

Published in Journals: Applied Sciences, Journal of Sensor and Actuator Networks, Sensors, Photonics and Telecom

Topic Reprint

Machine Learning in Communication Systems and Networks, 2nd Edition

Edited by
Yichuang Sun, Haeyoung Lee and Oluyomi Simpson

mdpi.com/topics



Machine Learning in Communication Systems and Networks, 2nd Edition

Machine Learning in Communication Systems and Networks, 2nd Edition

Topic Editors

Yichuang Sun

Haeyoung Lee

Oluyomi Simpson



Basel • Beijing • Wuhan • Barcelona • Belgrade • Novi Sad • Cluj • Manchester

Topic Editors

Yichuang Sun

School of Physics, Engineering
and Computer Science

University of Hertfordshire

Hatfield

UK

Haeyoung Lee

School of Physics, Engineering
and Computer Science

University of Hertfordshire

Hatfield

UK

Oluyomi Simpson

School of Physics, Engineering
and Computer Science

University of Hertfordshire

Hatfield

UK

Editorial Office

MDPI AG

Grosspeteranlage 5

4052 Basel, Switzerland

This is a reprint of the Topic, published open access by the journals *Applied Sciences* (ISSN 2076-3417), *Journal of Sensor and Actuator Networks* (ISSN 2224-2708), *Sensors* (ISSN 1424-8220), *Photonics* (ISSN 2304-6732) and *Telecom* (ISSN 2673-4001), freely accessible at: <https://www.mdpi.com/topics/62GBDT9S0F>.

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

Lastname, A.A.; Lastname, B.B. Article Title. <i>Journal Name</i> Year , Volume Number, Page Range.
--

ISBN 978-3-7258-8819-1 (Hbk)

ISBN 978-3-7258-8820-7 (PDF)

<https://doi.org/10.3390/books978-3-7258-8820-7>

Contents

About the Editors	vii
-----------------------------	-----

Preface	ix
-------------------	----

Haeyoung Lee, Yichuang Sun and Oluyomi Simpson

Machine Learning for Intelligent and Adaptive Communication Systems: From Optimization to Emerging Paradigms

Reprinted from: <i>Sensors</i> 2026 , 26, 2882, https://doi.org/10.3390/s26092882	1
--	---

Christian Szczerba, Enrique Dávalos, Ariel Leiva and Juan Pinto-Ríos

Early Fault Detection in a Real Scenario of Hybrid Fiber–Coaxial Networks Using Machine Learning: An Approach Based on Decision Trees and Random Forests

Reprinted from: <i>Applied Sciences</i> 2025 , 15, 10442, https://doi.org/10.3390/app151910442	6
--	---

Anya Apavatjirut

Sensor-Driven RSSI Prediction via Adaptive Machine Learning and Environmental Sensing

Reprinted from: <i>Sensors</i> 2025 , 25, 5199, https://doi.org/10.3390/s25165199	18
--	----

Tianxiao Wang, Yingtao Niu and Zhanyang Zhou

Few-Shot Intelligent Anti-Jamming Access with Fast Convergence: A GAN-Enhanced Deep Reinforcement Learning Approach

Reprinted from: <i>Applied Sciences</i> 2025 , 15, 8654, https://doi.org/10.3390/app15158654	34
---	----

Chao Tang, Dashun He and Jianping Yao

Distributed Interference-Aware Power Optimization for Multi-Task Over-the-Air Federated Learning

Reprinted from: <i>Telecom</i> 2025 , 6, 51, https://doi.org/10.3390/telecom6030051	53
---	----

Mays A. Mawlood and Dhari Ali Mahmood

Optimizing Weighted Fair Queuing with Deep Reinforcement Learning for Dynamic Bandwidth Allocation

Reprinted from: <i>Telecom</i> 2025 , 6, 46, https://doi.org/10.3390/telecom6030046	78
---	----

Ricardo Suarez del Valle, Abdulkadir Kose and Haeyoung Lee

Context-Aware Beam Selection for IRS-Assisted mmWave V2I Communications

Reprinted from: <i>Sensors</i> 2025 , 25, 3924, https://doi.org/10.3390/s25133924	115
--	-----

Rui Chen, Yijun Ma, Zhong Wang and Shengli Sun

Incoherent Optical Neural Networks for Passive and Delay-Free Inference in Natural Light

Reprinted from: <i>Photonics</i> 2025 , 12, 278, https://doi.org/10.3390/photonics12030278	131
---	-----

Gongyu Xia, Jiacheng Liu, Qilin Hong, Pingyu Zhu, Ping Xu and Zhihong Zhu

An Efficient Frequency Encoding Scheme for Optical Convolution Accelerator

Reprinted from: <i>Photonics</i> 2025 , 12, 26, https://doi.org/10.3390/photonics12010026	147
--	-----

Pablo López-Muñoz, Luis Gimeno San Frutos, Christian Abarca, Francisco José Alegre, Jose Luis Calle and Jose F. Monserrat

Hybrid Artificial-Intelligence-Based System for Unmanned Aerial Vehicle Detection, Localization, and Tracking Using Software-Defined Radio and Computer Vision Techniques

Reprinted from: <i>Telecom</i> 2024 , 5, 64, https://doi.org/10.3390/telecom5040064	157
---	-----

Tahereh Shah Mansouri, Gennady Lubarsky, Dewar Finlay and James McLaughlin Machine Learning-Based Structural Health Monitoring Technique for Crack Detection and Localisation Using Bluetooth Strain Gauge Sensor Network Reprinted from: <i>J. Sens. Actuator Netw.</i> 2024 , 13, 79, https://doi.org/10.3390/jsan13060079 . . .	180
Zainab AlMania, Tarek Sheltami, Gamil Ahmed, Ashraf Mahmoud and Abdulaziz Barnawi Energy-Efficient Online Path Planning for Internet of Drones Using Reinforcement Learning Reprinted from: <i>J. Sens. Actuator Netw.</i> 2024 , 13, 50, https://doi.org/10.3390/jsan13050050 . . .	196
Zhao Ma, Shengliang Fang, Youchen Fan, Shunhu Hou and Zhaojing Xu Tackling Few-Shot Challenges in Automatic Modulation Recognition: A Multi-Level Comparative Relation Network Combining Class Reconstruction Strategy Reprinted from: <i>Sensors</i> 2024 , 24, 4421, https://doi.org/10.3390/s24134421	222
Xin Zheng, Ying He, Chong Zhang and Pu Miao VLCMnet-Based Modulation Format Recognition for Indoor Visible Light Communication Systems Reprinted from: <i>Photonics</i> 2024 , 11, 403, https://doi.org/10.3390/photonics11050403	242
Bianca S. de C. da Silva, Victoria D. P. Souto, Richard D. Souza and Luciano L. Mendes A Survey of PAPR Techniques Based on Machine Learning Reprinted from: <i>Sensors</i> 2024 , 24, 1918, https://doi.org/10.3390/s24061918	260

About the Editors

Yichuang Sun

Yichuang Sun is a Professor of Communications and Electronics in the School of Physics, Engineering and Computer Science at the University of Hertfordshire, UK. His research expertise and interests are in the areas of wireless and mobile communications, 6G and beyond technologies, RF systems and circuits, analogue and mixed-signal circuits, memristor circuits and systems, neural networks, machine learning and neuromorphic computing. He has published 450 papers, 10 chapters and 4 books. He has been a Book Series Editor for IEE (now IET) Press, a journal Editor of 8 IEEE and international journals, and the sole or lead Guest Editor of 15 Special Issues or Topics of IEEE, IEE/IET and international journals.

Haeyoung Lee

Haeyoung Lee is a Senior Lecturer of Wireless Communication at the School of Physics, Engineering and Computer Science, University of Hertfordshire. Prior to joining the University of Hertfordshire, she held the position of Senior Research Fellow at the 5G Innovation Centre at the University of Surrey. She has contributed to several European research projects, including Dedicat6G, 5G-Heart, Clear5G, Speed-5G, and OneFIT. Before entering academia, she worked on CDMA mobile platform development at Samsung Electronics. She also served as a Research Officer at the national Radio Research Agency of the Ministry of Information and Communication (MIC), South Korea. She has organized multiple IEEE workshops on 5G/6G technologies and IEEE Conferences. She currently serves as an Associate Editor for the *IEEE Open Journal of Vehicular Technology*. Her research interests encompass 6G NTN/TN, random access for massive IoT and intelligent Vehicular networks.

Oluyomi Simpson

Oluyomi Simpson is an Associate Professor of Communications and Electronics, Head of Electrical and Electronic Engineering, and a member of the Wireless Communications Research Lab in the School of Physics, Engineering and Computer Science at the University of Hertfordshire, UK. His research and technical expertise span algorithms, optimization, hardware design, and the implementation of communication systems, including wireless communications, RF systems, and digital systems. He has published one book, over 65 research papers, and holds one patent. Dr Simpson is also an Associate Editor for *IEEE Communications Letters* and has served as a Guest Editor for five Special Issues.

Preface

The communication systems and networks landscape is rapidly transforming due to the widespread use of mobile devices and growing data transmission demands. The ubiquity of mobile devices, coupled with the popularity of applications, has led to unprecedented data traffic levels, presenting substantial challenges in managing infrastructure complexity and optimizing user experiences.

In addressing these challenges, machine learning has emerged as a promising solution. The incorporation of machine learning, driven by powerful computing platforms, is posited to introduce innovative problem-solving approaches in dynamic and heterogeneous communication environments. This integration envisions making significant contributions to intelligent system management and optimization through predictive capabilities and data-driven decision-making.

This Topic endeavours to explore the intersection of machine learning and communication research, presenting a collection of state-of-the-art contributions that underscore the potential of machine learning as a catalyst for adaptive and intelligent communication. The manuscripts presented in this Topic have undergone a rigorous peer-review process and were selected for publication in the Topic “Machine Learning in Communication Systems and Networks, 2nd Edition” by MDPI journals, including *Applied Sciences*, *Sensors*, *Photonics*, *Journal of Sensor and Actuator Networks*, and *Telecom*. Comprising fifteen articles, including an editorial and fourteen research papers, this Topic offers insights into current challenges and innovative solution approaches involving machine learning adaptation in communications and networks.

Yichuang Sun, Haeyoung Lee, and Oluyomi Simpson

Topic Editors

Machine Learning for Intelligent and Adaptive Communication Systems: From Optimization to Emerging Paradigms

Haeyoung Lee *, Yichuang Sun and Oluyomi Simpson

School of Physics, Engineering and Computer Science, University of Hertfordshire, Hatfield AL10 9AB, UK;
y.sun@herts.ac.uk (Y.S.); o.simpson@herts.ac.uk (O.S.)

* Correspondence: h.lee@herts.ac.uk

1. Introduction

Machine learning (ML) has been increasingly considered for various communication applications, demonstrating promising feasibility and effectiveness in enhancing system intelligence, adaptability, and operational efficiency [1]. By enabling data-driven modeling and decision-making, ML has been successfully applied to a wide range of communication tasks, including signal processing, resource allocation, and network management [2,3]. These developments have not only improved overall system performance but also strengthened the role of ML as a key enabler in the evolution of future communication systems, moving beyond conventional optimization toward fully autonomous and cognitive network operation [4].

At the same time, communication environments are undergoing rapid transformation, driven by emerging applications and increasingly complex system requirements. Communication systems are expanding across diverse domains, including satellite networks, unmanned aerial vehicle (UAV) systems, vehicular communications, and large-scale Internet of Things (IoT) deployments [5–7]. These environments are inherently dynamic, heterogeneous, and data-intensive, making traditional model-based approaches increasingly insufficient. As a result, there is a growing need for intelligent mechanisms capable of adapting to rapidly changing conditions [1]. In this context, ML is expected to take on a more active role in enabling the real-time, adaptive, and context-aware optimization of communication systems.

Despite its strong potential, the application of ML in communication systems also introduces several critical challenges [8,9]. These include the limited availability of high-quality data, concerns related to data privacy and security, and the need for efficient learning across distributed and resource-constrained environments [10,11]. Addressing these challenges has led to the emergence of new learning paradigms, such as federated learning and edge learning, which enable distributed and privacy-preserving model training. In addition, large-scale AI models and privacy-preserving techniques are being explored to enhance learning capability while maintaining scalability and data confidentiality, further supporting the deployment of ML in complex communication scenarios.

Beyond algorithmic advancements, ML is also driving the development of novel communication and computing architectures. In particular, emerging hardware-oriented solutions, such as optical neural networks and optical convolution accelerators, are being investigated to overcome the limitations of conventional electronic processing [12,13]. These architectures leverage the properties of light to perform computation, offering significant improvements in processing speed and energy efficiency. Such innovations highlight the

growing convergence of communication, computation, and intelligence, and point toward new system designs that are inherently optimized for ML-driven operations.

Given these developments, the interest in integrating ML into communication systems is expected to continue growing. In this context, and following the progress established in the first edition of this Topic [14], this second edition further extends the exploration of ML toward more complex, large-scale, and heterogeneous communication environments. Compared to the first edition, which demonstrated the feasibility and broad applicability of ML across core communication tasks, this edition places greater emphasis on advanced learning paradigms, scalability, and integration into next-generation communication and computing architectures, reflecting a natural progression from foundational studies toward more practical and intelligent system designs.

Accordingly, this Editorial presents fourteen contributions that collectively explore the role of ML across a diverse range of applications. The featured works address key aspects including intelligent network optimization, data-driven sensing and monitoring, advanced learning mechanisms, and emerging communication and computing architectures. Together, these contributions provide valuable insights into research directions and highlight the transformative impact of ML on next-generation communication systems.

2. An Overview of Published Articles

ML has emerged as a powerful enabler for transforming conventional communication systems into adaptive and intelligent infrastructures. In this context, the following contributions demonstrate how ML can be leveraged to address a broad range of communication challenges, such as signal prediction, dynamic resource allocation, interference mitigation, encompassing anti-jamming strategies and path planning. Apavatjirut (contribution 2) developed a Received Signal Strength Indicator (RSSI) prediction framework by leveraging adaptive supervised learning models that incorporate environmental sensing data to capture non-linear signal variations. The effectiveness of the proposed approach is validated in a LoRa (Long-Range) network environment using real-world sensor measurements. Wang et al. (contribution 3) address anti-jamming communication by integrating generative adversarial networks (GANs) with deep Q-networks (DQNs), enabling data augmentation and improved reinforcement learning convergence. Mawlood et al. (contribution 5) propose a dynamic bandwidth allocation scheme using deep reinforcement learning, specifically a Soft Actor–Critic (SAC) model enhanced with continual learning to adapt to changing traffic conditions, with the aim of improving fairness and overall efficiency in network resource allocation. Suarez del Valle et al. (contribution 6) employ a contextual multi-armed bandit approach to optimize beam selection in mmWave vehicular networks, enabling online learning under dynamic environments. AlMania et al. (contribution 11) designed an energy-efficient UAV path planning framework by combining reinforcement learning with particle swarm optimization (PSO), allowing adaptive navigation in complex environments. Zheng et al. (contribution 13) enhanced modulation recognition in visible light communication (VLC) systems using a deep learning architecture that integrates Convolutional Neural Networks (CNNs), attention mechanisms, and Long Short-Term Memory (LSTM) networks. This integrated design aims to jointly capture spatial features, temporal dependencies, and salient signal characteristics, thereby improving robustness and recognition accuracy under noise, distortion, and dynamic channel conditions. Finally, da Silva et al. (contribution 14) provide a survey of ML-based approaches for PAPR (Peak to Average Power Ratio) reduction in OFDM (Orthogonal Frequency Division Multiplexing) systems.

Building upon these optimization-focused studies, machine learning is further applied to enable intelligent monitoring, detection, and sensing capabilities across communication-related systems. In this direction, four contributions illustrate how data-driven approaches

can replace traditional, labor-intensive, or less accurate techniques with more efficient and scalable solutions. Szcerba et al. (contribution 1) applied supervised learning to detect faults in hybrid fiber-coaxial networks, improving service reliability through proactive maintenance. López-Muñoz et al. (contribution 9) present a hybrid system combining RF signal analysis and computer vision for accurate UAV (unmanned aerial vehicle) detection, localization, and tracking, effectively improving detection accuracy and reducing false positives in drone monitoring for security applications. In the context of structural health monitoring, Shah Mansouri et al. (contribution 10) utilized machine learning and statistical techniques to detect and localize cracks in materials using sensor networks. Meanwhile, Ma et al. (contribution 12) address automatic modulation recognition, a critical sensing task in wireless communications, by employing meta-learning to overcome data scarcity challenges. Together, these studies illustrate the growing importance of machine learning in enabling intelligent perception and diagnostics within modern communication ecosystems.

Beyond these application-driven studies, recent efforts have explored advanced machine learning paradigms to address fundamental limitations in communication systems, particularly in terms of data availability and distributed operation. Tang et al. (contribution 4) introduce an over-the-air federated learning framework that enables efficient and privacy-preserving distributed model training while mitigating interference in multi-cell environments. Complementing this, Ma et al. (contribution 12) propose a meta-learning-based approach for automatic modulation recognition, allowing the model to learn how to learn and rapidly adapt to new modulation types with limited training data. These contributions demonstrate how emerging learning paradigms such as federated learning and meta-learning provide scalable, distributed, and data-efficient solutions, paving the way for more intelligent and collaborative communication systems.

Looking ahead, machine learning is also driving the development of novel communication and computing architectures, pointing toward the future of intelligent systems. The study by Tang et al. (contribution 4) represents a shift toward integrated communication learning frameworks through over-the-air federated learning, while AlMania et al. (contribution 11) extend this vision to distributed and autonomous Internet of Drones systems. At the hardware level, Chen et al. (contribution 7) and Xia et al. (contribution 8) introduce optical neural networks and optical convolution accelerators, respectively, offering new computing architectures that significantly improve speed and energy efficiency compared to traditional electronic processing. Chen et al. performed inference directly using light in passive optical systems, while Xia et al. leveraged frequency-domain encoding to enable highly parallel and efficient convolutional operations. Together, these works highlight the potential of optical computing to overcome the performance and energy limitations of conventional ML hardware. Collectively, these advances demonstrate how machine learning is not only enhancing existing systems but also enabling fundamentally new architectures that redefine the boundaries of communication and computation.

3. Conclusions

In summary, the contributions presented in this collection demonstrate the transformative impact of machine learning across multiple layers of communication systems, from optimization and control to sensing, learning, and architectural design. The reviewed works highlight how ML enables more adaptive, efficient, and intelligent solutions to longstanding challenges in wireless communications, network management, and system monitoring. At the same time, emerging paradigms such as federated learning and optical computing signal a shift toward more distributed, scalable, and energy-efficient architectures. Despite these advances, several open challenges remain, including model generalization, real-time deployment constraints, and the integration of heterogeneous systems. Future research is

expected to further explore the synergy between machine learning and communication technologies, ultimately paving the way for fully autonomous, intelligent, and resilient communication ecosystems.

Author Contributions: Conceptualization, H.L.; writing—original draft preparation, H.L.; writing—review, Y.S. and O.S. and editing, H.L. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

List of Contributions:

1. Szczerba, C.; Dávalos, E.; Leiva, A.; Pinto-Ríos, J. Early Fault Detection in a Real Scenario of Hybrid Fiber–Coaxial Networks Using Machine Learning: An Approach Based on Decision Trees and Random Forests. *Appl. Sci.* **2025**, *15*, 10442. <https://doi.org/10.3390/app151910442>.
2. Apavatjirut, A. Sensor-Driven RSSI Prediction via Adaptive Machine Learning and Environmental Sensing. *Sensors* **2025**, *25*, 5199. <https://doi.org/10.3390/s25165199>.
3. Wang, T.; Niu, Y.; Zhou, Z. Few-Shot Intelligent Anti-Jamming Access with Fast Convergence: A GAN-Enhanced Deep Reinforcement Learning Approach. *Appl. Sci.* **2025**, *15*, 8654. <https://doi.org/10.3390/app15158654>.
4. Tang, C.; He, D.; Yao, J. Distributed Interference-Aware Power Optimization for Multi-Task Over-the-Air Federated Learning. *Telecom* **2025**, *6*, 51. <https://doi.org/10.3390/telecom6030051>.
5. Mawlood, M.A.; Mahmood, D.A. Optimizing Weighted Fair Queuing with Deep Reinforcement Learning for Dynamic Bandwidth Allocation. *Telecom* **2025**, *6*, 46. <https://doi.org/10.3390/telecom6030046>.
6. Suarez del Valle, R.; Kose, A.; Lee, H. Context-Aware Beam Selection for IRS-Assisted mmWave V2I Communications. *Sensors* **2025**, *25*, 3924. <https://doi.org/10.3390/s25133924>.
7. Chen, R.; Ma, Y.; Wang, Z.; Sun, S. Incoherent Optical Neural Networks for Passive and Delay-Free Inference in Natural Light. *Photonics* **2025**, *12*, 278. <https://doi.org/10.3390/photonics12030278>.
8. Xia, G.; Liu, J.; Hong, Q.; Zhu, P.; Xu, P.; Zhu, Z. An Efficient Frequency Encoding Scheme for Optical Convolution Accelerator. *Photonics* **2025**, *12*, 26. <https://doi.org/10.3390/photonics12010026>.
9. López-Muñoz, P.; San Frutos, L.G.; Abarca, C.; Alegre, F.J.; Calle, J.L.; Monserrat, J.F. Hybrid Artificial-Intelligence-Based System for Unmanned Aerial Vehicle Detection, Localization, and Tracking Using Software-Defined Radio and Computer Vision Techniques. *Telecom* **2024**, *5*, 1286–1308. <https://doi.org/10.3390/telecom5040064>.
10. Shah Mansouri, T.; Lubarsky, G.; Finlay, D.; McLaughlin, J. Machine Learning-Based Structural Health Monitoring Technique for Crack Detection and Localisation Using Bluetooth Strain Gauge Sensor Network. *J. Sens. Actuator Netw.* **2024**, *13*, 79. <https://doi.org/10.3390/jsan13060079>.
11. AlMania, Z.; Sheltami, T.; Ahmed, G.; Mahmoud, A.; Barnawi, A. Energy-Efficient Online Path Planning for Internet of Drones Using Reinforcement Learning. *J. Sens. Actuator Netw.* **2024**, *13*, 50. <https://doi.org/10.3390/jsan13050050>.
12. Ma, Z.; Fang, S.; Fan, Y.; Hou, S.; Xu, Z. Tackling Few-Shot Challenges in Automatic Modulation Recognition: A Multi-Level Comparative Relation Network Combining Class Reconstruction Strategy. *Sensors* **2024**, *24*, 4421. <https://doi.org/10.3390/s24134421>.
13. Zheng, X.; He, Y.; Zhang, C.; Miao, P. VLCMnet-Based Modulation Format Recognition for Indoor Visible Light Communication Systems. *Photonics* **2024**, *11*, 403. <https://doi.org/10.3390/photonics11050403>.
14. da Silva, B.S.d.C.; Souto, V.D.P.; Souza, R.D.; Mendes, L.L. A Survey of PAPR Techniques Based on Machine Learning. *Sensors* **2024**, *24*, 1918. <https://doi.org/10.3390/s24061918>.

References

1. Tera, P.S.; Chinthajinjala, R.; Pau, G.; Kim, H.K. Toward 6G: An Overview of the Next Generation of Intelligent Network Connectivity. *IEEE Access* **2025**, *13*, 925–961. [CrossRef]
2. Morocho-Cayamcela, M.E.; Lee, H.; Lim, W. Machine Learning for 5G/B5G Mobile and Wireless Communications: Potential, Limitations, and Future Directions. *IEEE Access* **2019**, *7*, 137184–137206. [CrossRef]

3. Jiang, F.; Pan, C.; Dong, L.; Wang, K.; Debbah, M.; Niyato, D.; Han, Z. A Comprehensive Survey of Large AI Models for Future Communications: Foundations, Applications, and Challenges. *IEEE Commun. Surv. Tutor.* **2026**, *28*, 4731–4764. [CrossRef]
4. Pivoto, D.G.S.; Figueiredo, F.A.P.d.; Cavdar, C.; Tejerina, G.R.d.L.; Mendes, L.L. A Comprehensive Survey of Machine Learning Applied to Resource Allocation in Wireless Communications. *IEEE Commun. Surv. Tutor.* **2026**, *28*, 1986–2053. [CrossRef]
5. Mahboob, S.; Liu, L. Revolutionizing Future Connectivity: A Contemporary Survey on AI-Empowered Satellite-Based Non-Terrestrial Networks in 6G. *IEEE Commun. Surv. Tutor.* **2024**, *26*, 1279–1321. [CrossRef]
6. Noor-A-Rahim, M.; Liu, Z.; Lee, H.; Khyam, M.O.; He, J.; Pesch, D.; Moessner, K.; Saad, W.; Poor, H.V. 6G for Vehicle-to-Everything (V2X) Communications: Enabling Technologies, Challenges, and Opportunities. *Proc. IEEE* **2022**, *110*, 712–734. [CrossRef]
7. Latreche, S.; Bellahsene, H. A comprehensive survey on 6G: Enabling technologies, key applications, and future challenges. *Frankl. Open* **2026**, *15*, 100559. [CrossRef]
8. Akbar, M.S.; Hussain, Z.; Ikram, M.; Sheng, Q.Z.; Mukhopadhyay, S.C. On challenges of sixth-generation (6G) wireless networks: A comprehensive survey of requirements, applications, and security issues. *J. Netw. Comp. Appl.* **2025**, *233*, 104040. [CrossRef]
9. Kato, N.; Mao, B.; Tang, F.; Kawamoto, Y.; Liu, J. Ten Challenges in Advancing Machine Learning Technologies toward 6G. *IEEE Wirel. Commun.* **2020**, *27*, 96–103. [CrossRef]
10. Fouda, M.M.; Fadlullah, Z.M.; Ibrahim, M.I.; Kato, N. Privacy-Preserving Data-Driven Learning Models for Emerging Communication Networks: A Comprehensive Survey. *IEEE Commun. Surv. Tutor.* **2025**, *27*, 2505–2542. [CrossRef]
11. Liu, X.; Deng, Y.; Nallanathan, A.; Bennis, M. Federated Learning and Meta Learning: Approaches, Applications, and Directions. *IEEE Commun. Surv. Tutor.* **2024**, *26*, 571–618. [CrossRef]
12. Fu, T.; Zhang, J.; Sun, R.; Huang, Y.; Xu, W.; Yang, S.; Zhu, Z.; Chen, H. Optical neural networks: Progress and challenges. *Light Sci. Appl.* **2024**, *13*, 263. [CrossRef] [PubMed]
13. Sunny, F.; Shafiee, A.; Balasubramaniam, A.; Nikdast, M.; Pasricha, S. OPIMA: Optical Processing-in-Memory for Convolutional Neural Network Acceleration. *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.* **2024**, *43*, 3888–3899. [CrossRef]
14. Sun, Y.; Lee, H.; Simpson, O. Machine Learning in Communication Systems and Networks. *Sensors* **2024**, *24*, 1925. [CrossRef] [PubMed]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Early Fault Detection in a Real Scenario of Hybrid Fiber–Coaxial Networks Using Machine Learning: An Approach Based on Decision Trees and Random Forests

Christian Szczerba ¹, Enrique Dávalos ^{1,*}, Ariel Leiva ² and Juan Pinto-Ríos ²

¹ Polytechnic Faculty, Universidad Nacional de Asunción, Campus Universitario UNA, San Lorenzo 111421, Paraguay; cszczerba@fpuna.edu.py

² School of Electrical Engineering, Pontificia Universidad Católica de Valparaíso, Av. Brasil 2950, Valparaíso 2340000, Chile; ariel.leiva@pucv.cl (A.L.); juan.pinto.r@pucv.cl (J.P.-R.)

* Correspondence: edavalos@pol.una.py

Abstract: Cable service providers face significant challenges in managing Hybrid Fiber–Coaxial (HFC) networks due to the growing demand for high-speed services. Ensuring high service availability is critical to preventing customer attrition. This study employs machine learning techniques, specifically Decision Tree and Random Forest models, for proactive fault detection in HFC networks using data from the Simple Network Management Protocol (SNMP). Two operational scenarios were considered: a network-wide model and node-specific models. The dataset for fault detection exhibited a severe class imbalance, with outage events being extremely rare. To address this, the Synthetic Minority Oversampling Technique (SMOTE), which generates synthetic samples of the minority class to balance the dataset, was applied. This significantly improved recall and F1-scores—the harmonic mean of precision and recall—while maintaining high precision. The results demonstrate that these machine learning algorithms achieve up to 98% accuracy, and the SMOTE-enhanced models provide more reliable detection of connectivity faults. This approach is highly effective for cable operators in maintaining quality of service, enabling proactive management of problems and enhancement of network performance.

Keywords: Data Over Cable Service Interface Specification; Hybrid Fiber–Coaxial; decision tree; random forest; machine learning

1. Introduction

Despite the growing adoption of fiber optic services globally [1], many Internet Service Providers (ISPs) still rely on DOCSIS (Data Over Cable Service Interface Specification) standards [2] over Hybrid Fiber–Coaxial (HFC) networks for high-quality broadcasts and Internet services [3]. Maintaining high service reliability is essential to mitigating customer attrition, and the implementation of proactive fault-prevention strategies can play a pivotal role in enhancing overall user experience (UX).

Aging infrastructures, cable corrosion, and network heterogeneity further complicate operations [4], and HFC networks require continuous upgrades to maintain service quality [5]. Upstream link problems caused by interference can degrade performance, often requiring labor-intensive manual interventions that disrupt services [6]. In this context, machine learning (ML) offers a promising approach by detecting early signs of faults from large volumes of network data. Supervised learning algorithms such as Decision Trees and Random Forests are particularly effective for this task [7].

This study examines two complementary scenarios for early fault detection in HFC networks using machine learning. *Scenario 1* adopts a network-wide perspective, aggregating data across the entire network to identify global patterns, although with inherent limitations in isolating specific faults. In this scenario, due to the nature of the data acquisition, the target variable representing service outages is highly imbalanced. To address this issue, the Synthetic Minority Oversampling Technique (SMOTE) [8] was applied to generate synthetic samples of the minority class, thereby improving the robustness of the classification models. *Scenario 2* conducts analysis at the node level, enabling more granular and sensitive detection of individual faults. This dual approach provides a comprehensive framework for evaluating the trade-offs and benefits associated with different levels of monitoring granularity in HFC networks. The HFC network analyzed has been built progressively over more than three decades, resulting in a heterogeneous infrastructure with components from different technological generations and natural wear and tear. Furthermore, optimal signal levels vary by geographic area, influenced by factors such as topography, subscriber density, and environmental conditions.

This article investigates the application of machine learning models for early fault detection in a real Hybrid Fiber–Coaxial (HFC) network operated by a telecommunications provider. The study leverages data collected through the Simple Network Management Protocol (SNMP) monitoring system. Using this real-world dataset, two fault detection models were developed: (i) a general model covering the entire network and (ii) a node-specific model targeting localized service issues. Both models, implemented using Decision Tree and Random Forest algorithms, are designed to improve the accuracy and efficiency of fault detection. The study further compares models trained on the original dataset and on the SMOTE-balanced dataset to evaluate the impact of resampling on outage detection performance.

The methodology comprises dataset creation, data preprocessing, model training, and performance evaluation. Section 2 presents a review of previous work on signal-level monitoring tools and the application of machine learning (ML) in HFC networks. Section 3 defines the research problem, while Section 4 details the conceptual framework, including network subdivision, DOCSIS standards, and ML algorithms. Section 5 outlines the methodology in depth, including business understanding, data pre-processing, and model performance assessment.

2. Bibliographical Analysis

Proactive fault detection is essential for ensuring the performance and reliability of HFC networks, and ML is increasingly being explored as a viable solution. Several studies have identified network monitoring variables capable of predicting faults. For example, in ref. [9] highlights the potential of ML in automating cable modem data collection and analysis. In ref. [10], Decision Trees are applied to detect radio frequency behaviors in Long-Term Evolution (LTE) networks, offering insights relevant for HFC data analysis. Similarly, in ref. [11] employs DOCSIS-based tools to visualize key network parameters, aiding in fault detection.

Research specifically targeting HFC networks has demonstrated promising results. For instance, in ref. [12] uses Full-Band Capture (FBC) data to cluster cable modems based on signal patterns, enabling proactive repairs of weak network segments. Unsupervised techniques like k-means and Gaussian mixture models have also been employed for anomaly detection [13], with principal component analysis improving accuracy. Other works, such as [14], propose big data platforms for fault detection, although adjustments are needed for non-tree topologies. Active monitoring approaches using pattern recognition have also

proven effective in anomaly detection, with both FBC [12] and big data frameworks [14] serving as key enablers.

This review underscores the need for research tailored specifically to HFC networks. Our approach addresses this gap by applying ML to early fault detection using signal-level data, thereby advancing both fault detection and localization techniques for HFC network environments.

3. Problem Description

Cable service providers face significant challenges in managing Hybrid Fiber–Coaxial (HFC) networks, particularly due to the technical heterogeneity arising from network mergers. Older coaxial cables are especially susceptible to corrosion, which can lead to a degradation in connection quality [4].

The deployed network architecture employs distinct channels for upstream and downstream signal transmission. Upstream channels, which carry data from customer premises to network nodes, are particularly vulnerable to a phenomenon known as noise channeling, in which interference originating from a single household can propagate and disrupt multiple connections [6], thereby compromising uplink reliability. In operational practice, field technicians often use a binary search approach to locate malfunctioning devices, a process that involves disconnecting amplifiers and temporarily interrupting service during troubleshooting.

Proactive Network Maintenance (PNM) tools can assist in diagnosing network problems; however, they frequently produce excessive alarms, potentially overwhelming technicians [5]. Consequently, predicting failures prior to field intervention is critical to preventing service disruptions, underscoring the importance of predictive network management.

4. Conceptual Framework

This section outlines the conceptual framework adopted in this study, including the subdivision of the HFC network, the DOCSIS standard, machine learning, Decision Trees, and Random Forest.

4.1. HFC Network

Cable operators initially introduced high-speed data services over existing television infrastructures, with the bandwidth demand increasing significantly due to the rise of symmetric data applications, thereby enabling bidirectional IP traffic over HFC networks [2].

An HFC network is typically divided into two main sections: (i) the optical segment, which connects the central office to nodes via fiber optics, and (ii) the coaxial segment, which employs line amplifiers and splitters to deliver service to end users. Signals originate at the master headend and pass through the Cable Modem Termination System (CMTS) before reaching optical nodes, where they are converted to Radio Frequency (RF) for coaxial transmission [12]. This architecture facilitates fault detection, particularly within the segment between the CMTS and the optical nodes.

As illustrated in Figure 1, fiber nodes connect the network to customer cable modems, with amplifiers extending coverage to the final mile. Faults occurring in shared upstream channels can impact all devices connected to a given node.

4.2. DOCSIS

The development of HFC networks enabled broadband Internet access through DOCSIS, a set of standards ensuring high-speed data transmission over cable networks. Developed by CableLabs, DOCSIS supports a variety of services, including Internet access,

high-definition television (HDTV), Voice over IP (VoIP), the Internet of Things (IoT), and Virtual Reality (VR) [2].

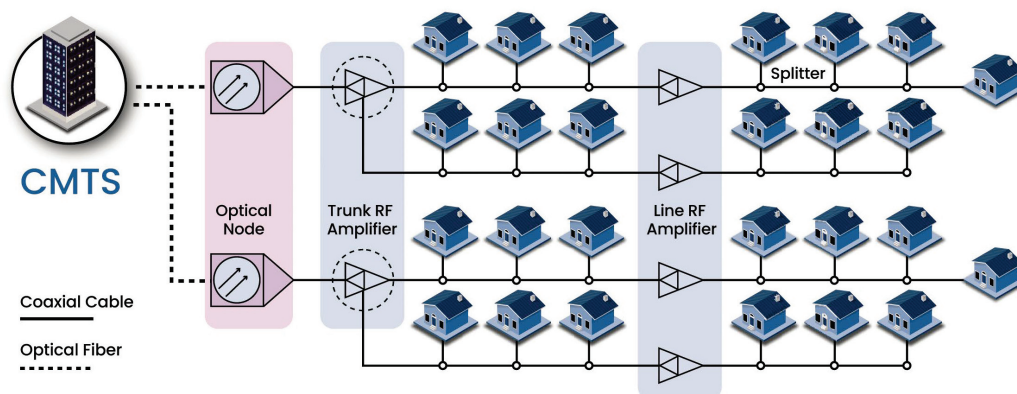


Figure 1. Overview of the Hybrid Fiber-Coaxial (HFC) network architecture, showing the hierarchical structure and key data collection points used for SNMP-based fault detection [2].

DOCSIS has evolved to address increasing bandwidth demands, with DOCSIS 3.1 introducing Orthogonal Frequency-Division Multiplexing (OFDM) for downstream and Orthogonal Frequency-Division Multiple Access (OFDMA) for upstream transmissions [5]. However, upstream links remain susceptible to noise and distortion [10], which can disrupt service for all modems within a fiber-node area [6]. The DOCSIS Full Band Capture (FBC) feature [12] enables remote spectrum analysis via the Simple Network Management Protocol (SNMP), offering cost advantages over traditional field analyzers. SNMP monitoring captures signal-level data from customer modems, including transmission and reception (TX/RX) power, which must remain within DOCSIS-defined thresholds to avoid communication issues. Additional performance metrics—such as error counts, corrected and uncorrected packets, Signal-to-Noise Ratios (SNRs), and system uptime—are also recorded to support automated analysis and troubleshooting.

4.3. Machine Learning

Machine learning (ML), a branch of Artificial Intelligence (AI) [15], is a powerful approach for analyzing large datasets and generating predictions for various variables [10]. It has been widely applied in domains such as medicine, technology, and telecommunications [16]. ML applications can generally be categorized into regression, classification, and decision-making tasks [17], with classification being the primary focus in this study for fault detection. ML has demonstrated notable success in industrial processes [18] and optical networks [19]. The main learning paradigms include supervised learning, which uses labeled data for prediction [20]; unsupervised learning; and reinforcement learning [21].

This study applies supervised learning techniques, comparing Decision Tree and Random Forest models for early fault detection by analyzing network variables.

4.4. Decision Trees

Achieving real-time fault detection requires the adoption of ML techniques capable of delivering rapid and accurate classification responses. Decision Trees (DTs) have been proposed for this purpose. DTs classify instances by evaluating feature values, where each internal node represents a feature and the branches denote possible outcomes [20]. A Decision Tree consists of a root node, intermediate decision nodes, and leaf nodes, which correspond to predicted class labels. This structure facilitates classification and prediction based on input variables and features [20].

4.5. Random Forest

Using bagging techniques, Random Forests generate multiple Decision Trees by randomly selecting subsets of the data during node splitting. This ensemble approach enhances accuracy by aggregating predictions from multiple trees. However, determining the optimal number of trees often requires empirical tuning, as increasing the number does not always yield significant performance gains [22]. Nonetheless, this characteristic allows the model to be readily adjusted to optimize performance.

5. Methodology for Early Fault Detection in HFC Networks Using ML

To develop a machine learning-based fault detection system, the CRoss Industry Standard Process for Data Mining (CRISP-DM) framework [23] was adopted. This methodology provides a structured yet flexible approach to data mining. Figure 2 summarizes the adaptation of CRISP-DM to the design of the proposed ML-based fault detector. The process comprises the following stages: (I) business understanding, (II) data understanding, (III) data preparation, (IV) modeling, and (V) evaluation. The implementation stage is not addressed in this study, as deployment conditions are beyond its scope.

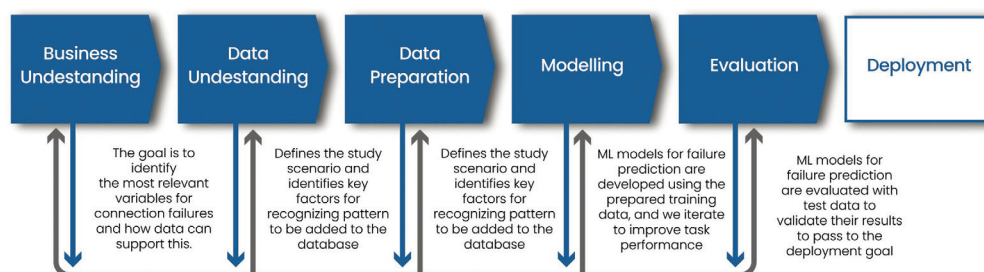


Figure 2. Methodology schematic outlining the five stages of the CRISP-DM framework applied: business understanding, data understanding, data preparation, modeling, and evaluation. Deployment is not covered in this study.

In constructing the dataset for model training, the CRISP-DM methodology was applied as follows:

5.1. Business Understanding

In this phase, key operational factors in the DOCSIS network are identified from the perspective of the telecommunications operator. The objective is to determine the most relevant variables associated with connection failures and to establish how available data can support their detection.

Each cable modem is assigned a service IP address via DHCP for internet access and a management IP for remote administration through the Simple Network Management Protocol (SNMP), which enables signal monitoring and configuration changes. Leveraging the DOCSIS Management Information Base (MIB) structure [24], data such as MAC addresses, RX/TX power levels, Signal-to-Noise Ratios (SNRs), and system uptime are collected periodically.

5.2. Data Understanding

This stage defines the study scenario and determines key factors for identifying patterns in network failures, based on the variables established in the previous phase. The goal is to collect and explore relevant data, retaining patterns of interest while discarding irrelevant information.

To detect outages, logs from modems associated with affected nodes were captured. Fault flags were incorporated into the SNMP dataset, enabling identification of modems and signal readings corresponding to periods of failure.

Figure 3 presents an example of signal levels recorded for a cable modem, highlighting an outage period beginning at the timestamp 14:29. The horizontal axis represents the measurement timestamps, while the left vertical axis indicates the number of corrected packets and the right vertical axis shows the corresponding signal levels.

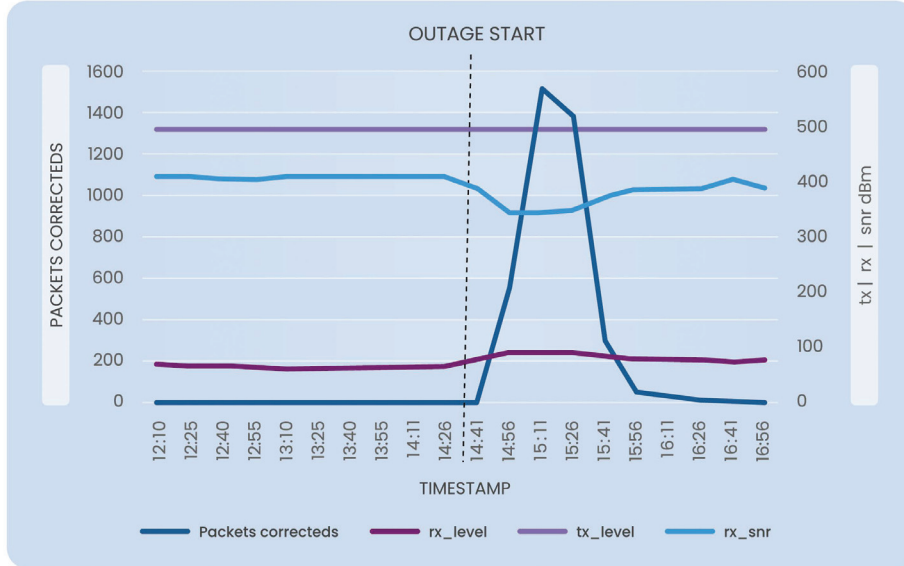


Figure 3. Example of signal-level measurements showing their behavior before and during a node outage event, highlighting key SNMP metrics relevant for fault detection.

In Scenario 1, the dataset exhibits a significant imbalance between the majority class (normal operation) and the minority class (outage). To mitigate this issue, the Synthetic Minority Oversampling Technique (SMOTE) was employed to generate synthetic samples of the minority class, thereby improving the capacity of machine learning models to learn patterns related to rare outage events [8].

Given the large volume of data collected by monitoring systems, not all records can be directly associated with the outage event. For the purposes of this study, only the variables exhibiting the strongest correlation with the failure were considered as key factors, as summarized in Figure 4. A correlation analysis was performed to validate the selection of these features.

5.3. Data Preparation

This stage involves preprocessing the dataset by extracting key features, handling missing values, and removing irrelevant attributes. Techniques such as normalization, scaling, and encoding are applied to ensure data consistency. The dataset is subsequently divided into training and testing subsets, following a common split ratio of 70% for training and 30% for evaluation.

Given the severe class imbalance in the dataset, with outage events being extremely rare, SMOTE was considered to generate synthetic samples of the minority class. This approach enhances the ability of the models to detect rare failures, improving recall and F1-scores while maintaining precision [25]. Preprocessing also includes filtering relevant variables, eliminating erroneous data, and adding features to enhance learning [10]. This stage is critical for the development of machine learning models, as the quality of input data directly influences prediction accuracy. Raw datasets often contain noise, outliers, and

anomalies that can degrade performance. To address this, a flag column is introduced to identify records within the outage threshold, and invalid records are removed before the final split into training and evaluation sets.

PARAMETER	DESCRIPTION
Outage	Service degradation status within the node
Status	Online / Offline status of the cable modem
Rx_level	Reception level, the power in decibels per milliwatts (dBm) of the received signal
Tx_level	Transmission level, power in dBm at which a cable modem transmits its signal
Rx_snr	Cable modem signal-to-noise ratio
Ping_ms	Ping measured in milliseconds
Noerrors	Number of packets transmitted without error
Correcteds	Number of packets with errors corrected

Figure 4. Key HFC network factors for the database compiled during the business understanding phase.

5.4. Modeling

In this stage, machine learning (ML) models for failure prediction were developed using the prepared dataset. Classification techniques—encompassing algorithm selection, hyperparameter tuning, and cross-validation—were applied to mitigate overfitting and balance model complexity for improved generalization. Model performance was optimized using evaluation metrics such as accuracy, recall, and F1-score [16].

Decision Tree and Random Forest models were trained using binary labels (“non-outage” and “outage”). Two data collection scenarios were considered: (i) Scenario 1, which represents normal network operation using aggregated data from all nodes, and (ii) Scenario 2, which focuses on network failures by including only data from the affected node. Four models were thus trained: two Decision Tree models and two Random Forest models, one for each scenario.

For Scenario 1, each classifier was trained and evaluated twice: once using the original imbalanced dataset, and once using the SMOTE-balanced dataset. This comparison allowed the identification of the best-performing model in terms of its ability to detect outage events despite the imbalance. Scenario 2, by contrast, did not require resampling given its higher outage density.

The effectiveness of these models is presented in the following section.

5.5. Evaluation

In the final stage, the proposed ML models for failure prediction were evaluated against the reserved test data to validate their performance. This step ensures that the models meet the business objective of accurately predicting failures within the HFC network.

The test sets set aside during preprocessing were used to assess performance, and confusion matrices were generated for both Scenario 1 and Scenario 2. In Figures 5 and 6, the left panel shows the confusion matrix for the Decision Tree model, while the right panel presents the results for the Random Forest model. Figure 7 displays the confusion matrices for both models when classifying events as “outage” or “no outage.” Each matrix is color-coded to represent true positives, false positives, false negatives, and true negatives.

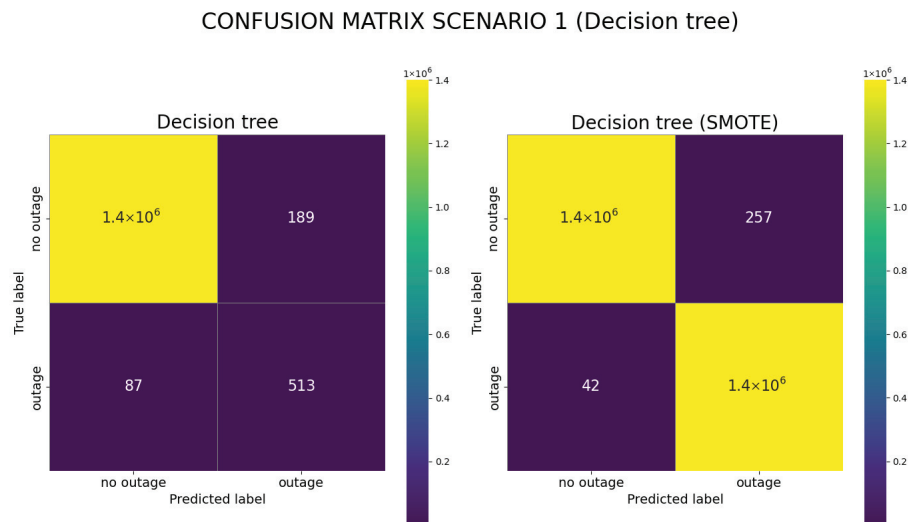


Figure 5. Confusion matrices for Scenario 1 for proposed Decision Tree models.

