the last mile of artificial intelligence with edge computing. *Proceedings of the IEEE*, 107(8), 1738–1762.
9. Osifeko, M. O., & Hancke, G. P. (2020). Artificial intelligence techniques for cognitive sensing in future IoT: State-of-the-art, potentials, and challenges. *Journal of Sensor and Actuator Networks*, 9(3), 21.
10. Gasser, R., Rossetto, L., & Schuldt, H. (2019). *Towards an All-Purpose Content-Based Multimedia Information Retrieval System*. *arXiv preprint arXiv:1902.03878*.
11. Ferziger, J. H., Perić, M., & Street, R. L. (2019). *Computational Methods for Fluid Dynamics*. Springer.
12. Rajab, H., & Cinkelnr, T. (2018). IoT-based smart cities. In *2018 International Symposium on Networks, Computers, and Communications (ISNCC)* (pp. 1–4). IEEE.
13. Zhang, W., Zhang, R., Wu, C., Goh, A. T., Lacasse, S., Liu, Z., & Liu, H. (2020). State-of-the-art review of soft computing applications in underground excavations. *Geoscience Frontiers*, 11(4), 1095–1106.
14. Paracha, K. N., Rahim, S. K., Soh, P. J., & Khalily, M. (2019). Wearable antennas: A review of materials, structures, and innovative features for autonomous communication and sensing. *IEEE Access*, 7, 56694–56712.
15. Xie, Y., Takikawa, T., Saito, S., Litany, O., Yan, S., Khan, N., Tombari, F., Tompkin, J., Sitzmann, V., & Sridhar, S. (2022). Neural fields in visual computing and beyond. In *Computer Graphics Forum* (Vol. 41, No. 2, pp. 641–676).
16. Varghese, B., Wang, N., Barbhuiya, S., Kilpatrick, P., & Nikolopoulos, D. S. (2016). Challenges and opportunities in edge computing. In *Proceedings of the 2016 IEEE International Conference on Smart Cloud (SmartCloud)* (pp. 20–26).
17. Jones, D. E., Alimi, T. O., Pordell, P., Tangka, F. K., Blumenthal, W., Jones, S. F., Rogers, J. D., Benard, V. B., & Richardson, L. C. (2021). Pursuing data modernization in cancer surveillance by developing a cloud-based computing platform: Real-time cancer case collection. *JCO Clinical Cancer Informatics*, 5, 24–29.
18. Alam, A. (2022). Cloud-based e-learning: Scaffolding the environment for adaptive e-learning ecosystem based on cloud computing infrastructure. In *Computer Communication, Networking and IoT: Proceedings of 5th ICICC 2021* (Vol. 2 pp. 1–9). Springer Nature Singapore.
19. Cao, K., Liu, Y., Meng, G., & Sun, Q. (2020). An overview on edge computing research. *IEEE Access*, 8, 85714–85728.
20. Zhang, X., Chen, H., Zhao, Y., Ma, Z., Xu, Y., Huang, H., Yin, H., & Wu, D. O. (2019). Improving cloud gaming experience through mobile edge computing. *IEEE Wireless Communications*, 26(4), 178–183.
21. Chen, J., & Ran, X. (2019). Deep learning with edge computing: A review. *Proceedings of the IEEE*, 107(8), 1655–1674.
22. Deng, S., Zhao, H., Fang, W., Yin, J., Dustdar, S., & Zomaya, A. Y. (2020). Edge intelligence: The confluence of edge computing and artificial intelligence. *IEEE Internet of Things Journal*, 7(8), 7457–7469.

23. Shakarami, A., Shahidinejad, A., & Ghobaei-Arani, M. (2020). A review on the computation offloading approaches in mobile edge computing: A game-theoretic perspective. *Software: Practice and Experience*, 50(9), 1719–1759.
24. Luo, Q., Hu, S., Li, C., Li, G., & Shi, W. (2021). Resource scheduling in edge computing: A survey. *IEEE Communications Surveys & Tutorials*, 23(4), 2131–2165.
25. Khan, F., Khan, M. A., Abbas, S., Athar, A., Siddiqui, S. Y., Khan, A. H., Saeed, M. A., & Hussain, M. (2020). Cloud-based breast cancer prediction empowered with soft computing approaches. *Journal of Healthcare Engineering*, 2020.
26. Sadeeq, M. M., Abdulkareem, N. M., Zeebaree, S. R., Ahmed, D. M., Sami, A. S., & Zebari, R. R. (2021). IoT and cloud computing issues, challenges and opportunities: A review. *Qubahan Academic Journal*, 1(2), 1–7.
27. Anthes, C., García-Hernández, R. J., Wiedemann, M., & Kranzlmüller, D. (2016). State of the art of virtual reality technology. In *2016 IEEE Aerospace Conference* (pp. 1–19). IEEE.
28. Badue, C., Guidolini, R., Carneiro, R. V., Azevedo, P., Cardoso, V. B., Forechi, A., Jesus, L., Berriel, R., Paixao, T. M., Mutz, F., de Paula Veronese, L. (2021). Self-driving cars: A survey. *Expert Systems with Applications*, 165, 113816.
29. Almurisi, N., & Tadisetty, S. (2022). Cloud-based virtualization environment for IoT-based WSN: Solutions, approaches, and challenges. *Journal of Ambient Intelligence and Humanized Computing*, 13(10), 4681–4703.
30. Bashar, D. A. (2020). Sensor cloud-based architecture with efficient data computation and security implantation for Internet of Things applications. *Journal of IoT in Social, Mobile, Analytics, and Cloud*, 2(2), 96–105.
31. Čolaković, A., & Hadžialić, M. (2018). Internet of Things (IoT): A review of enabling technologies, challenges, and open research issues. *Computer Networks*, 144, 17–39.
32. Noor, T. H., Zeadally, S., Alfazi, A., & Sheng, Q. Z. (2018). Mobile cloud computing: Challenges and future research directions. *Journal of Network and Computer Applications*, 115, 70–85.
33. Al-Janabi, S., Al-Shourbaji, I., Shojafar, M., & Abdelhag, M. (2017). Mobile cloud computing: Challenges and future research directions. In *2017 10th International Conference on Developments in E-systems Engineering (DeSE)* (pp. 62–67). IEEE.
34. Tanwar, S., Tyagi, S., Budhiraja, I., & Kumar, N. (2019). Tactile internet for autonomous vehicles: Latency and reliability analysis. *IEEE Wireless Communications*, 26(4), 66–72.
35. Mehrabi, A., Siekkinen, M., Kämäräinen, T., & Ylä-Jääski, A. (2021). Multi-tier cloud: Leveraging edge computing in remote rendered virtual reality. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 17(2), 1–24.
36. Naha, R. K., Garg, S., Georgakopoulos, D., Jayaraman, P. P., Gao, L., Xiang, Y., & Ranjan, R. (2018). Fog computing: Survey of trends, architectures, requirements, and research directions. *IEEE Access*, 6, 47980–48009.
37. Fraga-Lamas, P., Lopes, S. I., & Fernández-Caramés, T. M. (2021). Green IoT and edge AI as key technological enablers for a sustainable digital transition towards a smart circular economy: An Industry 5.0 use case. *Sensors*, 21(17), 5745.
38. Mohapatra, S. K., Nayak, P., Mishra, S., & Bisoy, S. K. (2019). Green computing: A step towards eco-friendly computing. In *Emerging Trends and Applications in Cognitive Computing* (pp. 124–149). IGI Global.

39. Shakarami, A., Ghobaei-Arani, M., Masdari, M., & Hosseinzadeh, M. (2020). A survey on the computation offloading approaches in mobile edge/cloud computing environment: A stochastic-based perspective. *Journal of Grid Computing*, 18(4), 639–671.
40. Rahbari, D., & Nickray, M. (2019). Low-latency and energy-efficient scheduling in fog-based IoT applications. *Turkish Journal of Electrical Engineering and Computer Sciences*, 27(2), 1406–1427.
41. Jia, Z., Yu, J., Ai, X., Xu, X., & Yang, D. (2018). Cooperative multiple task assignment problem with stochastic velocities and time windows for heterogeneous unmanned aerial vehicles using a genetic algorithm. *Aerospace Science and Technology*, 76, 112–125.
42. Sen, T., & Shen, H. (2019). Machine learning based timeliness-guaranteed and energy-efficient task assignment in edge computing systems. In *2019 IEEE 3rd International Conference on Fog and Edge Computing (ICFEC)* (pp. 1–10). IEEE.
43. Mechalikh, C., Taktak, H., & Moussa, F. (2019). PureEdgeSim: A simulation toolkit for performance evaluation of cloud, fog, and pure edge computing environments. In *2019 International Conference on High-Performance Computing & Simulation (HPCS)* (pp. 700–707). IEEE.
44. Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. *IEEE Internet of Things Journal*, 3(5), 637–646.
45. Shi, W., & Dustdar, S. (2016). The promise of edge computing. *Computer*, 49(5), 78–81.
46. Satyanarayanan, M. (2017). The emergence of edge computing. *Computer*, 50(1), 30–39.
47. Chen, J., & Ran, X. (2019). Deep learning with edge computing: A review. *Proceedings of the IEEE*, 107(8), 1655–1674.
48. Bentaha, M. L., Battaia, O., & Dolgui, A. (2015). An exact solution approach for disassembly line balancing problem under uncertainty of the task processing times. *International Journal of Production Research*, 53(6), 1807–1818.
49. Islam, A., Debnath, A., Ghose, M., & Chakraborty, S. (2021). A survey on task offloading in multi-access edge computing. *Journal of Systems Architecture*, 118, 102225.
50. Ngo, S. T., Jaafar, J., Aziz, I. A., & Anh, B. N. (2021). A compromise programming for multi-objective task assignment problem. *Computers*, 10(2), 15.
51. Zhu, M., Du, X., Zhang, X., Luo, H., & Wang, G. (2019). Multi-UAV rapid-assessment task-assignment problem in a post-earthquake scenario. *IEEE Access*, 7, 74542–74557.
52. El-Ashmawi, W. H., & Ali, A. F. (2020). A modified salp swarm algorithm for the task assignment problem. *Applied Soft Computing*, 94, 106445.
53. Arratia-Martinez, N. M., Maya-Padron, C., & Avila-Torres, P. A. (2021). University course timetabling problem with professor assignment. *Mathematical Problems in Engineering*, 2021, 1–9.
54. Zhao, Y., Cheng, G., Duan, Y., Gu, Z., Zhou, Y., & Tang, L. (2021). Secure IoT edge: Threat situation awareness based on network traffic. *Computer Networks*, 201, 108525.

55. Lv, Z. (2021). Security of Internet of Things edge devices. *Software: Practice and Experience*, 51(12), 2446–2456.
56. Yi, S., Hao, Z., Qin, Z., & Li, Q. (2015). Fog computing: Platform and applications. In *2015 Third IEEE Workshop on Hot Topics in Web Systems and Technologies (HotWeb)* (pp. 73–78). IEEE.
57. Duggal, A., Jenkins, F., Shilane, P., Chinthekindi, R., Shah, R., & Kamat, M. (2019). Data Domain Cloud Tier: Backup here, backup there, deduplicated everywhere! In *2019 USENIX Annual Technical Conference (USENIX ATC 19)* (pp. 647–660).
58. Ganchev, I., Ji, Z. H., & O'Droma, M. Á. (2019). Designing a cloud tier for the IoT platform EMULSION. *WSEAS Transactions on Systems and Control*, 14(46), 375–383.
59. Kukunuri, R., Aglawe, A., Chauhan, J., Bhagtani, K., Patil, R., Walia, S., & Batra, N. (2020). Edgenilm: Towards NILM on edge devices. In *Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation* (pp. 90–99).
60. Jhang, C. J., Xue, C. X., Hung, J. M., Chang, F. C., & Chang, M. F. (2021). Challenges and trends of SRAM-based computing-in-memory for AI edge devices. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 68(5), 1773–1786.
61. Van Hasselt, H., Guez, A., & Silver, D. (2016). Deep reinforcement learning with double Q-learning. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 30, No. 1).
62. Kumar, A., Zhou, A., Tucker, G., & Levine, S. (2020). Conservative Q-learning for offline reinforcement learning. *Advances in Neural Information Processing Systems*, 33, 1179–1191.
63. Naimi, A. I., & Balzer, L. B. (2018). Stacked generalization: An introduction to super learning. *European Journal of Epidemiology*, 33, 459–464.
64. Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. MIT Press.
65. Matsuo, Y., LeCun, Y., Sahani, M., Precup, D., Silver, D., Sugiyama, M., Uchibe, E., & Morimoto, J. (2022). Deep learning, reinforcement learning, and world models. *Neural Networks*, 152, 267–275.
66. Moerland, T. M., Broekens, J., Plaat, A., & Jonker, C. M. (2023). Model-based reinforcement learning: A survey. *Foundations and Trends® in Machine Learning*, 16(1), 1–18.
67. Tran-Dang, H., Bhardwaj, S., Rahim, T., Musaddiq, A., & Kim, D. S. (2022). Reinforcement learning based resource management for fog computing environment: Literature review, challenges, and open issues. *Journal of Communications and Networks*, 24(1), 83–98.
68. Talaat, F. M., Saraya, M. S., Saleh, A. I., Ali, H. A., & Ali, S. H. (2020). A load balancing and optimization strategy (LBOS) using reinforcement learning in a fog computing environment. *Journal of Ambient Intelligence and Humanized Computing*, 11, 4951–4966.
69. Wang, Y., Wang, K., Huang, H., Miyazaki, T., & Guo, S. (2018). Traffic and computation co-offloading with reinforcement learning in fog computing for industrial applications. *IEEE Transactions on Industrial Informatics*, 15(2), 976–986.

70. Zabihi, Z., Eftekhari Moghadam, A. M., & Rezvani, M. H. (2023). Reinforcement learning methods for computation offloading: A systematic review. *ACM Computing Surveys*, 56(1), 1–41.
71. Qu, G., Wu, H., Li, R., & Jiao, P. (2021). DMRO: A deep meta reinforcement learning-based task offloading framework for edge-cloud computing. *IEEE Transactions on Network and Service Management*, 18(3), 3448–3459.
72. Wang, J., Hu, J., Min, G., Zomaya, A. Y., & Georgalas, N. (2020). Fast adaptive task offloading in edge computing based on meta reinforcement learning. *IEEE Transactions on Parallel and Distributed Systems*, 32(1), 242–253.
73. Zhu, X., Luo, Y., Liu, A., Bhuiyan, M. Z., & Zhang, S. (2020). Multiagent deep reinforcement learning for vehicular computation offloading in IoT. *IEEE Internet of Things Journal*, 8(12), 9763–9773.
74. Yang, Y., Hao, J., Chen, G., Tang, H., Chen, Y., Hu, Y., Fan, C., & Wei, Z. (2020). Q-value path decomposition for deep multiagent reinforcement learning. In *International Conference on Machine Learning* (pp. 10706–10715). PMLR.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

CHAPTER 10

Applications, Enabling Technologies, Vision of Edge Artificial Intelligence for 6G

TAMIRAT TAGESSE TAKORE,¹ KONDA HARI KRISHNA,²
K. SURESH KUMAR,³ R. GANESH KUMAR,⁴ AMIT JAIN,⁵ and
T. KEERTHIKA⁶

*¹Department of Electrical and Computer Engineering, Wachemo
University, India*

*²Department of CSE, School of Computing, Mohan Babu University,
Tirupati, Andhra Pradesh, India*

*³MBA Department, Panimalar Engineering College, Varadarajapuram,
Poonamallee, Chennai, Tamil Nadu, India*

*⁴Department of Computer Science and Engineering, CHRIST (Deemed
to be University), School of Engineering and Technology, Kengeri
Campus, Kanminike, Bangalore, Karnataka, India*

*⁵Computer Science and Engineering Department, OP Jindal University,
Raigarh, Chhattisgarh, India*

*⁶Associate Professor, Department of IT, Sri Krishna College of Engineering
and Technology, Coimbatore, India*

ABSTRACT

The advancement in applications of artificial intelligence technologies that propels the ongoing development of wireless networks. Envisioned

Advances in AI for Cloud, Edge, and Mobile Computing Applications.

Shrikaant Kulkarni and P. William (Eds.)

© 2025 Apple Academic Press, Inc. Co-published with CRC Press (Taylor & Francis)

as a transformative force, 6G is expected to shift wireless evolution from “connecting devices” to “enabling intelligent connections.” However, cutting-edge AI systems relying on deep learning and extensive data analytics demand significant computational and communication resources, leading to notable issues such as privacy breaches, network congestion, energy consumption and latency during both inference and training phases. Edge AI emerges as a disruptive technology for 6G by incorporating inference capabilities and model training at the network’s edge. This approach seamlessly integrates intelligence, computation, communication and sensing enhancing the security, privacy, effectiveness and efficiency of the 6G technologies. The study includes the applications, enabling technologies dependable and scalable edge AI technology through an integrated approach that encompasses decentralized models of machine learning and strategies of wireless communication. We will delve into new wireless network design principles, optimization methods for resource allocation driven by service requirements, and an all-encompassing architecture of end-to-end system supporting edge AI. Additionally, we will touch upon application scenarios, hardware and software platforms, and standardization for facilitating the commercialization and industrial adoption of edge AI systems.

10.1 INTRODUCTION

With the global implementation of 5G networks through standardization efforts, researchers, governments and corporations, have initiated discussions about the vision, potential applications, and groundbreaking technologies for the future 6G networks. Notably, countries like the China, European Union and United States have recently allocated funds for 6G projects [1–3], all sharing the common objective of making possible networked intelligence. The international telecommunication union (ITU) has also published Network 2030’s guiding principles and system criteria [4].

Numerous businesses are looking at new use cases to improve real-time immersive experiences, interactions, and speed up the integration of intelligence into digital twin technologies and the Industrial Internet of Things. For instance, focusing on the effective use of 5G technologies in the present world that include improved broadband using mobile network communication of low latency that are ultra-reliable and communication of massive machine type [5, 6]. Many recent proposed technologies are introduced in Huawei that include application scenarios of three additional

types with context of 5.5 G technologies. This type of situations involves broadband communication of real time uplink centric synchronized sensing and communication as well as broadband communication [7].

10.2 LITERATURE REVIEW

The anticipation for 6G extends beyond mobile internet capabilities, with a focus on supporting ubiquitous Internet of Everything and Artificial Intelligence applications [1–4, 8]. These applications encompass various domains such as metaverse, extended reality, high-fidelity holographic society, haptic and tactile internet, digital twins, brain computer interfaces, connected autonomous system and sustainable cities [9], e-health, and more. Both academic researchers and industries have proposed numerous visionary ideas for 6G [10–12], aiming to provide a more comprehensive understanding of, and interaction with, the physical universe. Particularly, event-defined URLLC, contextuality agile CMBB communications, and computation-oriented communications have all been proposed as unique application services for 6G [12]. Figure 10.1 depiction of the progression of 6G’s futuristic use cases highlights the incorporation of computation, sensing, coordination, and intelligence in building a connected cyber-physical world by drawing on the aforementioned utilization scenarios.

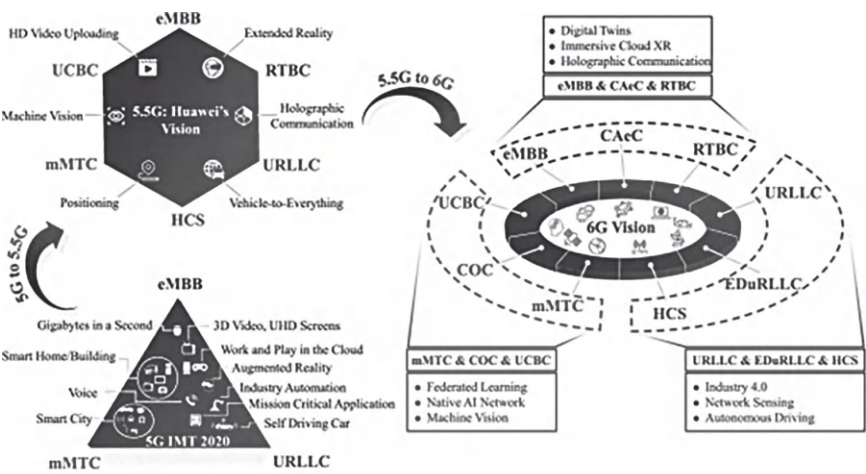


FIGURE 10.1 Evolution of 5G technology to 6G technology.

For shaping the landscape of 6G use cases in 2030, it is imperative to engage in multidisciplinary research and leverage a range of groundbreaking technologies. These technologies encompass spectrum exploration techniques, advancements in circuit and device technologies, and enhancements in learning, sensing, computing and networking functionalities. Notably, artificial intelligence (AI), with a particular emphasis on deep learning, emerges as a transformative conceptualization and improvement strategy for 6G wireless networks at different layer levels, including the application layer, medium access layer as well as physical layer [12, 13].

In particular, DL presents a novel approach for optimizing the overall communication algorithms, radio environment [14, 15], application integration as well as hardware infrastructure in a cogent manner [16, 17] in order to shape the 6G air interface. Recent developments in fields including task-specific communication [19], joint source-channel coding (JSCC) [18], and semantic communication [21] have been influenced by this strategy. Furthermore, machine learning (ML) offers a paradigm shift by enabling the automatic learning of algorithms for rapid optimization and high-performance for addressing resource allocation challenges in wireless communication system [22–25]. This includes the incorporation of domain-specific knowledge, such as theoretical tools and optimization models into the framework of deep learning (DL) to enhance the optimization of low latency and ultra-reliable communication networks [26]. ML methodologies have also been applied to tackle security, networking and communication issues within vehicular applications [27]. With the continuous development of hardware and software platforms, learning algorithms and models and techniques for wireless data collection we anticipate that AI will evolve into an intrinsic tool for shaping groundbreaking wireless technologies. This evolution will accelerate the processes of commercialization, standardization and design of 6G networks.

10.3 METHODOLOGY

In this section, we will introduce a range of distributed optimization algorithms focused on efficient communication for edge training. Subsequently, we will discuss promising wireless techniques that can facilitate the implementation of edge learning algorithms and models.

10.3.1 EDGE LEARNING ALGORITHMS AND MODELS

To customize a global model using decentralized data from various smart devices, the training approach for edge AI models typically involves minimizing an empirical or loss risk function. Distributed optimization targets the global loss function, denoted by L , which can be expressed as a reduction of R_d , where R_d represents the model parameters. For device K , the localized loss function is denoted by L_K and is given by:

$$L_K, L() := \sum_k w_k L_k(; D_k) D_k,$$

Here, D_k represents the local dataset for device K , S denotes the collection of participating edge nodes, and w_k are the weights assigned to each localized loss function, with $w_k \geq 0$ and $\sum w_k = 1$.

The complexity of the network architecture for edge training, the heterogeneity of local datasets D_k , the variability in device participation within SSS, the compute environment, dynamic communication, privacy issues, and possible adversarial threats make it essential to develop highly efficient and reliable distributed optimization algorithms.

10.3.2 FEDERATED EDGE AI LEARNING

It represents the collaborative framework of machine learning designed for training the unified model for statistical nature without any necessity to access the private raw data residing on edge devices. In this framework, a specialized edge server takes on the responsibility of collecting updates from local learning models and distributing updates for the global learning model [28]. “FL has gained adoption among several industry players, including its utilization in NVIDIA’s use in forecasting oxygen demands for COVID-19 patients, Apple’s QuickType keyboard for voice classification, Google’s Gboard mobile keyboard for features like next-word prediction and emoji recommendations, and WeBank’s implementation for spotting money laundering activities.”

10.3.3 DISTRIBUTED LEARNING

Distributed machine learning involves the acquisition of based on peer-to-peer interactions across the Network topologies and communication

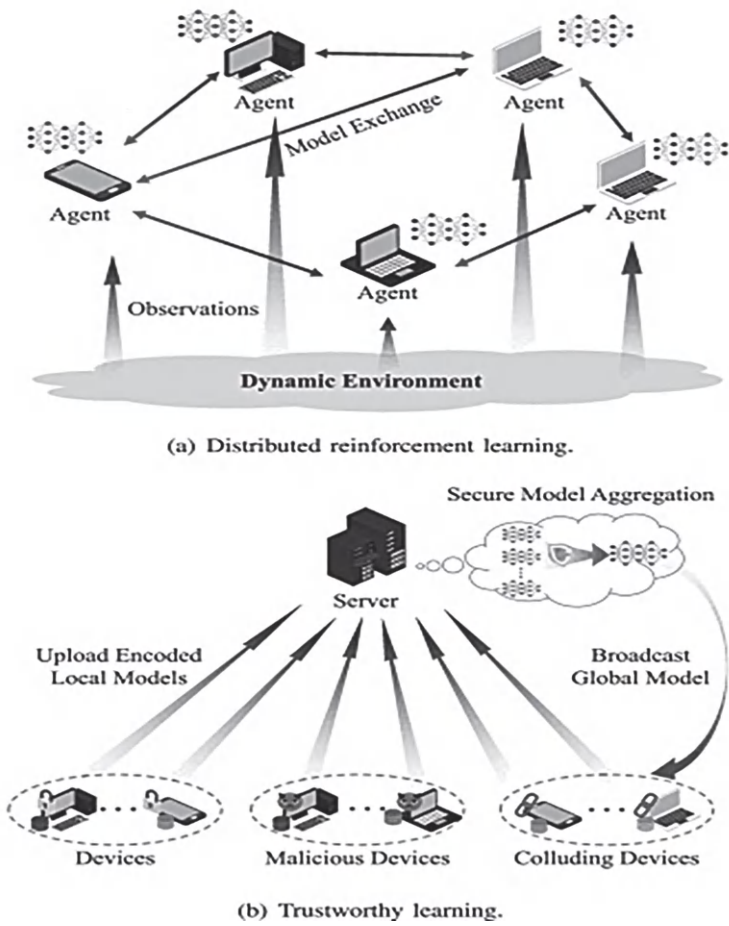


FIGURE 10.2 Edge Learning Algorithms and Models.

that underlies and naturally decentralized data structures, all without the need for a central authority. This approach holds significant promise for applications within autonomous industrial systems, including cooperative robotics, collaborative simultaneous mapping and localization, cooperative automated driving in advanced settings of manufacturing sector [29]. The architecture of distributed learning capitalizes on advantages such as improved efficiency of communication, enhanced data locality and computational scalability. Notably, swarm learning [30] offers a fully decentralized AI solution.

10.3.4 PARTITIONED MODEL LEARNING

Partitioned model learning facilitates a process of collaborative learning involving edge servers and edge devices by distributing the parameters of the proposed model that includes the edge nodes. In this setup, each edge node, which includes both edge servers and edge devices, is responsible solely for updating its designated θ_k within the overall model $\theta=[\theta_1, \theta_2, \dots, \theta_s]$ as described in Equation (1). This model partitioning architecture offers the advantages of heightened privacy and improved trade-offs between communication and computation. As a result, it finds particular relevance in DL scenarios with large model parameter sizes, as opposed to data partition-based training methods like federated learning (FL), which typically necessitate local updates of an entire replicated global model for each participation of edge device. The approaches and parameters [31] for the partitioned model for the edge learning involves training using the specific block of the parameters of model using the methods like descent coordination for the machine learning decomposable models [32] or alternating minimization for general DL models [33].

10.3.5 DECENTRALIZED REINFORCEMENT LEARNING

It serves as a versatile framework for making sequential decisions in dynamic environments through interactions with such environments, as illustrated in Figure 10.2(a). These scenarios are often conceptualized In Markov decision process for learning process and decision-making process [24]. Common RL algorithms encompass model-based methods, policy-based approaches (such as natural policy gradients), value-based techniques (like Q-learning), and the actor-critic method. Notably, utilization of parallel computing, an asynchronous method was introduced in [34] to address non convex RL, large scale problems.

Though, in contemporary intelligent applications like robotics and autonomous driving, it becomes imperative to incorporate the concept of reward-based multi-agent learning. In MARL, several agents cooperate in a shared environment to accomplish shared goals and optimize collective rewards, even though they may have distinct local action spaces.

1. **Reliable Learning:** For acquiring and deploying AI models in high-stakes applications such as autonomous driving, ensuring privacy, security, interpretability, accountability, resilience, and

equity throughout the edge learning processes is of paramount importance. However, the inherent diversity of decentralized datasets and massive scale edge systems poses unique challenges when it comes to designing dependable edge AI methodologies. Despite federated learning (FL) addressing local data confidentiality by keeping datasets on local devices, shared model updates can still lead to significant privacy risks, including threats like model inversion attacks. Moreover, the global model learned through FL may be vulnerable to collusion by malicious actors [35, 36], and edge devices themselves can potentially act as adversarial attackers, engaging in activities like data or model poisoning. These circumstances necessitate the development of robust privacy-preserving mechanisms and secure aggregation protocols.

2. **Wireless Approaches for Edge Training:** Given that the primary focus of edge AI communication has shifted from traditional data rates to learning performance, we aim to leverage the aligning the intrinsic algorithms and models of edge AI with the concepts and frameworks of wireless networks. In order to build a co-design principle that merges learning and communication for 6G networks and enable native AI capabilities within 6G, this method intends to increase the effectiveness of edge training within wireless networks. In order to accommodate the dynamic engagement of a large We will offer next-generation multiple access strategies (such big random access and AirComp) during the training process. Brand-new methods of using multiple antennas, like RIS and cell-free massive MIMO will also are used by us to allow the interchange of updates for high dimensional models. Additionally, in order to support various edge training models and topologies, we will investigate new network architectures.
3. **Wireless Computation Techniques:** In edge training scenarios, in order to update a global model, the activities often entail computing aggregation functions based on numerous local model changes. The procedure involves sending local updates from the edge devices, whether it's to achieve weighted average aggregation in federated learning, decentralized learning techniques like consensus aggregation or reliable aggregation techniques like

geometric median. The edge server then processes these updates to determine the appropriate aggregation function. However, dealing with a large number of edge devices that need to send their local model updates for global aggregation presents a substantial difficulty due to the constrained bandwidth and resources offered by wireless networks. In this situation, AirComp presents a revolutionary multiple access strategy intended to facilitate the aggregation of low-latency model data.

4. **Ultra-Large MIMO:** Utilizing extensive arrays of antennas represents a crucial wireless technology that enables energy and high spectral efficiency. This technology is in 6G, it is anticipated to scale up dramatically by an order of magnitude. Massive MIMO can now be used in practice even over greater frequency bands thanks to recent developments in hybrid beam forming, analog beam forming, and digital beam forming. Massive MIMO has been shown to have significant benefits for edge training systems. These benefits include achieving data rates and high model aggregation accuracy, as well as offering dependable communication for a large number of devices. By utilizing spatial diversity, massive MIMO can specifically give great computational accuracy for model aggregation [37]. It also enables ultra-fast model aggregation while simultaneously performing multiple computational tasks through spatial multiplexing [38].
5. **Adaptive Smart Surfaces:** In order to achieve the conduct magnitude alignment by changing the transmit signals, such as by employing methods like channel inversion, it is often necessary to perform the needed average functions through AirComp. This modification aims to lessen the interruptions brought on by changes in the channel conditions [39]. But because edge devices have limited resources and channels have non-uniform fading characteristics, the less-than-ideal signal propagation environment inexorably leads to a loss of misalignment and signal intensity, thereby affecting the quality of model aggregation. This, in turn, leads to a decline in the performance in the process of edge training. Additionally, when dealing with a large number of edge servers that has intermittent access for the edge devices, some of these devices may be situated in areas with poor

network coverage, making it challenging to detect device activity, especially for weak channel connections.

6. **Integrated Space-Air-Ground Networks:** The standard Space-Air-Ground Integrated Network (SAGIN) [40, 41] offers a unified platform for space-based information exchange that spans satellite networks, including vehicular communications, terrestrial communication, UAV communication, aerial networks and miniaturized satellites [42–44]. This integrated setup aims to provide seamless connectivity for different types of configurations for edge g-training. Performing edge learning within a Vehicle to everything network is of utmost importance, particularly for enabling applications in autonomous driving with real-time requirements [45]. In this context, the timely and dependable aggregation of local model updates among neighboring vehicles via vehicle-to-vehicle communications becomes crucial.

10.4 RESULTS AND DISCUSSION

In this section, we introduce strategies for achieving efficient communication in edge inference tasks while ensuring low latency and reliability. To categorize federated learning (FL) Yang et al. [46] proposed differentiating between horizontal FL based on the characteristics of dataset distribution.

10.4.1 HORIZONTAL EDGE INFERENCE

10.4.1.1 INFERENCE DISTRIBUTED ACROSS EDGE DEVICES

Substantial efforts have been invested in the development of TinyML, involving techniques like DL model compression and neural network architecture optimization. These endeavors aim to facilitate energy-efficient and low latency model inference on individual devices with restricted storage and computational capabilities [48]. Nevertheless, owing to the limited storage capacity of edge devices, achieving inference computations on a single device becomes exceedingly challenging, particularly for applications like mobile navigation that entail processing vast datasets containing extensive map information [49].

10.4.1.2 COLLABORATIVE INFERENCE VIA EDGE SERVERS

DL models with extensive parameters can deliver highly accurate intelligent services. Nevertheless, deploying these large AI models directly onto IoT devices poses significant challenges because of the severe constraints on energy resources, storage and onboard computation. An alternative promising solution emerges by DL model deployment and operation on edge servers. Although the wireless bandwidth between edge servers and edge devices is constrained emerges as a critical bottleneck [50, 51].

10.4.2 VERTICAL EDGE INFERENCE

In the upcoming sections, we will first introduce effective strategies to enable vertical edge inference that is communication-efficient in these two cases. Next, we offer task-oriented communication, a novel overarching design approach adapted for resource-constrained vertical edge inference.

10.4.2.1 EDGE DEVICE-SERVER COLLABORATIVE INFERENCE

It has a low latency, but its accuracy is restricted by its computing power and bandwidth limitations. However, edge server cooperative inference may result in data leakage and lengthy communication delays while still achieving high accuracy with DL models. For applications that require speed and secrecy, this strategy might not be appropriate. Edge device-server co-inference, as an alternative to horizontal edge inference, shows promise in reducing communication cost while retaining high accuracy and privacy in DNN model inference, enabling the delivery of ubiquitous AI services across many application scenarios.

10.4.3 TASK-ORIENTED COMMUNICATION

There is a built-in trade-off between communication and computation in resource-constrained inference of a vertical edge, as displayed in [52]. DL-based inference's data amplification problem, where the dimension of intermediate features may be bigger than the amount of the input data, is what essentially motivates this trade-off. Therefore, using a neural network

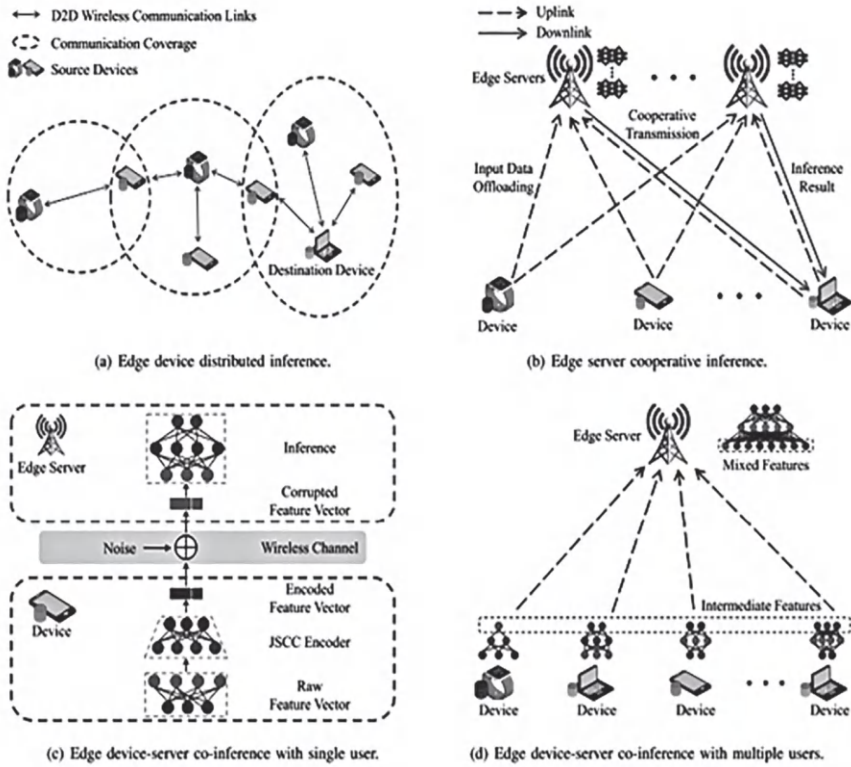


FIGURE 10.3 Edge AI Architecture and Enabling Technologies.

with only a few layers on an edge device could provide output features that are larger than the input data, which would increase communication overhead.

10.4.4 URLLC

Very low packet lengths may be required when transmitting extracted output characteristics from edge devices [53], which increase the likelihood of decoding errors in the finite block length regime. Additionally, for mission-critical applications, the edge devices must get inference results obtained from edge devices with sufficient reliability and latency. When system dynamics like task arrivals at the wireless channel and network layer variations at the physical layer are taken into account, cross-layer

optimization becomes essential to reducing co inference of server and end to end delay in the edged devices.

Edge Inference on the Vertical We look at two different vertical edge inference scenarios, one with a single edge device and the other with numerous edge devices, as shown in Figure 10.3. We will first discuss effective methods for communication-conscious vertical edge inference in these two situations in the following sections. Then, we'll propose task-oriented communication, a brand-new overarching design paradigm for resource-limited vertical edge inference.

10.4.5 *EDGE AI ARCHITECTURE*

In this section, we present a revolutionary mobile network architecture specifically designed for cutting-edge artificial intelligence (AI) systems. This architecture builds on existing wireless network infrastructures and service-driven resource allocations as part of an end-to-end design. It encompasses applications, network administration, network functionality, data governance, and network infrastructure.

10.4.5.1 *COMPREHENSIVE ARCHITECTURE*

With each new generation of mobile networks, there has been a need to introduce fresh services and capabilities at the architectural level to address increasingly stringent demands. Originally, mobile networks were designed primarily for voice services. However, over time, Mobile network deployment and architecture have developed into a centralized and hierarchical paradigm that mostly reflects the voice and packet features of mobile internet services. 6G is prepared to challenge and modify these existing paradigms in order to realize the goal of “connected intelligence,” ushering in a fresh design that is in line with new requirements for fully integrating sensing, computation, AI, communication, and at the network edge. This transformation will be underpinned by evolutionary and revolutionary enabling technologies.

10.4.5.2 *DATA GOVERNANCE*

We believe that in future 6G networks, data will need to be stored, processed, gathered, and used mostly at the network edge due to the

projected considerable energy consumption and the urgent privacy and security issues. The necessity for an effective and uniform data governance system at the architectural level is compelling given that data and AI applications in 6G are anticipated to be more varied than ever. This framework must encompass aspects beyond conventional data storage and collection, taking into account legal implications, knowledge management, data sovereignty, quality and data availability.

10.4.5.3 INDEPENDENT DATA PLANE

In order to support “a new network data analytics function (NWDAF) is now part of the core network’s AI-based network automation for 5G. The NWDAF enhances the quality of all user services by optimizing a variety of network operations, such as AI-based mobility management. Its main goal is to collect and examine data from other 5G network components in order to train AI models and apply AI-based inference to automate and scale network optimization.”

Various Player Roles Data stewards, data owners, data providers, and data clients are just a few of the roles that are involved in the data governance ecosystem. The same or various business organizations, including individual users, may play these functions. As a result, a typical scenario involving numerous stakeholders is data governance.