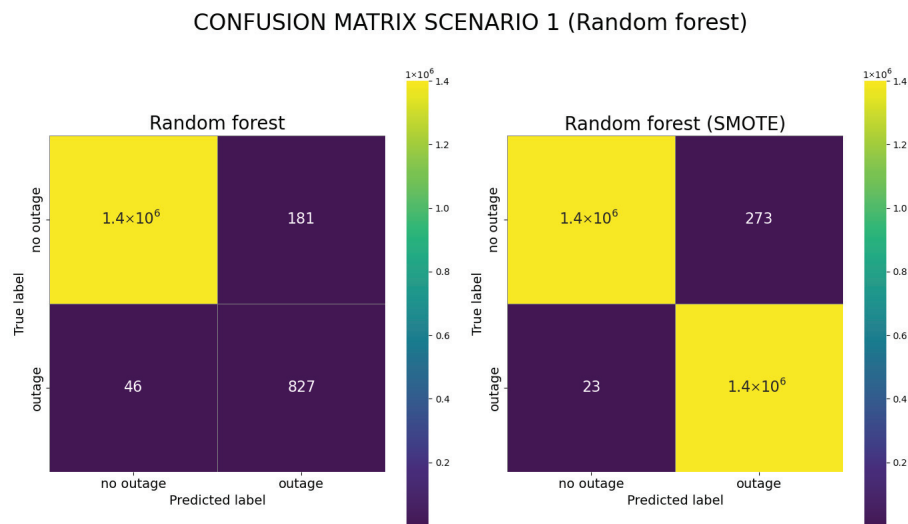


Figure 6. Confusion matrices for Scenario 1 for proposed Random Forest models.

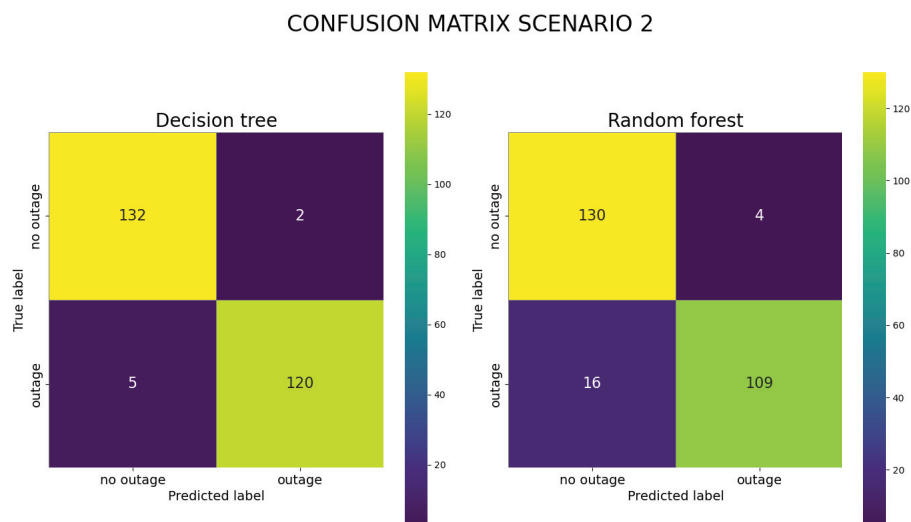


Figure 7. Confusion matrices for Scenario 2 for Decision Tree and Random Forest models.

This evaluation framework enables the calculation of key performance metrics such as precision, recall, and F1-score. The F1-score, defined as the harmonic mean of precision

and recall, assigns equal weight to both metrics. It ranges from 0 to 1, where 1 represents perfect classification and 0 indicates the worst possible performance [20].

In addition to the models trained on the original dataset, SMOTE-based variants were also evaluated. These models incorporated synthetic oversampling of the minority outage class, allowing a more balanced training distribution. As a result, both Decision Tree and Random Forest classifiers trained with SMOTE achieved substantially higher recall and F1-scores compared to their original counterparts, while maintaining competitive precision.

The methodology described in the previous section was applied to a real Hybrid Fiber-Coaxial (HFC) network operating under a DOCSIS system. A total of 6,352,750 samples were collected by the monitoring system from the live network during the business understanding and data understanding phases. In Scenario 1, which incorporates signal level readings from all cable modems within a 5 h observation window, the Decision Tree model achieved high accuracy (99.98%) but low fault detection precision (19%). The Random Forest model obtained slightly higher accuracy (99.99%) but exhibited poor performance in identifying true positives, with 100% precision and very low recall (0.2%). This seemingly perfect precision is misleading, as it results from a severe class imbalance: only 0.0089% of the samples corresponded to outage events. This imbalance biased the model predictions and led to a low F1-score. By contrast, the SMOTE-enhanced Decision Tree model achieved 99.91% recall and 94.65% F1-score, while the Random Forest model obtained 99.97% recall and 94.68% F1-score, demonstrating the effectiveness of oversampling in mitigating class imbalance.

Figure 8 summarizes the performance of the evaluated models under both scenarios, using a diverse set of metrics to capture different aspects of classification quality. Accuracy reports the overall proportion of correctly classified samples, while Precision indicates the reliability of the model in identifying outages without producing false alarms. Recall measures the ability to detect actual outage events, which is critical for minimizing undetected failures. The F1-score balances precision and recall into a single indicator, making it particularly relevant in the context of imbalanced datasets [26]. The Area Under the Curve (AUC) further evaluates the discriminatory capacity of the models across thresholds. Additionally, Cohen's Kappa accounts for agreement beyond chance, and the Matthews Correlation Coefficient (MCC) provides a balanced assessment of prediction quality, even in the presence of class imbalance [27,28].

MODEL PERFORMANCE COMPARISON								
	MODEL	ACCURACY	PRECISION	RECALL	F1	AUC	KAPPA	MCC
SCENARIO 1	Decision Tree Classifier (Original)	99.90%	5.39%	4.39%	4.84%	64.19%	4.79%	4.81%
	Random Forest Classifier (Original)	99.94%	77.78%	0.25%	0.50%	77.49%	0.50%	4.26%
	Decision Tree Classifier (SMOTE)	94.35%	89.91%	99.91%	94.65%	97.71%	88.70%	89.25%
	Random Forest Classifier (SMOTE)	94.38%	89.92%	99.97%	94.68%	98.11%	88.76%	89.32%
SCENARIO 2	Decision Trees (Original)	97.29%	98%	96%	97%	97%	94%	94%
	Random Forest (Original)	92.27%	96%	87%	92%	92%	84%	84%

Figure 8. Performance metrics of Decision Tree and Random Forest classifiers on original and SMOTE-enhanced datasets, including accuracy, precision, recall, F1-score, AUC, Cohen's Kappa, and MCC to assess classification under class imbalance.

It is important to note that the low recall observed in Scenario 1 reflects the inherent challenges of modeling fault detection at a holistic network level when working with a large, highly imbalanced dataset in which failure events are extremely rare. In contrast, Scenario 2—by focusing on a specific node—enabled the models to achieve a substantially higher detection performance by leveraging localized data patterns. Moreover, the SMOTE-

enhanced models demonstrated that oversampling transforms impractical classifiers into effective tools, yielding actionable insights for early fault detection in real-world operations.

These findings emphasize a critical point: models trained on the original, highly imbalanced dataset—especially in Scenario 1—exhibited recall values so low that they rendered the approach impractical for real-world fault detection, as most failure events were systematically missed. By contrast, incorporating SMOTE transformed these same classifiers into viable solutions, markedly increasing recall and F1-scores while maintaining competitive precision. This demonstrates that oversampling techniques are not merely optional adjustments, but rather essential steps when addressing the intrinsic rarity of failure events in large-scale HFC networks.

6. Conclusions

This study investigated fault detection in Hybrid Fiber–Coaxial (HFC) networks using Decision Tree and Random Forest models, focusing on dataset preprocessing, operational scenarios, and performance evaluation. In Scenario 1, models trained on the original dataset achieved high accuracy but suffered from low recall due to a severe class imbalance, whereas Scenario 2, focused on a single node, benefited from localized patterns and yielded a higher detection performance. The application of SMOTE effectively mitigated class imbalance, substantially improving recall and F1-scores and transforming models into practical tools for real-world deployment. As future work, extending the methodology to a multiclass framework by labeling different types of failure events—such as partial degradations, intermittent outages, or complete service interruptions—would enable classifiers to provide more granular insights, supporting proactive maintenance and resource optimization. Overall, these results demonstrate that combining robust machine learning models with appropriate preprocessing techniques—such as SMOTE—bridges the gap between theoretical performance and operational applicability, laying the foundation for proactive fault management and enhanced service quality in HFC networks.

Author Contributions: Conceptualization, J.P.-R.; methodology, J.P.-R.; software, C.S.; validation, J.P.-R. and C.S.; formal analysis, J.P.-R.; investigation, J.P.-R.; resources, A.L. and E.D.; data curation, C.S.; writing—original draft preparation, J.P.-R., C.S., A.L. and E.D.; writing—review and editing, J.P.-R., C.S., A.L. and E.D.; visualization, J.P.-R. and C.S.; supervision, A.L. and E.D.; project administration, A.L. and E.D.; funding acquisition, A.L. and E.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Agencia Nacional de Investigación y Desarrollo (ANID-Chile) Doctorado Nacional, grant number 2022-21220867, and Agencia Nacional de Investigación y Desarrollo (ANID-Chile) FONDECYT REGULAR, grant number 1241362. The APC was funded by ANID.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy reasons.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. OECD. Percentage of Fibre Connections in Total Broadband. December 2023. Available online: <https://www.oecd.org/en/topics/sub-issues/broadband-statistics.html> (accessed on 4 October 2024).
2. CableLabs. DOCSIS 4.0 MAC and Upper Layer Protocols Interface Specification. Specifications CM-SP-MULPIv4.0, CableLabs. 2023. Available online: <https://www.cablelabs.com/specifications/CM-SP-MULPIv4.0> (accessed on 13 May 2024).

3. Cisco. Cisco Annual Internet Report (2018–2023) White Paper. Cisco, 2020. Available online: <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html> (accessed on 24 September 2025).
4. Ji, R.; Gao, J.; Xie, G.; Flowers, G.T.; Jin, Q. The impact of coaxial connector failures on high frequency signal transmission. In Proceedings of the 2015 IEEE 61st Holm Conference on Electrical Contacts (Holm), San Diego, CA, USA, 11–14 October 2015; pp. 298–303. [CrossRef]
5. CableLabs. PNM Best Practices Primer: HFC Networks (DOCSIS 3.1). Specifications CM-GL-PNM-3.1-V01-200506, CableLabs. 2024. Available online: <https://www.scribd.com/document/563981463/CM-GL-PNM-3-1-V01-200506-3> (accessed on 13 May 2024).
6. Williams, T.H.; Hunter, D. Isolating an Upstream Noise Source in a Cable Television Network. U.S. Patent 9,729,257, 8 August 2017.
7. Mahmoud, H.H.H.; Ismail, T. A review of machine learning use-cases in telecommunication industry in the 5G Era. In Proceedings of the 2020 16th International Computer Engineering Conference (ICENCO), Cairo, Egypt, 29–30 December 2020; pp. 159–163.
8. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority over-Sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [CrossRef]
9. Abedin, S.; Ben Ghorbel, M.; Hossain, M.J.; Berscheid, B.; Howlett, C. A Novel Approach for Profile Optimization in DOCSIS 3.1 Networks Exploiting Traffic Information. *IEEE Trans. Netw. Serv. Manag.* **2019**, *16*, 578–590. [CrossRef]
10. Villamar, V.; Rocha, C.; Navarrete, H.; Lupera-Morillo, P. A Predictive Handover Approach in LTE Networks with Measurements and Decision Tree Algorithms (Case Study City of Quito). *Rev. Politécnica* **2023**, *52*, 15–24. [CrossRef]
11. Benhavan, T.; Songwatana, K. HFC network performance monitoring system using DOCSIS cable modem operation data in a 3 dimensional analysis. In Proceedings of the 4th Joint International Conference on Information and Communication Technology, Electronic and Electrical Engineering (JICTEE), Chiang Rai, Thailand, 5–8 March 2014; pp. 1–5. [CrossRef]
12. Gibellini, E.; Righetti, C.E. Unsupervised Learning for Detection of Leakage from the HFC Network. In Proceedings of the 2018 ITU Kaleidoscope: Machine Learning for a 5G Future (ITU K), Santa Fe, Argentina, 26–28 November 2018; pp. 1–8.
13. Millicom. *Millicom Earnings Release Q1 2024*; Technical Report; Millicom: Luxembourg, 2024.
14. Simakovic, M.; Cica, Z. Detection and localization of failures in hybrid fiber–coaxial network using big data platform. *Electronics* **2021**, *10*, 2906. [CrossRef]
15. Jordan, M.I.; Mitchell, T.M. Machine learning: Trends, perspectives, and prospects. *Science* **2015**, *349*, 255–260. [CrossRef]
16. Choudhary, R.; Gianey, H.K. Comprehensive review on supervised machine learning algorithms. In Proceedings of the 2017 International Conference on Machine Learning and Data Science (MLDS), Noida, India, 14–15 December 2017; pp. 37–43.
17. Dargan, S.; Kumar, M.; Ayyagari, M.R.; Kumar, G. A Survey of Deep Learning and Its Applications: A New Paradigm to Machine Learning. *Arch. Comput. Methods Eng.* **2020**, *27*, 1071–1092. [CrossRef]
18. Lei, Y.; Yang, B.; Jiang, X.; Jia, F.; Li, N.; Nandi, A.K. Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mech. Syst. Signal Process.* **2020**, *138*, 106587. [CrossRef]
19. Wang, D.; Zhang, C.; Chen, W.; Yang, H.; Zhang, M.; Lau, A.P.T. A review of machine learning-based failure management in optical networks. *Sci. China Inf. Sci.* **2022**, *65*, 211302. [CrossRef]
20. Kotsiantis, S.B.; Zaharakis, I.; Pintelas, P. Supervised machine learning: A review of classification techniques. *Emerg. Artif. Intell. Appl. Comput. Eng.* **2007**, *160*, 3–24.
21. Arulkumar, K.; Deisenroth, M.P.; Brundage, M.; Bharath, A.A. Deep Reinforcement Learning: A Brief Survey. *IEEE Signal Process. Mag.* **2017**, *34*, 26–38. [CrossRef]
22. Oshiro, T.M.; Perez, P.S.; Baranauskas, J.A. How many trees in a random forest? In *Machine Learning and Data Mining in Pattern Recognition, Proceedings of the 8th International Conference, MLDM 2012, Berlin, Germany, 13–20 July 2012*; Proceedings 8; Springer: Berlin/Heidelberg, Germany, 2012; pp. 154–168.
23. Elkabalawy, M.; Al-Sakkaf, A.; Abdelkader, E.M.; Alfalah, G. CRISP-DM-Based Data-Driven Approach for Building Energy Prediction Utilizing Indoor and Environmental Factors. *Sustainability* **2024**, *16*, 7249. [CrossRef]
24. Woundy, R.; Marez, K. Cable Device Management Information Base for Data-Over-Cable Service Interface Specification (DOCSIS) Compliant Cable Modems and Cable Modem Termination Systems. RFC 4639, Internet Engineering Task Force (IETF), December 2006. Available online: <https://www.rfc-editor.org/info/rfc4639> (accessed on 24 September 2025).
25. Matharaarachchi, S.; Domaratzi, M.; Muthukumarana, S. Enhancing SMOTE for imbalanced data with abnormal minority instances. *Mach. Learn. Appl.* **2024**, *18*, 100597. [CrossRef]
26. Williams, C.K.I. The Effect of Class Imbalance on Precision-Recall Curves. *Neural Comput.* **2021**, *33*, 853–857. [CrossRef]

27. Chicco, D.; Tötsch, N.; Jurman, G. The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Mining* **2021**, *14*, 13. [CrossRef]
28. Chicco, D.; Jurman, G. A statistical comparison between Matthews correlation coefficient (MCC), prevalence threshold, and Fowlkes–Mallows index. *J. Biomed. Inform.* **2023**, *132*, 104426. [CrossRef] [PubMed]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Sensor-Driven RSSI Prediction via Adaptive Machine Learning and Environmental Sensing

Anya Apavatjirut

OASYS Research Group, Computer Engineering, Chiang Mai University, Chiang Mai 50200, Thailand;
anya@eng.cmu.ac.th

Abstract: Received Signal Strength Indicator (RSSI) prediction is valuable for network planning and optimization as it helps determine the optimal placements of wireless access points and enables better coverage planning. It is also crucial for efficient handover management between cells or access points, reducing dropped connections and improving service quality. Additionally, RSSI prediction supports indoor positioning systems, power management optimization, and cost-efficient network deployment. Path loss models have historically served as the foundation for RSSI prediction, providing a theoretical framework for estimating signal strength degradation. However, modern machine learning approaches have emerged as a revolutionary solution for network optimization, providing more versatile and data-driven methods to enhance wireless network performance. In this paper, an adaptive machine learning framework integrating environmental sensing parameters such as temperature, relative humidity, barometric pressure, and particulate matter for RSSI prediction is proposed. Performance analysis reveals that RSSI values are influenced by environmental factors through complex, non-linear interactions, thereby challenging the conventional linear assumptions of traditional path loss models. The proposed model demonstrates improved predictive accuracy over the baseline, with relative increases in variance explained of 6.02% and 2.04% compared to the baseline model excluding and including environmental parameters, respectively. Additionally, the root mean squared error is reduced to 1.40 dB. These results demonstrate that cognitive methods incorporating environmental data can substantially enhance RSSI prediction accuracy in wireless communications.

Keywords: machine learning; RSSI; LoRaWAN; radio propagation; sensors; RSSI prediction; RSSI estimation; path loss model

1. Introduction

A Received Signal Strength Indicator (RSSI) is a crucial metric that measures the power level of a radio signal received by a wireless device. Expressed in decibels (dB) or decibel-milliwatts (dBm), RSSI quantifies the signal strength at the receiving antenna and serves as an indicator of link quality and reliability.

In recent wireless technologies, the RSSI plays a vital role in optimizing network performance. It enables the identification and mitigation of interference sources, enhancing the overall quality of services [1,2]. By leveraging RSSI data, network operators can anticipate signal strength variations and dynamically adapt transmission power, routing strategies, and sleep cycles. This optimization leads to reduced power consumption and improved network reliability [3,4]. Additionally, RSSI information facilitates intelligent decision-making regarding base station placement, antenna configurations, and power

management [1,5]. In scenarios where GPS signals are ineffective, the RSSI serves as a valuable alternative for range-based localization [6,7].

Traditional RSSI estimation at varying distances relies on path loss models to characterize signal attenuation between the transmitter and the receiver. However, these models face several limitations in accurately predicting real-world signal behavior. First, radio signals are not only susceptible to path loss but also to various other factors that affect signal degradation, such as fading and multi-path propagation. Additionally, the path loss model's performance typically depends on the specific environment where it is deployed. As a result, specialized knowledge of radio propagation is required to identify an appropriate model for each specific location. Both empirical models, which rely on statistically collected data, and linear regression methods have been introduced to optimize path loss parameters.

In contrast to conventional approaches that rely on regression to adjust environment-specific parameters in path loss models, this paper introduces a fully dynamic and adaptive machine learning approach that operates independently of traditional modeling, offering greater flexibility and adaptability. Building on the study by González-Palacio et al. [8], which employed multiple linear regressions to model path loss and establish a predictive baseline, this paper introduces a novel adaptive machine learning framework for RSSI prediction using the same dataset. The results demonstrate that the proposed technique improves RSSI prediction accuracy, and this improvement is further enhanced when environmental factors such as temperature, humidity, barometric pressure, and particulate matter (PM2.5) collected from sensors are included as features. These findings suggest that the relationship between the RSSI and environmental parameters may be more complex than previously assumed. Although individual linear correlations between parameters might be weak, analyzing them collectively through machine learning reveals intricate interactions, highlighting the cognitive nature of path loss modeling. This innovative approach enables more accurate and adaptive RSSI prediction, allowing network operators to optimize their systems based on real-time environmental conditions.

2. Literature Review

Traditional path loss models have been widely used for RSSI prediction in wireless communications. These models initially focused on basic parameters such as distance and frequency as the primary sources of signal attenuation [9,10]. Over time, researchers incorporated additional factors into these models, such as fading and multi-path effects from obstacles and environmental conditions, through extra path loss exponents in the log-distance path loss model [10–12].

The evolution of path loss modeling has led to increasingly specialized models for various environments and conditions. Researchers have developed specific models for urban, forest, and coastal areas [13], as well as indoor, campus, city, and suburban environments [14]. These models account for various factors including antenna characteristics [15], human presence and movement [16], terrain profiles with diffraction and fading [17], foliage thickness [18], tree height, and electromagnetic wave polarization [19].

Despite numerous attempts to create highly accurate path loss models, a significant limitation remains: these models are typically constrained by location-specific environmental characteristics, as multiple factors can affect signal propagation. Empirical path loss models were later developed to account for various environmental factors using extrapolated measurement data [13,20].

Machine learning has revolutionized wireless communication optimization in recent years, offering powerful new solutions across multiple domains. One particularly promising application is in path loss modeling, where machine learning algorithms were used to

perform sophisticated regression analysis to optimize environmental parameters to enhance prediction accuracy. Early implementations successfully integrated machine learning techniques with conventional path loss models, enabling more precise parameter tuning based on specific environmental conditions. This hybrid approach represents an advancement in modeling wireless signal propagation and environmental interactions. There are several previous research studies that employed various techniques such as linear regression, support vector regression, random forests [21], and artificial neural networks [22] to account for environmental data and antenna characteristics [8,19,23,24]. Although regression-based path loss modeling may introduce accuracy improvements, the path loss model still relies on basic mathematical forms such as linear or exponential equations, which may not fully capture the complexity of real-world signal behavior.

More recently, researchers have introduced fully adaptive machine learning approaches for RSSI prediction. For instance, in [25], the authors focused on cellular configuration, leveraging antenna characteristics to predict the reference signal received power (RSRP). This method bypasses traditional path loss modeling, offering greater flexibility by avoiding the constraints of conventional models. Machine learning enables the identification of complex, non-linear relationships among environmental variables, resulting in more accurate and adaptive signal strength predictions.

In [26], Goldoni et al. collected an empirical dataset from LoRaWAN sensor nodes deployed in an agricultural field, which included environmental measurements such as temperature, relative humidity, barometric pressure, and soil characteristics. Their study underscored the potential of these variables for developing more accurate path loss and shadowing models. While the paper presented basic analyses of the relationships among these variables, it did not propose a unified model that integrates them for RSSI prediction.

The importance of incorporating environmental parameters into path loss modeling was further emphasized in [8], where multiple linear regressions were applied to enhance traditional log-distance path loss models. The study demonstrated the benefits of integrating environmental measurements to improve prediction accuracy and introduced two distinct modeling approaches to achieve this. Expanding on this research work, which employed multiple linear regression-based path loss model for estimation, this paper adopts a fundamentally different strategy by leveraging the dynamic and adaptive capabilities of machine learning to construct a purely data-driven framework for RSSI prediction. Instead of explicitly predicting the RSSI by modeling path loss, this paper proposes using trained models to capture complex, non-linear interactions among variables, thereby eliminating the reliance on predefined propagation models. This methodology aligns with the approach introduced by Fauzi et al. [25], where antenna characteristics were incorporated as predictive features within a machine learning framework. In contrast, the present study emphasizes environmental parameters as the primary features. Utilizing the same dataset examined by González-Palacio et al. [8], the selected environmental features include temperature, relative humidity, barometric pressure, and particulate matter (PM_{2.5}).

3. Methodology Overview

This section outlines the experimental framework, including both hardware and software components. It provides an overview of the dataset, including the collection procedures and parameter characteristics. The specifications of the baseline model are described, followed by the implemented machine learning algorithms, their configurations, and the evaluation metrics employed in this study.

3.1. Experimental Framework

The dataset used in this study is derived from the data provided by González-Palacio et al. [8]. The data is collected from a LoRaWAN network consisting of four LoRa nodes communicating with a single gateway using the LoRa radio frequency protocol. A detailed description of the hardware and dataset is presented in this section.

3.1.1. Hardware Description

According to the hardware description in [8], the four LoRa nodes communicated with a gateway using a radio frequency protocol operating in the 902–928 MHz band. The transmission bandwidth is 125 kHz and cycles through various sub-bands in the ISM US915 band (903.9, 904.1, 904.3, 904.7, 904.9, 905, and 905.1 MHz). The LoRa nodes were positioned at distances of $d_1 = 2.11$, $d_2 = 3.42$, $d_3 = 5.32$, and $d_4 = 8.21$ km, respectively, from the gateway, as illustrated in Figure 1.

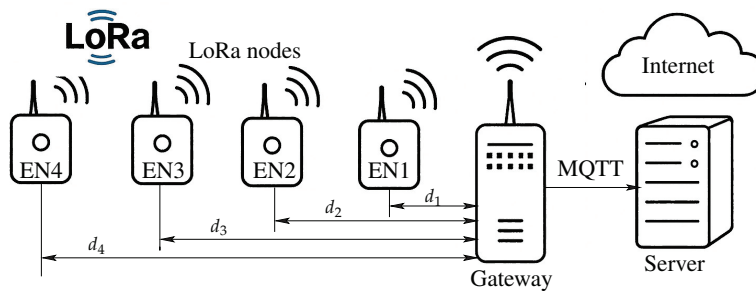


Figure 1. LoRaWAN architecture for dataset retrieval.

The LoRa nodes, also referred to as end nodes in the referenced paper, were based on the PyCom LoPy4 platform. These nodes were equipped with an SX1276 (Semtech Corp., Camarillo, CA, USA) radio for LoRa transmission. The antenna used is an omnidirectional PSKN3-900 (Mobile Mark Inc., Itasca, IL, USA) with a peak gain of 3 dBi. The nodes measured environmental variables using various sensors, including a DHT22 (Aosong Electronics Co., Guangzhou, China) for temperature and relative humidity, a BMP280 (Bosch Sensortec GmbH, Reutlingen, Germany) for barometric pressure, an HPM11500 (Honeywell International Inc., Charlotte, NC, USA) for particulate matter (PM_{2.5}), and an INA219 (Texas Instruments Inc., Dallas, TX, USA) for energy consumption.

The LG308 (Dragino Technology Co., Shenzhen, China) gateway functions as an intermediary between the LoRa nodes and the network server. It is equipped with two radios SX1257 and SX1301 (Semtech Corp., Camarillo, CA, USA) and features an antenna with a peak gain of 4.4 dBi. Upon receiving data from the nodes, the gateway transmits the information to a cloud-hosted message queuing telemetry transport (MQTT) broker, where the payloads are published for subsequent processing.

3.1.2. Dataset Description

The provided dataset contains 930,753 observations recorded in Medellín, Colombia, from October 2021 to March 2022, with a mean sampling interval of 60 s. The data includes the following information, as presented in Table 1 [8].

Based on the dataset described in Table 1, eight relevant fields were selected for this study. The selected fields include the target variable for prediction, the Received Signal Strength Indicator (*rss_i*), and key features commonly used in path loss models, such as frequency (*f*) and distance (*d*). Additionally, environmental parameters including temperature (*T*), relative humidity (*rh*), barometric pressure (*bp*), and particulate matter (*pm2_5*) are incorporated to enhance model accuracy. Additionally, the signal-to-noise ratio (*snr*) is included as part of the training data. Although correlation analysis indicates a very

strong relationship between the SNR and the RSSI, suggesting that the SNR alone could serve as an effective predictor of the RSSI, this study includes the SNR as an optional feature. This decision is made to maintain consistency and comparability with the baseline path loss regression model, as described in Equation (3), which already incorporates the SNR as one of its parameters. An initial dataset analysis was conducted in this paper to explore the relationships among these variables, focusing on their correlation and distribution, as illustrated in Figures 2 and 3, respectively.

Table 1. Dataset details and descriptions.

Column Name	Description
index	Sequential number identifying the corresponding observation.
timestamp	Date and time of the observation in the format of yyyy-mm-dd hh:mm:ss.
device_id	Identifier of the LoRa node defined as EN1, EN2, EN3, or EN4.
distance (<i>d</i>)	Distance between the corresponding LoRa node and the gateway in meters.
ht	Antenna height of the transmitter (LoRa node) in meters.
hr	Antenna height of the receiver (gateway) in meters.
ptx	Transmitted radiated power in dBm (fixed at 20 dBm).
ltx	Loss associated with cables and connectors at the transmitter in dB.
gtx	Antenna gain of the transmitter (LoRa node) in dBi.
lrx	Loss associated with cables and connectors at the receiver (gateway) in dB (measured as 4.25 dB).
grx	Antenna gain of the receiver (gateway) measured as 4.161 dBi.
frequency (<i>f</i>)	Carrier frequency in Hz, operating in the US 902–928 MHz ISM band.
frame_length	Number of bytes in the current transmission payload.
temperature (<i>T</i>)	Temperature in °C.
rh	Relative humidity in %.
bp	Barometric pressure in hPa.
pm2_5	Particulate matter (PM2.5) in $\mu\text{g}/\text{m}^3$.
rss_i	The Received Signal Strength Indicator measured at the gateway in dBm.
snr	The signal-to-noise ratio in dB.
toa	Time on air in seconds.
experimental_pl	Experimental path loss in dB, calculated as $\text{ptx} + \text{gtx} + \text{grx} - \text{lrx} - \text{rss}_i$.
energy	Energy consumed during the current transmission in Joules.
esp	Effective signal power of the current transmission in dBm.
pn	Noise power in dBm.

Figure 2 presents the correlation coefficients between variables. Notably, barometric pressure exhibits a strong positive correlation with the RSSI, with a coefficient of 0.91. This relationship may be influenced by the varying antenna installation heights, as described in the data collection setup. A higher barometric pressure can enhance radio

wave propagation through atmospheric ducting effects, which could explain the observed correlation. Additionally, air quality, as represented by the PM2.5 concentration, shows a moderate positive correlation with the RSSI, suggesting that environmental conditions affect signal behavior.

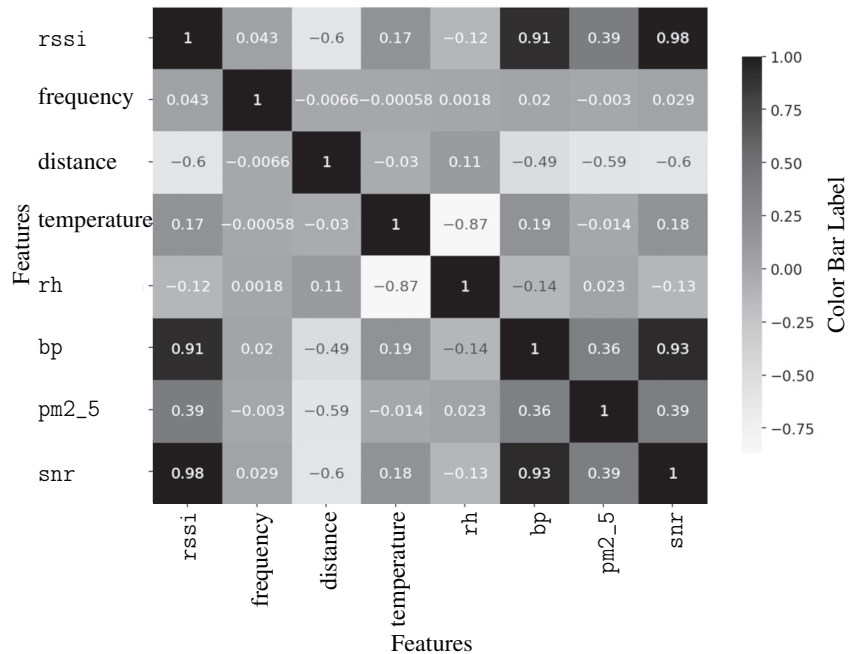


Figure 2. Correlation matrix of variables of interest.

Figure 3 illustrates the distribution profiles of the variables of interest, with the y-axis indicating the frequency of occurrence for each value. The analysis reveals that all variables deviate from normal distributions, with PM2.5 showing the most extreme departure through high positive skewness (1.91) and heavy tails, indicating frequent pollution spikes. Humidity also has some outlier measurements, but they are much fewer. Temperature shows slight positive skewness, indicating occasional hot extremes, while humidity exhibits negative skewness, suggesting more frequent high-humidity conditions.

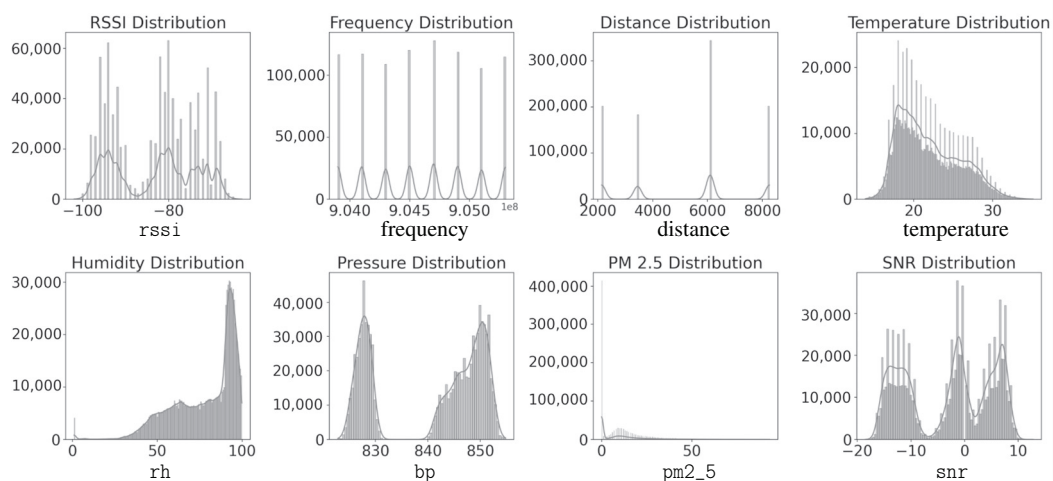


Figure 3. Distribution profiles of variables of interest.

3.2. Baseline Model

Traditionally, the Received Signal Strength Indicator (RSSI) is derived from path loss models, where the log-distance equation serves as the fundamental formulation. To

improve the accuracy of predictions, several additional parameters have been introduced, typically involving antenna configurations and environmental factors. Such parameters, as proposed in [8], aim to better align the model with real-world conditions.

The first path loss model mentioned in [8] involves using a multiple linear regression algorithm to determine the coefficients of the well-known *Log-Distance Path Loss Model* (LDPLM), which accounts for path loss along with a random component representing shadowing effects, as shown in Equation (1):

$$\text{LDPLM} = K + 10 \cdot \gamma \cdot \log_{10} \left(\frac{d}{d_0} \right) + \psi. \quad (1)$$

In this equation, K represents a scaling factor that adjusts the received signal strength based on system-specific parameters such as antenna gain, system losses, and environmental characteristics; d denotes the distance between the transmitter and the receiver; d_0 is the reference distance, typically set within the far-field region; γ is the path loss exponent, which characterizes the rate at which the signal attenuates with distance; and ψ is a random variable representing shadowing effects, whose statistical distribution is defined as follows in Equation (2):

$$p(\psi) = \frac{\xi}{\sqrt{2\pi} \cdot \sigma_\psi \cdot \psi} \cdot \exp \left(-\frac{(10 \cdot \log_{10}(\psi) - \mu_\psi)^2}{2 \cdot \sigma_\psi^2} \right), \quad \psi > 0, \quad (2)$$

where $\xi = \frac{10}{\ln(10)}$, σ_ψ is the standard deviation of ψ , and μ_ψ is its mean. Based on the results of linear regression analysis, the optimal coefficients that best fit the dataset were specified as $\gamma = 2.739$ and $K = 1.75$.

The second model proposed in [8] is the *Multiple Linear Regression Model* (MLRM), which extends the LDPLM by incorporating environmental variables such as temperature (T), relative humidity (rh), barometric pressure (bp), particulate matter ($\text{pm}_{2.5}$), and the signal-to-noise ratio (snr). These variables are empirically determined through regression analysis to better align the model with real-world signal behavior. The model is expressed as follows:

$$\text{MLRM} = \hat{\beta}_0 + 10\gamma \log_{10}(d) + 20 \log_{10}(f) + \hat{\beta}_1 T + \hat{\beta}_2 \text{rh} + \hat{\beta}_3 \text{bp} + \hat{\beta}_4 \text{pm}_{2.5} + \hat{\beta}_5 \text{snr} + \psi. \quad (3)$$

In this equation, $\hat{\beta}_0$ represents the offset coefficient, which is analogous to K in Equation (1). The coefficients $\hat{\beta}_i$ for $i \in \{1, \dots, 5\}$ were determined through regression analysis. The model predictors include f , the frequency in Hertz (Hz); T , the temperature in degrees Celsius; rh , the relative humidity in percentage (%); bp , the barometric pressure in hectopascals (hPa); $\text{pm}_{2.5}$, the concentration of particulate matter with a diameter less than $2.5 \mu\text{m}$ ($\mu\text{g}/\text{m}^3$); and snr , the signal-to-noise ratio. The term ψ denotes the shadow fading component, as defined in Equation (2). The coefficient values for each predictor in the MLRM, based on the approach proposed in [25], are summarized in Table 2.

Table 2. Model predictors and coefficient values for the Multiple Linear Regression Model.

Predictor	Variable	Value	Unit
Intercept	$\hat{K}$	−505.640	dB
Path loss exponent	$\hat{\gamma}$	2.203	-
Temperature	$\hat{\beta}_1$	0.123	dB/°C
Relative humidity	$\hat{\beta}_2$	0.011	dB/%
Barometric pressure	$\hat{\beta}_3$	0.407	dB/hPa
PM2.5	$\hat{\beta}_4$	0.002	dB/ $\mu\text{g}/\text{m}^3$
SNR	$\hat{\beta}_5$	−0.635	dB/dB

3.3. Proposed Model

The model proposed in this paper adopts a novel methodology that utilizes the adaptive capabilities of machine learning to develop a fully data-driven framework for RSSI estimation, referred to as the *Adaptive Machine Learning Model* (AMLM). Rather than relying on traditional path loss equations, the approach employs trained algorithms to capture intricate, non-linear relationships among influencing factors, thereby eliminating the need for predefined signal propagation models.

3.3.1. Model Selection

Low-complexity models, particularly tree-based ensemble methods, were primarily selected in this paper due to the resource constraints of the embedded LoRa gateway framework. This choice is supported by findings reported in [27], indicating that lightweight random forest regressors outperform neural networks with respect to RSSI estimation accuracy, reliability, and computational efficiency. The models explored in this paper range from basic approaches such as linear regression to more advanced regression techniques.

The first basic model selected is *linear regression* (LR) [28], which assumes a linear relationship between input features and the target variable, aiming to determine the optimal line that best fits the data. Next, *decision trees* (DTs) [29] are included. These models build hierarchical structures by recursively splitting the dataset based on feature values, offering interpretable decision-making capabilities, although they are prone to overfitting.

Tree-based ensemble methods considered such as the *random forest* (RF) [30], *Gradient Boosting* [31], *Extreme Gradient Boosting* (XGBoost or XGB) [32], and *Categorical Boosting* (CatBoost or CB) [33] combine multiple weak learners to form a strong predictive model. The random forest is an ensemble of decision trees that mitigate overfitting and enhance generalization. Gradient Boosting constructs models in a stage-wise manner, iteratively minimizing a loss function to improve performance. XGBoost and CatBoost are advanced implementations of Gradient Boosting, offering additional features and optimizations that improve speed and accuracy. CatBoost was specifically developed to efficiently handle categorical data.

The objective of this study is to evaluate these machine learning algorithms based on their effectiveness in predicting RSSI values. This evaluation aims to assess how well these algorithms can model and capture non-linear relationships among environmental and signal-related variables.

3.3.2. Model Setup

In this study, the framework is implemented using Python libraries, as outlined in Table 3. For ensemble and boosted tree methods, a total of 100 base estimators with a maximum depth of 5 are employed. To ensure reproducibility, the random seed is fixed at 42. The learning rate for both XGBoost and CatBoost is set to 0.1.

Table 3. Machine learning libraries.

Technique	Library
LR	<code>sklearn.linear_model.LinearRegression()</code>
DT	<code>sklearn.tree.DecisionTreeRegressor()</code>
RF	<code>sklearn.ensemble.RandomForestRegressor()</code>
GB	<code>sklearn.ensemble.GradientBoostingRegressor()</code>
XGB	<code>xgboost.XGBRegressor()</code>
CB	<code>catboost.CatBoostRegressor()</code>

The proposed framework was specifically trained to estimate received signal strength at the gateway. To ensure a robust evaluation, the dataset is divided into the training (75%) and testing (25%) subsets.

3.4. Evaluation Metrics

To evaluate the performance of the machine learning algorithms, two key metrics were selected: the Percentage of Variance (PoV) and the Root Mean Squared Error (RMSE). In addition, the Mean Absolute Percentage Error (MAPE) is also presented to provide further insight into the prediction accuracy across different models.

3.4.1. Percentage of Variance

The *Percentage of Variance* (PoV) assesses the percentage of total variance in observed data that can be accounted for by a model's predictions, providing an indication of the model's capability to capture underlying patterns and relationships in the data. A higher PoV value suggests a better fit and more precise predictions.

The PoV is calculated using the following equation:

$$\text{PoV} = \left(1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \right) \times 100\%, \quad (4)$$

where n represents the number of observations in the dataset, y_i is the actual (observed) value of the dependent variable for the i -th observation, $\hat{y}_i$ denotes the predicted value given by the model, and $\bar{y}$ is the mean of the observed values.

3.4.2. Root Mean Squared Error

The *Root Mean Squared Error* (RMSE) measures the average magnitude of the errors between the predicted and actual values. In the context of RSSI prediction, a lower RMSE value indicates that the predicted distances are, on average, closer to the actual distances.

The RMSE is calculated using the following formula:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (5)$$

where n is the number of observations, y_i represents the actual (observed) value of the i -th observation, and $\hat{y}_i$ is the predicted value of the i -th observation.

3.4.3. Mean Absolute Percentage Error

The *Mean Absolute Percentage Error* (MAPE) quantifies the average magnitude of prediction errors relative to the actual values and is expressed as a percentage. This characteristic makes it particularly effective for comparing prediction performance across different models. The MAPE is calculated as follows:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\%, \quad (6)$$

where n is the total number of observations, y_i denotes the actual value of the i -th observation, and $\hat{y}_i$ represents the corresponding predicted value.

4. Results and Evaluation

This section presents a performance analysis of the proposed Adaptive Machine Learning Model (AMLML) model. The first subsection evaluates the prediction performance of the framework across various machine learning algorithms, followed by a comparative analysis of their results. Subsequently, the most effective models are selected and benchmarked

against predictive results based on the baseline Multiple Linear Regression Model (MLRM) to assess their relative performance.

4.1. Performance Analysis of the Proposed Model

The proposed Adaptive Machine Learning Model (AMLM) for RSSI prediction is developed based on several algorithms, including linear regression (LR), decision trees (DTs), random forest (RFs), Gradient Boosting (GB), Extreme Gradient Boosting (XGBoost or XGB), and Categorical Boosting (CatBoost or CB). The models are trained on input features derived from traditional path loss models, specifically distance (d) and frequency (f), as well as environmental variables that may influence the RSSI, such as temperature (T), relative humidity (rh), barometric pressure (bp), and fine particulate matter with a diameter of 2.5 micrometers or less (pm2_5). Additionally, the signal-to-noise ratio (snr) is incorporated to ensure consistency with the baseline model, thereby enabling fair and meaningful comparisons. Model performance is evaluated using multiple metrics, including the Percentage of Variance (PoV), the Root Mean Square Error (RMSE), and execution time, as presented in Tables 4–6, respectively.

Table 4. PoV-based evaluation predictions across different feature selection methods.

Feature Selection	LR	DT	RF	GB	XGB	CB
d	36.48	93.53	93.53	93.53	93.53	93.53
f	0.19	0.56	0.56	0.56	0.56	0.56
d, f	36.64	93.78	93.78	93.80	93.81	93.81
d, f, T	38.92	93.94	93.95	94.06	94.12	94.07
d, f, rh	36.97	94.00	94.01	94.12	94.19	94.15
d, f, bp	86.35	93.82	93.82	93.91	93.95	93.93
$d, f, pm2_5$	36.79	94.50	94.51	94.58	94.63	94.61
$d, f, T, rh, bp, pm2_5$	86.62	94.80	94.82	95.16	95.69	95.38
$d, f, snr, T, rh, bp, pm2_5$	96.88	97.54	97.58	97.79	97.86	97.79

Note: In Tables 4–6, two rows are highlighted in gray to indicate key configurations used for comparison in Section 4.2. Additionally, results from the XGBoost model are highlighted in bold to help quickly identify the most relevant findings discussed in the analysis.

Table 5. RMSE-based evaluation predictions across different feature selection methods.

Feature Selection	LR	DT	RF	GB	XGB	CB
d	7.61	2.43	2.43	2.43	2.43	2.43
f	9.55	9.53	9.53	9.53	9.53	9.53
d, f	7.61	2.38	2.38	2.38	2.38	2.38
d, f	7.61	2.38	2.38	2.38	2.38	2.38
d, f, T	7.47	2.35	2.35	2.33	2.32	2.33
d, f, rh	7.59	2.34	2.34	2.32	2.30	2.31
d, f, bp	3.53	2.38	2.38	2.36	2.35	2.35
$d, f, pm2_5$	7.60	2.24	2.24	2.22	2.21	2.22
$d, f, T, rh, bar, pm2_5$	3.49	2.18	2.17	2.10	1.98	2.05
$d, f, T, rh, bp, pm2_5$	3.49	2.18	2.17	2.10	1.98	2.05
$d, f, snr, T, rh, bp, pm2_5$	1.69	1.50	1.49	1.42	1.40	1.42

4.1.1. Percentage of Variance Analysis

The Percentage of Variance (PoV) results in Table 4 demonstrate that model performance improves significantly with the inclusion of additional environmental features. Traditionally, RSSI prediction relies on distance (d) and frequency (f), which are core components of path loss models. Using distance alone yields PoV values over 93% across most machine learning algorithms, except for linear regression, which performs poorly. In

contrast, frequency (f) has minimal predictive value, consistently resulting in PoV values below 1%.

Table 6. Time delay (s) from RSSI estimation with various feature selections.

Feature Selection	LR	DT	RF	GB	XGB	CB
d	0.02	0.03	4.72	5.48	0.87	3.46
f	0.02	0.07	9.58	8.80	1.23	3.58
d, f	0.04	0.15	15.46	16.75	2.91	4.61
d, f, T	0.05	0.32	27.18	28.89	1.67	4.87
d, f, rh	0.05	0.39	30.58	31.02	1.69	4.84
d, f, bp	0.05	0.31	26.66	28.38	1.55	5.00
$d, f, pm2_5$	0.05	0.26	22.86	24.99	1.64	5.26
$d, f, T, rh, bp, pm2_5$	0.11	0.82	58.69	63.21	2.67	5.65
$d, f, snr, T, rh, bp, pm2_5$	0.14	0.96	68.25	74.20	1.90	6.14

Combining distance and frequency offers a slight performance boost, but the most notable improvements occur when environmental parameters such as temperature, relative humidity, barometric pressure, and particulate matter are included. Among these, relative humidity and particulate matter are particularly influential, likely due to their effects on radio wave absorption and scattering. Incorporating all environmental features leads to the highest overall accuracy, emphasizing the value of a comprehensive feature set.

Extreme Gradient Boosting (XGBoost) stands out as the best-performing algorithm thanks to its gradient boosting framework enhanced by optimized tree pruning and regularization. Since the dataset contains only numerical features, Categorical Boosting is not applicable and offer no added benefit. Accordingly, XGBoost results are highlighted in bold in Table 4.

Using only the fundamental variables commonly found in traditional path loss models, such as frequency (f) and distance (d), XGBoost achieves a prediction accuracy (PoV) of 93.81%. When environmental variables are added (f, d , temperature (T), relative humidity (rh), barometric pressure (bp), and particulate matter ($pm2_5$)), accuracy improves to 95.69%. Incorporating the full feature set as defined in Equation (3), which includes the signal-to-noise ratio (snr), further boosts the PoV to 97.86%, demonstrating superior model generalization and fit. Two rows in Table 4 are highlighted in gray to indicate key configurations. These configurations will be referenced in the next section for comparison with RSSI predictions generated using the Multiple Linear Regression Model (MLRM) of path loss, which serves as the baseline.

4.1.2. Root Mean Square Error Analysis

The Root Mean Square Error (RMSE) values presented in Table 5 exhibit a trend consistent with the PoV results. Simpler models and limited feature sets result in higher prediction errors, with linear regression performing the worst. As more features are incorporated, the RMSE decreases across all models. XGBoost consistently achieves the lowest RMSE, particularly when the full feature set is utilized. Specifically, it reaches an RMSE of 1.40 when all parameters defined in Equation (3) are included, indicating minimal prediction error. This outcome reinforces XGBoost's robustness in handling complex, multi-dimensional data.

4.1.3. Execution Time

Execution time (Table 6) reveals a trade-off between accuracy and computational efficiency. While tree-based ensemble methods like the random forest and Gradient Boosting offer high accuracy, they incur significantly higher delays (e.g., over 68 s with full features).

In contrast, XGBoost maintains a balance, achieving top performance with relatively low execution time (e.g., 1.90 s with all features). Linear models are the fastest (e.g., 0.02–0.14 s), but their predictive performance is substantially lower. CatBoost also shows moderate execution time (up to 6.14 s) but does not outperform XGBoost in terms of accuracy.

The implementation based on a Python (version 3.11) script was executed on an Intel Core i7-4790S processor (Intel Corporation, Santa Clara, CA, USA) featuring four physical cores with hyper-threading (eight logical cores total), operating at a 3.20 GHz base frequency with turbo boost up to 4.00 GHz. The CPU includes a multi-level cache system with an 8 MiB shared L3 cache and supports an x86_64 architecture with advanced instruction sets including AVX2 for enhanced mathematical computations.