10.4.5.4 DEEPLY INTEGRATED COMMUNICATION AND COMPUTATION

In 5G, exceptional performance was achieved by incorporating AI capabilities into the Radio Access Network (RAN) [54, 55]. For example, machine learning methods were used to mitigating interference and optimizing scheduling of radio resource. This integration of AI into 5G networks is often referred to as “AI for networks.” However, the goals of edge AI encompass not only “AI for networks” but also “networks for AI.”

10.4.5.5 ORCHESTRATION AND MANAGEMENT OF EDGE AI

Edge AI encompasses a wide array of machine learning algorithms, models, intricate collaborations and network infrastructure among intelligence, computation and communication. Establishing a framework for managing

and orchestrating edge AI becomes a crucial aspect of architectural design. This framework must facilitate the seamless deployment and integration of AI services, particularly those from third-party providers. Achieving this entails optimizing, maintaining, deploying and planning decentralized machine learning models, algorithms, and edge network functions and infrastructures.

10.4.6 EDGE AI RESOURCE ALLOCATION

We lay forth the technical requirements for developing robust and scalable edge AI systems. Resources must be allocated in accordance with the needs of edge AI systems to ensure accurate intelligence distillation within the edge network while functioning at ultra-low power and minimal latency.

1. **Precision:** The creation of global iterates, represented as $[t]$, where ‘t’ stands for the iteration index, is the central component of the edge training process. With a nearly insignificant optimality gap for the given problem, this method seeks for minimizing the loss function of empirical nature at the quick convergence rate (1). Characterizing the convergence behaviors of these global iterates becomes crucial when creating effective resource allocation plans for edge training systems. These behaviors typically depend on choices made for underlying algorithms, function landscapes, propagation conditions, network topologies, aggregation techniques, local updates and scheduled devices. In the context of edge training systems employing AirComp, it’s crucial to acknowledge that aggregation errors for the global model that stems for wireless channel noise and fading can lead to degradation in learning performance [56].
2. **Latency:** In the context of edge training, latency encompasses both communication latency and computation latency. The processing power of the edge servers and devices, together with the dimensions of the underlying models and data sets, all have a significant impact on computation latency. Contrarily, communication latency is the ratio of the entire number of learning rounds needed to train the global model until convergence to the cumulative time it takes for one round of communication.

The training round, which includes the scheduling of edge devices within a typical communication latency, wireless communication protocols, available bandwidth, power limitations, and wireless channel conditions are the main factors affecting both uplink and downlink broadcasts for the update of model. Using techniques such as the huge deviation theory and approximation of Saddle Point, Li et al. [58] approximated the distribution. Delays for federated learning over arbitrary fading channels. Additionally, studies have shed light on the trading off between latency per-round and convergence speed, showing that an increase in scheduled devices can hasten convergence rates while lengthening the time required for local updates to be uploaded in each iteration due to the limited availability of radio resources [59].

3. **Energy Usage:** In the context of edge training, energy consumption is composed of both computational and communication processes. For instance, AlphaGo, as an example, requires approximately 280 GPUs and incurs an electricity cost of around \$3,000 per game [60]. Therefore, it is imperative to develop technologies for cutting-edge training that are energy-efficient and aligned with goals for carbon neutrality. The size of the training models, the algorithms used to train the models, wireless transmission techniques, hardware components (“for example, cutting-edge technologies like scaled SiGe bipolar technology [61]”), and Hardware and architecture of edge computing are all important considerations in the design of such systems. In an effort to reduce the amount of energy used for energy production, both the computational energy used for updates of local model and the used energy for transmitting these model updates are concurrently optimized, as demonstrated in [62]. This optimization process takes into account factors like learning accuracy and latency constraints within the context of wireless federated learning.
4. **Reliability:** Reliability stands as a central motivator in the advancement of AI technologies of next generation. Specifically, the AI algorithms and models under development must possess qualities such as privacy preservation, resilience against adversarial attacks, robustness, fairness, optimality, and interpretability [63]. In the context of edge training, privacy considerations

predominantly revolve around how raw data and intermediate features are offloaded or encoded. An effective and direct approach to preserving user privacy in federated learning is by retaining datasets on the individual devices.

Additionally, the presence of wireless channel noise introduces a degree of noise into the model aggregation process through AirComp. This inherent feature of AirComp serves as a system that protects privacy, enhancing the differential features of privacy belongs to every edge device. Furthermore, an Adaptive technique of power control has been created in this over-the-air federated learning system [64]. This method manages the levels of differential privacy while preventing any degradation in learning performance.

Resource Management Aligned with Service Needs: Edge AI systems must seamlessly blend different communication approach and architecture for wireless network by merging communication and computation. This transformation leads to the creation of a highly intricate and ever-changing network environment, necessitating the development of innovative technologies and solutions. The presence of diverse use cases, such as smart healthcare, Industrial IoT and autonomous driving each with varying demands regarding trustworthiness, energy efficiency, latency and accuracy only further amplifies the complexity associated with resource allocation in the system of edge AI in 6G technology.

Furthermore, the intricacy of edge servers and base stations results in significant energy consumption, presenting formidable challenges in achieving optimal energy efficiency.

10.4.7 APPLICATIONS

We explore application scenarios empowered by edge AI, which serve as a source of inspiration for the development of fresh communication algorithms, optimization algorithms for resource allocation, and innovative data processing methods.

10.4.7.1 AUTONOMOUS DRIVING

Autonomous driving involves vehicles that can operate without human intervention. This concept integrates a range of cutting-edge technologies,

including advanced sensor systems, future vehicular networks, electric automobiles, and next-generation AI technology. Driverless vehicles have the potential to greatly energy efficiency, collision avoidance, logistics and travel efficiency, passenger comfort and enhanced safety. Edge AI plays a critical role in enabling real time data analysis, intelligent networking, ultra-low latency communication, and robust security for intelligent vehicles, [65]. In a recent study, a comprehensive DL framework was introduced to facilitate vehicle communication that is extremely dependable and responsive, utilizing tools from information theory and cross-layer optimization strategies, among another domain-specific knowledge. A federated learning (FL) approach was created [66] to learn the distribution of queue lengths' tails in order to reduce the queuing latency for cars. High mobility and different network structures provide difficulties in vehicle networks, although DL shows promise in managing network traffic and dynamic resource allocation. Edge AI approaches, in particular distributed reinforcement learning (RL) [67, 68], decentralized graph neural networks (GNN), and distributed deep neural networks (DNN) with binarized output layers [69], show the ability to learn and carry out distributed resource allocation policies automatically and in real-time.

10.4.7.2 INTERNET OF THINGS

The concept of artificial intelligence of things (AIoT) harnesses the power of AI technologies within IoT infrastructures to enhance interactions between humans and machines, facilitating multi-agent collaborations and communication. AIoT transcends traditional communication methods for data, video and audio transfer. It introduces semantic communication [70], a means of exchanging meaningful data shared between agents. There are three levels of communication according to Shannon and Weaver: transmittal rate focused on accurate symbol transmission), semantic level (aimed at precise conveyance of intended meanings, and effectiveness level (concentrating on the successful execution of desired actions) [71]. Semantic communication plays a pivotal role in optimizing communication efficiency by transmitting only pertinent information, extracted for semantic information semantic mistake as the performance metric for delivery jobs. For enabling semantic communication of low latency in IoT network, a distributed edge DL technique was recently proposed [70].

In this method, wireless fading channels are taken into account while compressing quantizing DNN-based receivers at the edge server and DNN-based transmitters at the edge IoT devices are jointly optimized.

10.4.7.3 INTELLIGENT HEALTHCARE

The concept of intelligent healthcare works to create a unified platform that encourages intelligent health monitoring, efficient and individualized healthcare, promoting the growth of precision medicine through collaboration between various stakeholders, such as doctors, patients, hospitals, and research institutions. Promoting the growth of precision medicine through collaboration between various stakeholders, such as doctors, patients, hospitals, and research institutions. This objective is accomplished through the use of cutting-edge technologies including DL [72, 73], the tactile internet, the internet of things (IoT), edge AI, and wireless connectivity. The Internet of Medical Things (IoMT) and precision medicine in particular have shown remarkable potential for improvement in terms of dependability, accuracy, scalability, privacy, and security [74], covering areas like medical imaging, drug development, and the management of chronic conditions and diseases [75]. To illustrate, Kaissis et al. [57] introduced a federated learning strategy for medical imaging that protects privacy, safeguards against any dataset or learning algorithm attacks, and ensures accuracy. A decentralized and private solution for clinical disease identification (such as COVID-19, tuberculosis, and leukemia) is also provided by recent advancements in swarm learning [30]. It was accomplished using edge computing blockchain technology and 6G feature approaches for providing private and safe framework for decentralized learning while preserving the local level anonymity of medical data.

10.5 CONCLUSION

Incorporating intelligent capabilities that are energy-efficient, low-latency, dependable, and secure at the network edge is an unavoidable development that will significantly alter both academic and industrial research. Edge AI functions as a decentralized neural network, bringing intelligent connectivity to 6G and seamless interactions and enabling technologies between the physical world, human activities, and the digital world. The difficulties

in creating complete edge AI ecosystems are diverse and span multiple fields, including wireless communications, machine learning, operations research, specific application domains, regulatory compliance, and ethical considerations. To design scalable as well as reliable edge AI systems, we have examined key 6G technologies for communication, efficient resource management methodologies, and comprehensive network designs in this study. We have also explored aspects related to standardization, platforms, and potential applications, all with the intention of advancing the development and commercialization of edge AI solutions. Our aspiration is for the present research study to serve as a valuable point of reference and a practical guide for future opportunities in the realm of edge AI.

KEYWORDS

- **intelligent healthcare**
- **edge AI**
- **precision medicine**
- **internet of medical things (IoMT)**
- **federated learning**
- **6G technologies**

REFERENCES

1. National Science Foundation. (2021). *Resilient and Intelligent NextG Systems (RINGS)*. Available: <https://www.nsf.gov/pubs/2021/nsf21581/nsf21581.pdf> (accessed on 17 October 2024).
2. Hexa-X. (2021). *Expanded 6G Vision Use Cases and Societal Values*. Available: https://hexa-x.eu/wp-content/uploads/2021/05/Hexa-X_D1.2.pdf (accessed on 17 October 2024).
3. IMT-2030 (6G) Promotion Group. (2021). *6G Vision and Candidate Technologies*. Available: <http://www.caict.ac.cn/english/news/202106/P020210608349616163475.pdf> (accessed on 17 October 2024).
4. Network 2030. (2019). *A Blueprint of Technology Applications and Market Drivers Towards the Year 2030 and Beyond*.
5. Andrews, J. G., et al. (2014). What will 5G be? *IEEE Journal on Selected Areas in Communications*, 32(6), 1065–1082.

6. Shafi, M., et al. (2017). 5G: A tutorial overview of standards, trials, challenges, deployment, and practice. *IEEE Journal on Selected Areas in Communications*, 35(6), 1201–1221.
7. Huawei. (2020). *Huawei 5.5G*. Available: <https://www.huawei.com/en/news/2020/11/mbbf-shanghai-huawei-david-wang-5dot5g> (accessed on 17 October 2024).
8. Tataria, H., Shafi, M., Molisch, A. F., Dohler, M., Sjöland, H., & Tufvesson, F. (2021). 6G wireless systems: Vision, requirements, challenges, insights, and opportunities. *Proceedings of the IEEE*, 109(7), 1–34.
9. Maier, M., Ebrahimzadeh, A., Rostami, S., & Benicche, A. (2020). The internet of no things: Making the internet disappear and ‘see the invisible.’ *IEEE Communications Magazine*, 58(11), 76–82.
10. You, X. H., et al. (2021). Towards 6G wireless communication networks: Vision, enabling technologies, and new paradigm shifts. *Science China Information Sciences*, 64(1), 1–74.
11. Saad, W., Bennis, M., & Chen, M. (2020). A vision of 6G wireless systems: Applications, trends, technologies, and open research problems. *IEEE Network*, 34(3), 134–142.
12. Letaief, K. B., Chen, W., Shi, Y., Zhang, J., & Zhang, Y. J. A. (2019). The roadmap to 6G: AI empowered wireless networks. *IEEE Communications Magazine*, 57(8), 84–90.
13. Ali, S., Saad, W., Steinbach, D., Ahmad, I., & Huusko, J. (2020). White paper on machine learning in wireless communication networks. *6G Research Vision*, 7, 1–36.
14. Wang, J., et al. (2021). Interplay between RIS and AI in wireless communications: Fundamentals, architectures, applications, and open research problems. *IEEE Journal on Selected Areas in Communications*, 39(8), 2271–2288.
15. Kim, H., Jiang, Y., Rana, R. B., Kannan, S., Oh, S., & Viswanath, P. (2018). Communication algorithms via deep learning. In *Proceedings of the International Conference on Learning Representations (ICLR)* (pp. 1–19).
16. O’Shea, T., & Hoydis, J. (2017). An introduction to deep learning for the physical layer. *IEEE Transactions on Cognitive Communications and Networking*, 3(4), 563–575.
17. Hoydis, J., Aoudia, F. A., Valcarce, A., & Viswanathan, H. (2021). Toward a 6G AI-native air interface. *IEEE Communications Magazine*, 59(5), 76–81.
18. Saidutta, Y. M., Abdi, A., & Fekri, F. (2021). Joint source-channel coding over additive noise analog channels using a mixture of variational autoencoders. *IEEE Journal on Selected Areas in Communications*, 39(7), 2000–2013.
19. Strinati, E. C., & Barbarossa, S. (2021). 6G networks: Beyond Shannon towards semantic and goal-oriented communications. *Computer Networks*, 190, 108118.
20. Shao, J., Mao, Y., & Zhang, J. (2021). *Learning Task-Oriented Communication for Edge Inference: An Information Bottleneck Approach*. *arXiv:2102.04170*.
21. Xie, H., Qin, Z., Li, G. Y., & Juang, B. H. (2021). Deep learning enabled semantic communication systems. *IEEE Transactions on Signal Processing*, 69, 2663–2675.
22. Shen, Y., Shi, Y., Zhang, J., & Letaief, K. B. (2020). LORM: Learning to optimize for resource management in wireless networks with few training samples. *IEEE Transactions on Wireless Communications*, 19(1), 665–679.

23. Shen, Y., Shi, Y., Zhang, J., & Letaief, K. B. (2021). Graph neural networks for scalable radio resource management: Architecture design and theoretical analysis. *IEEE Journal on Selected Areas in Communications*, 39(1), 101–115.
24. Liang, L., Ye, H., Yu, G., & Li, G. Y. (2020). Deep-learning-based wireless resource allocation with application to vehicular networks. *Proceedings of the IEEE*, 108(2), 341–356.
25. Jiang, T., Cheng, H. V., & Yu, W. (2021). Learning to reflect and to beamform for intelligent reflecting surface with implicit channel estimation. *IEEE Journal on Selected Areas in Communications*, 39(7), 1931–1945.
26. She, C., et al. (2021). A tutorial on ultrareliable and low-latency communications in 6G: Integrating domain knowledge into deep learning. *Proceedings of the IEEE*, 109(3), 204–246.
27. Tang, F., Kawamoto, Y., Kato, N., & Liu, J. (2020). Future intelligent and secure vehicular network toward 6G: Machine-learning approaches. *Proceedings of the IEEE*, 108(2), 292–307.
28. Kairouz, E. B. P., & McMahan, H. B. (2021). Advances and open problems in federated learning. *Foundations and Trends in Machine Learning*, 14(1), 1–210.
29. Savazzi, S., Nicoli, M., Bennis, M., Kianoush, S., & Barbieri, L. (2021). Opportunities of federated learning in connected, cooperative, and automated industrial systems. *IEEE Communications Magazine*, 59(2), 16–21.
30. Warnat-Herresthal, S., et al. (2021). Swarm learning for decentralized and confidential clinical machine learning. *Nature*, 594, 265–270.
31. Li, M., Andersen, D. G., Smola, A. J., & Yu, K. (2014). Communication efficient distributed machine learning with the parameter server. In *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)* (Vol. 27, pp. 19–27).
32. Wright, S. J. (2015). Coordinate descent algorithms. *Mathematical Programming*, 151(1), 3–34.
33. Choromanska, A., et al. (2019). Beyond backprop: Online alternating minimization with auxiliary variables. In *Proceedings of the International Conference on Machine Learning (ICML)* (Vol. 97, pp. 1193–1202).
34. Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A. (2017). Deep reinforcement learning: A brief survey. *IEEE Signal Processing Magazine*, 34(6), 26–38.
35. Mnih, V., et al. (2016). Asynchronous methods for deep reinforcement learning. In *Proceedings of the International Conference on Machine Learning (ICML)* (pp. 1928–1937).
36. Dong, Y., Cheng, J., Hossain, M. J., & Leung, V. C. M. (2019). Secure distributed on-device learning networks with Byzantine adversaries. *IEEE Network*, 33(6), 180–187.
37. Damaskinos, G., El Mhamdi, E. M., Guerraoui, R., Guirguis, A. H. A., & Rouault, S. L. A. (2019). AGGREGATHOR: Byzantine machine learning via robust gradient aggregation. In *Proceedings of the Conference on Systems and Machine Learning (SysML)* (pp. 1–19).
38. Zhai, X., Chen, X., Xu, J., & Ng, D. W. K. (2021). Hybrid beamforming for massive MIMO over-the-air computation. *IEEE Transactions on Communications*, 69(4), 2737–2751.

39. Zhu, G., & Huang, K. (2019). MIMO over-the-air computation for high-mobility multimodal sensing. *IEEE Internet of Things Journal*, 6(4), 6089–6103.
40. Zhu, G., Xu, J., Huang, K., & Cui, S. (2020). *Over-the-Air Computing for Wireless Data Aggregation in Massive IoT*. *arXiv:2009.02181*.
41. Liu, J., Shi, Y., Fadlullah, Z. M., & Kato, N. (2018). Space-air-ground integrated network: A survey. *IEEE Communications Surveys & Tutorials*, 20(4), 2714–2741.
42. Zhu, X., Jiang, C., Yin, L., Kuang, L., Ge, N., & Lu, J. (2018). Cooperative multigroup multicast transmission in integrated terrestrial-satellite networks. *IEEE Journal on Selected Areas in Communications*, 36(5), 981–992.
43. Saeed, N., Elzanaty, A., Almorad, H., Dahrouj, H., Al-Naffouri, T. Y., & Alouini, M. S. (2020). CubeSat communications: Recent advances and future challenges. *IEEE Communications Surveys & Tutorials*, 22(3), 1839–1862.
44. Zeng, Y., Wu, Q., & Zhang, R. (2019). Accessing from the sky: A tutorial on UAV communications for 5G and beyond. *Proceedings of the IEEE*, 107(12), 2327–2375.
45. Zhou, H., Xu, W., Chen, J., & Wang, W. (2020). Evolutionary V2X technologies toward the internet of vehicles: Challenges and opportunities. *Proceedings of the IEEE*, 108(2), 308–323.
46. Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology*, 10(2), 1–19.
47. Hosseinalipour, S., Brinton, C. G., Aggarwal, V., Dai, H., & Chiang, M. (2020). From federated to fog learning: Distributed machine learning over heterogeneous wireless networks. *IEEE Communications Magazine*, 58(12), 41–47.
48. Cheng, Y., Wang, D., Zhou, P., & Zhang, T. (2018). Model compression and acceleration for deep neural networks: The principles, progress, and challenges. *IEEE Signal Processing Magazine*, 35(1), 126–136.
49. Li, S., Yu, Q., Maddah-Ali, M. A., & Avestimehr, A. S. (2017). A scalable framework for wireless distributed computing. *IEEE/ACM Transactions on Networking*, 25(5), 2643–2654.
50. Yang, K., Shi, Y., Yu, W., & Ding, Z. (2020). Energy-efficient processing and robust wireless cooperative transmission for edge inference. *IEEE Internet of Things Journal*, 7(10), 9456–9470.
51. Hua, S., Zhou, Y., Yang, K., Shi, Y., & Wang, K. (2021). Reconfigurable intelligent surface for green edge inference. *IEEE Transactions on Green Communications and Networking*, 5(2), 964–979.
52. Shao, J., & Zhang, J. (2020). Communication-computation trade-off in resource-constrained edge inference. *IEEE Communications Magazine*, 58(12), 20–26.
53. Polyanskiy, Y., Poor, H. V., & Verdú, S. (2010). Channel coding rate in the finite blocklength regime. *IEEE Transactions on Information Theory*, 56(5), 2307–2359.
54. Mao, Q., Hu, F., & Hao, Q. (2018). Deep learning for intelligent wireless networks: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, 20(4), 2595–2621.
55. Luong, N. C., et al. (2019). Applications of deep reinforcement learning in communications and networking: A survey. *IEEE Communications Surveys & Tutorials*, 21(4), 3133–3174.

56. Amiri, M. M., & Gündüz, D. (2020). Machine learning at the wireless edge: Distributed stochastic gradient descent over-the-air. *IEEE Transactions on Signal Processing*, 68, 2155–2169.
57. Kaissis, G. A., Makowski, M. R., Rückert, D., & Braren, R. F. (2020). Secure, privacy-preserving and federated machine learning in medical imaging. *Nature Machine Intelligence*, 2(6), 305–311.
58. Li, L., et al. (2021). *Delay Analysis of Wireless Federated Learning Based on Saddle Point Approximation and Large Deviation Theory*. arXiv:2103.16994.
59. Shi, W., Zhou, S., Niu, Z., Jiang, M., & Geng, L. (2021). Joint device scheduling and resource allocation for latency constrained wireless federated learning. *IEEE Transactions on Wireless Communications*, 20(1), 453–467.
60. de Lima, T. F., et al. (2019). Machine learning with neuromorphic photonics. *Journal of Lightwave Technology*, 37(5), 1515–1534.
61. Alioto, M., De, V., & Marongiu, A. (2018). Energy-quality scalable integrated circuits and systems: Continuing energy scaling in the twilight of Moore's law. *IEEE Transactions on Emerging Topics in Circuits and Systems*, 8(4), 653–678.
62. Yang, Z., Chen, M., Saad, W., Hong, C. S., & Shikh-Bahaei, M. (2021). Energy efficient federated learning over wireless communication networks. *IEEE Transactions on Wireless Communications*, 20(3), 1935–1949.
63. Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., & Müller, K. R. (2021). Explaining deep neural networks and beyond: A review of methods and applications. *Proceedings of the IEEE*, 109(3), 247–278.
64. Liu, D., & Simeone, O. (2021). Privacy for free: Wireless federated learning via uncoded transmission with adaptive power control. *IEEE Journal on Selected Areas in Communications*, 39(1), 170–185.
65. Zhang, J., & Letaief, K. B. (2020). Mobile edge intelligence and computing for the internet of vehicles. *Proceedings of the IEEE*, 108(2), 246–261.
66. Samarakoon, S., Bennis, M., Saad, W., & Debbah, M. (2020). Distributed federated learning for ultra-reliable low-latency vehicular communications. *IEEE Transactions on Communications*, 68(2), 1146–1159.
67. Nasir, Y. S., & Guo, D. (2019). Multi-agent deep reinforcement learning for dynamic power allocation in wireless networks. *IEEE Journal on Selected Areas in Communications*, 37(10), 2239–2250.
68. Hu, Y., Chen, M., Saad, W., Poor, H. V., & Cui, S. (2021). Distributed multi-agent meta learning for trajectory design in wireless drone networks. *IEEE Journal on Selected Areas in Communications*, 39(10), 3177–3192.
69. Lee, H., Lee, S. H., & Quek, T. Q. S. (2019). Deep learning for distributed optimization: Applications to wireless resource management. *IEEE Journal on Selected Areas in Communications*, 37(10), 2251–2266.
70. Xie, H., & Qin, Z. (2021). A lite distributed semantic communication system for Internet of Things. *IEEE Journal on Selected Areas in Communications*, 39(1), 142–153.
71. Shannon, C. E., & Weaver, W. (1949). *The Mathematical Theory of Communication* (Vol. 96). Urbana, IL, USA: University of Illinois Press.
72. Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—Big data, machine learning, and clinical medicine. *New England Journal of Medicine*, 375(13), 1216.

73. Zhou, S. K., et al. (2021). A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. *Proceedings of the IEEE*, 109(5), 820–838.
74. Sun, Y., Lo, F., & Lo, B. (2019). Security and privacy for the internet of medical things enabled healthcare systems: A survey. *IEEE Access*, 7, 183339–183355.
75. Subramanian, M., et al. (2020). Precision medicine in the era of artificial intelligence: Implications in chronic disease management. *Journal of Translational Medicine*, 18(1), 1–12.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

PART IV

Computational Intelligence: IoT Systems



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

CHAPTER 11

AI Products in E-Commerce and Commercial Bank

M. NITHYA,¹ PIYUSH MEHTA,² VEENA PRASAD VEMURI,³
MEETA JOSHI,⁴ and TRIPTI TIWARI⁵

¹Computer Science and Engineering (5 Year Integrated), Sri Sairam Engineering College, Chennai, Tamil Nadu, India

²United World School of Business, Karnavati University Uvarsad, Gandhinagar, Gujarat, India

³I/C Principal, Commerce/Management, NKES College of Arts Commerce and Science, Mumbai University, India

⁴Faculty of Management Studies, Marwadi University, Gujarat, India

⁵Department of Management Studies, Bharati Vidyapeeth (Deemed to be University), Institute of Management and Research, Delhi, India

ABSTRACT

As more people enter the financial industry, commercial banks are adopting artificial intelligence more widely in finance. This study looked into how commercial banks were using AI applications to improve their financial processes. The investigation into artificial intelligence (AI) in online trading is also summarized in this paper, and this study stream may benefit from understanding how information systems function. AI, also known as intelligent machines on occasion, is a replication of human intelligence. Robots have an inbuilt form of cognition that is different from human comprehension. The primary subjects of this study include

Advances in AI for Cloud, Edge, and Mobile Computing Applications.

Shrikaant Kulkarni and P. William (Eds.)

© 2025 Apple Academic Press, Inc. Co-published with CRC Press (Taylor & Francis)

the potential utilization of AI in the financial sector, how technology has affected finance, how it has transformed human work, and so on. We are all aware that mistakes are made by people, but as the world changes and technologies follow suit, there aren't enough highly trained professionals to oversee the automation. Considering the industry's significant and quick evolution, the public's opinion of banking in India is changing as a result of the present applications of AI in banking, which are highlighted by this research. Secondary data were used in the study on how AI is affecting commercial bank operations. According to study, AI is employed in the banking industry for a number of tasks, including Chatbot, risk evaluation, money laundering prevention, detection of fraud, and algorithms. The implementation of AI in banking is always changing, but three goals that have a consequence on daily operations are the focus of the bank's overall AI application trend. The implications of artificial intelligence on banking operations, particularly those on client satisfaction, bank performance and costs, risk management, and compliance.

11.1 INTRODUCTION

The creation of intelligent machines is the focus of artificial intelligence, one of computer science's most significant and well-known subfields today. These smart machines are made to be able to carry out activities that call for human intelligence. Additionally, we claim that artificial intelligence as a concept is utilized in homes to projects like online studying and shopping [1, 2]. It is AI that enables businesses to conduct real-time investigations, increase productivity, and contribute to national safety and security [4, 5]. It can be said that the e-commerce industry, in the present period, makes the best use relating to artificial intelligence by accumulating a sizable clientele, grasping customer need, carrying out real-time analysis, offering thorough solutions to problems, and much more. The e-commerce sector uses a variety of AI solutions, including Chatbot, customer relationship management, enterprise resource planning (ER), product content management (PCM) and many others. Among these, AI is the most recent [6–10]. Due to its capacity to “correctly interpret external data, learn from such data, and use those learning to achieve specific goals and tasks through flexible adaptation,” it is revolutionizing e-commerce.

Over the past 60 years that artificial intelligence has been developed, various ground-breaking achievements have been made and our economy and society have benefited greatly from the organization's discoveries. Artificial intelligence was all but assumed after Lee Se-dol, the finest human professional go player, was defeated by Google Alpha G in 2016. The State Council publicly established federal strategic aim for the advancement of the latest AI era by announcing the formulation of a development plan for it in 2017 [11–16]. Artificial intelligence is now present in e-commerce, having already shown useful in industries like marketing, financial services, health care, and education.

The sophistication of smart algorithm and artificial intelligence technologies has grown, and they now have the ability to transform mainstream banking processes and force the transition to faster, flexible, and customer-focused digital banking [17–22]. AI has the ability to change all parts of banking, making it possible for a smarter, speedier, and more efficient way for us to trade and spend And more client-centric.

According to author, banks may gain from digitization and compete with FinTech companies by combining cognitive technology and AI. In fact, 32% of companies that offer financial services now use AI-based technology, including recognition of speech and analytics for prediction, among other things [23–27]. By lowering fraud and raising compliance, the application of sophisticated data analysis by artificial intelligence will change banking in the future. AI programs execute activities preventing money laundering just a few seconds instead of hours or days. Thanks to artificial intelligence (AI), banks can manage enormous amounts of data at breakneck rates with the aim to extract valuable information from it. Using tools like AI robots, online financial planners, and fingerprint systems that detect fraud, a broader customer base may become access to better services [28–30]. All of these lead to increased revenue, decreased costs, and increased profitability.

- **Evolution:** When Alan Turing published an article on the possibility of autonomous computers in the 1950s, the history of AI officially began, even though we have only lately witnessed AI in action. Although the phrase “artificial intelligence” had only just been coined, neither the idea nor the method had ever been used in practice before the late 1990s [31–33]. The rapid development of artificial intelligence only proceeded to pick up after 2011, when

big tech companies like Instagram, IBM, Microsoft, and Google began using AI and machine learning for commercial purposes.

- **Adoption:** At the moment, a variety of techniques are used in AI implementations, such as algorithm monitoring, recognizing faces, and recognition of optical characters. A number of commercial fields, including genetics, accounting, insurance, the internet, transportation, and the aviation sector, are currently using artificial intelligence [34–36]. The potential for deep neural networks, speech recognition, emotion detection, and picture and voice recognition were improved by innovative methods that concentrate on AI-related activities. Then, in an effort to generate interest in the market, other start-ups embraced it.
- **Financial Services with AI:** The asset management, hiring, and customer service procedures in the banking sector have also seen a number of improvements. For instance, modern finance and stock investment primarily rely on technical know-how and good fortune. But in the future, computers, crowd-sourced data, and sentiment evaluation will allow us to handle money in a totally different way.

11.2 ARTIFICIAL INTELLIGENCE AND ELECTRONIC TRADE

E-commerce has increased significantly over the past years; instance 2018, size of online retail business continued to increase, with sales reaching 9 trillion. Sales of real goods in online stores among them reached 7 trillion in 2017, up 25.4% from the previous year. Additionally, they tracked all retail purchases of products for social consumption, which increased by 7%. E-commerce is becoming more and more convenient for consumers. The electronic commerce platform regularly investigates innovative approaches to meet more consumer expectations in order to improve service quality and operational efficiency [37–39]. New ideas have emerged as a result of the advancement of artificial intelligence and e-commerce will expand in both directions, adding value to the industry in both directions.

- **Electronic Trade:** The management of whatever kind of trade, business, or other form of information utilizing the internet and contemporary communication methods is known as electronic

commerce (EC). The foundation of online shopping is a business strategy based on computer hardware and network technology. E-commerce automates all company procedure involving both internal and external ones. Supply chain management (SCM), enterprise resource planning (ERP), management information system (MIS), CRM (customer relationship management), and HRM are examples of internal procedures [40, 41]. Network advertising, electronic payments, transportation, and shipping are examples of external processes. Internet, extranet, intranet, electronic mail, electronic data interchange (EDI), data bases, development of websites tools, and electronic transactions are the major technological pillars of e-commerce.

E-commerce is the outcome of advancements in economics, science, technology, and culture. It is a revolution in both technology and economy. It is based on the Internet, supported by computer network technology, and achieves effective integration of management, information, and business technologies. The development of e-commerce has had a substantial positive impact on the status of the global economy by changing how firms operate and how consumers shop.

- **Machine Intelligence:** A relatively new area of technology study, artificial intelligence (AI), develops conceptual models, instruments, and software to duplicate and improve human intelligence. In the 1956, some researcher led a gathering of creative young scientists to look at and talk about some relevant topics relating to utilizing machines to simulate intelligence [42–45]. During this time, the phrase “artificial intelligence” was first used. Artificial intelligence, which has developed into a significant, interdisciplinary frontier science, can now be considered to have officially begun.

Technology that emulates and enhances human intelligence is known as artificial intelligence. It can automate mental processes. Robots, voice-activated interfaces, photo recognition systems, and other devices with intelligence equivalent to human thought processes are generated as a result of artificial intelligence technology, which based its work on intelligent technology. Biological sciences, computer psychological research, language study, medicine and other fields are strongly tied to AI technology. Mechanical devices

today possess heightened intelligence that can carry out commands from humans because of the advancements in social science and technology in the modern era. As a result, artificial intelligence is now used in more industries. The evolution of computer technology and artificial intelligence are closely intertwined, sharing comparable inspirations and innovations [46, 47]. The foundational elements of artificial intelligence technology today are learning through interaction and machine learning. As a cross-disciplinary investigation of artificial intelligence, recognition of patterns, data mining, the theory of probability, statistics, and other topics, a subfield of computer science is machine learning. Numerical optimization and machine learning share a lot of similarities.

11.3 ARTIFICIAL INTELLIGENCE TECHNOLOGY'S USE IN ECOMMERCE

The technology of artificial intelligence has steadily advanced into potent instrument to increase sales and improve e-commerce operations, particularly in the area of internet commerce [48, 49]. The development of artificial intelligence technologies is fundamentally altering how people work and live today. The following are the key uses of AI now being used in the area of electronic commerce:

- a. Intelligent Assistant Using AI:** A Chatbot with artificial intelligence (AI) that's primary function is to deliver product recommendations, execute basic voice instructions, and automatically react to consumer enquiries. On mobile and e-commerce websites, machine learning algorithms are utilized to customize customer dialogue in chat dialogs.

Customers can utilize Chatbot to find the right products, verify product availability, contrast products, and, in the end, assist customers with payment. Customers can also use the Chatbot to assist them in contacting the appropriate service provider if they have any issues or inquiries [50, 51]. After getting merchant approval and debugging, Chatbot can change a few client services, significantly lower labor expenses, enhance service quality, boost night traffic recovery, and support customer service in resolving difficulties with repeated consultations.

- b. Intelligent Logistics:** In order to substitute persons who use technical apparatus, the term “intelligent logistics” refers to a physical development method where the use of information technology is made equipment and control intelligent. When compared to traditional logistics, logical logistics may considerably improve service quality and operating efficiency. In 2009, IBM originally proposed the concept of logical logistics [52–54]. The idea of a smart supply chain was first put up by IBM, which called for the use of gauges, tags with RFID, the brakes, GPS, and other hardware and systems to deliver real-time data. The logistics and back-end supply chain connections are the areas where AI most obviously has a direct impact.
- c. Optimal Pricing:** Even for online shops with little inventories, the contemporary electronic commerce business is booming, creating a huge challenge for this kind of long-term continual pricing change [55–57]. Artificial intelligence technology, which can quickly process enormous volumes of data, has essentially addressed the issue of automatic pricing for a variety of goods. The final comprehensive ranking results will be influenced by the item’s evaluation, logistical cost, and service quality. Businesses struggle to set prices that are optimal, but artificial intelligence is well suited to researching pricing concerns thoroughly. AI technology uses extremely sophisticated machine learning algorithms to continuously monitor market’s dynamics and adjust the context of environment in order to address issues of optimum costs.
- d. Customer Relationships Management:** You are living in a world that has long since passed if you still believe that the human resources department’s only responsibility is to manage and maintain the customer relationship. In the contemporary world, a range of information data can be obtained and assessed with the use of artificial intelligence to guarantee that potential customers are provided with the best-in-class services. Thanks to AI, it is now simple to transfer, use, and exchange massive amounts of data. To be able to ensure acceptable and secure contact, this information is utilized to monitor client purchasing trends, preferences, and other factors that may affect their decision to make a purchase. The e-commerce industry needs CRM because only the CRM platform

enables for in-depth research on customer buying patterns, usage patterns, and other factors, allowing for the creation of predictions that are both accurate and profitable [58].

- e. **Management of product content (PCM):** Every e-commerce business wants to give their customers the greatest possible experience, and AI can help them do so. Starting with cataloging, product design, and ensuring that the customer has the best experience possible. One may argue that the development of the greatest and highest-quality PCM (Product Content Management) is made possible by artificial intelligence itself [59].
- f. **Customer Services:** No matter how big or small an e-commerce company is, its consumers are its lifeblood and foundation. Therefore, it is crucial to make sure your consumers are getting the best services and experiences possible from you. One of the options that can help you do everything so that your customers are satisfied all the time is artificial intelligence. At last, AI is able to assist the business in beginning the journey toward flawless and outstanding customer service. Artificial intelligence enables one to learn about customer satisfaction and how to meet their needs regardless of the situation or time.
- g. **AI Helps in Automation:** The majority mistakenly people assume that the moment we say “automation,” we mean robots would eventually replace humans in every task. Organizations may learn what their clients want regardless of a time difference or other obstacle thanks to automation. To guarantee the finest client experience and high revenue, sales personnel from all over the world can communicate conveniently, collaborate, and connect with clients as much as possible with the aid of artificial intelligence [60].

11.4 ELECTRONIC COMMERCE INTELLIGENT SYSTEM PRACTICE CASE ANALYSIS

As the technology in the field advances, AI is being implemented in business in an expanding variety of ways. This essay uses an example from Baidu’s takeaway system to emphasize the importance of smart technology

for e-commerce. Using the Online to Offline (OTO) retailing concept, Baidu Takeout uses a platform to connect customers and merchants and offers services including delivery, new retail, medication, doing errands, and so forth. Delivery services that adhere to punctuality and security are offered using a targeted intra-city logistics model. The user puts an order, food is sent, and courier delivery is the three essential steps in the take-out business basic procedure. System flow's primary objective is to facilitate user order placement and on-time delivery. The secret is to increase allocation efficiency while generating more value for customers and businesses. Smart technology is at the heart of this industry's competitiveness.

The first takeout system in the nation to employ streaming computing and a clever scheduling algorithm was Baidu's. The two key systems that make up Baidu Takeout are the recommendations system and scheduling system. Adaptive data from the Order history, delivery status, current scene, and click history are processed by the suggestion and scheduling system to accomplish the delivery in real-time and effectively. Accurate recommendations are made possible by combining static data such as portraits of user, nutrition facts, retailer portraits, scene preferences, and real-time scenario. The average user wait time can be cut by 28% using intelligent scheduling and recommendations, and order delivery efficiency can be raised by 100% [61–68].

11.4.1 SMART SCHEDULING SYSTEM

A crucial component of the delivery system is the smart time coding system. Smart sequencing and the planning of paths is its main components. In the early days of e-commerce, the dispatcher was primarily in charge of scheduling. A dispatcher is thought to be able to handle around a thousand orders every day. The number of dispatchers needed to fulfill a million orders each day makes for an extremely complex operational structure. A smart and totally automatic scheduling system is needed in the OTO era since the system must handle a lot of orders and data.

Figure 11.1 depicts the architecture of the Baidu takeaway system. Algorithms for flow computation, forecasting, putting orders based on correlation, and dynamic preparation such as designing a route, automated order allocation, smart routing, and so on. are used to schedule tasks in real-time, within constraints created by collection of the courier's GIS service, orders, and relevant user data, among other things.

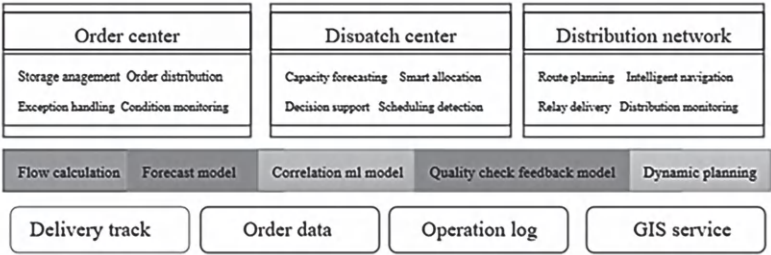


FIGURE 11.1 Architecture diagram of delivery system.

To enhance actual time scheduling, the system additionally makes use of offline tracking, forecasting, and delivery technologies. An offline forecast among them, makes use of intelligent methods to anticipate hotspots and capacity for carrying. That knowledge could be integrated with schedule software and offline scenario analysis to produce the ideal path for dissemination. Additionally, it may be used to automatically carry out business layer tasks such logistic management, retailer administration, schedule identification, and distribution monitoring [62–69]. It additionally has the potential to continuously monitor and modify the supply chain.

11.4.2 INTELLIGENT RECOMMENDATION SYSTEM

To satiate the requirements and objectives of each and every user of the business take-out platform, a variety of orders from merchants and dishes are provided for a variety of cases likes diverse users, regions, and time periods. Through a clever recommendation engine, Baidu Takeout can also analyze data, circumstances, and user profiles, opening up new possibilities.

Figure 11.2 depicts the six main parts of Baidu Takeout’s recommendation system: the program layer, the user simulation layer, the algorithmic layer, the strategic layer, the layer for modeling businesses, and the layer for data processing.

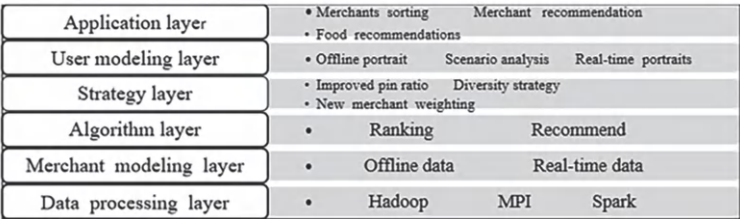


FIGURE 11.2 An intelligent recommendation system’s schematic diagram.

Offline and real-time components are included in merchant and user modeling. Merchant rating, the suitable scenario, Retailer attributes and retailer tag mining are all a part of the merchant model's offline data. Real-time data comprises operational data, merchant activity, inventory data, and capacity data. Both the long-term and short-term portraits of the user are included in the user modeling layer's off-line data. Real-time data covers things like real-time user photos and analysis of purchasing circumstances. The recommendation system's efficient real-time computing technology analyzes the enormous volumes of data from Baidu's users and merchants, creates models of the merchants and users, and implements diverse strategies to achieve the merchant's exact order, the merchant's advice, and dish recommendations [10–13, 65, 66].

11.5 APPLICATION OF AI IN THE BANKING INDUSTRY

11.5.1 CHATBOT

A Chatbot is created using Chatbot and a natural language processor. The key advantages of Chatbot in the banking sector include relationship improvement, relationship development, and the promotion of more mobile bank services. In order to retain consumer knowledge, banks will create a question resource when a Chatbot interacts with customers. This will improve the Chatbot's response time when dealing with more difficult inquiries [67]. As more customers and businesses are being established, cut costs while maintaining and improving customer service levels. Through enabling the customer support representatives to concentrate on more valuable activities, job satisfaction rises.

11.5.2 RISK ASSESSMENT

- **Credit Risk:** Identification of credit risk is the most important step in credit risk management. The most popular traditional way of determining credit risk was developed by Edward Altman and is called credit score analysis. This strategy is increasingly more sophisticated [55]. AI supports banks' analysis of consumer data, both quantitatively and qualitatively.
- **Operational Risks:** Identifying fraudulent activity is one operational risk area where AI may make a substantial difference.

Banks are now placing a high premium on reducing fraud [68]. This is the most effective way of fraud detection in banks has been discovered as using AI applications.

- **Availability Risk:** An essential component in the financial management process is cash flow forecasting. The analytical power of AI can analyze current transaction data that affects a bank's liquidity to identify and predict future consequences [18]. Automatic reaction mechanisms are then activated by AI to assist in resolving such instances.

11.5.3 ALGORITHMIC

Algorithmic trading, or “black-box” trading, has been used by financial institutions for trading, particularly the trading of hedge funds and proprietary traders. This is the use of mathematical models created to react to shifting market conditions instantly, opening and closing positions according to predetermined criteria without involving a human. Major financial institutions, including investment banks and funds, frequently use algorithms to help them operate more quickly and efficiently, which benefits both the business and its clients. These transactions have several benefits, including greater precision, fewer mistakes, investment funds, simultaneous automated evaluation of various market conditions, and a reduction in mistakes due to psychological or emotional factors.

11.5.4 CROSS-SELLING AND SUPPORT MANAGEMENT

For banks, cross-selling has emerged as one of their primary revenue streams. AI uses information from social media activity to assist in comprehension and education over time. The financial institution can recognize patterns, relate with their prior knowledge, and derive new understanding before acting.

This AI function has many benefits for banks, including the ability to offer new items to existing customers and make customized consumer decisions. By supporting new sales channels, enhance the consumer experience. Banks will earn more money, and client retention will go up.

11.6 THE ROLE OF THE BANKING INDUSTRY IN AI

Due to their considerable involvement in the modern economy and their ability to control the value of money, credit scores, and other financial resources, banks are often referring to the phrase “lifeblood” of the economy. Banks encourage and help customers in saving money and earning interest for a more comfortable future. Some of the methods that banks use for business include ATMs, emails, smart phone banking, internet banking, and cell banking. Only because banks use AI is the ideal concept of banking over computers and networks practical.

- **SBI:** State Bank of India (SBI) is the main public sector supplier of banking services in the country. To provide effective financial services, the bank makes use of artificial intelligence. With the help of artificial intelligence (AI), SBI Intelligent Assistant (SIA), a smart conversation assistant, helps customers with routine banking duties similar to how a human would.
- **Baroda Bank:** Baroda Bank, a different public sector lender, focuses on using AI to improve customer service while expanding banking services and cutting expenses associated with account management. The bank makes use of cutting-edge technology, including the Wi-Fi-free Digital Lab and the Baroda Brainy robot.
- **HDFC:** Another Indian financial services and banking company that use AI is HDFC, with its headquarters in Mumbai. Eva, a sophisticated Chatbot created by the bank, uses Google Assistant on millions of Android devices to respond to consumer questions and improve services. Eva was created by the Bengaluru-based Sensing Forward AI Technology and has allegedly stated that it had answered over 5 million user queries accurately (more than 85% of the time). In addition, HDFC provides On Chat, a Chatbot with AI that made its debut on Facebook Messenger in 2016.

11.7 METHODOLOGY

A wide notion, AI is connected to the development of intelligent computers. A few of the various terms used to characterize AI include recognition of images, the processing of natural language (NLP), artificial intelligence boosters, and increased intelligence.

The effects of AI on how the commercial bank system functions are examined using qualitative research methods. To conduct the research, the techniques of document analysis, secondary data collection, and interviewing were chosen. A poll of 65 commercial bank executives from chosen commercial banks was used to collect the primary data.

11.8 ANALYSIS AND INTERPRETATION

The researchers polled a few commercial bank administrators to evaluate how AI will impact the way that commercial banks operate. There were 65 questionnaires conducted, and there were 60 responses. This information is really helpful.

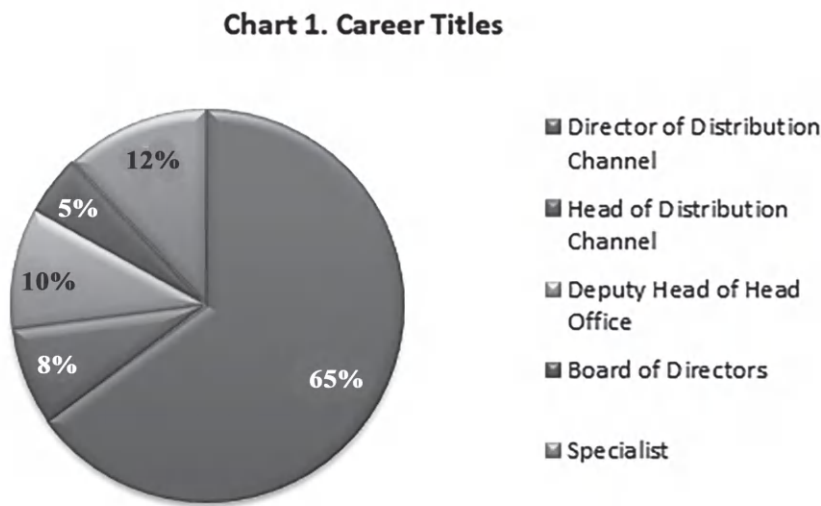


FIGURE 11.3 Career chart.

Data in Figure 11.3 shows that around 90% of respondents, including senior and intermediate managers at commercial banks, held significant management positions within the bank, confirming the validity of this survey’s findings.

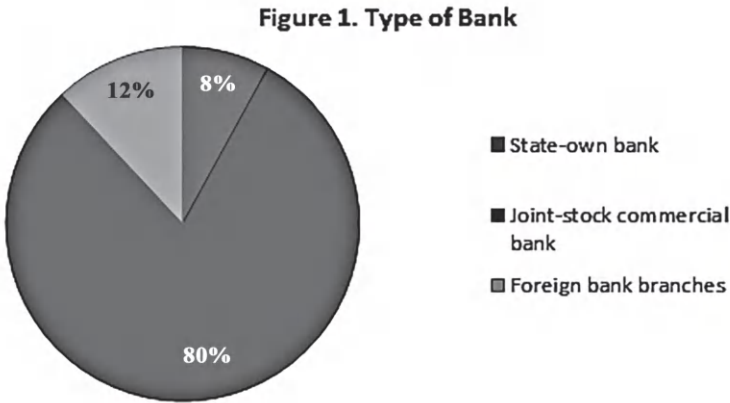


FIGURE 11.4 Types of bank.

Due to the prevalence of joint-stock commercial banks, which make up a significant share of commercial banks, nearly 80% of the sample, as seen in Figure 11.4, came from commercial banks 90% of survey participants were over the age of 31, and 54% of them were over the age of 40 (Figure 11.5).

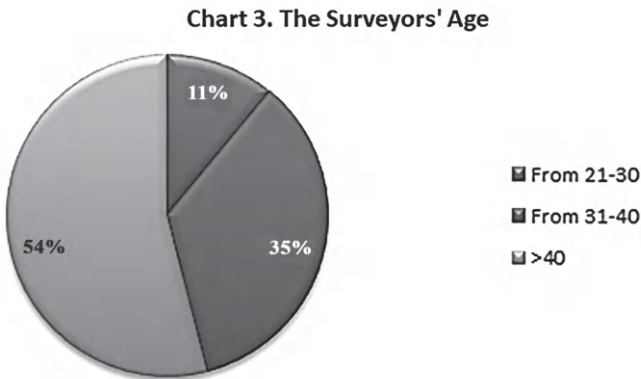


FIGURE 11.5 Surveyors age.

This time period was correct and wise enough to recognize the issue. The sample's working years range mainly between 5 and 10 years; 39% from sample's employment years are more than 10 years. Figure 11.6 shows that surveyor judgments are more trustworthy the longer they have been in the field, the more experience they have, and the more years they have worked.

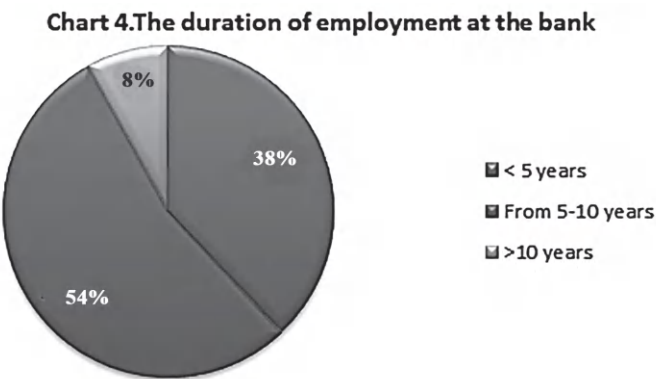


FIGURE 11.6 Duration of employment at the bank.

The study’s findings indicate that the sample’s level of access to AI technology is accessible, with the majority of respondents having some knowledge of it but not in-depth understanding it. Only around 18% of respondents had a thorough understanding of AI, according to up to 55% of the sample (Figure 11.7).

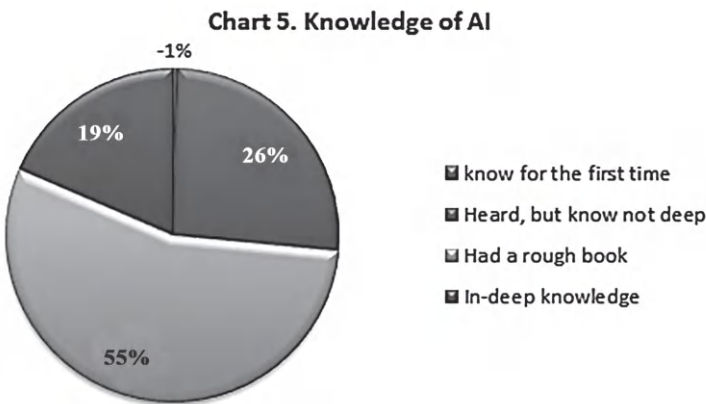


FIGURE 11.7 Knowledge of AI.

Though some commercial banks have only begun to use the most basic apps, the technological gap presents several potentials for banks. According to the statistical sample, banks have reportedly used this technology in 89% of survey samples (Figure 11.8).

Commercial banks mostly use AI technology to help lending analysis, particularly for loans to individuals. Additionally, this technology is employed to combat theft because it puts banks at great risk. When client needs are growing, banks with a big clientele collaborate with other companies to cross-sell products, which provide income for banks.

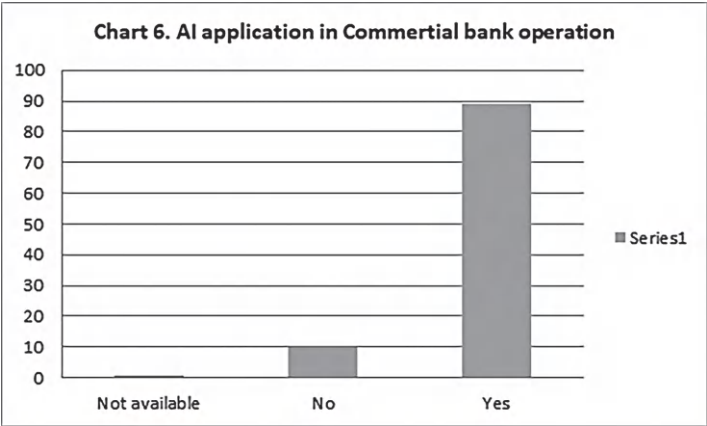


FIGURE 11.8 AI application in commercial bank operation.

Market segmentation, client spending research, and marketing analysis are all made possible with the use of AI. The risk model application had the lowest density, with only around 10% of applications being made in total (Figure 11.9).

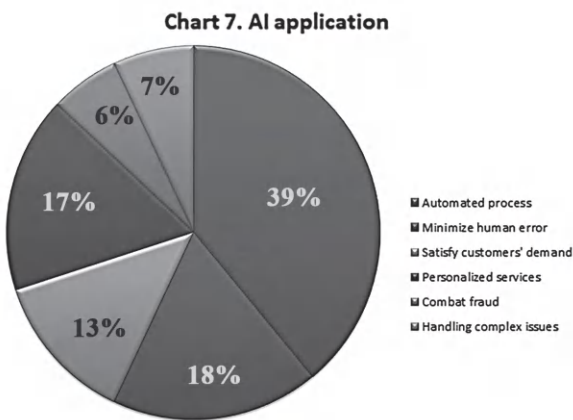


FIGURE 11.9 AI application.

The use of AI technology having a consequence on how bank work, according to greater than 50% of the poll sample; 59% of respondents think this impact is beneficial, and while 29% think there is no positive impact. This may be a result of the technology's relatively costly initial investment cost, but banks stand to gain in the long run. You may get a consensus rate of 67% by streamlining the service, of which 50% strongly agree. Also quite high, accounting for 75% of the total, is the anti-fraud rate. Only 19% of the total was accounted for by the effect on customer satisfaction.

More than 50% of respondents in a study on the influence of AI on risk management agreed these technological advances influenced asset management, risk analysis, and recommendation of acceptable goods for clients based on previous data. The degree of agreement is fairly high, especially in regards to helping to lower transaction costs, comprehend consumer behavior, and assist in automatically complying with banking procedures. This technology's effect on client segmentation is more glaring.

11.9 CONCLUSION

Findings from studies on how AI affects commercial banking Operations demonstrate that three key uses of AI are to reduce human mistake, handle automatic submission, and provide tailored services. Artificial intelligence's effects on how sophisticated goods and services are, lowering fraud and improving the ATM's performance are some of the most noticeable effects on bank operations, which are more beneficial than bad. AI also affects consumer segmentation, behavioral surveillance, more accurate credit evaluation, and bank credit risk control.

To increase competition, e-commerce goliaths are currently aggressively improving their own e-commerce platforms and integrating artificial intelligence technologies. Advanced service robots were introduced by Jingdong, Alibaba, and Amazon. In the logistics industry, E-commerce businesses have also released their own goods. In addition to the character recognition technology, Jingdong also introduced a visual artificial intelligence platform and the Zhong Kui technology, an image information platform for Alibaba called DT PAI. To put it another way, e-commerce businesses each have their own advantages regarding the particular use of artificial intelligence.

KEYWORDS

- **artificial intelligence**
- **bank performance**
- **commercial bank**
- **e-commerce**
- **FinTech**

REFERENCES

1. Nadikattu, R. R. (2020). Implementation of new ways of artificial intelligence in sports. *Journal of Xidian University*, 14(5), 5983–5997.
2. Nadikattu, R. R. (2019). New ways in artificial intelligence. *International Journal of Computer Trends and Technology*.
3. Mohammad, S. M. (2020). Artificial intelligence in information technology.
4. Soni, V. D. (2020). Challenges and solutions for artificial intelligence in cybersecurity of the USA.
5. Peng, M. (2017). Research progress and application overview of artificial intelligence. *Science and Technology Square*, 10.
6. David2AI. (2018). 5 ways artificial intelligence is shaping the future of e-commerce. *CSDN*.
7. Chen, S., Jiang, X., & Qin, W. (2015). Intelligent practice of mobile e-commerce. *Information and Communication Technology*, 9(4), 9–14.
8. Zhou, X. (2018). *Artificial Intelligence: Ushering in the Era of Subversive Intelligence*. Beijing: Taihai Press.
9. Lin, J. (2019). Application of artificial intelligence technology in e-commerce. *China Business Review*, 2019(2), 19, 20.
10. Sun, L., & Huang, Q. (2017). *Practical Machine Learning*. Beijing: Posts and Telecommunications Press.
11. Wang, T. (2017). *Artificial Intelligence Revolution: History, Present and Future*. Beijing: Beijing Times Chinese Book Company.
12. Yang, A., Bu, X., & Yanjia, X. (2018). *Artificial Intelligence era: The Future has Come*. Beijing: Posts and Telecommunications Press.
13. Analysis: What is AI doing for e-commerce? (2017). *Net Economy Agency*.
14. Inventory: Top 10 e-commerce solutions in 2018. (2018). *Web.com*.
15. Jenner, G. (2017). AI-enabled world will change reality: Airlines underestimating the likely future role of artificial intelligence in day-to-day operations might not only lose out on cost savings and other efficiencies but also revenue opportunities. *Flight Airline Business*.

16. Joshi, N. (2019). *7 Types of Artificial Intelligence*. Retrieved from: <https://www.forbes.com/sites/cognitiveworld/2019/06/19/7-types-of-artificial-intelligence/#103c8bcb233e> (accessed on 17 October 2024).
17. Kaya, O., Schilbach, J., AG, D. B., & Schneider, S. (2019). Artificial intelligence in banking. *Artificial Intelligence*.
18. Schwab, K. (2016). *The Fourth Industrial Revolution: What it Means, How to Respond*. Retrieved from: <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/> (accessed on 17 October 2024).
19. Gokul, B. (2018). Artificial intelligence in financial services. *Sansmaran Research Journal*, 8(1), 3–5.
20. Nilsson, N. J. (2009). *The Quest for Artificial Intelligence*. Cambridge University Press.
21. Srivastava, U., & Gopalkrishnan, S. (2015). Impact of big data analytics on the banking sector: Learning for Indian banks. *Procedia Computer Science*, 50, 643–652.
22. Takyar, A. (2018). *What is Artificial Intelligence? Understand Artificial Intelligence in 5 Minutes*. Retrieved from: <https://www.leewayhertz.com/what-is-artificial-intelligence-understand-ai-in-5-minutes/> (accessed on 17 October 2024).
23. Yaninen. (2017). Artificial intelligence and the accounting profession in 2030. Retrieved from: https://cpapng.org.pg/data/documents/CPA-Presentation-Artificial-Intelligence-and-the-Accounting-Profession-in-2030_1.pdf.
24. Vives, X. (2017). The impact of FinTech on banking. *European Economy*, (2), 97–105.
25. Jewandah, S. (2018). How artificial intelligence is changing the banking sector: A case study of top four commercial Indian banks. *International Journal of Management, Technology and Engineering*. Retrieved from: <http://ijamtes.org/gallery/66.july%20ijmte%20-%20711.pdf> (accessed on 17 October 2024).
26. Thim, C. K., & Seah, E. (2010). Optimizing portfolio construction using artificial intelligence. In *5th International Conference on Computer Sciences and Convergence Information Technology* (pp. 338–343).
27. Pathak, M. A., Kamlakar, K., Mohapatra, S., & Nagaraj, U. (2016). Development of control software for stair detection in a mobile robot using artificial intelligence and image processing. *International Journal of Computers Engineering and Technology*, 7(3), 93–98.
28. Dhanabalan, T., & Sathish, A. (2018). Transforming Indian industries through artificial intelligence and robotics in Industry 4.0. *International Journal of Mechanical Engineering and Technology*, 9(10), 835–845.
29. Kallem, S. R. (2013). Artificial intelligence in the movement of mobile agent (robotic). *International Journal of Computer Engineering and Technology*, 4(6), 394–404.
30. Abbasi, A., Zhang, Z., Zimbra, D., Chen, H., & Nunamaker, J. F. (2010). Detecting fake websites: The contribution of statistical learning theory. *MIS Quarterly*, 34(3), 435–461. <https://doi.org/10.2307/25750686>.
31. Abdel-Karim, B. M., Pfeuffer, N., & Hinz, O. (2021). Machine learning in information systems: A bibliographic review and open research issues. *Electronic Markets*, 31(3), 643–670. <https://doi.org/10.1007/s12525-021-00459-2>.

32. Adomavicius, G., Bockstedt, J. C., Curley, S. P., & Zhang, J. (2013). Do recommender systems manipulate consumer preferences? A study of anchoring effects. *Information Systems Research*, 24(4), 956–975. <https://doi.org/10.1057/isre.2013.0497>.
33. Adomavicius, G., Sankaranarayanan, R., Sen, S., & Tuzhilin, A. (2005). Incorporating contextual information in recommender systems using a multidimensional approach. *ACM Transactions on Information Systems*, 23(1), 103–145. <https://doi.org/10.1145/1055709.1055714>.
34. Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), 734–749. <https://doi.org/10.1109/TKDE.2005.99>.
35. Ågerfalk, P. J. (2020). Artificial intelligence as digital agency. *European Journal of Information Systems*, 29(1), 1–8. <https://doi.org/10.1080/0960085X.2020.1721947>.
36. Bag, S., Kumar, S. K., & Tiwari, M. K. (2019). An efficient recommendation generation using relevant Jaccard similarity. *Information Sciences*, 483, 53–64. <https://doi.org/10.1016/j.ins.2019.01.023>.
37. Baptista, J., Stein, M. K., Klein, S., Watson-Manheim, M. B., & Lee, J. (2020). Digital work and organizational transformation: Emergent digital/human work configurations in modern organizations. *The Journal of Strategic Information Systems*, 29(2), 101618. <https://doi.org/10.1016/j.jsis.2020.101618>.
38. Barzegar Nozari, R., & Koochi, H. (2020). A novel group recommender system based on members' influence and leader impact. *Knowledge-Based Systems*, 205, 106296. <https://doi.org/10.1016/j.knosys.2020.106296>.
39. Bassano, C., Gaeta, M., Piciocchi, P., & Spohrer, J. C. (2017). Learning the models of customer behavior: From television advertising to online marketing. *International Journal of Electronic Commerce*, 21(4), 572–604. <https://doi.org/10.1080/10864415.2016.1355654>.
40. Cheung, K. W., Kwok, J. T., Law, M. H., & Tsui, K. C. (2003). Mining customer product ratings for personalized marketing. *Decision Support Systems*, 35(2), 231–243. [https://doi.org/10.1016/S0167-9236\(02\)00108-2](https://doi.org/10.1016/S0167-9236(02)00108-2).
41. Bauer, J., & Jannach, D. (2018). Optimal pricing in e-commerce based on sparse and noisy data. *Decision Support Systems*, 106, 53–63. <https://doi.org/10.1016/j.dss.2017.12.002>.
42. Chiarini Tremblay, M., Kohli, R., & Forsgren, N. (2021). Theories in flux: Reimagining theory building in the age of machine learning. *MIS Quarterly*, 45(1), 455–459.
43. Ghose, A., & Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, 23(10), 1498–1512. <https://doi.org/10.1109/TKDE.2010.188>.
44. Gielens, K., & Steen, J. B. E. M. (2019). Branding in the era of digital (dis) intermediation. *International Journal of Research in Marketing*, 36(3), 367–384. <https://doi.org/10.1016/j.ijresmar.2019.01.005>.
45. Bawack, R. E., Wamba, S. F., & Carillo, K. (2021). A framework for understanding artificial intelligence research: Insights from practice. *Journal of Enterprise Information Management*, 34(2), 645–678. <https://doi.org/10.1108/JEIM-07-2020-0284>.

46. Bai, X., Marsden, J. R., Ross, W. T., & Wang, G. (2020). A note on the impact of daily deals on local retailers' online reputation: Mediation effects of the consumer experience. *Information Systems Research*, 31(4), 1132–1143. <https://doi.org/10.1287/isre.2020.0935>.
47. Aghaei Chadegani, A., Salehi, H., Md Yunus, M. M., Farhadi, H., Fooladi, M., Farhadi, M., & Ale Ebrahim, N. (2013). A comparison between two main academic literature collections: Web of Science and Scopus databases. *Asian Social Science*, 9(5), 18–26.
48. Balabanović, M., & Shoham, Y. (1997). Content-based, collaborative recommendation. *Communications of the ACM*, 40(3), 66–72. <https://doi.org/10.1145/245108.245124>.
49. Bandyopadhyay, S., Rees, J., & Barron, J. M. (2008). Reverse auctions with multiple reinforcement learning agents. *Decision Sciences*, 39(1), 33–63. <https://doi.org/10.1111/j.1540-5915.2008.00181.x>.
50. De Carolis, B., de Gemmis, M., Lops, P., & Palestra, G. (2017). Recognizing users' feedback from non-verbal communicative acts in conversational recommender systems. *Pattern Recognition Letters*, 99, 87–95.
51. De Smedt, J., Lacka, E., Nita, S., Kohls, H. H., & Paton, R. (2021). Session stitching using sequence fingerprinting for web page visits. *Decision Support Systems*, 150, 113579. <https://doi.org/10.1016/j.dss.2021.113579>.
52. MacKenzie, D. (2008). *An Engine, not a Camera: How Financial Models Shape Markets*. MIT Press.
53. Manju, K. (2019). *Artificial Intelligence in Finance: Understanding How Automation and Machine Learning is Transforming the Financial Industry*. Master's thesis, Centrica University of Applied Sciences, India.
54. Dumasia, J. (2021). *5 Applications of Artificial Intelligence in Banking*. Retrieved from: <https://ibsiintelligence.com/ibsi-news/5-applications-of-artificial-intelligence-in-banking/> (accessed on 17 October 2024).
55. Huang, A. H., Leavy, R., Zang, A. Y., & Zheng, R. (2018). Analyst information discovery and interpretation roles: A topic modeling approach. *Management Science*, 64, 2472–2486. <https://doi.org/10.1287/mnsc.2017.2796>.
56. Kumar, S. (2021). Use of artificial intelligence in banking world today. Retrieved from: <https://www.finextra.com/blogposting/20688/use-of-artificial-intelligence-in-banking-world-today> (accessed on 17 October 2024).
57. Lerner, J., & Tufano, P. (2011). The consequences of financial innovation: A counterfactual research agenda. *Annual Review of Financial Economics*, 3, 4185. <https://doi.org/10.1146/annurev.financial.050808.114326>.
58. Sota, S., Chaudhry, H., Chamaria, A., & Chauhan, A. (2018). Customer relationship management research from 2007 to 2016: An academic literature review. *Journal of Relationship Marketing*. Published online: <https://doi.org/10.1080/15332667.2018.1440148>.
59. Trappey, A. J. C., & Trappey, C. V. (2004). Global content management services for product providers and purchasers. *Computers in Industry*, 53, 39–58.
60. Mohammad, S. M. (2020). AI automation and application in diverse sectors. *International Journal of Computer Trends and Technology (IJCTT)*, 68(1). Available at SSRN: <https://ssrn.com/abstract=3630577> (accessed on 17 October 2024).

61. David2AI. (2018). 5 ways artificial intelligence is shaping the future of e-commerce [EB/OL]. CSDN.
62. Chen, S., Jiang, X., & Qin, W. (2015). Intelligent practice of mobile e-commerce. *Information and Communication Technology*, 9(04), 9–14.
63. Zhou, X. (2018). *Artificial Intelligence: Ushering in the Era of Subversive Intelligence*. Beijing: Taihai Press.
64. Lin, J. (2019). Application of artificial intelligence technology in e-commerce. *China Business Review*, (02), 19–20.
65. Peng, M. (2017). Research progress and application overview of artificial intelligence. *Science and Technology Square*, (10).
66. Inventory: Top 10 AI e-commerce solutions in 2018 [EB/OL]. (2018). Web.com.
67. Eryk, L. (2020). *Artificial Intelligence in Finance: Opportunities and Challenges*. Retrieved from: <https://towardsdatascience.com/artificial-intelligence-in-finance-opportunities-and-challenges-ccc94f2f3858> (accessed on 17 October 2024).
68. Kaya, O., Schildbach, J., AG, D. B., & Schneider, S. (2019). Artificial intelligence in banking. *Artificial Intelligence*.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

CHAPTER 12

Sustainable Green Human Resource Management Using Artificial Intelligence

ASHUTOSH SINGH,¹ NAKSHATRESH KAUSHIK,² MEETA JOSHI,³
A. PRABHU,⁴ and RUPESH ROSHAN SINGH⁵

*¹Institute of Business Management, GLA University, Mathura,
Uttar Pradesh, India*

*²Management, Lloyd Business School, Greater Noida, Uttar Pradesh,
India*

³Faculty of Management Studies, Marwadi University, Gujarat, India

*⁴Electrical and Electronics Engineering, K. Ramakrishnan College of
Engineering, Trichy, Tamil Nadu, India*

⁵Mittal School of Business, Lovely Professional University, Punjab, India

ABSTRACT

The techniques and business operations of human resource management (HRM) have been dramatically changed by artificial intelligence (AI). Nowadays, every business is developing cutting-edge, environmentally responsible businesses. Green HRM is necessary for combining corporate environmental management with human resource management. Since it aims to “green” businesses and make them eco-friendly, GHRM is one of these areas that is seeing a lot of study attention. The Scopus

database contains 320 GHRM-related papers that were released between 2008 and 2021 and are the subject of this investigation. In the study, free (unstructured) text found in databases and documents is transformed into standardized, organized information that can be utilized for evaluation or as input for machine learning algorithms through the use of latent semantic analysis (LSA), network analysis, and text mining. To build a connection among the qualitative and quantitative literature on GHRM and to investigate current research trends in the area, KNIME and VOSviewer tools were used in the study's execution. The study uses K-mean clustering to determine five current research trends in GHRM. Future academics may build on these patterns to address environmental problems, make the world eco-friendlier, and encourage businesses to adopt GHRM.

12.1 INTRODUCTION

The nature of work is changing, and we are living through it. Brain jobs are expanding while muscle jobs are dwindling, and labor-based sectors will eventually need to give way to knowledge-based industries to replace them. According to several studies, businesses are to blame for these environmental problems since they consistently produce carbons and other microscopic particles that combine with the air to produce hazardous gases that affect the health of humans and animals [1]. Stakeholders have so begun to concentrate on these issues. The organization has also changed its emphasis to include sustainable development and green activities. Human behavior [2] typically contributes to climate change [3], since people waste more resources and pollutes the air, water, and environment. Since a result, businesses now place a higher priority on creating policies that encourage green behavior [4]. One strategy for promoting green behaviors and efforts is green HRM [5, 6].

Traditional human resource management techniques are dated and ineffective in adjusting to environmental changes. Organizations are concentrating on adopting green practices these days due to their efficacy. Organizations are using more technology to concentrate on green practices, which is why brain jobs are replacing physical labor. AI has becoming more important [7] in the adoption of ethical methods for managing human resources.

GHRM is described by researchers as “Activities of human resource management which improve the outcome of the environment” [8]. In order to accomplish sustainability, reducing waste, social duty, and small advantage, technology is used using human resources. This is achieved through carrying out ecological targets and procedures that are connected to the association’s arrangements and objectives [9], as well as by proceeding to learn and fill here. An employee’s attitudes, knowledge, and behavior in the workplace can be evaluated and influenced through human resource management.

The company may now use HRM to execute green initiatives and communicate them [10]. It contains two essential elements, including knowledge of capital conservation and environmentally sustainable human resource methods. These two elements sustain the degree of employee involvement and retention while lowering costs and improving efficiency. Selection, appointment, enrollment, training, motivation, development, and performance are all included in GHRM and may be very important in problem-solving and are linked to the environment. In order to reduce carbon emissions, minimize travel, and promote teleconferences, several organizations are endorsing green policies [11].

Global warming is a problem that is exacerbated by industrialization. Global warming is a result of a multitude of factors, including human activity and mechanization [12]. The COVID-19 virus shocked the whole globe in 2019 to the point that it was deemed a worldwide pandemic in 2020. Each nation’s economy is still being weakened by this illness. To slow the spread of the virus, countries established limits on international travel, but these measures only slightly improved the environment. Many environmentalists, scientists, governments, and organizations have stepped up to make a variety of initiatives. Organizations that have made specific suggestions for such issues have been shown by environmental concerns on international forums [13, 14].

Organizations, legislatures, and both the general population and confidential areas have vowed to embrace green human asset the executives as a response to ecological issues [15, 16]. In accordance with environmental concerns, HRM has modified its approach to GHRM [18, 19]. The culture of a company has changed from HRM to GHRM as a result of several international organizations dealing with environmental concerns [20]. The term “Green” was first used in HRM by some studies while editing a book “Greening people: human resource and environmental management” [21].

After the 1990s, the idea of GHRM began to appear in research papers. The people are a vital resource that may be exploited to find a solution for each of the challenges covered above. Environmental problems can only be solved through mindfulness and awareness. The administration of HR should be a piece of each and every association's methodology for encouraging development among its individuals and elements [22]. Any firm may increase employee productivity and commitment by using HRM policies to affect their drive, aptitude, and availability for work-related duties [23]. Every industry that supports the economy has benefited from HRM's contributions in the past.

12.1.1 PURPOSE OF RESEARCH

The method used to conduct the systematic literature review is what makes this research special. Previous studies focused on thematic analysis of informational texts that discussed GHRM and the theories, procedures, and guidelines that the organization followed. We focus on using data mining techniques to extract details from unstructured data instead. As a result, it presents important phrases from the corpus alongside a suitable score for every word. Additionally, it creates a semantic connection among the concepts, which helps in the future when explaining contemporary trends to academics. Therefore, employing natural language processing, this study will assist researchers in locating keywords, important subjects, and developing research fields in GHRM.

12.1.2 OBJECTIVE OF RESEARCH

1. Discuss a number of environmentally friendly procedures that may be used to establish a green workplace.
2. Give readers a fundamental knowledge of green HRM.
3. To identify and forecast current trends and GHRM's key areas.

12.1.3 RESEARCH QUESTION

- RQ1: Who are the top publishers and researchers in GHRM?
- RQ2: What connection exists between GHRM's qualitative and quantitative literature?

- RQ3: What recent developments and prospective developments are there in GHRM?

12.2 LITERATURE REVIEW

12.2.1 AI AND GHRM

“Green human resource management” is an expression used to demonstrate the incorporation of environmental manageability into approaches of human resource management [24, 25]. This might include things like putting into effect environmentally friendly policies and procedures, encouraging staff to behave sustainably, and integrating sustainability into the recruiting and recruitment process. By using GRHM principles, businesses [26] may concentrate on lessening their environmental impact as well as enhancing employee engagement and happiness, which may provide them a competitive edge in the market. AI is the term used to describe the reproduction of human intellect in robots that are built to think and learn similarly to people. Decision-making, identification of images, and natural language processing are just a few of the activities that artificial intelligence [27] is employed for. Depending on the degree of the system’s sophistication, there are many forms of AI.

12.2.2 AI IN SUSTAINABLE DEVELOPMENT AND GREEN HRM

As organizations work to get naturally cognizant and manageable, the job of artificial intelligence in Green HRM [28] is turning out to be progressively pivotal. The most common way of fulfilling present needs without forfeiting the limit of people in the future to fulfill their own requirements is known as maintainable turn of events. This includes elements pertaining to the environment, society, and economics. The definition of “sustainable development” was initially given in the Brundtland Report, which said that it was “development that meets the requirements of now while safeguarding the capacity of generations to come to satisfy their own requirements” [29].

In addition to automating and speeding up HR processes, AI technology could be used to evaluate data on employee commute patterns and energy use at work. It can also be used to spot potential for telecommuting and other options. Organizations may lower their carbon footprint and enhance their overall environmental performance by using AI to encourage sustainable practices [30]. AI may also promote staff involvement and environmental

education, which will assist to create a sustainable organizational culture. By and large, incorporating artificial intelligence into green HRM is a vital stage toward acknowledging feasible turn of events and accomplishing the SDGs framed by the UN.

AI may also be used to measure and analyze how much energy is utilized at work and to identify areas for improvement. The use of paper and other resources may be decreased by using AI to automate and simplify HR activities like training and recruiting. Overall, by strategically using technology, AI may assist enterprises in becoming more environmentally friendly and sustainable [31].

Together, the ideas of sustainable development (SD) and green human resource management (GRHM) may construct a more manageable future [32]. GRHM centers explicitly around how associations can integrate ecological supportability into their human asset the executives [20] rehearses. In contrast to sustainable development, which considers deeper social problems and how to satisfy the requirements of today's generation while safeguarding the ability of future generations, GRHM looks at more specific organizational issues?

12.2.3 PAST STUDIES

In this part, the strategies currently employed in research for green human resource management are presented. Businesses from all across the world are also essential to saving the environment. To guarantee sustainability, these businesses are implementing new procedures and regulations. The sustainable culture is created, implemented, and sustained by the human resource management (HRM) division [34]. Additionally, according to studies, the HRM department is involved in hiring individuals, developing strategies, and giving workers access to facilities for the good of the company [35].