Overall, based on the proposed adaptive machine learning framework, Extreme Gradient Boosting (XGBoost) emerges as the most effective algorithm for RSSI prediction, offering the best trade-off in terms of accuracy (the highest PoV and the lowest RMSE). XGBoost demonstrates the most favorable balance between execution time and model accuracy, making it the most computationally efficient choice among the evaluated algorithms.

The inclusion of environmental and signal-related features significantly enhances model performance, underscoring the importance of comprehensive feature engineering in wireless signal modeling. Using XGBoost, the experimental results demonstrate that incorporating all environmental parameters as input features yields the highest prediction accuracy among all feature selection strategies, achieving a PoV of 97.86% and an RMSE of 1.40.

4.2. Evaluation Relative to the Baseline Model

This section evaluates the RSSI prediction results obtained from the proposed Adaptive Machine Learning Model (AMLM), with a particular emphasis on the Extreme Gradient Boosting (XGBoost) algorithm, which achieved the highest RSSI prediction accuracy among all evaluated models. These results are compared with the predictive results using the Multiple Linear Regression Model (MLRM) mentioned in [8].

The MLRM model, based on Equations (2) and (3), enabled path loss prediction with a PoV of 92.50% and an RMSE of 1.84 dB [8]. Unlike path loss prediction, the focus of this paper is on the Received Signal Strength Indicator (RSSI), which can be derived from path loss using a conversion equation also provided by [8]. The equation is as follows:

$$PL = ptx + gtx + grx - lrx - rssi. \quad (7)$$

The performance comparison of RSSI prediction between the two models is illustrated in Figure 4. The performance is evaluated under two metrics, namely the Percentage of Variance (PoV) and the Root Mean Square Error (RMSE), and is conducted under two scenarios: one that includes environmental variables and one that excludes them.

Figure 4 indicates that the proposed AMLM demonstrated superior performance, particularly when environmental data is incorporated. Compared to the baseline, the higher PoV achieved by the proposed model indicates its ability to explain a greater percentage of variance in the observed data and to more effectively capture underlying signal patterns. This improvement is consistent across both scenarios regardless of whether environmental factors are considered or not. Regarding the PoV, compared to the prediction when environmental features are excluded (i.e., the MLRM without environmental data, which is equivalent to the LDPLM described in Equation (1)), the PoV increases from 92.30% to 95.90%. When environmental features are included, the PoV further improves from 93.53% to 97.86%, indicating a strong correlation between predicted and actual RSSI values. According to the interpretation guidelines in [34], this level of PoV approaches near-perfect prediction, highlighting the model's high predictive power. Additionally, the

lower RMSE values down to 1.40 dB further validate the enhanced accuracy, indicating that the predicted RSSI values are, on average, closer to the actual measurements.

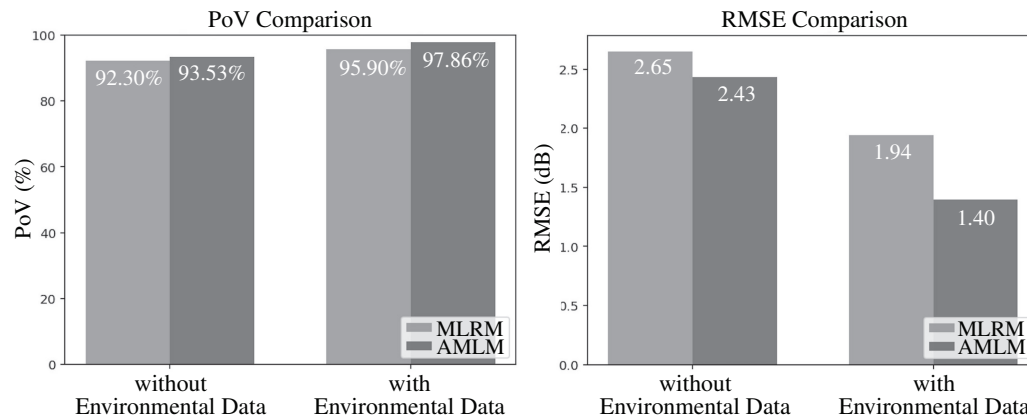


Figure 4. Comparison of the PoV and the RMSE Between the MLRM and the AMLM with and without environmental data for RSSI prediction

Figures 5 and 6 present scatter plots comparing actual and predicted RSSI values using the Multiple Linear Regression Model (MLRM) and the Advanced Machine Learning Model (AMLM), respectively, under identical conditions. The left plot shows predictions without considering environmental factors, while the right plot incorporates sensing data. In both plots, actual RSSI values are displayed on the x-axis and predicted values on the y-axis. Ideally, perfect predictions would align along the diagonal line, indicating a one-to-one correspondence. Compared to Figure 5, Figure 6 exhibits a tighter clustering around the diagonal line, especially when environmental data is included, highlighting superior predictive accuracy when environmental sensing data is considered.

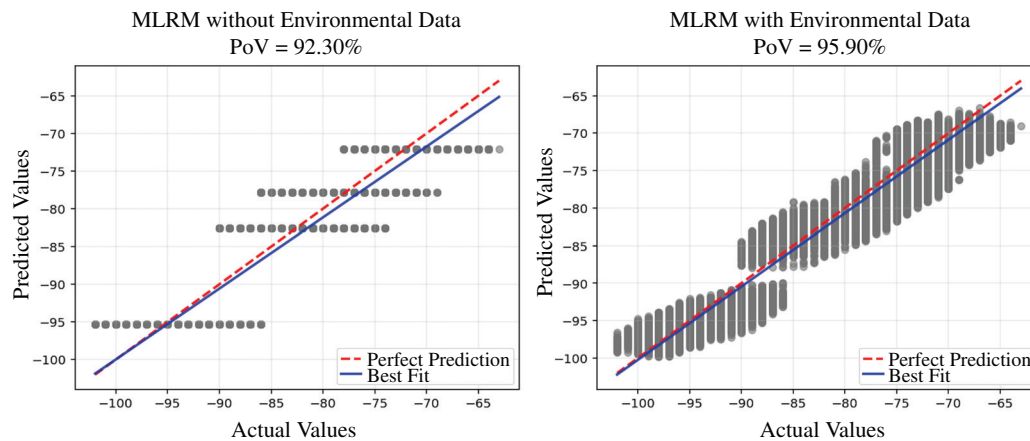


Figure 5. Scatter plot of actual RSSI values vs. predicted RSSI values of the baseline MLRM.

The integration of environmental factors into the adaptive machine learning model using XGBoost enhances its ability to capture the complex interactions between wireless signal propagation and the surrounding environment. By incorporating variables such as temperature, relative humidity, barometric pressure, and particulate matter (PM2.5), the model can better account for fluctuations in signal strength influenced by these conditions, leading to more accurate RSSI predictions.

Figure 7 compares the distribution of prediction errors across different models. Each box shows how far predicted values deviate from actual values, with the y-axis representing the prediction error. The plot helps assess model accuracy and consistency by highlighting bias, spread, and outliers in prediction performance. The accuracy improvement is evidenced in Figure 7, where the Mean Absolute Percentage Error (MAPE) of the AMLM

decreases from 2.42% to 1.37% when environmental factors are included, compared to the baseline model, which shows a reduction from 2.68% to 1.93%. This result demonstrates a meaningful and reliable improvement in performance.

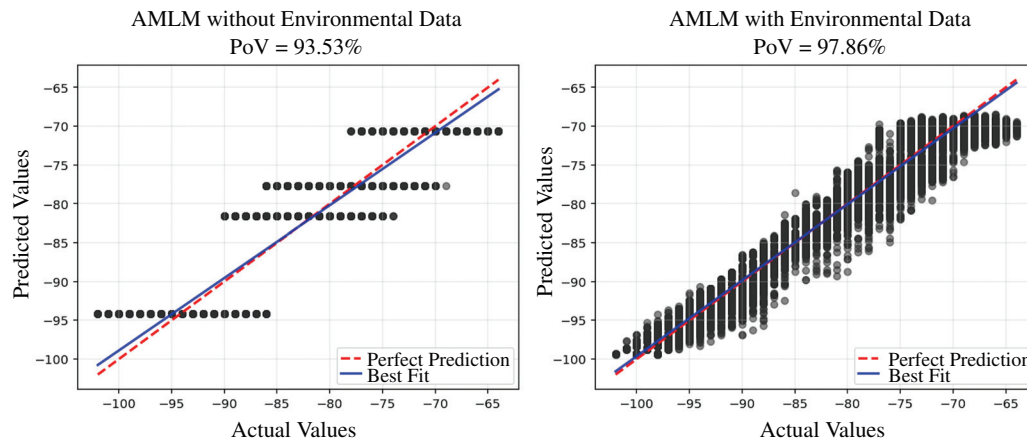


Figure 6. Scatter plot of actual RSSI values vs. predicted RSSI values of the proposed AMLM.

The proposed adaptive machine learning model demonstrates the capability to capture complex, non-linear relationships and interactions among features, characteristics that are often present in real-world wireless signal environments. Unlike traditional path loss modeling, which typically assumes linear relationships and is sensitive to outliers, machine learning approaches such as XGBoost can learn adaptively from data, handle noise more effectively, and automatically model intricate feature interactions. This leads to a better predictive performance, as evidenced by higher PoV and lower RMSE values.

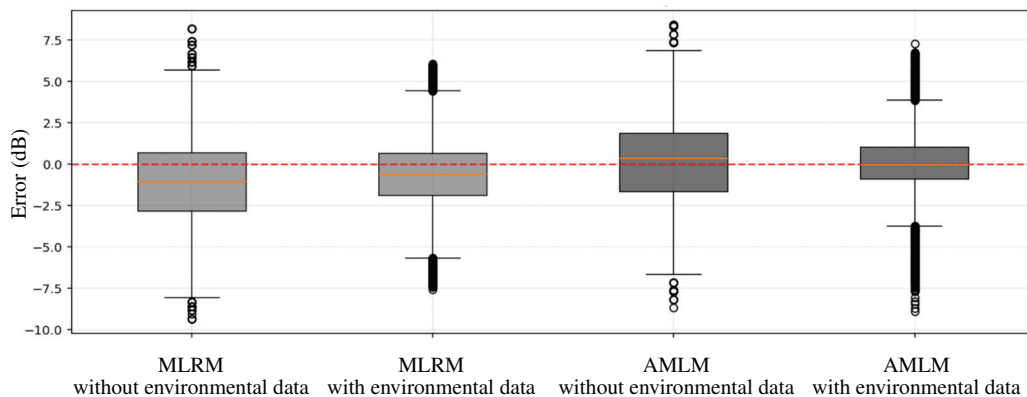


Figure 7. Comparison of prediction error distributions across various models.

5. Conclusions

This paper proposes an innovative approach for predicting the Received Signal Strength Indicator (RSSI) using adaptive machine learning techniques and environmental sensing data. Unlike conventional methods that rely on path loss modeling, which is typically based on linear assumptions, this sensor-driven framework dynamically learns complex, non-linear relationships among environmental variables. These include temperature, relative humidity, barometric pressure, and fine particulate matter (PM2.5), which are collected via a LoRa network.

The adaptive learning capability of the proposed model enables it to capture intricate dependencies that are often overlooked by traditional models. Experimental results demonstrate that the Extreme Gradient Boosting (XGBoost) algorithm achieves strong predictive performance, explaining 97.86% of the variance in the RSSI with a root mean square error of 1.40 dBm. The proposed model demonstrates improved predictive accuracy over the

baseline, with relative increases in variance explained of 6.02% (from 92.30% to 97.86%) and 2.04% (from 95.90% to 97.86%) when compared to the baseline model excluding and including environmental parameters, respectively. This outperforms the prediction using a baseline regression-based path loss model. By integrating advanced algorithms with multi-sensor data, the model offers enhanced accuracy and greater adaptability across diverse environmental conditions.

Despite its promising performance, this study's scope may be constrained by the use of data from only four LoRa nodes, each placed at a distinct distance. Future research can extend this framework by incorporating a wider range of distance measurements and environmental conditions to enhance generalizability. Feature selection may include physical obstacles, atmospheric conditions, electromagnetic interference, and environmental dynamics. Additionally, factors such as antenna characteristics and positioning could further improve the model's robustness and predictive accuracy.

Overall, this approach introduces a new direction for integrating environmental data into wireless communication systems. Through the use of multiple sensor inputs and sophisticated learning techniques, it offers a deeper understanding of how environmental factors influence signal propagation, with the potential to significantly improve network reliability and overall system performance.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Aiboud, Y.; Elhassani, I.; Griguer, H. RSSI optimization method for indoor positioning systems. In Proceedings of the International Conference on Wireless Networks and Mobile Communications (WINCOM), Rabat, Morocco, 1–4 November 2017; pp. 246–248.
2. Raj, N.; Wang, X. IndoorRSSINet—Deep learning based 2D RSSI map prediction for indoor environments with application to wireless localization. In Proceedings of the 15th International Conference on COMMunication Systems & NETWORKS (COMSNETS), Bangalore, India, 3–8 January 2023; pp. 609–616.
3. Bashyam, S.L.R.; Chandra, J.; Devi, S.R. An Efficient Node to Node Coverage and Connectivity with RSSI Using Grey-Wolf Prediction Optimization Algorithm in Remote Low Accessibility Area. In Proceedings of the 2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN), Vellore, India, 5–6 May 2023; pp. 1–6.
4. Li, M.; de Oliveira, T.E.A. EP-FPG applied to RSSI-Based Wireless Indoor Localization. In Proceedings of the IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), London, ON, Canada, 30 August–2 September 2020; pp. 1–6.
5. Ogawa, Y.; Umehira, M.; Wang, X. Kriging-based RSSI Prediction for Cell Coverage Discovery using Spectrum Database in 5G Multi-band Cellular Networks. In Proceedings of the 17th International Symposium on Communications and Information Technologies (ISCIT), Perth, WA, Australia, 11–13 December 2017; pp. 1–6.
6. Apavatjirut, A.; Boonyasiwapong, E. Indoor WIFI localization on embedded systems. In *Information Science and Applications*; Springer: Berlin/Heidelberg, Germany, 2015; Volume 339, ISBN 978-3-662-46577-6.
7. Daiya, V.; Ebenezer, J.; Murty, S.A.V.S.; Raj, B. Experimental Analysis of RSSI for Distance and Position Estimation. In Proceedings of the International Conference on Recent Trends in Information Technology (ICRTIT), Chennai, India, 3–5 June 2011; pp. 1093–1098.
8. González-Palacio, M.; Tobón-Vallejo, D.; Sepúlveda-Cano, L.M.; Rúa, S.; Pau, G.; Le, L.B. LoRaWAN Path Loss Measurements in an Urban Scenario including Environmental Effects. *Data* **2023**, *8*, 4. [CrossRef]
9. Molisch, A.F. *Wireless Communications*; John Wiley & Sons: Hoboken, NJ, USA, 2012.
10. Rademacher, M.; Linka, H.; Horstmann, T.; Henze, M. Path Loss in Urban LoRa Networks: A Large-Scale Measurement Study. In Proceedings of the 2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall), Norman, OK, USA, 27–30 September 2021; pp. 1–6.
11. Kabiri, M.; Cimarelli, C.; Bavle, H.; Sanchez-Lopez, J.L.; Voos, H. A Review of Radio Frequency Based Localisation for Aerial and Ground Robots with 5G Future Perspectives. *Sensors* **2023**, *23*, 188. [CrossRef] [PubMed]

12. Raj, N. Indoor RSSI Prediction using Machine Learning for Wireless Networks. In Proceedings of the International Conference on COMMunication Systems & NETworkS (COMSNETS), Bangalore, India, 5–9 January 2021; pp. 372–374.
13. Callebaut, G.; Perre, L.V. Characterization of LoRa Point-to-Point Path Loss: Measurement Campaigns and Modeling Considering Censored Data. *IEEE Internet Things J.* **2020**, *7*, 1910–1918. [CrossRef]
14. Chall, R.E.; Lahoud, S.; Helou, M.E. LoRaWAN Network: Radio Propagation Models and Performance Evaluation in Various Environments in Lebanon. *IEEE Internet Things J.* **2019**, *6*, 2366–2378. [CrossRef]
15. Alobaidy, A.A.H.; Mandeep, J.S.; Behjati, M.; Nordin, R.; Abdullah, N.F. Wireless Transmissions, Propagation and Channel Modeling for IoT Technologies: Applications and Challenges. *IEEE Access* **2021**, *10*, 24095–24131. [CrossRef]
16. Obeidat, H.; Shuaieb, W.; Obeidat, O.; Abd-Alhameed, R. A Review of Indoor Localization Techniques and Wireless Technologies. *Wirel. Pers. Commun.* **2021**, *119*, 289–327. [CrossRef]
17. Seybold, J.S. *Introduction to RF Propagation*; John Wiley & Sons: Hoboken, NJ, USA, 2005; p. 143, ISBN 0-471-65596-1.
18. Quispetupa, D.A.; Fernandez, J.P.; Tacas, R.Q. An experimental evaluation of RSSI-LoRa in an olive tree plantations. In Proceedings of the IEEE Engineering International Research Conference (EIRCON), Lima, Peru, 26–28 October 2022; pp. 1–4.
19. Barbosa, B.S.d.S.; Cruz, H.A.O.; Macedo, A.S.; Cardoso, C.M.M.; Fernandes, F.C.; Eras, L.E.C.; Araújo, J.P.L.d.; Calvacante, G.P.S.; Barros, F.J.B. Application of Artificial Neural Networks for Prediction of Received Signal Strength Indication and Signal-to-Noise Ratio in Amazonian Wooded Environments. *Sensors* **2024**, *24*, 2542. [CrossRef] [PubMed]
20. Hata, M. Empirical Formula for Propagation Loss in Land Mobile Radio Services. *IEEE Trans. Veh. Technol.* **1980**, *29*, 317–325. [CrossRef]
21. Sousa, M.; Alves, A.; Vieira, P.; Queluz, M.P.; Rodrigues, A. Analysis and Optimization of 5G Coverage Predictions Using a Beamforming Antenna Model and Real Drive Test Measurements. *IEEE Access* **2021**, *9*, 101787–101808. [CrossRef]
22. Jo, H.-S.; Park, C.; Lee, E.; Choi, H.K.; Park, J. Path Loss Prediction Based on Machine Learning Techniques: Principal Component Analysis, Artificial Neural Network, and Gaussian Process. *Sensors* **2020**, *20*, 1927. [CrossRef] [PubMed]
23. Moraitis, N.; Tsipi, L.; Gkioni, D.V.A.; Louvros, S. Performance evaluation of machine learning methods for path loss prediction in rural environment at 3.7 GHz. *Wirel. Netw.* **2021**, *27*, 4169–4188. [CrossRef]
24. He, R.; Gong, Y.; Bai, W.; Li, Y.; Wang, X. Random Forests Based Path Loss Prediction in Mobile Communication Systems. In Proceedings of the 2020 IEEE 6th International Conference on Computer and Communications (ICCC), Chengdu, China, 11–14 December 2020; pp. 1246–1250.
25. Fauzi, M.F.A.; Nordin, R.; Abdullah, N.F.; Alobaidy, H.A.H. Mobile Network Coverage Prediction Based on Supervised Machine Learning Algorithms. *IEEE Access* **2022**, *10*, 55782–55793. [CrossRef]
26. Goldoni, E.; Savazzi, P.; Favalli, L.; Vizziello, A. Correlation between weather and signal strength in LoRaWAN networks: An extensive dataset. *Comput. Netw.* **2022**, *202*, 108627. [CrossRef]
27. Arif, S.; Khan, M.A.; Rehman, S.U. A Lightweight Received Signal Strength Indicator Estimation Model for Low-Power Internet of Things Devices in Constrained Indoor Networks. *Appl. Sci.* **2025**, *15*, 3535. [CrossRef]
28. Freedman, D.A. *Statistical Models: Theory and Practice*; Cambridge University Press: Cambridge, UK, 2009.
29. Quinlan, J.R. Induction of Decision Trees. *Mach. Learn.* **1986**, *1*, 81–106. [CrossRef]
30. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
31. Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
32. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
33. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased Boosting with Categorical Features. In Proceedings of the 32nd International Conference on Neural Information Processing Systems, Montreal, QC, Canada, 3–8 December 2018; pp. 6639–6649.
34. Turney, S. Coefficient of Determination (R^2) | Calculation & Interpretation. Scribbr. 2022. Available online: <https://www.scribbr.com/statistics/coefficient-of-determination> (accessed on 1 July 2024).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Few-Shot Intelligent Anti-Jamming Access with Fast Convergence: A GAN-Enhanced Deep Reinforcement Learning Approach

Tianxiao Wang ^{1,2}, Yingtao Niu ^{2,*} and Zhanyang Zhou ^{1,2}

¹ The Sixty-Third Research Institute, National University of Defense Technology, Nanjing 210007, China; 15179823980@163.com (T.W.); zzy285518423@hotmail.com (Z.Z.)

² The College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China

* Correspondence: niuyingtao78@hotmail.com

Abstract: To address the small-sample training bottleneck and inadequate convergence efficiency of Deep Reinforcement Learning (DRL)-based communication anti-jamming methods in complex electromagnetic environments, this paper proposes a Generative Adversarial Network-enhanced Deep Q-Network (GA-DQN) anti-jamming method. The method constructs a Generative Adversarial Network (GAN) to learn the time–frequency distribution characteristics of short-period jamming and to generate high-fidelity mixed samples. Furthermore, it screens qualified samples using the Pearson correlation coefficient to form a sample set, which is input into the DQN network model for pre-training to expand the experience replay buffer, effectively improving the convergence speed and decision accuracy of DQN. Our simulation results show that under periodic jamming, compared with the DQN algorithm, this algorithm significantly reduces the number of interference occurrences in the early communication stage and improves the convergence speed, to a certain extent. Under dynamic jamming and intelligent jamming, the algorithm significantly outperforms the DQN, Proximal Policy Optimization (PPO), and Q-learning (QL) algorithms.

Keywords: wireless communication; anti-jamming; generative adversarial network; deep reinforcement learning; Deep Q-Network

1. Introduction

With rapid development in economies and societies, wireless communication applications have become increasingly widespread. However, due to the openness of wireless channels, wireless communication is vulnerable to unintentional and malicious human-made jamming. Therefore, wireless communication anti-jamming technology has been subject to long-term research and extensive application, with a history of more than 50 years so far [1]. Traditional communication anti-jamming technologies, such as spread spectrum anti-jamming technology, have fixed anti-jamming strategies that cannot effectively ensure the reliability and effectiveness of information transmission in the face of complex and diverse jamming.

Early spread spectrum technologies improved anti-jamming capabilities through spectrum expansion, but essentially belonged to a “blind anti-jamming” mechanism without environmental perception, failing when the jamming band exceeded 1/3 [2]. Subsequently, spread spectrum technologies evolved from conventional spread spectrum to adaptive

spread spectrum, including technologies such as frequency/power-adaptive frequency hopping and direct-sequence spread spectrum based on narrowband jamming adaptive suppression/power adaptation [3]. Although these adaptive anti-jamming technologies introduce closed-loop parameter adjustment, they are still limited by the fixed strategies of pre-set rule bases and have difficulty coping with rapidly time-varying jamming threats. This limitation has prompted researchers to explore higher-level anti-jamming intelligent technologies.

In recent years, with the rapid development of artificial intelligence and machine learning, intelligent communication anti-jamming has become a current research hotspot. Artificial intelligence algorithms represented by DRL provide new methods for intelligent anti-jamming technologies by virtue of their iterative learning ability of “perception–decision optimization”. At present, DRL-based communication anti-jamming technologies widely adopt the DQN and its improved methods to solve complex communication anti-jamming problems.

Bailin et al. [4] proposed a dynamic ϵ -DQN intelligent decision algorithm that can optimally select ϵ values according to the state of the decision network, significantly improving the convergence speed and decision success rate. Liang et al. [5] presented a cross-domain anti-jamming algorithm based on the DQN, enabling mobile nodes to learn optimal position adjustment and power control strategies in unknown dynamic-jamming environments to achieve reliable transmission. Boyu et al. [6] proposed a joint frequency–power domain anti-jamming algorithm based on the Prioritized Experience Replay Deep Q-Network, which combines the DQN method with the prioritized replay strategy to realize channel selection and power control in wireless communication systems, improve the fault tolerance of wireless communication systems, and effectively enhance their transmission success rate. Feng et al. [7] modeled the communication anti-jamming process in a limited channel state as a partially observable Markov process and proposed a deep recurrent Q-network algorithm suitable for this scenario, whose convergence performance is significantly better than that of the ordinary DQN algorithm. Yangyang et al. [8] proposed a practical and fast-converging DRL anti-jamming algorithm that uses soft labels instead of rewards to obtain more information entropy, and which does not require random exploration, significantly accelerating the algorithm convergence speed. Yangyang et al. [9] proposed a parallel DRL method, which decomposes complex action spaces through a parallel network architecture, significantly improving the convergence speed and the effectiveness of anti-jamming strategies. Gang et al. [10] proposed a sample information entropy-assisted DRL anti-jamming method, which uses information entropy to predict and finely select training samples for the anti-jamming strategy network so as to improve the quality of training samples and enhance the online decision-making ability and generalization performance of anti-jamming strategies.

In summary, although AI algorithms markedly enhance anti-jamming efficacy by constructing a three-tier “environmental cognition–decision reasoning–waveform reconstruction” architecture, the existing DRL approaches are still confronted by three fundamental challenges in contested communications:

- (1) Small-sample training bottleneck: In complex electromagnetic environments, jamming patterns are typically unknown and scarce, causing conventional DQN to suffer from insufficient samples and slow convergence during training [1].
- (2) Limited convergence efficiency: In high-dimensional action–state spaces, the exploration–exploitation trade-off is easily destabilized and traditional policies require lengthy exploration steps before discovering optimal strategies. This latency risks rendering anti-jamming decisions obsolete before adapting to rapidly evolving jamming.

- (3) Poor transferability: Methods leveraging transfer learning or knowledge graphs perform better when confronting identical or similar jamming styles, whereas the traditional DQN still demands extensive retraining.

To address the above challenges, this paper investigates how Generative Adversarial Networks (GANs) can enhance the convergence speed of DQN anti-jamming algorithms based on the traditional DQN framework. As a generative model, GANs can solve the problem of insufficient samples by generating similar samples with the same distribution as real samples, as has been widely applied in fields such as image processing [11,12], target detection [13,14], and audio generation [15–17]. This paper applies GANs to the field of communication anti-jamming and proposes a small-sample anti-jamming decision method based on a GAN-enhanced Deep Q-Network (GA-DQN). The method uses a GAN to independently learn from collected jamming signals, capture the latent distribution of jamming signals, and generate training samples that approximate the jamming signals distribution to construct a mixed experience replay buffer. Then, a stochastic gradient descent method is adopted to train the policy network, which significantly improves the number and diversity of samples obtained in each algorithm iteration and accelerates the training process of the policy network.

2. Related Work

2.1. The Principle of GANs

GANs were proposed by Goodfellow et al. [18], as a generative model for unsupervised learning. The birth of the GAN idea was inspired by the two-person zero-sum game in game theory, in which two participants fight against each other and improve each other. Compared with traditional generative models, such as autoencoders and variational autoencoders, there are two models in a GAN: the generative model G and the discriminant model D . The generator G aims to generate samples $G(z)$ that follow the real data distribution P_{data} as closely as possible, based on the input noise sequence z , so as to deceive the discriminator D . On the other hand, the discriminator D is used to determine whether the input sample is a real sample x or a generated sample $G(z)$. G and D confront and promote each other. During the continuous confrontation, both will reach a Nash equilibrium. Eventually, G can capture the distribution of x and generate a $G(z)$ that can pass for real, while D , due to its inability to correctly distinguish between x and $G(z)$, will output a value approaching a fixed constant.

The loss functions of the discriminator and the generator are as follows:

$$\max_D V(D, G) = E_{x \sim P_{\text{data}}(x)} [\log(D(x))] + E_{z \sim P_z(z)} [\log(1 - D(G(z)))] \quad (1)$$

$$\max_G V(D, G) = E_{z \sim P_z(z)} [\log(D(G(z)))], \quad (2)$$

where $D(x)$ is the probability that D judges the input x as a real sample; $D(G(z))$ is the probability that the generated sample is judged as a real sample after passing through the discriminator.

2.2. GAN Applied to Data Augmentation

Data augmentation (DA) is an effective method for expanding a dataset scale, enabling models to learn the original data distribution to generate new data with consistent distribution. In recent years, GANs have been extensively applied in data augmentation. By learning real sample distributions from random noise inputs, GANs generate novel sample data, serving as an effective solution for few-sample data augmentation. GANs have been utilized for data augmentation across numerous domains. Zhang et al. [19] proposed a

Dual-Attention Deep Convolutional Generative Adversarial Network (DADCGAN), which incorporates efficient channel attention and external attention mechanisms to capture fundamental feature information from channel and spatial dimensions of images, respectively. Additionally, spectral normalization and dual-timescale update rule strategies are integrated to stabilize the training process. Ye et al. [20] constructed a cycle-consistent GAN for domain transformation, using prior-knowledge-based attention modules, leveraging self-attention and channel attention mechanisms to facilitate conversion, thereby achieving targeted generation of defective insulator samples. Seon et al. [21] introduced a novel GAN-based method for augmenting time series data to address dataset imbalance in IoT applications. Embedding, restoration, and supervised networks are adopted to encapsulate temporal dynamics in the time series data. Furthermore, incorporating mutual information and MSE terms into the loss function while implementing spectral normalization enhances GAN training stability. These diverse improved GAN variants provide a solid theoretical foundation for few-sample data augmentation.

2.3. Methods of Combining GAN with DRL

The performance of Deep Reinforcement Learning (DRL) relies on extensive, high-quality, and diverse interaction data. However, acquiring such data in real-world scenarios often proves costly, inefficient, or even infeasible. For instance, in communication anti-jamming problems, obtaining specific jamming information before encountering jamming is challenging, and DRL exhibits prolonged convergence times in complex jamming environments. The core functionality of a GAN lies in learning from data distributions to generate realistic samples, effectively alleviating data scarcity bottlenecks. Han et al. [22] designed a GAN-based spectrum completion network to reconstruct missing spectral data, which was subsequently used to train DRL-based channel networks, effectively addressing spectrum access issues caused by incomplete sensing information. Strickland et al. [23] employed a GAN to generate synthetic network traffic data, integrating synthetic and real data into DRL models to enhance the overall performance of intrusion detection systems. Wang et al. [24] proposed a GAN-based algorithm to recover traffic data from adjacent intersections, addressing the problem of insufficient traffic data caused by limited communication bandwidth, and they introduced a value-decomposition-based multi-agent DRL algorithm for cooperative traffic-light control; the proposed method significantly improved traffic efficiency. Shi et al. [25] developed a DRL approach that integrates a GAN with DDPG, constructing a DDPG model together with an environment-predictive GAN module to facilitate efficient agent–environment interaction; compared with conventional DRL algorithms, the method achieved substantial performance gains in dormitory environment scenarios. Guo et al. [26] introduced a GAN-augmented Meta-DRL intelligent routing framework for dynamic network environments, employing a learnable GAN feature encoder to generate high-quality rare traffic features that markedly accelerate convergence and improved reward performance. Huang et al. [27] presented a DRL-based risk-averse stochastic dynamic scheduling method to enhance both the robustness and economic efficiency of power system operation in critical risk scenarios. They further proposed KEG-GAN to synthesize additional critical scenario samples, which not only elevated the economic efficiency of power system operation but also reduced the potential high costs associated with critical scenarios.

In summary, GAN-enhanced DRL approaches have been widely adopted across diverse domains, leveraging synthetic data to boost the performance of DRL algorithms. Our work follows this paradigm by integrating GANs into a DQN-based anti-jamming decision framework for communications: we employ the GAN to generate a richer variety

of jamming samples for pre-training the DQN, thereby substantially accelerating and improving the learning process of the agent.

3. System Model

3.1. Communication Model

As shown in Figure 1, the communication scenario consists of a pair of transceivers and an active jammer, where the transmitter and receiver continuously communicate. The intelligent agent at the receiver end perceives the electromagnetic environment in real time to acquire the features of the jamming signals, and the transmitter dynamically selects anti-jamming strategies. When the transmitter sends signals to the receiver, the jammer performs electromagnetic jamming on the receiver's channel access, blocking legitimate communications between the two parties.

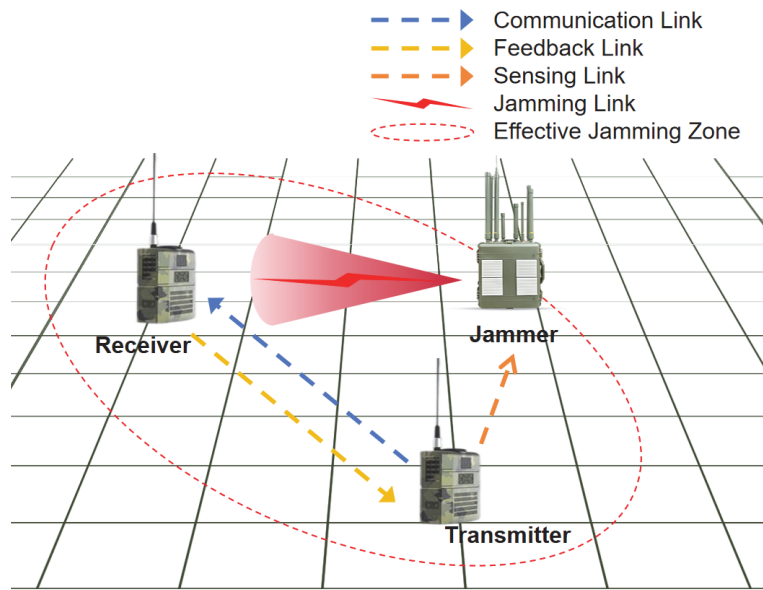


Figure 1. The schematic diagram of the communication confrontation model.

The system uniformly divides the communication frequency band into N continuous non-overlapping channels with a bandwidth of b , and the channel set is $f \in \{0, 1, 2, \dots, N-1\}$. In the discrete time model, the communication time is divided into multiple time slots with a length of T . A time slot is the minimum time unit for continuous data transmission, and the system can communicate within the given wireless communication frequency band by changing the communication channel in the time slot. The receiver determines the jamming degree of this communication according to the Signal-to-Jamming-plus-Noise Ratio (SJNR) of the received signal and evaluates the communication effect, which can be represented by

$$\theta_t = 10 \lg \frac{P_S}{P_S + P_J + P_N}. \quad (3)$$

where P_S is the transmission power of the signal transmitter, P_J is the transmission power of the jammer, and P_N is the radio frequency environmental noise. The SJNR in the communication system needs to satisfy $\theta_t \geq \theta_{th}$ in any time slot T of the time period to ensure successful signal transmission. Here, θ_{th} represents the SJNR threshold.

3.2. Problem Formulation

In this study, we modeled the anti-jamming problem using a Markov Decision Process (MDP). The MDP model involves a set of interacting entities, namely, an agent and an environment. Within a time slot t , the agent observes the current state $\mathbf{s}_t$ from the envi-

ronment, derives an action $\mathbf{a}_t$ based on this current state, and inputs the action into the environment. The environment then determines the next state according to $\mathbf{s}_t$, and $\mathbf{a}_t$ yields a corresponding reward r based on the state value function. The following are the five key elements of the MDP model for solving decision-making problems:

1. State space $\mathbf{S}$ represents the set of all channel jamming states; $\mathbf{s}_t \in \mathbf{S}$ is an element of the environmental state space, denoting the set of jamming states of all channels at time slot t . The actual state of the entire target frequency band can be expressed as $\mathbf{s}_t = [s_t^1, s_t^2, \dots, s_t^N]$.
2. Action space $\mathbf{A}$ is the set of all executable anti-jamming actions $\mathbf{a}_t \in \mathbf{A}$ of the wireless communication system. The transmission power of the communication transmitter is constant. At time slot t , the action of the system is to select one channel from N channels for communication; $\mathbf{a}_t = [a_t^1, a_t^2, \dots, a_t^N]$ represents the transmission channel selected by the transmitter at time slot t .
3. Reward function r represents the return obtained from the environment after the wireless communication system selects and executes the anti-jamming action $\mathbf{a}_t$. This reward function r is used to calculate the immediate reward. The immediate reward is measured by the normalized throughput and is defined as

$$r(\mathbf{s}_t, \mathbf{a}_t) = \delta(\theta_t \geq \theta_{th}), \quad (4)$$

where θ_t is the average SJNR received by the receiver; $\delta(\cdot)$ is an indicator function. The average normalized throughput is used to evaluate the performance of the decision-making algorithm from time slot t_1 to t_2 :

$$r_{t_1, t_2} = \sum_{t_1}^{t_2} r(\mathbf{s}_t, \mathbf{a}_t) / (t_2 - t_1). \quad (5)$$

4. The state transition probability p is the probability $\mathbf{P}(\mathbf{s}_{t+1} | \mathbf{s}_t, \mathbf{a}_t) \rightarrow [0, 1]$ that the intelligent anti-jamming communication system transitions to $\mathbf{s}_{t+1}$ after selecting and executing the action $\mathbf{a}_t$ in the jamming environment state $\mathbf{s}_t$.
5. Policy $\pi(\mathbf{a}_t, \mathbf{s}_t)$ represents the conditional probability distribution of selecting different anti-jamming actions $\mathbf{a}_t$ in the current jamming state $\mathbf{s}_t$.

4. Experimental Principles

4.1. LACGAN-Based Small-Sample Data-Augmentation Method

4.1.1. Auxiliary-Classifer Generative Adversarial Networks

To address the issue of potential multi-type communication jamming in complex electromagnetic environments, this paper proposes a jamming data-augmentation method using a Layer-normalized Auxiliary-Classifer Generative Adversarial Network (LACGAN) to supplement samples that follow the same distribution as real samples. A traditional GAN is unsupervised learning and cannot generate interference samples of specific categories. The LACGAN is a multi-task learning method based on conditional Generative Adversarial Networks. It can utilize the auxiliary classifier in the discriminator to achieve precise control over the category attributes of generated samples. By combining adversarial loss and classification loss, the generator can synthesize realistic samples with specified features.

The LACGAN inputs both the jamming signal x and the label l into G simultaneously. In D , while determining whether the input sample is a real sample through the auxiliary classifier, it also accomplishes the classification task. When a single sample is input into the LACGAN, the discriminator will output two values: one is the adversarial loss L_{adv} that measures the distribution difference between the generated sample and the real sample; the other is the classification loss L_D that distinguishes the category label of the generated

sample from the target jamming type. Similarly, the loss function of the LACGAN consists of the discriminative loss L_{adv} and the classification loss L_D .

The overall loss after summarization is as follows:

$$L_{adv} = E_{x \sim P_{data}(x)}[\log(D_{ac}(x))] + E_{z \sim P_z(z)}[\log(1 - D_{ac}(G(z, l)))] \quad (6)$$

$$L_{cls} = E_{x \sim P_{data}(x)}[\log(L_{cls}(l_x|x))] + E_{z \sim P_z(z)}[\log L_{cls}(l|G(z, l))], \quad (7)$$

where $D_{ac}(x)$ is the probability that D judges the input x as a real sample; $D_{ac}(G(z, y))$ is the probability that the generated sample is judged as a real sample after passing through the discriminator; $L_{cls}(l_x|x)$ represents the classification loss of D for x , and l_x is the real label of x ; and $L_{cls}(y|G(z, l))$ represents the classification loss for the generated sample.

The overall loss after summarization is as follows:

$$\max_D V(D, G) = L_{adv} + L_D \quad (8)$$

$$\max_G V(D, G) = -L_{adv} + L_D. \quad (9)$$

4.1.2. Network Architecture Design

The specific design of the LACGAN network is shown in Figure 2. Based on a traditional GAN, mainly the following improvements have been made:

1. A label-embedding layer is added to the input of the generator. It maps discrete category labels into continuous vectors, which are then concatenated with jamming vectors to form combined inputs. This enables the generator to clearly distinguish the feature distributions of different jamming types in the latent space, achieving precise control over jamming types.
2. Instance normalization (InstanceNorm) is adopted in the discriminator. The discriminator of a traditional GAN only outputs the authenticity probability and uses batch normalization (BatchNorm), which leads to instability in small-batch training. Instance normalization performs normalization on the feature dimension of each sample and does not depend on the batch size, improving the fine-grained classification of samples.
3. The LACGAN introduces learnable scaling factors at the output layer of the generator by setting learnable intensity parameters. It can automatically learn the optimal intensity range for different jamming types through backpropagation, dynamically adjusting the jamming intensity. This enhances the adaptability of the generator in different Signal-to-Jamming-plus-Noise Ratio (SJNR) conditions.

4.1.3. Sample Evaluation of Sample Evaluation Metrics

This paper evaluated and screened the jamming samples generated by a LACGAN using the Pearson correlation coefficient. In the process of improving the success rate of DQN anti-jamming decision-making through data augmentation, if the similarity between the generated signal and the real signal is extremely high and the added training data fails to bring diversity to the recognition model then the augmentation effect exerted by the generated signal will be limited. Conversely, if the similarity between the generated signal and the real signal is too low then it will mislead the DQN model. Not only will it fail to achieve an augmentation effect but it will also reduce the decision-making accuracy of the DQN.

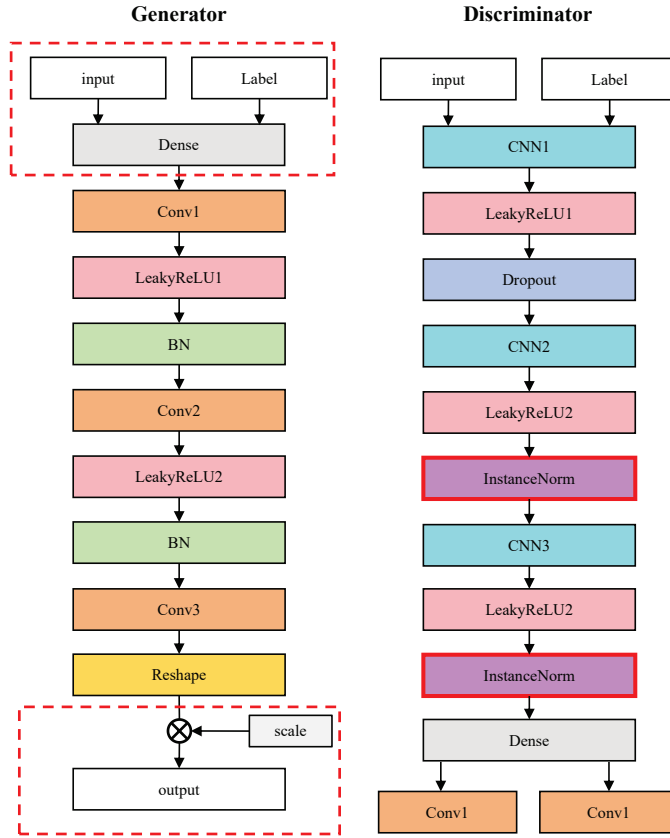


Figure 2. The network architectures of the generator and discriminator in a LACGAN (the red border is the improvement).

This paper applied the Pearson correlation coefficient [28] to evaluate the linear correlation between the original samples and the generated jamming samples. Its value range is $[0, 1]$, where the closer the absolute value is to 1 the higher the similarity between the two. The calculation formula is as follows:

$$r(x_l, y_l) = \frac{\sum_{i=1}^N (x_{l_i} - \bar{x}_l)(y_{l_i} - \bar{y}_l)}{\sqrt{\sum_{i=1}^N (x_{l_i} - \bar{x}_l)^2 \sum_{i=1}^N (y_{l_i} - \bar{y}_l)^2}}. \quad (10)$$

where r represents the Pearson correlation coefficient; x_{l_i} and y_{l_i} are the original sample of category l ; and $\bar{x}_l$ and $\bar{y}_l$ are the mean values of the generated samples of category l .

Finally, according to the calculation results of the Pearson correlation coefficient, the signals in the generated sample set $S_{coarse-grained}$ are sorted in descending order. The five samples with the highest similarity and the five samples with the lowest similarity are selected, and the final generated jamming sample result $S_{fine-grained}$ is obtained.

4.2. GA-DQN Anti-Jamming Decision-Making Method

4.2.1. DQN Algorithm

In the MDP, the goal of the intelligent anti-jamming communication system is to find an optimal policy π^* , such that in any given state the selected action can maximize the long-term cumulative reward. To achieve this goal, a neural network is used to approximate

the state-action value function $Q(\mathbf{s}, \mathbf{a}; \theta)$, where θ represents the network parameters. Thus, it can be expressed using the optimal Q-function as

$$Q^*(s, a) = E^\pi \left[\sum_{\tau=0}^{\infty} \gamma^\tau r(s_{t+\tau}, a_{t+\tau}) | s_t = s, a_t = a \right], \quad (11)$$

where $E^\pi[\cdot]$ is the mathematical expectation operator and τ denotes the subsequent steps starting from time slot t .

To enhance training stability, an independent target Q-network is employed to compute the target Q-value. The parameters θ^- of the target network are periodically copied from the main network, with the update frequency controlled by hyperparameters. The loss function of the DQN is defined as the Mean Squared Error (MSE):

$$L(\theta) = E \left[(r_t + \gamma \max_{\mathbf{a}'} Q(\mathbf{s}_{t+1}, \mathbf{a}'; \theta^-) - Q(\mathbf{s}_t, \mathbf{a}_t; \theta))^2 \right]. \quad (12)$$

During the algorithm training phase, the ε -greedy policy is adopted for training updates. The optimal policy $\pi^*(a|\mathbf{s})$ can then be inferred as follows:

$$\pi^*(a|\mathbf{s}) = \begin{cases} \underset{a \in A}{\operatorname{argmax}} Q^*(\mathbf{s}, \mathbf{a}), & 0 \leq x \leq 1 - \varepsilon \\ \text{random}, & 1 - \varepsilon \leq x \leq 1 \end{cases}. \quad (13)$$

4.2.2. DQN Algorithm

For the anti-jamming decision-making problem in small-sample conditions, firstly, LACGAN is used to complete the data augmentation and concatenation of the small-sample jamming signals. Then, the augmented samples are input into the DQN model as training samples. The entire training process for generating jamming samples is shown in Algorithm 1.

Algorithm 1 GAN Algorithm for Synthetic Sample Generation

```

1: Initialize  $G, D$  with random weights;
2: Initialize Adam optimizers for  $G$  and  $D$ ;
3: Initialize learning rate, epochs  $N$ , batch size  $B$ ;
4: for  $epoch \leftarrow 1$  to  $N$  do
5:   for  $batch \leftarrow 1$  to  $B$  do
6:     Sample real jamming batch  $x_{\text{real}}$ ; sample noise batch  $z$ ;
7:     Generate fake data  $x_{\text{fake}} \leftarrow G(z)$ ;
8:     Train Discriminator  $D$ :
9:       Update  $D$  parameters via Adam to minimize  $\mathcal{L}_D$ ;
10:    Sample new noise batch  $z'$ ; generate new fake data  $x'_{\text{fake}}$ ;
11:    Train Generator  $G$ :
12:      Update  $G$  parameters via Adam to minimize  $\mathcal{L}_G$ ;
13:    end for
14:    Compute Pearson correlation coefficients  $r$ ; concatenate selected samples to form augmented dataset;
15:  end for
16: return Augmented dataset.

```

To accelerate the convergence speed and enhance the decision-making accuracy of the DQN anti-jamming algorithm, this paper proposes a fast anti-jamming algorithm based on a Generative Adversarial DQN. The DQN model in this algorithm adopts the classical conventional architecture [29], as shown in Figure 3, and it incorporates Algorithm 1, enabling the GAN to generate additional experience samples every K time slots. This approach improves the efficiency of sample collection during the training phase of the

algorithm, thereby accelerating the learning process. The specific procedure is detailed in Algorithm 2.

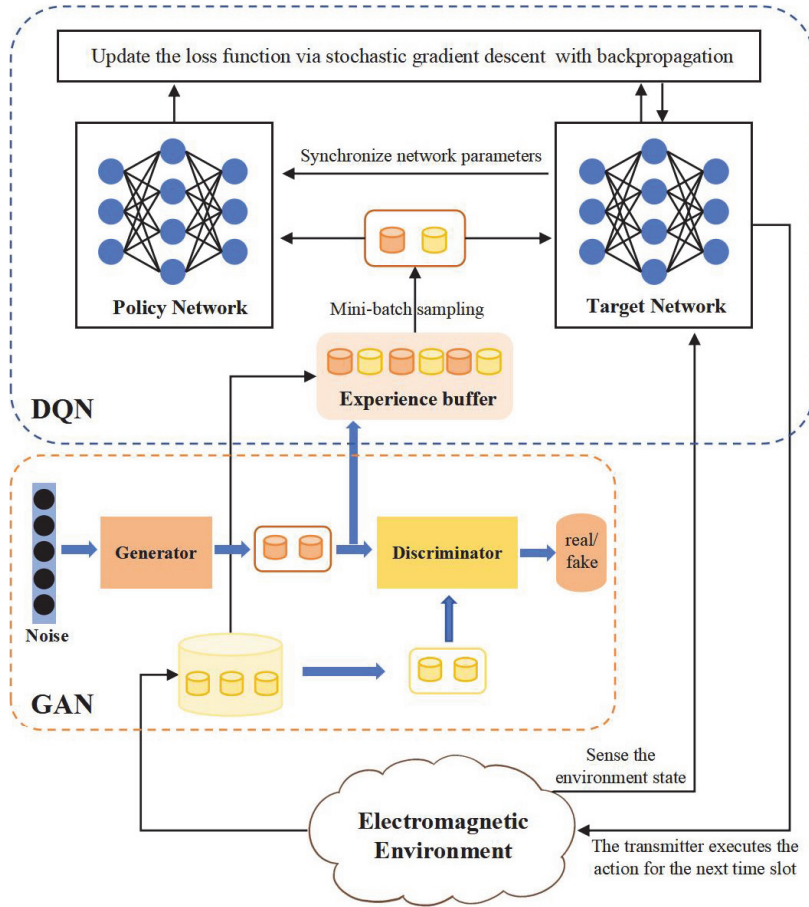


Figure 3. Algorithm architecture.