Renwick and accomplices described GHRM as the joining of corporate environment management and human resource management [36]. Human resource management (HRM) encompasses the processes of hiring, performance, benefits, recognition, training, and development, as well as other tasks related to employees [37]. Green HRM is also thought to combine green policies with human resource policies. For every one of these cycles, the public authority and worldwide associations give green arrangements

that incorporate the arrangement of supportability and resolving natural issues [38, 39]. According to studies revision of the concept of GHRM, it all revolves around greening HRM practices in relation to their functional and competitive aspects [40]. Ecological sustainability is influenced by the creation, application, and maintenance of a system to make workers and organizations environmentally friendly in order to fulfill environmental objectives [41]. Green and other disciplines have been combined to create a number of ground-breaking projects, including green marketing [42], green finance [43], green retailing [44], and eco-friendly in the incorporation of HRM, which is always changing.

In GHRM, there have been a few writing surveys embraced [45–47]. The creator declares that GHRM rehearses delayed affect worker inspiration to take part in natural exercises in light off's how associations might interpret developing green abilities and giving open doors to representatives to take an interest in ecological administration [48]. Associations might be ruining EM improvement endeavors by not carrying out all GHRM best practices. Analysts examine the motivations behind different HRM practices and inclination the business to offer more to natural maintainability [49]. They inspect the various open doors for crucial HRM and environment the board coordination. Studies examine the association among GHRM and green stock administration as a pivotal HRM subject [50]. This study recommended a structure for GHRM and GSCM incorporation as well as an examination plan. Moreover, the significance of coordinating GHRM and GSCM for analysts, chiefs, and specialists in the fields of hierarchical supportability and feasible stock chains was accentuated. Researchers' rationale for the introduction of GHRM was effective business environment management [51]. They gave hypothetical underpinnings to GHRM; observational turn of events, measurements, and components that lead to GHRM rehearses. The study created a model that accounts for all of the variables that affect it and placed an emphasis on the value of understanding and quantification.

The significant goal of the exploration project is to execute a semi-mechanized writing survey that is less one-sided in examining the latest thing in the relevant issue. A semi-computerized and quantitative examination of the data gained from Scopus has been performed utilizing dormant semantic investigation, a piece of natural language processing (NLP) [52]. The KNIME tool, which is accessible to all academics as open-source software, is used by the author to create the semi-automatic analysis

method [53]. The research community will be assisted in identifying the key topics and current trends in GHRM by the analysis conducted in this study. The present orientations and key regions of several fields have previously been shown using this method [54–56, 58]. There aren't much empirical research papers that discuss current GHRM trends, to maximize the extent of the researchers' knowledge.

12.3 METHODOLOGY

12.3.1 CHOOSING A STRING

The development of a string produced in accordance with the rules is the initial stage in a systematic literature study [59]. The string was built using AND and OR Boolean operations, keywords from relevant publications, and key phrases from the topic selection. Execution of the AND procedure to the catchphrases in the pursuit term considers the consideration of simply those articles in results from look through that contain the critical watchwords in general. However, any articles that contain one or more of the chosen keywords are included when OR are used. The vital watchwords for this exploration are “green” and human asset the executives. The last pursuit string, made by joining various watchword mixes with Boolean tasks, is “green human asset” OR “green human asset the board” OR “ghrm” OR “green hrm.”

12.3.2 SOURCE OF DATA

Although there are several digital libraries where you may get publications about Green HRM, Scopus was taken into consideration for this study. Moreover, prior investigations of Scopus, Web of Science, and other scholarly data sets, for example, [60–62], uncover that Scopus has a more extensive inclusion of distributions, enlightening the way that it is a complete and dependable information source [63], supporting its qualification for this survey. The subsequent inquiry question was thus used to look through the Scopus information base. This led to the retrieval of 325 findings, of which 320 studies were kept for analysis. Each article's title, authors, year, publication source, search terms, and abstract are included in an example of the produced corpus for efficient LSA analysis.

12.3.3 IMPLEMENT

Since it is an open-source application with text handling capacities and the review depends on text mining, the KNIME instrument was decided to do this request [64]. The program is user-friendly and enables process sharing amongst writers [65]. The open-source software VOSviewer program is used for network analysis.

12.3.4 KNIME PROCESS

The text must be pre-processed in KMINE in order to use text mining. POS defined as Part of Speech labeling, eliminating the report's numerals, eliminating stop words, and isolating the words are a portion of the pre-handling methods [66]. The document's token of words are constructed using POS tagging, which also produces a BOW defined as Bag of Words. The LSA is carried out using a word dictionary called BOW. Utilizing the case conversion node in KNIME, all of the papers are converted to lower- or uppercase during the normalization step, getting them ready for text mining applications inside the corpus. To eliminate any prejudice resulting from case sensitivity, this case conversion is necessary.

Additionally, stop words and any other textual components that lack information are removed during stop word filtering [67]. The redundancy in the data is eventually eliminated using the stemming procedure to boost the efficacy of the text mining procedure with LSA [68].

12.4 RESULTS AND DISCUSSION

12.4.1 META ANALYSIS

The meta-investigation fundamentally centers on descriptive data from the corpus of papers. Since the corpus contains data on these and different qualities, different metadata may be built for each article, year, subject, writer, and so on. Table 12.1 records the best 9 GHRM scientists along with the number of articles and references they have in the chosen corpus. This shows that Jabbour is the researcher of Green HRM with the most distributions and references. Jabbour has 26 publications in GHRM, compared to fewer than 10 for all the other major writers.

TABLE 12.1 Top 9 Researchers

Name of Author	Document	Citation
Jabbour	26	1705
Teixeira	6	177
Yusliza	6	176
Guerci	6	379
Renwick	6	345
Yong	5	141
Paille	4	393
Pham	4	152
Raut	3	36

Figure 12.1 demonstrates that the year 2008 marked the start of the development of papers in the subject of GHRM. A steady increase in articles submitted each year has also been seen. In 2020, there may be a maximum of 102 publications in this subject. It suggests that GHRM is a hotly debated topic right now. As a means of achieving sustainability, managing environmental challenges, and other goals, academics are becoming more and more interested in integrating green practices into HRM.

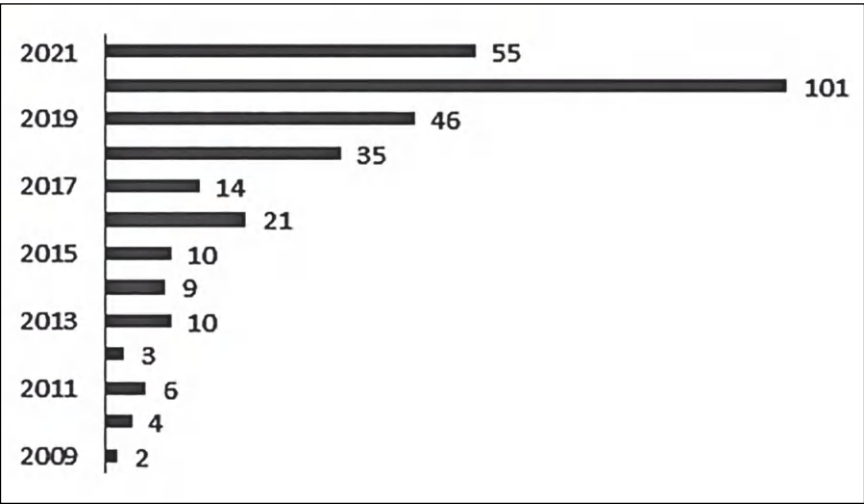


FIGURE 12.1 Publication analysis year wise.

Figure 12.2 displays the top 9 journals by publications. The majority of publications—27 articles—come from the Journal of Cleaner Production,

accounting for almost 10% of all the papers included for analysis. Following maintainability, which has distributed a few articles, are the Global Diary of HRM and the Worldwide Diary of Labor supply.



FIGURE 12.2 Top 9 GHRM-dominating journals.

To assist with text analysis, a document term matrix is created from the corpus. It contains information about the frequency, absence, and presence of various words in a given document. Popular words or frequent terms may be retrieved using this matrix. The phrase frequency-inverse document frequency (TF-IDF) score is developed with this objective in mind. In order to determine which article was weighted more heavily than the others and to choose the terms with a high TF-IDF score, this study computed TF-IDF scores from the corpus. NLP practitioners classify and summarize text using the TF-IDF approach [68–71].

Using the extracted words from the chosen corpus, this approach determines the significance of phrases. The keyword in a particular text serves as the basis for the weights that TF-IDF applies to the words from the BOW corpus [58].

Using the acronym TF, which stands for frequency, one may determine how often a certain word appears in a document. Concentrates on ensure that not permanently set up as the extent of the articulation’s repeat to unquestionably the quantity of various events in the report [33]. The corpus has 320 articles that were collected straight from Scopus, and

BOW created a word dictionary using 3135 distinct tokens. The top 19 frequently occurring terms out of these 3135 tokens are visually shown in Figure 12.3.

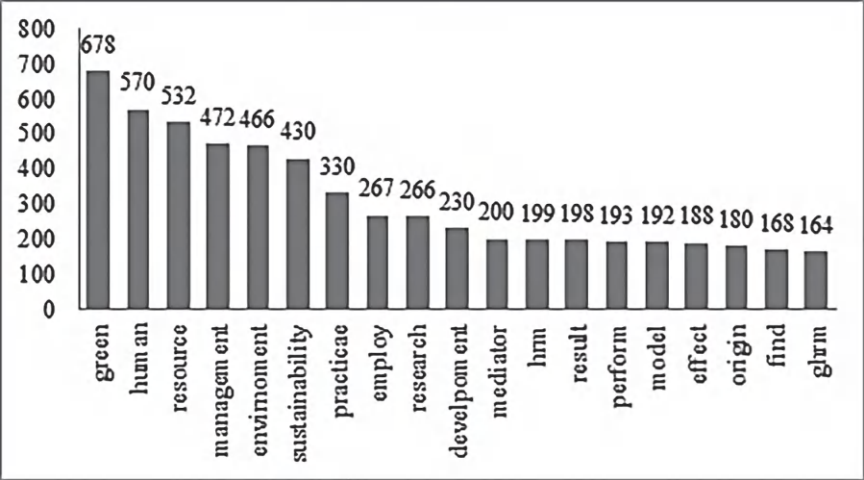


FIGURE 12.3 Top 19 frequent term.

To get data from the investigations accumulated from the Scopus data set on GHRM, the examination discoveries might be sorted, and subjects may then be allotted in view of terms that heap into these bunches. K-mean clustering is used to examine the data after BOW extracts the K number of phrases from the corpus and generates a topic solution. Deerwester contends that five is the best number of subjects for a corpus of 320 documents, despite the fact that it may be challenging to select the ideal or right number of themes [57]. Table 12.2 displays the optimal five themes that were retrieved using KNIME and had high loading words. Based on their TF-IDF ratings, these five subjects or clusters may now be identified.

12.4.2 ANALYSIS OF NETWORK

The relationship between the writers and their co-authors is shown by the co-authorship analysis. Such analysis is essential for determining which authors are acting as main researchers and for networking purposes in order to disseminate the findings. This may be very helpful in the strategic design of programs that aim to improve research and development

TABLE 12.2 Ideal Subject Label with Plenty of Phrases to Load

Cluster	Topic	Terms Used
Cluster 1	Global Approaches to GHRM	Malaysia, Sustainable, system, systematic, environmental management, analysis, information
Cluster 2	Organizational and Environmental Sustainability Techniques	‘Green,’ ‘sustainable’ capital, ‘fuzzy obstacles,’ ‘leadership,’ ‘organizational methods,’ ‘quality,’ ‘employee skills’
Cluster 3	Responsibilities for eco-innovation	Management, organizational behavior, growth and development, eco-innovation, dependent on resources citizenship, business, sustainable identity, and goods
Cluster 4	Behavior Control and Performance Elements	Employee, social, education, environmental responsibility, behavior, management, creative policies, staff, human, and social reward
Cluster 5	Health and Ethical Advantages	Hospital, environmental pandemic, efficiency, policy, green ethical standards, participants, diagnosis, quality, advantages

capability [72]. Figure 12.4 displays the co-authorship network for this study, which was created using the VOSviewer program. One network of writers is represented by each color code. Jabbour is the primary contributor with the most extensive networking because he is in bold and is at the center of the network. Figure 12.4 shows network density in a manner similar to that of node sizes.

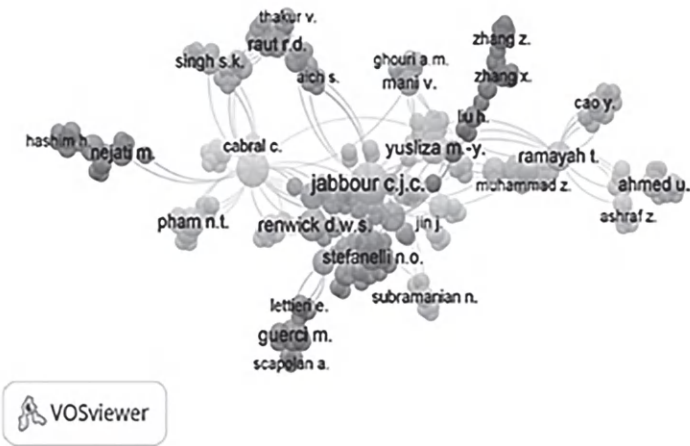


FIGURE 12.4 Network of research.

12.5 CONCLUSION

Human resource management's primary objective is to find the best applicants at the right time and place. Human resource management continually places an emphasis on the effectiveness, participation, efficiency, inventiveness, and sustainability of the business and its employees. Advertising, recruiting, selection, training, deployment, performance, awards, and so on are all handled by the human resources department. The hiring procedure is not particularly environmentally friendly or concerned with long-term environmental sustainability. Our environment is deteriorating every day as a result of climate change, prompting businesses and individuals to embrace green policies. Less paper and environmentally friendly recruiting methods are employed since all human resource operations are conducted online. The focus of green human resources is ecological sustainability, however because this is a new discipline, future researchers will focus on the areas that require more investigation. Based on the keywords that were recovered using latent semantic analysis, the author of this article described current trends; nevertheless, further research may be necessary. Practices for Organizational and Environmental Sustainability, Global GHRM Strategies, Behavior Management, Productivity variables, Ethics and Wellness Advantages, and Eco-Innovation Duties are some areas that require more emphasis or where researchers may conduct studies. Today's scholars are becoming increasingly interested in text mining and developing fields like natural language processing (NLP). To conduct this study, text mining and LSA are employed in NLP.

This study examines all features of GHRM using the referenced research domains. The results of this study support every topic of study relevant to GHRM and beneficial to environmental sustainability. Through their study, organizations may be given progressive policies and methods to address environmental challenges. Additionally, GHRM material that raises awareness of the adoption of green practices should be found. Utilizing text mining, LSA, and network examination, agents will actually want to detect GHRM obstacles and work toward carrying out green practices to settle ecological issues.

KEYWORDS

- **artificial intelligence**
- **GHRM**
- **NLP**
- **sustainability**
- **text mining**

REFERENCES

1. Rajabpour, E., Fathi, M. R., & Torabi, M. (2022). Analysis of factors affecting the implementation of green human resource management using a hybrid fuzzy AHP and type-2 fuzzy DEMATEL approach. *Environmental Science and Pollution Research*, 29(32), 48720–48735.
2. Zoogah, D. B. (2011). The dynamics of green HRM behaviors: A cognitive social information processing approach. *German Journal of Human Resource Management*, 25(2), 117–139.
3. Gatersleben, B. (2018). Measuring environmental behaviour. In *Environmental Psychology: An Introduction* (pp. 155–166).
4. Dumont, J., Shen, J., & Deng, X. (2017). Effects of green HRM practices on employee workplace green behavior: The role of psychological green climate and employee green values. *Human Resource Management*, 56(4), 613–627.
5. Shafaei, A., Nejati, M., & Yusoff, Y. M. (2020). Green human resource management: A two-study investigation of antecedents and outcomes. *International Journal of Manpower*, 41(7), 1041–1060.
6. Gupta, M., & Haque, M. M. (2015). Talent retention: A major concern for organizations. *Proceedings of ICRBS*, 2015.
7. Alzyoud, A. A. Y. (2022). Artificial intelligence for sustaining green human resource management: A literature review. In *2022 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETISIS)* (pp. 321–326). IEEE.
8. Renwick, D. W., Redman, T., & Maguire, S. (2013). Green human resource management: A review and research agenda. *International Journal of Management Reviews*, 15(1), 1–14.
9. Le Blanc, D. (2015). Towards integration at last? The sustainable development goals as a network of targets. *Sustainable Development*, 23(3), 176–187.
10. Saeed, B. B., Afsar, B., Hafeez, S., Khan, I., Tahir, M., & Afridi, M. A. (2019). Promoting employees' pro-environmental behavior through green human resource management practices. *Corporate Social Responsibility and Environmental Management*, 26(2), 424–438.

11. El Dessouky, N. F., & Alquaiti, H. O. (2020). Impact of green human resource management (GHRM) practices on organizational performance. In *2020 Second International Sustainability and Resilience Conference: Technology and Innovation in Building Designs* (51154) (pp. 1–4). IEEE.
12. Olhoff, A., & Christensen, J. M. (2020). *Emissions Gap Report 2020*.
13. United Nations Framework Convention on Climate Change (UNCC). (1997). *Kyoto Protocol*. Retrieved from: https://unfccc.int/kyoto_protocol (accessed on 17 October 2024).
14. United Nations Framework Convention on Climate Change (UNFCCC). (2007). *Bali Climate Change Conference*. Retrieved from: <https://unfccc.int/process-and-meetings/conferences/past-conferences/bali-climate-change-conference-december-2007/bali-climate-change-conference-december-2007-0> (accessed on 17 October 2024).
15. Agreement, Paris. (2016). *The Paris Agreement*. Retrieved from: <https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement> (accessed on 17 October 2024).
16. Unit, B., Star, E., & SmartWay, E. P. A. (2015). *Corporate Social Responsibility*.
17. Victor, D. G. (2011). *The Collapse of the Kyoto Protocol and the Struggle to Slow Global Warming*. Princeton University Press.
18. Wang, H., Tong, L., Takeuchi, R., & George, G. (2016). Corporate social responsibility: An overview and new research directions. *Academy of Management Journal*, 59(2), 534–544.
19. Gardiner, S. M. (2004). The global warming tragedy and the dangerous illusion of the Kyoto Protocol. *Ethics & International Affairs*, 18(1), 23–39.
20. Wehrmeyer, W. (2017). *Greening People: Human Resources and Environmental Management*. Routledge.
21. Milliman, J., & Clair, J. (2017). Best environmental HRM practices in the U.S. In W. Wehrmeyer (Ed.), *Greening People* (pp. 49–73). Routledge.
22. Vuong, B. N., & Suntrayuth, S. (2019). The impact of human resource management practices on employee engagement and the moderating role of gender and marital status: Evidence from the Vietnamese banking industry. *Management Science Letters*, 10(7), 2020.
23. Sharma, C., Ahmad, S., & Singh, S. (2022). Impact of human resource practices on individual and organizational growth. In *2022 International Conference on Decision Aid Sciences and Applications (DASA)* (pp. 937–941). IEEE.
24. Islam, M. A., Jantan, A. H., Yusoff, Y. M., Chong, C. W., & Hossain, M. S. (2020). Green human resource management (GHRM) practices and millennial employees' turnover intentions in the tourism industry in Malaysia: Moderating role of work environment. *Global Business Review*, 0972150920907000.
25. Sharma, K. (2016). Conceptualization of green HRM and green HRM practices: Commitment to environmental sustainability. *International Journal of Advanced Scientific Research and Management*, 1(8), 74–81.
26. Chan, E. S., Hon, A. H., Chan, W., & Okumus, F. (2014). What drives employees' intentions to implement green practices in hotels? The role of knowledge, awareness, concern, and ecological behavior. *International Journal of Hospitality Management*, 40, 20–28.

27. Gupta, M. (2021). The role of artificial intelligence in adopting green HRM practices. In *Reinventing Manufacturing and Business Processes Through Artificial Intelligence* (pp. 1–18).
28. Ansari, N. Y., Farrukh, M., & Raza, A. (2021). Green human resource management and employees' pro-environmental behaviors: Examining the underlying mechanism. *Corporate Social Responsibility and Environmental Management*, 28(1), 229–238.
29. White, M. A. (2013). Sustainability: I know it when I see it. *Ecological Economics*, 86, 213–217.
30. Gupta, H. (2018). Assessing organizations' performance on the basis of GHRM practices using BWM and fuzzy TOPSIS. *Journal of Environmental Management*, 226, 201–216.
31. Garg, V., Srivastav, S., & Gupta, A. (2018). Application of artificial intelligence for sustaining green human resource management. In *2018 International Conference on Automation and Computational Engineering (ICACE)* (pp. 113–116). IEEE.
32. Jyoti, K. (2019). Green HRM–people management commitment to environmental sustainability. In *Proceedings of the 10th International Conference on Digital Strategies for Organizational Success*.
33. Artama, M., Sukajaya, I. N., & Indrawan, G. (2020). Classification of official letters using TF-IDF method. In *Journal of Physics: Conference Series* (Vol. 1516, No. 1, p. 012001). IOP Publishing.
34. Collings, D. G., Wood, G. T., & Szamosi, L. T. (2018). *Human Resource Management: A Critical Approach*. Routledge.
35. Heneman, R. L., Tansky, J. W., & Camp, S. M. (2000). Human resource management practices in small and medium-sized enterprises: Unanswered questions and future research perspectives. *Entrepreneurship Theory and Practice*, 25(1), 11–26.
36. Renwick, D., Redman, T., & Maguire, S. (2008). Green HRM: A review, process model, and research agenda. *University of Sheffield Management School Discussion Paper*, 1(1), 1–46.
37. Jamal, T., Zahid, M., Martins, J. M., Mata, M. N., Rahman, H. U., & Mata, P. N. (2021). Perceived green human resource management practices and corporate sustainability: Multigroup analysis and major industries perspectives. *Sustainability*, 13(6), 3045.
38. Jabbour, C. J. C., Santos, F. C. A., & Nagano, M. S. (2010). Contributions of HRM throughout the stages of environmental management: Methodological triangulation applied to companies in Brazil. *The International Journal of Human Resource Management*, 21(7), 1049–1089.
39. Daily, B. F., & Huang, S. C. (2001). Achieving sustainability through attention to human resource factors in environmental management. *International Journal of Operations & Production Management*, 21(12), 1539–1552.
40. Jackson, S. E., Renwick, D. W., Jabbour, C. J., & Muller-Camen, M. (2011). State-of-the-art and future directions for green human resource management: Introduction to the special issue. *German Journal of Human Resource Management*, 25(2), 99–116.
41. Yong, J. Y., Yusliza, M. Y., & Fawehinmi, O. O. (2020). Green human resource management: A systematic literature review from 2007 to 2019. *Benchmarking: An International Journal*, 27(7), 2005–2027.

42. Grant, J. (2008). Green marketing. *Strategic Direction*, 24(6), 25–27. <https://doi.org/10.1108/02580540810868041>.
43. Bebbington, J. (2001). Sustainable development: A review of the international development, business and accounting literature. *Accounting Forum*, 25(2), 128–157. Taylor & Francis.
44. Lai, K. H., Cheng, T. C. E., & Tang, A. K. (2010). Green retailing: Factors for success. *California Management Review*, 52(2), 6–31.
45. Paulet, R., Holland, P., & Morgan, D. (2021). A meta-review of 10 years of green human resource management: Is Green HRM headed towards a roadblock or a revitalization? *Asia Pacific Journal of Human Resources*, 59(2), 159–183.
46. Teng-Calleja, M., Presbitero, A., & de Guzman, M. M. (2022). Organizational direction, expectations, and employees' intention for Green HRM practices in the Philippines: A signaling theory perspective. *Asian Business & Management*, 1–27.
47. Shahriari, B., Hassanpoor, A., Navehebrahim, A., & Jafarinia, S. (2019). A systematic review of green human resource management. *Sustainability*, 11(21), 6027.
48. Rayner, J., & Morgan, D. (2018). An empirical study of 'green' workplace behaviors: Ability, motivation, and opportunity. *Asia Pacific Journal of Human Resources*, 56(1), 56–78.
49. Renwick, D. W., Jabbour, C. J., Muller-Camen, M., Redman, T., & Wilkinson, A. (2016). Contemporary developments in Green (environmental) HRM scholarship. *The International Journal of Human Resource Management*, 27(2), 114–128.
50. Jabbour, C. J. C., & de Sousa Jabbour, A. B. L. (2016). Green human resource management and green supply chain management: Linking two emerging agendas. *Journal of Cleaner Production*, 112, 1824–1833.
51. Ren, S., Tang, G., & Jackson, S. E. (2018). Green human resource management research in emergence: A review and future directions. *Asia Pacific Journal of Management*, 35, 769–803.
52. Hoblos, J. (2020). Experimenting with latent semantic analysis and latent Dirichlet allocation on automated essay grading. In *2020 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS)* (pp. 1–7). IEEE.
53. Alam, S., & Yao, N. (2019). The impact of preprocessing steps on the accuracy of machine learning algorithms in sentiment analysis. *Computational and Mathematical Organization Theory*, 25, 319–335.
54. Fortuna, B., Mladenich, D., & Grobelnik, M. (2005). Semi-automatic construction of topic ontologies. In *European Web Mining Forum* (pp. 121–131). Berlin, Heidelberg: Springer Berlin Heidelberg.
55. Kulkarni, S. S., Apte, U. M., & Evangelopoulos, N. E. (2014). The use of latent semantic analysis in operations management research. *Decision Sciences*, 45(5), 971–994.
56. Kundu, A., Jain, V., Kumar, S., & Chandra, C. (2015). A journey from normative to behavioral operations in supply chain management: A review using latent semantic analysis. *Expert Systems with Applications*, 42(2), 796–809.
57. Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6), 391–407.

58. Yalcinkaya, M., & Singh, V. (2015). Patterns and trends in building information modeling (BIM) research: A latent semantic analysis. *Automation in Construction*, 59, 68–80.
59. Kitchenham, B. A. (2012). Systematic review in software engineering: Where we are and where we should be going. In *Proceedings of the 2nd International Workshop on Evidential Assessment of Software Technologies* (pp. 1–2).
60. Singh, V. K., Singh, P., Karmakar, M., Leta, J., & Mayr, P. (2021). The journal coverage of Web of Science, Scopus, and Dimensions: A comparative analysis. *Scientometrics*, 126, 5113–5142.
61. Zhu, J., & Liu, W. (2020). A tale of two databases: The use of Web of Science and Scopus in academic papers. *Scientometrics*, 123(1), 321–335.
62. Harzing, A. W., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: A longitudinal and cross-disciplinary comparison. *Scientometrics*, 106, 787–804.
63. Tseng, M. L., Islam, M. S., Karia, N., Fauzi, F. A., & Afrin, S. (2019). A literature review on green supply chain management: Trends and future challenges. *Resources, Conservation and Recycling*, 141, 145–162.
64. Fillbrunn, A., Dietz, C., Pfeuffer, J., Rahn, R., Landrum, G. A., & Berthold, M. R. (2017). KNIME for reproducible cross-domain analysis of life science data. *Journal of Biotechnology*, 261, 149–156.
65. Dietz, C., & Berthold, M. R. (2016). KNIME for open-source bioimage analysis: A tutorial. *Focus on Bio-Image Informatics*, 179–197.
66. Islam, S., Karia, N., Fauzi, F. B. A., & Soliman, M. (2017). A review on green supply chain aspects and practices. *Management & Marketing. Challenges for the Knowledge Society*, 12(1), 12–36.
67. Evangelopoulos, N., Zhang, X., & Prybutok, V. R. (2012). Latent semantic analysis: Five methodological recommendations. *European Journal of Information Systems*, 21(1), 70–86.
68. Feldman, R., & Sanger, J. (2006). *The Text Mining Handbook: A Practical Guide to Text Mining*. Cambridge University Press.
69. Spärck Jones, K. (2004). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 60(5), 493–502.
70. Ramos, J. (2003). Using TF-IDF to determine word relevance in document queries. In *Proceedings of the First Instructional Conference on Machine Learning* (Vol. 242, No. 1, pp. 29–48).
71. Wu, H. C., Luk, R. W. P., Wong, K. F., & Kwok, K. L. (2008). Interpreting TF-IDF term weights as making relevance decisions. *ACM Transactions on Information Systems (TOIS)*, 26(3), 1–37.
72. Morel, C. M., Serruya, S. J., Penna, G. O., & Guimarães, R. (2009). Co-authorship network analysis: A powerful tool for strategic planning of research, development, and capacity building programs on neglected diseases. *PLoS Neglected Tropical Diseases*, 3(8), e501.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

CHAPTER 13

Natural Resource Management and Artificial Intelligence for Sustainable Development

ANKUR KULSHRESTHA,¹ DIKSHA SRIVASTAVA,² M. ANGULAKSHMI,³
SACHIN TRIPATHI,⁴ and M. Z. M. NOMANI⁵

¹*Department of Mechanical and Aerospace Engineering,
NIMS University, Jaipur, Rajasthan, India*

²*Department of Biotechnology and Chemical Engineering,
Manipal University, Jaipur, Rajasthan, India*

³*School of Computer Science Engineering and Information Systems,
Vellore Institute of Technology, Vellore, Tamil Nadu, India*

⁴*Symbiosis Law School, Symbiosis International (Deemed) University
(SIU), WATHODA, Nagpur, Maharashtra, India*

⁵*Faculty of Law, Aligarh Muslim University Aligarh, Uttar Pradesh,
India*

ABSTRACT

Natural resource management is a critical facet of achieving sustainable development in an era marked by increasing population growth, urbanization, and environmental degradation. As humanity faces pressing challenges such as climate change, biodiversity loss, and resource depletion, the integration of artificial intelligence (AI) emerges as a transformative tool to address these issues while promoting sustainable development. This

abstract explores the synergistic relationship between natural resource management and AI, elucidating how AI technologies can play a pivotal role in shaping a more sustainable future. AI technologies, including machine learning, data analytics, and remote sensing, have the potential to revolutionize natural resource management by enhancing our capacity to monitor, analyze, and make informed decisions about our environment. Furthermore, AI-driven predictive models can aid in forecasting resource availability and demand, allowing for more efficient resource allocation and management. In conclusion, the fusion of natural resource management and artificial intelligence holds immense promise for advancing sustainable development goals. As we navigate the intricate web of challenges posed by climate change and resource scarcity, AI emerges as a formidable ally in our pursuit of a more sustainable and equitable future.

13.1 INTRODUCTION

Building intelligent devices that can augment, automate, and expedite important daily tasks that traditionally need human intelligence is the objective of AI, a broad area of computer science. It entails identifying anomalies, identifying patterns, and projecting “future states.” The advancements in computing, technology, and study in the domain of AI have led to a rise in the applications of AI in every category, especially the forest and wildlife industries, as well as in e-commerce, social networks, agriculture, ecosystem sustainability, medical, social care, and urban planning [1–8]. The following variables have fueled the developments: (i) a surge in the accessibility of information at a faster rate has given rise to the possibility of generating significant insights and decision-making; (ii) a decrease in the cost of data storage and computing due to the accessibility of cloud techniques has allowed us to develop intricate solutions at scale to hasten experiments and studies; and (iii) the accessibility of comprehensive data sources due to the accessibility of (a) high-resolution. The effects of AI technology have been inconsistent, favoring industries with substantial economic returns more often than forestry [9, 10] and biodiversity conservation [11–14] with fewer uses. Although the use of AI in the handling of forests and natural resources dates back 30 years [15], forestry research and the use of AI technologies lag other industries like health, transportation, and agriculture [3].

In the 21st century, the pursuit of sustainable development has emerged as a pressing worldwide necessity.

It is abundantly obvious that conventional approaches to resource management are unable to handle the complex problems of our day as our planet struggles to meet the ever-increasing needs of an expanding population and the effects of unrestrained resource depletion [50, 70–74]. Fortunately, there is some optimism in sight thanks to the quick development of AI. The incorporation of AI into natural resource management offers an unrivalled chance to not only alleviate the negative effects of resource exploitation but also to build a future in which mankind coexists peacefully with the earth in the face of rising environmental issues.

Natural resources, encompassing everything from clean water and arable land to minerals and energy sources, are the lifeblood of human civilization. However, the unchecked exploitation and mismanagement of these resources have placed unprecedented stress on ecosystems, contributed to climate change, and threatened the livelihoods of countless communities worldwide. Traditional approaches to resource management have often fallen short, plagued by inefficiencies, inadequate monitoring, and insufficient adaptation to changing environmental conditions.

About 30% of the world's surface area is covered by forests, which are the primary terrestrial ecosystems and home to 90% of the terrestrial biodiversity [16, 17]. As they offer various essential services for sustaining life, like protective functions and environmental services, forests are essential to the healthy operation of our planet. For example, they stabilize greenhouse gases in the atmosphere, so supplying clean, breathable air, and they also absorb up to 30% of the yearly worldwide atmospheric CO₂ emissions [18]. By assisting pollinators, and the hydrological cycle, they significantly contribute to global food security. They provide a significant portion of the world's renewable energy (biofuels) and vital sources of medicinal herbs. They are essential for maintaining the hydrological integrity of different ecosystems and help make people and ecosystems more resilient to catastrophic events like floods and droughts. Approximately 1.6 billion people also rely on trees for their livelihood [19], which is another factor.

Regrettably, forests too encounter numerous difficulties. Due to logging, expanding agriculture, and urbanization, forests are degrading quickly on a global scale. Wildfires and other climate change effects are aggravating the condition of the world's forests [20]. Global forest loss in the past

25 years was 129 million hectares, which decreased the global carbon pool by 17.4 Gt [21]. The present global trend of forest loss and carbon emissions is expected to continue soon. The forest industry still employs conventional management methods despite the advantages the forest offers and the types of risks it faces. In fact, it is one of the few industries with a slow acceptance rate for new technologies [22]. For instance, doing forest inventories still requires using a pen and paper in many nations, especially India. Such conventional approaches have several shortcomings, including the introduction of personal bias, sluggish data collecting and processing, and lack of scalability [23]. Other industries, such as agriculture, which is one of the main causes of deforestation, have adopted technological advancements more quickly than the forestry industry [24]. Organizations like *sagarobotics.com* and *FarmBot*, which utilize robots for exact de-weeding, precise fertilization, or precise pesticide treatments, have contributed to precision agriculture by increasing yields per acre.

In especially at a time when many governments are seeking for ways to bring about transformative change in order to achieve their NDCs, forests, which are distinct from agricultural systems in that they are dynamic, must be treated as such. The inadequacies of traditional data collection and analytics techniques, which are necessary for effective forest management and preservation, are where technologies may make a significant impact [25–28]. The forest industry would greatly benefit from technology's inherent ability to advance knowledge and spread technology across various locations, on several scales, and at a far faster rate.

In contrast, AI, with its capabilities in data analysis, machine learning, and predictive modeling, has emerged as a potent ally in the pursuit of sustainable development. This research paper delves into the transformative potential of AI in natural resource management, elucidating how it can empower decision-makers, enhance environmental conservation efforts, and drive sustainable development across the globe [29].

13.2 LITERATURE REVIEW

AI's ability to process and analyze vast datasets is a game-changer in natural resource management. The collection of environmental data has never been more extensive, thanks to satellites, sensors, and other monitoring technologies. However, without AI, this data is often overwhelming

and underutilized. AI-driven analytics can harness the power of big data to provide real-time insights into ecosystems, weather patterns, and resource availability [30–32]. By doing so, it equips policymakers and resource managers with the information necessary to make informed decisions that balance economic development with environmental preservation.

Moreover, AI's predictive capabilities have the potential to revolutionize resource management strategies. Machine learning algorithms can forecast changes in resource availability and demand, enabling proactive planning and adaptation. For example, in agriculture, AI can predict crop yields and optimize planting and harvesting times, reducing waste, and maximizing productivity. In water resource management, AI can predict droughts and water quality issues, enabling authorities to implement conservation measures and ensure safe drinking water for communities [33].

The integration of AI into natural resource management also holds promise for wildlife conservation and habitat preservation. AI-powered monitoring systems can track the movement of endangered species, detect illegal poaching activities, and even analyze the health of ecosystems [34]. These applications not only protect biodiversity but also contribute to the delicate balance of ecosystems that support life on Earth.

One of the most compelling aspects of AI in sustainable development is its potential to democratize information and decision-making. As AI systems become more accessible and user-friendly, communities and individuals can participate in resource management discussions, advocate for sustainable practices, and hold governments and corporations accountable for their environmental impact [35]. This democratization of knowledge and influence can help ensure that sustainability is a collective effort involving all stakeholders.

In conclusion, the intersection of natural resource management and artificial intelligence is poised to reshape the future of our planet. As we confront the pressing challenges of resource depletion, climate change, and environmental degradation, AI offers a beacon of hope. This research paper explores the multifaceted ways in which AI can catalyze sustainable development, from optimizing resource allocation to preserving biodiversity and fostering inclusive decision-making processes. As we embark on this transformative journey, we must recognize the immense potential of AI while remaining vigilant about ethical considerations and unintended consequences [36–39]. The path to sustainable development is illuminated

by the promise of AI, but its success will ultimately depend on our collective wisdom and commitment to a more harmonious coexistence with the natural world.

While the intersection of natural resource management and AI has gained increasing attention in recent years, there exists a notable research gap that demands further exploration to effectively harness AI for sustainable development in the context of natural resource management. Specifically, there is a need for more comprehensive investigations into the following key areas: Most existing research in this domain focuses on high-income countries or regions with robust technological infrastructures. A significant research gap exists in understanding how AI can be adapted and made accessible in resource-constrained regions, where sustainable development challenges are often more acute. Bridging this gap is crucial to ensure that AI-driven solutions can truly benefit the global effort towards sustainability [40]. The integration of AI into natural resource management often requires interdisciplinary collaboration between experts in AI, environmental science, and policy-making. A research gap exists in identifying effective models of interdisciplinary collaboration and understanding the challenges and opportunities associated with such collaborations. Evaluating the long-term impact of AI applications in natural resource management is essential. Research should aim to develop methodologies for assessing the sustainability and effectiveness of AI-driven solutions over extended periods, ensuring that short-term gains do not lead to unintended negative consequences.

13.3 METHODOLOGY

This research aims to investigate the role of AI in the domain of natural resource management for achieving sustainable development goals. The study will employ a mixed-method research approach, combining qualitative and quantitative methods to provide a comprehensive understanding of the subject matter. The research paper aims to investigate the role of AI in enhancing natural resource management for sustainable development. This research methodology outlines the approach and methods that will be used to collect, analyze, and interpret data to address the research objectives.

1. **Research Design:** This study will adopt a mixed-methods research design, which combines both quantitative and qualitative research approaches to provide a comprehensive understanding of the topic. The quantitative aspect will encompass data collection through surveys and statistical analysis, while the qualitative aspect will utilize interviews and content analysis.
2. **Data Collection:** a. **Surveys:** Online surveys will be distributed to a diverse sample of professionals, policymakers, and experts in the fields of natural resource management and AI. The survey questionnaire will be designed to gather quantitative data on perceptions, opinions, and experiences related to AI and sustainable development. In-depth semi-structured interviews will be conducted with key stakeholders, involving government officials, AI experts, and environmentalists. These interviews will provide qualitative insights into the practical applications, challenges, and opportunities of AI in natural resource management. Relevant literature, reports, and documents related to AI applications in natural resource management and sustainable development will be systematically reviewed and analyzed to supplement the primary data collected. For the survey, a stratified random sampling technique will be employed to ensure representation from different sectors and geographic regions. Key informants for interviews will be selected purposively, targeting individuals with expertise and experience in the subject matter. **Quantitative Data:** Survey data will be assessed using statistical software like SPSS. Descriptive statistics, like frequencies and percentages, will be used to summarize responses. Inferential statistics, including regression analysis, will be applied to identify relationships and patterns in the data. **Qualitative Data:** Interview transcripts and content from documents will undergo thematic analysis to identify recurring themes, patterns, and key insights. Coding will be done to categorize data, and NVivo or similar software may be used to facilitate the analysis process. Ethical approval will be sought from the relevant institutional review board. Informed consent will be obtained from survey respondents and interviewees, ensuring privacy and data protection. To enhance the credibility and validity of the findings, triangulation will be employed by cross-referencing quantitative and qualitative data. Member checking, where

participants review the findings, will also be conducted to ensure accuracy. Possible limitations include sampling bias, respondent bias, and the rapidly evolving nature of AI technology. To mitigate these, rigorous sampling techniques will be used, and efforts will be made to ensure diversity in the sample. The research findings will be disseminated through a research paper, presentations at conferences, and policy briefs targeting relevant stakeholders. The goal is to contribute valuable insights to the field of natural resource management and sustainable development. This research methodology outlines a comprehensive approach to examine the intersection of AI and natural resource management for sustainable development. By employing both quantitative and qualitative methods, the research aims to offer a holistic understanding of the subject and offer practical suggestions for policymakers and practitioners.

The proposed research methodology aims to provide a rigorous and comprehensive investigation into the integration of AI in natural resource management for sustainable development. By employing a mixed-methods approach, ethical considerations, and data analysis techniques, this study seeks to contribute valuable insights to both academia and practice in these critical fields.

13.4 RESULTS AND DISCUSSIONS

According to the previous guideline, the main data gathering method should be utilized to analyze the outcomes of the usage of artificial intelligence in managing the natural ecosystem in relation to sustainable development [17]. In order to get correct data, this portion of the study is responsible for managing data from the primary data gathering phase 50 responders who oversee operations in the natural ecosystem will participate in the study and provide information.

What are the ways artificial intelligence can help create a sustainable natural ecosystem?

➤ Survey Questions for Research

Question 1: Do you agree that artificial intelligence advances its sustainability level while protecting natural ecosystem (Table 13.1; Figure 13.1).

TABLE 13.1 Application of Artificial Intelligence to Manage Natural Ecosystem

Determinants	Responses Obtained	Response Percentage (%)
Agreed strongly	9	18
Agreed	17	34
Neutral	2	4
Disagreed	13	26
Disagreed strongly	9	18

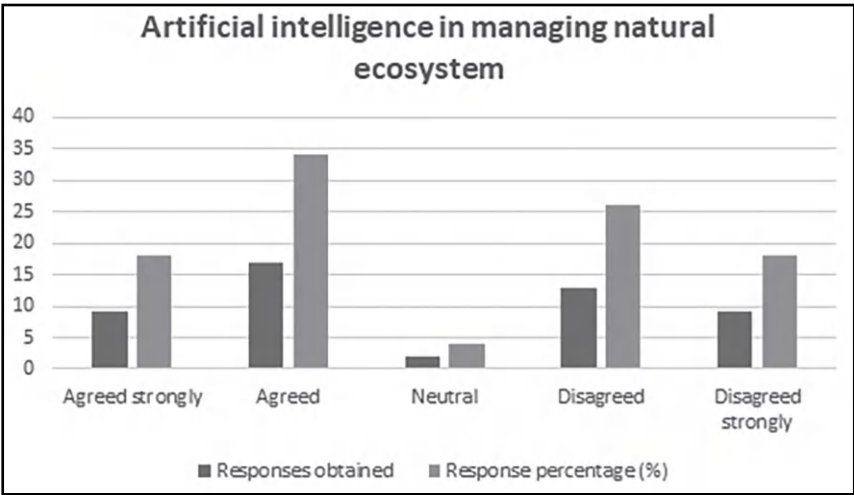


FIGURE 13.1 Managing natural ecosystem with artificial intelligence.

Question 2: Do you agree that artificial intelligence be effectively used to monitor and manage wildlife populations in natural ecosystems? (Table 13.2; Figure 13.2).

TABLE 13.2 Artificial Intelligence to Manage Wildlife Populations

Determinants	Responses Obtained	Response Percentage (%)
Agreed strongly	10	20
Agreed	18	36
Neutral	1	2
Disagreed	14	28
Disagreed strongly	7	14

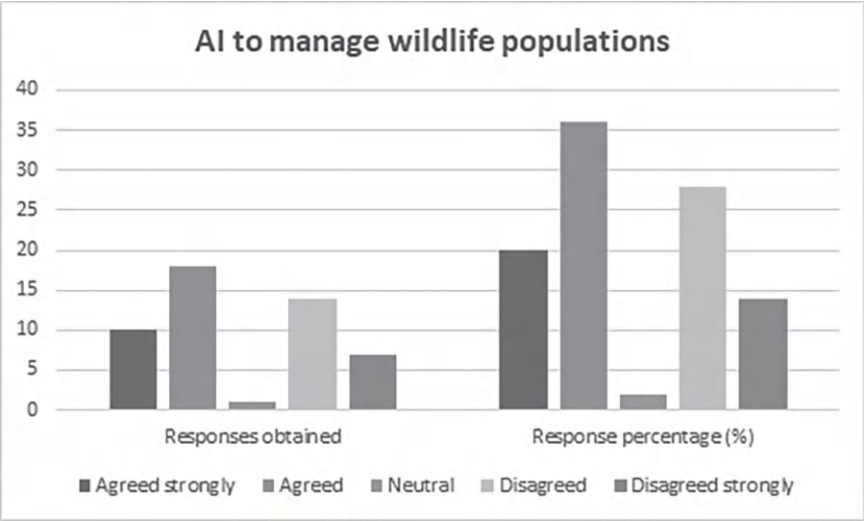


FIGURE 13.2 Managing wildlife populations with artificial intelligence.

The conservation of wildlife and natural ecosystems is a global concern in the face of climate change, habitat loss, and poaching. To address these challenges, there is a growing consensus that artificial intelligence (AI) can play a pivotal role in monitoring and managing wildlife populations. AI offers the potential to revolutionize conservation efforts by providing real-time data, enhancing efficiency, and enabling proactive interventions.

Questions 3: Is AI adoption in natural resource management cost-effective and feasible for developing countries striving for sustainable development? (Table 13.3; Figure 13.3).

TABLE 13.3 Cost-Effectiveness and Feasibility of AI Adoption in Natural Resource Management

Determinants	Responses Obtained	Response Percentage (%)
Agreed strongly	12	24
Agreed	16	32
Neutral	3	6
Disagreed	13	26
Disagreed strongly	6	12

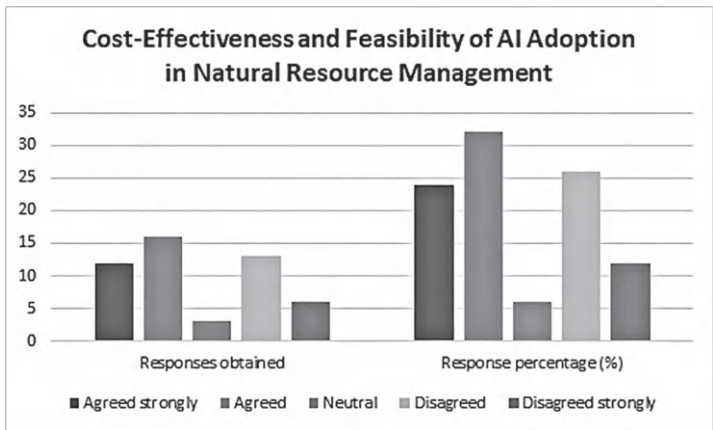


FIGURE 13.3 Cost-effectiveness and feasibility of AI adoption in natural resource management.

Developing countries often face the formidable challenge of balancing economic growth with sustainable development. Natural resource management plays a pivotal role in this delicate equilibrium, as these nations are often rich in natural resources but must also ensure their responsible utilization. AI has emerged as a transformative tool that can aid developing countries in achieving sustainable development goals by enhancing the efficiency, effectiveness, and cost-effectiveness of natural resource management efforts.

Natural resource management plays a pivotal role in achieving sustainable development, as it involves the responsible utilization and preservation of our planet’s finite resources. With the growing global population and increasing demands for resources, it is imperative to employ innovative approaches to ensure a balance between resource exploitation and conservation. AI has emerged as a powerful tool in this endeavor, offering the potential to revolutionize how we manage and protect natural resources for a sustainable future.

AI enables the collection and analysis of vast amounts of data from diverse sources, such as remote sensing satellites, sensors, and field surveys. This data can be used to predict changes in resource availability and environmental conditions [41–44]. For instance, machine learning algorithms can forecast water scarcity, deforestation rates, or soil erosion, allowing governments and organizations to take proactive measures to address these issues. In agriculture, AI-driven precision farming techniques

optimize resource use by precisely managing irrigation, fertilization, and pesticide application. This not only enhances crop yields but also reduces resource waste and environmental pollution. AI-powered drones and autonomous tractors are becoming integral in monitoring and managing agricultural operations efficiently. AI-powered camera traps and acoustic sensors aid in wildlife conservation efforts by automatically identifying and tracking endangered species. This technology assists researchers and conservationists in understanding animal behavior, population dynamics, and migration patterns, enabling targeted conservation strategies [45].

AI can improve the management of renewable energy resources like solar and wind power. AI algorithms can forecast energy production, optimize grid management, and enhance energy storage systems. These advancements are crucial for transitioning to clean and sustainable energy sources. AI-based models can predict and respond to natural disasters more effectively. For instance, AI-driven early warning systems can provide timely alerts for hurricanes, earthquakes, or forest fires, minimizing their impact on human lives and the environment [25, 46–49].

13.5 CONCLUSION

In the pursuit of sustainable development, the integration of Natural Resource Management (NRM) and AI has emerged as a powerful synergy with the potential to reshape our approach to resource conservation, environmental protection, and economic growth [51]. This research paper has delved into the multifaceted relationship between these two domains, exploring the ways in which AI technologies can enhance NRM practices and contribute to the achievement of sustainable development goals.

As our world grapples with the pressing challenges of climate change, biodiversity loss, and resource scarcity, the importance of effective NRM cannot be overstated. Traditionally, NRM has relied on human expertise, which, while invaluable, has its limitations in handling the vast and complex datasets that underpin sustainable resource management [52–57]. This is where AI steps in as a game-changer, offering the ability to process and analyze data at scales and speeds previously unimaginable. Machine learning algorithms, remote sensing technologies, and data-driven decision support systems have enabled NRM practitioners to make more informed choices, optimize resource allocation, and monitor ecological systems with unprecedented precision.

One of the notable contributions of AI in NRM is its role in predictive modeling. By harnessing historical data and utilizing advanced algorithms, AI can forecast environmental changes, habitat shifts, and the potential impact of human activities on natural resources. This predictive capacity empowers policymakers and conservationists to proactively address emerging issues, adapt strategies, and implement interventions that can mitigate or even prevent ecological damage [58]. Furthermore, AI-driven modeling has proven instrumental in disaster preparedness and response, allowing for early warning systems that can save lives and reduce the environmental fallout of natural catastrophes.

The efficient use of resources is a cornerstone of sustainability, and AI plays a pivotal role in optimizing resource management. In agriculture, for instance, AI-driven precision farming techniques can optimize irrigation, fertilizer use, and pest control, resulting in increased crop yields while minimizing environmental harm. In forestry, AI assists in sustainable logging practices, ensuring that deforestation is curbed and reforestation efforts are maximally effective [59]. In the energy sector, AI aids in the optimization of renewable energy production and distribution, making clean energy sources more economically viable.

While the potential benefits of AI in NRM are immense, it is essential to acknowledge the ethical and societal implications. The democratization of AI technologies must be accompanied by safeguards to ensure that marginalized communities are not left behind and that AI applications do not exacerbate existing environmental injustices [60]. Moreover, the use of AI in NRM raises questions about data privacy, algorithmic bias, and the need for robust regulatory frameworks.

In conclusion, the marriage of Natural Resource Management and Artificial Intelligence offers a promising path towards sustainable development. The fusion of human expertise with AI-driven insights empowers us to address environmental challenges with greater precision and efficiency. It allows us to make informed decisions, optimize resource utilization, and adapt to a rapidly changing world. However, as we embrace AI as a powerful tool in NRM, we must do so with a commitment to ethics, inclusivity, and responsible governance [61–63]. Only through a holistic and conscientious approach can we harness the full potential of AI to create a more sustainable and equitable future for generations to come. The journey towards sustainable development is ongoing, and AI is a vital companion on this path, helping us navigate the complex terrain of our planet's natural resources.

Natural resource management (NRM) is a critical component of achieving sustainable development, as it involves the responsible use and conservation of our planet's finite resources. In recent years, AI has emerged as a promising tool to enhance NRM practices, offering potential solutions to complex environmental challenges. While the integration of AI into NRM holds great promise, it is essential to recognize and address its limitations to ensure the success of sustainable development initiatives [64].

- **Data Limitations:** One of the primary challenges in implementing AI for NRM is the availability and quality of data. AI systems heavily rely on large datasets for training and decision-making. In many regions, especially in developing countries, data collection infrastructure may be inadequate, leading to gaps in information [65]. Additionally, biased or incomplete data can result in AI models making inaccurate predictions or recommendations, undermining the effectiveness of NRM efforts.
- **Lack of Accessibility:** AI technologies can be expensive to develop and deploy, limiting their accessibility to certain organizations and regions. Smaller communities and developing countries may struggle to adopt AI solutions for NRM due to financial constraints, technical expertise requirements, and the need for specialized hardware [66]. This accessibility gap can exacerbate existing disparities in resource management and hinder the achievement of sustainable development goals.
- **Ethical and Social Concerns:** The use of AI in NRM raises ethical concerns related to privacy, transparency, and accountability. AI algorithms may inadvertently infringe upon individuals' privacy rights when collecting and analyzing data. Moreover, decision-making processes driven by AI models can be perceived as opaque, making it challenging for affected communities to understand and trust the recommendations [67]. Ensuring ethical AI practices in NRM is essential to maintain public trust and support.
- **Limited Generalization:** AI models trained on specific datasets and conditions may struggle to generalize their findings to new or diverse environments. This limitation is particularly relevant in NRM, where ecosystems vary significantly across regions and

can change rapidly due to climate change and human activities. Over-reliance on AI models without considering local context and expert knowledge can lead to ineffective resource management strategies.

- **Vulnerability to Biases:** AI algorithms can perpetuate and amplify biases present in the data they are trained on. In the context of NRM, biases related to historical resource use, land tenure, and socio-economic factors can result in unjust or unsustainable resource management practices [68]. It is crucial to continuously monitor and mitigate these biases in AI systems to ensure equitable and sustainable development.
- **Environmental Impact:** The environmental impact of AI technologies, including the production and disposal of hardware components, energy consumption, and data storage, can be significant. While AI has the potential to optimize resource management and reduce waste, the sustainability of AI itself should be carefully considered to avoid inadvertently exacerbating environmental challenges.
- **Human-AI Collaboration Challenges:** Achieving the full potential of AI in NRM requires effective collaboration between AI systems and human experts [69]. This collaboration can be hindered by challenges in communication, trust-building, and the integration of AI recommendations into decision-making processes. Striking the right balance between human expertise and AI assistance is an ongoing challenge.

While AI holds promise for revolutionizing natural resource management and advancing sustainable development goals, it is essential to acknowledge and address these limitations. Overcoming data challenges, ensuring accessibility and equity, addressing ethical concerns, and adapting to diverse environments are crucial steps in harnessing AI's potential for the benefit of both the environment and society. By recognizing and actively mitigating these limitations, researchers and practitioners can work towards more effective and sustainable NRM practices through the integration of artificial intelligence.

KEYWORDS

- **artificial intelligence**
- **data limitations**
- **lack of accessibility**
- **natural resource management**
- **sustainable development**

REFERENCES

1. Kim, K., & Park, J. (2009). A survey of applications of artificial intelligence algorithms in eco-environmental modelling. *Environmental Engineering Research*, 14(2), 102–110.
2. Gomes, C., Dietterich, T., Barrett, C., Conrad, J., Dilkina, B., Ermon, S., & Zeeman, M. L. (2019). Computational sustainability: Computing for a better world and a sustainable future. *Communications of the ACM*, 62(9), 56–65.
3. Shi, Z. R., Wang, C., & Fang, F. (2020). *Artificial Intelligence for Social Good: A Survey*. *arXiv preprint arXiv:2001.01818*.
4. Christin, S., Hervet, É., & Lecomte, N. (2019). Applications for deep learning in ecology. *Methods in Ecology and Evolution*, 10(10), 1632–1644.
5. Jha, K., Doshi, A., Patel, P., & Shah, M. (2019). A comprehensive review on automation in agriculture using artificial intelligence. *Artificial Intelligence in Agriculture*, 2, 1–12.
6. Lamba, A., Cassey, P., Segaran, R. R., & Koh, L. P. (2019). Deep learning for environmental conservation. *Current Biology*, 29(19), R977–R982.
7. Imada, A. (2014). A literature review: Forest management with neural network and artificial intelligence. In *International Conference on Neural Networks and Artificial Intelligence* (pp. 9–21). Cham: Springer International Publishing.
8. Liu, Z., Peng, C., Xiang, W., Tian, D., Deng, X., & Zhao, M. (2010). Application of artificial neural networks in global climate change and ecological research: An overview. *Chinese Science Bulletin*, 55, 3853–3863.
9. Khan, S., & Gupta, P. K. (2018). Comparative study of tree counting algorithms in dense and sparse vegetative regions. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 801–808.
10. Fromm, M., Schubert, M., Castilla, G., Linke, J., & McDermid, G. (2019). Automated detection of conifer seedlings in drone imagery using convolutional neural networks. *Remote Sensing*, 11(21), 2585.
11. Metcalf, O. C., Ewen, J. G., McCready, M., Williams, E. M., & Rowcliffe, J. M. (2019). A novel method for using ecoacoustics to monitor post-translocation behavior in an endangered passerine. *Methods in Ecology and Evolution*, 10(5), 626–636.

12. Nay, J., Burchfield, E., & Gilligan, J. (2018). A machine-learning approach to forecasting remotely sensed vegetation health. *International Journal of Remote Sensing*, 39(6), 1800–1816.
13. Burivalova, Z., Game, E. T., & Butler, R. A. (2019). The sound of a tropical forest. *Science*, 363(6422), 28–29.
14. Wood, C. M., Gutiérrez, R. J., & Peery, M. Z. (2019). Acoustic monitoring reveals a diverse forest owl community, illustrating its potential for basic and applied ecology. *Ecology*, 100(9), 1–3.
15. Coulson, R. N., Folse, L. J., & Loh, D. K. (1987). Artificial intelligence and natural resource management. *Science*, 237(4812), 262–267.
16. Novotny, V., Drozd, P., Miller, S. E., Kulfan, M., Janda, M., Basset, Y., & Weiblen, G. D. (2006). Why are there so many species of herbivorous insects in tropical rainforests? *Science*, 313(5790), 1115–1118.
17. Schmitt, C. B., Burgess, N. D., Coad, L., Belokurov, A., Besançon, C., Boisrobert, L., & Winkel, G. (2009). Global analysis of the protection status of the world's forests. *Biological Conservation*, 142(10), 2122–2130.
18. Bellassen, V., & Luyssaert, S. (2014). Carbon sequestration: Managing forests in uncertain times. *Nature*, 506(7487), 153–155.
19. IPCC. (2014). *Climate Change 2014: Synthesis Report—Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Edited by The Core Writing Team, R. K. Pachauri, & L. Meyer. Available at: http://www.ipcc.ch/pdf/assessment-report/ar5/syr/SYR_AR5_LONGERREPORT.pdf (accessed on 17 October 2024).
20. Curtis, P. G., Slay, C. M., Harris, N. L., Tyukavina, A., & Hansen, M. C. (2018). Classifying drivers of global forest loss. *Science*, 361(6407), 1108–1111.
21. Sharma, J. V. (2017). Forestry sector in India is net source of green house gases (GHGs). *Journal of Environmental Science and Engineering Technology*, 5, 2–7.
22. Innes, J. L. (2009). The promotion of ‘innovation’ in forestry: A role for government or others? *Environmental Sciences*, 6(3), 201–215.
23. Näyhä, A., Pelli, P., & Hetemäki, L. (2015). Services in the forest-based sector—unexplored futures. *Foresight*, 17(4), 378–398.
24. Garske, B., Bau, A., & Ekardt, F. (2021). Digitalization and AI in European agriculture: A strategy for achieving climate and biodiversity targets? *Sustainability*, 13(9), 4652.
25. Rykiel Jr, E. J. (1989). Artificial intelligence and expert systems in ecology and natural resource management. *Ecological Modelling*, 46(1, 2), 3–8.
26. Wang, F., Wong, W. K., Ortiz, G. G. R., Al Shraah, A., Mabrouk, F., Li, J., & Li, Z. (2023). Economic analysis of sustainable exports value addition through natural resource management and artificial intelligence. *Resources Policy*, 82, 103541.
27. Wang, Y., & Li, Y. (2023). Chinese economic growth and sustainable development: Role of artificial intelligence and natural resource management. *Resources Policy*, 85, 103996.
28. Stock, M. W., & Rauscher, H. M. (1996). Artificial intelligence and decision support in natural resource management. *New Zealand Journal of Forestry Science*, 26(1/2), 145–157.

29. Davis, J. R., Nanninga, P. M., Hoare, J. R. L., & Press, A. J. (1989). Transferring scientific knowledge to natural resource managers using artificial intelligence concepts. *Ecological Modelling*, 46(1, 2), 73–89.
30. Wang, J., Wang, K., Dong, K., & Zhang, S. (2023). Assessing the role of financial development in natural resource utilization efficiency: Does artificial intelligence technology matter? *Resources Policy*, 85, 103877.
31. Humphries, G. R., Magness, D. R., & Huettmann, F. (2018). *Machine Learning for Ecology and Sustainable Natural Resource Management* (p. 3). Switzerland: Springer.
32. Peng, X., Mousa, S., Sarfraz, M., & Haffar, M. (2023). Improving mineral resource management by accurate financial management: Studying through artificial intelligence tools. *Resources Policy*, 81, 103323.
33. Li, M., Waheed, R., Kirikkaleli, D., & Aziz, G. (2023). Relevance of hybrid artificial intelligence for improving the forecasting accuracy of natural resource prices. *Geoscience Frontiers*, 101670.
34. Fleming, S. W., Garen, D. C., Goodbody, A. G., McCarthy, C. S., & Landers, L. C. (2021). Assessing the new Natural Resources Conservation Service water supply forecast model for the American West: A challenging test of explainable, automated, ensemble artificial intelligence. *Journal of Hydrology*, 602, 126782.
35. Li, A. (2023). An application of extenics, spatial factors, and natural resource market in China: The role of artificial intelligence and geopolitical risk. *Resources Policy*, 81, 103289.
36. Anderson, J., & Evans, M. (1994). Intelligent agent modelling for natural resource management. *Mathematical and Computer Modeling*, 20(8), 109–119.
37. Kala, A. K., & Kumar, M. (2021). Role of geospatial technologies in natural resource management. In *Climate Impacts on Sustainable Natural Resource Management* (pp. 19–34).
38. Saia, S. M., Nelson, N. G., Huseth, A. S., Grieger, K., & Reich, B. J. (2020). Transitioning machine learning from theory to practice in natural resources management.
39. Frey, U. J. (2020). Putting machine learning to use in natural resource management—improving model performance. *Ecology and Society*.
40. Bansal, S., Singh, S., & Nangia, P. (2022). Assessing the role of natural resource utilization in attaining select sustainable development goals in the era of digitalization. *Resources Policy*, 79, 103040.
41. Schmoltdt, D. L., & Rauscher, H. M. (2012). *Building Knowledge-Based Systems for Natural Resource Management*. Springer Science & Business Media.
42. Katkani, D., Babbar, A., Mishra, V. K., Trivedi, A., Tiwari, S., & Kumawat, R. K. (2022). A review on applications and utility of remote sensing and geographic information systems in agriculture and natural resource management. *International Journal of Environment and Climate Change*, 12(4), 1–18.
43. Reddy, G. O., Raval, M. S., Adinarayana, J., & Chaudhary, S. (2022). *Data Science in Agriculture and Natural Resource Management*. Springer.
44. Xiang, X., Li, Q., Khan, S., & Khalaf, O. I. (2021). Urban water resource management for sustainable environment planning using artificial intelligence techniques. *Environmental Impact Assessment Review*, 86, 106515.

45. Dunster, K. (2011). *Dictionary of Natural Resource Management*. UBC Press.
46. Gawdiya, S., Malik, N., Dohling, P. N. K., & Yadav, P. (2022). Chapter 6: Role of artificial intelligence for natural resource management. In Dr. R. K. Naresh (Ed.), *Natural Resource Management* (pp. 83).
47. Trébuil, G., Bousquet, F., Baron, C., & Shinawatra-Ekasingh, B. (2002). Collective creation of artificial worlds can help govern concrete natural resource management problems: A Northern Thailand experience.
48. Stone, N. D., & Engel, B. A. (1990). Knowledge-based systems in agriculture and natural resource management. *AI Magazine*, 11(3), 20–20.
49. Prasad, R., & Sinha, A. K. (2009). Role of expert system in natural resources management. *Geospatial Application Papers on Natural Resource Management*. Retrieved from: <http://www.gisdevelopment.net/application/nrm/overview/ma03130.htm> (accessed on 17 October 2024).
50. Saputra, M. H., & Lee, H. S. (2019). Prediction of land use and land cover changes for North Sumatra, Indonesia, using an artificial-neural-network-based cellular automaton. *Sustainability*, 11(11), 3024.
51. Krishnan, S. R., Nallakaruppan, M. K., Chengoden, R., Koppu, S., Iyapparaja, M., Sadhasivam, J., & Sethuraman, S. (2022). Smart water resource management using artificial intelligence—A review. *Sustainability*, 14(20), 13384.
52. Arnold, G. W., & De Wit, C. T. (1976). *Critical Evaluation of Systems Analysis in Ecosystems Research and Management*. PUDOC.
53. Loehle, C. (1987). Applying artificial intelligence techniques to ecological modeling. *Ecological Modeling*, 38(3, 4), 191–212.
54. Epp, H., Kalin, M., & Miller, D. (1988). PC software for artificial intelligence applications. *Science*, 240(4853), 824–830.
55. Li, A. (2023). An application of extenics, spatial factors, and natural resource market in China: The role of artificial intelligence and geopolitical risk. *Resources Policy*, 81, 103289.
56. Manu, D. S., & Thalla, A. K. (2017). Artificial intelligence models for predicting the performance of biological wastewater treatment plant in the removal of Kjeldahl nitrogen from wastewater. *Applied Water Science*, 7, 3783–3791.
57. Nasser, A. A., Rashad, M. Z., & Hussein, S. E. (2020). A two-layer water demand prediction system in urban areas based on micro-services and LSTM neural networks. *IEEE Access*, 8, 147647–147661.
58. Navarathna, P. J., & Malagi, V. P. (2018). Artificial intelligence in smart city analysis. In *2018 International Conference on Smart Systems and Inventive Technology (ICSSIT)* (pp. 44–47). IEEE.
59. Agrawal, A., Cashore, B., Hardin, R., Shepherd, G., Benson, C., & Miller, D. (2013). Economic contributions of forests. *Background Paper*, 1, 1–127.
60. Bowler, D. E., Buyung-Ali, L. M., Healey, J. R., Jones, J. P., Knight, T. M., & Pullin, A. S. (2012). Does community forest management provide global environmental benefits and improve local welfare? *Frontiers in Ecology and the Environment*, 10(1), 29–36.
61. Lund, J. F., Rutt, R. L., & Ribot, J. (2018). Trends in research on forestry decentralization policies. *Current Opinion in Environmental Sustainability*, 32, 17–22.

62. Ziran, Z. (1999). Natural resources planning, management, and sustainable use in China. *Resources Policy*, 25(4), 211–220.
63. Chen, P., Xu, H., Chen, S., Shang, Y., Zhang, B., & Zhao, X. (2022). Developing a system framework for China's natural resources balance sheet from the perspective of sustainable development. *Frontiers in Environmental Science*, 9, 661.
64. Bekun, F. V., Alola, A. A., & Sarkodie, S. A. (2019). Toward a sustainable environment: Nexus between CO₂ emissions, resource rent, renewable and nonrenewable energy in 16-EU countries. *Science of the Total Environment*, 657, 1023–1029.
65. Chatti, W. (2021). Moving towards environmental sustainability: Information and communication technology (ICT), freight transport, and CO₂ emissions. *Heliyon*, 7(10).
66. Chien, F., Anwar, A., Hsu, C. C., Sharif, A., Razzaq, A., & Sinha, A. (2021). The role of information and communication technology in encountering environmental degradation: Proposing an SDG framework for the BRICS countries. *Technology in Society*, 65, 101587.
67. Saenz-de-Miera, O., & Rosselló, J. (2014). Modeling tourism impacts on air pollution: The case study of PM10 in Mallorca. *Tourism Management*, 40, 273–281.
68. Aeknarajindawat, N. (2020). The impact of natural resources, renewable energy, and economic growth on carbon dioxide emissions in Malaysia. *International Journal of Energy Economics and Policy*.
69. Cenek, M., & Franklin, M. (2017). An adaptable agent-based model for guiding multi-species Pacific salmon fisheries management within a SES framework. *Ecological Modelling*, 360, 132–149.
70. Ching, T., Himmelstein, D. S., Beaulieu-Jones, B. K., Kalinin, A. A., Do, B. T., Way, G. P., & Greene, C. S. (2018). Opportunities and obstacles for deep learning in biology and medicine. *Journal of The Royal Society Interface*, 15(141), 20170387.
71. Elith, J., Graham, C. H., Anderson, R., Dudík, M., Ferrier, S., Guisan, A., & Zimmermann, N. E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29(2), 129–151.
72. Le, Q. V. (2013). Building high-level features using large scale unsupervised learning. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 8595–8598). IEEE.
73. LeDell, E., Gill, N., Aiello, S., Fu, A., Candel, A., Click, C., & Malohlava, M. (2020). h2o: R interface for the 'H₂O' scalable machine learning platform. *R package version*, 3(0.1), 2.
74. Paruelo, J., & Tomasel, F. (1997). Prediction of functional characteristics of ecosystems: A comparison of artificial neural networks and regression models. *Ecological Modelling*, 98(2, 3), 173–186.

CHAPTER 14

A Sustainable Development Goal Perspective Using Business Models and Artificial Intelligence

ALPANA SHARMA,¹ VIDHIKA TIWARI,² RAJESH DEB BARMAN,³
KOUSIK BORO,³ GANESH P. SHUKLA,⁴ and P. WILLIAM⁵

¹*Commerce, Bhilai Mahila Mahavidyalaya, Bhilai, Chhattisgarh, India*

²*Department of Mechanical Engineering, DPGITM, Gurugram, Haryana, India*

³*Department of Commerce, Bodoland University, Kokrajhar, Assam, India*

⁴*Department of Industrial and Production Engineering, Guru Ghasidas Vishwavidyalaya (A Central University), Bilaspur, Chhattisgarh, India*

⁵*Research and Development, Department of Information Technology, Sanjivani College of Engineering, SPPU, Pune, Maharashtra, India*

ABSTRACT

In recent years, a great deal of studies has been done on how to develop the basis of an artificial intelligence business model. One matter, though, remained unresolved and was difficult to comprehend. How organizations alter and refine their AI business model architecture to achieve ongoing value generation has been essential for aspirant and progressive managers and researchers alike. Organizations that successfully create value over long periods of time plan, adapt, and restart their business models using

AI technologies to support this value creation. A summary of findings from a study project on continuously growing organizations is provided. This paper classifies three risky competencies, namely, a focus on identifying and utilizing creative business opportunities, a well-adjusted use of resources, and achieving clarity between management, ethos, and employee responsibilities. Taken together, these competencies determine key maneuvering schedules. In our conclusion, we highlight the implications for AI business model research and specialists and offer executives a tool that enables them to replicate and categorize risky issues related to evolving and changing their business model to sustain value generation. In this essay, we attempt to examine the essential components of the commercial Model Framework, which will completely alter the environment of the commercial market.

14.1 INTRODUCTION

The United Nations General Assembly fixed a set of 17 Sustainable Development Goals (SDGs) in 2015 as a way for everyone to communicate the most pressing issues affecting humankind's ability to sustain its social, economic, and environmental systems [31]. They address a wide range of issues, including eradicating impoverishment gender equality, and climatic change.

Organizations have a significant impact on the society in which they operate, and their dedication to the SDGs is essential to the entire success of the Sustainable Development Agenda [32–34]. In fact, the private sector can make a substantial contribution to the SDGs' realization through technological advancement, strategic thinking, and the deployment of specialized resources and talents [35]. In addition, the SDGs give businesses a clear set of objectives that help them identify, prioritize, and plan for a wide range of unfilled stakeholder and consumer requirements. GRI, UN Global Compact, and WBCSD [36] state that “businesses” that integrate their objectives with the SDGs can boost the involvement of consumers, workers, and other stakeholders;” those who do not will be subjected to rising legal, reputational, and other economic concerns.

The company has the potential to open enormous business possibilities with a potential value of \$12 trillion in company savings and income by 2030. Additionally, they will help create 380 million new jobs across four

different economic systems. Following the SDGs is therefore anticipated to reveal a wide range of chances to find untapped market demands, resulting in the creation of innovative business models, competitive benefits, and sustainable development [35, 37, 38].

As a strategy to endorse the attainment of the SDGs, business model innovation has already been used in practice and studied in the research papers [40]. Novel possibilities for sustainable value proposition, value capture, and value creation can then initiate or be initiated by this process [41–43]. As a result, the SDGs can, on the one hand, offer beneficial guidance that spurs the development of sustainable business models [44]. On the other side, innovative business models can be a major force in bringing about the change required to achieve the SDGs [45, 46].

14.2 LITERATURE REVIEW

In recent years, artificial intelligence (AI) has advanced from being a purely theoretical field to potentially finding applications in many different fields and aspects of daily life [9]. AI is especially useful for difficult judgments that either need a lot of physical effort at a high cost or are either too advanced for standard traditional programming [10]. AI can be used, for instance, to do workforce and data analytics, which are mentioned as crucial issues that firms must address to remain competing and in a better economic condition. According to eight different papers, using workforce analytics more efficiently can help managers and other company leaders accomplish their strategic and operational objectives more successfully [11].

Academic researchers have returned to the study of AI in the last 50 years. Developing a machine operate in ways that would be considered intelligent if a human being did so is what the Dartmouth Research Project classified as AI [12].

Since external input must be appropriately interpreted and used to achieve specified goals and objectives by a flexible configuration [13]. Although connected, the Internet of Things (IoT) and Big Data are distinct concepts from AI. While Big Data encompasses data acquired by any method, the IoT permits the gathering of external data to be used as input for AI [15, 16]. Additionally, because humans possess “cognitive, emotional, and social intelligence,” intelligent systems may accurately mimic human behaviors [17]. In a similar vein, there are various ways to

access AI and machine learning [18]. Even so, the goal is to deliver and operate intelligent goods, and engagements through the exchange of data in order to foster collaboration or generate the most valuable and long-lasting value [19]. Though AI is still in its beginnings, it is challenging to forecast what will happen to it in the future. The world must consider AI demands and standards, such as regulation, employment, ethics, education, entente, and development, in order to better understand and apply AI.

As a result, AI has altered more than only how information is produced and used for decision-making [20]. AI has also transformed how businesses are conducted [21], altering trade methods in various enterprises that provide progressively competing and sustainable goods and services [22–27]. In fact, the relationship between AI and human intelligence is dependent on algorithms that should aid managers in making the best choices, leading to a cultural shift in which numerous data points, linkages, and encounters are incorporated into the typical management of companies [28]. These mathematical frameworks simplify the work of managers. Previous studies have even shown that these methods are often more successful than human assessments owing to their neatly sorted and structured knowledge sets [29]. Although AI has implications to enterprises, that is, intelligence operations. Sousa, and Rocha [30] established a method of talents that disruptive company managers ought to emerge: creativity, authority, and administration.

The more we use AI to implement SDGs, including autonomous vehicles [1], healthcare applications [2], and smart electrical grids [3], the more crucial it is to fund the AI safety research required to keep these systems reliable and helpful, and to avoid them from breaking down or being compromised [4]. A crucial study field for a secure deployment of AI is the identification of crises that might be enabled by a systemic defect in AI technology. Based on a recent world economic forum (WEF) study [5], one such risk is the application of AI in the banking sector. Therefore, in a society that depends increasingly on this innovation, it is critical to promote awareness of the risks associated with probable AI system failures. Additionally, even though we were able to locate many studies indicating that AI may be able to facilitate the achievement of many SDG targets and indicators, most of these investigations [6–8] were performed in controlled laboratory settings, used small datasets, or were prototype-based. As an outcome, extrapolating this data to evaluate the consequences in the real world is frequently difficult.

14.3 RESEARCH METHODOLOGY

In this part, we outline our strategy for achieving the SDGs dependent on surveys that rank various digital techniques. We found some fundamental details regarding the framework that has been employed to evaluate AI-driven DT preference correlations in recent literature [47–49]. Then, we specify the instruments to be used in developing the survey instrument. Lastly, in order to develop our method and capture decision-making that ranks various digital techniques to accomplish the SDGs of Germany and Poland individuals, we selected to use the suggested paths of action for AI-driven DT given in [50] as a reference.

In scientific research, the phase of instrument design and adaption is typical for carrying out a study to address a research issue [51]. This research, which was quantitative in nature, made use of a questionnaire because of its acceptable quality, capacity to assess the parameters for which it was developed, and adaptability to the topic. In order to assist this stage and serve more people while lowering expenses and response time, Survey Monkey was utilized as an online digital medium. The availability to reach more individuals was taken into consideration.

The platform offered by Survey Monkey has the benefit of enabling us to have an infinite number of respondents at various degrees of response.

Its interaction with statistical software also makes it simple to exchange data. The questionnaire was divided into two sections: the first asked for the significance of each SDG in terms of its impact on the economy, society, and environment; the second asked for preferences for AI-driven DT. Prior to starting the survey, respondents received explanations of the SDGs and AI-driven DT terminologies. To acquire comments for decision-making estimates, our suggested survey was easier and quicker.

There was no need to enter any personal information because the surveys were done independently. To obtain the survey's outcomes, respondents might opt to enter their email address. A website visit invitation campaign was started, targeting academics and professionals from various industries with varying levels of AI-driven DT expertise and knowledge. We also got in touch with organizations with solid ties.

Respondents in the poll that started in June 2021 and ended in October 2021 are given terminology and details on the SDGs and AI-driven DT in the questionnaire. Sampling is a crucial stage in the study procedures since

it informs the validity of the inferences the investigators draw and should produce enough information about the occurrence of interest to permit a rich terminology, expanding both the validity [52]. Since G-Power software is user-friendly and covers a variety of projects, we determined the sample size using it [53] regarding get high-efficient information in terms of absolute numbers.

Once we had enough responses for each survey, we classified the SDGs according to each dimension to comprehend their importance in the two countries using an upward Likert-scale, where 1 represented the least priority level and 5 represented the highest. Using a scale of very low to very high contribution, we sequentially recognized each AI-driven DT in each SDG. From there, we tallied all the data to figure out the mean scores for each DT regarding its performance and their choice to endorse each SDG. The Levene's test was performed to evaluate the equality of variances to prevent homoscedasticity [54], corroborating our investigation and opinions regarding the suggestions. Nevertheless, these mean scores differed amongst the two groups.