Algorithm 2 DQN Anti-Jamming Decision-Making Algorithm

Input: Environmental state s_t .

Output: Anti-jamming transmission policy π .

- 1: **Initialization:** $Q, \hat{Q}$ (set $Q = \hat{Q}$), and parameter K .
- 2: **for** $T = 1, 2, \dots, \infty$ **do**
- 3: The transmitter executes the transmission action a_{T+1} according to the action instruction.
- 4: Calculate the transmission action reward $r(s_T, a_{T+1})$.
- 5: Obtain the channel state set s^{T+1} of the next time slot and collect channel states.
- 6: **if** $T \% K = 0$ **then**
- 7: Execute Algorithm 1, generate additional experience samples. After pre-training by the DQN, store them in the experience pool $\mathcal{E}$.
- 8: **end if**
- 9: Batch sample from the experience pool $\mathcal{E}$.
- 10: Input samples into the target network and the policy network. The former outputs $\max \hat{Q}(s', a'; \theta)$, and the latter outputs $Q(s, a; \theta)$.
- 11: Use stochastic gradient descent + backpropagation to update the network parameters θ .
- 12: At the end of the time slot, sense the spectrum to obtain the state s_{T+1} of the next time slot.
- 13: Continue to execute the policy according to the previous time slot.
- 14: **end for**

5. Experimental Results and Analysis

The analysis of our experimental results is mainly divided into two parts. The first part focuses on the performance analysis of jamming sample signals generated by the

LACGAN network, verifying whether a GAN can learn the time–frequency features of jamming signals when applied in communication environments. The second part aims to verify the anti-jamming decision-making effect of a GA-DQN, testing whether the DQN model can improve its decision-making accuracy and convergence speed after pre-training with experience sample signals generated by a GAN.

5.1. Simulation Setup

To support the performance analysis and comparison of small-sample jamming signal data augmentation in this paper, four types of jamming signals were provided for sample generation by a GAN: Multi-Tone Jamming (MTJ), Frequency-Hopping Jamming (FH), Comb-shaped Sweeping Jamming (CSJ), and Linear Sweeping Jamming (LSJ). The specific parameters are shown in Table 1. In the dataset, the jamming signals are transmitted through a random noise channel. The JNR can be selected within the range of -4 dB to 16 dB (with a step size of 2). The shape is set as 12×100 , where the number of channels is 12 and the jamming time slots number 100 .

- **Multi-Tone Jamming:** The energy of this jamming is evenly concentrated on multiple fixed channels. The interval between the jamming channels is 4 channels, and the initial jamming position is channel 2 . It continuously acts on multiple fixed channels within 100 time slots.
- **Frequency-Hopping Jamming:** This type of jamming adopts a pseudo-random hopping mode, performing time-varying frequency point switching within 12 channels. Only 1 channel is activated in each time slot, and the activation position is determined cyclically by the frequency-hopping pattern map.
- **Comb-shaped Sweeping Jamming:** This type of jamming presents a dynamically changing comb-tooth shape in the time–frequency domain. The interference channels are spaced 4 channels apart, starting from channel 2 , and the sweep is at a rate of 1 channel per time slot.
- **Linear Sweeping Jamming:** This jamming exhibits continuous frequency sweeping characteristics. The jamming starts from a random channel number and scans at a constant rate of 1 channel per time slot. It reflects at the spectrum boundary, forming a saw-toothed jamming path.

In the experiment, firstly, the jamming signals are input into the LACGAN for data augmentation. Then, the generated jamming samples are subjected to time-stitching data expansion, input into the DQN model for pre-training, and stored in the experience pool, so as to improve the anti-jamming capability of the DQN.

Table 1. GAN-enhanced DRL models.

Ref	Application	Method	Key Feature
[22]	Dynamic spectrum access	SWGAN	Simulates the spectrum environment for offline training.
[23]	Network intrusion detection	CTGAN	Generates datasets via a GAN to pre-train DRL.
[24]	Traffic signal control	GAN	Reconstructs neighboring traffic data, using traffic statistics.
[25]	Dormitory environment design	GAN	Rapidly predicts indoor performance with a GAN.
[26]	Network routing optimization	GAN	Produces rare traffic features; optimizes the discriminator boundary.
[27]	Power system scheduling	KEG-GAN	Synthesizes additional critical scenario samples.

5.2. Experiment on the Generalization of Network Application

Intelligent anti-jamming methods based on Deep Reinforcement Learning (DRL) determine optimal communication anti-jamming strategies autonomously by learning the patterns and characteristics of detected jamming signals, thus evading jamming signals emitted by jammers. However, a distinctive feature of such methods is that they require acquiring jamming information over sufficient time slots for strategy exploration. In contrast, the intelligent anti-jamming decision-making method based on the GA-DQN proposed in this paper directly leverages a GAN to learn the time–frequency characteristics of sudden jamming signals from short-term attacks. It then generates samples with similar characteristics as the jamming signals. When facing unfamiliar and unknown jamming, this enables the DQN to converge more rapidly. Compared with traditional intelligent anti-jamming methods, it exhibits stronger generalization. The following presents an experimental investigation into the application effects of GANs under different jamming signals, different JNRs, and different iteration counts, to verify its generalization.

Figure 4 shows the process of the simultaneous convergence of the loss functions of the generator and the discriminator. Due to the mutual antagonism between the two networks, in the initial stage of training the oscillation of the generator loss is relatively obvious, reflecting that the learning of the time–frequency characteristics of the jamming samples by the generator of the GAN network is still unstable. From the overall trend, the loss function of the discriminator network shows a decreasing trend, while the loss function of the generator shows an increasing trend, indicating that the discrimination ability of the discriminator is improving. When the number of iterations exceeds 250, both the generator loss and the discriminator loss function tend to be stable, and the network can converge.

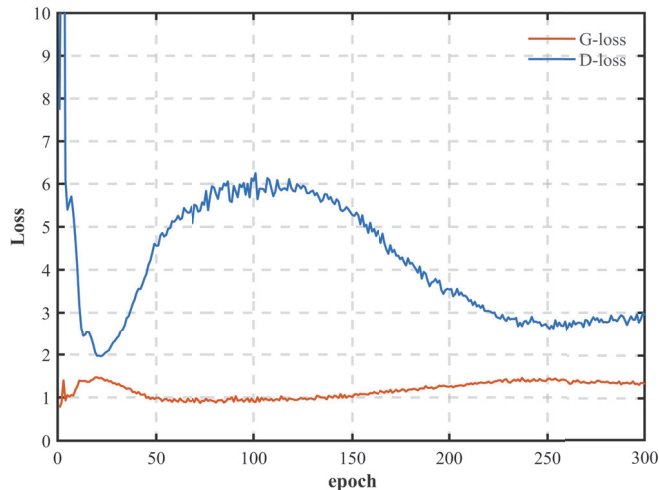


Figure 4. Generator and discriminator loss function variations during GAN training.

In experiment 1, time–frequency feature learning was performed on the CSJ signal, and the signal was input into the GAN network for training. To illustrate that a GAN also has a good generation effect on different types of jamming signals, MTJ, FH, and LSJ were, respectively, selected as the target jamming signals. All the signals were normalized, and the JNR was set to 10 and then input into the GAN for adversarial training.

As can be seen from Figure 5, for different types of jamming signals the network learned their time–frequency features. Among these, the jamming samples generated from the MTI and CSJ signals had the greatest similarity and almost overlapped with the original samples. The jamming samples generated from the LSJ signal possessed the time–frequency characteristics of the original jamming, but the start time slot of the attack

was delayed by 50 time slots. The jamming samples generated from the FH signal had relatively large differences from the original samples, but they conformed to the signal characteristics of Frequency-Hopping Jamming with fast carrier-frequency switching. The GAN could not only generate jamming samples with very high similarity to the original samples, meeting the subsequent needs of improving the decision-making accuracy and convergence speed of the DQN, but it could also generate jamming samples with relatively low similarity but similar jamming strategies. The added training data can bring diversity to the decision-making model and exert a better enhancement effect.

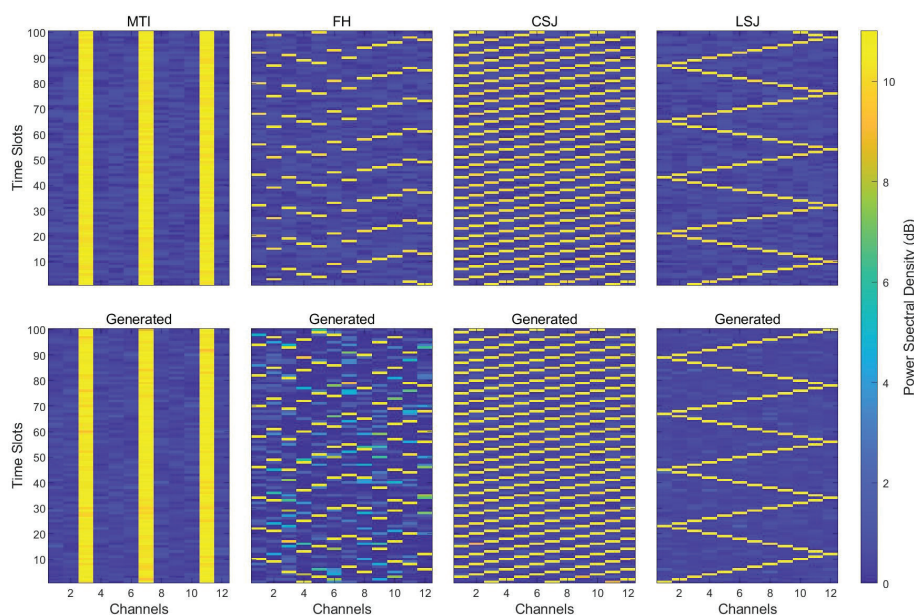


Figure 5. Comparison between generated samples and original samples.

Figure 6 presents the Pearson correlation coefficients between the samples generated under the JNR and iteration counts and the original jamming samples, demonstrating the generalization ability of the LACGAN. It can be seen from the figure that the Pearson correlation coefficient was basically positively correlated with the iteration count and JNR. As the iteration count or the JNR increased, the Pearson correlation coefficient also increased. In subsequent experiments, the JNR was set to 10, and when the iteration count reached 90 the Pearson correlation coefficient had exceeded 0.99. Therefore, the iteration count of the GAN was set to 90 in the subsequent experiments.

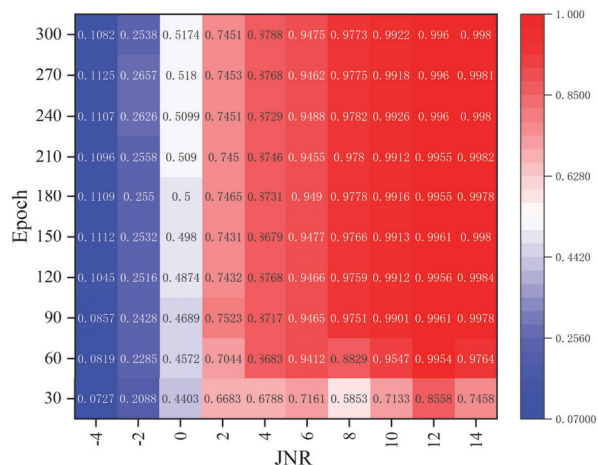


Figure 6. Heatmap of Pearson correlation coefficients.

5.3. Simulation Analysis of GA-DQN

To verify the anti-jamming decision-making capability of the GA-DQN algorithm, this paper compares the GA-DQN algorithm with the conventional DQN algorithm [29], the Proximal Policy Optimization algorithm (PPO) [30], and the Q-learning algorithm [31]. For an objective analysis of the performance of the proposed algorithm, the parameters of the algorithms involved in the comparison were all set to the same values as those of the proposed algorithm. The maximum capacity of the experience replay buffer was 10,000, and the batch processing size was 64. The other simulation parameters were set as shown in Table 2. In addition, the normalized throughput of rewards was adopted as the evaluation index for the anti-jamming decision-making capability of the intelligent algorithm, and the sliding average method (with a window length of 50) was used for curve plotting to obtain smooth curves.

Table 2. Simulation parameter settings.

Parameter	Value
Length of communication time slot T_s	350
Number of channels M	12
Threshold of JSR θ_{th}	8
Discount rate γ	0.3
Communication power p_L/dBm	10
Jamming power p'_L/dBm	10

In the performance comparison simulation, four distinct jamming environments were configured to analyze the anti-jamming performance and robustness of the proposed algorithm, as can be seen from Figure 7:

- Single-jamming environment: The jammer employs only single comb-shaped sweeping jamming to disrupt the receiver;
- Composite-jamming environment: The jammer simultaneously deploys comb-shaped sweeping jamming and multi-channel jamming to challenge the receiver;
- Dynamic-jamming environment: The jammer switches jamming types every 100 time slots in the sequence of CSJ $\rightarrow$ FH $\rightarrow$ MTJ $\rightarrow$ LSJ. This dynamic pattern prevents the DQN from fully exploring and adapting to any single jamming type;
- Intelligent-jamming environment: The jammer updates its target channels every 10 time slots, selecting the channels occupied by the communication signals in the previous 10 slots plus 4 adjacent channels. This adaptive strategy mimics intelligent adversarial behavior.

In neural network training, the learning rate is of great importance to the performance of the neural network. To set appropriate learning rate parameters, this paper analyzed the performance of the DQN algorithm at different learning rates in a single-jamming environment. Figure 8 compares the normalized throughput performance of the algorithm proposed in this paper when the learning rates were 0.01, 0.001, and 0.0001, respectively. It can be seen from the figure that when the learning rate was 0.01, due to the too-fast convergence of the neural network parameters, the DQN calculation was unstable and problems such as gradient explosion occurred; when the learning rate was 0.0001, the convergence speed of the DQN network was slow and the ideal performance was not achieved; when the learning rate was set to 0.001, the DQN converged well and could achieve the ideal performance. Therefore, this learning rate was used in the subsequent simulations.

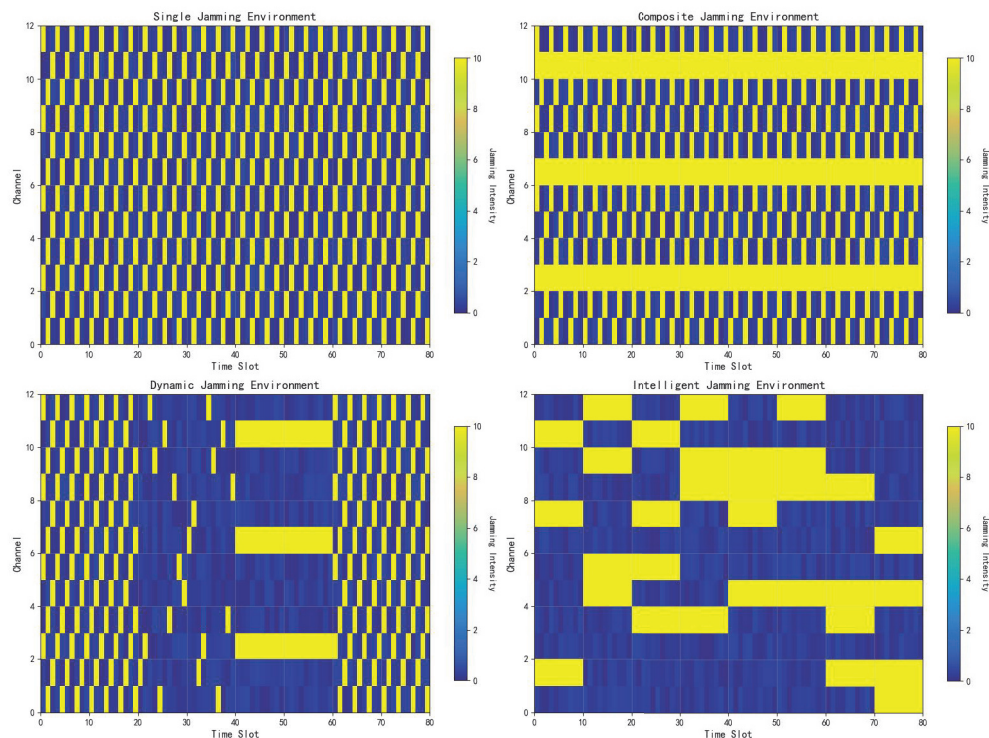


Figure 7. Time–frequency schematic diagram of jamming environment.

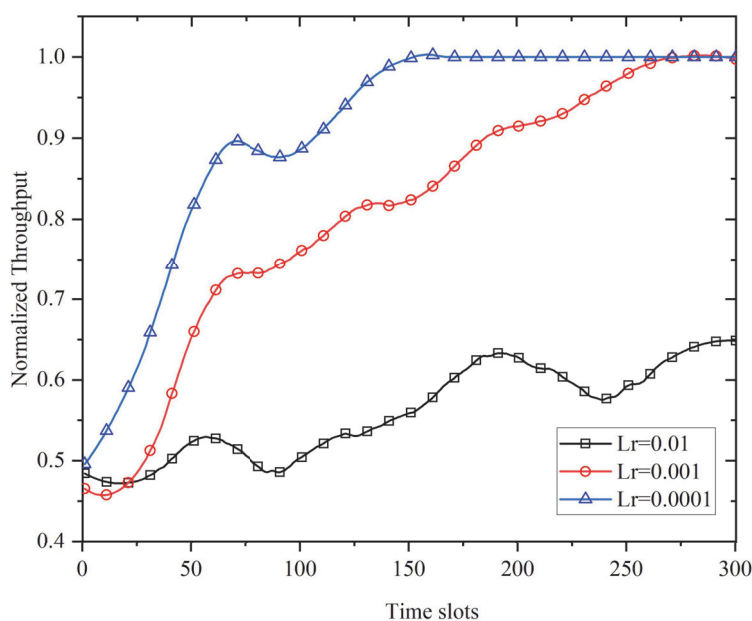


Figure 8. Comparison of DQN algorithm at different learning rates.

In this paper, the experience sample generation cycle of the GA-DQN algorithm was set to k time slots. To achieve optimal performance, it was necessary to validate the impact of different generation cycles K on the algorithm's performance. Figure 9 compares the normalized performance of the proposed algorithm in different generation cycles ($k \in \{0, 10, 20, 30, 40, 50\}$) in a single-jamming environment. When $k = 0$, it corresponded to the traditional DQN algorithm; when $k = 10$, the GAN could not effectively learn the jamming signal characteristics, due to the excessively short generation cycle, which affected the algorithm's convergence; when $k = 30, 40, 50$, the longer generation cycles failed to achieve the optimal performance; when $k = 20$, the GAN could both learn the

jamming signal characteristics and effectively improve the convergence speed, achieving the ideal effect. Therefore, the experience sample generation cycle was set to 20 in the subsequent experiments.

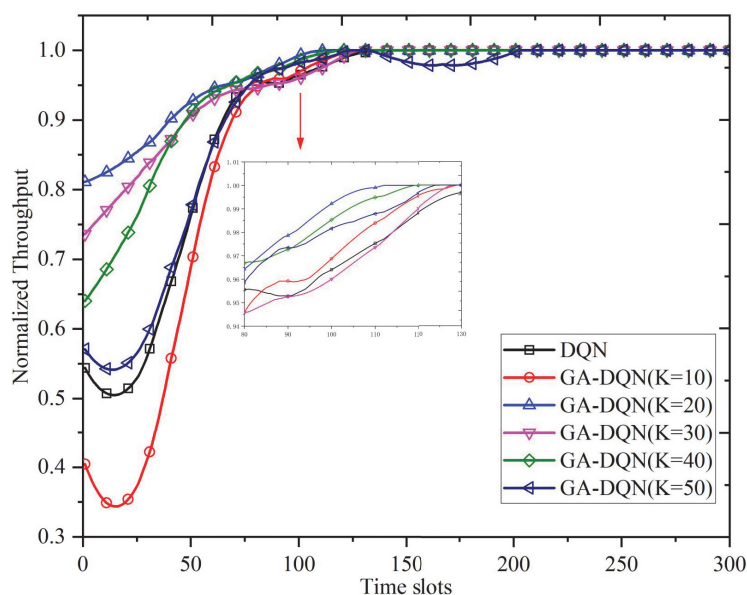
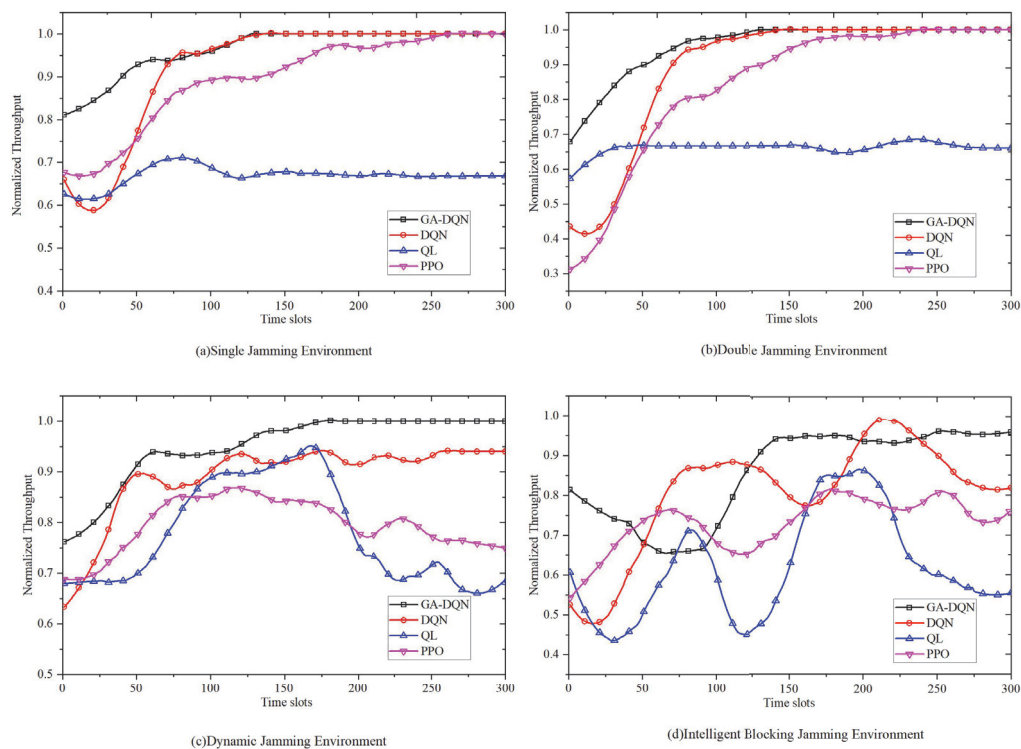


Figure 9. Comparison of the proposed algorithm in different generation cycles.

Figure 10 shows the performance comparison between the GA-DQN algorithm and the three other comparative algorithms in the four types of jamming environments. It can be seen from the figure that the GA-DQN algorithm had the fastest convergence speed and the highest convergence value. In single- and composite-jamming environments, the GA-DQN algorithm outperformed the traditional Deep Reinforcement Learning algorithms such as the DQN and PPO in convergence, with significantly reduced jamming times in the early communication stage; the DQN algorithm converged slightly slower than the GA-DQN but could reach around 1; the PPO algorithm performed slightly worse than the DQN and needed about 275 time slots to converge, while the QL algorithm showed the worst performance and could only converge to about 0.65. In the dynamic-jamming environment, the GA-DQN algorithm could converge at 170 time slots and resist the last fixed multi-channel jamming; the DQN algorithm could only converge to about 0.94, and the PPO and QL algorithms had poor convergence with severe jamming. In the intelligent-jamming environment, the normalized throughput performance of the GA-DQN algorithm decreased at the beginning of communication, started to rise at 100 time slots, and converged to about 0.95 at 130 time slots; due to the intelligent jamming changing every 10 time slots, the other three algorithms did not achieve convergence within 300 time slots. Table 3 aggregates the anti-jamming performance metrics (convergence slots and throughput performance) of our proposed algorithm alongside the three benchmark algorithms across the four distinct jamming environments. The comparative analysis visually demonstrates our algorithm's superior convergence speed and enhanced throughput efficiency relative to the alternatives.

Table 3. The performance index of the anti-jamming algorithm based on comparing the convergence slot (CS) and the throughput performance (TP) of the four types of algorithms.

Model	Single Jamming		Double Jamming		Dynamic Jamming		Intelligent Jamming	
	C S	T P	C S	T P	C S	T P	C S	T P
GA-DQN	126	1	131	1	185	1	245	0.95
DQN	147	1	156	1	256	0.94	280	0.82
PPO	263	1	252	1	280	0.75	300	0.75
QL	149	0.67	211	0.67	300	0.68	300	0.55

**Figure 10.** Comparison of multiple algorithms in different jamming environments.

6. Conclusions

In this paper, we addressed the convergence speed issue of DRL algorithms in complex jamming environments by proposing a GAN-enhanced DQN fast anti-jamming algorithm. The algorithm leverages a GAN to learn the time–frequency characteristics of jamming signals, generate additional experience samples, and input them into the DQN model for pre-training, thereby accelerating the algorithm’s learning process and convergence speed. Furthermore, this paper simulated and compared three other anti-jamming algorithms. The proposed algorithm demonstrated faster convergence, higher convergence values, and more stable communication across diverse jamming environments. The algorithm also holds the potential to be applied to more complex jamming scenarios, effectively improving communication reliability.

Author Contributions: Conceptualization, T.W. and Y.N.; methodology, T.W.; software, T.W. and Z.Z.; validation, T.W. and Z.Z.; formal analysis, T.W.; writing—original draft preparation, T.W.; writing—review and editing, T.W. and Y.N.; project administration, Y.N.; funding acquisition, Y.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (NSFC) under No.62371461.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article, and further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zhou, Q.; Niu, Y. From Adaptive Communication Anti-Jamming to Intelligent Communication Anti-Jamming: 50 Years of Evolution. *Adv. Intell. Syst.* **2024**, *6*, 2300853.
2. Torrieri, D. *Principles of Spread-Spectrum Communication Systems*; Springer: Berlin, Germany, 2018.
3. Fuqiang, Y. *Communication Anti-Jamming Engineering and Practice*, 3rd ed.; Publishing House of Electronics Industry: Beijing, China, 2025.
4. Song, B.; Xu, H.; Jiang, L.; Rao, N. An intelligent decision-making method for anti-jamming communication based on deep reinforcement learning. *J. Northwestern Polytech. Univ.* **2021**, *39*, 641–649.
5. Xiao, L.; Jiang, D.; Xu, D.; Zhu, H.; Zhang, Y.; Poor, H.V. Two-dimensional Anti-jamming Mobile Communication Based on Reinforcement Learning. *IEEE Trans. Veh. Technol.* **2017**, *67*, 9499–9512.
6. Wan, B.; Niu, Y.; Chen, C.; Zhou, Z.; Xiang, P. A novel algorithm of joint frequency–power domain anti-jamming based on PER-DQN. *Neural Comput. Appl.* **2025**, *37*, 7823–7840.
7. Zhang, F.; Niu, Y.; Zhou, Q.; Chen, Q. Intelligent anti-jamming decision algorithm for wireless communication under limited channel state information conditions. *Sci. Rep.* **2025**, *15*, 6271.
8. Li, Y.; Xu, Y.; Li, G.; Gong, Y.; Liu, X.; Wang, H.; Li, W. Dynamic Spectrum Anti-Jamming Access with Fast Convergence: A Labeled Deep Reinforcement Learning Approach. *IEEE Trans. Inf. Forensics Secur.* **2023**, *18*, 5447–5458.
9. Li, Y.; Xu, Y.; Li, W.; Li, G.; Feng, Z.; Liu, S.; Du, J.; Li, X. Achieving Hiding and Smart Anti-Jamming Communication: A Parallel DRL Approach against Moving Reactive Jammer. *IEEE Trans. Commun.* **2025**, *1*, 0090–6778.
10. Li, G.; Wu, Q.; Wang, X.; Luo, H.; Li, L.; Jing, X.; Chen, Q. Deep reinforcement learning-empowered anti-jamming strategy aided by sample information entropy. *J. Commun.* **2024**, *45*, 115–128.
11. Hou, Y.; Zhang, W.; Zhu, Z.; Yu, H. CLIP-GAN: Stacking CLIPs and GAN for Efficient and Controllable Text-to-Image Synthesis. *IEEE Trans. Multimed.* **2025**, *27*, 3702–3715.
12. Gu, B.; Wang, X.; Liu, W.; Wang, Y. MDA-GAN: Multi-dimensional Attention Guided Concurrent-Single-Image-GAN. *Circuits Syst. Signal Process.* **2025**, *44*, 1075–1102.
13. Fang, F.; Zhang, P.; Zhou, B.; Qian, K.; Gan, Y. Atten-GAN: Pedestrian Trajectory Prediction with GAN Based on Attention Mechanism. *Cogn. Comput.* **2022**, *14*, 2296–2305.
14. Kapoor, P.; Arora, S. Optic-GAN: A generalized data augmentation model to enhance the diabetic retinopathy detection. *Int. J. Inf. Technol.* **2025**, *17*, 2251–2269.
15. Airale, L.; Alameda-Pineda, X.; Lathuilière, S.; Vaufraydaz, D. Autoregressive GAN for Semantic Unconditional Head Motion Generation. *ACM Trans. Multimed. Comput. Commun. Appl.* **2025**, *21*, 14.
16. Singh, R.; Sethi, A.; Saini, K.; Saurav, S.; Tiwari, A.; Singh, S. VALD-GAN: Video anomaly detection using latent discriminator augmented GAN. *Signal Image Video Process.* **2024**, *18*, 821–831.
17. Ma, D.; Zhang, F.; Bull, D.R. CVEGAN: A perceptually-inspired GAN for Compressed Video Enhancement. *Signal Process. Image Commun.* **2024**, *127*, 117–127.
18. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the 28th International Conference on Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 2672–2680.
19. Zhang, K.; Yang, X.; Xu, L.; Thé, J.; Tan, Z.; Yu, H. Enhancing coal-gangue object detection using GAN-based data augmentation strategy with dual attention mechanism. *Energy* **2024**, *287*, 129654.
20. Ye, R.; Boukerche, A.; Yu, X.-S.; Zhang, C.; Yan, B.; Zhou, X.-J. Data augmentation method for insulators based on Cycle-GAN. *J. Electron. Sci. Technol.* **2024**, *22*, 36–47.
21. Seon, J.; Lee, S.; Sun, Y.G.; Kim, S.H.; Kim, D.I.; Kim, J.Y. Least Information Spectral GAN with Time-Series Data Augmentation for Industrial IoT. *IEEE Trans. Emerg. Top. Comput. Intell.* **2025**, *9*, 757–769.
22. Han, H.; Wang, X.; Gu, F.; Li, W.; Cai, Y.; Xu, Y.; Xu, Y. Better Late Than Never: GAN-Enhanced Dynamic Anti-Jamming Spectrum Access with Incomplete Sensing Information. *IEEE Wirel. Commun. Lett.* **2021**, *10*, 1800–1804.

23. Strickland, C.; Zakar, M.; Saha, C.; Soltani Nejad, S.; Tasnim, N.; Lizotte, D.J.; Haque, A. DRL-GAN: A Hybrid Approach for Binary and Multiclass Network Intrusion Detection. *Sensors* **2024**, *24*, 2746.
24. Wang, Z.; Zhu, H.; He, M.; Zhou, Y.; Luo, X.; Zhang, N. GAN and Multi-Agent DRL Based Decentralized Traffic Light Signal Control. *IEEE Trans. Veh. Technol.* **2022**, *71*, 1333–1348.
25. Shi, Z.; Huang, C.; Wang, J.; Yu, Z.; Fu, J.; Yao, J. Enhancing performance and generalization in dormitory optimization using deep reinforcement learning with embedded surrogate model. *Build. Environ.* **2025**, *276*, 112864.
26. Guo, Q.; Zhao, W.; Lyu, Z.; Zhao, T. A GAN enhanced meta-deep reinforcement learning approach for DCN routing optimization. *Inf. Fusion* **2025**, *121*, 103160.
27. Huang, W.; Dai, Z.; Hou, J.; Liang, L.; Chen, Y.; Chen, Z.; Pan, Z. Risk-averse stochastic dynamic power dispatch based on deep reinforcement learning with risk-oriented Graph-Gan sampling. *Front. Energy Res.* **2023**, *11*, 1272216.
28. Cai, J.; Xu, K.; Zhu, Y.; Hu, F.; Li, L. Prediction and analysis of net ecosystem carbon exchange based on gradient boosting regression and random forest. *Appl. Energy* **2020**, *262*, 114566.
29. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.; Fidjeland, A.K.; Ostrovski, G.; et al. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533.
30. Chen, Y.; Zhao, Z. Intelligent anti-jamming decision algorithm of bivariate frequency hopping pattern based on ET-PPO. *Telecommun. Sci.* **2022**, *38*, 86–95.
31. Yao, F.; Jia, L. A Collaborative Multi-agent Reinforcement Learning Anti-jamming Algorithm in Wireless Networks. *IEEE Wirel. Commun. Lett.* **2019**, *8*, 1024–1027.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Distributed Interference-Aware Power Optimization for Multi-Task Over-the-Air Federated Learning

Chao Tang, Dashun He and Jianping Yao *

School of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China;
2112303042@mail2.gdut.edu.cn (C.T.); 2112203101@mail2.gdut.edu.cn (D.H.)

* Correspondence: yaojp@gdut.edu.cn

Abstract: Over-the-air federated learning (Air-FL) has emerged as a promising paradigm that integrates communication and learning, which offers significant potential to enhance model training efficiency and optimize communication resource utilization. This paper addresses the challenge of interference management in multi-cell Air-FL systems, focusing on parallel multi-task scenarios where each cell independently executes distinct training tasks. We begin by analyzing the impact of aggregation errors on local model performance within each cell, aiming to minimize the cumulative optimality gap across all cells. To this end, we formulate an optimization framework that jointly optimizes device transmit power and denoising factors. Leveraging the Pareto boundary theory, we design a centralized optimization scheme that characterizes the trade-offs in system performance. Building upon this, we propose a distributed power control optimization scheme based on interference temperature (IT). This approach decomposes the globally coupled problem into locally solvable subproblems, thereby enabling each cell to adjust its transmit power independently using only local channel state information (CSI). To tackle the non-convexity inherent in these subproblems, we first transform them into convex problems and then develop an analytical solution framework grounded in Lagrangian duality theory. Coupled with a dynamic IT update mechanism, our method iteratively approximates the Pareto optimal boundary. The simulation results demonstrate that the proposed scheme outperforms baseline methods in terms of training convergence speed, cross-cell performance balance, and test accuracy. Moreover, it achieves stable convergence within a limited number of iterations, which validates its practicality and effectiveness in multi-task edge intelligence systems.

Keywords: over-the-air federated learning (Air-FL); multi-task; interference management; distributed algorithm

1. Introduction

In the era of ubiquitous edge intelligence, federated learning (FL) has emerged as a pivotal framework for privacy-preserving distributed model training [1,2]. By enabling edge devices to collaboratively train models without sharing raw data, FL addresses concerns related to data privacy and communication overhead [3,4]. However, traditional FL frameworks often rely on digital communication methods, which can be inefficient in resource-constrained wireless environments due to the substantial communication overhead associated with transmitting high-dimensional model updates [5–7].

To mitigate these challenges, over-the-air computation (AirComp) has been integrated with FL, giving rise to over-the-air federated learning (Air-FL) [8,9]. AirComp leverages

the superposition property of wireless multiple-access channels to aggregate data directly in the air, significantly reducing the communication latency and enhancing bandwidth efficiency [10,11]. This synergy between communication and computation positions Air-FL as a promising paradigm for efficient model aggregation in wireless networks [12,13].

1.1. Related Work

While Air-FL offers notable substantial benefits, its deployment in multi-cell wireless networks introduces significant challenges [14]. In particular, concurrent FL tasks among multiple cells can cause severe inter-cell interference, which undermines the quality of aggregated updates and slows down model convergence [15,16]. This issue is especially acute in parallel multi-task scenarios, where each cell tackles different learning tasks with heterogeneous data and distinct model objectives [17,18].

Existing strategies to tackle interference in multi-cell Air-FL primarily involve centralized optimization [19,20]. For example, the authors in [19,20] characterized the Pareto boundary of the error-induced gap region to quantify learning performance trade-offs among different FL tasks, which formulates optimization problems to minimize the sum of error-induced gaps across all cells. Although such centralized solutions achieve near-optimal performance, they suffer from scalability and latency issues in large or dynamic networks, due to the dependence on global channel state information (CSI) and centralized coordination.

To address these concerns, other works have introduced innovations such as hierarchical personalized FL frameworks with optimized beamforming [21], and dynamic clustering combined with power control for two-tier FL systems [22]. Additionally, methods based on compressed sensing [23] have been proposed to enable cross-cell model aggregation with privacy preservation and a reduced need for strict synchronization. The authors in [24] investigated a theoretically guaranteed two-stage search algorithm for joint edge association and aggregation optimization, extended with flexible bandwidth allocation. However, these methods still require some inter-cell coordination.

Despite these advances, centralized methods still face limitations in real-world deployments due to the signaling overhead and latency. This paper addresses that gap by proposing a distributed power control framework guided by interference temperature (IT) parameters, offering near-Pareto optimal performance without requiring centralized coordination. Compared with existing centralized interference management strategies, such as joint power control and beamforming [19,20], our approach introduces several key advancements. First, prior works typically rely on centralized optimization frameworks that require global CSI and full inter-cell coordination, which pose challenges in terms of scalability, latency, and signaling burden. In contrast, our proposed scheme adopts a distributed optimization perspective based on IT, which enables each cell to manage its transmit power using only local CSI and lightweight inter-cell communication. Second, while [19,20] focus on minimizing global error-induced gaps via Pareto boundary characterization under centralized control, our framework introduces a novel dynamic IT update mechanism that allows the system to iteratively approximate Pareto optimality in a decentralized manner. These features make our approach more suitable for large-scale, heterogeneous, and dynamic wireless learning scenarios. A comparison of the focuses, contributions, and potential limitations among highly relevant studies is provided in Table 1.

Table 1. Summary of related work.

Reference	Focuses	Contributions	Limitations
[19]	Achieved efficient downlink and uplink model aggregation in multi-cell Air-FL.	Constructed the Pareto boundary to characterize performance trade-offs among multiple tasks.	Do not fully consider the long-term effect of cumulative aggregation errors on convergence.
[20]	Addressed inter-cell interference in multi-cell using simultaneously transmitting and reflecting reconfigurable intelligent surface (STAR-RIS) assisted Air-FL.	Characterized Pareto-optimal gaps for inter-cell trade-offs and demonstrated mean squared error (MSE) reduction in uplink/downlink via experiments.	Assumed low noise, neglected higher-order errors, and experimented only cover two-cell networks.
[21]	Addressed data heterogeneity in hierarchical FL.	Derived the convergence bound under inter-cluster interference and data heterogeneity.	AF with lower communication overhead was not considered.
[22]	Optimized the learning performance of two-tier Air-FL.	Derived the impact of aggregation errors on convergence performance.	The impact of inter-cluster interference was not considered.
[23]	Addressed the issues of low communication efficiency and weak privacy protection in Air-FL spectrum sharing.	Proposed a compressed sensing-based Air-FL framework to achieve efficient and secure aggregation that is noise-free/encryption-free.	Intra-group nodes require strict synchronization; pseudo-transmitters add redundancy.
[24]	Optimized the joint edge aggregation and association decision-making for Air-FL.	Proposed a theoretically guaranteed two-stage search algorithm, reconstructed the supermodular function, and extended a flexible bandwidth allocation scheme.	The algorithm complexity increases significantly with network scale.

1.2. Contributions

This paper proposes an interference-aware distributed power control scheme tailored for parallel multi-task Air-FL scenarios. By leveraging the local CSI and IT constraints to enable each cell to autonomously manage its transmit power, our approach mitigates inter-cell interference and enhances overall system performance. Through rigorous analysis and extensive simulations, we demonstrate that our proposed scheme achieves near-optimal performance, which offers a practical and scalable solution for interference management in multi-cell Air-FL deployments.

The primary contributions of this work are as follows:

- **System modeling and centralized optimization:** We construct a multi-cell parallel Air-FL system model, detail the AirComp architecture between devices and APs, and establish the interference mechanisms under shared spectrum conditions. We analyze the impact of aggregation errors on the convergence of the optimality gap and formulate a power control optimization problem aimed at minimizing the total optimality gap across all cells. To characterize performance trade-offs, we employ the Pareto boundary theory and design a centralized power control algorithm to delineate the Pareto boundary.
- **Distributed optimization via IT:** We propose a distributed optimization scheme based on IT, which decouples the globally coupled problem into locally solvable subproblems. Each cell independently adjusts its transmit power using local CSI. To address the non-convexity of the subproblems, we first transform them into convex

problems and then develop an analytical solution framework grounded in Lagrangian duality theory and implement a dynamic IT update mechanism to iteratively approach the Pareto boundary.

- **Simulation validation:** Through numerical simulations, we validate the efficacy of our proposed scheme. The results demonstrate that our approach surpasses baseline methods in terms of training convergence speed, cross-cell performance balance, and test accuracy. Moreover, it achieves stable convergence within a limited number of iterations, underscoring its practicality and effectiveness in complex multi-task edge intelligence scenarios.

The remainder of this paper is structured as follows. Section 2 introduces the parallel multi-task Air-FL system model. Section 3 examines the influence of aggregation errors on cell-level local optimality gaps, formulates a power control optimization framework, and presents the concept of the Pareto boundary. Section 4 proposes both a centralized power control optimization scheme and a distributed power control optimization scheme based on IT constraints. Section 5 provides numerical results to validate the performance of our proposed scheme. Section 6 concludes the paper.

2. System Model

As illustrated in Figure 1, we consider a parallel multi-task Air-FL system composed of M cells. Each cell includes a dedicated AP, indexed by $m \in \mathcal{M} \triangleq \{1, \dots, M\}$, and serves K single-antenna edge devices, indexed by $k \in \mathcal{K}_m \triangleq \{1, \dots, K\}$. Define the global device set as $\mathcal{K} \triangleq \{\mathcal{K}_1, \dots, \mathcal{K}_M\}$. Each cell independently collaborates to train a distinct machine learning model, which highlights the heterogeneity across tasks.

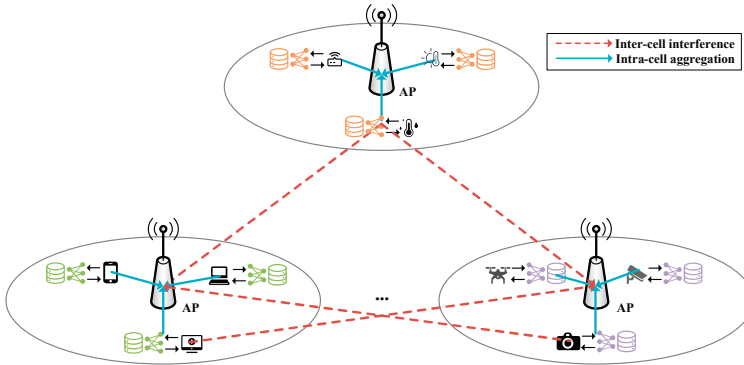


Figure 1. Parallel multi-task Air-FL system.

2.1. Federated Learning Model

Each AP–device cluster forms a standalone FL system, with the m -th cell learning a cell-specific model parameterized by the vector $\mathbf{e}_m = [e_{m,1}, \dots, e_{m,C}]^T$, where C is the model dimension. Each device $k \in \mathcal{K}_m$ holds a private local dataset $\mathcal{D}_k$. The sample loss function evaluates the prediction error of model parameter $\mathbf{e}_m$ on an input-label pair $(\mathbf{x}_i, \tilde{\zeta}_i) \in \mathcal{D}_k$. Accordingly, the local loss function on device k is defined as

$$H_k(\mathbf{e}_m) = \frac{1}{|\mathcal{D}_k|} \sum_{(\mathbf{x}_i, \tilde{\zeta}_i) \in \mathcal{D}_k} H_k(\mathbf{e}_m, \mathbf{x}_i, \tilde{\zeta}_i). \quad (1)$$

Assuming uniform dataset sizes within each cell, i.e., $|\mathcal{D}_j| = |\mathcal{D}_k|, \forall j, k \in \mathcal{K}_m$, the union of all local datasets in cell $m \in \mathcal{M}$ is denoted as $\mathcal{D}_m = \cup_{k \in \mathcal{K}_m} \mathcal{D}_k$. Then, the corresponding global loss function for cell $m \in \mathcal{M}$ is expressed as

$$H_m(\mathbf{e}_m) = \frac{|\mathcal{D}_k|}{|\mathcal{D}_m|} \sum_{k \in \mathcal{K}_m} H_k(\mathbf{e}_m) = \frac{1}{K} \sum_{k \in \mathcal{K}_m} H_k(\mathbf{e}_m). \quad (2)$$

The training goal of cell $m \in \mathcal{M}$ is to find the optimal model parameter $\mathbf{e}_m^*$ that minimizes the global loss function $H_m(\mathbf{e}_m)$, formulated as

$$\mathbf{e}_m^* = \arg \min_{\mathbf{e}_m \in \mathbb{R}^C} H_m(\mathbf{e}_m). \quad (3)$$

2.2. Communication Model

Each cell adopts the federated stochastic gradient descent (FedSGD) for model training, where devices compute and transmit gradients to the AP via AirComp-enabled uplink aggregation. The training is conducted over N global communication rounds, indexed by $n \in \mathcal{N} \triangleq \{1, \dots, N\}$. At the beginning of round $n = 1$, the aggregation center broadcasts a common initialization $\mathbf{e}^{(1)}$ to all APs, which then distribute it to their associated devices, setting their local model parameters $\mathbf{e}_k^{(1)} = \mathbf{e}^{(1)}, \forall k \in \mathcal{K}_m$. We assume that the downlink is error-free.

During each round n , device $k \in \mathcal{K}_m$ computes its local gradient estimate $\mathbf{a}_k^{(n)}$ over a randomly sampled mini-batch $\tilde{\mathcal{D}}_k^{(n)} \subset \mathcal{D}_k$, expressed as

$$\mathbf{a}_k^{(n)} = \frac{1}{|\tilde{\mathcal{D}}_k^{(n)}|} \sum_{x_i, \xi_i \in \tilde{\mathcal{D}}_k^{(n)}} \nabla H_k(\mathbf{e}_k^{(n)}, x_i, \xi_i). \quad (4)$$

To align power levels and mitigate signal distortion, each device applies gradient normalization, defined as

$$\bar{\mathbf{a}}_k^{(n)} = \frac{1}{C} \sum_{c=1}^C a_{k,c}^{(n)}, \quad (5)$$

$$(v_k^{(n)})^2 = \frac{1}{C} \sum_{c=1}^C \left(a_{k,c}^{(n)} - \bar{\mathbf{a}}_k^{(n)} \right)^2, \quad (6)$$

$$s_{k,c}^{(n)} = \frac{a_{k,c}^{(n)} - \bar{\mathbf{a}}_k^{(n)}}{v_k^{(n)}}, \quad (7)$$

where $\bar{\mathbf{a}}_k^{(n)} \in \mathbb{R}$ and $v_k^{(n)} \in \mathbb{R}_+$ denote the mean and standard deviation of the C dimensional local gradient $\mathbf{a}_k^{(n)}$, respectively. $s_{k,c}^{(n)}$ is the normalized signal to be transmitted, satisfying $\mathbb{E}[s_{k,c}^{(n)}] = 0$ and $\mathbb{E}[(s_{k,c}^{(n)})^2] = 1$. Gradients from different devices are assumed to be statistically independent.

Each device transmits its signal over a quasi-static channel. Let $h_{m,k}^{(n)}$ denote the channel from device $k \in \mathcal{K}_m$ to its AP $m \in \mathcal{M}$, and $h_{m,i}^{(n)}$ denote the channel from interfering device $i \in \mathcal{K}_j, j \neq m$ to AP m . The transmission coefficient is set as $b_k^{(n)} = \sqrt{p_k^{(n)} \frac{(h_{m,k}^{(n)})^H}{|h_{m,k}^{(n)}|}}$ with

power $p_k^{(n)} \leq P_k^{max}$, where P_k^{max} is the instantaneous power budget on each device. The signal received at AP $m \in \mathcal{M}$ for the c -th model parameter is expressed as

$$r_{m,c}^{(n)} = \sum_{k \in \mathcal{K}_m} \sqrt{p_k^{(n)}} |h_{m,k}^{(n)}| s_{k,c}^{(n)} + Z_{m,c}^{(n)} + \sum_{j \in \mathcal{M} \setminus \{m\}} \sum_{i \in \mathcal{K}_j} \frac{\sqrt{p_i^{(n)}} h_{m,i}^{(n)} (h_{j,i}^{(n)})^H}{|h_{j,i}^{(n)}|} s_{i,c}^{(n)}, \quad (8)$$

where $Z_{m,c}^{(n)} \sim \mathcal{CN}(0, \sigma_m^2)$ denotes the additive white Gaussian noise (AWGN) with power σ_m^2 at the receiver of AP $m \in \mathcal{M}$.