14.4 RESULTS AND DISCUSSION

The SDGs that have been presented as priorities in each dimension (Table 11.1) have some of the highest mean score (MS) rates. We have a relatively even distribution of individual votes since the variance (F) rates between the scores correspond to the mean variances among Germany and Poland. Industry, Development, and Framework is chosen as the most important SDG for both countries in terms of the commercial aspects. "Decent Work and Economic Growth" was Germany's top priority in terms of the economy, while "Good Health and Well-Being" came in second. In Poland, ("Zero Hunger") came in second place, next to SDG3 ("Good Health and Well-Being"). "Partnerships for the Goals" and "Peace, Justice, and Solid Institutions" were the two SDGs that received the most priority in both countries when it came to the social dimension. "Life below Water" and "Life on Land" were Poland's and Germany's top environmental priorities, respectively. For both nations, "Climate Action" came in third. There were no significant differences in the t scores, which indicate that Poland and Germany have somewhat similar SDG objectives.

TABLE 14.1 SDG Ranking Priority Based on *T* Scores for Dimensions

Dimension	SDG	Germany 1	Poland 2	F 3
Economic Dimension	SDG1: No Poverty	1.621	1.645	2.578
	SDG2: Zero Hunger	1.568	3.785	2.946
	SDG3: Good Health and Well-Being	2.421	1.568	0.538
	SDG8: Decent Work and Economic development	3.689	2.849	1.429
	SDG9: Industry, development, and Infrastructure	1.976	2.378	2.589
Social Dimension	SDG4: Quality Education	1.639	1.639	1.978
	SDG5: Gender Equality	2.578	3.475	2.745
	SDG10: Minimized Inequalities	2.378	2.737	1.986
	SDG11: Sustainable Cities and Communities	1.589	2.723	2.078
	SDG16: Peace, Justice, and Solid Organizations	2.674	2.942	1.872
Environmental Dimension	SDG17: Partnerships for the Objectives	1.531	3.648	2.678
	SDG6: Clear Water and Sanitation	1.438	1.237	3.738
	SDG7: Affordable and Clean Energy	2.273	2.956	3.785
	SDG12: Responsible Production and Consumption	3.937	2.903	2.856
	SDG13: Climate Action	2.863	5.283	1.875
	SDG14: Life Below Water	2.192	2.638	2.623
	SDG15: Life on Land	3.973	2.289	1.629

There was an affinity for three of the five AI-driven DT suggestions to approach the targets underpinning this objective, with a large MS variance. This implies that AI-driven DT aspects fusing and regulating the three tactics would be regarded as well-liked comprehensive remedies: (i) that offer advice to families on financial and economic literacy; and (ii) AI that uses big data to analyze population economic data [55] (*see* Table.14.2).

TABLE 14.2 AI-Driven DT Preference *t* scores in SDG1

AI-Driven DT	Germany 1	Poland 2	F 3
Machine learning in government data governance and poverty-fighting	3.579	1.689	2.487
Assessing digital channels for the purpose of measuring poverty	1.273	1.682	3.645
AI leveraging big data for population economic data analysis	2.289	3.684	2.587
Informational advisory services for families' financial and commercial education	2.957	2.582	1.789
Free online resources backed by personal money management	2.273	2.579	2.579

A generally symmetrical distribution of individual votes is the consequence of the related variance, which is the most popular alternative. This variance does not indicate a very strong inclination for or against the possibilities. Even if SDG2.2 and SDG2.3 were met in Poland, there were just enough variances between the last two possibilities to make them stand out. Respondents expressed a high interest in learning about healthy eating standards, without considering any cultural elements, food availability, or special eating habits.

In the five suggested choices (*see* Table 14.3) for achieving “Good health and well-being for all,” there is no clear hierarchy of choices among the nations. The use of virtual reality technologies and mobile apps for teaching medical workers and healthcare prediction and diagnosis with AI were less well-liked in the health landscape. An intermediate choice was given to free public digital apps that provide personalized advice for healthy behaviors. This result was expected given that the COVID-19 pandemic emerged in populations that exhibited specifically inconsistent and disorganized behavior [56], especially in Germany, in relation to regulating infectious diseases, and that treatment was uniform across many nations.

TABLE 14.3 AI-Driven DT Preference *t* Scores in SDG3

AI-Driven DT	Germany 1	Poland 2	F 3
Artificial intelligence-based diagnostics and medical forecasting	1.478	3.266	1.689
Using mobile apps and virtual/augmented reality technology to train healthcare personnel	1.378	3.278	1.279
Public digital applications offering individualized suggestions for healthy behaviors	1.289	3.289	0.678
Technologies that provide information on infectious disease detection, prevention, and safety	2.168	1.678	0.256
Multiple sources of medical and health data homogenized via big data	2.276	3.092	0.378

The proposals for SDG 4 are shown in Table 14.4. Two well-liked suggestions were noted. It is ideal to use technologies that connect with more social strata and combine expenditures to offer less expensive or even free education. Employing the same logic of universal education, the second choice, is something that is highly related to national realities, particularly considering how prevalent learning concepts are currently and how “inclusive education” is a top priority for many learning specialists [47]. Big Data for managing student data for public education policy

decision-making and conversation facilitators for administration and teaching effectiveness evaluation occupied the final spots.

TABLE 14.4 AI-Driven DT Preference *t* scores in SDG4

AI-Driven DT	Germany 1	Poland 2	F 3
Open and free educational portals with suggested materials to promote lifelong learning	1.647	1.542	1.478
Big Data processing student information for public education policy decision-making	3.589	3.675	1.635
Conversation facilitators for administration and teaching effectiveness evaluation	1.637	3.896	1.945
AI-based personalized learning and inclusive classrooms	3.529	1.537	2.756
Utilizing mobile applications and virtual/augmented reality technologies to transmit technical material and enhance teaching at various training levels	2.723	2.658	3.589

Open DT for discussions on more effective legislation in defense of girls in Germany were the suggestions made for this SDG that were highly favored. The options centered on citizens’ readiness to prevent social and legal discrimination. The second and third options, accordingly, were AI deploying tracking digital networks to discover gender bias in hiring processes and deliberate bureaucracy, more directed to discrimination in work environments. It appears that political initiatives, social manifestations, new laws, and education are insufficient. The concern is whether AI-driven DT will be able to eradicate people's biases [58]. Consequently, all forms of AI-driven DT are valid, independent of the scores related here between Germany and Poland (see Table 14.5).

TABLE 14.5 AI-Driven DT Preference *t* Scores in SDG5

AI-Driven DT	Germany 1	Poland 2	F 3
Girls and women can receive specialized training in emerging techniques online.	1.678	3.268	1.268
Blockchain technology and AI are being used to identify offensive or discriminatory behavior and information directed at girls and women on social media.	3.459	4.279	2.378
AI tracking digital networks to find gender prejudice in hiring procedures or deliberate bureaucracy	2.548	3.280	4.282
Conversations about better laws to protect women and girls should be open on DT.	2.469	1.792	3.876
Using AI and Virtual Reality, we can help women and girls achieve mental and commercial independence.	2.534	6.728	2.376

It is widely accepted that AI-driven DT is used to solve obstacles, and its concept has been examined from a variety of perspectives [13]. Its future will depend on our ability to build solutions for each problem that is discovered and to navigate the significant ethical, legal, and societal challenges that arise from these applications of AI-driven DT [13, 39].

To determine whether AI methods and related digital technologies merit more consideration and funding in order to help Germany and Poland achieve the SDGs, various AI technologies and correlated digital techniques were examined. The results showed that, despite the SDG priority classification appearing to be distinct; there were similar elements in the AI-driven DT choices across the countries.

Concerning the choices of AI-driven DT, digital educational platforms have made a significant effort to revolutionize various disciplines in order to achieve the SDG, in part due to its extensive reach among populations in various societal strata [57, 61]. This data is a resource that may be used to inform future decisions, anticipate variations, and recurrently monitor the development of SDG indicators.

The respondents also favored using big data [62, 63] to develop new policies depending on the data at hand. It is evident that both public and private institutions must make significant efforts to ensure data accuracy for better decision-making in order to process and analyze a big amount of information. Depending on its distributed methodology and data integrity properties, blockchain is often cited as a helpful technology for governability [64, 65]. The similar trend of fewer contributing techniques is followed by Digital Twins, which may be less well recognized among respondents since they do not find it as fascinating as other technologies.

14.4.1 CHALLENGES

The integration of AI-driven digital technologies into the pursuit of Sustainable Development Goals (SDGs) represents a promising avenue for addressing complex global challenges. Germany and Poland, two European countries with distinct economic and technological landscapes, offer valuable perspectives on the challenges associated with this endeavor.

- 1. Data Privacy and Security Concerns:** One of the foremost challenges in implementing AI for SDGs is ensuring data privacy and security. Germany has stringent data protection regulations, such as the General Data Protection Regulation (GDPR), which

can pose hurdles to data-sharing necessary for AI models. Poland also has data protection laws but may face different challenges in enforcement and compliance. Balancing the need for data accessibility with privacy concerns is a persistent challenge.

2. **Data Quality and Availability:** Effective AI models require large and high-quality datasets. Both Germany and Poland face challenges in obtaining comprehensive and reliable data for various SDG-related domains. Inconsistent data collection methods, incomplete datasets, and data silos hinder AI applications.
3. **Resource Allocation:** Integrating AI technologies into SDG initiatives requires significant financial and human resources. Germany, as a developed nation, may have more resources at its disposal, but ensuring equitable distribution and access to AI-driven solutions within the country is still challenging. In Poland, resource constraints might be more pronounced, making it challenging to adopt and deploy AI technologies effectively.
4. **Technological Infrastructure:** The deployment of AI technologies relies on robust technological infrastructure. Germany generally has a more advanced infrastructure compared to Poland. Ensuring that Poland has the necessary infrastructure to support AI-driven SDG initiatives is a significant challenge, requiring investments in both urban and rural areas.
5. **Ethical and Bias Concerns:** AI systems can inadvertently perpetuate biases present in training data. Ensuring fairness and equity in AI-driven SDG implementations is essential. Both countries need to address the ethical implications and biases that may arise from AI systems and establish guidelines for responsible AI use.
6. **Skills Gap:** Germany may have a more developed pool of AI talent, but the demand for skilled professionals still exceeds the supply. Poland faces a similar issue, and both countries need to invest in education and training programs to bridge the AI skills gap.
7. **Interdisciplinary Collaboration:** Achieving the SDGs requires collaboration across various sectors and disciplines. Effective coordination between government agencies, academia, private sector, and civil society is challenging but crucial for leveraging AI for sustainable development.

8. **Regulatory Frameworks:** Developing and adapting regulations to accommodate AI technologies can be a slow and complex process. Both countries need to establish clear guidelines and regulations for the use of AI in SDG projects while allowing flexibility for innovation.
9. **Monitoring and Evaluation:** Assessing the impact of AI-driven interventions on SDGs is essential for accountability. Developing robust monitoring and evaluation frameworks that capture both intended and unintended consequences is a challenge in both countries.
10. **Digital Divide:** Ensuring that the benefits of AI-driven technologies reach marginalized and underserved populations is a pressing challenge. Addressing the digital divide within and between countries is essential to avoid exacerbating inequalities.

In conclusion, the integration of AI-driven digital technologies into SDG implementation is a complex and multifaceted process that presents several challenges. Germany and Poland, while facing different contexts, share common obstacles related to data, ethics, resources, and regulation. Overcoming these challenges requires a concerted effort, collaborative approaches, and a commitment to harnessing the potential of AI for sustainable development while addressing its limitations and risks.

14.5 CONCLUSION

In this study, numerous AI-driven DT suggestions are ranked in relation to Brazil and Portugal's execution of the UN's SDGs. These suggestions came from a prior study that examined the specialized literature to determine the most recent developments in the use of those technologies in the SDG sector. This research can serve as assistance for policymakers in promoting DT to support SDGs and in developing better expenditure plans will spur more development for SDG accomplishments.

We emphasize that the rankings of the suggestions are only a way of comparison within and among the examined countries, thus the most recent preferred technology should not be disregarded. We looked at which DT contributed the most to the top-ranked suggestions among the 17 ranks.

These AI-driven DT suggestions were created as basic approaches to advance the SDGs while considering citizen viewpoints after an extensive review of the research in the field. This material should be helpful for individuals who are eager to introduce innovative DT in their nations. The value of AI-driven DT in educating people about the SDGs, indicating that it is worthwhile to invest in this entire family of technologies in order to increase the likelihood that Earth will survive.

We stress once again how crucial it is for businesses, governments, and individuals to monitor the battle for more AI-driven DT expenditures to the public, permitting remedies to adhere to the SDG 2030 agenda—particularly for healthier manufacturing, which has a big influence on further economic dimensions. Furthermore, this study may contribute the progress of the digital technology concept from a distinctive viewpoint and serve to elucidate current technological developments related AI-driven DT to promote SDGs in conjunction with institutions and research centers.

Because of their socioeconomic settings and criteria, which are limited to Germany and Poland, this research is not without limits. The insights of offering AI-driven DT inclinations to achieve the SDG may seem excessively determined and other types of input, like qualitative studies with AI-driven DT innovators focusing on these quantitative outcomes with policymakers. The collection may serve as a buffer to lessen organizational bottlenecks in creating procedures even though it does not encompass every social level. For this reason, we advise its replication in other nations, particularly in those that are making fewer efforts to achieve the SDGs, in order to make it easier for people to understand how AI-driven DT will be used and implemented.

In the pursuit of achieving the Sustainable Development Goals (SDGs), the integration of artificial intelligence-driven digital technologies has emerged as a transformative force. Germany and Poland, as dynamic and forward-thinking nations within the European Union, have made significant strides in harnessing the power of AI to address pressing global challenges. In this perspective, we have explored the ways in which AI-driven digital technologies are contributing to the implementation of the SDGs in both countries. As we conclude our examination, it is evident that AI holds tremendous promise and potential for shaping a sustainable future, but also poses unique challenges and considerations.

Firstly, it is essential to acknowledge the substantial progress made in Germany and Poland in utilizing AI for sustainable development. Both

countries have demonstrated a commitment to leveraging AI in areas such as healthcare, transportation, energy, and agriculture, aligning with various SDGs. In Germany, initiatives like Industry 4.0 have revolutionized manufacturing processes, improving efficiency and reducing environmental impacts. Poland, on the other hand, has focused on utilizing AI in healthcare systems, which has not only enhanced patient care but also contributed to SDG 3 (Good Health and Well-being).

One of the noteworthy aspects of AI implementation in Germany and Poland is the emphasis on collaboration between governments, industry, academia, and civil society. Multi stakeholder partnerships have played a crucial role in driving innovation and ensuring that AI technologies align with sustainable development objectives. This collaborative spirit has fostered a conducive ecosystem for AI research, development, and deployment, ultimately accelerating progress towards the SDGs.

However, challenges and considerations must not be overlooked. One pressing concern is the ethical and regulatory framework surrounding AI. Ensuring that AI technologies are developed and used responsibly is paramount. Both Germany and Poland must continue to refine their regulatory landscapes to address issues related to bias, transparency, accountability, and data privacy. Ethical AI practices should be at the core of their strategies to prevent unintended consequences that may hinder sustainable development efforts.

Furthermore, the digital divide remains a significant challenge, not only within these countries but also globally. While AI has the potential to drive progress, it must be accessible to all segments of society. Efforts should be made to bridge the gap, ensuring that marginalized communities can benefit from AI-driven solutions. This inclusivity is in line with the principle of “Leave No One Behind,” a fundamental aspect of the SDGs.

In conclusion, the integration of artificial intelligence-driven digital technologies into the pursuit of the Sustainable Development Goals is a journey that Germany and Poland have embraced with enthusiasm and dedication. These nations have demonstrated that AI has the power to enhance economic growth, improve living standards, and address environmental challenges, all while advancing social equity and justice.

As the implementation of AI continues to evolve, both countries should remain committed to fostering collaboration, ethical practices, and inclusivity. By doing so, they can unlock the full potential of AI in achieving the SDGs, leading not only by example within Europe but also inspiring

the global community to harness the transformative power of AI for a more sustainable and equitable world.

In a rapidly changing landscape where technology intersects with global challenges, Germany and Poland stand as beacons of progress, showcasing how artificial intelligence-driven digital technologies can be harnessed to build a more prosperous, sustainable, and inclusive future for all. It is our hope that their experiences and successes will serve as valuable lessons for nations worldwide, motivating them to embark on similar journeys towards a brighter, more sustainable tomorrow.

KEYWORDS

- **artificial intelligence**
- **business model**
- **general data protection regulation**
- **sustainable development goals**

REFERENCES

1. Bonnefon, J. F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, 352(6293), 1573–1576.
2. De Fauw, J., Ledsam, J. R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., & Ronneberger, O. (2018). Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature Medicine*, 24(9), 1342–1350.
3. International Energy Agency (IEA). (2017). *Digitalization & energy*. IEA.
4. Russell, S., Dewey, D., & Tegmark, M. (2015). Research priorities for robust and beneficial artificial intelligence. *AI Magazine*, 36(4), 105–114.
5. Uykur, D. (2018). The new physics of financial services: Understanding how artificial intelligence is transforming the financial ecosystem. Part of the Future of Financial Services series. World Economic Forum.
6. Gandhi, N., Armstrong, L. J., & Nandawadekar, M. (2017). Application of data mining techniques for predicting rice crop yield in semi-arid climatic zone of India. In *2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR)* (pp. 116–120). IEEE.
7. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Correction: Corrigendum: Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 546(7660), 686.

8. Cao, Y., Li, Y., Coleman, S., Belatreche, A., & McGinnity, T. M. (2014). Detecting price manipulation in the financial market. In *2014 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFER)* (pp. 77–84). IEEE.
9. Williams, O. (2021). Towards human-centred explainable AI: A systematic literature review. *Master's Thesis*.
10. Boullart, L., Krijgsman, A., & Vingerhoeds, R. A. (2013). *Application of Artificial Intelligence in Process Control: Lecture Notes Erasmus Intensive Course*. Elsevier.
11. Huselid, M. A. (2018). The science and practice of workforce analytics: Introduction to the HRM special issue. *Human Resource Management*, 57(3), 679–684.
12. McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. *AI Magazine*, 27(4), 12.
13. Nilsson, N. J. (1983). Artificial intelligence prepares for 2001. *AI Magazine*, 4(4), 7.
14. Kaplan, A., & Haenlein, M. (2019). Digital transformation and disruption: On big data, blockchain, artificial intelligence, and other things. *Business Horizons*, 62(6), 679–681.
15. Kaplan, A. M. (2012). If you love something, let it go mobile: Mobile marketing and mobile social media 4x4. *Business Horizons*, 55(2), 129–139.
16. Kaplan, A. M., & Haenlein, M. (2016). Higher education and the digital revolution: About MOOCs, SPOCs, social media, and the Cookie Monster. *Business Horizons*, 59(4), 441–450.
17. Haenlein, M., & Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California Management Review*, 61(4), 5–14.
18. Brynjolfsson, E., & McAfee, A. (2017). The business of artificial intelligence: How AI fits into your data science team. *Harvard Business Review*, 98(4), 1–20.
19. Gretzel, U., Sigala, M., Xiang, Z., & Koo, C. (2015). Smart tourism: Foundations and developments. *Electronic Markets*, 25, 179–188.
20. Mikalef, P., Framnes, V. A., Danielsen, F., Krogstie, J., & Olsen, D. (2017). Big data analytics capability: Antecedents and business value.
21. Schneider, S., & Leyer, M. (2019). Me or information technology? Adoption of artificial intelligence in the delegation of personal strategic decisions. *Managerial and Decision Economics*, 40(3), 223–231.
22. Govindan, K., Jafarian, A., & Nourbakhsh, V. (2019). Designing a sustainable supply chain network integrated with vehicle routing: A comparison of hybrid swarm intelligence metaheuristics. *Computers & Operations Research*, 110, 220–235.
23. Kuo, T. C., & Smith, S. (2018). A systematic review of technologies involving eco-innovation for enterprises moving towards sustainability. *Journal of Cleaner Production*, 192, 207–220.
24. Ding, S., Zhang, Y., & Duygun, M. (2019). Modeling price volatility based on a genetic programming approach. *British Journal of Management*, 30(2), 328–340.
25. Garbuio, M., & Lin, N. (2019). Artificial intelligence as a growth engine for health care startups: Emerging business models. *California Management Review*, 61(2), 59–83.
26. Tuan, M. N. D., Thanh, N. N., & Le Tuan, L. (2019). Applying a mindfulness-based reliability strategy to the Internet of Things in healthcare: A business model in the Vietnamese market. *Technological Forecasting and Social Change*, 140, 54–68.

27. Wirtz, B. W., & Müller, W. M. (2019). An integrated artificial intelligence framework for public management. *Public Management Review*, 21(7), 1076–1100.
28. Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, 126, 3–13.
29. Kauppi, K., Salmi, A., & You, W. (2018). Sourcing from Africa: A systematic review and a research agenda. *International Journal of Management Reviews*, 20(2), 627–650.
30. Sousa, M. J., & Rocha, Á. (2019). Skills for disruptive digital business. *Journal of Business Research*, 94, 257–263.
31. United Nations General Assembly (2015). Transforming our world: The 2030 Agenda for Sustainable Development. UN Doc. A/RES/70/1 (September 25, 2015).
32. Kolk, A., Kourula, A., & Pisani, N. (2017). Multinational enterprises and the sustainable development goals: What do we know and how to proceed? *Transnational Corporations*, 24(3), 9–32.
33. Pizzi, S., Caputo, A., Corvino, A., & Venturelli, A. (2020). Management research and the UN sustainable development goals (SDGs): A bibliometric investigation and systematic review. *Journal of Cleaner Production*, 276, 124033.
34. Sachs, J. D., Schmidt-Traub, G., Mazzucato, M., Messner, D., Nakicenovic, N., & Rockström, J. (2019). Six transformations to achieve the sustainable development goals. *Nature Sustainability*, 2(9), 805–814.
35. Scheyvens, R., Banks, G., & Hughes, E. (2016). The private sector and the SDGs: The need to move beyond ‘business as usual.’ *Sustainable Development*, 24(6), 371–382.
36. Rasche, A. (2020). The United Nations Global Compact and the Sustainable Development Goals. In *Research Handbook of Responsible Management* (pp. 228).
37. Muff, K., Kapalka, A., & Dyllick, T. (2017). The Gap Frame: Translating the SDGs into relevant national grand challenges for strategic business opportunities. *The International Journal of Management Education*, 15(2), 363–383.
38. Rosati, F., & Faria, L. G. (2019). Addressing the SDGs in sustainability reports: The relationship with institutional factors. *Journal of Cleaner Production*, 215, 1312–1326.
39. Floridi, L. (2014). Open data, data protection, and group privacy. *Philosophy & Technology*, 27, 1–3.
40. Breuer, H., Lüdeke-Freund, F., & Brick, C. (2018). Business model innovation in the era of sustainable development goals. In *ISPIM Conference Proceedings* (pp. 1–27). The International Society for Professional Innovation Management (ISPIM).
41. Bocken, N. M., Short, S. W., Rana, P., & Evans, S. (2014). A literature and practice review to develop sustainable business model archetypes. *Journal of Cleaner Production*, 65, 42–56.
42. Lüdeke-Freund, F., Carroux, S., Joyce, A., Massa, L., & Breuer, H. (2018). The sustainable business model pattern taxonomy—45 patterns to support sustainability-oriented business model innovation. *Sustainable Production and Consumption*, 15, 145–162.
43. Pedersen, E. R. G., Gwozdz, W., & Hvass, K. K. (2018). Exploring the relationship between business model innovation, corporate sustainability, and organizational values within the fashion industry. *Journal of Business Ethics*, 149, 267–284.

44. Morioka, S. N., Bolis, I., Evans, S., & Carvalho, M. M. (2017). Transforming sustainability challenges into competitive advantage: Multiple case studies kaleidoscope converging into sustainable business models. *Journal of Cleaner Production*, 167, 723–738.
45. Santa, R., Ferrer, M., Bretherton, P., & Hyland, P. (2010). Contribution of cross-functional teams to the improvement in operational performance. *Team Performance Management: An International Journal*, 16(3/4), 148–168.
46. Sarin, S., & Mahajan, V. (2001). The effect of reward structures on the performance of cross-functional product development teams. *Journal of Marketing*, 65(2), 35–53.
47. Alonso, S., Montes, R., Molina, D., Palomares, I., Martínez-Cámara, E., Chiachio, M., & Herrera, F. (2021). Ordering artificial intelligence-based recommendations to tackle the SDGs with a decision-making model based on surveys. *Sustainability*, 13(11), 6038.
48. Alonso, S., Chiclana, F., Herrera, F., Herrera-Viedma, E., Alcalá-Fdez, J., & Porcel, C. (2008). A consistency-based procedure to estimate missing pairwise preference values. *International Journal of Intelligent Systems*, 23(2), 155–175.
49. Herrera-Viedma, E., Herrera, F., & Chiclana, F. (2002). A consensus model for multiperson decision making with different preference structures. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 32(3), 394–402.
50. Palomares, I., Martínez-Cámara, E., Montes, R., García-Moral, P., Chiachio, M., Chiachio, J., & Herrera, F. (2021). A panoramic view and SWOT analysis of artificial intelligence for achieving the sustainable development goals by 2030: Progress and prospects. *Applied Intelligence*, 51, 6497–6527.
51. Ali, S., Hussain, T., Zhang, G., Nurunnabi, M., & Li, B. (2018). The implementation of sustainable development goals in “BRICS” countries. *Sustainability*, 10(7), 2513.
52. Loughran, R. J., Wallbrink, P. J., Walling, D. E., & Appleby, P. G. (2003). Sampling methods. In *Handbook for the Assessment of Soil Erosion and Sedimentation Using Environmental Radionuclides* (pp. 41–57).
53. Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191.
54. Pan, G. (1999). On a Levene type test for equality of two variances. *Journal of Statistical Computation and Simulation*, 63(1), 59–71.
55. Tilak, J. B. (2002). Education and poverty. *Journal of Human Development*, 3(2), 191–207.
56. Cosci, F., & Guidi, J. (2021). The role of illness behavior in the COVID-19 pandemic. *Psychotherapy and Psychosomatics*, 90(3), 156–159.
57. Tagliabue, L. C., Cecconi, F. R., Maltese, S., Rinaldi, S., Ciribini, A. L. C., & Flammini, A. (2021). Leveraging digital twin for sustainability assessment of an educational building. *Sustainability*, 13(2), 480.
58. Noriega, M. (2020). The application of artificial intelligence in police interrogations: An analysis addressing the proposed effect AI has on racial and gender bias, cooperation, and false confessions. *Futures*, 117, 102510.

Different Requirements in Quality of Service Using an Adaptive Network Algorithm

J. JAYAGANESH,¹ K. SURESH KUMAR,² KONDA HARI KRISHNA,³
MOHIT TIWARI,⁴ R. VETTRISELVAN,⁵ and CHETAN SHELKE⁶

¹Department of Computer Science, Government Arts and Science College, Perumbakkam, Chennai, Tamil Nadu, India

²MBA Department, Panimalar Engineering College, Varadarajapuram, Poonamallee, Chennai, Tamil Nadu, India

³Department of CSE, School of Computing, Mohan Babu University, Sree Sainath Nagar, Rangampeta, Tirupati, Andhra Pradesh, India

⁴Department of Computer Science and Engineering, Bharati Vidyapeeth's College of Engineering, Paschim Vihar, Delhi India

⁵AMET Business School, Academy of Maritime Education and Training (AMET University), Chennai, Tamil Nadu, India

⁶Alliance College of Engineering and Design, Alliance University, Bangalore, Karnataka, India

ABSTRACT

The availability of video services online has increased lately. Service suppliers nevertheless struggle to deliver high-quality video broadcasting, though, as a result of the programming effectiveness issue and the significant end-to-end latency variations caused by the multiple paths algorithms

that are now in use. Networks are anticipated to be able to effectively manage the networking units for achieving quality-of-service (QoS) different requirements to address these two issues that have an impact on the standard of video delivery, which is an issue of concern for broadcast services. For centralized routing calculations and real-time networking to indicate upgrading in digital media uses, this paper suggests an active and adaptive multiple paths routing algorithm (AMRA) beneath 3 limitations (package loss, time latency, and bandwidth) that depend on software-described networks (SDN). In comparison to other multiple paths routing concepts, active and AMRA made effective utilization of the most recent broad networking condition data obtained by the Open Flow regulator and actively generated the ideal routes based on the link's real-time condition data. Furthermore, the suggested technique may substantially decrease the controller's computing expense while performing extremely fine stream equilibrium. The empirical outcomes reveal that active and AMRA greatly surpass other present scheduling algorithms in terms of improving the QoS by 35%–70%.

15.1 INTRODUCTION

Quality of service (QoS), congestion stream management, and intermediary networking operations are just a few of the services that the conventional layout of networked computers is unable to provide to the present standards [1]. To fit with the requirements of the future Internet, networked computer design must be completely updated and changed [2].

As broadcast multimedia services like internet protocol television (IPTV) and video-on-demand have become more common, the QoS-adaptive method has become increasingly important in packet-switching networking [3]. With QoS assistance, service suppliers may offer clients QoS supply to satisfy even more stringent service requirements and ensure an appropriate degree of efficiency (e.g., minimal package loss rate) regarding data flows, such as excessive bit rates (throughput) for real-time multimedia applications, minimal latency for voice over IP (VoIP), or multiplayer games that has minimal jitter (delay variance) for video calling or VoIP [4, 5, 7, 8]. Nevertheless, QoS systems still haven't been properly implemented since current ones, like IntServ [9] and Diffserv [10], are built on best-effort networking [11] and don't have a comprehensive

understanding of all available networking sources. As a result, numerous QoS routing projects are nevertheless ongoing.

As video information is in such high request due to the Internet's exponential expansion, a sizable portion of network capacity is typically taken up by video broadcasting [12]. Lots of packets will be lost or even crucial traffic will be disrupted on the link when connection traffic develops across the network [13]. The bandwidth supplies of the additional connections outside of the transportation route are idle when using the conventional single-path routing, making it impossible to fully utilize them when video broadcasting is being transmitted [14]. The innovation of Web connectivity has advanced significantly with the rise of network gadgets, and multiple paths transfer is now feasible due to the abundance of available routes between units [15].

In contrast to single-path routing, multiple paths have become the web's preferred innovation because it can implement traffic engineering through load leveling [16]. Particularly, it has the potential to employ numerous pathways for the transfer of video traffic over the web, efficiently utilizing the idle connection's bandwidth and enhancing the video transfer's dependability [17]. Nevertheless, due to the Web's explosive expansion in video data, the fundamental data network that underpins multimedia services requires significant improvement to prevent becoming a bottleneck for mutating video congestion [18]. It is critically necessary to develop an intriguing network structure with a more knowledgeable and effective video-transmitting network to meet these problems and obstacles [19].

The flaws in conventional networks can be fixed by Open-Flow. The routing's management and sending layers are separated by this [20]. Using the software-described networking (SDN) control system, the division of administration and data transmission enables more complex data administration and tailored applications to be implemented to enhance network efficiency [21, 22]. Additionally, it offers the capability to govern the actions of the whole network and makes the network customizable and application-aware. Because of its adaptability and programming, the novel routing component may be expanded depending on pre-existing algorithms to enable effective video broadcasting to diverse users and satisfy the various QoS needs of the service supply approach [23, 24]. Due to these 3 different requirements, this study suggests an active and adaptive multiple paths routing algorithm (AMRA) for QoS adaptable video single-cast that relies on SDN [25]. This paper offers an innovative

adaptable scheduling method using the Lagrange relaxation-oriented aggregating cost (LARAC) algorithm to find the optimum routes and most effective utilization of networks while taking into account greater than two restrictions, setting it apart from other QoS techniques [26–28].

Unavoidably, there will be an additional charge for single-cast routing (point-to-point) transmitting video from a single user to another user [29]. Determining a practical route between 2 users while keeping in mind the limits is mostly what single-cast routing entails. To accomplish this, the constrained shortest pathway (CSP) method is a broadly applicable method for web-based multimedia applications that takes into account latency as the sole restriction and the two measurements of jitter and package loss as the expense variables [30–32]. The optimum route that meets the latency requirement in various kinds of traffic may not always be computed using this technique [33]. This paper, therefore investigates 3 restriction variables based on known methods that may determine the outcome of a route near to the real value.

15.2 LITERATURE REVIEW

When routing choices are being made in conventional routing, just a single measurement, like package loss, latency, and latency variance, must be taken into account. For a path selected from an initial point to a location, various measures (constraints) including package loss, latency, and latency variance must all be met at the same time in QoS routing [34]. Therefore, in addition to the shortest routes connecting end-to-end nodes, the expense of optimal routes for QoS routing must take into account the aforementioned cumulative measures [35]. A routing issue, nevertheless, is NP-complete if there are greater than 2 QoS-based measures [36].

To address the CSP and multiple-constrained pathway (MCP) issues, a few QoS routing techniques are suggested. Enhanced selection of path algorithms was developed by [37, 38], both of which concentrated on MCP issues. A solution to the delay-constrained lowest-cost (DCLC) issue was suggested by [39] employing the LARACE technique, along with a conceptual minimum constraint. For the CSP issue, [40] offered 2 strategies, however, because of the longer operating duration; the efficiency was not greater than that in [41]. To compute paths of routing and actively adjust network routing, [42] introduced an effective structure for

QoS routing using SDN. This paradigm gathers data about the networking status. This explains why [43]'s efficiency is better than [44] through [45].

While the condition or topology of the networking constantly shifts, it is feasible in SDN networking for the control system to alter the route of data streams among switches and to set various routing principles for certain data streams. According to a unique classification system used by [46], streams of data were divided into 3 categories: QoS level-1, QoS level-2, and best-effort streams. Each category received a unique routing protocol. In networks, streams of data may be diverted away from their existing path to satisfy a specific set of QoS criteria [47].

On multiple paths routing over an Open-Flow network, a few other academics have researched [48]. Developed better choosing path algorithms by concentrating on MCP issues. With the aid of SDN, [49] proposed an optimum system for QoS routing that gathers data on the current condition of the network to determine routing pathways and adjust networking routing on demand. They did not, nevertheless, discuss a multiple paths routing mechanism in their study. For the CSP issue, [50] offered two strategies, however, because of the longer operating duration, the efficiency was not greater than that in [42, 51]. The researchers created an approach named equal-cost multiple paths routing (ECMP), which may distribute end-to-end congestion employing a randomized way to various transmitting channels. However, a larger throughput is not possible with this approach. The Round Robin Multiple-Path technique was suggested by the researchers in [53] to select a distinct transmitting route from each end-to-end link. The multiple paths routing method built on Open-Flow that the researchers developed in [54] does not efficiently combine network bandwidth constraints or take into consideration the network's capacities [55]. Presented a non-SDN-oriented approach of multiple paths routing in traffic bottlenecks depending on dynamic optic networking. Multiple paths transmit management protocol routing method depending on SDN was suggested by [56], and it makes utilization of the controller to determine the many least routes between the 2 units. There are a similar number of information sub-flows created from the data flow. There is a similar set of information sub-flows created from the data flow. Additionally, the control system creates stream tables and transmits these via several pathways to an Open Flow switching device. According to their study, this approach failed to increase network throughput because the algorithms just managed to obtain a portion of the data about the system topology

and traffic dispersion, which resulted in an absence of adaptability in the routing planning and real-time variable acquiring of the overall networking state. Therefore, the network throughput will decrease as a result of this. Multiple-path routing is suggested in [57] as a way to more efficiently combine networking bandwidth supplies, which can reduce connection traffic and increase networking throughput. The findings also demonstrated that, in terms of the total capability of video streaming, a multiple-path video broadcasting method may be more economically advantageous than a single-path technique. This study takes 3 restrictions into account while determining the link loads, which is how the method varies from theirs.

15.3 METHODOLOGY

15.3.1 SYSTEM FRAMEWORK

It is challenging to actively and adaptively rebuild multiple distinct routing pathways as feasible on a per-flow level with the present Web framework. For instance, a router examines its routing table data when a package enters and then transmits it by established rules set up by networking managers. This shortcoming is primarily addressed by Open-Flow, a novel framework that enables us to create various routing principles linked to streams of data, enabling users to quickly change the design of the networking divisions and the traffic streams, much like in an SDN setting [58, 59]. The Open-Flow specification is used by all of the exchanges in the SDN-oriented networking. These controls are all linked to a single controlling device. The regulator can obtain the current networking state, including the latency, package loss rate, and bandwidth accessible, from the information plane (the system and the user). After collecting the QoS metrics, it will evaluate them and determine the video broadcasting strategy based on the bandwidth aggregating rule [60]. The suggested system framework, shown in Figure 15.1, provides a variety of ports and features, a few of which were previously only found in routers in the conventional Web paradigm.

The Open Flow regulations, which in Figure 15.1 are supported by the fundamental network, are primarily made up of 3 modules: the controller, routing, and monitoring component. The routing component, which depends on the floodlight system, may offer data on the condition of the network globally and determine the best route and bandwidth accessible for every stream [61, 62]. It works together with the topology component,

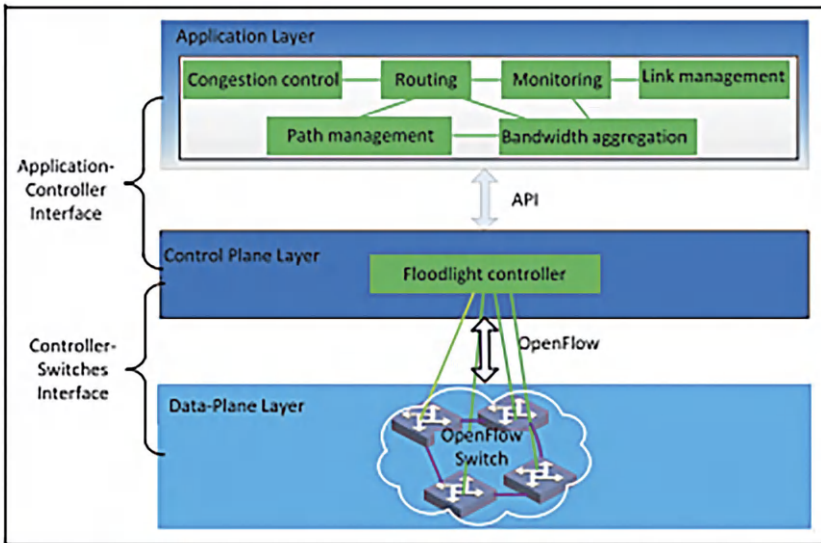


FIGURE 15.1 System framework.

which can obtain the networking topology and real-time upgrade the topology. The bandwidth aggregating component must reevaluate the routing whenever the network condition does not comply with the demands of the specified QoS criteria. The controlling device creates the stream table in the meantime and transmits it to the appropriate valves.

Figure 15.1 shows the Application of controller and switch-controller connections, which are the two principal connections. To deliver flow tables related to data runs, learn the topology of the system, and obtain traffic condition data, the controller connects to a system of switches using the encrypted Open Flow regulations in the first case. About the latter, the control system offers an accessible (non-real moment), secured interface to the program's suppliers so they can reserve novel information divisions and even describe novel routing principles related to the divisions [63, 64].

To interface with the data plane via the Open-Flow regulations, the controller component is in charge of controlling and keeping track of network-related data of the entire network for more efficient use of network supplies. Additionally, based on the user's demand message, it creates and transmits the stream data to an Open-Flow switching. These are the major components of its operations:

The link layered discovery standard (LLDS) is used by the control system to verify the present network condition and to keep track of broad topology. In this deployment, the control system initially sends an LLDS package through a Packet-out signal to each associated switch [65, 66]. After receiving the signal, these switching devices will forward LLDS packages to each of the connections. When an Open-Flow switch collects LLDS packages, it notifies the controlling device via the Packet in signal about the connection details between the 2 switches. The broad topology can be built by the control system after it has determined which switching devices are directly linked to one another after receiving the package in a signal with LLDS. Furthermore, if the user continues to transmit request signals, the study leads to the guidelines (i.e., seven percent of the mean Floodlight capability) to avoid mutual interference between the procedures (User, Floodlight, Mininet) [67, 68]. In contrast, the application tier is linked to the North-bound-application software interface and consists of the following important components:

Network surveillance is critical for network administration since it provides statistics and tracking of traffic in terms of the present networking condition, comprising details concerning connection gridlock, bandwidth accessible, time latency, and package loss ratio depending on the Open Flow regulations. This component may also completely surveillance the networking topology and routing condition, and alters the path choosing in real-time in response to any modifications in the networking condition. This study utilizes the standard sampling technique to retrieve connection data using the control system. This study can obtain s bytes in an interface by doing a T cycle using sampling switching devices. As a result, this study may get the approximated connection bandwidth, s/T .

Based on the duration stamp technique, the latency in transmitting a package is primarily regarded in the connection's latency, and the average time latency is the proportion of the overall duration of the obtained data package to the count of successive data. Following the collection of surveillance data, the system launches the traffic management component to address the issue of how to prevent and identify traffic based on the traffic degree.

The routing component is in charge of gathering data about the broad networking state for the control system as well as computing the traffic allocation ratio of different pathways to give the optimum routing regulations for the application tier.

The bandwidth aggregating component is in charge of gathering data from the surveillance component and generating bandwidth aggregating routing based on the surveillance data.

Path supply and path computation are handled by the path administration component. Path supply works with switches to control service supply methods and process Open-Flow signals. When it receives a Packet-in signal, it analyzes it and sends the data to the path computing module for path calculation. After acquiring the assignment from path supply, the other is in charge of computing the service supply plan for every user and collecting the present network condition.

15.3.2 ACTIVE AND AMRA PLAN

This part describes an active and AMRA optimizing algorithm for designing an Open-Flow control system that takes maximum advantage of networking supplies and prevents network traffic, efficiently lowering the package loss ratio during the transmission procedure.

15.3.3 ACTIVE AND AMRA ROUTING COMPONENT

The core component of the suggested system is called active and AMRA; it combines bandwidth depending on surveillance data and prevents data package loss brought on by connection breakdowns or link traffic, hence guaranteeing the transmission's dependability. A greater equilibrium of networking resource consumption results from the data stream being actively assigned to several efficient pathways between any 2 nodes. Due to the significant reduction in traffic brought on by the transfer of data, both the end-to-end latency and the package loss ratio will also be reduced [69].

The foundation for the execution of a multiple paths algorithm in an Open Flow system is provided by the controller's ability to obtain data on the system's architecture, connections, and data regarding busy and idle connections. The execution of the multiple paths algorithm is highly versatile and diverse as a result of the customizable control system, which may ensure the scaling of the current Open-Flow platform by expanding the required component.

The purpose of this technique is to determine the way with the lowest expense that meets both the latency and bandwidth restrictions. The suggested algorithm’s primary concept is that the optimizing issue changes to a linear function by employing the function’s goal variables switched by the restriction circumstances, which simplifies the resolving procedure and reduces the period complication of acquiring the optimum or more effective solution.

The algorithm in this paper is in charge of computing the aggregated bandwidth based on real-time surveillance data and a regular scheduling technique that can enhance the link value. Furthermore, when traffic arises, the suggested method may discover the compatible principles regionally and evaluate if the package has been transmitted or discarded. In comparison to previous multiple paths routing techniques, this may decrease the overhead generated by the control system constantly tracking the network. To see the controller-triggered expenses, we evaluated the latency of the initial package acquired by host 4. It was 0.014 milliseconds. Nevertheless, the 2nd one was reduced to 0.004 ms, demonstrating that the strategy may decrease the expense of the control system analyzing the packages in Table 15.1.

TABLE 15.1 The Control System Overhead

	Initial Package Processing Duration (Ms)	2nd Package Processing Duration (Ms)
Active and AMRA	0.014	0.004
ECMP	0.8	1.2

15.3.4 ACTIVE AND AMRA ROUTING

Historically, the major components of a network depending on dispersed methodologies for resilience have been switching and routing architectures. In SDN systems, a single video streams across the shortest route determined by the control system depending on route expense between any 2 end units s and t. In other words, the 2 streams are initially directed down the shortest route [70]. They will have the option to be redirected on a viable path determined by the routing component depending on the route’s latency if the shortest route does not meet the time latency restriction.

The topology of the system will dynamically alter as a result of package losses and congestion surges. As a result, the network’s efficiency

will be negatively impacted by the aforementioned issues. To address this type of issue, the control system sends portions of the flow table, which is updated depending on the table forwarded. A novel topology and a novel setup stream table are also computed by the control system using the algorithm described in this paper. The control system initially eliminates any relevant stream tables from the upgrading procedure if they fall short of the latency limitation. While this is going on, the switch will just transmit the Package-in signal to the control system if instructed to perform so by the control system or if the switch's stream table is missing the proper entry [71, 72]. The control system does not respond to the switch right away after collecting the data. Nevertheless, to provide the bandwidth aggregated regulation in real time, the control system asks for the QoS variable data of every connection in the fundamental network.

To secure the outstanding dependability and adaptation of the stream transmission, it is also crucial to ensure the efficiency of the stream transmitting the bandwidth aggregating regulations, which may alter the routing algorithm promptly, and actively [52, 73].

15.3.5 THE SCHEDULING ALGORITHM'S EXECUTION PROCEDURE

The complete flowchart for the approach suggested in this paper was created using the results of the evaluation mentioned above. The scheduling algorithm described in this paper seeks to identify the best paths for traffic streams with the least amount of latency, package loss, and throughput at the connections. As a result, this study works on ways to reduce the network's optimum connection consumption. Link values are used by the suggested method to optimize the network design. By including the routing component as a floodlight control system application in the suggested technique, this study allows the SDN network to pick the routing autonomously with the topology. The algorithms have also been applied to SDN routes on the Mininet simulation (version 2.1.0).

The control system frequently checks the QoS characteristics of the base system after determining whether the optimum route should be sought at program startup. For surveillance of the data from the base layer from the viewpoint of Open v Switch, the query time of the control system has been configured to 5 seconds in this instance.

The control system initially determines all the open pathways and their respective bandwidths when a package comes to the control system. The

stream's transmitting route will then be the one that uses the largest amount of bandwidth. The route will be the best option and the streams of broadcasting video won't be redirected if it complies with the latency restriction. If not, the control system must use the active and AMRA method to determine the best course of action. This study will not redirect any of them if there is no best route since none of the possible paths can satisfy the time restriction. Otherwise, the control system collects the QoS data from the surveillance component and determines the connection values between every pair of units if an ideal route is present. The control system then changes every link's link-state condition data on the route wherein the stream resides. Employing the topology administration component from the floodlight control system, this study can retrieve the link-state details of each connection, including the linked units and ports, saved to clusters to determine the routing path. In addition, by visiting every node and its associated nodes, we may also discover the complete route from the supply node to the target node. The best route that meets the latency restriction can be found in the interim.

15.4 RESULTS AND DISCUSSION

In this part, this study uses the Mininet [52] and lighting simulator [6] test system to determine the suggested active and AMRA's efficacy and contrast it with single-path and ECMP transmitting techniques.

This study employed Mininet to design the networking topology and established the speed of every connection to 100 Mbps. Additionally, a remote-control system running Flood-light 1.2 was linked to each unit in the actively constructed architecture. The control system is in charge of controlling the flow, including decision-making over redirecting and route computation. This study put several streams on the control system to verify the efficacy of active and AMRA depending on the networking topology, and this study configured the duration of the control system to be five seconds for surveillance of the underneath networking. The evaluation elements comprise the latency, throughput, and package loss ratio.

In this paper, the suggested algorithm is also contrasted with a single-path approach and an ECMP technique depending on broad networking. Single-path methods, like those developed by Dijkstra and Bellman-Ford, are frequently employed in the route of current data networking. Furthermore, it chooses a portion of the pathways to transport video broadcasting depending on the additive expense property (such as hop

values). Conversely, the ECMP method distributes traffic evenly across several equal-expense pathways and comprises every networking connection. ECMP can, in some cases, outperform single-path links in terms of performance and dependability, however, because of the hashing collision, this technique is unable to fully use the bandwidth at hand. The ratios of the flow dispersion cannot be altered by this algorithm in response to modifications in the connection condition.

Also, active and AMRA operate better than ECMP. The cause is that ECMP just enables equal expense for the briefest routes to be chosen for an even dispersion of the traffic and causes no effort to alter the connection expenses, which makes it challenging to pinpoint the proper routes. In contrast, active and AMRA use multi-path innovation to utilize the accessible capability and reduce the expense of video delivery, which results in an ensured QoS. This study carried out numerous studies using 3 distinct methodologies to verify the efficacy of the 3 algorithms, and the findings have been examined in this study.

15.4.1 LATENCY

This study starts by looking at the time latency’s efficiency, as seen in Figure 15.2(a). Varied use approaches lead to varied traffic dispersions along every path, which alters the time latency. Because of the receipt of the initial data stream, the time latencies as determined by the 3 methodologies are all increasing at the start of the study. Additionally, the floodlight control system receives the initial data package from the Open-Flow switch. Following the routing has been determined by the routing component, the control system transmits and installations the flow tables to the Open-Flow switching device. Additionally, when employing multiple-path approaches, the latency of certain packages may rise due to route switching.

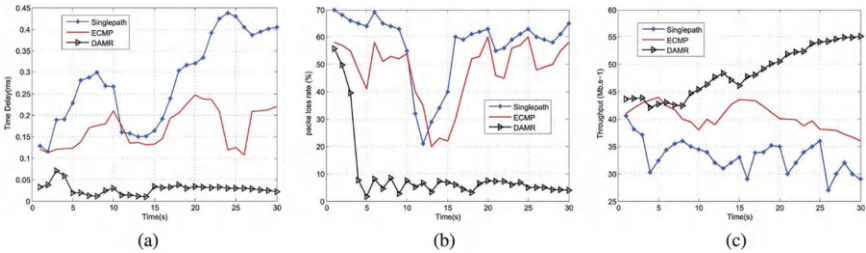


FIGURE 15.2 The study outcomes.

According to Figure 15.2(a), single-path delays are the most unstable because of networking traffic. As a result, a data package will be lost as well as a decline in the efficiency of some connections in the system. Even though ECMP may, to a certain degree, decrease time latencies contrasted to single-path, it also causes increased jitter over extended periods, which impairs network efficiency. The suggested active and AMRA, in general, is jitter-free and smooth, demonstrating its supremacy over ECMP and single-path because of its capability to fully utilize the aggregating path capability per stream and offer bandwidth assurance.

15.4.2 PACKAGE LOSS RATIO

Figure 15.2(b) depicts the results of testing the package loss ratio in the same testing scenario. Due to the active and AMRA technique's utilization of idle routes for transmitting the video, which minimizes the negative impacts of congestion brought on by just one connection throughout the entire procedure of video delivery, the variation of the package loss rate is comparatively smooth, and is lower than 10% when employing it, as may be observed in the figure. With the help of Figure 15.2(c), this study can deduce that different forms of flow on the common path may have additional package loss. Because of the lengthy time it takes to generate the flow table employing the 3 distinct approaches, the study's initial phase results in an elevated package loss ratio. Compared to single-path, ECMP can somewhat lower the package loss rate, but it struggles with large volatility and lowers networking efficiency over the entire research procedure.

The use of active and AMRA, which can actively adjust the network architecture to achieve resource transfer in a multi-path setting, is suggested as a way to reduce the negative impacts of video broadcasts by resolving the bandwidth supply constraint. The package loss is connected to the networking connection state, according to the entire set of research findings. Single-path routing fails fault tolerant because it only takes into account a single route between a supply unit and a target in comparison to active and AMRA. Additionally, active and AMRA operate better than ECMP and single-path. The fundamental cause of this is the ability of multiple paths of data transfer to select an alternative route from among those that are accessible. By doing so, networking traffic can be avoided and the package loss ratio of transmitting video can be somewhat reduced.

As a result, utilizing the active and AMRA approach greatly reduces the rate of package loss during data transfer.

15.4.3 THROUGHPUT

Because of networking latency or route switching employing the 3 ways, there are significant swings in Figure 15.2(c). Active and AMRA can transfer data across numerous routes as opposed to single-path and ECMP, which indirectly enhances the framework's overall bandwidth and boosts throughput. The ECMP method's efficiency is also not consistently steady, even though it can increase throughput.

The 3 strategies can produce identical efficiency when there is no traffic on the network connections. When there is traffic, the active and AMRA method has excellent flexibility to deal with the variance in bandwidth and can direct some traffic to lines that are less busy or even not crowded at all, which can enhance the efficient use of network supplies. Discovering many routes to satisfy a specific aggregated bandwidth demand while incurring the least amount of cost is one way to approach the multiple paths issue. Without taking into account multiple-cast algorithms, this study concentrates on single-cast methods for active and AMRA to provide end-to-end QoS. Furthermore, we have put forth a practical approach depending on multiple path transmissions that may actively alter the sending route with the connection state over SDNs.

Lastly, take note that the model's duration in these studies is represented by the x-axis in Figures 15.2(a), (b), and (c).

15.5 CONCLUSION

The three restrictions of CSP over SDNs were covered in this study's summary of active and AMRA methods. Open-Flow's centralized management function, which can achieve resource allocation optimization, is fully utilized by active and AMRA. Furthermore, by utilizing all aspects of the optimization concept, it can efficiently combine bandwidth supplies for optimizing the connection's values. The empirical findings show that the suggested approach is capable of overcoming the regulator's processing cost and successfully adapting to network dynamics. As a result, it can enhance both the connection's usage and the caliber of customer service.

The future scope of this study will then be expanded to consider the case where distinct streams have varied targets, like emphasizing a particular stream and employing various QoS techniques to address bandwidth or stream of traffic latency constraints.

KEYWORDS

- **adaptive multiple paths routing algorithm**
- **broadcasts**
- **quality of service**
- **software-described networks**
- **video services**

REFERENCES

1. Kusumadewi, R. N., & Karyono, O. (2019). Impact of service quality and service innovations on competitive advantage in retailing. *Budapest International Research and Critics Institute-Journal (BIRCI-Journal)*, 2(2), 366–374.
2. Mroue, H., Parrein, B., Hamrioui, S., Bakowski, P., Nasser, A., Cruz, E. M., & Vince, W. (2020). LoRa+: An extension of LoRaWAN protocol to reduce infrastructure costs by improving the quality of service. *Internet of Things*, 9, 100176.
3. Alvarez, F., Breitgand, D., Griffin, D., Andriani, P., Rizou, S., Zioulis, N., Moscatelli, F., Serrano, J., Keltsch, M., Trakadas, P., & Phan, T. K. (2019). An edge-to-cloud virtualized multimedia service platform for 5G networks. *IEEE Transactions on Broadcasting*, 65(2), 369–380.
4. Noreikis, M., Xiao, Y., & Ylä-Jääski, A. (2017). QoS-oriented capacity planning for edge computing. In *2017 IEEE International Conference on Communications (ICC)* (pp. 1–6). IEEE.
5. Yousefpour, A., Patil, A., Ishigaki, G., Kim, I., Wang, X., Cankaya, H. C., Zhang, Q., Xie, W., & Jue, J. P. (2019). Fogplan: A lightweight QoS-aware dynamic fog service provisioning framework. *IEEE Internet of Things Journal*, 6(3), 5080–5096.
6. Floodlight. (n.d.). Retrieved from: <http://www.projectfloodlight.org/download/>.
7. Barmponakis, S., Maroulis, N., Koursiompas, N., Kousaridas, A., Kalamari, A., Kontopoulos, P., & Alonistioti, N. (2022). AI-driven, QoS prediction for V2X communications in beyond 5G systems. *Computer Networks*, 217, 109341.
8. Lin, Y., & Shen, H. (2016). CloudFog: Leveraging fog to extend cloud gaming for thin-client MMOG with high quality of service. *IEEE Transactions on Parallel and Distributed Systems*, 28(2), 431–445.

9. Liu, J. (2021). Design and implementation of VoIP QoS model combining IntServ and DiffServ based on network processor IXP2400. In *2021 7th Annual International Conference on Network and Information Systems for Computers (ICNISC)* (pp. 60–64). IEEE.
10. Mohamad, A. H., & Hussein, H. A. (2023). Control dynamic system and QoS manager agent over IPv6 networks: IntServ and DiffServ approach in access nodes.
11. Ghani, M. A. (2019). Analisis performansi quality of service (QoS) pada jaringan multi protocol label switching dengan metode IntServ. *Jurnal Manajemen Informatika*, 9(2).
12. Yan, F. Y., Ayers, H., Zhu, C., Fouladi, S., Hong, J., Zhang, K., Levis, P., & Weinstein, K. (2020). Learning *in situ*: A randomized experiment in video streaming. In *17th USENIX Symposium on Networked Systems Design and Implementation (NSDI 20)* (pp. 495–511).
13. Dasari, M., Bhattacharya, A., Vargas, S., Sahu, P., Balasubramanian, A., & Das, S. R. (2020). Streaming 360-degree videos using super-resolution. In *IEEE INFOCOM 2020-IEEE Conference on Computer Communications* (pp. 1977–1986). IEEE.
14. Xu, M., Li, C., Zhang, S., & Le Callet, P. (2020). State-of-the-art in 360 video/image processing: Perception, assessment and compression. *IEEE Journal of Selected Topics in Signal Processing*, 14(1), 5–26.
15. Spiteri, K., Urgaonkar, R., & Sitaraman, R. K. (2020). BOLA: Near-optimal bitrate adaptation for online videos. *IEEE/ACM Transactions on Networking*, 28(4), 1698–1711.
16. Srilakshmi, U., Veeraiah, N., Alotaibi, Y., Alghamdi, S. A., Khalaf, O. I., & Subbayamma, B. V. (2021). An improved hybrid secure multipath routing protocol for MANET. *IEEE Access*, 9, 163043–163053.
17. Zhang, D. G., Chen, L., Zhang, J., Chen, J., Zhang, T., Tang, Y. M., & Qiu, J. N. (2020). A multi-path routing protocol based on link lifetime and energy consumption prediction for mobile edge computing. *IEEE Access*, 8, 69058–69071.
18. Mercat, A., Viitanen, M., & Vanne, J. (2020). UVG dataset: 50/120fps 4K sequences for video codec analysis and development. In *Proceedings of the 11th ACM Multimedia Systems Conference* (pp. 297–302).
19. Kumar, D. T. (2020). Video-based traffic forecasting using convolution neural network model and transfer learning techniques. *Journal of Innovative Image Processing*, 2(3), 128–134.
20. Li, Y., Song, L., Chen, Y., Li, Z., Zhang, X., Wang, X., & Sun, J. (2020). Learning dynamic routing for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 8553–8562).
21. Rafique, W., Qi, L., Yaqoob, I., Imran, M., Rasool, R. U., & Dou, W. (2020). Complementing IoT services through software-defined networking and edge computing: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, 22(3), 1761–1804.
22. Rehmani, M. H., Davy, A., Jennings, B., & Assi, C. (2019). Software defined networks-based smart grid communication: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, 21(3), 2637–2670.

23. Alaya, B., Khan, R., Moulahi, T., & Khediri, S. E. (2021). Study on QoS management for video streaming in vehicular ad hoc network (VANET). *Wireless Personal Communications*, 118, 2175–2207.
24. Raj, J. S., & Joe, C. V. (2021). Wi-Fi network profiling and QoS assessment for real-time video streaming. *IRO Journal on Sustainable Wireless Systems*, 3(1), 21–30.
25. Serhani, A., Naja, N., & Jamali, A. (2020). AQ-Routing: Mobility-, stability-aware adaptive routing protocol for data routing in MANET–IoT systems. *Cluster Computing*, 23(1), 13–27.
26. Varyani, N., Zhang, Z. L., Rangachari, M., & Dai, D. (2019). LADEQ: A fast Lagrangian relaxation-based algorithm for destination-based QoS routing. In *2019 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)* (pp. 462–468). IEEE.
27. Zhao, M., Li, X., Yin, J., Cui, J., Yang, L., & An, S. (2018). An integrated framework for electric vehicle rebalancing and staff relocation in one-way carsharing systems: Model formulation and Lagrangian relaxation-based solution approach. *Transportation Research Part B: Methodological*, 117, 542–572.
28. BinSahq, A., Sheltami, T., Mahmoud, A., & Nasser, N. (2021). Bootstrapped LARAC algorithm for fast delay-sensitive QoS provisioning in SDN networks. *International Journal of Communication Systems*, 34(11), e4880.
29. Horváth, M., & Kis, T. (2016). Solving resource-constrained shortest path problems with LP-based methods. *Computers & Operations Research*, 73, 150–164.
30. Alhussein, O., Do, P. T., Ye, Q., Li, J., Shi, W., Zhuang, W., Shen, X., Li, X., & Rao, J. (2020). A virtual network customization framework for multicast services in NFV-enabled core networks. *IEEE Journal on Selected Areas in Communications*, 38(6), 1025–1039.
31. Jiang, S., Zhang, Y., Liu, R., Jafari, M., & Kharbeche, M. (2022). Data-driven optimization for dynamic shortest path problem considering traffic safety. *IEEE Transactions on Intelligent Transportation Systems*, 23(10), 18237–18252.
32. Kucukoglu, I., Dewil, R., & Cattrysse, D. (2021). The electric vehicle routing problem and its variations: A literature review. *Computers & Industrial Engineering*, 161, 107650.
33. Tilk, C., Rothenbächer, A. K., Gschwind, T., & Irnich, S. (2017). Asymmetry matters: Dynamic half-way points in bidirectional labeling for solving shortest path problems with resource constraints faster. *European Journal of Operational Research*, 261(2), 530–539.
34. Belamri, F., Boulfekhar, S., & Aissani, D. (2021). A survey on QoS routing protocols in Vehicular Ad Hoc Network (VANET). *Telecommunication Systems*, 78(1), 117–153.
35. Gawas, M. A., & Govekar, S. (2021). State-of-art and open issues of cross-layer design and QoS routing in the internet of vehicles. *Wireless Personal Communications*, 116(3), 2261–2297.
36. Dhanalakshmi, B., SaiRamesh, L., & Selvakumar, K. (2021). Intelligent energy-aware and secured QoS routing protocol with dynamic mobility estimation for wireless sensor networks. *Wireless Networks*, 27, 1503–1514.
37. Hou, R., Chang, Y., & Yang, L. (2017). Multi-constrained QoS routing based on PSO for named data networking. *IET Communications*, 11(8), 1251–1255.

38. Eiza, M. H., Owens, T., & Ni, Q. (2015). Secure and robust multi-constrained QoS aware routing algorithm for VANETs. *IEEE Transactions on Dependable and Secure Computing*, 13(1), 32–45.
39. Huynh, T. T., Dinh-Duc, A. V., & Tran, C. H. (2016). Delay-constrained energy-efficient cluster-based multi-hop routing in wireless sensor networks. *Journal of Communications and Networks*, 18(4), 580–588.
40. Liang, M., Gao, C., & Zhang, Z. (2017). A new genetic algorithm based on modified Physarum network model for bandwidth-delay constrained least-cost multicast routing. *Natural Computing*, 16, 85–98.
41. Marinakis, Y., Migdalas, A., & Sifaleras, A. (2017). A hybrid particle swarm optimization–Variable neighborhood search algorithm for constrained shortest path problems. *European Journal of Operational Research*, 261(3), 819–834.
42. Karakus, M., & Duresi, A. (2017). Quality of service (QoS) in software-defined networking (SDN): A survey. *Journal of Network and Computer Applications*, 80, 200–218.
43. Keshari, S. K., Kansal, V., & Kumar, S. (2021). A systematic review of quality of services (QoS) in software-defined networking (SDN). *Wireless Personal Communications*, 116, 2593–2614.
44. Bari, M. F., Chowdhury, S. R., Ahmed, R., & Boutaba, R. (2013). PolicyCop: An autonomic QoS policy enforcement framework for software-defined networks. In *2013 IEEE SDN for Future Networks and Services (SDN4FNS)* (pp. 1–7). IEEE.
45. Elbasheer, M. O., Aldegheshem, A., Lloret, J., & Alrajeh, N. (2021). A QoS-based routing algorithm over software-defined networks. *Journal of Network and Computer Applications*, 194, 103215.
46. Moreira, D. C., Guerreiro, I. M., Sun, W., Cavalcante, C. C., & Sousa, D. A. (2020). QoS predictability in V2X communication with machine learning. In *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)* (pp. 1–5). IEEE.
47. Tan, X., Zhao, H., Han, G., Zhang, W., & Zhu, T. (2019). QSDN-WISE: A new QoS-based routing protocol for software-defined wireless sensor networks. *IEEE Access*, 7, 61070–61082.
48. Jaiswal, K., & Anand, V. (2021). A grey-wolf based optimized clustering approach to improve QoS in wireless sensor networks for IoT applications. *Peer-to-Peer Networking and Applications*, 14, 1943–1962.
49. Mammeri, Z. (2019). Reinforcement learning based routing in networks: Review and classification of approaches. *IEEE Access*, 7, 55916–55950.
50. Guck, J. W., Van Bemten, A., Reisslein, M., & Kellerer, W. (2017). Unicast QoS routing algorithms for SDN: A comprehensive survey and performance evaluation. *IEEE Communications Surveys & Tutorials*, 20(1), 388–415.
51. Sahhaf, S., Tavernier, W., Colle, D., & Pickavet, M. (2017). Adaptive and reliable multipath provisioning for media transfer in SDN-based overlay networks. *Computer Communications*, 106, 107–116.
52. Bhardwaj, A., & El-Ocla, H. (2020). Multipath routing protocol using genetic algorithm in mobile ad hoc networks. *IEEE Access*, 8, 177534–177548.
53. Sharma, A., & Kumar, R. (2019). Service-level agreement—Energy cooperative quickest ambulance routing for critical healthcare services. *Arabian Journal for Science and Engineering*, 44(4), 3831–3848.

54. Chemeritskiy, E., & Smelansky, R. (2014). On QoS management in SDN by multipath routing. In *2014 International Science and Technology Conference (Modern Networking Technologies) (MoNeTeC)* (pp. 1–6). IEEE.
55. Besta, M., Domke, J., Schneider, M., Konieczny, M., Di Girolamo, S., Schneider, T., Singla, A., & Hoefler, T. (2020). High-performance routing with multipathing and path diversity in ethernet and HPC networks. *IEEE Transactions on Parallel and Distributed Systems*, 32(4), 943–959.
56. Chiesa, M., Kindler, G., & Schapira, M. (2016). Traffic engineering with equal-cost-multipath: An algorithmic perspective. *IEEE/ACM Transactions on Networking*, 25(2), 779–792.
57. Robinson, Y. H., Balaji, S., & Golden, J. E. (2019). PSOBLAP: Particle swarm optimization-based bandwidth and link availability prediction algorithm for multipath routing in mobile ad hoc networks. *Wireless Personal Communications*, 106, 2261–2289.
58. Khairi, H. H., Ariffin, S. H., Latiff, N. A., Yusof, K. M., Hassan, M. K., & Rava, M. (2020). The impact of firewall on TCP and UDP throughput in an OpenFlow software-defined network. *Indonesian Journal of Electrical Engineering and Computer Science*, 20(1), 256–263.
59. Nguyen, T. G., Phan, T. V., Hoang, D. T., Nguyen, T. N., & So-In, C. (2020). Efficient SDN-based traffic monitoring in IoT networks with double deep Q-network. In *International Conference on Computational Data and Social Networks* (pp. 26–38). Cham: Springer International Publishing.
60. Garg, S., Kaur, K., Ahmed, S. H., Bradai, A., Kaddoum, G., & Atiquzzaman, M. (2019). MobQoS: Mobility-aware and QoS-driven SDN framework for autonomous vehicles. *IEEE Wireless Communications*, 26(4), 12–20.
61. Priya, A. V., & Radhika, N. (2019). Performance comparison of SDN OpenFlow controllers. *International Journal of Computer Aided Engineering and Technology*, 11(4, 5), 467–479.
62. Li, W., Meng, W., & Kwok, L. F. (2016). A survey on OpenFlow-based Software Defined Networks: Security challenges and countermeasures. *Journal of Network and Computer Applications*, 68, 126–139.
63. Zhang, L., Lin, H., Huan, W., & Bi, W. (2022). Software-defined network flow rule conflict detection system based on OpenFlow. *Journal of Computer Applications*, 42(2), 528.
64. Aryan, R., Yazidi, A., Kure, Ø., & Engelstad, P. E. (2021). A parallel approach for detecting OpenFlow rule anomalies based on a general formalism. *Concurrency and Computation: Practice and Experience*, 33(15), e5907.
65. Popic, S., Vuleta, M., Cvjetkovic, P., & Todorović, B. M. (2020). Secure topology detection in software-defined networking with network configuration protocol and link layer discovery protocol. In *2020 International Symposium on Industrial Electronics and Applications (INDEL)* (pp. 1–5). IEEE.
66. Jiménez, Y., Cervelló-Pastor, C., & García, A. (2015). Dynamic resource discovery protocol for software-defined networks. *IEEE Communications Letters*, 19(5), 743–746.

67. Suguna, P., & Velusamy, R. L. (2022). LARIC: Latency-aware QoS routing for interactive communication in software-defined multimedia. *International Journal of Information Technology*, 14(6), 2941–2950.
68. Nehra, A., Tripathi, M., Gaur, M. S., Battula, R. B., & Lal, C. (2019). SLDP: A secure and lightweight link discovery protocol for software-defined networking. *Computer Networks*, 150, 102–116.
69. Lu, Y., Chen, Y., Xu, X., Fu, Q., Chen, J., & Liu, L. (2023). A sub-flow adaptive multipath routing algorithm for data center network. *International Journal of Computational Intelligence Systems*, 16(1), 25.
70. Huang, Y., Cao, W., Liu, X., Jiang, X., Yang, J., & Yang, F. (2021). An adaptive multipath routing for LEO satellite network. In *2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)* (Vol. 4, pp. 1536–1541). IEEE.
71. Venkatasubramanian, S. (2022). Fruit-fly algorithm based dynamic source routing algorithm for energy efficient multipath routing in MANET. In *2022 International Conference on Computer Communication and Informatics (ICI)* (pp. 1–8). IEEE.
72. Chen, Z., Zhou, W., Wu, S., & Cheng, L. (2020). An adaptive on-demand multipath routing protocol with QoS support for high-speed MANET. *IEEE Access*, 8, 44760–44773.
73. Jiawei, W., Xiuquan, Q., & Guoshun, N. (2018). Dynamic and adaptive multi-path routing algorithm based on software-defined network. *International Journal of Distributed Sensor Networks*, 14(10), 1550147718805689.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Index

A

- Academy of Maritime Education and Training (AMET), 323
- Accountability, 219, 298, 316, 318
- Accurate
 - credit evaluation, 258
 - intelligent services, 223
- Actor-critic method, 219
- Adaptability, 82, 87, 98, 107, 159, 160, 174, 190, 197, 309, 325, 328
- Adaptable
 - allocating resources, 158
 - architecture, 158
 - radio, 159
- Adaptation to dynamic environments, 109
- Adaptive
 - multiple paths routing algorithm (AMRA), 324, 325, 331, 332, 334–338
 - optimizing algorithm, 331
 - smart surfaces, 221
- Adversarial
 - attacks, 228
 - convolution techniques, 43
 - training, 26, 27, 29, 32, 33, 39, 43
 - framework, 27
- Agent's regional information, 196
- Aggregation techniques, 227
- AIoT (artificial intelligence of things), 75, 230
- Air pollution, 158
- AirComp, 220, 221, 227, 229
- Algorithm, 8, 9, 14, 15, 28, 100, 102, 104, 105, 109, 120, 134, 168, 171, 173, 174, 178, 181, 182, 216, 219, 220, 227–229, 242, 252, 289, 297, 308, 323–327, 333, 335, 337
 - evaluation, 36
 - monitoring, 244
 - primary concept, 332
 - robustness, 109
- Algorithmic
 - biases, 108
 - trading, 252
- Alibaba, 258
- Allocation efficiency, 249
- AlphaGo, 228
- Amazon, 52, 122, 168, 258
 - AWS deepLens, 122
 - machine image, 172
- Analytics techniques, 288
- Anomaly
 - detection (AE), 25–33, 35–37, 40, 42, 43
 - accuracy, 29
 - techniques, 28, 36
 - explanation, 27–29, 32, 35, 43
- Ant colony optimization (ACO), 53
- Anti-fraud rate, 258
- Apple timepieces, 145
- Application
 - oriented variables, 97
 - safety demands, 146
- Artificial intelligence (AI), 1, 3–19, 25, 29, 41–43, 51, 52, 65, 71, 73, 74, 77, 81, 86–89, 95, 96, 98–100, 105, 107–110, 117–126, 140, 141, 143, 147, 149, 167–169, 171, 173, 182, 189–191, 205, 206, 213, 214, 216–220, 223–232, 241–248, 251–254, 256–259, 265, 266, 269, 270, 279, 285–300, 305–319, 323
- algorithms, 96, 99, 100, 105, 107–109, 169, 171, 182, 228, 296, 298, 299
- anomaly detection, 107
- applications, 215
- bazaar, 4, 6–10, 12, 15–17, 19
 - computer trading framework, 6
 - equilibrium, 6
- boosters, 253
- business model research, 306
- driven
 - analytics, 289

- data compression, 107
- decisions, 100
- digital technologies, 316, 317
- enhancements, 109, 110
- modeling, 297
- precision farming techniques, 295, 297
- SDG initiatives, 315
- solutions, 98, 107, 290, 315, 318
- technologies, 316
- financial services, 244
- inference, 8, 10
- mobility management, 226
- operational platforms, 81
- powered
 - drones, 296
 - monitoring systems, 289
- service providers, 7
- technology, 18, 98, 213, 228, 230, 243, 245–247, 258, 286, 296, 298, 299, 306, 314–316, 318
- democratization, 297
- training, 8
- Asset
 - booking calculations, 60
 - management, 121, 244, 258
- Auto-encoders, 32
- Automate payments, 18
- Automated
 - optical inspection (AOI), 125
 - order allocation, 249
- Automatic
 - reaction mechanisms, 252
 - scheduling system, 249
- Automation, 96, 226, 242, 248
- Autonomous
 - adaptable resource allocation, 171
 - automobiles, 122
 - driving, 219, 222, 229
 - mapping network, 123
 - sample T-test, 152
 - tractors, 296
- Availability risk, 252

B

- Back-end supply chain connections, 247
- Backhaul connections, 157
- Back-propagation NN (BPNN), 74, 76, 82, 84–86, 89

- Baidu takeaway system, 249
- Bandwidth
 - consumption, 191
 - delays, 168
 - limitations, 223
 - management, 42
- Bank
 - credit risk control, 258
 - operations, 242, 258
 - performance, 242, 259
- Banking
 - industry, 242, 251, 252
 - AI application, 251
 - chatbot, 251
 - cross-selling, 252
 - risk assessment, 251
 - support management, 252
 - operations, 242
 - sector, 244, 251, 308
- Baroda bank, 253
- Base station (BS), 159
- Basic voice instructions, 246
- Behavior management, 278
- Behavioral surveillance, 258
- Best-in-class services, 247
- Big data, 19, 79, 105, 106, 149, 157, 159, 169, 289, 307, 311–314
 - analysis, 160
 - technology, 106
- Biodiversity loss, 285, 296
- Biological sciences, 245
- Black-box, 147, 252
- Block reward conditions, 15
- Blockchain, 5, 8, 9, 18, 19, 313
 - consensus protocols, 6
 - mining, 6, 10, 16, 17
 - technology, 4, 5, 8, 15, 18, 19, 146, 231, 313
- Bluetooth, 151
- Brain computer interface (BCI), 215
- Broadband communication, 215
- Broadcast, 11, 228, 336, 338
 - multimedia services, 324
- Business model, 305–307, 319
 - innovation, 307

C

C.V. Raman Global University (CGU), 73

Capital

conservation, 267

expenditures (CapEx), 52

Carbon

emissions, 267, 288

footprint, 52, 269

impression, 53

neutrality, 228

Cardiac monitoring devices, 151

Catastrophic events, 287

Cell

banking, 253

phones, 118, 125

Center for automation and robotics

(ANRO), 73

Central processing unit (CPU) operations,
156

abnormalities, 26

related metrics, 26

user space utilization, 26

Character recognition technology, 258

Chatbot, 242, 246, 251, 253

Chemical

engineering, 170

variations, 119

China's industrialization, 97

Chinese multinational technology, 122

City

environmental security, 84

public privacy, 73

risk evaluation, 76

Client

experience (CX), 248

retention, 252

satisfaction, 121, 242

segmentation, 258

spending research, 257

Climate

change, 266, 278, 285–287, 289, 294,
296, 299

evaluations, 158

Clinical disease identification, 231

Cloud

centralized nodes, 8

computation, 79, 144, 148, 150–152,
159, 160, 168, 169, 172, 190

cooperative big data assessment, 149

computing

amenities, 168

field, 190

system, 3, 4, 19, 26, 41, 51–54, 65,
117, 119, 120, 124, 126, 128–134,
144, 145, 152, 168, 169, 182

data

center (CDC), 51, 52, 56–58, 121,
126, 130

transmission, 125

databases, 42

environments, 41, 54

fuzzy environment, 25, 28, 32, 36, 37, 43

infrastructures, 25

organization dependability, 65

processing

resources, 150

techniques, 190

service, 51, 54

sim toolset, 52

system, 191, 201

utilization assistance, 148

Clustered workloads, 62

Co-authorship analysis, 276

Cognitive radio (CR), 159

Collision avoidance, 230

Commercial

bank, 242, 254, 257, 259

management platforms, 171

market, 306

model framework, 306

Communication, 156

computation platforms, 160

latency, 227, 228

spectrum, 145

technologies, 75, 77

Complex data administration, 325

Complicated computational requirements, 5

Comprehensive

architecture, 225

data sources, 286

Comprehensive network designs, 232

Computation, 29, 83, 109, 120, 126, 128,
133, 144, 146–150, 152, 154, 156–158,
160, 168, 169, 171, 172, 174–176, 178,
179, 189–192, 206, 214, 215, 219, 223,
225–227, 229, 334

assistance, 148, 156

- expenses, 158
- latency, 227
- resources, 156
- techniques, 154
- Computational
 - authority, 171
 - complexity, 134
 - power, 12, 28, 121, 145, 198
 - resources, 8, 104, 105, 108, 156
 - services, 126
 - simulations, 192
 - strength, 87
- Computer
 - implementation outcomes, 153
 - network technology, 245
 - power, 12, 131, 143
 - psychological research, 245
 - room air conditioning (CRAC), 58
 - services, 26
- Computerized metropolis, 75
- Computing
 - power
 - members, 12, 15
 - provider (CPP), 4–10, 12–15, 17, 19
 - trade framework, 5, 6
- Congestion stream management, 324
- Connection's latency, 330
- Conscious cloud resourceplanning, 54
- Consensus protocol, 6, 11
- Constrained
 - computation, 169
 - shortest pathway (CSP), 326, 327, 337
- Consumer
 - electronics, 156
 - experience (CX), 252
 - knowledge, 251
 - requirements, 306
 - segmentation, 258
- Contemporary
 - communication methods, 244
 - corporate management, 97
- Content distribution networks (CDN), 159
- Conventional
 - assessment, 180
 - cloud computing paradigm, 167
 - commercial manufacturing, 145
 - data designs, 170
 - management methods, 288
 - measurements, 176
 - networks, 325
 - NN method, 74, 85, 86
 - routing, 326
 - single-path routing, 325
 - studies, 76
 - web paradigm, 328
 - work scheduling methods, 134
- Convex loss function, 178
- Convolutional
 - auto-encoder, 32
 - neural network (CNN), 27, 30, 30, 33, 37, 43, 118, 134
- Cooling
 - management, 58, 61, 63
 - systems, 53, 58
- Cooperative automated driving (CDA), 218
- Coordinated resource administration, 156
- Corporate environmental management (CEM), 265
- Cosmology, 148
- Credit
 - risk, 251
 - score analysis, 251
- Crisis public security administration, 78
- Critical data transmission, 104
- Cross
 - layer optimization, 230
 - referencing, 291
 - tier disruption, 156
- Cryptographic
 - hashing, 18
 - security, 8
- Cuckoo
 - optimization (CO), 52, 53, 56, 59, 60, 65, 287
 - resource management (CORM), 56
 - search meta-heuristic (CSMH), 56, 57, 61–63
- Cultural elements, 312
- Customer
 - buying patterns, 248
 - focused digital banking, 243
 - relationship management, 242, 245, 247
 - satisfaction, 54, 248, 258
 - services, 248