The denoising factor $\theta_m^{(n)}$ is used to equalize signal power and suppress noise/interference. To reconstruct the aggregated gradient, the AP performs gradient-based linear estimation post-processing on the received signal $r_{m,c}^{(n)}$, given as

$$\begin{aligned} y_{m,c}^{(n)} &= \frac{1}{K} \left(\frac{r_{m,c}^{(n)}}{\sqrt{\theta_m^{(n)}}} + \sum_{k \in \mathcal{K}_m} \bar{a}_k^{(n)} \right) \\ &= \frac{1}{K} \left(\sum_{k \in \mathcal{K}_m} a_{k,c}^{(n)} + \frac{r_{m,c}^{(n)}}{\sqrt{\theta_m^{(n)}}} - \sum_{k \in \mathcal{K}_m} (a_{k,c}^{(n)} - a_k^{(n)}) \right) \\ &= a_{m,c}^{(n)} + \varepsilon_{m,c}^{(n)}, \end{aligned} \quad (9)$$

where $a_{m,c}^{(n)} = \frac{1}{K} \sum_{k \in \mathcal{K}_m} a_{k,c}^{(n)}$ is the desired average gradient; $\varepsilon_{m,c}^{(n)}$ represents the aggregation error due to noise, power mismatch, and inter-cell interference, given as

$$\varepsilon_{m,c}^{(n)} = \frac{1}{K} \left(\sum_{k \in \mathcal{K}_m} \left(\frac{\sqrt{p_k^{(n)}} |h_{m,k}^{(n)}|}{\sqrt{\theta_m^{(n)}}} - v_k^{(n)} \right) s_{k,c}^{(n)} + \sum_{j \in \mathcal{M} \setminus \{m\}} \sum_{i \in \mathcal{K}_j} \frac{\sqrt{p_i^{(n)}} h_{m,i}^{(n)} (h_{j,i}^{(n)})^H}{\sqrt{\theta_m^{(n)}} |h_{j,i}^{(n)}|} s_{i,c}^{(n)} + \frac{Z_{m,c}^{(n)}}{\sqrt{\theta_m^{(n)}}} \right). \quad (10)$$

By collecting all C dimension model parameters, the average global gradient received by AP $m \in \mathcal{M}$ is given by

$$\hat{\mathbf{a}}_m^{(n)} = \Re\{\mathbf{y}_m^{(n)}\} = \left(\mathbf{a}_m^{(n)} + \Re\{\boldsymbol{\varepsilon}_m^{(n)}\} \right), \quad (11)$$

where $\mathbf{a}_m^{(n)} = [a_{m,1}^{(n)}, \dots, a_{m,c}^{(n)}]^T$, $\mathbf{y}_m^{(n)} = [y_{m,1}^{(n)}, \dots, y_{m,c}^{(n)}]^T$, $\boldsymbol{\varepsilon}_m^{(n)} = [\varepsilon_{m,1}^{(n)}, \dots, \varepsilon_{m,c}^{(n)}]^T$.

Finally, the global model of cell $m \in \mathcal{M}$ is updated as

$$\mathbf{e}_m^{(n+1)} = \mathbf{e}_m^{(n)} - \eta_m^{(n)} \hat{\mathbf{a}}_m^{(n)}, \quad (12)$$

where $\eta_m^{(n)}$ is the learning rate. This process continues until convergence or until the maximum number of communication rounds is reached.

3. Convergence Analysis and Problem Formulation

To guide the design of effective power control strategies in the proposed multi-task Air-FL system, we develop a convergence analysis framework that explicitly accounts for the impact of aggregation errors. In this framework, we adopt the optimality gap as a key performance indicator, which quantifies the deviation between the current global loss function value and the corresponding optimal value for each federated task during the n -th communication round.

3.1. Assumptions

As a preliminary step for our convergence analysis, we introduce several foundational assumptions, which are in line with those commonly utilized in prior studies [19,25–29]. These include the smoothness of local loss functions, bounded variance of stochastic gradients, and Polyak–Łojasiewicz condition to ensure the stability of gradient-based updates and allow the analytical derivation of learning dynamics under communication constraints.

Assumption 1 (Lipschitz Smoothness). *For any cell $m \in \mathcal{M}$, the global loss function $H_m(\mathbf{e})$ is assumed to be differentiable and L_m -smooth. That is, there exists a non-negative constant $L_m > 0$ such that for all $\mathbf{e}, \mathbf{v} \in \mathbb{R}^C$, the following condition holds [19,25]:*

$$\|\nabla H_m(\mathbf{e}) - \nabla H_m(\mathbf{v})\| \leq L_m \|\mathbf{e} - \mathbf{v}\|, \quad (13)$$

where $\nabla H_m(\mathbf{e})$ denotes the gradient of the loss function. Equivalently, the following upper bound holds for the loss function:

$$H_m(\mathbf{e}) \leq H_m(\mathbf{v}) + \nabla H_m(\mathbf{v})^T (\mathbf{e} - \mathbf{v}) + \frac{L_m}{2} \|\mathbf{e} - \mathbf{v}\|^2. \quad (14)$$

Assumption 2 (Bounded Variance). *The local gradient estimates $\mathbf{a}_k^{(n)}$ are assumed to be independent and unbiased estimates of the true gradient $\nabla H_k(\mathbf{e}_k)$ and possess bounded variance, i.e., [26,27]*

$$\mathbb{E}[\mathbf{a}_k^{(n)}] = \nabla H_k(\mathbf{e}_k^{(n)}), \forall k \in \mathcal{K}, \forall n \in \mathcal{N}, \quad (15)$$

$$\mathbb{E}\left(\left\|\mathbf{a}_k^{(n)} - \nabla H_k(\mathbf{e}_k^{(n)})\right\|^2\right) \leq \frac{\phi_k^2}{n_b}, \forall k \in \mathcal{K}, \forall n \in \mathcal{N}, \quad (16)$$

where $\phi_k \geq 0$ denotes the variance of stochastic gradients at each device, and n_b denotes the mini-batch size used for gradient computation. (While we assume uniform mini-batch sizes and dataset sizes across devices for analytical tractability, the stochastic gradient variance term ϕ_k captures statistical heterogeneity arising from non-identically distributed (non-IID) local data distributions. Extensions to support heterogeneous dataset sizes and computation budgets can be incorporated by adjusting the weighting factors in the loss function and optimization problem.)

Assumption 3 (Polyak–Łojasiewicz Condition). *Let H_m^* denote the global optimal loss function value for cell m . The Polyak–Łojasiewicz condition asserts that there exists a constant $\mu_m \geq 0$ such that $\nabla H_m(\mathbf{e})$ satisfies [28,29]*

$$\|\nabla H_m(\mathbf{e})\|^2 \geq 2\mu_m (H_m(\mathbf{e}) - H_m^*). \quad (17)$$

3.2. Optimality Gap vs. Aggregation Error

Building upon the preceding assumptions, this subsection investigates the convergence behavior of parallel multi-task Air-FL, with a particular focus on how aggregation errors influence the optimality gap.

Let $\mathbf{e}_m^{(n+1)}$ denote the updated global model parameter for cell $m \in \mathcal{M}$ after global communication round n . The corresponding optimality gap is defined as $\mathbb{E}[H_m(\mathbf{e}_m^{(n+1)})] - H_m^*$. Let $\mathbb{E}\left[\Re\{\boldsymbol{\varepsilon}_{m,c}^{(n)}\}^2\right]$ represent the MSE of the global gradient estimate. With an appropriately designed diminishing learning rate, we establish the following result to characterize the convergence bound for each task.

Theorem 1. Consider a parallel multi-task Air-FL setup where each cell $m \in \mathcal{M}$ adopts a diminishing learning rate of the form $0 \leq \eta_m^{(n)} = \frac{u_m}{n+v_m} \leq \frac{1}{L_m} \leq \frac{1}{\mu_m}, \forall n \in \mathcal{N}$ with constants $v_m > 0, u_m > \frac{1}{\mu_m}$. Suppose that each device employs a fixed mini-batch size $n_b = N$. Under these settings, the expected optimality gap for each cell $m \in \mathcal{M}$ after N communication rounds is bounded by

$$\begin{aligned} \mathbb{E}[H_m(\mathbf{e}_m^{(N+1)})] - H_m^* &\leq \prod_{n \in \mathcal{N}} D_m^{(n)} (\mathbb{E}[H_m(\mathbf{e}_m^1)] - H_m^*) \\ &\quad + \sum_{n \in \mathcal{N}} J_m^{(n)} (\eta_m^{(n)} B_m + L_m (\eta_m^{(n)})^2 \sum_{c=1}^C \mathbb{E}[\Re\{\boldsymbol{\varepsilon}_{m,c}^{(n)}\}^2]), \end{aligned} \quad (18)$$

where $D_m^{(n)} \triangleq 1 - \mu_m \eta_m^{(n)}$ satisfies $0 < D_m^{(n)} < 1, \forall n \in \mathcal{N}$, $J_m^{(n)} \triangleq \frac{\prod_{i=n}^N D_m^{(i)}}{2D_m^{(n)}}$, and the gradient variance is defined as $B_m \triangleq \frac{1}{K} \sum_{k \in \mathcal{K}_m} \frac{\phi_k^2}{N}$.

Remark 1. The proof of Theorem 1 is similar to reference [19]. Please refer to reference [19] for more details. The convergence upper bound of the optimality gap in Theorem 1 consists of three distinct components. The first component is the initial optimality gap $D_m^{(n)} \mathbb{E}[H_m(\mathbf{e}_m^1)] - H_m^*$, determined by the discrepancy between the initial point of the global loss function and its optimal value in cell $m \in \mathcal{M}$. The second component is the gradient variance term $J_m^{(n)} \eta_m^{(n)} B_m$, capturing the influence of stochastic gradient noise arising from statistical heterogeneity within the local datasets of cell m . The third component is the aggregation error term $J_m^{(n)} L_m (\eta_m^{(n)})^2 \sum_{c=1}^C \mathbb{E}[\Re\{\boldsymbol{\varepsilon}_{m,c}^{(n)}\}^2]$, which quantifies the cumulative effect of model aggregation errors induced by channel noise, channel fading, and inter-cell interference. In real-world deployments, the number of communication rounds is finite, and the learning rate cannot be reduced indefinitely. As a result, each cell's optimization process converges to a neighborhood of the global minimum, characterized by a small but non-zero steady-state optimality gap.

3.3. Problem Formulation

In this subsection, building on the convergence analysis from the previous subsection, we derive the optimality gap expression and further formulate a power control optimization problem aimed at minimizing the optimality gap in each cell. From the derived upper bound, it is evident that the initial gap and gradient variance terms remain constant once the system configuration is fixed. Therefore, our attention centers on the aggregation error, which varies with the power control variables and dominates the effective optimality gap to be minimized, formulated as

$$\sum_{n \in \mathcal{N}} \sum_{c=1}^C J_m^{(n)} L_m (\eta_m^{(n)})^2 \mathbb{E}[\Re\{\boldsymbol{\varepsilon}_{m,c}^{(n)}\}^2]. \quad (19)$$

To enable tractable optimization in parallel multi-task scenarios, we decouple the problem into N subproblems, each corresponding to a single communication round. Furthermore, for clarity, we focus on the aggregation error in a single dimension and omit the superscript n without loss of generality. Based on Equation (9), the MSE for the c -th model parameter in cell m is given by

$$\mathbb{E}[\Re\{\boldsymbol{\varepsilon}_{m,c}\}^2] = \frac{1}{K^2} \left(\sum_{k \in \mathcal{K}_m} \left(\frac{\sqrt{p_k} |h_{m,k}|}{\sqrt{\theta_m}} - v_k \right)^2 + \sum_{j \in \mathcal{M} \setminus \{m\}} \sum_{i \in \mathcal{K}_j} \frac{p_i \Re\{h_{m,i}(h_{j,i})^H\}^2}{\theta_m |h_{j,i}|^2} + \frac{\sigma_m^2}{2\theta_m} \right). \quad (20)$$

Substituting this expression into (19), we obtain the following formulation for the effective optimality gap as a function of the power control variables, written as

$$\Phi(\{p_k\}, \{\theta_m\}) \triangleq E_m \left(\sum_{k \in \mathcal{K}_m} \left(\frac{\sqrt{p_k} |h_{m,k}|}{\sqrt{\theta_m}} - v_k \right)^2 + \sum_{j \in \mathcal{M} \setminus \{m\}} \sum_{i \in \mathcal{K}_j} \frac{p_i \Re\{h_{m,i} h_{j,i}^H\}^2}{\theta_m |h_{j,i}|^2} + \frac{\sigma_m^2}{2\theta_m} \right), \quad (21)$$

where $E_m = \frac{J_m L_m \eta_m^2}{K^2}$. Based on this, the optimality gap minimization problem is formulated as

$$(P0) : \min_{\{p_k, \theta_m\}} \{\Phi(\{p_k\}, \{\theta_m\})\} \quad (22)$$

$$\text{s.t. } 0 \leq \theta_m, \quad (23)$$

$$0 \leq p_k \leq P_k^{max}, \forall k \in \mathcal{K}_m, m \in \mathcal{M}. \quad (24)$$

Minimizing the objective function $\Phi(\{p_k\}, \{\theta_m\})$ effectively reduces the optimality gap for each cell. However, in multi-cell wireless systems, independently optimizing each cell may inadvertently amplify inter-cell interference, thus degrading the learning performance of neighboring cells.

To address this challenge and ensure fairness across distributed tasks, it is essential to introduce a theoretical framework that captures the global trade-offs among all cells' learning performances. To this end, the next subsection introduces the concept of the Pareto boundary to characterize the optimal trade-off set of optimality gaps among cells under shared resource constraints, thereby laying a theoretical foundation for subsequent coordinated power control strategies across cells.

3.4. Pareto Boundary Definition and Characterization

To achieve balanced learning performance across different cells, we introduce the concept of the optimality gap region, denoted by $\mathcal{G}$. This region represents the set of all achievable tuples $(\Delta_1, \Delta_2, \dots, \Delta_M)$, where each component Δ_m corresponds to the optimality gap of cell $m \in \mathcal{M}$ under the given system constraints such as per-device power budgets. Formally, the optimality gap region is defined as

$$\mathcal{G} = \bigcup \{(\Delta_1, \Delta_2, \dots, \Delta_M) \mid \Delta_m \geq \text{Gap}_m, \forall m \in \mathcal{M}\}, \quad (25)$$

where $\text{Gap}_m \triangleq J_m L_m \eta_m^2 \mathbb{E}[\Re\{\varepsilon_{m,c}\}^2]$ represents the minimal achievable optimality gap for cell m .

The Pareto boundary of region $\mathcal{G}$ is defined as the set of all Pareto optimal points, for which it is impossible to reduce any component Δ_m without increasing at least one of the others. These points characterize the optimal trade-offs between the learning performance results of different cells under shared resource constraints and provide a theoretical limit to how well system-wide performance can be balanced.

To mathematically characterize the Pareto boundary, we adopt the rate profile method proposed in [30]. This method enables joint optimization across all APs by formulating a scalarized minimization problem, where the individual performance weights are specified by a profile vector $\kappa = [\kappa_1, \kappa_2, \dots, \kappa_M]$. Each weight κ_m reflects the relative importance or performance share assigned to cell m , subject to the normalization constraint $\sum_{m \in \mathcal{M}} \kappa_m = 1$.

The optimization problem for characterizing the Pareto boundary is then given by

$$(P1) : \min_{\{p_k, 0 < \theta_m, 0 < \varsigma\}} \varsigma \quad (26)$$

$$\text{s.t. } \text{Gap}_m \leq \kappa_m \varsigma, \forall m \in \mathcal{M}, \quad (27)$$

$$0 \leq p_k \leq P_k^{\max}, \forall k \in \mathcal{K}_m, m \in \mathcal{M}, \quad (28)$$

where ς serves as an upper bound that balances the individual optimality gaps according to the profile vector κ .

For a given profile vector κ , the solution ς^* to problem (P1) determines a Pareto optimal tuple as $\kappa\varsigma^*$. Geometrically, this point lies on the intersection between the ray defined by direction κ and the Pareto boundary of region $\mathcal{G}$. By varying κ over the unit simplex, the entire Pareto boundary can be traced as illustrated in Figure 2, thereby fully revealing the trade-off structure among competing tasks.

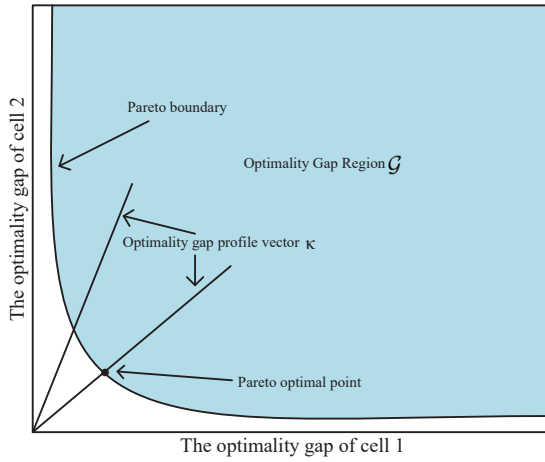


Figure 2. Pareto boundary.

4. Proposed Method

Building upon the previous section, this section develops efficient power control algorithms for parallel multi-task Air-FL from both centralized and distributed perspectives. We begin with a centralized approach that leverages full global CSI to iteratively transform and solve the non-convex optimization problem through convex reformulations. This approach enables accurate characterization of the Pareto boundary via joint optimization across all cells. However, due to the high overhead and poor scalability of centralized solutions in large-scale systems, we then propose a distributed power control scheme based on IT. The distributed approach uses only local CSI and IT constraints, and achieves near-Pareto optimal performance through inter-cell coordination without requiring a global controller.

4.1. Centralized Scheme

The main challenge in solving problem (P1) lies in the coupling between the transmit power p_k of devices and the denoising factor θ_m within the expression of the optimality gap Gap_m , which renders the original problem non-convex. To address this, we propose an iterative centralized algorithm. Specifically, we first fix the transmit powers p_k and optimize the denoising factor θ_m for each cell, then substitute the resulting θ_m back into the original problem to solve for the optimal transmit powers. This alternating optimization (AO) continues until convergence.

4.1.1. Denoising Factor Optimization

Under fixed device transmit power, problem (P1) is decomposed into M independent subproblems, each corresponding to one cell, expressed as

$$(P2) : \min_{\{0 < \theta_m\}} \left\{ \sum_{k \in \mathcal{K}_m} \left(\frac{\sqrt{p_k} |h_{m,k}|}{\sqrt{\theta_m}} - v_k \right)^2 + \sum_{j \in \mathcal{M} \setminus \{m\}} \sum_{i \in \mathcal{K}_j} \frac{p_i \Re\{h_{m,i}(h_{j,i})^H\}^2}{\theta_m |h_{j,i}|^2} + \frac{\sigma_m^2}{2\theta_m} \right\}. \quad (29)$$

By introducing auxiliary variables $\gamma_m = 1/\sqrt{\theta_m}$, the aforementioned problem can be reformulated into a quadratic equation, thereby enabling the straightforward derivation of the closed-form solution for the denoising factor, given as

$$\theta_m = \left(\frac{\sum_{k \in \mathcal{K}_m} p_k |h_{m,k}|^2 + \sum_{j \in \mathcal{M} \setminus \{m\}} \sum_{i \in \mathcal{K}_j} \frac{p_i \Re\{h_{m,i}(h_{j,i})^H\}^2}{|h_{j,i}|^2} + \frac{\sigma_m^2}{2}}{\sum_{k \in \mathcal{K}_m} \sqrt{p_k} |h_{m,k}| v_k} \right)^2. \quad (30)$$

4.1.2. Device Transmit Power Optimization

Substituting the solution for θ_m in Equation (30) back into problem (P1), problem (P1) is rewritten as

$$(P3) : \min_{\{p_k, 0 < \varsigma\}} \varsigma \quad (31)$$

$$\begin{aligned} \text{s.t. } & \left(\sum_{k \in \mathcal{K}_m} v_k^2 - \frac{\kappa_m \varsigma}{E_m} \right) \left(\sum_{k \in \mathcal{K}_m} p_k |h_{m,k}|^2 + \sum_{j \in \mathcal{M} \setminus \{m\}} \sum_{i \in \mathcal{K}_j} \frac{p_i \Re\{h_{m,i}(h_{j,i})^H\}^2}{|h_{j,i}|^2} + \frac{\sigma_m^2}{2} \right) \\ & \leq \left(\sum_{k \in \mathcal{K}_m} \sqrt{p_k} |h_{m,k}| v_k \right)^2, \forall m \in \mathcal{M}, \end{aligned} \quad (32)$$

$$0 \leq p_k \leq P_k^{\max}, \forall k \in \mathcal{K}_m, m \in \mathcal{M}. \quad (33)$$

For any given ς , problem (P3) can be reformulated as

$$(P4) : \text{Find } \{p_k\} \quad (34)$$

$$\begin{aligned} \text{s.t. } & \psi_m \left(\sum_{k \in \mathcal{K}_m} p_k |h_{m,k}|^2 + \sum_{j \in \mathcal{M} \setminus \{m\}} \sum_{i \in \mathcal{K}_j} \frac{p_i \Re\{h_{m,i}(h_{j,i})^H\}^2}{|h_{j,i}|^2} + \frac{\sigma_m^2}{2} \right) \\ & \leq \left(\sum_{k \in \mathcal{K}_m} \sqrt{p_k} |h_{m,k}| v_k \right)^2, \forall m \in \mathcal{M}, \end{aligned} \quad (35)$$

$$0 \leq p_k \leq P_k^{\max}, \forall k \in \mathcal{K}_m, m \in \mathcal{M}, \quad (36)$$

where $\psi_m = \sum_{k \in \mathcal{K}_m} v_k^2 - \frac{\kappa_m \varsigma}{E_m}$. Introduce the interference matrix $\Lambda_m \triangleq [\Lambda_{m,1}, \Lambda_{m,2}, \dots, \Lambda_{m,K_{\text{tot}}}]$ with $K_{\text{tot}} = KM$, defined as

$$\Lambda_{m,k} = \begin{cases} |h_{m,k}|, & k \in \mathcal{K}_m, \\ \frac{\Re\{h_{m,j}(h_{j,i})^H\}}{|h_{j,i}|}, & k \in \mathcal{K}_j, j \in \mathcal{M} \setminus \{m\}. \end{cases} \quad (37)$$

Constraint (35) can be equivalently expressed as

$$\sqrt{\psi_m \left(\sum_{m \in \mathcal{M}} \sum_{k \in \mathcal{K}_m} \Lambda_{m,k}^2 p_k + \frac{\sigma_m^2}{2} \right)} \leq \sum_{k \in \mathcal{K}_m} \sqrt{p_k} |h_{m,k}| v_k. \quad (38)$$

Define $\mathbf{q} = [\sqrt{p_1}, \sqrt{p_2}, \dots, \sqrt{p_{K_{\text{tot}}}}]^T$, $\boldsymbol{\alpha}_m = \text{diag}([\sigma_m^2/2, \Lambda_m])$, $\boldsymbol{\beta}_m = [\beta_{m,1}, \beta_{m,2}, \dots, \beta_{m,K_{\text{tot}}}]^T$ with $\beta_{m,k}$ satisfying:

$$\beta_{m,k} = \begin{cases} |h_{m,k}| v_k, & k \in K_m, \\ 0, & k \in K_j, j \in \mathcal{M} \setminus \{m\}. \end{cases} \quad (39)$$

Therefore, problem (P4) can be transformed into a standard second-order cone program (SOCP) problem, given as

$$(\text{P5}) : \text{Find } \{\mathbf{q}\} \quad (40)$$

$$\text{s.t. } \sqrt{\psi_m} \left\| [1; \mathbf{q}]^T \boldsymbol{\alpha}_m \right\| \leq \mathbf{q}^T \boldsymbol{\beta}_m, \forall m \in \mathcal{M}, \quad (41)$$

$$0 \leq q_k \leq \sqrt{P_k^{\max}}, \forall k \in \mathcal{K}_m, m \in \mathcal{M}. \quad (42)$$

Problem (P5) can be efficiently solved using standard convex optimization solvers (e.g., CVX [31]). Once the optimal solution $\mathbf{q}^*$ is obtained, the transmit powers follow as $p_k^* = (q_k^*)^2$. To determine the optimal ς^* , we employ a bisection search over feasible ς values and solve the corresponding SOCP at each step. The complete process is summarized in Algorithm 1.

Algorithm 1 Centralized scheme for solving problem (P1)

- 1: **Input:** Local gradient variance v_k , profile vector $\boldsymbol{\kappa}$, convergence threshold ι , and maximum power budget $P_k^{\max}$.
 - 2: Set $\varsigma^{\text{low}} \leftarrow 0$, $\varsigma^{\text{up}} \leftarrow \min_{m \in \mathcal{M}} (E_m \sum_{k \in \mathcal{K}_m} v_k^2 / \kappa_m)$.
 - 3: **while** $|\varsigma^{\text{up}} - \varsigma^{\text{low}}| < \iota$ **do**
 - 4: $\varsigma \leftarrow \frac{\varsigma^{\text{low}} + \varsigma^{\text{up}}}{2}$.
 - 5: Solve problem (P5) to obtain p_k^* .
 - 6: **if** problem (P5) is feasible **then**
 - 7: $\varsigma^{\text{up}} \leftarrow \varsigma$.
 - 8: **else**
 - 9: $\varsigma^{\text{low}} \leftarrow \varsigma$.
 - 10: **end if**
 - 11: **end while**
 - 12: Obtain θ_m^* based on (30).
 - 13: **Output:** Optimal solutions $\{p_k^*, \theta_m^*, \varsigma^*\}$.
-

While the centralized scheme achieves globally optimal solutions under full CSI, it requires an aggregation center to collect information from all APs, including intra- and inter-cell channels. This introduces significant signaling overhead and limits the system's scalability in real-world implementations. To address this, the next subsection proposes a distributed power control algorithm based on IT. In this framework, each AP requires only local CSI and interference thresholds, thereby enabling decentralized optimization while still approximating the global Pareto boundary through iterative coordination.

4.2. Distributed Scheme

This subsection investigates the decentralized power control strategy for parallel multi-task systems under IT constraints to characterize the optimality gap and the Pareto boundary.

We define $\Gamma_{j,m}$ as the IT value representing the maximum allowable interference power from devices in cell m to AP j . This constraint ensures that the interference from cell m does not exceed a tolerable threshold at AP j , thereby enabling independent local optimization while maintaining global coordination.

Let $\mathbf{\Gamma}$ be a vector of size $M(M-1) \times 1$ containing all $\Gamma_{j,m}$, and $\mathbf{\Gamma}_m$ be a vector of size $2(M-1) \times 1$ containing both $\Gamma_{j,m}$ and $\Gamma_{m,j}$. For AP $m \in \mathcal{M}$, replace the interference term $\sum_{i \in \mathcal{K}_j} \frac{p_i \Re\{h_{m,i}(h_{j,i})^H\}^2}{|h_{j,i}|^2}$ with $\Gamma_{m,j}$. Define the interference channel from device $k, \forall k \in \mathcal{K}_m$ to neighboring cells as $|g_{j,k}|^2 = \frac{\Re\{h_{j,k}(h_{m,k})^H\}^2}{|h_{m,k}|^2}$. Consequently, the optimality gap minimization problem can be independently addressed at AP $m \in \mathcal{M}$, formulated as

$$(P6.m) : \min_{\{p_k, \theta_m \geq 0\}} E_m \left(\sum_{k \in \mathcal{K}_m} \left(\frac{\sqrt{p_k} |h_{m,k}|}{\sqrt{\theta_m}} - v_k \right)^2 + \sum_{j \in \mathcal{M} \setminus \{m\}} \frac{\Gamma_{m,j}}{\theta_m} + \frac{\sigma_m^2}{2\theta_m} \right) \quad (43)$$

$$\text{s.t. } \sum_{k \in \mathcal{K}_m} p_k |g_{j,k}|^2 \leq \Gamma_{j,m}, j \in \mathcal{M} \setminus \{m\}, \quad (44)$$

$$0 \leq p_k \leq P_k^{\max}, \forall k \in \mathcal{K}_m, m \in \mathcal{M}. \quad (45)$$

Due to the complex coupling relationship between the transmit power variables p_k and the denoising factor θ_m in the objective function of problem (P6.m), the problem exhibits a non-convex structure and is challenging to solve. To address this, we introduce the inverse denoising factor $\gamma_m = 1/\theta_m$ and an auxiliary variable $Q_k = \sqrt{p_k \gamma_m}$, thereby transforming problem (P6.m) as

$$(P7) : \min_{\{Q_k, \gamma_m \geq 0\}} E_m \left(\sum_{k \in \mathcal{K}_m} (Q_k |h_{m,k}| - v_k)^2 + \gamma_m \left(\sum_{j \in \mathcal{M} \setminus \{m\}} \Gamma_{m,j} + \frac{\sigma_m^2}{2} \right) \right) \quad (46)$$

$$\text{s.t. } \sum_{k \in \mathcal{K}_m} |Q_k g_{j,k}|^2 \leq \Gamma_{j,m} \gamma_m, j \in \mathcal{M} \setminus \{m\}, \quad (47)$$

$$Q_k^2 \leq P_k^{\max} \gamma_m, \forall k \in \mathcal{K}_m, m \in \mathcal{M}. \quad (48)$$

Since problem (P7) is a standard convex optimization problem, it can be directly solved using common convex optimization tools (e.g., CVX [31]). To further derive a closed-form expression and enhance theoretical understanding of the problem structure, we employ the Lagrangian dual method for analysis.

Let $\{\lambda_{j,m}\}_{j \in \mathcal{M} \setminus \{m\}} \geq 0$ represent the Lagrange multipliers corresponding to the j -th IT constraint, and φ_k denote the Lagrange multiplier associated with the k -th edge device's power constraint. Thus, the partial Lagrangian function for AP $m \in \mathcal{M}$ is expressed as

$$\begin{aligned} \mathcal{L}_m(Q_k, \gamma_m, \{\lambda_{j,m}\}_{j \in \mathcal{M} \setminus \{m\}}, \varphi_k) = & E_m \left(\sum_{k \in \mathcal{K}_m} (Q_k |h_{m,k}| - v_k)^2 \right. \\ & \left. + \gamma_m \left(\sum_{j \in \mathcal{M} \setminus \{m\}} \Gamma_{m,j} + \frac{\sigma_m^2}{2} \right) \right) + \sum_{k \in \mathcal{K}_m} \varphi_k (Q_k^2 - P_k^{\max} \gamma_m) \\ & + \sum_{j \in \mathcal{M} \setminus \{m\}} \lambda_{j,m} \left(\sum_{k \in \mathcal{K}_m} |Q_k g_{j,k}|^2 - \Gamma_{j,m} \gamma_m \right). \end{aligned} \quad (49)$$

This leads to the dual function

$$\mathcal{R}_m\left(\{\lambda_{j,m}\}_{j \in \mathcal{M} \setminus \{m\}}\right) = \min_{\{Q_k \geq 0, \gamma_m \geq 0\}} \mathcal{L}_m\left(Q_k, \gamma_m, \{\lambda_{j,m}\}_{j \in \mathcal{M} \setminus \{m\}}, \varphi_k\right) \quad (50)$$

$$\text{s.t. } Q_k^2 \leq P_k^{\max} \gamma_m, \forall k \in \mathcal{K}_m. \quad (51)$$

Therefore, the dual problem is formulated as

$$(P8) : \max_{\{\lambda_{j,m} \geq 0, \varphi_k > 0\}} \mathcal{R}_m\left(\{\lambda_{j,m}\}_{j \in \mathcal{M} \setminus \{m\}}, \varphi_k\right). \quad (52)$$

Since problem (P7) satisfies Slater's condition, strong duality holds between its dual problem and the primal problem, which allows us to apply the Karush–Kuhn–Tucker (KKT) optimality conditions [32] to solve the problem. Define the optimal solutions of the dual problem as $\{\lambda_{j,m}^*\}_{j \in \mathcal{M} \setminus \{m\}}$ and φ_k^* . According to the KKT optimality conditions, the variables $\{Q_k^*, \gamma_m^*, \{\lambda_{j,m}^*\}_{j \in \mathcal{M} \setminus \{m\}}, \varphi_k^*\}$ satisfy the following complementary slackness conditions, given as

$$\lambda_{j,m}^* \geq 0, \forall j \in \mathcal{M} \setminus \{m\}, \varphi_k^* \geq 0, \forall k \in \mathcal{K}_m, \quad (53)$$

$$\lambda_{j,m} \left(\sum_{k \in \mathcal{K}_m} |Q_k^* g_{j,k}|^2 - \Gamma_{j,m} \gamma_m^* \right) = 0, \forall j \in \mathcal{M} \setminus \{m\}, \quad (54)$$

$$\varphi_k^* \left((Q_k^*)^2 - P_k^{\max} \gamma_m^* \right) = 0, \forall k \in \mathcal{K}_m, \quad (55)$$

$$\nabla_{Q_k} \mathcal{L}_m \left(Q_k, \gamma_m^*, \{\lambda_{j,m}^*\}_{j \in \mathcal{M} \setminus \{m\}}, \varphi_k^* \right) |_{Q_k=Q_k^*} = 0, \quad (56)$$

$$\nabla_{\gamma_m} \mathcal{L}_m \left(Q_k^*, \gamma_m, \{\lambda_{j,m}^*\}_{j \in \mathcal{M} \setminus \{m\}}, \varphi_k^* \right) |_{\gamma_m=\gamma_m^*} = 0. \quad (57)$$

To determine the optimal auxiliary variable Q_k of each device, we begin by decomposing the Lagrangian dual function into K_m independent subproblems, each corresponding to a device's power optimization task. This decomposition allows for parallel optimization across devices.

The complementary slackness conditions indicate that when $\varphi_k^* > 0$, the power constraint is strictly tight, and the device transmits at its maximum power $P_k^{\max}$. Conversely, if $\varphi_k^* = 0$, the device transmit power satisfies $Q_k^2 \leq P_k^{\max} \gamma_m$. Accordingly, each subproblem is equivalently formulated as

$$(P9) : \min_{\{0 \leq Q_k\}} E_m(Q_k |h_{m,k}| - v_k)^2 + \sum_{j \in \mathcal{M} \setminus \{m\}} \lambda_{j,m} |Q_k g_{j,k}|^2 \quad (58)$$

$$\text{s.t. } Q_k^2 \leq P_k^{\max} \gamma_m. \quad (59)$$

Furthermore, according to the stationarity condition (56) for Q_k , we obtain

$$\begin{aligned} & \nabla_{Q_k} \mathcal{L}_m \left(Q_k, \gamma_m^*, \{\lambda_{j,m}^*\}_{j \in \mathcal{M} \setminus \{m\}}, \varphi_k^* \right) |_{Q_k=Q_k^*} \\ &= 2E_m \left(|h_{m,k}|^2 Q_k - |h_{m,k}| v_k \right) + 2Q_k \sum_{j \in \mathcal{M} \setminus \{m\}} \lambda_{j,m}^* |g_{j,k}|^2 = 0. \end{aligned} \quad (60)$$

By rearranging the equations and considering the maximum device power constraint, the optimal solution for Q_k is derived as

$$Q_k^* = \min \left(\frac{E_m v_k |h_{m,k}|}{E_m |h_{m,k}|^2 + \sum_{j \in \mathcal{M} \setminus \{m\}} \lambda_{j,m} |g_{j,k}|^2}, \sqrt{P_k^{\max} \gamma_m} \right). \quad (61)$$

Similarly, to determine the optimal inverse denoising factor γ_m , we consider the following optimization problem:

$$(P10) : \min_{\{0 \leq \gamma_m\}} E_m \gamma_m \left(\sum_{j \in \mathcal{M} \setminus \{m\}} \Gamma_{m,j} + \frac{\sigma_m^2}{2} \right) - \sum_{j \in \mathcal{M} \setminus \{m\}} \lambda_{j,m} \Gamma_{j,m} \gamma_m - \sum_{k \in \mathcal{K}_m} \varphi_k P_k^{\max} \gamma_m. \quad (62)$$

Applying the stationarity condition (57) with respect to γ_m , we can obtain

$$\begin{aligned} & \nabla_{\gamma_m} \mathcal{L}_m \left(Q_k^*, \gamma_m, \{\lambda_{j,m}^*\}_{j \in \mathcal{M} \setminus \{m\}}, \varphi_k^* \right) |_{\gamma_m = \gamma_m^*} \\ &= E_m \left(\sum_{j \in \mathcal{M} \setminus \{m\}} \Gamma_{m,j} + \frac{\sigma_m^2}{2} \right) - \sum_{j \in \mathcal{M} \setminus \{m\}} \lambda_{j,m}^* \Gamma_{j,m} - \sum_{k \in \mathcal{K}_m} \varphi_k^* P_k^{\max} = 0. \end{aligned} \quad (63)$$

Therefore, the optimal solution for γ_m is derived as

$$\gamma_m^* = \frac{E_m \left(\sum_{j \in \mathcal{M} \setminus \{m\}} \Gamma_{m,j} + \frac{\sigma_m^2}{2} \right)}{\sum_{j \in \mathcal{M} \setminus \{m\}} \lambda_{j,m}^* \Gamma_{j,m} + \sum_{k \in \mathcal{K}_m} \varphi_k^* P_k^{\max}}. \quad (64)$$

To find the optimal dual variables $\{\lambda_{j,m}^*\}_{j \in \mathcal{M} \setminus \{m\}}$ and φ_k^* , we can employ the ellipsoid method [33], which utilizes subgradients to iteratively refine the estimates of the dual variables. The subgradient for $\{\lambda_{j,m}\}_{j \in \mathcal{M} \setminus \{m\}}$ is $\sum_{k \in \mathcal{K}_m} \left| Q_k^* g_{j,k} \right|^2 - \Gamma_{j,m} \gamma_m^*$, and the subgradient for φ_k^* is $(Q_k^*)^2 - P_k^{\max} \gamma_m^*$. By iteratively updating $\{\lambda_{j,m}\}_{j \in \mathcal{M} \setminus \{m\}}$ and φ_k^* using these subgradients, the ellipsoid method converges to the optimal dual variables. Once the optimal dual variables are obtained, these variables can be substituted back into the expressions (61) and (64) to determine the optimal primal variables.

Finally, the optimal transmit power p_k^* and the denoising factor θ_m^* are computed as

$$p_k^* = \min \left(\left(\frac{E_m v_k |h_{m,k}| \theta_m^*}{E_m |h_{m,k}|^2 + \sum_{j \in \mathcal{M} \setminus \{m\}} \lambda_{j,m}^* |g_{j,k}|^2} \right)^2, P_k^{\max} \right), \quad (65)$$

$$\theta_m^* = \frac{\sum_{j \in \mathcal{M} \setminus \{m\}} \lambda_{j,m}^* \Gamma_{j,m} + \sum_{k \in \mathcal{K}_m} \varphi_k^* P_k^{\max}}{E_m \left(\sum_{j \in \mathcal{M} \setminus \{m\}} \Gamma_{m,j} + \frac{\sigma_m^2}{2} \right)}. \quad (66)$$

These results indicate that when $\varphi_k^* > 0$, the device transmits at its maximum power $P_k^{\max}$. Otherwise, the transmit power is adjusted based on the regularized inverse power

transmission strategy, where the regularization term $\sum_{j \in \mathcal{M} \setminus \{m\}} \lambda_{j,m}^* |g_{j,k}|^2$ accounts for inter-cell interference. The denoising factor θ_m^* is influenced by both inter-cell interference and the power constraints of all devices. While these solutions satisfy the inter-cell interference constraints, they are limited to scenarios with fixed interference thresholds and may not achieve Pareto optimality. Therefore, an iterative algorithm for optimizing inter-cell interference thresholds is proposed to enhance overall system performance.

To facilitate efficient distributed power control in parallel multi-task Air-FL systems, this paper introduces the concept of IT as a pivotal coordination variable to establish a local information-driven distributed optimization framework. (The proposed analysis and algorithms assume quasi-static channel conditions, where wireless links remain constant over each communication round. This assumption enables tractable optimization and reliable gradient aggregation using AirComp. In practical scenarios with mobility or fading, channel variations may affect aggregation accuracy. Nonetheless, our distributed scheme relies only on local CSI and permits iterative IT updates, offering inherent adaptability to slow channel variations. Future work may consider extending the framework to incorporate robust optimization techniques for time-varying or uncertain CSI, thereby improving resilience under mobility and fading.) Building upon this foundation, an iterative collaborative algorithm is proposed, wherein APs dynamically adjust mutual IT values through peer-to-peer signaling interactions.

In each iteration, the system selects a pair of APs for local collaborative updates. These updates are designed to ensure that the optimality gap of the local cell does not increase, nor does the performance of other cells degrade, thereby guaranteeing non-decreasing overall system performance. To implement this mechanism, it is assumed that APs possess basic backhaul link support for essential information sharing and synchronization of interference parameters.

However, determining whether the system achieves Pareto optimality under arbitrary IT configurations remains challenging. To address this, a lemma is proposed to characterize the necessary constraints between IT and the cell optimality gap under Pareto optimality conditions. This lemma provides a theoretical foundation for subsequent algorithm design.

Lemma 1 (Necessary Condition for Pareto Optimality). *Under any given IT configuration Γ , if the optimality gap $\overline{\Delta}_m(\Gamma_m)$ has reached a Pareto optimal state, then for any pair of APs m and j , the determinant of the following 2×2 matrix $\mathbf{T}_{j,m}$ must be zero, given as [34]*

$$\mathbf{T}_{j,m} = \begin{bmatrix} \frac{\partial \overline{\Delta}_m(\Gamma_m)}{\partial \Gamma_{j,m}} & \frac{\partial \overline{\Delta}_m(\Gamma_m)}{\partial \Gamma_{m,j}} \\ \frac{\partial \overline{\Delta}_j(\Gamma_j)}{\partial \Gamma_{j,m}} & \frac{\partial \overline{\Delta}_j(\Gamma_j)}{\partial \Gamma_{m,j}} \end{bmatrix}. \quad (67)$$

Building upon the previously established analytical framework, we now delve into the derivation of the matrix elements within the IT optimization context. By solving both the primal and dual formulations of the problem, we obtain explicit expressions for each component of the matrix $\mathbf{T}_{j,m}$, which encapsulates the sensitivity of the optimality gaps to variations in IT parameters.

Specifically, the partial derivatives of the optimality gap $\overline{\Delta}_m(\Gamma_m)$ with respect to the IT parameters $\Gamma_{j,m}$ and $\Gamma_{m,j}$ are given by

$$\frac{\partial \overline{\Delta}_m(\Gamma_m)}{\partial \Gamma_{j,m}} = -\lambda_{j,m}^* \gamma_m^*, \quad \frac{\partial \overline{\Delta}_j(\Gamma_j)}{\partial \Gamma_{m,j}} = -\lambda_{m,j}^* \gamma_j^*, \quad (68)$$

where $\{\lambda_{j,m}^*, \lambda_{m,j}^*\}$ respectively denote the optimal dual multipliers associated with the IT constraints in the primal problems of cell m and cell j , and $\{\gamma_m^*, \gamma_j^*\}$ are the corresponding optimal inverse denoising factors.

Additionally, the derivatives with respect to the reciprocal IT parameters are written as

$$\frac{\partial \bar{\Delta}_m(\Gamma_m)}{\partial \Gamma_{m,j}} = E_m \gamma_m^*, \quad \frac{\partial \bar{\Delta}_j(\Gamma_j)}{\partial \Gamma_{j,m}} = E_j \gamma_j^*. \quad (69)$$

To iteratively refine the IT vector Γ , we focus on updating only the mutual IT parameters between a selected pair of APs m and j , while keeping the rest unchanged, which is defined as

$$[\Gamma'_{j,m}, \Gamma'_{m,j}]^T = [\Gamma_{j,m}, \Gamma_{m,j}]^T + \delta_{j,m} \cdot \mathbf{t}_{j,m}, \quad (70)$$

where Γ' represents the updated IT vector; $\delta_{j,m}$ denotes a suitably small step size; and $\mathbf{t}_{j,m}$ is a vector ensuring that the product $\mathbf{T}_{j,m} \mathbf{t}_{j,m} < 0$, thereby guaranteeing a reduction in the optimality gaps.

Assuming $\mathbf{T}_{j,m} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, a feasible solution for $\mathbf{t}_{j,m}$ is computed as

$$\mathbf{t}_{j,m} = \text{sign}(bc - ad) \cdot [\ell_{j,m}d - b, a - \ell_{j,m}c]^T, \quad (71)$$

where $\ell_{j,m}$ is a ratio parameter that modulates the relative reduction rate of the optimality gap between APs m and j ; $\text{sign}(\cdot)$ denotes the sign function.

By adjusting $\ell_{j,m}$, we can control the balance of improvement between the two APs. Specifically, setting $\ell_{j,m} \geq 1$ prioritizes AP m , while $\ell_{j,m} \leq 1$ favors AP j . Through the careful selection of $\delta_{j,m}$ and $\ell_{j,m}$, we can navigate the Pareto boundary to identify configurations where no AP performance can be improved without degrading the others.

This iterative approach culminates in a distributed power control algorithm that leverages IT optimization to enhance overall system efficiency. The complete algorithm is encapsulated in Algorithm 2, which systematically updates the IT parameters to converge towards Pareto-optimal configurations. (The proposed distributed scheme currently assumes uniform data and computational resources across devices. In practice, edge devices may exhibit heterogeneity in dataset sizes, local data distributions, and processing capabilities. Our framework can accommodate such variations by incorporating per-device weighting or fairness constraints in the optimization objectives. Extending the algorithm to explicitly model non-IID data and device-level heterogeneity is a promising direction for future research).

The distributed algorithm outlined above incrementally minimizes the global optimality gap across all APs through iterative, pairwise updates of IT parameters. (Although our distributed power optimization framework shares a conceptual similarity with some interference-constrained optimization methods (e.g., [35]), there are key distinctions. In [35], the objective is to enhance location privacy and spectrum efficiency in cognitive radio networks through spectrum sharing policies, typically under single-task settings. In contrast, our work addresses a fundamentally different problem of minimizing the learning optimality gap in multi-task Air-FL systems. Furthermore, our framework integrates gradient aggregation distortion, Pareto boundary trade-offs, and an iterative IT update mechanism that collectively enable decentralized learning coordination, which is different from [35].) In each iteration, a selected pair of APs adjusts their mutual IT levels to reduce their respective optimality gaps. Crucially, these adjustments are designed to ensure that the

optimality gaps of other APs in the network remain unaffected, thereby preserving overall system stability.

Algorithm 2 IT-based decentralized scheme for solving problem (P1)

- 1: **Input:** IT vector Γ , convergence threshold ι .
 - 2: **for** any pair of APs m and $j, j \in \mathcal{M} \setminus \{m\}$ **do**
 - 3: **if** $|T_{j,m}| > \iota$ **then**
 - 4: APs m and j synchronously update their current IT values via a reliable backhaul link.
 - 5: APs m and j solve problem (P7) to obtain optimal solution $\{p_k^*\}_{k \in \mathcal{K}_m}, \{p_i^*\}_{i \in \mathcal{K}_j}, \{\theta_m^*, \theta_j^*\}, \{\lambda_{j,m}^*, \lambda_{m,j}^*\}$.
 - 6: APs m and j update the elements in $T_{j,m}$ based on (68) and (69)
 - 7: APs m and j exchange elements in $T_{j,m}$ via the backhaul link, reconstruct $T_{j,m}$ via (67), and compute $t_{j,m}$ via (71).
 - 8: APs m and j update their IT values $\Gamma'_{j,m}$ according to (70).
 - 9: **end if**
 - 10: **end for**
 - 11: **Output:** Optimal device transmit power p_k^* and optimal denoising factor θ_m^* .
-

This iterative update mechanism guides the system's performance toward the Pareto boundary of the optimality gap region. By carefully selecting step sizes and direction vectors for IT adjustments, the algorithm ensures that each update leads to a non-increasing optimality gap for the involved APs. Over successive iterations, this process converges to a state where no further reductions in optimality gaps are possible without adversely affecting other APs, which signifies Pareto optimality.

Compared to the initial configurations or schemes where each AP optimizes independently, this cooperative approach achieves a more efficient global performance. It encourages APs to participate in collaborative adjustments, even when such changes may not immediately benefit their individual performance metrics. This collaborative behavior is facilitated by the algorithm's design, which ensures that any trade-offs made by individual APs contribute to the overall reduction in the system's optimality gap. Such approaches highlight the potential of decentralized coordination mechanisms in enhancing the efficiency and fairness of distributed wireless systems.