Cutting-edge
 innovations, 81
 methods, 26
 networking innovation, 96
 technology, 81, 228, 229, 231, 253
 Cyber production, 172

D

D2D (device-to-device), 156
 Dartmouth Research Project, 307
 Data
 accessibility, 174, 315
 aggregation techniques, 96
 analysis techniques, 292
 analytics, 119, 214, 226, 286, 307
 assessment, 146, 150, 158, 176, 179
 breaches, 27
 caches, 157
 classification, 82, 145
 collection, 124, 291
 infrastructure, 298
 driven decision support system (DSS),
 296
 evaluation, 85, 149, 150, 152
 gathering, 148, 150, 160, 292
 governance, 225, 226, 311
 limitations, 298, 300
 locality, 218
 management, 81, 86, 89, 105, 106
 platform, 86, 89
 mining (DM), 169, 246, 268
 networks, 82
 owners, 226
 packet delay efficiency, 101
 preprocessing, 150
 privacy and security, 109, 314
 processing, 74, 80, 104, 119, 148, 149,
 169, 176, 190, 191, 250
 protection, 291, 314, 315
 quality, 173, 315
 resource management (DRM), 76
 routing, 98, 107
 sharing, 18, 81, 96
 sovereignty, 226
 stream and management hazards, 78
 system architecture, 168

transmission, 99, 107, 119, 325
 delays, 99
 latency, 119
 transport, 172, 175
 Database assistance, 148
 Datacenter
 cooling systems (DCS), 58, 62, 63, 65
 operational activities, 53
 Dataset visualization, 182
 Day-to-day technology, 125
 Debugging, 246
 Decentralization, 9, 169
 Decentralized
 blockchain manner, 6
 data structures, 218
 decision-making, 18
 learning techniques, 220, 231
 ledger system, 17
 neural network, 231
 Decision
 making, 75, 76, 98, 123, 125, 194, 202,
 206, 219, 269, 286, 289, 298, 299,
 308, 309, 313, 314, 334
 framework, 194
 processes, 75, 289, 298, 299
 speed and accuracy, 125
 trees (DT), 100, 153, 177, 258,
 309–314, 316, 317
 Deep
 feature extraction techniques, 30
 learning (DL), 9, 87, 88, 98, 100, 107,
 158, 169, 171, 216, 219, 223, 230, 231
 algorithms, 118
 capable video camera, 122
 methods, 35, 37, 122
 technology, 3, 4
 neural network (DNN), 118, 223, 230,
 231
 systems, 118
 transmitters, 231
 Deeper neural network layers, 118
 Deeply integrated communication and
 computation, 226
 Deforestation rates, 295
 Delay-constrained lowest-cost (DCLC), 326
 Demographic
 density, 76
 dimension, 76

Department of CSE (DS), 189
 Dependability, 42, 52, 59, 63, 64, 86, 87, 96, 98, 148, 174, 231, 331, 333, 335
 Deployment processes, 88
 Detrimental incidents, 27
 Deviation theory, 228
 Device participation, 217
 Device
 cloud cooperation mode, 119
 processing capacity, 132
 Digital
 beam forming (DBF), 221
 earth growth, 75
 educational platforms, 314
 government industries, 148
 libraries, 272
 twins, 314
 Dimensionality reduction, 30, 35
 Disaster preparedness, 297
 Discrete workloads, 62
 Dispersed
 computation system, 175
 computing resources, 7
 decision-making platforms, 206
 IoT platforms, 157, 160
 Distributed
 data
 points, 145
 storage, 9
 machine learning, 217
 optimization algorithms, 216, 217
 resource allocation policies, 230
 Distribution monitoring, 250
 Document analysis, 254
 Domain
 dependent applications, 146
 specific knowledge, 216, 230
 Dormant semantic investigation, 271
 Drilling smart sensors, 175
 Driverless vehicles, 230
 Drug development, 231
 Duration stamp technique, 330
 Dynamic
 communication, 217
 environment, 13
 mobile cloud environment, 133
 network connection, 133

node clustering, 107
 optic networking, 327
 resource allocation, 230
 routing, 96
 software development ML, 191
 thresholding, 40

E

Earthquakes, 296
 Eco-innovation duties, 278
 Ecological
 supportability, 270
 sustainability, 271, 278
 Ecosystem sustainability, 286
 Edge
 AI
 architecture, 225
 communication, 220
 enabled autonomous vehicles, 125
 hardware, 125
 resource allocation, 227
 appliances, 126
 applications, 126
 approach, 154
 cloud
 collaboration, 119
 design, 197
 fog, 189
 computation, 144–150, 152–154, 157, 160, 170–172, 175, 190, 191, 196, 197, 205
 assistance, 148
 capacities, 147
 design, 191
 infrastructures, 189
 innovation, 144
 instruments, 171
 patents, 152
 platforms, 192
 system framework detection, 147
 computers, 126
 computing (EC), 3, 4, 14, 19, 117–124, 126–129, 132, 134, 143–146, 148, 149, 152, 153, 160, 167–179, 181, 182, 189–191, 197, 206, 228, 231, 245
 applications, 143
 architecture, 169

- components, 182
- device, 128, 175, 176
- equipment, 169, 174, 175, 181
- framework, 190
- gadgets, 145
- innovation, 144
- jobs' offloading positions, 131
- resources, 133
- system, 133, 148, 149
- designs, 146
- device, 8, 120, 121, 126, 128–130, 132, 133, 217, 219–225, 228
- equipment, 132, 169, 171, 191, 193, 196–198, 203, 204
- gadgets, 126
- g-training, 222
- information processing innovation, 144
- IoT applications, 160
- learning
 - algorithms, 216
 - processes, 220
- machines, 171
- network functions, 227
- scheduler, 128–131, 133
- server
 - actual time conditions, 129
 - computing resources, 132
 - cooperative inference, 223
- technology, 122, 129
- training
 - process, 227
 - systems, 221, 227
 - wireless approaches, 220
- Education, 81, 84, 97, 105, 243, 252, 270, 277, 308, 311–313, 315
- over time, 252
- Educational framework, 79
- Effective
 - business environment management, 271
 - cloud resource management, 52
 - forest management, 288
 - video-transmitting network, 325
- Efficient resource utilization, 105
- E-health, 215
- Electric automobiles, 230
- Electronic
 - commerce (E-commerce), 242–249, 258, 259, 286
 - business, 247, 258
 - goliaths, 258
 - industry, 242, 247
 - intelligent system, 248–250
 - operations, 246
 - platform, 244, 258
 - sector, 242
 - data interchange (EDI), 245
 - mail, 245
 - payments, 245
 - trade, 244
 - transactions, 245
- Emails, 253
- Embedded
 - computer information, 96
 - mobile strength, 151
- Emotion detection, 244
- Empirical orientation, 79
- Employee productivity, 268
- Enabling technologies, 214, 225, 231, 232
- Encrypted data transmission, 82
- Endangered species, 289, 296
- End-to-end latency (E2E latency), 159, 323, 331
- Energy
 - aware resource
 - management strategies, 59
 - scheduling, 54
 - consumption, 43, 52, 54, 55, 58, 59, 63, 96, 108, 214, 226, 228, 229, 299
 - effectiveness, 156, 171
 - efficient, 96, 98, 99, 107, 109, 122, 123, 229, 230
 - algorithms, 110
 - protocols, 96
 - management, 62
 - model, 58
 - production, 228
 - productivity, 56, 65
 - resources, 98, 99, 107, 108, 223, 296
 - reuse
 - effectiveness, 58
 - proficiency, 62
 - trading framework, 4
 - usage, 228
 - utilization, 52, 55, 57, 65
 - viability, 53

Enhance energy storage systems, 296
 Enrollment, 267
 Enterprise
 performance management (EPM), 97
 resource planning (ERP), 97, 175, 176, 242, 245
 Environmental
 degradation, 285, 289
 manageability, 269
 management, 267, 277
 pollution, 43, 296
 privacy and security, 76
 protection, 296
 science, 290
 services, 287
 sustainability, 278
 systems, 306
 Environmentally sustainable human
 resource methods, 267
 Equal-cost multiple paths routing (ECMP), 327, 332, 334–337
 Esperanto technologies, 122
 Ethical
 AI practices, 298
 considerations, 108, 232, 289, 292
 implications, 315
 practices, 318
 Ethics, 297, 308, 316
 European Union (EU), 214, 317
 Experimental dataset, 26, 39
 Exploration data assessment (EDA), 176
 Extranet, 245
 Extreme gradient boosting (EGB), 178–180

F

FarmBot, 288
 Far-off data centers, 143
 Fault
 detection, 63, 64, 98
 management, 63
 Feature extraction, 30, 42
 Federated learning (FL), 8, 206, 217, 219, 220, 222, 230
 Femtocells, 159
 Fertilization, 296
 Field surveys, 295

Financial
 constraints, 298
 industry, 241
 institutions, 252
 management process, 252
 services sector, 19
 stability, 81
 Fintech, 259
 Fitbit fitness monitors, 145
 Floodlight, 330
 control system, 333–335
 system, 328
 Fog-to-cloud architecture, 156
 Food availability, 312
 Forecast
 energy production, 296
 environmental changes, 297
 water scarcity, 295
 Forecasting, 80, 85, 124, 217, 249, 250, 252, 312
 resource availability, 286
 Forestry industry, 288
 Fourth Industrial Revolution (4IR), 181
 Framework energy, 59
 Full offloading, 132
 Future innovations, 144, 146, 160

G

Gadget density, 190
 Game-changing technological advancement, 18
 Gaussian regression (GR), 30
 General data protection regulation (GDPR), 314, 319
 Generalized assignment issue (GAI), 191
 Genetic methods, 192
 Geographical
 data systems, 75
 study statistics, 80
 Geology, 148
 Geometric median, 221
 Global
 aggregation, 221
 GHRM strategies, 278
 implementation, 214
 learning model, 217

- population, 295
- warming, 267
- Good health and well-being, 310, 311
- Google, 52, 145, 168, 217, 243, 244, 253
 - cloud, 168
- Government
 - administration, 74
 - agencies, 315
 - officials, 291
- Gradient-oriented methods, 179
- Graph neural networks (GNN), 230
- Graphical representation, 15
- Graphics processing units (GPUs), 118, 123, 228
- Green
 - computation theory, 191
 - finance, 271
 - human resource management (GHRM), 265–279
 - marketing, 271
 - retailing, 271
- Greenhouse gases, 287
- Ground-breaking
 - projects, 271
 - wireless technologies, 216
- Gyroscopes, 151

H

- Habitat loss, 294
- Hardware and software platforms, 214, 216
- Hardware setup, 153
- Hash encoding credentials, 42
- Healthcare, 19, 74, 97, 108, 148, 229, 308, 312, 318
- Heterogeneous
 - devices, 127
 - edge equipment, 193
- Heuristic-oriented
 - methods, 169
 - options, 190
- Hidden Markov model (HMM), 30
- High
 - computational power, 28
 - fidelity holographic society, 215
 - performance
 - computers, 144
 - computing, 56

- quality datasets, 315
- resolution, 286
- stakes applications, 219
- Holistic approach (HA), 55
- Holistic
 - energy optimization framework (HEOF), 55
 - workload scaling (HWS), 55
- Horizontal edge inference, 222, 223
- Host heterogeneity, 53
- Hosting application services, 51
- Household wealth, 81
- Human
 - action acknowledgment, 151
 - activity, 144, 150, 151, 160, 267
 - AI collaboration challenges, 299
 - assessments, 308
 - body acknowledgment, 150
 - comprehension, 241
 - intelligence, 242, 245, 286, 308
 - replication, 241
 - intervention, 28, 229
 - machine interaction (HMI), 119
 - mobility, 152
 - resource
 - department, 247, 278
 - management (HRM), 245, 265–275, 278
 - policies, 270
- Hurricanes, 296
- Hybrid beam forming, 221
- Hydrological cycle, 287
- Hypothetical underpinnings, 271

I

- IaaS (infrastructure as a service), 57
- ICT (information and communication technologies), 75, 81
- I-Fog-Sim, 192
- Illegal accessibility, 87
- Image processing, 27
- Inconsistent data collection methods, 315
- Independent data plane, 226
- Indian financial services, 253
- Industrial
 - attackers, 82

- automation, 95, 108
 - software sensors, 180
 - water treatment, 36
 - Industrialization, 97, 267
 - Industry
 - competitiveness, 249
 - stakeholders, 108
 - Information
 - driven devices, 172
 - science, 121
 - technology, 75, 97, 247
 - transmission, 206
 - Infrared visual collection devices, 151
 - Infrastructure upgrades, 123
 - Innovative
 - business models, 307
 - data processing methods, 229
 - Institutional review board (IRB), 291
 - Integrated
 - smart IIoT infrastructure, 42
 - space-air-ground networks (SAGINs), 222
 - Intellectual property (IP), 147
 - Intelligence operations, 308
 - Intelligent
 - AI-enabled systems, 126
 - assistant using AI, 246
 - automation, 118
 - connectivity, 231
 - data gathering devices, 42
 - factories, 106
 - healthcare, 231
 - logistics, 247
 - machines, 119, 241, 242
 - networking, 230
 - optimization algorithms, 134
 - technology, 245
 - transportation management (ITM), 123, 124
 - vehicles, 230
 - Interactive
 - applications, 143
 - cloud-based gaming, 191
 - Inter-cluster connectivity, 158
 - Intercontinental connection, 42
 - Interdisciplinary
 - collaboration, 290, 315
 - frontier science, 245
 - Intermediary networking operations, 324
 - International
 - domains, 96
 - forums, 267
 - organizations, 267
 - telecommunication union (ITU), 214
 - Internet, 3, 4, 52, 59, 106, 121, 123, 126, 143, 144, 148, 172, 214, 215, 245, 307, 324, 325
 - banking, 253
 - commerce, 246
 - computing facility, 125, 126
 - protocol television (IPTV), 324
 - Interoperability, 110
 - Interpretability, 219, 228
 - Intranet, 245
 - Intrusion
 - detection system (IDS), 42
 - prevention systems (IPS), 42
 - Inventory data, 251
 - IoMT (internet of medical things), 231
 - IoT (internet of things), 3, 4, 8, 52, 81, 87, 96, 105, 106, 117, 118, 121, 123, 125, 126, 143–146, 148, 156–160, 168–172, 174, 175, 189–193, 197, 214, 223, 229–231, 239, 307
 - applications, 143, 145, 146, 157
 - devices grouping, 158
 - IoV (internet of vehicles), 169
- J**
- Jingdong, 258
 - Job execution status, 128
 - Joint source-channel coding (JSCC), 216
 - Joint stretching, 119
 - Journalism, 148
- K**
- Kilowatt hours (kWh), 62
 - K-mean clustering, 266, 276
 - Knapsack-oriented scheduling system, 192
 - K-nearest neighbors (kNN), 101–105
 - Knowledge management, 226
- L**
- Lab
 - advancement scientists, 174
 - data administration structure (LDAS), 175

- Labeled data, 27, 30, 31
 - Labor
 - force, 81
 - intensive task, 28
 - sectors, 266
 - Lagrange relaxation-oriented aggregating cost (LARAC), 326
 - Landslides, 82
 - Language study, 245
 - Large scale
 - deployments, 99
 - user environment, 17
 - visual recognition challenge (LSVRC), 118
 - Latency, 4, 55, 65, 78, 96, 98–100, 105, 107, 109, 122, 126–129, 144, 145, 148, 158, 171, 214, 221–224, 227–231, 324, 326, 328, 330, 332–335, 338
 - constraints, 128, 228
 - restriction, 145, 334
 - sensitive needs, 158
 - variance, 326
 - Latent semantic analysis (LSA), 266, 272, 273, 278
 - Learning
 - accuracy, 228
 - algorithms, 15, 216
 - blockchain, 9
 - challenge, 11
 - mining service, 9
 - oriented methods, 158
 - puzzle, 11
 - Legal
 - implications, 88, 226
 - ramifications, 88
 - Legislatures, 267
 - Leukemia, 231
 - Limited
 - generalization, 298
 - machine learning power, 124
 - Link layered discovery standard (LLDS), 330
 - Linkages, 308
 - Linux operating system, 82
 - Load balancing, 107
 - Local
 - government employees, 88
 - implementation, 132
 - Logical logistics, 247
 - Logistic management, 250
 - Long
 - short-term memory (LSTM), 30
 - term
 - continual pricing change, 247
 - environmental sustainability, 278
 - Low
 - cost 6G devices, 106
 - error designs, 176
 - latency communication, 105
- M**
- Machine
 - intelligence, 245
 - learning (ML), 27–29, 31, 37, 41, 77, 87, 96, 98, 100, 107, 121–124, 145, 149, 150, 157, 158, 168–177, 179–182, 189, 190, 195, 205, 206, 214, 216, 217, 219, 226, 227, 232, 244, 246, 247, 266, 286, 288, 295, 308
 - algorithm, 27–29, 37, 168, 170, 171, 173, 174, 176–182, 226, 246, 247, 266, 295, 296
 - assistance, 174
 - methods, 37, 226
 - RL convergence, 191
 - Major financial institutions, 252
 - Management information system (MIS), 245
 - Managerial decision-making prejudice, 76
 - Market
 - research, 30
 - segmentation, 257
 - Markov decision procedure (MDP), 194
 - Mars science laboratory (MSL), 36
 - Massive scale edge systems, 220
 - Mathematical
 - frameworks, 308
 - programming, 134
 - Mean
 - floodlight capability, 330
 - latency per packet, 129
 - score (MS), 310, 311
 - Medical
 - care, 78, 96, 119, 145

- crisis administration, 78
 - imaging, 231
 - professionals, 19
- Medicinal herbs, 287
- Merchant
 - activity, 251
 - rating, 251
- Metaheuristic computations, 192
- Methodology, 7–9, 29, 55, 77, 79, 99, 100, 125, 126, 128–130, 146, 147, 149, 151, 152, 172–175, 182, 192, 194, 195, 216, 217, 219, 268, 272, 273, 290, 292, 314, 328, 331–333
 - data, 99, 151, 152
 - collection, 149
 - source, 151
 - cloud computing, 130
 - DRM, 76, 78–80, 82, 89
 - edge
 - cloud computing, 129
 - computing basics, 126
 - learning algorithms, 217
 - federated edge AI learning, 217
 - framework distribution, 7
 - implement, 273
 - KNIME process, 273
 - logic board scenarios, 126
 - partitioned model learning, 219
 - power provider layer, 8
 - problem formulation, 9
 - secured data collection, 174
 - system
 - framework, 147, 328
 - model, 192
- Microscopic particles, 266
- Microsoft, 52, 122, 168, 190, 244
- Minimal
 - energy usage, 150, 204
 - equipment density situations, 201
 - package loss rate, 324
- MiniMax Q-learning, 15
- Mininet simulation, 333
- Mining risk costs, 12
- Mission-critical applications, 224
- Mixed-method research
 - approach, 290
 - design, 291
- Mobile
 - applications, 156, 313
 - internet capabilities, 215
 - internet services, 225
 - network, 225
 - deployment, 225
 - restrictions, 160
 - sensor nodes, 106
- Mobility
 - administration, 159
 - properties, 159
- Model
 - aggregation, 221, 229
 - creation, 175, 181
 - inversion attacks, 220
 - methods, 219
 - poisoning, 220
 - surveillance, 174
 - verification, 182
- Money laundering
 - activities, 217
 - prevention, 242
- Monopolistic tendencies, 5
- Moral awareness, 84
- Movement modeling, 171
- Multi-agent collaborations, 230
- Multi-cloud storage framework, 55
- Multifunctional management strategy, 81
- Multi-head attention mechanism, 43
- Multi-layer
 - methodology, 195
 - RL approach, 190, 202, 203, 205
 - technique, 202, 205
- Multimedia
 - databases, 190
 - services, 325
- Multi-path
 - innovation, 335
 - setting, 336
- Multiple
 - constrained pathway (MCP), 326, 327
 - network services, 159
 - path
 - algorithm, 331
 - routing techniques, 332
 - routing, 328
 - transmissions, 337
 - sensor imaging platform, 76

Multi-role capability, 6, 7
Multivariate time series, 26, 32, 36
data, 36

N

Naive Bayes (NB), 153, 154
Nash equilibrium, 6
Nation's economy, 267
Natural
catastrophe, 78, 297
surveillance, 79
catastrophic events, 82
ecosystem, 292–294
language processing (NLP), 8, 253, 268,
269, 271, 275, 278, 279
maintainability, 271
resource management (NRM), 285, 287,
289–292, 295–300

Network
administration, 158, 225
advertising, 245
applications, 26, 42
bandwidth, 64, 119, 132, 143, 327
congestion, 214
data collection situations, 144
dynamics, 107, 337
efficiency, 325, 336
energy usage, 98
examination, 278
functionality, 225
infrastructure, 225, 226
intrusion detection, 42
latency, 145
layer variations, 224
longevity, 107
optimum connection consumption, 333
performance, 105, 109
related data, 329
scalability, 99
sensor nodes' energy consumption, 106
simulations, 100
specifications, 146
state, 128
storage, 144
surveillance, 330
throughput, 327
topologies, 100, 217, 227

Networked
financial system, 81
intelligence, 214
populations, 81
Networking
connection state, 336
divisions, 328
latency, 337
resource, 168
consumption, 331
topology, 329, 330, 334
virtualization, 159
Neural network (NN), 8, 9, 11, 27, 43, 74,
76, 85, 86, 89, 100, 122, 123, 158, 191,
206, 230, 244
architecture optimization, 222
New network data analytics function
(NWDAF), 226
Nitrogen design, 176
Non-destructive forms, 176
Non-inexhaustible sources, 58
Nonlinear iteration
assessment, 179
PLS (NIPLS), 179
Non-uniform fading characteristics, 221
North-bound-application software interface
(NBI), 330
Novel
computational paradigm, 167
computing frameworks, 8
Numerical optimization, 246

O

Object identification, 30
Offloaded activities, 200
Offloading decision process, 205
Omni anomaly, 35, 37
On-demand networking, 172
Online
application, 180
digital medium, 309
surveys, 291
trading, 241
Open
flow
centralized management function, 337
control system, 331

- platform, 331
 - regulations, 328–330
 - regulator, 159, 324
 - specification, 328
 - switch, 330, 335
 - switching device, 327, 329, 335
 - source
 - cloud platform, 41
 - software, 271, 273
 - stack, 41
 - Operational
 - control and surveillance, 170
 - data, 251
 - efficiency, 244
 - risks, 251
 - Operations research (OR), 232
 - Optical automation inspection (AOI), 125
 - Optimal pricing, 247
 - Optimality, 227, 228
 - Optimization targets, 121, 133, 217
 - Optimize grid management, 296
 - Optimum routing regulations, 330
 - Order
 - delivery efficiency, 249
 - history, 249
 - Organizational
 - assistance, 105
 - bottlenecks, 317
 - environmental sustainability, 277, 278
 - management techniques, 97
 - Osmotic computing features, 172
 - OTO (online to offline), 249
 - Outlier identification, 176
 - Overhead, 109, 332
 - Overpopulation, 88
- P**
- Package loss ratio, 330, 331, 334, 336
 - Packet
 - delivery delay, 102, 104, 105
 - switching networking, 324
 - Partial
 - least squares (PLS), 179
 - offloading, 133
 - Participative smart environment develop-
ment, 75
 - Particle swarm optimization (PSO), 53
 - Partitioned model learning, 219
 - Patent, 144, 145, 147, 152, 160
 - Path computation, 331
 - Peace, justice, and solid institutions, 310
 - Peaks over threshold (POT), 26, 27, 29,
40, 43
 - Peer-to-peer interactions, 217
 - Performance
 - assessment, 59, 196
 - metrics, 100
 - reward factor's (PRF), 17
 - surveillance, 182
 - Personal computer (PC), 175, 176
 - Personalized requirements, 19
 - Person-centered preventive activities, 78
 - Photo recognition systems, 245
 - Physical
 - devices/sensors, 168
 - hardware based machines, 41
 - Picture production, 118
 - Piezoelectric and tri-bioelectric bases, 119
 - Pillar arrangement, 42
 - Planet's finite resources, 295, 298
 - Platform as a service (PAAS), 57, 159
 - Poaching, 289, 294
 - PoDL (proof of deep leaning), 6
 - Poisson distribution, 11
 - Policy
 - approaches, 219
 - making, 290
 - Population dynamics, 296
 - Portable innovations, 191
 - Potent computation architecture, 160
 - Potential applications, 100, 214, 232
 - PoW (proof of work), 11
 - Power
 - allocation, 7, 16
 - capacities, 106
 - computing resources, 8
 - effectiveness, 88
 - limitations, 228
 - management units (PMUs), 57
 - provider layer, 8
 - standard unit price, 10
 - trading, 3–9, 12–14, 19
 - computing mechanisms, 3

- framework, 5–7, 14, 19
- networking paradigm, 6
- usage restrictions, 171
- utilization, 204
- Practical
 - applications, 153, 291
 - implementations, 19
- Precise fertilization, 288
- Precise
 - pesticide treatments, 288
 - spatiotemporal data, 157
- Precision agriculture (PA), 288
- Predictive maintenance (PdM), 96, 121
- Price policy, 13, 14
- Primary
 - revenue streams, 252
 - terrestrial ecosystems, 287
- Principal component analysis (PCA), 30
- Privacy and security prioritization, 87
- Private sector, 306, 315
- Probabilistic long-term system effectiveness, 158
- Problem formulation, 9
- Procedure systems engineering (PSE), 170
- Procurement strategies, 9
- Product
 - content management (PCM), 242, 248
 - recommendations, 246
- Productivity variables, 278
- Profit distribution, 5
- Profit-sharing mechanism, 5
- Propagation conditions, 227
- Property management, 74
- Proprietary traders, 252
- Pseudo-labeled samples, 31
- Public
 - assets, 88
 - authority, 270
 - cloud providers, 52
 - healthcare crisis administration, 78
 - privacy, 73–80, 82–84, 86, 88, 89
 - safety administration, 78
 - utilities, 74
- Pure-Edge-Sim, 197
 - simulation, 190, 205
 - standards, 198
- Python, 14, 36, 82, 83

Q

- Q-learning, 13–15, 100, 192, 194, 206, 219
 - standard algorithm, 15
- QoS (Quality of service), 25, 54, 57, 63, 65, 95–100, 107–110, 127, 129, 156, 157, 323–329, 333–335, 337, 338
 - adaptable video single-cast, 325
 - adaptive method, 324
 - application, 63
- Qualitative
 - data, 291
 - studies, 317
- Quality
 - assurance (QA), 97, 168, 172, 173, 182
 - framework, 170
 - management, 5
- Quantifiable performance metrics, 30
- Quantitative
 - data, 291
 - gathering devices, 151
 - evaluations, 174
 - examination, 271
- Queuing latency, 230

R

- Radio
 - access network (RAN), 156, 226
 - connections, 193
 - resource, 159
 - administration, 77
- Random forest (RF), 153, 177–179
 - regression (RFR), 178, 180
- Randomization, 81
- Real
 - time
 - analysis, 242
 - applications, 98, 104
 - conditions, 96
 - data analysis, 109, 148, 150, 230, 247, 251, 294
 - efficiency, 202
 - evaluation, 87
 - immersive experiences, 214
 - information transfer, 148
 - monitoring systems, 82, 105
 - multimedia applications, 324

- reaction, 144
 - requirements, 222
 - training information, 158
 - video analysis, 191
 - world
 - applications, 8
 - hardware, 125
 - WSNs, 98
 - Recognizing faces, 244
 - Reconstruction errors, 32
 - Record transactions, 18
 - Recruiting, 269, 270, 278
 - Recursive support vector regression (RSVR), 101–105
 - Reduce energy usage, 57
 - Referenced research domains, 278
 - Regular data points, 32
 - Regulatory
 - compliance, 232
 - frameworks, 316
 - Reinforcement learning (RL), 8, 13, 14, 98, 100, 190–192, 194–197, 200, 202–206, 219, 230
 - Relationship
 - development, 251
 - improvement, 251
 - Reliability, 26, 28, 52, 55, 56, 59, 60, 63, 65, 99, 100, 106, 107, 153, 169, 222, 224, 228
 - Reliable
 - aggregation techniques, 220
 - learning, 219
 - Remote sensing, 286
 - data, 36
 - satellites, 295
 - technologies, 296
 - Renewable energy, 58, 61, 62, 287, 296
 - production, 297
 - Reproduction threshold, 178
 - Research
 - allocation, 5, 59, 88, 107, 214, 216, 227, 229, 286, 289, 296, 315, 337
 - availability, 132, 289, 295
 - conservation, 296
 - constrained
 - computing nodes, 6
 - environments, 105, 107, 108
 - inference, 223
 - nodes, 108
 - vertical edge inference, 223
 - wireless sensor networks, 104
 - constraints, 108
 - depletion, 285, 287, 289
 - design, 291
 - distribution, 145
 - exploitation, 287, 295
 - institutions, 231
 - intensive hurdle, 28
 - limited vertical edge inference, 225
 - management, 52–54, 59, 61, 65, 76, 86, 232, 265, 270, 287–292, 294–299
 - aligned, 229
 - solution, 52
 - strategies, 289, 299
 - methodology, 99
 - data collection, 99
 - scarcity, 286, 296
 - utilization, 204, 297
 - Restricted resources, 106
 - Retailer administration, 250
 - Retaining essential information, 39
 - Reward-based multi-agent learning, 219
 - Risk
 - alert accuracy, 86
 - mitigation strategies, 78
 - Robotic vehicles, 125
 - Robots, 241, 245
 - Robust, 4, 26, 29, 36, 109, 174, 228
 - technological infrastructures, 290
 - Role-playing ratio, 16, 17
 - Root-mean-square deviation (RMSD), 178
 - Round robin multiple-path technique, 327
- ## S
- SAAS (software as a service), 57
 - Safeguards data processing, 123
 - Safety management, 105
 - Sampling bias, 292
 - Scalability, 28, 42, 96, 98, 99, 107, 109, 121, 122, 218, 231, 288
 - Scalable edgeAI technology, 214
 - Scale network optimization, 226
 - Scaled SiGe bipolar technology, 228

- Schedule
 - devices, 227, 228
 - identification, 250
- Scope quantification, 61
- Scopus, 272
- Secondary data, 242
 - collection, 254
- Security
 - administration, 76–80, 89
 - platform, 76
 - communications network, 77
 - data
 - indications, 74
 - resources management platform, 74
 - management
 - platform, 86
 - studies, 89
 - performance, 104
 - reinforcement, 96
- Selection technology, 29, 267
- Self
 - centeredness, 5
 - driving cars, 125, 191
 - healing capabilities, 107
 - interested behavior, 3
 - organization, 98, 107
 - resilient computation structures, 160
- Semantic communication, 216, 230
- Semi-supervised learning (SSL), 29, 31
- Sensitivity, 29, 41, 43, 179, 273
 - analysis of parameters, 35
 - assessments, 179
- Server-based cloud-empowered design (SCA), 55
- Service
 - consolidation, 52
 - driven resource allocations, 225
 - level compliance scoring, 145
- Signal processing operations, 157
- Single
 - agent reinforcement learning (SARL), 13
 - cast routing, 326
 - path methods, 334
- Skills gap, 315
- Small
 - cell cloud-enhanced eNodeB, 157
 - scale experimental settings, 99
- Smart
 - city, 74, 75, 81, 96, 107
 - monitoring, 80
 - planning, 75
 - contracts, 19
 - edge-IoT applications, 157
 - electrical grids, 308
 - environment, 74–77, 79–82, 86–89
 - assessment study, 87
 - economies, 77
 - installations, 81
 - platform, 87
 - public privacy, 76
 - services, 81
 - technologies, 79
 - homes, 18
 - houses, 145, 148
 - multimedia, 171
 - parking management (SPM), 123
 - phone banking, 253
 - technology, 248
 - transportation, 123
- Social
 - care, 286
 - consumption, 244
 - dimension, 310
 - discrimination, 313
 - equity, 318
 - intelligence, 307
 - manifestations, 313
 - networks, 286
- Societal
 - dissatisfaction, 76
 - hazards worldwide, 78
 - movement, 76
 - privacy, 81
 - security, 77, 81
 - threats, 77
- Software
 - described networking (SDN), 146, 148, 159, 324, 325, 327, 328, 332, 333, 338
 - sensor, 169, 170, 175, 176, 180–182
 - development, 168, 182
- Soil erosion, 295
- Solar and wind power, 296
- Space-air-ground integrated network (SAGIN), 222

- Spatial multiplexing (SM), 221
 - Specific application domains, 232
 - Speech
 - labeling, 273
 - recognition, 8, 244
 - Standard
 - deviation (SD), 129, 152
 - sampling technique, 330
 - traditional programming, 307
 - Standardized dataset, 11
 - State Bank of India (SBI), 253
 - intelligent assistant (SIA), 253
 - Static thresholds, 27
 - Statistical
 - analysis, 156, 291
 - definitions, 194
 - Statistics, 35, 74, 80, 84, 87, 144, 168, 174, 176, 246, 291, 330
 - Storage capacity, 131, 153, 222
 - Stream of traffic latency constraints, 338
 - Substantial
 - delay, 168
 - financial losses, 27
 - Supervised learning, 29–31, 195
 - Supply chain management (SCM), 245
 - systems, 106
 - Support vector machine (SVM), 30, 101–105, 158
 - Surveillance
 - data, 80, 330–332
 - footage, 28, 30
 - video anomaly detection system, 28
 - Survey questionnaire, 291, 292
 - Sustainability, 42, 51, 54, 57, 75, 88, 267, 269, 270, 274, 278, 279, 289, 290, 292, 297, 299
 - Sustainable
 - cities, 215
 - development (SD), 152, 266, 269, 270, 285–292, 294–300, 307, 315–319
 - agenda, 306
 - goals (SDGs), 270, 286, 290, 295, 296, 298, 299, 306–310, 314–319
 - goods and services, 308
 - logging practices, 297
 - organizational culture, 270
 - value proposition, 307
 - Swarm learning, 218, 231
 - Symbiosis International (Deemed) University (SIU), 285
 - Synergistic relationship, 286
 - System
 - adaptability, 87
 - architecture, 331
 - computing resources, 204
 - durability, 173
 - edge computation, 150
 - performance, 148, 159, 205
 - running, 174
 - System-wide information management (SWIM), 157
- ## T
- Tamper-proof nature, 5
 - Target edge technology, 129
 - Task
 - assignment issue (TAI), 190–193, 205
 - completion time, 127
 - implementation
 - duration, 205
 - performance rate, 201
 - offloading, 121, 157, 205, 206
 - oriented communication, 223, 225
 - performance rate, 199
 - Technical
 - apparatus, 247
 - expertise requirements, 298
 - innovation capacity, 147
 - Technological infrastructure, 315
 - Teleconferences, 267
 - Telemetric box, 124
 - Temperature assessments, 53
 - Terrestrial biodiversity, 287
 - Terrestrial communication, 222
 - Text mining, 266, 273, 278, 279
 - applications, 273
 - The phrase frequency-inverse document frequency (TF-IDF), 275, 276
 - Three convolutional neural networks, 26
 - Time
 - complexity, 103–105
 - latency
 - efficiency, 335
 - restriction, 332
 - sensitive wireless connections, 160
 - stamped records, 18

Total
 computation resources, 156
 equipment efficiency, 180
Track markets, 18
Trade-off factor-beta, 204
Traditional
 communication methods, 230
 corporate cost management methods, 97
 data collection, 288
 human resource management techniques, 266
 linear approaches, 43
 logistics, 247
Traffic
 congestion, 123
 data surveillance, 83
 management component, 330
 monitoring/ prediction (TMP), 123
Training-to-testing ratio, 36
Train-test splits, 176
Transactional processes, 19
Transmission sequence numbers (TSNs), 101
Transport layer security (TLS), 42
Troubleshooting, 174
Troublesome variables, 181
Tuberculosis (TB), 231
Two-stage Stackelberg game, 12
Typhoon, 82

U

Ultrafast data processing, 125
Ultra-large MIMO (XX-MIMO), 221
Ultra- reliable low latency communication (URLLC), 216, 230
Unintentional catastrophes, 78
Uninterruptible power supply (UPS), 58
United Nations General Assembly (UNGA), 306
Unlabeled data, 31
Unsupervised learning, 29–31
Untapped market, 307
Upstream drilling, 180
Urban
 air quality, 84
 crisis, 77
 data, 79, 84
 diplomacy security, 84
 planning, 286

 populace, 88
 public
 privacy, 74, 76–78
 security, 76, 78
 security measures, 76
Urbanization, 73, 75, 76, 88, 285, 287

V

V2V (vehicle-to-vehicle), 222
Value-based techniques, 219
Vehicle
 communication, 230
 edge-cloud architecture, 124
Vehicular communications, 222
Vertical edge inference, 223, 225
Video
 anomaly detection, 28
 broadcasting, 323, 325, 328, 334
 congestion, 325
 footage, 28
 services, 323, 338
 surveillance, 28
 transfer's dependability, 325
Virtual
 machine (VM), 41, 56, 58, 159
 scheduling (VMS), 54
 object system (VOS), 273, 277
 reality (VR), 191, 312
Virtualization, 62
 cloud datacenter, 55
 productivity, 62
 technologies, 26
Vision-safety technologies, 122
Visual artificial intelligence platform, 258
Visualization instrument, 176
Voice over IP (VoIP), 324
Vulnerable
 security feature, 124
 sensor nodes, 106

W

Water
 cooling management model, 58
 grade, 78
 quality issues, 289
 resource management, 289

- Wearable gadgets, 159
 - Weather patterns, 289
 - Web
 - application, 172, 180
 - connectivity, 325
 - multimedia applications, 326
 - services, 148
 - Wi-Fi-free digital lab, 253
 - Wildlife
 - conservation, 289, 296
 - populations, 293, 294
 - Wireless
 - channel conditions, 228
 - communication, 107, 214, 216, 228
 - computation techniques, 220
 - connectivity, 231
 - detectors, 148
 - federated learning, 228
 - network, 104, 105, 213, 220, 221
 - infrastructures, 225
 - sensor
 - environments, 98
 - network (WSN), 81, 95–100, 104, 105, 107–110
 - signal identification, 172
 - technology, 221
 - transmission techniques, 228
 - Wolff's principle, 4, 13, 15, 19
 - World economic forum (WEF), 308
 - World's security risk administration, 80
 - Worldwide associations, 270
 - Worldwide Diary of Labor supply, 275
- Z**
- Zero hunger, 310, 311
 - Zhong Kui technology, 258