5. Numerical Results

In this section, we present the simulation results to evaluate the performance of the proposed distributed scheme in multi-task Air-FL scenarios. Specifically, we consider two representative tasks: ridge regression on synthetic datasets and handwritten digit classification using the Modified National Institute of Standards and Technology (MNIST) dataset. These experiments aim to assess the effectiveness of the distributed scheme in managing interference and optimizing learning performance across multiple tasks. (While MNIST and synthetic regression tasks are used in our simulations for clarity and analytical convenience, the proposed distributed power control framework is compatible with more complex datasets and deep neural models. Our goal in this paper is to highlight the impact of interference-aware power optimization on FL performance under multi-task and multi-cell settings.)

5.1. Simulation Setup and Benchmark Schemes

We simulate a two-cell network, each hosting a distinct FL task. The first cell is centered at coordinates (0,0), and the second at (40,0). Devices within each cell are randomly distributed within a 20-meter radius around their respective APs. The wireless channels

between devices and APs follow a distance-dependent Rayleigh fading model. (In this study, we adopt Rayleigh fading to model the wireless channels between APs and devices. This choice captures a rich-scattering, non-line-of-sight (NLoS) environment and allows for tractable performance evaluation. However, we acknowledge that practical deployments may involve line-of-sight (LoS) components, in which case Rician or Nakagami- m fading models may provide more accurate characterizations. Incorporating such models to study the impact of LoS-dominant fading on Air-FL performance is an important direction for future work.) The channel gain between device k and AP m is modeled as $h_{m,k} = \sqrt{\Omega_0 d_{m,k}^{-\zeta}} h_0$, where Ω_0 denotes the path loss at a reference distance of 1 meter; ζ is the path loss exponent; $d_{m,k}$ is the distance between device k and AP m ; and $h_0 \sim \mathcal{CN}(0, I)$ is a complex Gaussian random variable with zero mean and unit variance. Interference channels between devices and non-associated APs are modeled similarly. Simulation parameters are detailed in Table 2.

Table 2. Simulation parameters.

Parameter	Value
K	10
K_{tot}	20
Ω_0	−60 dB
ζ	3
σ_m^2	10^{-7} W
p_k^{max}	1 W
ι	10^{-9}
$\ell_{j,m}$	0.5
$\eta_1^{(n)}$	$2/(n+8)$
$\eta_2^{(n)}$	$2/(n+10)$

In terms of performance metrics, the optimality gap and prediction error are used to evaluate the ridge regression task on the synthetic dataset. To benchmark the proposed distributed scheme, we compare it against the following baseline methods:

- **Benchmark with maximum power:** In this scheme, all devices transmit at their maximum power levels, i.e., $p_k = p_k^{\text{max}}$. This scheme requires no CSI collection and represents the simplest power control strategy.
- **Benchmark without AirComp:** In this scheme, all devices transmit their local model updates to their respective APs, which perform aggregation without any interference. This scenario assumes an ideal communication environment, serving as an upper performance bound.
- **Benchmark without interference:** In this scheme, each AP optimizes the device transmit power and denoising factor based solely on intra-cell CSI, without coordinating with other APs. The optimization problem for AP m is formulated as

$$\min_{\{0 \leq p_k \leq p_k^{\text{max}}, 0 \leq \theta_m\}} E_m \left(\sum_{k \in \mathcal{K}_m} \left(\frac{\sqrt{p_k} |h_{m,k}|}{\sqrt{\theta_m}} - v_k \right)^2 + \frac{\sigma_m^2}{2\theta_m} \right). \quad (72)$$

5.2. Multi-Task Ridge Regression Performance

We consider that each cell independently trains a ridge regression model on distinct datasets. Each cell's dataset follows a unique distribution, which reflects heterogeneous data environments. The sample loss function for cell $m \in \mathcal{M}$ is defined as

$$H_m(e_m, x_m, \tau_m) = \frac{1}{2} \|x_m^T e_m - \tau_m\|^2 + \rho R(e_m), \quad (73)$$

where $\rho = 5 \times 10^{-5}$ is the regularization hyperparameter. The input sample vector $x_m \in \mathbb{R}^C$ for each cell is drawn from a standard normal distribution, i.e., $x_m \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, with model dimension $C = 20$. The target labels for each cell are generated as $\tau_1 = x_1(2) + 3x_1(5) + 0.2z$ and $\tau_2 = x_2(3) + 2x_2(8) + 4x_2(10) + 0.3z$, where z represents standard Gaussian noise. Each device within a cell holds $D_k = 500, \forall k \in \mathcal{K}$ data samples, resulting in a total of $D_m = \sum_{k \in \mathcal{K}} D_k = 5000$ samples per cell.

To evaluate the convergence behavior, we compute the smoothness parameter L_m and the Polyak–Łojasiewicz parameter μ_m for each cell as in Assumptions 1 and 3, based on the eigenvalues of the sample covariance matrix $X_m^T X_m / D_m + 10^{-4} \mathbf{I}$ [36].

The optimal model parameters e_m^* are obtained as $e_m^* = (X_m^T X_m + \rho \mathbf{I})^{-1} X_m^T \tau_m$ with $\tau_m = [\tau_1, \dots, \tau_{D_m}]^T$. The initial model for each cell is initialized as a zero vector.

Figure 3 presents the evolution of the optimality gap and prediction error across communication rounds N for various power control strategies in both cells. The proposed distributed scheme demonstrates a consistent decline in both metrics as N increases, with the prediction error stabilizing in later rounds. This trend indicates that the local models of both cells progressively converge towards their respective optimal solutions on the training set and ultimately achieve strong generalization performance on the test set. Analyzing the performance disparity between the two cells, Cell 1 consistently exhibits superior outcomes in both optimality gap and prediction error. This advantage is primarily due to the simpler label structure and lower data noise in Cell 1's task, thereby enabling more efficient learning and faster convergence under identical training conditions. Comparative assessments reveal that the proposed distributed scheme significantly outperforms both the benchmark with maximum power and benchmark without interference in both cells. Specifically, the proposed scheme not only accelerates the reduction in training errors to bring the model closer to the optimal training solution but also enhances generalization capability as evidenced by the lower prediction errors on the test set. This performance underscores its effectiveness in balancing rapid convergence with robust generalization. Furthermore, Figure 3 also includes the benchmark without AirComp as an ideal performance upper bound. In later communication rounds, the prediction error achieved by the proposed distributed scheme closely approaches this ideal benchmark with only a marginal gap. This proximity illustrates the scheme's capacity to approximate optimal performance despite practical challenges such as communication losses and inter-cell interference. In contrast, other baseline schemes continue to exhibit significant error at high communication rounds, particularly the benchmark with maximum power, which highlights their limitations in achieving efficient convergence and generalization.

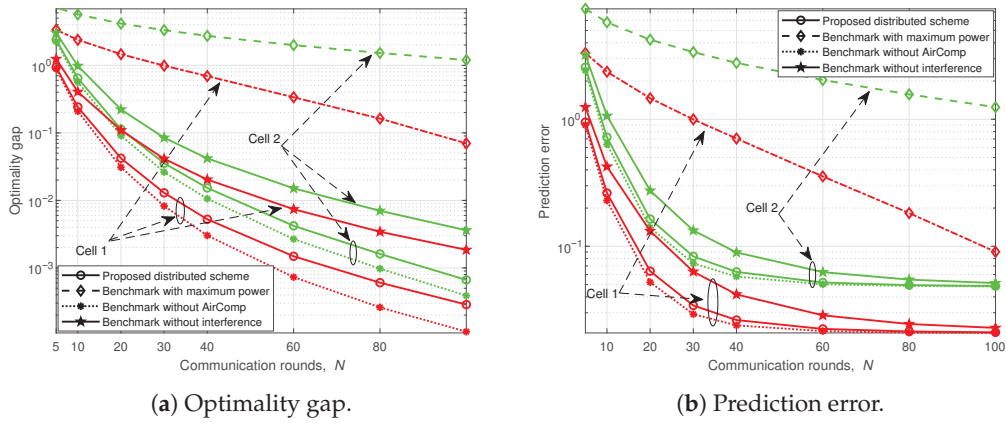


Figure 3. Learning performance on synthetic datasets versus communication rounds.

Figure 4 illustrates the achievable regions of the optimality gap for two cells at communication rounds $N = 100$. The Pareto boundary is delineated by varying the profile vector of the centralized power control scheme as $\kappa = [\kappa_m, 1 - \kappa_m]$, where $\kappa_m = [0.001, 0.01, 0.1, 0.3, 0.4, 0.5, 0.6, 0.7, 0.9, 0.99, 0.999]$. The bottom-leftmost point on the Pareto boundary by the centralized scheme represents the Pareto-optimal point, while the bottom-leftmost point by the benchmark without AirComp denotes the idealized learning performance, which serves as a benchmark for evaluating the proximity of practical schemes to optimal performance. It is observed that the Pareto-optimal point lies closest to this idealized benchmark, indicating that the centralized scheme achieves the best practical performance. Furthermore, the proposed distributed scheme demonstrates a performance remarkably closer to the Pareto-optimal point, with deviations within 10^{-4} , which demonstrates its efficacy in approaching centralized optimality. This proximity underscores its practical applicability, especially in scenarios constrained by communication resources. In contrast, the benchmark without interference exhibits inferior overall performance, which highlights the importance of considering inter-cell dynamics. The benchmark with maximum power performs the worst due to a lack of power adjustment or interference management.

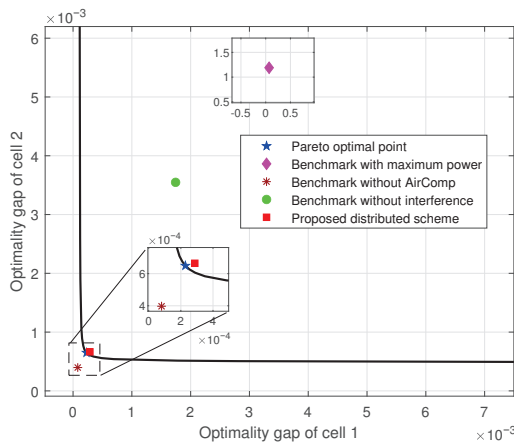


Figure 4. Achievable optimality gap region.

Figure 5 demonstrates the convergence behavior of the proposed distributed scheme by depicting the evolution of the optimality gap for Cell 1 and Cell 2 across iterative rounds with $K = 10$ and $K = 20$ devices per cell. The results demonstrate that all schemes converge within 30 rounds. (While our scheme demonstrates convergence within approximately

30 global rounds, we do not compare directly with FedAvg or FedProx, as these methods operate under fundamentally different assumptions. As such, we evaluate against baselines that are aligned with our system model to ensure a fair and meaningful comparison.) Comparative analysis between the two device configurations reveals that, under the same number of iterations, the scenario with $K = 20$ devices per cell attains a significantly lower optimality gap than the $K = 10$ scenario. This suggests that incorporating more devices into the optimization process enhances the overall system performance, which drives the learning outcomes closer to the optimal solution. Such scalability is crucial for practical applications, where the number of participating devices can vary. While an increased number of devices introduces greater complexity in collaborative computation and power adjustment, potentially leading to a slight reduction in convergence speed, the algorithm maintains stable convergence within a limited number of iterations. This resilience confirms the effectiveness of the proposed distributed scheme in managing complex multi-cell scenarios, thereby ensuring reliable performance even as the network scales.

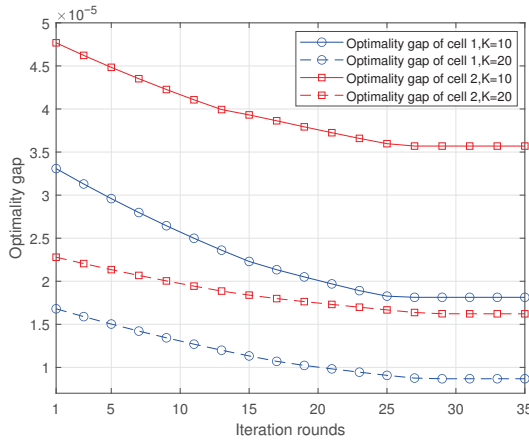


Figure 5. Convergence behavior of the proposed distributed scheme.

5.3. Performance on Multi-Task MNIST Dataset

We establish a parallel multi-task handwritten digit classification scenario utilizing the MNIST dataset. Specifically, Cell 1 and Cell 2 are assigned distinct subsets of digit classes, labeled as 0–4 and 5–9, respectively. This configuration introduces task heterogeneity, as each cell processes non-overlapping label distributions. Both cells employ an identical convolutional neural network (CNN) architecture for local model training. The architecture comprises two convolutional layers with rectified linear unit (ReLU) activation functions and kernel sizes of 5×5 , containing 32 and 64 channels, respectively. Each convolutional layer is followed by a 2×2 max-pooling operation. Subsequently, the network includes a fully connected layer with 1024 units, culminating in a softmax output layer for classification. During local training, all devices utilize a fixed mini-batch size of $m_b = 128$. The performance of the models is evaluated using the average test accuracy and loss function values of both cells to provide a comprehensive assessment of classification efficacy and convergence behavior within the multi-task learning system.

Figure 6 illustrates the impact of communication rounds N on the overall performance of parallel multi-task Air-FL over the MNIST dataset. From Figure 6a, all schemes demonstrate a continuous improvement in average test accuracy as N increases, with convergence observed in the later stages. The benchmark without AirComp serves as the performance upper bound, which achieves optimal accuracy. Notably, the proposed distributed scheme closely approaches this ideal benchmark throughout the training process, with an optimal test accuracy difference within 1%. This proximity demonstrates its near-optimal classi-

fication performance, which surpasses both the benchmark without interference and the benchmark with maximum power, thereby validating its superior convergence speed and learning efficacy. Figure 6b reveals that the average loss function value decreases across all schemes, albeit with varying rates of decline and final convergence levels. The proposed distributed scheme achieves a loss function value closer to the optimal performance of the benchmark without AirComp compared to other baseline schemes. This outcome highlights its capability to effectively enhance the model's generalization performance.

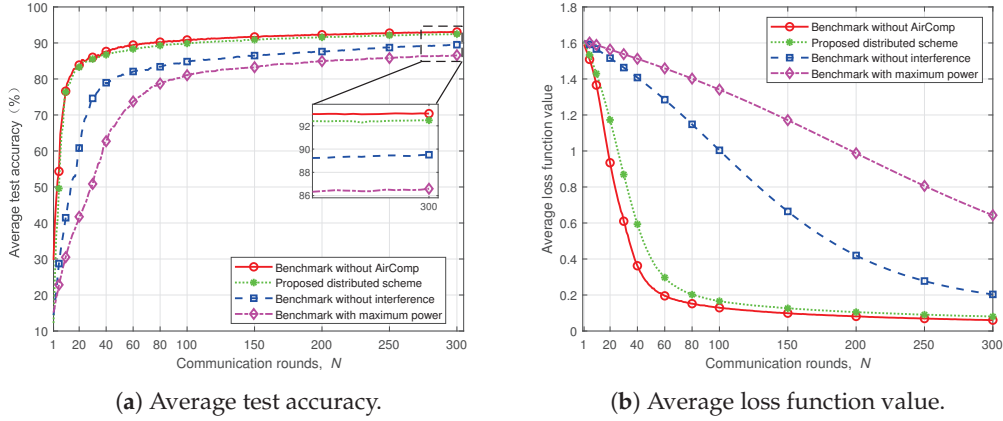


Figure 6. Learning performance on MNIST dataset versus communication rounds.

6. Conclusions

This paper addressed the challenge of inter-cell interference in parallel multi-task Air-FL systems, where each cell independently trains a distinct model while simultaneously performing over-the-air aggregation. We analyzed how aggregation errors impact the local optimality gap and proposed a joint optimization framework for device transmit power and denoising factors aimed at minimizing the cumulative optimality gap across all cells. To characterize the trade-offs in multi-cell performance, we employed Pareto boundary theory and designed a centralized optimization scheme that serves as a performance upper bound. Building upon this, we introduced a distributed power control strategy based on IT, which decouples the globally coupled problem into locally solvable subproblems. This approach allows each cell to independently adjust its transmit power using only local CSI. To enhance computational efficiency, we derived closed-form solutions for the subproblems through Lagrangian duality theory and implemented a dynamic IT update mechanism to effectively approach the Pareto boundary. Simulation results demonstrated that the proposed distributed scheme outperforms benchmark methods in terms of training convergence speed, cross-cell performance balance, and test accuracy. Future work will extend our framework to larger-scale and more heterogeneous datasets such as CIFAR-10, which will enable validation under non-convex learning objectives and more realistic visual data distributions. This will further demonstrate the scalability and robustness of our distributed scheme.

Author Contributions: Conceptualization, C.T. and J.Y.; methodology, C.T. and J.Y.; software, C.T. and D.H.; validation, C.T., J.Y. and D.H.; formal analysis, C.T. and J.Y.; investigation, C.T. and J.Y.; resources, J.Y.; data curation, C.T. and D.H.; writing—original draft preparation, C.T. and J.Y.; writing—review and editing, C.T. and J.Y.; visualization, J.Y. and D.H.; supervision, J.Y.; project administration, J.Y.; funding acquisition, J.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

FL	Federated learning
AirComp	Over-the-air computation
Air-FL	Over-the-air federated learning
IT	Interference temperature
CSI	Channel state information
STAR-RIS	simultaneously transmitting and reflecting reconfigurable intelligent surface
MSE	mean squared error
AO	Alternating optimization
AP	Access point
FedSGD	Federated stochastic gradient descent
AWGN	Additive white Gaussian noise
SOCP	Second-order cone program
KKT	Karush–Kuhn–Tucker
non-IID	non-identically distributed
MNIST	Modified national institute of standard and technology
LoS	line-of-sight
NLoS	non-line-of-sight
CNN	Convolutional neural network
ReLU	Rectified linear unit

References

1. Verbraeken, J.; Wolting, M.; Katzy, J.; Kloppenburg, J.; Verbelen, T.; Rellermeyer, J.S. A survey on distributed machine learning. *ACM Comput. Surv.* **2020**, *53*, 1–33. [CrossRef]
2. Wang, S.; Tuor, T.; Salonidis, T.; Leung, K.K.; Makaya, C.; He, T.; Chan, K. When edge meets learning: Adaptive control for resource-constrained distributed machine learning. In Proceedings of the IEEE Conference on Computer Communications (INFOCOM), Honolulu, HI, USA, 16–19 April 2018; pp. 63–71. [CrossRef]
3. McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; Arcas, B.A. Communication-Efficient Learning of Deep Networks from Decentralized Data. *arXiv* **2017**, arXiv:1602.05629v4.
4. Imteaj, A.; Thakker, U.; Wang, S.; Li, J.; Amini, M.H. A survey on federated learning for resource-constrained IoT devices. *IEEE Internet Things J.* **2022**, *9*, 1–24. [CrossRef]
5. Zhou, C.; Liu, J.; Jia, J.; Zhou, J.; Zhou, Y.; Dai, H.; Dou, D. Efficient device scheduling with multi-job federated learning. In Proceedings of the AAAI Conference on Artificial Intelligence, Online, 22 February–1 March 2022; pp. 9971–9979. [CrossRef]
6. Ma, H.; Guo, H.; Lau, V.K.N. Communication-efficient federated multitask learning over wireless networks. *IEEE Internet Things J.* **2023**, *10*, 609–624. [CrossRef]
7. Sami, H.U.; Güler, B. Over-the-air clustered federated learning. *IEEE Trans. Wirel. Commun.* **2023**, *23*, 7877–7893. [CrossRef]
8. Cao, X.; Başar, T.; Diggavi, S.; Eldar, Y.C.; Letaief, K.B.; Poor, H.V.; Zhang, J. Communication-efficient distributed learning: An overview. *IEEE J. Sel. Areas Commun.* **2023**, *41*, 851–873. [CrossRef]
9. Cao, X.; Lyu, Z.; Zhu, G.; Xu, J.; Xu, L.; Cui, S. An overview on over-the-air federated edge learning. *IEEE Wirel. Commun.* **2024**, *31*, 202–210. [CrossRef]
10. Wang, Z.; Zhao, Y.; Zhou, Y.; Shi, Y.; Jiang, C.; Letaief, K.B. Over-the-air computation for 6G: Foundations, technologies, and applications. *IEEE Internet Things J.* **2024**, *11*, 24634–24658. [CrossRef]
11. Şahin, A.; Yang, R. A survey on over-the-air computation. *IEEE Commun. Surv. Tuts.* **2023**, *25*, 1877–1908. [CrossRef]
12. Zhu, J.; Shi, Y.; Zhou, Y.; Jiang, C.; Chen, W.; Letaief, K.B. Over-the-air federated learning and optimization. *IEEE Internet Things J.* **2024**, *11*, 16996–17020. [CrossRef]
13. Azimi-Abarghouyi, S.M.; Fodor, V. Scalable hierarchical over-the-air federated learning. *IEEE Trans. Wirel. Commun.* **2024**, *23*, 8480–8496. [CrossRef]
14. Aygün, O.; Kazemi, M.; Gündüz, D.; Duman, T.M. Over-the-air federated edge learning with hierarchical clustering. *IEEE Trans. Wirel. Commun.* **2024**, *23*, 17856–17871. [CrossRef]
15. Asaad, S.; Wang, P.; Tabassum, H. Over-the-air FEEL with integrated sensing: Joint scheduling and beamforming design. *IEEE Trans. Wirel. Commun.* **2025**, *24*, 3273–3288. [CrossRef]

16. Liang, Y.; Chen, Q.; Zhu, G.; Jiang, H.; Eldar, Y.C.; Cui, S. Communication-and-energy efficient over-the-air federated learning. *IEEE Trans. Wirel. Commun.* **2025**, *24*, 767–782. [CrossRef]
17. Zhong, C.; Yang, H.; Yuan, X. Over-the-air federated multi-task learning over MIMO multiple access channels. *IEEE Trans. Wirel. Commun.* **2023**, *22*, 3853–3868. [CrossRef]
18. Li, F.; Ye, Q.; Fapi, E.T.; Sun, W.; Jiang, Y. Multi-cell over-the-air computation systems with spectrum sharing: A perspective from α -fairness. *IEEE Trans. Veh. Technol.* **2023**, *72*, 16249–16265. [CrossRef]
19. Wang, Z.; Zhou, Y.; Shi, Y.; Zhuang, W. Interference management for over-the-air federated learning in multi-cell wireless networks. *IEEE J. Sel. Areas Commun.* **2022**, *40*, 2361–2377. [CrossRef]
20. Zeng, X.; Mao, Y.; Shi, Y. STAR-RIS assisted over-the-air vertical federated learning in multi-cell wireless networks. In Proceedings of the IEEE International Conference on Communications Workshops (ICC Wkshps), Rome, Italy, 28 May–1 June 2023; pp. 361–366. [CrossRef]
21. Zhou, F.; Wang, Z.; Shan, H.; Wu, L.; Tian, X.; Shi, Y.; Zhou, Y. Over-the-air hierarchical personalized federated learning. *IEEE Trans. Veh. Technol.* **2025**, *74*, 5006–5021. [CrossRef]
22. Guo, W.; Huang, C.; Qin, X.; Yang, L.; Zhang, W. Dynamic clustering and power control for two-tier wireless federated learning. *IEEE Trans. Wirel. Commun.* **2024**, *23*, 1356–1371. [CrossRef]
23. Li, W.; Chen, G.; Zhang, X.; Wang, N.; Ouyang, D.; Chen, C. Efficient and secure aggregation framework for federated-learning-based spectrum sharing. *IEEE Internet Things J.* **2024**, *11*, 17223–17236. [CrossRef]
24. Wu, T.; Qu, Y.; Liu, C.; Dai, H.; Dong, C.; Cao, J. Cost-efficient federated learning for edge intelligence in multi-cell networks. *IEEE/ACM Trans. Netw.* **2024**, *32*, 4472–4487. [CrossRef]
25. Li, X.; Huang, K.; Yang, W.; Wang, S.; Zhang, Z. On the Convergence of Fedavg on Non-Iid Data. *arXiv* **2019**, arXiv:1907.02189v4.
26. Bottou, L.; Curtis, F.E.; Nocedal, J. Optimization methods for large-scale machine learning. *SIAM Rev.* **2018**, *60*, 223–311. [CrossRef]
27. Feng, C.; Yang, H.H.; Hu, D.; Zhao, Z.; Quek, T.Q.S.; Min, G. Mobility-aware cluster federated learning in hierarchical wireless networks. *IEEE Trans. Wirel. Commun.* **2022**, *21*, 8441–8458. [CrossRef]
28. Wang, S.; Tuor, T.; Salonidis, T.; Leung, K.K.; Makaya, C.; He, T.; Chan, K. Adaptive federated learning in resource constrained edge computing systems. *IEEE J. Sel. Areas Commun.* **2019**, *37*, 1205–1221. [CrossRef]
29. Cao, X.; Zhu, G.; Xu, J.; Wang, Z.; Cui, S. Optimized power control design for over-the-air federated edge learning. *IEEE J. Sel. Areas Commun.* **2022**, *40*, 342–358. [CrossRef]
30. Lan, Q.; Kang, H.S.; Huang, K. Simultaneous signal-and-interference alignment for two-cell over-the-air computation. *IEEE Wirel. Commun. Lett.* **2020**, *9*, 1342–1345. [CrossRef]
31. Grant, M.; Boyd, S. CVX: MATLAB Software for Disciplined Convex Programming. 2016. Available online: <http://cvxr.com/cvx> (accessed on 3 July 2025).
32. Boyd, S.P.; Vandenberghe, L. Convex Optimization. 2004. Available online: <https://web.stanford.edu/~boyd/cvxbook/> (accessed on 3 July 2025).
33. Xu, J.; Yao, J. Exploiting physical-layer security for multiuser multicarrier computation offloading. *IEEE Wirel. Commun. Lett.* **2019**, *8*, 9–12. [CrossRef]
34. Cao, X.; Zhu, G.; Xu, J.; Huang, K. Cooperative interference management for over-the-air computation networks. *IEEE Trans. Wirel. Commun.* **2021**, *20*, 2634–2651. [CrossRef]
35. Jiao, L.; Ge, Y.; Zeng, K.; Hilburn, B. Location privacy and spectrum efficiency enhancement in spectrum sharing systems. *IEEE Trans. Cogn. Commun. Netw.* **2023**, *9*, 1472–1488. [CrossRef]
36. Liu, D.; Simeone, O. Privacy for free: Wireless federated learning via uncoded transmission with adaptive power control. *IEEE J. Sel. Areas Commun.* **2021**, *39*, 170–185. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Optimizing Weighted Fair Queuing with Deep Reinforcement Learning for Dynamic Bandwidth Allocation

Mays A. Mawlood * and Dhari Ali Mahmood

Department of Computer Engineering, University of Technology-Iraq, Baghdad 10066, Iraq;
dhari.a.mahmood@uotechnology.edu.iq

* Correspondence: ce.22.02@grad.uotechnology.edu.iq

Abstract: The rapid growth of high-quality telecommunications demands enhanced queueing system performance. Traditional bandwidth distribution often struggles to adapt to dynamic changes, network conditions, and erratic traffic patterns. Internet traffic fluctuates over time, causing resource underutilization. To address these challenges, this paper proposes a new adaptive algorithm called Weighted Fair Queues continual Deep Reinforcement Learning (WFQ continual-DRL), which integrates the advanced deep reinforcement learning Soft Actor-Critic (SAC) algorithm with the Elastic Weight Consolidation (EWC) approach. This technique is designed to overcome neural networks' catastrophic forgetting, thereby enhancing network routers' dynamic bandwidth allocation. The agent is trained to allocate bandwidth weights for multiple queues dynamically by interacting with the environment to observe queue lengths. The performance of the proposed adaptive algorithm was evaluated for eight queues until it expanded to twelve-queue systems. The model achieved higher cumulative rewards as compared to previous studies, indicating improved overall performance. The values of the Mean Squared Error (MSE) and Mean Absolute Error (MAE) decreased, suggesting effectively optimized bandwidth allocation. Reducing Root Mean Square Error (RMSE) indicated improved prediction accuracy and enhanced fairness computed by Jain's index. The proposed algorithm was validated by employing real-world network traffic data, ensuring a robust model under dynamic queueing requirements.

Keywords: multi-queue systems; adaptive algorithms; resource utilization; EWC; bandwidth allocation; WFQ; continual learning

1. Introduction

In recent years, the demand for high-speed data transmission, efficient resource management, and network communication has undergone a significant rise due to the rapid growth in the numbers of connected devices and data-intensive applications, leading to efforts to improve the performance of data transmission queueing systems and giving rise to congestion, latency, and dynamic bandwidth allocation challenges [1]. Due to the conflict between the growth in traffic and the constraints on bandwidth, it is crucial to utilize bandwidth resources efficiently to enhance the Quality of Service (QoS) [2]. Modern networks rely on advanced scheduling algorithms and optimized queueing mechanisms to address these issues, support performance, and ensure fair resource distribution. As networks evolve, achieving scalability, flexibility, and reliability remains an important goal for researchers and industry professionals [3]. Unfair resource allocation can lead to resource waste, starvation, or redundancy [4].

A system that allows devices to communicate and exchange data efficiently is known as a network. Networks vary in size and design, ranging from Local Area Networks (LANs) to Wide Area Networks (WANs). Routers serve as key components of network communication as they direct data packets between networks. Through network communication, routers manage traffic using queueing systems, which control how packets are processed and transmitted [5]. A queueing system is essential for handling varying traffic loads, reducing congestion, and ensuring smooth data flows. Different models, such as M/M/1, M/M/c, and G/G/1, determine how packets arrive, wait, and are then served in network queues. These models can optimize performance and reduce delays, thus supporting the design of efficient traffic management strategies [6].

Scheduling algorithms are implemented in routers to efficiently manage packets in queues [3]. These algorithms determine the packet processing order and aim to strike a balance between fairness and efficiency. Basic scheduling methods, such as First-In-First-Out (FIFO), process packets consecutively, whereas Priority Queuing (PQ) prioritizes essential traffic, such as real-time voice or video data. More advanced approaches, such as Weighted Fair Queuing (WFQ) [2,3] and Deficit Round Robin (DRR), ensure proportionate bandwidth allocation among traffic classes. These are crucial in achieving the desired QoS, optimizing bandwidth usage, and improving network reliability [3]. Reinforcement Learning (RL) is commonly used in resource allocation to optimize decision-making in dynamic and unsteady environments. Reinforcement learning-based resource allocation enables the efficient distribution of bandwidth, computing resources, scheduling priorities, and networking [1,4,7].

According to the aim of this paper, while effective in ensuring fair bandwidth distribution, traditional Weighted Fair Queuing (WFQ) systems struggle to adapt dynamically to varying network traffic patterns, such as bursty or periodic flows. Due to its adaptive learning capabilities, deep Reinforcement Learning (DRL) has become a profitable solution for dynamic bandwidth allocation. Regardless, typical DRL models suffer from catastrophic forgetting when exposed to new traffic scenarios, requiring frequent retraining. This limits scalability and robustness in real-time networking environments. Thus, there is a critical requirement to enhance DRL-based WFQ systems with continual learning mechanisms that can retain earlier obtained scheduling policies while adapting to new network conditions efficiently and without forgetting, therefore enabling a more intelligent and scalable bandwidth allocation strategy.

The remainder of this paper is organized as follows: Section 2 includes related work, Section 3 produces a contribution of this research, Section 4 describes a methodology for the proposed system model. Section 5 provides the results and analysis. Section 6 discusses the model parameters and the results. Section 7 includes the conclusions.

2. Related Work

Various scheduling disciplines are available for managing separate queues while maintaining fairness and balance in terms of Quality of Service (QoS). In communication networks, WFQ is a crucial scheduling discipline used to ensure fairness and QoS. Using a weight-based bandwidth allocation approach, WFQ assigns weights to various queues based on packet length, thereby ensuring equitable distribution of bandwidth [8]. A technique called PQ, which prioritizes the queue order, often causes resource starvation for lower-priority queues by serving the highest-priority option among non-empty queues. Simulation results show that PQ has a smaller memory queue size than WFQ and FIFO, and a more significant delay [9]. Scheduling disciplines such as WFQ cannot dynamically allocate bandwidth and must learn about the status of queues. The WFQ method has been extensively studied in the literature, with a focus on performance evaluation. Several

metrics are typically used with the WFQ system, including the mean queue length, system throughput, and mean response time. It would be interesting to compare the performance of a WFQ system with a finite buffer to that of a WFQ system with an infinite buffer [10].

The authors of [3] presented mechanisms and approaches to implementing QoS services in packet networks and created and examined a network model that facilitated data transfer associated with various traffic classes. Traffic shaping techniques based on the WFQ method were utilized for QoS. In individual work, the study in [11] thoroughly investigated traffic scheduling in IPv4 and IPv6 networks for multimedia applications, examining the effectiveness of various scheduling methods for different types of traffic, including HTTP and FTP. Among the different approaches considered (WFQ, FIFO, PQ), WFQ was found to be the best option in terms of its improved performance over the other algorithms, as it offered suitable outcomes over both IPv4 and IPv6. In [12], a packet-scheduling technique was developed that combined WFQ with PQ to decrease latency, significantly improving network performance compared to scenarios where packet scheduling was not employed. By accounting for the extra waiting periods caused by variations in arrival rates, the proposed model was able to lower the estimation error. Even when traffic utilizes the exact network channel or session, it is necessary to consider a complex mix of network and application traffic. The authors also proposed an approximate analytical technique for this purpose. In these research works, specified weights were allocated based on traffic class or priority due to the absence of adaptability to real-time network requirements.

In [13], the QoS performance of the WFQ algorithm was investigated and evaluated, and various QoS results were examined. The average throughput, packet arrival ratio, packet loss ratio, and average latency were calculated. The study in [14] considered QoS factors such as jitter, delay, and packet loss probability and evaluated these for algorithms using PQ, FIFO, custom queuing, fair queueing, and WFQ. Given that PQ is usually regarded as the least efficient choice, it is recommended to employ it sparingly in situations where there is little chance of channel erasure [15]. The simulation referenced the approximation terms presented by Mahmood and Horvath for the mean response times in a two-class (ideal) WFQ system with a Poisson arrival process and exponentially distributed service times. The approximation is usually based on an arithmetical approach, and some decisions on how to approximate the behavior occasionally appear to be made ad hoc [16]. These approaches described above rely on predefined controls for resource allocation, which may limit their ability to adapt dynamically to unexpected changes in traffic patterns. Furthermore, these approaches may face challenges in regard to optimizing real-time performance compared to techniques that leverage predictive analysis and adaptive learning.

Studies have been conducted on the application of reinforcement learning (RL) to queueing systems. This research field can be categorized further based on the RL techniques used; for example, in model-based RL, a specific queueing network model is considered to optimize the network's performance, and the focus is on the use of RL techniques and managing these limitations by exploring the effectiveness of this approach in other network architectures [17]. The authors propose a DRL-based AQM, Learning Fair Queue (LFQ), that dynamically adapts per-flow buffer sizes using a neural network. It addresses the challenge of managing real-time queues for diverse traffic conditions. Results display enhanced throughput fairness and lower latency compared to traditional AQMs [18]. The author presented a Deep Q-Network (DQN) algorithm employed for smart queue management. The procedure includes simulating traffic scenarios with two types of protocols, UDP and TCP, using a neural network. Specifically, the DQN agent can select to increase, decrease, or preserve weights for diverse classes of traffic: Gold, Silver, and Bronze. The

outcomes show that the DQN agent can effectively manage the weights for the traffic flows. The average reward stabilizes positively, showing convergence after approximately 1500 training episodes [19]. The authors of [20] outlined the use of a Deep Deterministic Policy Gradient (DDPG) algorithm to reduce the service rate and provide a probabilistic latency guarantee, finding that the proposed method offered probabilistic upper bounds on the system's end-to-end delay. This approach is more challenging than standard practice control methods, which only consider average performance metrics such as the mean delay. In the study in [21], a WFQ system achieved a reduction in the packet loss rate by utilizing a SAC technique, although this scheme did not account for latency. The solution in [22] employed deep RL (DRL) techniques for dynamic bandwidth allocation in WFQ routers. Specifically, it emphasized the application of algorithms such as DDPG and SAC to handle bandwidth allocation across multiple queues, with the goal of minimizing packet loss, reducing the average queue length, and decreasing the average delay.

3. Research Contributions

This paper contributes to the formulation of a $G/G/1/K$ parallel queuing system by developing the WFQ continual DRL framework. The proposed approach integrates DRL with continual learning to enhance bandwidth utilization in the WFQ algorithm. The model ensures fairness by dynamically adapting allocations. Continual learning is represented by Elastic Weight Consolidation (EWC), a technique that protects the important parameters of a neural network when the model is trained on new tasks. The system continuously learns and adjusts according to network demands and objectives, thereby maximizing the cumulative reward. In addition, continual learning reduces the retraining overhead, making the model more efficient in dynamic environments compared to traditional DRL-based approaches, as demonstrated in this paper. It uses EWC to mitigate catastrophic forgetting, enabling real-time adaptation. In comparison to existing literature, this work proposes a reinforcement learning method to optimize QoS by adjusting actions based on the queue state, such as packet loss or transmission decisions. While effective, its approach is limited to static traffic environments, and the service rates chosen are such that the system stays stable, does not get congested, and lacks mechanisms for long-term adaption while network conditions grow [20]. Also, the paper [22] proposes using DRL to allocate bandwidth dynamically across WFQ queues, improving performance under varying traffic loads. Still, it does not consider the need for continual adaptation or knowledge retention when system configurations of queue length.

This paper presents an adaptive algorithm that integrates DRL with EWC to support continual learning across varying network conditions. This approach handles adaptive bandwidth allocation under WFQ. It maintains previously obtained knowledge, providing stability and reducing catastrophic forgetting as the number of queues increases, which leads to faster convergence and makes it more suitable for real-time and complex network scenarios where traffic dynamics evolve over time. The main contribution of the proposed approach can be illustrated as described below:

1. The proposed model improves performance by integrating the SAC algorithm with EWC, enabling continual learning within the WFQ scheduling framework. Through integration, the system can handle dynamic traffic patterns, such as bursty or periodic flows, without forgetting previously learned scheduling policies, ensuring adaptability to new scenarios without frequent retraining.
2. The approach dynamically learns optimal bandwidth allocation across multi-queue systems in WFQ in real-time, adapting to high traffic and queue length variations.

3. EWC enables the agent to retain previously learned policies while learning new ones, eliminating the need to reprocess old data. This helps system stability and scalability in evolving network environments, particularly when facing unpredictable traffic patterns.
4. By integrating SAC and EWC within the WFQ scheduling algorithm, the model achieves efficient resource allocation and robustness to changing network requirements. That helps fair queuing and constant performance across varying traffic scenarios.
5. The new adaptive algorithm ensures high performance while increasing the queues, allowing the system to maximize performance.
6. A real traffic network dataset provides a more realistic and complex testbed, capturing burstiness, and variability across multiple queues.

4. Proposed Model

4.1. Model Definition

Using the WFQ discipline as a scheduling algorithm, the information system can be described as a $G/G/1/K$ parallel queueing system [22], where routers are modeled as such. Differential arrival flows arrive at each queue independently and are then sent at a rate determined by the appropriate bandwidth allocation. The arrival rates and patterns of incoming packets are erratic and change over time, whereas the router's overall bandwidth resources are constant. To enhance the system's performance, allocating limited bandwidth resources across queues with varying arrival attributes is crucial. The proposed system consists of n $G/G/1/K$ queueing systems that share standard bandwidths in parallel, as shown in Figure 1.

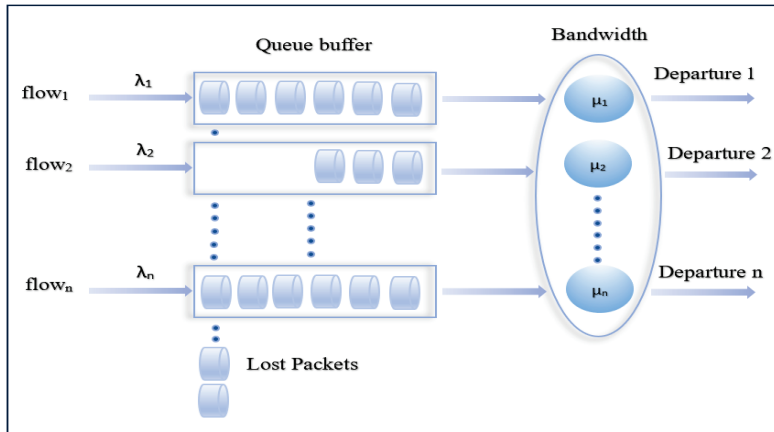


Figure 1. A model containing n $G/G/1/K$ queueing system [22].

In this study, it is assumed that the packets enter a matching buffer to wait for service after arriving at the system via a process with commonly dispersed interarrival periods, where λ_i represents the flow and i is the arrival rate. The flow i locates service times for packets from the associated buffer broadly and dispersedly. The service rate is $\mu_i = w_i \times b$, where b is the system's total bandwidth and w_i is the server's allocated weight. K_i is the buffer size for the queue _{i} . If the buffer is full, newly arriving packets will be discarded. The goal is to find a control policy that can effectively decrease the average latency and packet loss rate in a WFQ system.

4.2. Continual Learning

Continual learning, also known as incremental learning or lifelong learning, enables an agent, such as an artificial intelligence model, to continuously acquire new skills or knowledge while retaining previously learned information. A problem known as 'catastrophic forgetting' occurs when a model updates its parameters to learn a new task and, in

doing so, loses its ability to perform the previously learned tasks effectively. It also refers to the process by which the human brain learns and retains information. In humans, the neocortex is believed to have mechanisms that enable the integration of new information without erasing old memories. This biological principle is a key aspect when learning languages or playing musical instruments, where new skills must coexist with existing knowledge [23].

4.3. *Continual Learning in a Reinforcement Learning Field*

RL is a type of machine learning that enables machines to make decisions. An agent learns to make decisions by interacting with an environment and receives rewards or penalties based on its actions, allowing it to learn optimal behaviors through trial and error. The goal is to learn a policy that maximizes cumulative rewards over time. An example of RL is the SAC algorithm, which utilizes a maximum entropy framework. This algorithm learns stochastic policies through an objective that maximizes both the expected reward and the policy's entropy, thereby encouraging exploration [24]. To utilize the RL domain to address a more comprehensive set of conditions needed for successful continual learning systems, EWC can be used to retain previously acquired knowledge. In particular, higher-level mechanisms are needed to infer which task is currently being performed, to notice and incorporate new tasks as they are encountered, and to allow for rapid and elastic switching between tasks [23]. SAC is particularly effective for queueing and network scheduling tasks [22], as opposed to other algorithms such as Proximal Policy Optimization (PPO) and Twin Delayed Deep Deterministic Policy Gradient (TD3), because it balances sample efficiency, exploration, exploitation, and stability, which are essential in dynamic environments. Unlike PPO, which is an on-policy method requiring more data and computational resources, SAC's off-policy procedure enables faster learning from fewer samples, which is essential for real-time queue management, where quick adaptation is crucial. Compared to TD3, SAC incorporates entropy maximization, enabling better exploration that can arise from early convergence, a common problem in non-stationary traffic scenarios. Regarding to TD3, it employs deterministic policies, which can result in insufficient exploration, particularly in dynamic or non-stationary environments, such as varying network traffic [25].

4.4. *Continual Reinforcement Learning Framework for WFQ*

The process allocates bandwidth weights in the WFQ system, which operates under a control policy that takes into consideration the system's dynamic nature. The interaction between the algorithm and multi-queue systems is shown in Figure 2.

After observing the system's status, represented in the form of queue lengths, the agent decides to allocate bandwidth weights as an action. Until the next decision is taken, the queueing system will use the allocated bandwidth and send rewards back to the agent, allowing it to adjust its policy accordingly. This study focuses on dynamic control and slot dependence, where the decision time is t seconds. Since the state may vary significantly within a specific time slot, it is more practical to implement it in the engineering field, but this also makes it more challenging to determine the optimal policy. By integrating continual learning with DRL, the agent learns optimal policies, retains knowledge from past experiences, and adapts to evolving network conditions without catastrophic forgetting. This approach enables the system to efficiently generalize across different traffic patterns, thus reducing the need for comprehensive retraining. A continual RL-based agent can enhance bandwidth allocation over time by leveraging past knowledge while adapting to previously unnoticed traffic behaviors, leading to optimal stability, scalability, and long-term performance in dynamic networks.

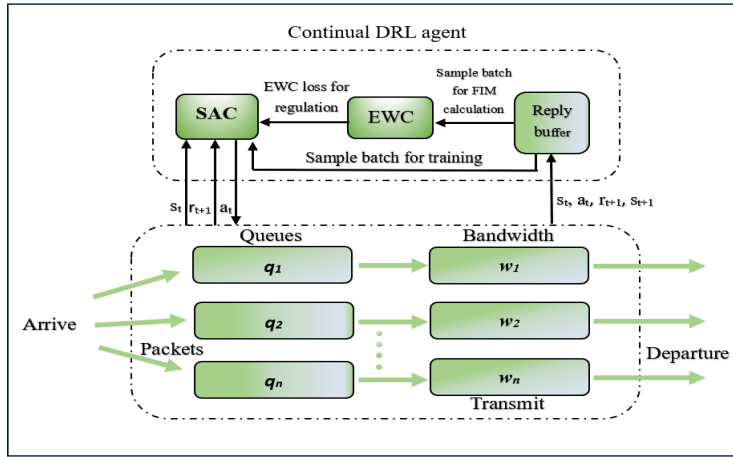


Figure 2. Interaction between the WFQ continual-DRL algorithm and the multi-queueing system.

For a continuous state space and a continuous action space, the continual DRL situations are as follows:

State: q_i is defined as the number of packets that are either serviced by the server i or are waiting in $queue_i$. $q_i = (q_1, q_2, \dots, q_n)$, $i = 1, 2, \dots, n$ also represents the state of the system. As it captures queue lengths, the continual DRL method is practical for use in telecommunications networks. If we assume $q_i \leq K_i$, then each queue has a limited buffer. $s = \{q = (q_1, q_2, \dots, q_n) \in N_n : q_i \leq K_i\}$ represents the state space. However, this is enhanced with historical states, thus allowing the model to be better generalized and to retain past traffic patterns. In the state of a multi-queue system, the input to the neural network is a vector representing the current queue lengths of the multi-queue queues as $s = \{q = (q_1, q_2, \dots, q_n) \in N_n : q_i \leq K_i\}$. To confirm stable and efficient learning, each queue length was normalized before being input into the neural networks. The normalization approach involves logarithmic scaling $\sum_{i=1}^n \lceil \log_2(K_i + 1) \rceil$, which corresponds to the current queue length of q_i . This transformation reduces the impact of large values, ensuring the state input remains within a manageable range while preserving relative differences and stabilizing the training process. This normalized state vector is then provided into fully connected layers of the neural networks for policy and Critic network estimation, encouraging the agent to learn QoS-aware scheduling decisions effectively.

Action: At the decision-making step, the bandwidth weights for each queue are allocated, and this defines the action a . a represents the action space, where $\{a = (w_1, w_2, \dots, w_n) \in \mathbb{R}_+^n : \sum_{i=1}^n w_i = 1\}$. In continual DRL, the model leverages history-learned policies, adapting to new traffic patterns without relearning from scratch and improving the efficiency of decision-making.

Reward: In queueing systems, minimizing the average delay is equivalent to reducing the average queue length. It is assumed that for every packet in the queue, the cost per unit of time is equal to the cost of charging for the delay. The reward function is as follows:

$$r(q, a) = - \sum_{i=1}^n (h_i q_i + c_i l_i) \quad (1)$$

where the symbol l_i represents the number of lost packets in a given unit of time. In addition, the packet loss l_i , for $i, 2, 3, \dots, n, q_i$, which includes the number of packets lost for queue length represented as a vector, where every packet loss results in a penalty c_i represented as a scalar. We also assume that each packet in q_i represents the current length of q_i a vector number of packets generates a holding cost h_i per unit of time, h_i the holding cost per packet in q_i representing a scalar equivalent to imposing a penalty for its delay. A counted continual DRL scheme includes

a mechanism to evaluate how well past learning is transferred to the new requirements, thus confirming the adaptation of requirements and optimal dynamic bandwidth allocation. Each element in the vector of queue length q_i is defined as the number of packets currently held in that queue. The packet loss is also described as a vector, where each element corresponds to the number of packets lost in the queue.

In this paper, we incorporate EWC into RL to prevent catastrophic forgetting when switching between diverse tasks, thereby ensuring that the agent retains knowledge of previous tasks while optimizing the current task. By assessing which parameters are critical for previously learned tasks and constraining their updates during training on new tasks, this approach helps maintain the performance on earlier tasks. This enables knowledge retention and adaptation to new traffic situations without catastrophic forgetting, thereby improving generalization to previously unnoticed workloads.

- EWC regulation: The loss function is modified by adding the EWC penalty to ensure stability when learning a new policy. The regularization term is added for each policy and critic losses, confirming that changes to parameters considered important for previous tasks are penalized. Thus, the final loss function employed in backpropagation is:

$$\mathcal{L}_{total} = \mathcal{L}_{RL} + \sum_i \frac{\lambda_{EWC}}{2} F_i (\theta_i - \theta_i^*)^2 \quad (2)$$

where $\mathcal{L}_{RL}$ is the standard loss function, $\sum_i \frac{\lambda_{EWC}}{2} F_i (\theta_i - \theta_i^*)^2$ is the term representing the EWC regulation function, λ_{EWC} reflects the significance of the old task compared to the new task, θ_i^* is the optimal parameter from the previous task, and i tags each parameter. F_i is the Fisher Information Matrix (FIM), which measures the importance of each parameter. The aim of EWC is to maintain the network parameters relative to the learned parameters of both the first and second tasks. When the model is applied to a third task, this constraint can be enforced either with two separate penalties or by combining them into one, since the sum of two quadratic penalties is also a quadratic penalty [23].

- FIM: The FIM provides essential parameters for preserving primary tasks. It is an essential concept that reflects the amount of information an observable random variable carries about an unknown parameter upon which the probability of that random variable depends. It assesses the sensitivity of the network's performance in terms of parameter variations θ_i^* and it quantifies the sensitivity of the model's parameters to modifications in the loss function, capturing the importance of each parameter for previously learned tasks. In practice, the FIM is estimated by computing the squared gradients of the loss for the policy and critic networks with respect to their parameters, averaged over samples from the replay buffer. Where θ is the optimal parameter θ_i^* set after learning task A, these gradients function as a practical approximation of the FIM and are calculated as:

$$F_\theta = \mathbb{E} \left[\left(\frac{\partial \log \pi_\varnothing(a|s)}{\partial \theta} \right)^2 \right] \quad (3)$$

Here, $\pi_\varnothing(a_t | s_t)$ is a policy network, where s represents the state, and a represents the action [26]. This formula is also applicable to the critic network Q.

The FIM computing is as follows:

1. Train networks (policy, two critic networks)
2. From the replay buffer, sample a batch of experiences.
3. Computing the loss function for the policy and critic.
4. Obtain gradients by performing backpropagation with respect to their weights.

5. Calculating the squared gradients and averaging them over the buffer size.
6. Store θ_i^* and FIM for compute EWC loss (regulation term).
- In this study, the algorithm adapts bandwidth allocation decisions based on queue states. The theoretical fairness share for a given flow at any step-in episode is computed using the following formula [27]:

$$Fairness = \frac{w_i}{\sum_{j=1}^i w_j} w_i \geq 0 \quad (4)$$

where w_i is the weight allocated for each q_i , and it is proportional to the total weights $\sum_{j=1}^i w_j$. During training in each step, the theoretical fairness computes how each queue in the system should allocate bandwidth as a proportion of the total bandwidth based on its state (i.e., queue length), representing the actual value. According to the prediction, value came out of the policy network π_ϕ , which is a vector of action $a = (w_1, w_2, \dots, w_n)$ that will be implemented in the environment to allocate bandwidth for each queue and then the service rate will be computed.

- Root Mean Square Error (RMSE): A metric used to measure the root of average deviation of prediction value from target value, the measure is defined as indicated in equation [28]:

$$RMSE = \sqrt{\frac{\sum_{N=1}^i (\hat{y} - y)^2}{N}} \quad (5)$$

where $\hat{y}$ is the prediction value denoting weights allocated for each q_i , which came from policy network π_ϕ , y is the actual value indicating the weights computed by fairness function. In this study, the RMSE metric can be used to assess the model during training, and it offers insight how the model accurately becomes close to the theoretical fairness, meaning that the deviation of prediction value (weight for each q_i came from policy network) from the target value (weights of each queue computed by fairness function) into each step.

- Jain's Fairness Index: It is a mathematical metric used to evaluate how fairly resources, such as bandwidth, are distributed among multiple flows or entities in a system. It is defined as [29]:

$$f(x) = \frac{[\sum_{i=1}^n x_i]^2}{N \sum_{i=1}^n x_i^2} \quad \text{where } x_i \geq 0 \quad (6)$$

where $\sum_{i=1}^n x_i$ is represents the squared of stochastic action values predicated by the policy network π_ϕ , reflect the actual network in bandwidth allocation decision for each queue. $\sum_{i=1}^n x_i^2$ is optimal fitness of the bandwidth allocation (i.e., weight for each queue relative to the current queue state computed theoretically by fairness function). Specifically, in this study, a value of the Jain index near 1 indicates that the policy network's predictions closely match the optimal fairness pattern. If the value of the Jain index is significantly less than 1, it indicates a deviation from fairness (optimal weights for each queue based on the current state computed by fairness function), indicating that some queues are over or under-served relative to their optimal share.

4.5. WFQ Continual-DRL Algorithm

This study integrated continual learning techniques with a DRL framework for weighted fair queues within the proposed WFQ continual-DRL algorithm to dynamically allocate bandwidths. In this approach, an agent explores the dynamics of the queueing

system through periodical, recursive interactions within the architecture of WFQ continual DRL. The agent gains new insights about the environment's surroundings and exploits them to obtain high rewards. The agent also retains and transfers history-learned policies to improve decision-making in dynamic traffic environments and accumulates knowledge over time. One challenge is that the huge buffer properties of the WFQ system and several queues contribute to the curse of dimensionality.

To address this issue, the WFQ continual-DRL algorithm is employed, providing the control policy estimates as input to the neural network parameters. As input for neural networks, the system's state is transformed into a binary vector, thus eliminating the need for storage space. This enables us to receive normalized state information for various buffers, preventing a gradient explosion problem when computing policy gradients and accessing normalized queue lengths. This improvement enhances the numerical stability, interpretability, and learning efficiency of the system, where K_i vary widely. This ensures that the transformed values are more fluent for the WFQ continual DRL algorithm. $\sum_{i=1}^n \lceil \log_2(K_i + 1) \rceil$ is the dimension of s . The off-policy WFQ continual-DRL algorithm is also used to achieve high levels of data efficiency. In essence, the off-policy technique breaks the correlation between consecutive data using a replay buffer, and can, therefore, achieve more data usage efficiency by repeatedly utilizing historical data. After each investigation stage, transition data $(s_t, a_t, r_{t+1}, s_{t+1})$ are gathered and saved in the replay buffer to be sampled later and used to modify the control strategy. In addition to the sampled updating policy strategy, the updated parameters of the neural network are used to compute the FIMs, to ensure that the most critical parameters from the past are preserved. Continual learning is used to overcome catastrophic forgetting.

The SAC procedure is a sophisticated technique that optimizes the expected reward and policy entropy simultaneously. The efficiency of the investigation is improved by enhancing the entropy and optimizing the expected reward by maximizing the entropy of the policy. SAC enhances the efficiency of self-exploration and has been shown to achieve stable performance in continuous action spaces. Stable performance in a continuous action space has also been demonstrated. SAC involves employing a policy network π , and a critic network Q to approximate a stochastic policy. The structure of the neural network integration of SAC with EWC for eight queues is illustrated in Figure 3.

Two critic networks are utilized in the SAC algorithm to mitigate the overestimation bias commonly found in DRL. To evaluate the Q -values corresponding to state-action pairs, the critics, represented as Q_1 and Q_2 , are accountable, while their target networks, $\bar{Q}_1$ and $\bar{Q}_2$, facilitate stable learning through soft updates. The control calculates the target Q -value by selecting the minimum estimate from the two critics to reduce the likelihood of overestimation during training. The critics are updated using the MSE, which is the loss between the predicted Q -values and the computed target [24].

In the proposed algorithm of the WFQ continual-DRL algorithm, the integration of continual learning, represented by EWC, with the DRL algorithm, particularly in the WFQ system, enhances the agent's ability to retain optimal scheduling knowledge across various traffic scenarios. EWC identifies important neural network weights from previous queue patterns and penalizes changes to these weights during training on new patterns, thereby decreasing catastrophic forgetting. This process helps preserve performance consistency across varying queue loads and quick convergence in the early training episodes. As a result, the continual-DRL agent adapts efficiently to changing traffic intensities while ensuring stable scheduling [23].

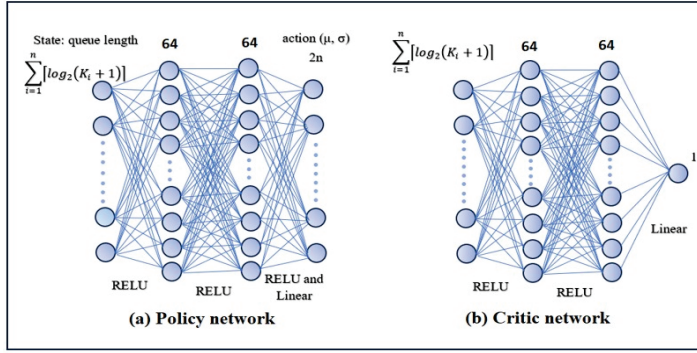


Figure 3. Structure of the neural network integrating SAC with EWC for eight queues.

In this study, a simulation was conducted for twelve queues as an enhancement. The parameter tuning process of the neural network for integration of SAC-EWC into the WFQ continual-DRL algorithm involved expanding the number of queues to twelve, which increased the state as input to the neural network and the action space as output, making the scheduling problem more dynamic and complex. This adjustment demands a more capable model to allocate bandwidth effectively across a larger number of queues. However, increasing the number of neurons in the network to 128 enhances the model's ability to approximate complex policies, resulting in improved decision-making. When extended to twelve queues, the input defined as state will be increased and output representing action will increase. Increasing the number of neurons from 64 to 128 to enhance the model's learning allows the network to capture more complex patterns in the data training, which is mainly valuable in dynamic environments such as multi-queue systems. This adjustment aims to improve performance and reach higher prediction accuracy. This enables the model to capture more complex pattern in the data through training data and avoid underfitting. Underfitting occurs when the model is too simple to capture the underlying patterns in the training data, resulting in inadequate and poor learning. This results in increased errors and invariably low rewards, even during training. A larger replay buffer enables the agent to retain past experiences for a longer period, thereby enhancing training stability by providing a diverse range of samples during training. Additionally, increasing the batch size to 256 contributes to more stable updates by averaging gradients over a larger set of experiences, thereby reducing variance in training. These enhancements improve learning efficiency while providing excellent stability in dynamic bandwidth allocation scenarios. The structure of the neural network integrating SAC with EWC for twelve queues is shown in Figure 4.

In the SAC structure of the critic network, the use of target critics stabilizes the learning process for both critic networks Q_1 and Q_2 . The π network produces the μ and σ of the distribution. In this approach, the action is displayed as a probability distribution (Gaussian distribution). The Gaussian distribution is represented as follows:

$$f(a|\sigma, \mu) = \frac{1}{\sigma\sqrt{2\pi}} \exp^{-\frac{1}{2} \left(\frac{a-\mu}{\sigma}\right)^2} \quad (7)$$

where μ, σ represents the input to the Gaussian distribution formula predicted by the policy network. a is the output from a Gaussian distribution. Whenever the action stays in the area around the mean, the action will be better [30].

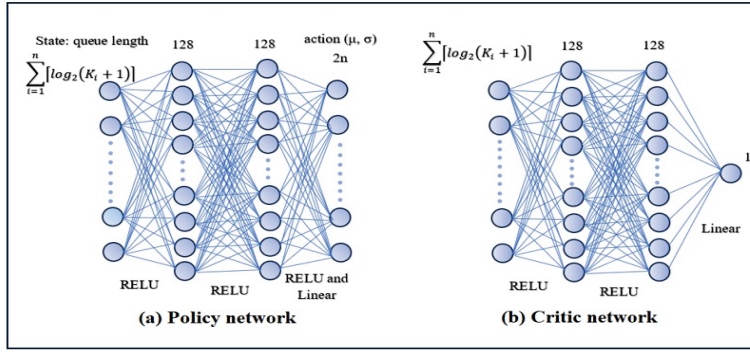


Figure 4. Structure of the neural network integrating SAC with EWC for twelve queues.

The action was sampled randomly from a probability distribution defined by μ, σ , which is known as a stochastic action or stochastic policy. A random noise ε is generated to confirm exploration, the value of the noise ε is drawn from a random distribution represented by $N(0, 1)$. The output action was sampled from $a = \mu + \sigma\varepsilon$, which internally is computed in Gaussian distribution. This process, known as reparameterization sampling of the action value, allows for effective policy optimization as follows [24]:

$$a_t = f_{\varnothing}(\varepsilon_t; s_t) \quad \text{Where } \varepsilon \sim N(0, 1) \quad (8)$$

where a_t represents the action, s_t represents the current state, and ε is the noise vector from the Gaussian distribution. This function $a_t = f_{\varnothing}(\varepsilon_t; s_t)$ means that internal noise addition allows the agent to produce stochastic actions necessary for exploration while trusting the training process is stable and efficient. The action reductions are set to be within the required range by applying the ReLU activation function, which is practical to eliminate negative values where only positive actions are valid, which is particularly useful in settings (bandwidth weights) for WFQ scheduling. The policy network effectively explores continuous action while a stochastic process balances between exploration and exploitation.

The approach presented in this study combines SAC with EWC. The goal is to integrate the EWC regulation mechanism into SAC to maintain stability across states while preserving knowledge learned from the first task through training. As usual, the end of the task includes calculating and storing the FIM and the parameters $(\theta_i^*, \varnothing_i^*)$. When reserving a new state, the training is initialized in that state. At the parameter update stage, the term representing the EWC loss function is added to the SAC losses. After completing the task, the Fisher matrix is updated, and the network parameters are saved. The combined losses that ensure the performance of the previous task are held constant while learning a new task, allowing the network to output the action [23].

In the WFQ continual-DRL algorithm, the significance of each model parameter is computed and evaluated. When the algorithm begins, the model is set to evaluation mode, and a dictionary is initialized to store Fisher information values for each parameter. Following this, a batch of experiences is sampled from the replay buffer, and the gradient is computed based on the loss function associated with either the actor or the critic network. In the actor network, the loss is defined as the mean of the predicted actions, whereas in the critic network, it is the mean squared error between the predicted Q-values and the target Q-values. After performing backpropagation, the Fisher information for each parameter is updated by averaging the squared gradients across the sampled batch. These values are then detached from the computation stored in a dictionary, to ensure that the learned values of the critical parameters are retained when the model is adapted to new tasks. The optimal parameter values are also stored to facilitate the EWC penalty calculations during training.

EWC characterizes itself from other continual learning methods, such as Gradient Episodic Memory (GEM) and Learning without Forgetting (LwF), by focusing on simplicity, memory efficiency, and biological probability. Unlike GEM, which requires high memory and computational costs, it stores data from previous tasks and performs confining on previous outputs or parameters, resulting in additional complexity and storage requirements. In contrast, EWC introduces a simple regularization based on the Fisher Information Matrix (FIM) to preserve important weights without requiring the preservation of old data or models, which makes EWC efficient, lightweight, and perfect for continual learning in dynamic resource settings, such as DRL-based queuing and scheduling systems [31].

The process of choosing the optimal value for λ_{EWC} is conducted in two steps. In the first step, six distinct ranges are defined using the random function, and each one is explored independently. Each range is run over 1000 episodes of network training, and random values for the new λ_{EWC} are sampled and evaluated. The performance of each sample is estimated based on the measure of the loss function and reward performance.

When the random function had been implemented, λ_{EWC} was obtained based on six ranges, for each range created five values (random value), which were operated over 1000 episodes. The model then iteratively trains the SAC with EWC agent for each λ_{EWC} , which evaluates the agent's performance. The loop stores tracking of the best-parameter λ_{EWC} , then updates the best λ_{EWC} by comparing with performance whenever a higher-performing result is found. This procedure enables observed tuning of the λ_{EWC} parameter to optimize learning stability and policy performance. Then, from each range, pick the best value of λ_{EWC} . In the second step, after all six ranges have been explored, obtain the optimal six λ_{EWC} value by conducting a structural evaluation in which a grid search process is applied to systematically compare the performance of the six optimal λ_{EWC} values that resulted from the random function used in training for 1000 episodes. The best-performing λ_{EWC} is selected based on the impact of the MSE value and the reward performance, ensuring optimal performance in continual learning. The model reached the best performance with a rather high λ_{EWC} value of 2271 for eight queues, 2275 for twelve queues, which led to significantly improved reward results and more stability in MSE and MAE. The high value of λ_{EWC} demonstrates that the model strongly penalizes changes to significant parameters, thus preserving previously learned knowledge more effectively. It reflects the model's dependance on stability over adaptivity, which is valuable in dynamic environments where continual learning is critical.

To cover the experimental characteristics of a random search, a combination of these two methods represented by the random function and grid search can control a wide range of possible values; this is followed by a refined grid-method evaluation to determine the best-performing λ_{EWC} .

The WFQ Continual-DRL Framework: Policy-Critic Update with EWC

The WFQ environment is designed to model multi-queue scheduling principles. It simulates multiple input traffic flows; each represented as a separate queue that shares a limited total bandwidth. Inter-arrival times and service times use gamma-distributed values that reflect the bursty nature of real network traffic. At every decision step (20 s), the state defined as queue length $s = \{q = (q_1, q_2, \dots, q_n) \in N_n : q_i \leq K_i\}$ takes as input to actor network to stabilize learning, and queue states normalize as $\sum_{i=1}^n \lceil \log_2(K_i + 1) \rceil$, ensuring the agent perceives state change on a normalized scale and queues receive an action vector from an actor network $a_t \sim \pi_{\mathcal{O}}(a_t | s_t)$, describing the ratio of total bandwidth allocated to each queue. This allocation determines the service rate for each queue $\mu_i = w_i \times b$, and the number of packets served is calculated based on gamma-distributed service times noting action defined as $\{a = (w_1, w_2, \dots, w_n) \in \mathbb{R}_+^n : \sum_{i=1}^n w_i = 1\}$, which account for

variability in processing or transmission delays. The served packets are removed from the queue, while newly arrived packets arrive, ensuring a fluid and realistic evolution of the queue states. The reward function contains two forms of penalties: firstly, holding cost h_i for unserved packets remaining in the queue, representing delay, and l_i represents packet lost due to buffer overflow.

Two critics $Q_{\theta_1}(s_k | a_k)$, $Q_{\theta_2}(s_k | a_k)$ functions are maintained, which are trained to minimize the error between the predicted Q-values and the target Q-value. In this approach, the EWC method is added to preserve performance on previous tasks during continual learning. For a batch of sampled transitions $(s_t, a_t, r_{t+1}, s_{t+1})$, the target value is computed as:

$$y_k = r_{k+1} + \gamma(1 - d_k) [\min_{\theta} \bar{Q}_{\theta_1}(s_{k+1}, a_{k+1}), \bar{Q}_{\theta_2}(s_{k+1}, a_{k+1})] - \alpha \log \pi_{\phi}(a_{k+1} | s_{k+1}) \quad (9)$$

where r_{k+1} is a reward function, γ is discount factor, $\min_{\theta} \bar{Q}_{\theta_1}(s_{k+1}, a_{k+1}), \bar{Q}_{\theta_2}(s_{k+1}, a_{k+1})$ represent minimum of two target Q-networks, the algorithm reduces the risk of over-estimation by selecting the smaller value from the two targets. The $\alpha \log \pi_{\phi}(a_{k+1} | s_{k+1})$ function, α is a temperature coefficient that controls the trade off between exploration and exploitation. $\log \pi_{\phi}(a_{k+1} | s_{k+1})$ is the probability of taking the next action from the next state observed. For two critic networks, the critic loss function along with computing an EWC regularization term is the given by:

$$J_Q = \frac{1}{m} \sum_{k=1}^m \frac{1}{2} [y_k - Q_{\theta}(s_k | a_k)]^2 + \sum_i \frac{\lambda_{EWC}}{2} F_i(\theta_i - \theta_i^*)^2 \quad (10)$$

where y_k is target value, $Q_{\theta}(s_k | a_k)$ represents the predication value, which estimates future reward action from current state s_t . λ_{EWC} is a hyperparameter that controls the strength of penalty. A higher value indicates that the actor is more constrained to preserve close to previous knowledge. $F_i(\theta_i - \theta_i^*)^2$ function to compute how much the current critic network parameters deviate from previously learned parameters, using FIM to take important weight.

The updated actor network is represented as follows:

$$J_{\pi} = \frac{1}{m} \sum_{k=1}^m [\alpha \log \pi_{\phi}(a_k | s_k) - Q_{\theta}(s_k | a_k)] + \sum_i \frac{\lambda_{EWC}}{2} F_i(\phi_i - \phi_i^*)^2 \quad (11)$$

where $\alpha \log \pi_{\phi}(a_k | s_k)$ represents the entropy term and actor sampled actions from probability distribution leading to high Q values (expected reward), where $Q_{\theta}(s_k | a_k)$ represents the predication value, which estimates future reward. $\sum_i \frac{\lambda_{EWC}}{2} F_i(\phi_i - \phi_i^*)^2$ is a regulation term to compute how much the current critic network parameters deviate from previously learned parameters, and FIM means to take important weight.

To perform gradient updates for actor and critic networks:

$$\phi \leftarrow \phi - \eta_{\pi} \nabla_{\phi} J_{\pi} \quad (12)$$

$$\theta_1 \leftarrow \theta_1 - \eta_{Q_1} \nabla_{\theta_1} J_Q \quad (13)$$

$$\theta_2 \leftarrow \theta_2 - \eta_{Q_2} \nabla_{\theta_2} J_Q \quad (14)$$

Slowly update the parameters of target Q networks to prevent instability caused by rapidly changing targets using a soft update:

$$\bar{\theta}_1 \leftarrow \tau \theta_1 - (1 - \tau) \bar{\theta}_1 \quad (15)$$

$$\bar{\theta}_2 \leftarrow \tau \theta_2 - (1 - \tau) \bar{\theta}_2 \quad (16)$$

The structure of the WFQ continual-DRL algorithm as illustrated in Figure 5.

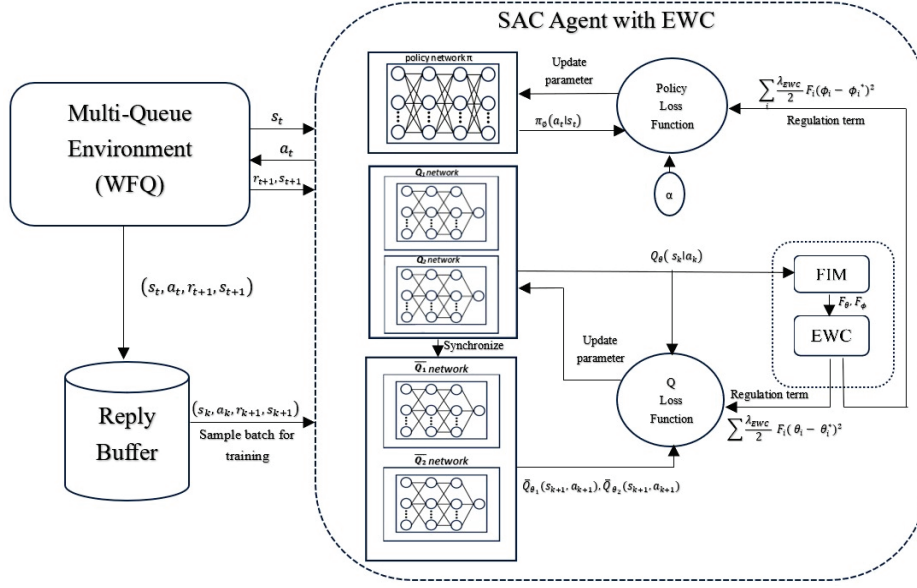


Figure 5. The structure of the WFQ continual-DRL algorithm.

The FIM is computed after finishing training on a task before starting on another task, including computing and storing the FIM. It is estimated by computing the squared gradients of the loss for the policy and critic networks with respect to their parameters, averaged over samples from the replay buffer size, where θ_i^* is the optimal parameter set after learning task A for future EWC loss computation. In our paper, the FIM was computed every 30 episodes, and then the parameters $(\theta_i^*, \mathcal{O}_i^*)$ and FIM were stored until the next 30 episodes, in an algorithm of SAC with EWC. This interval was chosen based on empirical test observe improve stability and adaptability; updating too often restricts the agent from learn new information, while updating too rarely weakens the term of regularization. In the WFQ continual-DRL framework, the 30-episode interval provides an effective tradeoff between preserving previous knowledge and learning new tasks. Figure 6 illustrates the steps of the adaptive algorithm of WFQ continual-DRL.

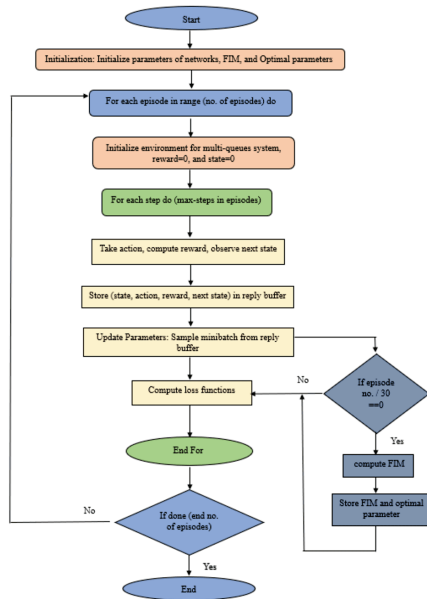


Figure 6. The diagram illustrates the steps of the WFQ continual-DRL algorithm.

The details of the WFQ continual-DRL algorithm are given in Algorithm 1.

Algorithm 1: The WFQ continual-DRL Algorithm (integration SAC with EWC)

1: Initialization:

Initialize the parameters of network, policy π , critics networks $Q_{\theta_1}, Q_{\theta_2}$, and target critics $\bar{Q}_{\theta_1}, \bar{Q}_{\theta_2}$, i.e., ϕ, θ_1, θ_2 , respectively.

Initialize Fisher Information Matrix F_θ and Optimal Parameters θ^*, ϕ^* .

2: For each episode do:

Initialize environment and initialize state s_0 .

Initialize episode reward to 0.

3: For each environment step do:

Select action: $a_t \sim \pi_\phi(a_t | s_t)$.

Execute action in environment and observe next state (s_{t+1}), reward (r_{t+1}), and done signal.

Store transition: $(s_t, a_t, r_{t+1}, s_{t+1})$ in replay buffer.

Store parameter θ^* , and compute Fisher Information Matrix F_θ every X episode.

4: Update Parameters:

Randomly sample minibatch of m transitions $(s_k, a_k, r_{k+1}, s_{k+1})$ from replay buffer.

Compute target Q-value using the minimum of two target Q-networks:

$$y_k = r_{k+1} + \gamma(1 - d_k) [\min_{\theta} \bar{Q}_{\theta_1}(s_{k+1}, a_{k+1}), \bar{Q}_{\theta_2}(s_{k+1}, a_{k+1})] - \alpha \log \pi_\phi(a_{k+1} | s_{k+1})$$

Update Critic networks $Q_{\theta_1}, Q_{\theta_2}$ minimizing loss:

$$J_Q = \frac{1}{m} \sum_{k=1}^m \frac{1}{2} [y_k - Q_\theta(s_k | a_k)]^2 + \sum_i \frac{\lambda_{EWC}}{2} F_i(\theta_i - \theta_i^*)^2$$

Where λ_{EWC} is the EWC regularization strength.

Update Actor network π_ϕ by minimizing the loss:

$$J_\pi = \frac{1}{m} \sum_{k=1}^m [\alpha \log \pi_\phi(a_k | s_k) - Q_\theta(s_k | a_k)] + \sum_i \frac{\lambda_{EWC}}{2} F_i(\phi_i - \phi_i^*)^2$$

Perform gradient updates:

$$\phi \leftarrow \phi - \eta_\pi \nabla_\phi J_\pi.$$

$$\theta_1 \leftarrow \theta_1 - \eta_{Q_1} \nabla_{\theta_1} J_Q.$$

$$\theta_2 \leftarrow \theta_2 - \eta_{Q_2} \nabla_{\theta_2} J_Q.$$

Soft update target networks:

$$\bar{\theta}_1 \leftarrow \tau \theta_1 + (1 - \tau) \bar{\theta}_1.$$

$$\bar{\theta}_2 \leftarrow \tau \theta_2 + (1 - \tau) \bar{\theta}_2.$$

5: Compute Fisher Information Matrix (FIM) Every X Episodes:

Compute gradient-based Fisher Information Matrix using log probabilities for the actor and critic losses:

$$F_\theta = \mathbb{E} \left[\left(\frac{\partial \log \pi_\phi(a | s)}{\partial \theta} \right)^2 \right]$$

Store optimal parameters θ^* and Fisher Information Matrix F_θ .

4.6. Real Traffic Implemented in Twelve Queues Experimental

This experiment was implemented with a real traffic dataset from a real telecommunication network to validate the efficiency of the proposed WFQ continual-DRL for real traffic applications. In this paper, the data set was collected from the Universidad Del Cauca, Popayán, Colombia network. Packet captures were performed at different morning and afternoon hours between April and June 2019. A total of 2,704,839 flow instances were gathered and stored [32].

Processing Real Traffic

In the first step of processing, a column of categories that consists of 24 categories and a column of their interarrival times are selected. Then, the category is split into multiple files.

After that, twelve categories from the total categories selected represent twelve sources, as differential arrival flows arrive at each queue independently. The selection is based on the most common and rich categories (chat, media, web, VoIP, software update, email, music, collaboration, cloud, network, download file transfer sharing, system). The interarrival rate for each category meeting is selected. According to the filtered data, removing rows indicates that invalid or missing data is present. The total dataset after processing is 1,265,319. Then, the arrival rate λ_i for each category is computed as in equation [33]:

$$\lambda = \frac{1}{T_A} \quad (17)$$

In addition to the initially selected twelve categories, the analysis was expanded to include the diversity of network traffic types. The 12 recent categories (Streaming, PRC, VPN, shopping, social media, data transfer, mining, game, remote access, database, video, unspecified) were also processed. The total dataset after processing is 128,458. Computing the arrival rate for each category, as in Equation (14), enables a more accurate characterization of traffic patterns, which is essential for simulating realistic queueing behavior. This measure enables us to compare performance metrics across a broader range of traffic resources, thereby improving the validation of the proposed adaptive algorithm, WFQ continual-DRL.

5. Results and Analysis

In this section, the proposed WFQ continual-DRL as the adaptive algorithm is evaluated in terms of optimizing bandwidth in the WFQ system. The experiments were implemented in Python 3.10 using PyTorch version: 2.6.0, a library for numerical computation in Python. TensorFlow is an open-source library facility for the computation and training of deep learning models. The network scheduling algorithm, represented as WFQ, manages the flow of packets across multi-queue systems. In this paper, the effectiveness of the proposed approach is evaluated through experiments conducted in an environment with the first eight and then twelve queues. A summary of the key parameters used in the implementation is given in Table 1.

Table 1. Key parameters of experimentally implemented.

Parameters	Value
Learning rate η	$\eta_\pi = 10^{-4}$, $\eta_{Q_1} = 10^{-3}$, $\eta_{Q_2} = 10^{-3}$
Update rate τ	10^{-2}
Discount factor γ	0.99
Entropy coefficient α	0.2
Regulation coefficient λ_{EWC}	λ_{EWC} for eight queues = 2271, λ_{EWC} for twelve queues = 2275
Reply buffer	eight queues = 10,000, twelve queues = 30,000
Batch size	eight queues = 128, twelve queues = 512
Optimizer	Adam

The holding cost is $h_i = \frac{1}{n}$, and the penalty for packet loss is $c_i = 1$, where $i = 1, 2, \dots, n$, and n is the number of queues. The reward function is represented as:

$$r(q, a) = -\sum_{i=1}^8 \left(\frac{1}{n} q_i + l_i \right)$$

5.1. Experiment 1: Eight Queues

In this section, we describe an experiment involving eight queues in the form of an arrival flow with arrival rates $\lambda_i = (0.03, 0.16, 0.2, 0.1, 0.08, 0.15, 0.06, 0.05)$ packets/s, where each arriving packet is classified and placed into its respective queue. The WFQ continual-DRL algorithm services queues in a manner that reflects their assigned weights. The interarrival times and service times were generated using a gamma distribution with a Squared Coefficient of Variation (SCV) of 0.8. The average packet size was 1 byte/packet, the buffers of queues were $K_1 = K_2 = \dots K_8 = 32$ packets, and the total finite bandwidth was 1 byte per second. The control policy allocated bandwidth weights satisfying $w_1 + w_2 + \dots + w_8 = 1$. The number of episodes was 1000, with a time length of 2000 s for each episode. The agent took action every 20 s.

When the random function had been implemented, λ_{EWC} was obtained, as explained in Section 2, based on six ranges, each of which was operated over 1000 episodes. Next, we set the six values in the grid to search for optimal values from them. The optimal value of λ_{EWC} was then obtained from the grid approach process, the FIM was computed every 30 episodes, and then the parameters θ_i^* and F_i were stored.

At the initial implementation step, a multi-queueing system consisting of eight queues was experimentally designed to evaluate the WFQ-DRL algorithm, as presented in [22]. This research aims to employ deep reinforcement learning for the allocation of dynamic bandwidth in router environments using a WFQ mechanism; each queue was modeled with identical arrival rates (λ) and buffer sizes, as detailed in the previous section. The implementation structure adopts a neural network containing policy and critic networks, as illustrated in Figure 3. The key hyperparameters of the network are summarized in Table 1.

After this, improving the WFQ-DRL algorithm by integrating continual learning allows the model to adapt over time without forgetting previously acquired knowledge. It primarily utilizes the EWC mechanism to preserve important network parameters as the model learns new tasks. The adaptive algorithm of the WFQ continual-DRL enables ongoing learning and adaptation to varying network conditions and objectives while maintaining the same structural configuration and parameter settings across the eight queues. This technique effectively maximizes the cumulative reward over time, ensuring robust performance in dynamic networking environments.

The average reward is shown in Figure 7 for the curve of SAC, and it can be seen that the curve exponentially converges after a 200-episode run, and the fluctuation range for average reward is between -10 and -6.5 , according to SAC with EWC curve of average reward for enhance algorithm resulting in consistent performance, and it can be seen that the curve exponentially converges after a 100-episode run. This indicates that the model was stable, with a steady range for the average rewards of between -2 and -1.5 , resulting in consistent performance. At the beginning of training, the controller typically explores different assignments as action strategies (bandwidth weights) through queues, leading to increased fluctuation. In this step, the controller has the potential to explore all the probabilities of processes to achieve an action that yields benefits. During the learning process, the controller exploits actions, resulting in a rise in the reward curve, and the policy gradually converges toward the optimal allocation bandwidth, reducing the variability in the reward curve. The rise indicates that the policy used in the WFQ continual-DRL algorithm improves performance and provides a suitable understanding of the environment dynamics. Eventually, the reward stabilizes during training, indicating that the controller consistently produces optimal performance rewards, thereby balancing exploration with exploitation while preserving past knowledge.

A comparison of the SAC and SAC with EWC rewards in Table 2, the standard SAC algorithm exhibits unsteady and fluctuating performance during training episodes from 50 to 1000. Initial rewards decline from -8.891 to -10.699 by episode 150, with only slight and inconsistent through episode 300. In middle training, rewards initially improve—reaching -7.381 at episode 650—but remain erratic. In the next steps, SAC achieves its best reward of -6.999 at episode 800; however, overall, SAC struggles to maintain stable and high rewards over time, reflecting challenges in long-term policy retention. The SAC with EWC consistently exceeds standard SAC across all training steps. At the same time, SAC shows unstable and suboptimal reward trends. The SAC with EWC maintains stable and significantly higher rewards, ranging nearly from -2.266 to -2 . The EWC-enhanced model exhibits better learning and improved convergence, especially during the middle to late training steps, where SAC rewards remain inconsistent; in contrast, SAC with EWC approach near-optimal values. This highlights the effectiveness of EWC in supporting stable and high-quality policy learning in DRL for the WFQ system over time.

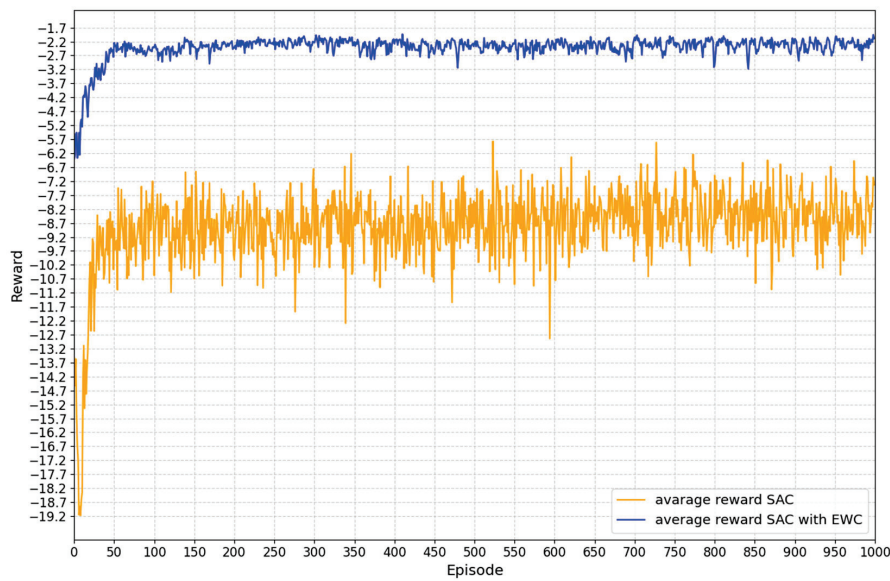


Figure 7. The average reward for eight queues in the WFQ DRL algorithm and the proposed WFQ continual-DRL algorithm.

Table 2. Comparison table of SAC and SAC with EWC rewards.

Episode Number	SAC Rewards	SAC with EWC Rewards
50	-8.891	-3.048
100	-10.579	-2.659
150	-10.699	-2.216
200	-10.029	-2.242
250	-9.929	-2.642
300	-10.286	-2.234
350	-8.322	-2.189
400	-7.872	-2.286
450	-9.479	-2.294
500	-8.709	-2.093
550	-10.896	-2.255
600	-9.008	-2.303

Table 2. Cont.

Episode Number	SAC Rewards	SAC with EWC Rewards
650	−7.381	−2.427
700	−7.226	−2.212
750	−7.311	−2.266
800	−6.999	−2.136
850	−8.165	−2.070
900	−8.011	−2.203
950	−7.897	−2.384
1000	−7.2	−2.048

Table 3 presents the MSE and MEA records for every 50 episodes during training for SAC with EWC in the WFQ system, using the WFQ continual-DRL algorithm, which enables an assessment of how well the environment supports adaptive training and learning. The MSE value is 0.612 at episode 50, which rises slightly to 0.639 and exhibits more fluctuations. At the start of training, the agent typically explores different assignment procedures as actions (bandwidth weights) for the queues, resulting in increased fluctuations in the system. That point to the agent explores all possible processes of the action it is learning to reach an action that yields benefits. Then, it decreases to 0.597 at episode 150, indicates that the model was successfully adjusted and improved its predictions. These fluctuations are common in reinforcement learning and reflect the balance between exploration and convergence.

It can be observed that the MSE values steadily decrease from 0.547 at episode 200 to 0.445 at episode 600, suggesting steady learning and improvement in the model's predictions. The results demonstrate successful training with continuous improvement in MSE between episodes 650 and 1000, and the model achieves its best performance while maintaining stable operation within these values. The slight increase in the MSE values after episode 900 remains within an acceptable range. The models reflect more complex or bursty traffic patterns that appear in training. Yet, the values of MSE remain within an acceptable range (near 0.47), meaning that the model has reached stability and effectively minimizes prediction error, although it has increased variability. The training process is well optimized, resulting in high accuracy.

From the column of MAE values, we see that this parameter is 0.583208 at episode 50, fluctuates slightly, and then improves to 0.535 at episode 300. Between episodes 400 and 700, the error remains controlled and stable, with a decline, indicating strong generalization capabilities. This suggests that the model is learning effectively and reducing prediction errors over time. The lowest MAE values are achieved at episode 1000 (0.510144), indicating that the model performs optimally towards the end of training. The learning process of the WFQ continual-DRL algorithm was efficient, as the estimation of errors was reduced over time, resulting in a well-optimized algorithm for dynamic traffic management.

Table 3. The MSE and MAE results for eight queues.

Episode Number	MSE	MAE
50	0.612392	0.583208
100	0.639665	0.614071
150	0.597392	0.601428
200	0.547177	0.572127

Table 3. Cont.

Episode Number	MSE	MAE
250	0.505351	0.553376
300	0.467858	0.53599
350	0.466958	0.539003
400	0.418296	0.510907
450	0.455936	0.532404
500	0.464072	0.53763
550	0.491030	0.552584
600	0.445696	0.526634
650	0.426627	0.513851
700	0.407303	0.502012
750	0.405632	0.500403
800	0.413735	0.506038
850	0.393311	0.492884
900	0.382233	0.486044
950	0.39783	0.497124
1000	0.419463	0.510144

In Figure 8, during training, the model with eight queues of WFQ initially has an MSE curve that rises to 0.7, indicating the model's adaptation to changing queue dynamics. Then, it decreases and fluctuates within the range of 0.35–0.45, indicating a change into a steady-state. These bounded fluctuations confirm robust error minimization and model stability under a dynamic environment.

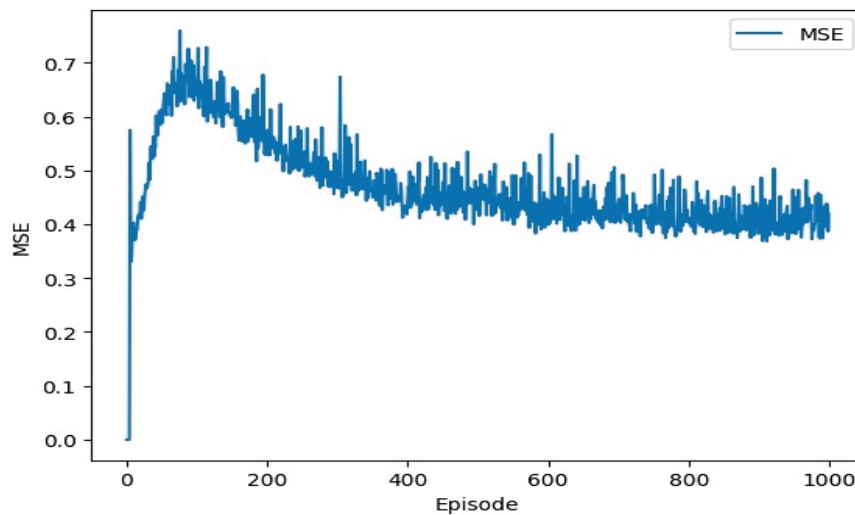


Figure 8. The Mean Square Error (MSE) of eight queues.

According to Figure 9, the MAE rises above 0.6, indicating an early training prediction error, and then gradually declines and stabilizes, with fluctuations in the range of nearly 0.5–0.55, reflecting steady fluctuations and demonstrating the model's consistent performance in its dynamic state.

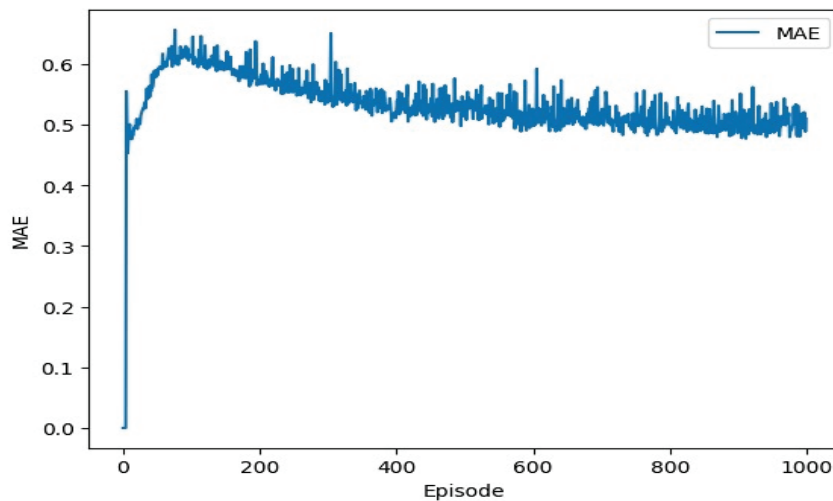


Figure 9. The Mean Absolute Error (MAE) of eight queues.

Table 4 represents a comparison between the RMSE and Jain's Index of the WFQ-DRL algorithm and the enhanced WFQ continual-DRL algorithm. SAC integration with EWC yields lower average and minimum RMSE and appears to have better prediction accuracy. The SAC with the EWC model achieved higher fairness in resource allocation compared to the baseline SAC, indicating that the enhanced model maintains a better allocation bandwidth among network flows. Jain's Index confirms that incorporating EWC leads to more stable and fair decision-making during training in dynamic environment.

Table 4. The comparison of RMSE and Jain's Index for eight queues.

Episode Number	RMSE for SAC	RMSE of SAC with EWC	Jain's Index of SAC	Jain's Index of SAC with EWC
50	0.115270	0.583208	0.451633	0.511600
100	0.095618	0.614071	0.517129	0.618827
150	0.082627	0.601428	0.548154	0.661623
200	0.060702	0.572127	0.590348	0.765714
250	0.084624	0.553376	0.557582	0.660985
300	0.057694	0.53599	0.673834	0.775280
350	0.054768	0.539003	0.702753	0.803767
400	0.046266	0.510907	0.645746	0.828189
450	0.050000	0.532404	0.668195	0.762000
500	0.052254	0.53763	0.668355	0.790373
550	0.062156	0.552584	0.640962	0.784969
600	0.060979	0.526634	0.664703	0.757960
650	0.059234	0.513851	0.654715	0.767560
700	0.054519	0.502012	0.648057	0.777945
750	0.061864	0.500403	0.666093	0.769028
800	0.068272	0.506038	0.680141	0.737892
850	0.069809	0.492884	0.658749	0.743764
900	0.068197	0.486044	0.630024	0.732841
950	0.065631	0.497124	0.650287	0.797796
1000	0.058266	0.510144	0.66	0.803811

Figure 10 shows a comparison between the curve of RMSE computed for SAC and SAC with EWC. After 200 episodes, the curve of SAC decreased to a stable range between 0.07 and 0.09. According to the curve for SAC, the EWC declines after 200 episodes, stabilizing between 0.05 and 0.07. The enhanced SAC with EWC exhibits higher learning performance, achieving a lower and more stable RMSE compared to the SAC. This suggests that EWC helps mitigate catastrophic forgetting learning, especially in complex or dynamic environments of the WFQ system. The curve of enhancement appeared to decrease in RMSE, reflecting improved steadiness and adaptability in decision-making. Lower RMSE indicates better accuracy in predicting actions in dynamic queue management (resource allocation) with respect to optimal resource allocation computed in the Fairness function.



Figure 10. The comparison of the Root Mean Square Error (RMSE) of eight queues.

As shown in Figure 11, the comparison of the Jain's index between SAC and SAC with EWC reveals that the results show the curve of SAC after a 200-episode rise and steady between 0.6 and 0.7, while the curve of SAC with EWC remains stable after a 200-episode rise and remains steady between 0.7 and 0.8. Whenever the Jain's Index value is close to 1, it indicates excellent fairness, meaning better exploitation of resources.



Figure 11. The comparison table of Jain's Index for eight queues.

Complexity Analysis

The computation and comparison of the time complexity of SAC were implemented in the WFQ-DRL algorithm [22], and SAC with EWC was implemented in the WFQ continual-

DRL algorithm in this paper. Analysis of the execution time for both average step and overhead represented by periodic FIM in DRL as a continual process, providing insight into the tradeoff between learning stability of performance model and computational cost.

For policy (π) network structure:

Input layer $8 \rightarrow 64$

Hidden layer $64 \rightarrow 64$

Two output layers (μ, σ) $64 \rightarrow 8$ for each layer

Calculating the parameter (P) for the policy (π) structure:

For the first fully connected layer $O(8 \times 64) \rightarrow O(512)$

Second fully connected layer $O(64 \times 64) \rightarrow O(4096)$

Output fully connected layer $2 \times (64 \times 8) \rightarrow O(1024)$

Total P for policy network (π): $O(512 + 4096 + 1024) = O(5632)$

Complexity model:

For policy network (π) is $O(1)$ because the dimension is constant.

For the critic network (Q) structure:

Input layer $(8 + 8) \rightarrow 16$

Hidden layer $64 \rightarrow 64$

The output layer is a $64 \rightarrow 1$ output layer in a neural network

Calculating the parameter (P) for the policy (π) structure:

For the first fully connected layer $O(16 \times 64) \rightarrow O(1024)$

Second fully connected layer $O(64 \times 64) \rightarrow O(4096)$

Output fully connected layer $(64 \times 1) \rightarrow O(64)$

Total (P) for one critic network (π): $O(1024 + 4096 + 64) = O(5184)$

Total P for both critic networks (Q1, Q2) and policy network (π)

$5632 + 2 \times 5184 = P(16,000)$

Since the network dimension remains constant, the per-step computational complexity remains $O(1)$. Over the course of N training steps (100 steps per episode), the total complexity linearity is $O(N)$; although, for WFQ continual-DRL using SAC with EWC step cost remains $O(1)$, an additional overhead of $O(P)$ arises for 30 episodes due to FIM, highlighting the computational trade-off between standard DRL algorithm and enhancement algorithm introduced by continual learning.

For the complexity model:

For critic network (Q): $O(1)$, because the dimension is constant.

Overall complexity for N training network steps:

Total time complexity $O(N \times 1) = O(N)$.

One episode has 100 steps, which is $O(100)$, where $N = 100$.

For 1000 episodes (100,000) where $N = 100,000$, 100 steps for each episode.

Where FIM computes every 30 episodes, 100 steps per episode, $P = 16,000$ parameters. Table 5 illustrates the complexity time comparison between SAC and SAC with EWC algorithm.

Table 5. Complexity time comparison between SAC and SAC with EWC.

Operation	SAC	SAC with EWC	Notes
Time per step	$O(1)$	$O(1)$	Constant time for each step in each algorithm
Time per episode cost	$O(100)$	$O(100) + O(16,000)/30$	FIM cost is computed every 30 episodes
Time per 30 episodes	$O(3000)$	$O(3000) + O(16,000)$	One full FIM computation

For the SAC algorithm, 1 step requires 0.004 s, so for 100 steps, the total time is 0.4 s.

For (SAC with EWC), FIM step time = 0.00294 s every 30 episodes. Total 30 episodes $(30 \times 0.4) + 0.00294 \text{ s} = 12.00294 \text{ s}$. Experimental time comparison between SAC and SAC with EWC algorithm is shown in Table 6.

Table 6. Experimental time comparison between SAC and SAC with EWC.

Operation	SAC	SAC with EWC
Time per step	O (0.004)	O (0.004)
Time per episode cost (100 steps)	O (0.4)	O (0.4)
Time per 30 episodes	O (12)	O (12.00294)

FIM updates every 30 episodes (3000 steps), making the overhead negligible, which is only added at a rate of 0.00294 s every 30 episodes (3000 steps). Total time added when computing FIM through 1000 episodes about 33 times $= 33 \times 0.000294 \text{ s} = 0.009702 \text{ s}$. By observation, the overhead caused by using FIM in the WFQ continual-DRL algorithm is negligible compared to the performance enhancement, indicating that the new algorithm, WFQ continual-DRL, functions effectively.

5.2. Experiment 2: Twelve Queues

The experiment was extended by increasing the number of queues to twelve to evaluate the robustness and scalability of the proposed WFQ DRL-continual approach. This expansion enabled a more comprehensive evaluation of the algorithm's capacity to allocate bandwidth dynamically in a complex network environment. The twelve arrival flows had arrival rates $\lambda_i = (0.0713, 0.0563, 0.0840, 0.0395, 0.0879, 0.0873, 0.0631, 0.0453, 0.0257, 0.0869, 0.0407, 0.1420)$ packets/s, and the interarrival times and service times were generated using a gamma distribution with an SCV of 0.8. The average packet size was assumed to be 1 byte per packet, and the total bandwidth was 1 byte/s. An individual queue has a limited buffer $K_1 = K_2 = \dots K_{12} = 32$ packet. The control policy allocated bandwidth weights to $w_1 + w_2 + \dots + w_{12} = 1$. The number of episodes was 1000, with each episode having a time length of 2000 s. Every 20 s, the agent took action.

After the random function was performed, λ_{EWC} was obtained as explained in Section 2, running over a range of 1000 episodes to find an optimal λ_{EWC} the FIM was computed every 30 episodes, and F_i were stored with the parameters θ_i^* .

A plot of the run-average reward for SAC with the EWC algorithm in the WFQ continual DRL algorithm is shown in Figure 12. It can be seen that convergence occurs after 200 episodes, and the range of reward fluctuation stabilizes between -2.5 and -1.9 , resulting in consistent performance. At the beginning of the reward curve, the agent typically explores different assignment strategies (bandwidth weights) for the queues, resulting in more fluctuation. This means that the controller explores all possible processes of the action it is learning in order to achieve an action that yields benefits. As learning progresses, the agent exploits the probabilities of action, resulting in an increase in the reward curve; the policy then gradually converges toward the optimal allocation bandwidth, reducing the variability in the reward curve.

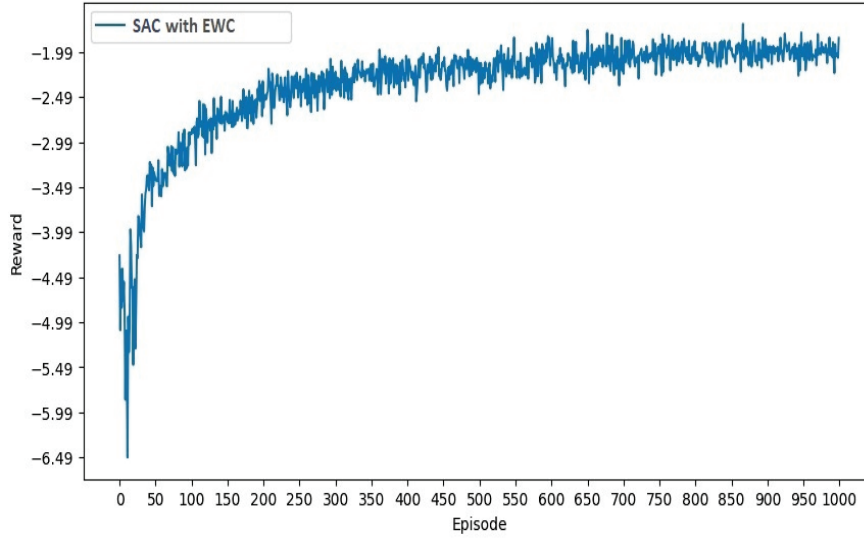


Figure 12. The average reward for twelve queues.

5.3. Training Real Traffic

In the experiment, twelve arrival flows with various flows (categories) arrive at the WFQ system of twelve queues. Twelve category groups of traffic data were used to train the agent. The average arrival rates of twelve sources in these twelve categories are shown in Table 7.

Table 7. The average arrival rate of the first twelve categories.

Queues	Categories	Average Arrival Rates (Packets/s)
1	Chat	0.047179
2	Media	0.038448
3	Web	0.031438
4	VoIP	0.030870
5	software update	0.023940
6	Email	0.023906
7	Music	0.022780
8	collaboration	0.021974
9	Cloud	0.015814
10	Network	0.011008
11	Download File Transfer, File Sharing	0.005254
12	System	0.002918

The remaining twelve arrival flows with different flows (categories) arrive at the WFQ system of twelve queues. Twelve category groups of traffic data were used to train the agent. The average arrival rates of twelve sources in these twelve categories are shown in Table 8.

The reward function is calculated as an equation:

$$r(q, a) = - \sum_{i=1}^8 \left(\frac{1}{n} q_i + l_i \right)$$

The episode time length, buffer size, and the total bandwidth resources are $T = 2000$ s, $K_1 = K_2 = \dots K_{12} = 32$, bandwidth = 1 byte/s, correspondingly. Running the 1000 episodes to train the agent by the SAC with EWC algorithms of twelve queues.

A plot of the run average reward comparison of the twelve queues simulation curves, along with two datasets according to SAC with the EWC algorithm in the WFQ continual-DRL algorithm, is shown in Figure 13. For the first simulation curve, it can be seen that convergence occurs after 200 episodes, and the range of reward fluctuation stabilizes between -2.5 and -2 , resulting in consistent performance. At the beginning of the reward curve, the agent typically explores different assignment strategies (bandwidth weights) for the queues, resulting in more fluctuation. This means that the controller explores all possible processes of the action it is learning in order to achieve an action that yields benefits. As learning processing, the agent exploits the probabilities of action, resulting in an increase in the reward curve; the policy then gradually converges toward the optimal bandwidth allocation, reducing the variability in the reward curve.

Referring to the convergence of the curve average reward after around 200 episodes for the first 12 categories of the real traffic 2 dataset, the range of the fluctuation curve gradually increases and shows steady fluctuation between -2.5 and -1.5 , indicating a stable performance. According to the recent 12 categories of the real traffic two of the dataset, the convergence of the average reward curve is before 200 episodes, and the range of the fluctuation curve is between -1.8 and -1 , reducing the fluctuation in the curve.

Table 8. The average arrival rate of the second twelve categories.

Queues	Categories	Average Arrival Rates (Packets/s)
1	streaming	0.051215
2	RPC	0.134050
3	VPN	0.017920
4	shopping	0.019674
5	social media	0.078182
6	data transfer	0.205775
7	Mining	0.016681
8	Game	0.030445
9	remote access	0.040550
10	database	0.016725
11	Video	0.026934
12	unspecified	0.036511

Table 9 presents the results of computing the MSE and MEA values for the twelve-queue scenario, which are used to measure the effectiveness of the environment in supporting adaptive training. The column containing the MSE values shows that at the start of training, the MSE is initially 0.8680; it then undergoes fluctuations before increasing to 1.0482 as the environment is explored and gradually improves during training. The slight variation suggests that the model is still learning, adjusting parameters, and refining predictions. By episode 300, the error is relatively steady compared to the prior fluctuations, indicating better convergence. In the intervening episodes, from 300 to 700, the error is more stable than the previous fluctuations, indicating better convergence. A noteworthy development is observed in the value of the MSE, which drops to 0.5302 at episode 400 and remains low, at approximately 0.5263, until episode 450. In the steady state, from episodes 400 to 700, the MSE values exhibit strong consistency over time, indicating that the model makes predictions that are consistent and accurate. Small fluctuations are expected during training due to the dynamic nature of the environment. The lowest MSE value of 0.4909 is observed at episode 1000, confirming the optimal performance of the model. The results for the last 200 episodes suggest that the model has achieved high

accuracy and stability, as the variation between episodes 750 and 950 remains controlled. From the column of MEA values at the start of training, showing that the MEA is 0.6437 at episode 50 and increases slightly to 0.6765 at episode 100. A slight increase is expected when the model adjusts at episode 300 during training, and the MEA value stabilizes, indicating that the model is starting to learn effectively. The MEA value decreases from episode 400 (0.5754) to episode 750 (0.5640), indicating that the model is learning to make more accurate predictions. The fluctuation between episodes 600 and 700 (between around 0.5994 and 0.6057) suggests that the model exhibits normal and healthy training behavior. At episode 1000, the lowest MAE value of 0.5528 is recorded, indicating optimal performance. From episode 850 (MAE values 0.5543 to 0.5528), the model maintains a constant learning rate with minimal fluctuations.

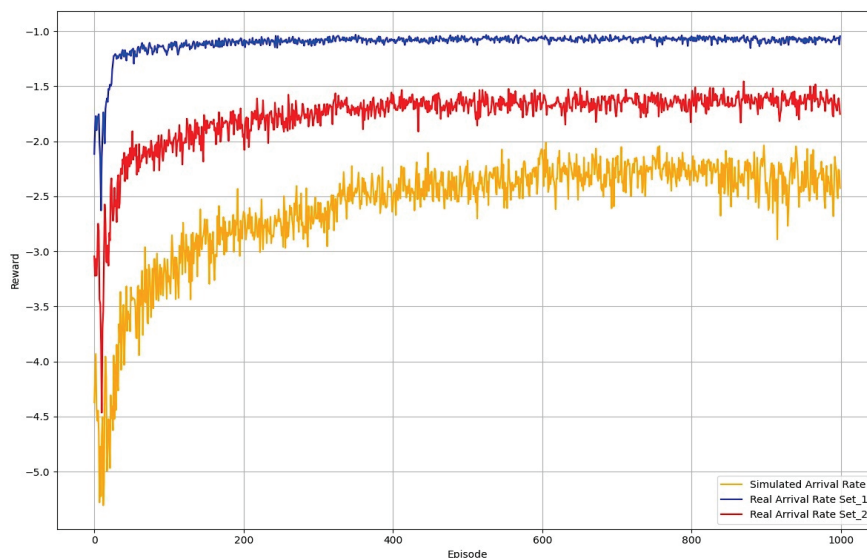


Figure 13. The comparison of 3 curves for twelve queues: simulated, real traffic 1, and real traffic 2 for dataset.

Table 9. Comparison of MSE and MAE values for all twelve queue scenarios.

Episodes	Twelve Queues of Simulation		Twelve Queues of Real Traffic 1		Twelve Queues of Real Traffic 2	
	MSE	MAE	MSE	MAE	MSE	MAE
50	0.868022	0.643713	0.533128	0.573652	0.788154	0.696236
100	1.048262	0.676523	0.508042	0.557808	0.724009	0.672481
150	0.868753	0.659184	0.441298	0.517397	0.634688	0.629618
200	0.670131	0.648144	0.412639	0.499913	0.614650	0.617652
250	0.797906	0.61379	0.392221	0.487977	0.551309	0.585325
300	0.88304	0.608288	0.344662	0.458005	0.475661	0.542329
350	0.727278	0.679143	0.317199	0.440590	0.511582	0.560313
400	0.530156	0.575429	0.356414	0.468477	0.423877	0.512770
450	0.52626	0.572744	0.346210	0.461428	0.409414	0.502853
500	0.570297	0.599337	0.321819	0.445289	0.525351	0.571535
550	0.769053	0.702485	0.351930	0.466554	0.496567	0.555185
600	0.498518	0.557347	0.350638	0.465073	0.423494	0.509751
650	0.583851	0.59942	0.310275	0.437058	0.429814	0.515980
700	0.589354	0.6057	0.322598	0.445472	0.400673	0.499607
750	0.51114	0.564016	0.353146	0.467203	0.510712	0.560117

Table 9. Cont.

Episodes	Twelve Queues of Simulation		Twelve Queues of Real Traffic 1		Twelve Queues of Real Traffic 2	
	MSE	MAE	MSE	MAE	MSE	MAE
800	0.535042	0.577101	0.310399	0.437901	0.390746	0.489732
850	0.496227	0.55432	0.349823	0.463259	0.381719	0.485158
900	0.544107	0.58117	0.308214	0.435233	0.660230	0.636922
950	0.515378	0.565431	0.328040	0.449288	0.423070	0.512251
1000	0.490912	0.552778	0.331264	0.453266	0.411846	0.505342

According to the evaluation of the twelve-queue real traffic for the first twelve categories using MSE and MAE metrics, there is an improvement in model stability over time. According to the MSE, a consistent decline across the training episodes indicates advanced improvement in the model's predictive accuracy. Beginning at 0.533 in episode 50, the MSE declines to 0.317 in episode 350, reflecting effective start learning. Between episodes 400 and 700, the MSE values fluctuate between 0.322 and 0.356, suggesting a stabilization phase where the model fine-tunes its parameters. In the final episode, from 750 to 1000, the MSE continues its downward trend, reaching 0.331, which highlights the model's capacity to minimize errors over time. The MAE shows below throughout the training process, suggesting improvements in the model's average prediction accuracy, which starts at 0.574 in episode 50, then decreases to 0.441 in episode 350, indicating successful initial learning. Between episodes 400 and 700, the MAE values exhibit small fluctuations, ranging from 0.445 to 0.468, which reflects the model's efforts to refine its predictions and handle more nuanced patterns in the data. From episode 750 to 1000, the MAE continues to decline, reaching 0.453, which underscores the model's steady improvement in minimizing average absolute errors.

Referring to the twelve-queue real traffic for the second twelve categories using MSE and MAE metrics, the MSE indicates a decreasing trend through the training process, showing advanced improvement in the model's predictive accuracy. Beginning at 0.788 in episode 50, the MSE declines to 0.511 by episode 350, which is effective in early-stage learning. Between episodes 400 and 700, the MSE values fluctuate within a range of around 0.400 to 0.525, suggesting a stabilization stage where the model fine-tunes its parameters. In the last phase, from 750 to 1000, the MSE continues its downward trend, reaching 0.412, which indicates the model's capacity to minimize errors over time. The MAE for real traffic of the recent twelve categories also displays a decreasing trend throughout the training process, indicating improvements in the model's average prediction accuracy. Initiating at 0.696 at episode 50, the MAE decreases to 0.560 at episode 350, indicating successful initial learning. Between episodes 400 and 700, the MAE values exhibit minor fluctuations, ranging from 0.499 to 0.571. From episode 750 to 1000, the MAE decreases, reaching 0.505, which highlights the model's steady improvement in minimizing average absolute errors.

In conclusion of comparison Table, the simulation has higher MSE and MAE than real traffic 1 and traffic 2 for twelve queues of the dataset; it generalized well to real-world conditions. A G/G/1/K WFQ system parallel queuing model with a gamma distribution was formulated to simulate complicated network traffic. That indicates that the simulation accurately captured realistic traffic variability and system conditions. Thus, the WFQ scheduler, as implemented in continual and DRL framework, demonstrated robust performance, validating its effectiveness.

According to Figure 14, the MSE curve initially rises to 0.8, indicating prediction error during early adaptation. In the last episodes of training, it then decreases and enters a steady fluctuating range nearly between 0.5 and 0.7. This bounded variability indicates that the model has stabilized and maintains reliable performance under dynamic multi-queue conditions, although in extended queues.

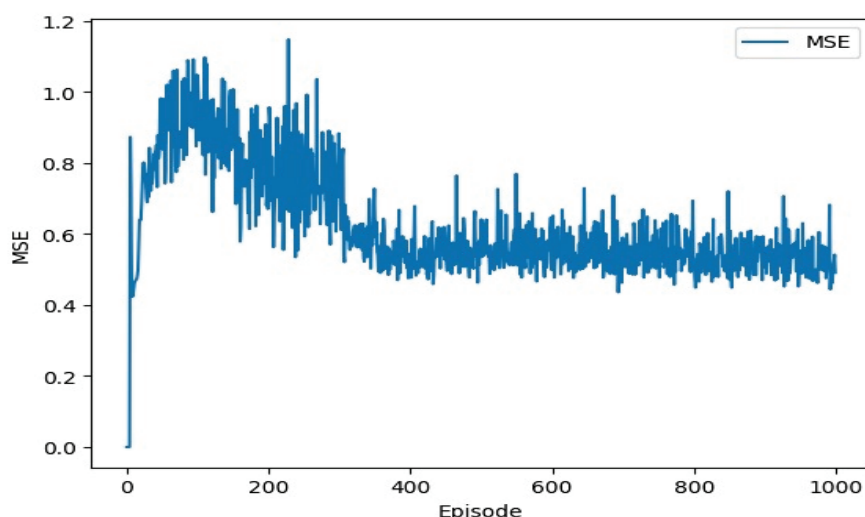


Figure 14. The Mean Square Error of twelve queues.

As shown in Figure 15, during the early training, the MAE curve rises to 0.7 in the initial step, indicating a high error. In the last episodes, it then drops and stabilizes within a fluctuating range between nearly 0.55 and 0.65. This behavior suggests the model's ability to generalize while maintaining consistent performance in a dynamic environment.

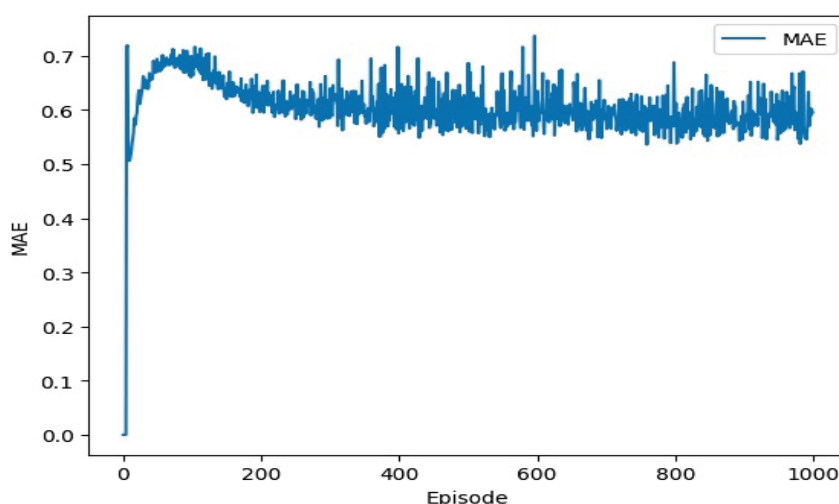


Figure 15. The Mean Absolute Error of twelve queues.

According to Figure 16, the MSE curve for the twelve-queue system trained on real traffic exhibits a high initial error due to the complexity of the traffic patterns. Still, it gradually decreases and stabilizes between 3.0 and 3.5, indicating that the model has effectively learned and adapted to real-world traffic.

For Figure 17, the MAE curve begins with a high error, reflecting meaningful early prediction deviations. As training advances, the error steadily declines. It stabilizes between nearly 4.0 and 4.5, demonstrating bounded fluctuations. That indicates the model has adapted well to real traffic dynamics and supports robust performance.

Regarding Figure 18, the MSE curve for the second real traffic dataset starts high, reflecting the initial instability of the prediction. It gradually decreases and stabilizes within the range of 4.5 to 5.5, showing enhanced model learning. This trend suggests that the model is evolving to accommodate more diverse traffic patterns over time. Despite the dynamic challenges, the error range suggests reliable convergence and robust performance.

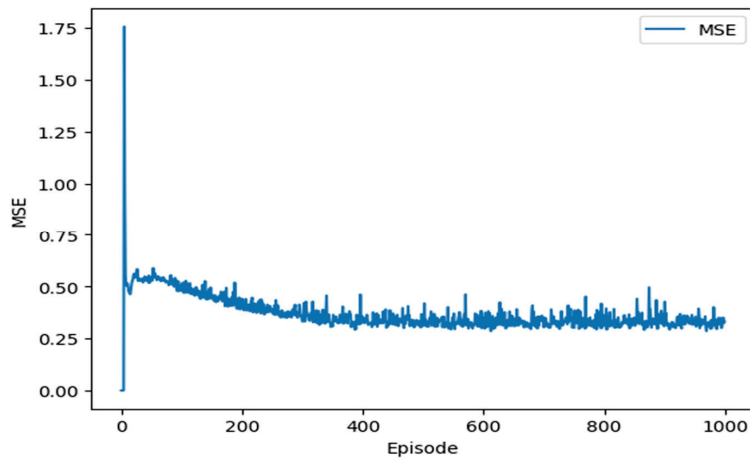


Figure 16. The Mean Square Error (MSE) of twelve queues for first real traffic.

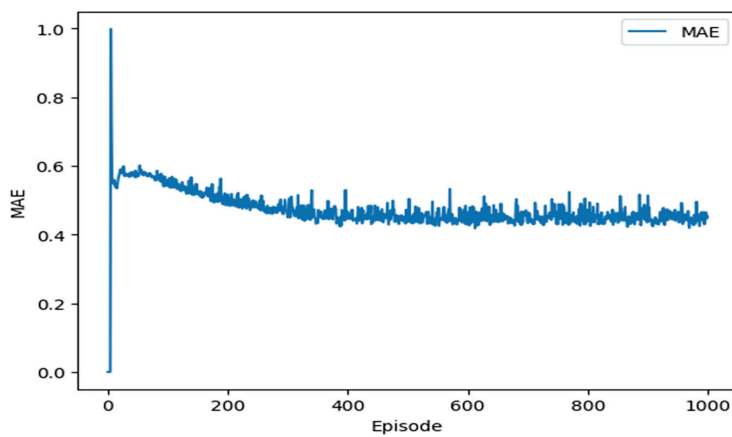


Figure 17. The Mean Absolute Error (MAE) for twelve queues for the first real traffic.

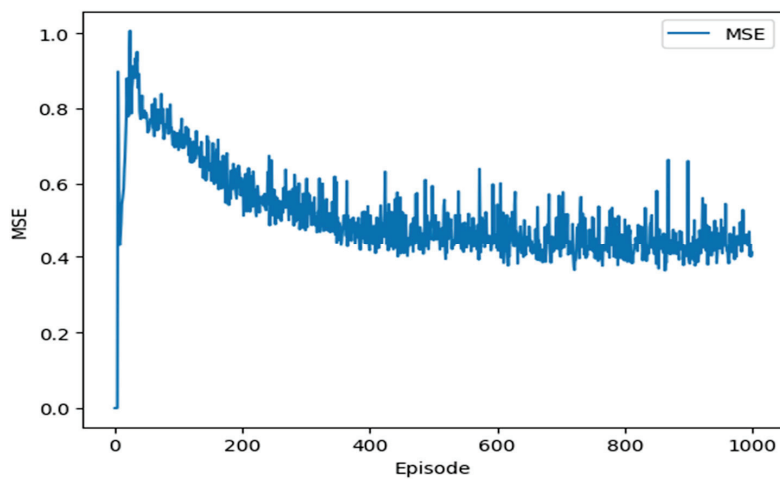


Figure 18. The Mean Square Error (MSE) of twelve queues for second real traffic.

Referring to Figure 19, the MAE curve for the second real traffic dataset initially increases sharply, exhibiting poor early prediction accuracy. As training progresses, the error decreases and stabilizes within a narrow range of 0.5 to 0.6. This steady fluctuation suggests the model has adapted better to the traffic dynamics. The bounded MAE range reflects consistent and robust performance under real-world dynamics.

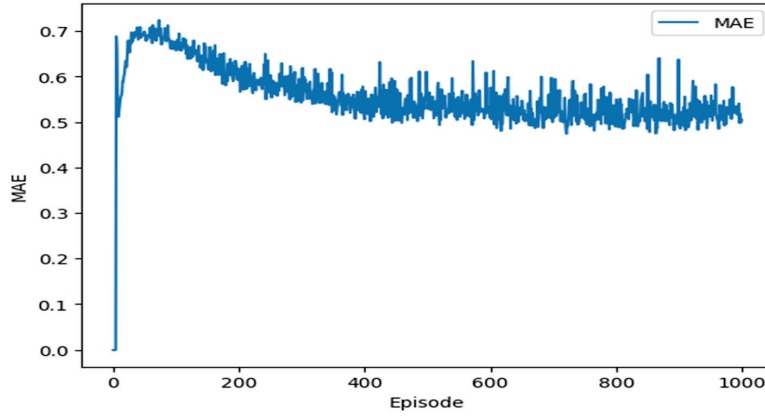


Figure 19. The Mean Absolute Error (MAE) of twelve queues for second real traffic.

The processing with statistical analysis addresses bursty real traffic for all categories, after which the dataset is equal to 1,393,777 [33].

Regarding the Square Coefficient of Variance CV^2 in the real traffic data, computed in equation [33]:

$$CV^2 = \frac{\sigma}{m} \quad (18)$$

where σ represent standard deviation, and m represents the mean. As illustrated in Figure 20, a higher CV^2 indicated that the real traffic data arrival patterns were more unpredictable and random. High variance reflects the bursty and unpartable nature of real traffic conditions.

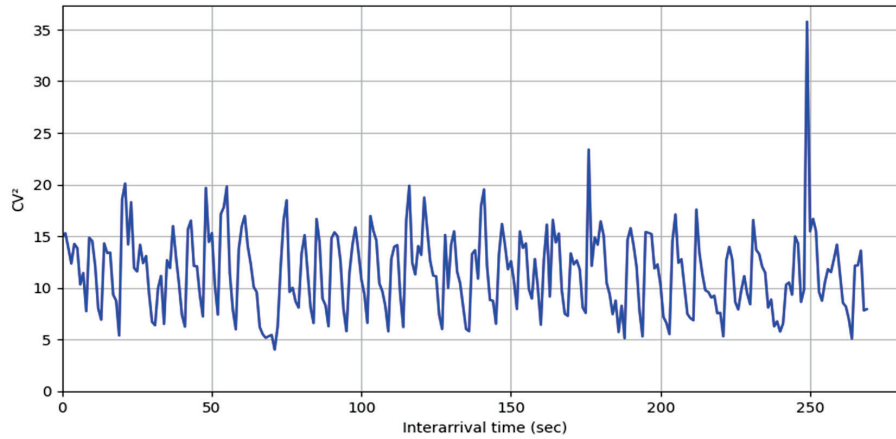


Figure 20. Square Coefficient of Variance (CV^2) in the real traffic data.

To compute correlation, the copy of interarrival time will be shifted to one position, and the correlation coefficient for real traffic will be computed in the equation of correlation coefficient [34]. The results shown in Figure 21 revealed that the real traffic data appear to have a more dynamic pattern in its correlation. The decline of correlation curves indicates more influence. The statistics clearly showed patterns of bursty traffic.

This paper presented knowledge that improved this algorithm by combining the continual learning technique with the DRL algorithm for the WFQ system. The development of the WFQ continual-DRL algorithm was achieved by contributing a continual learning method represented by the EWC technique, which helps to overcome catastrophic forgetting by remembering old tasks, selectively slowing down learning on the parameters important for those tasks, and improving the performance over time, which is combined with advanced deep reinforcement in the form of the SAC algorithm. In addition to developing this algorithm, the evaluation of the algorithm was expanded by increasing the number of queues to twelve. This expansion was considered to assess the scalability and ro-

business of the model in terms of managing higher traffic loads and more complex network conditions. The aim was to maximize the cumulative reward, and this approach achieved an improvement of approximately 76% in performance compared to the baseline reward curve in an experiment involving eight queues for WFQ-DRL. This finding demonstrates the effectiveness of our approach in regard to maintaining stable learning and adapting to new tasks while retaining previously learned knowledge. For the expansion in a second experiment involving twelve queues, the average reward curve shows an enhancement in performance by increasing reward with smother fluctuation, indicating that this new approach is more effective even with an expanded environment, is more robust, and is suitable for the nature of the queueing system, as it carries out dynamic bandwidth allocation control among WFQ parallel queues. The model shows higher reward than compared to simulation of twelve queues. This suggests that the model adapts well in real world traffic, and the result highlights the model's effectiveness under realistic traffic environments.

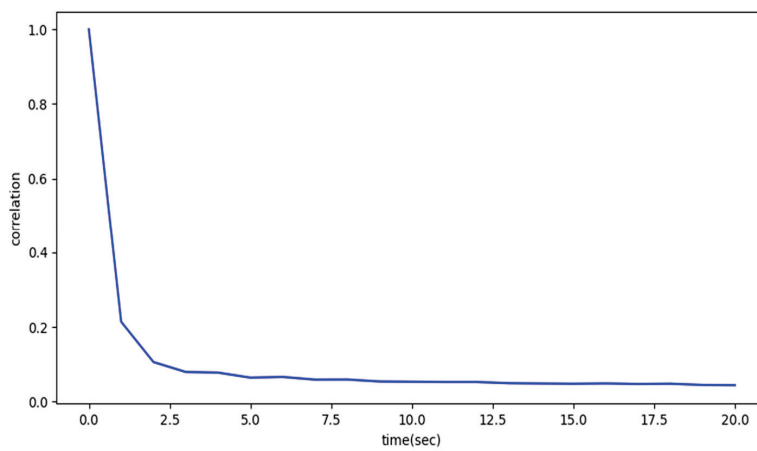


Figure 21. Correlation of real traffic data.

6. Discussion

The proposed approach in the form of the WFQ continual-DRL algorithm and its implementation in the system model to achieve dynamic bandwidth allocation leads to the following observations:

1. By dynamically adapting the weights assigned to each queue, the system can react to irregular traffic patterns, optimize bandwidth utilization, and maintain service quality.
2. Training the agent encourages the development of adaptive policies for weight assignment in WFQ systems, enabling real-time optimization based on monitored network states through interaction with the network environment.
3. Traffic patterns can change over time in dynamic network environments, meaning that the WFQ continual-DRL algorithm must continually adapt to these changes. Without an approach such as EWC, the agent may forget previously learned optimal policies, a phenomenon known as catastrophic forgetting. The EWC trains a regularization term added to the DRL loss function by penalizing significant deviations from parameters essential to previously learned tasks. This ensures that new learning does not overwrite valuable existing knowledge and maintains consistent performance across varying networks. In scenarios involving eight and twelve queues, a higher value of λ_{EWC} strengthens regularization, meaning that important knowledge persists as complexity increases. This improves performance by balancing adaptability and memory retention, ensuring the effectiveness of the WFQ continual-DRL learning algorithm.

4. Expanding the number of queues enables a more fine-grained allocation of bandwidth, representing a test of the adaptability of the WFQ continual-DRL approach in optimizing resource distribution under dynamic traffic scenarios. The results indicate that the model successfully generalizes to larger queue sets while maintaining stable performance, as reflected in the stabilization of the reward.
5. When the number of queues is expanded, the parameters of the neural network structure in SAC-EWC need to be tuned by increasing the number of neurons, batch size, and buffer size, which enables the model to capture more complex patterns in the data. As a result, it enables the model to make more informed decisions, thereby improving performance and reward.
6. The implementation of EWC as one mechanism of continual learning requires additional computational resources due to the regularization term in the loss function. It is, therefore, essential to balance the benefits of EWC with the available computational capacity. In this investigation, a random function was used to select an appropriate strength for the EWC regularization term. Appropriate tuning ensures that the model maintains stability without restricting its ability to learn new patterns in the network traffic.
7. The steady decrease in MSE and MAE in both the eight- and twelve-queue systems suggests effective learning and enhanced prediction accuracy, ensuring reliable decision-making and consistent performance.
8. The RMSE results clearly indicate that the SAC enhanced with EWC yields a lower prediction error compared to the baseline SAC model. After approximately 200 episodes, while the standard SAC stabilizes in the range of 0.07 to 0.09, the enhanced model further reduces the RMSE and stabilizes between 0.05 and 0.07. This improvement demonstrates the effectiveness of EWC in preserving important learned parameters, resulting in more accurate and stable policy predictions over time.
9. In addition to improving prediction accuracy, as shown by the RMSE results, SAC with EWC also achieves a higher Jain's Fairness Index, indicating better resource allocation across various network flows. The fairness values are closer to 1, not only showing the enhanced model performance but also maintaining fairness in scheduling decisions. The importance of employing EWC in environments with a WFQ system lies in the fact that fairness is as critical as efficiency.
10. The integration of EWC for continual learning within the DRL algorithm enhances accumulative rewards over time by preserving effective policies from prior network states, leading to enhanced performance as the agent encounters new traffic patterns. By mitigating the problem of catastrophic forgetting, the WFQ continual-DRL algorithm is an effective strategy.
11. The real traffic in a network revealed a higher reward for two real traffic of various categories. On the other hand, they are more bursty with higher variance. Furthermore, they offer unpredictable patterns that the SAC with EWC for the proposed algorithm can exploit, resulting in better learning and higher cumulative rewards than simulation traffic.
12. Continual learning through EWC helps retain practical knowledge across similar queue behaviors in real traffic, enhancing generalization and performance, meaning that the traffic generated in simulation is more aggressive, indicating that the model can handle various patterns and categories of real traffic.

Finally, integration of the EWC technique as a continual learning approach into DRL frameworks, realized here in the form of a new algorithm called WFQ continual-DRL for WFQ systems, facilitates adaptive dynamic bandwidth allocation, thereby enhancing the system's ability to maintain optimal performance across varying network conditions. This new approach enables more robust and reliable network management, effectively reducing

packet loss and queue lengths and enhancing the overall quality of service (QoS), although in real world conditions.

7. Conclusions

In this paper, a new integrated adaptive algorithm called WFQ continual-DRL was developed to enhance dynamic bandwidth allocation for each queue in a WFQ system. The proposed algorithm enables the system to obtain a policy agent that learns the optimal weights for different arrivals of network flows. Our contribution involves integrating the adopted reinforcement learning algorithm, known as the SAC algorithm, with the EWC approach, which supports continual learning in the more challenging field of DRL. To enable this implementation to be compared to the existing extensive telecommunication network, this investigation considered two scenarios: the first utilized eight queues, while the second employed twelve. The findings demonstrate that incorporating EWC effectively mitigates the issue of catastrophic forgetting in DRL models, thereby improving learning stability and the models' adaptability to new tasks without compromising prior performance. This achieved high performance by enabling efficient learning, rapid adaptation, and scalability, even when the number of queues increases. This integration significantly improved the performance of the measurement system, as reflected in cumulative reward. The results achieved an improvement in the reward curve in an experiment involving eight queues, and expanding to twelve queues appears to result in increased and smoother fluctuation of average reward, indicating improved overall performance of the MSE, and MAE and resulting in more reliable decision making within WFQ environments, particularly through dynamic bandwidth allocation. The comparative evaluation between the enhanced WFQ continual-DRL algorithm demonstrates that integrating EWC significantly improves accuracy, as indicated by a lower RMSE, and fairness, as reflected in a higher Jain's Index of the learning model. This enhancement of continual learning techniques in DRL and scheduling is important for resource allocation, mainly in complex network environments. Finally, training and evaluation of real traffic flows of the dataset indicate the agent to achieve better performance across diverse conditions, resulting in more adaptable and effective bandwidth allocation processes across twelve queues simulation, indicating that the model training is under aggressive traffic. An interesting direction for future study would involve integrating real network traffic data to effectively assess the algorithm's performance under real-world conditions in dynamic and complex network environments.

Author Contributions: Conceptualization, D.A.M. and M.A.M.; methodology, D.A.M. and M.A.M.; software, M.A.M.; validation, D.A.M. and M.A.M.; formal analysis, M.A.M.; investigation, M.A.M.; resources, M.A.M.; data curation, M.A.M.; writing—original draft preparation, M.A.M.; writing—review and editing, D.A.M. and M.A.M.; visualization, M.A.M.; supervision, D.A.M.; project administration, D.A.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data is contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Jay, N.; Rotman, N.H.; Godfrey, P.B.; Schapira, M.; Tamar, A. A deep reinforcement learning perspective on internet congestion control. In Proceedings of the 36th International Conference on Machine Learning (ICML), Long Beach, CA, USA, 9–15 May 2019; pp. 3050–3059.
2. Strzemiński, D.; Nafkha, R.; Zawislak, R. Performance analysis of a QoS system with WFQ queuing using temporal Petri nets. In Proceedings of the 20th International Conference (CISIM), Elk, Poland, 24–26 September 2021; pp. 462–476.

3. Zakariyya, I.; Rahman, M.N.A.; Ismail, M.N. Modelling the performance of class-based weighted fair queue using OPNET. *Int. J. Netw. Comput. Appl. Archit.* **2015**, *5*, 53–60. [CrossRef]
4. Brahim, S.B.; Zrelli, A.; Dardouri, S.; Bouallegue, R. Optimized architecture for efficient OFDMA network design. *Telecom* **2024**, *5*, 1051–1075. [CrossRef]
5. Zhang, J.; Tang, J.; Zhang, X.; Ouyang, W.; Wang, D. A survey of network traffic generation. In Proceedings of the Third International Conference on Cyberspace Technology (CCT 2015), Beijing, China, 17–18 October 2015; pp. 1–6.
6. Ghimire, S.; Thapa, G.B.; Ghinire, R.P.; Silvestrov, S. A survey on queueing systems with mathematical models and applications. *Am. J. Oper. Res.* **2017**, *1*, 1–14.
7. Attar, I.S.; Mahyuddin, N.M.; Hindia, M.H.D. A semi-distributed scheme for mode selection and resource allocation in device-to-device-enabled cellular networks using matching game and reinforcement learning. *Telecom* **2025**, *6*, 12. [CrossRef]
8. Zakariyya, I.; Rahman, M.N.A. Bandwidth guarantee using class-based weighted fair queue (CBWFQ) scheduling algorithm. *Int. J. Digit. Inf. Wirel. Commun.* **2015**, *3*, 152–157. [CrossRef]
9. Mustafa, M.E.; Talab, S.A. The effect of queueing mechanisms first in first out (FIFO), priority queueing (PQ) and weighted fair queueing (WFQ) on network's routers and applications. *Int. J. Digit. Inf. Wirel. Commun.* **2016**, *8*, 77–84. [CrossRef]
10. Al-Sawaai, A.; Awan, I.; Fretwell, R. Analysis of the weighted fair queueing system with two classes of customers with finite buffer. In Proceedings of the 2009 International Conference on Advanced Information Networking and Applications Workshops (WAINA), Bradford, UK, 26–29 May 2009; pp. 218–223.
11. Koyuncu, M.; Hayder, A.F. Comparison of scheduling algorithms for multimedia applications in IPv4 and IPv6. In Proceedings of the 2015 9th International Conference on Application of Information and Communication Technologies (AICT), Rostov-on-Don, Russia, 14–16 October 2015; pp. 418–422.
12. Assegie, T.A.; Bizuneh, H.D. Improving network performance with an integrated priority queue and weighted fair queue scheduling. *Indones. J. Electr. Eng. Comput. Sci.* **2020**, *19*, 241–247. [CrossRef]
13. Khanam, S.; Ahmedy, I.; Idris, M.Y.I. Performance evaluation of weighted fair queueing model for bandwidth allocation. In Proceedings of the International Conference on Innovative Computing and Communications (ICICC), New Delhi, India, 20–21 February 2021; Volume 1, pp. 175–183.
14. Elnaka, A.M.; Mahmoud, Q.H.; Li, X. Simulation-based comparative performance analysis of QoS traffic scheduling using fair and delay-adaptive scheduler (FDAS) versus WFQ and EDF. In Proceedings of the 2016 13th IEEE Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 9–12 January 2016; pp. 916–923.
15. Attar, H.; Khosravi, M.R.; Igorovich, S.S.; Georgiev, K.N.; Alhihi, M. Review and performance evaluation of FIFO, PQ, CQ, FQ, WFQ algorithms in multimedia wireless sensor networks. *Int. J. Distrib. Sens. Netw.* **2020**, *16*, 1550147720913233. [CrossRef]
16. Mahmood, D.A.; Horváth, G. A simple approximation for the response times in the two-class weighted fair queueing system. In Proceedings of the Analytical and Stochastic Modelling Techniques and Applications (ASMTA), Newcastle-upon-Tyne, UK, 10–11 July 2017; pp. 125–137.
17. Lillicrap, T.P.; Hunt, J.J.; Pritzel, A.; Heess, N.; Erez, T.; Tassa, Y.; Silver, D.; Wierstra, D. Continuous control with deep reinforcement learning. *arXiv* **2015**, arXiv:1509.02971.
18. Bachl, M.; Fabini, J.; Zseby, T. LFQ: Online Learning of Per-flow Queueing Policies using Deep Reinforcement Learning. In Proceedings of the 2020 IEEE 45th Conference on Local Computer Networks (LCN), Sydney, Australia, 16–19 November 2020; pp. 417–420. [CrossRef]
19. Fawaz, H.; Zeghlache, D.; Pham, T.A.; Leguay, J.; Medagliani, P. Deep Reinforcement Learning for Smart Queue Management. In Proceedings of the Conference on Networked Systems 2021 (NetSys 2021), Lübeck, Germany, 13–16 September 2021; Volume 80. [CrossRef]
20. Raeis, M.; Tizghadam, A.; Leon-Garcia, A. Queue-learning: A reinforcement learning approach for providing quality of service. In Proceedings of the AAAI Technical Track on Application Domains, Vancouver, BC, Canada, 2–9 February 2021; Volume 35, pp. 461–468.
21. Wang, P.; Jiang, Z.; Qi, M.; Dai, L.; Xu, H. Uncertainty-aware weighted fair queueing for routers based on deep reinforcement learning. In Proceedings of the 2021 IEEE 4th International Conference on Electronics and Communication Engineering (ICECE), Xi'an, China, 17–19 December 2021; pp. 1–7.
22. Pan, J.; Chen, G.; Wu, H.; Peng, X.; Xia, L. Deep reinforcement learning-based dynamic bandwidth allocation in weighted fair queues of routers. In Proceedings of the 2022 IEEE 18th International Conference on Automation Science and Engineering (CASE), Mexico City, Mexico, 22–26 August 2022; pp. 1580–1587. [CrossRef]
23. Kirkpatrick, J.; Pascanu, R.; Rabinowitz, N.; Veness, J.; Desjardins, G.; Rusu, A.A.; Milan, K.; Quan, J.; Ramalho, T.; Grabska-Barwinska, A.; et al. Overcoming catastrophic forgetting in neural networks. *Proc. Natl. Acad. Sci. USA* **2017**, *13*, 3521–3526. [CrossRef] [PubMed]

24. Haarnoja, T.; Zhou, A.; Abbeel, P.; Levine, S. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In Proceedings of the 35th International Conference on Machine Learning (ICML), Stockholm, Sweden, 10–15 July 2018; pp. 1861–1870.
25. Mock, J.W.; Muknahallipatna, S.S. A comparison of PPO, TD3 and SAC reinforcement algorithms for quadruped walking gait generation. *J. Intell. Learn. Syst. Appl.* **2023**, *1*, 36–56. [CrossRef]
26. Aich, A. Elastic weight consolidation (EWC): Nuts and bolts. *arXiv* **2021**, arXiv:2105.04093.
27. Shortle, J.F.; Fischer, M.J. Approximation for a two-class weighted fair queueing discipline. *Perform. Eval.* **2010**, *10*, 946–958. [CrossRef]
28. Rahimipour, S.; Moeinfar, R.; Hashemi, S.M. Traffic prediction using a self-adjusted evolutionary neural network. *J. Mod. Transp.* **2019**, *27*, 306–316. [CrossRef]
29. Rochim, A.F.; Muis, A.; Sari, R.F. A discrimination index based on Jain’s fairness index to differentiate researchers with identical H-index values. *J. Data Inf. Sci.* **2020**, *4*, 5–18. [CrossRef]
30. Wu, J. *Some Properties of the Normal Distribution*; LAMDA Tech. Rep., National Key Lab for Novel Software Technology, Nanjing University: Nanjing, China, 2019; pp. 1–8.
31. Wang, L.; Zhang, X.; Su, H.; Zhu, J. A comprehensive survey of continual learning: Theory, method and application. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 5362–5383. [CrossRef] [PubMed]
32. Rojas, J.S. Labeled Network Traffic Flows: 114 Applications. Available online: <https://www.kaggle.com/datasets/jsrojas/labeled-network-traffic-flows-114-applications> (accessed on 30 May 2025).
33. Mawlood, M.A.; Mahmood, D.A. Performance analysis of Weighted Fair Queuing (WFQ) scheduler algorithm through efficient resource allocation in network traffic modeling. *J. Commun. Softw. Syst.* **2024**, *3*, 266–277. [CrossRef]
34. Kumar, P. Application of correlation-regression method to correlate the statistical relationship between the parameters of some indexed journals. *Int. J. Eng. Adv. Technol.* **2019**, *9*, 505–509. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Context-Aware Beam Selection for IRS-Assisted mmWave V2I Communications

Ricardo Suarez del Valle ¹, Abdulkadir Kose ^{2,*} and Haeyoung Lee ³

¹ Department of Electronic Engineering, University of Surrey, Guildford GU2 7XH, UK; ricardo.su.dv@gmail.com

² Department of Computer Engineering, Abdullah Gül University, Kayseri 38080, Türkiye

³ School of Physics, Engineering and Computer Science, University of Hertfordshire, Hatfield AL10 9AB, UK; h.lee@herts.ac.uk

* Correspondence: abdulkdir.kose@agu.edu.tr

Abstract: Millimeter wave (mmWave) technology, with its ultra-high bandwidth and low latency, holds significant promise for vehicle-to-everything (V2X) communications. However, it faces challenges such as high propagation losses and limited coverage in dense urban vehicular environments. Intelligent Reflecting Surfaces (IRSs) help address these issues by enhancing mmWave signal paths around obstacles, thereby maintaining reliable communication. This paper introduces a novel Contextual Multi-Armed Bandit (C-MAB) algorithm designed to dynamically adapt beam and IRS selections based on real-time environmental context. Simulation results demonstrate that the proposed C-MAB approach significantly improves link stability, doubling average beam sojourn times compared to traditional SNR-based strategies and standard MAB methods, and achieving gains of up to four times the performance in scenarios with IRS assistance. This approach enables optimized resource allocation and significantly improves coverage, data rate, and resource utilization compared to conventional methods.

Keywords: mmWave; V2X; RIS; machine learning; multi-armed bandit

1. Introduction

The advent of millimeter-wave (mmWave) technology has sparked a transformative wave in next-generation wireless communication. In this dynamic landscape, Vehicle-to-Everything (V2X) communication emerges as a pivotal force poised to reshape the future of transportation systems [1]. By enabling seamless interaction among vehicles, infrastructure, pedestrians, and other road users, V2X holds the promise of fostering safer, more efficient, and smarter transportation networks [2]. However, leveraging mmWave for V2X introduces significant challenges due to the high-frequency nature of mmWave signals, including increased propagation loss, susceptibility to interference, and mobility constraints [3]. Additionally, mmWave small cells operate with highly directional narrow beams to maintain high data rates, but this results in frequent beam switching and handovers as vehicles rapidly move between beam sectors, adding further complexity to maintaining stable and reliable communication.

To fully harness mmWave's potential for V2X communication, researchers and engineers are delving into innovative solutions aimed at enhancing signal propagation, coverage, and reliability. One particularly promising avenue gaining momentum is the integration of Intelligent Reflecting Surfaces (IRSs) into mmWave V2X systems [4]. Outfitted with passive reflecting elements, these surfaces excel at redirecting signals when a direct

link between User Equipment (UE) and a mmWave Base Station (BS) is obstructed [5]. IRS technology emerges as a focal point in the research community, holding the potential to significantly boost spectral efficiency and expand communication range [6].

The efficacy of IRS-assisted mmWave V2X communication hinges not only on deploying IRSs but also on tailoring their functionality to the specific demands and dynamics of the mmWave V2X environment [7]. Context-awareness becomes paramount in this context, enabling the system to dynamically respond to real-time changes in vehicular mobility, traffic patterns, and environmental conditions [8]. Employing context-aware beam selection strategies alongside IRS technology proves instrumental in enhancing communication quality and coverage, minimizing latency, and ensuring the reliability of mmWave V2X communication across diverse scenarios [9].

2. Related Works

Beam management and selection, a fundamental aspect of IRS-assisted mmWave V2X systems, have been at the forefront of investigation. However, efficiently managing the beams in such systems, especially coordinating between the BS and IRS for two-hop transmission, is challenging. Several papers have contributed novel approaches to this domain, enhancing the efficiency of mmWave V2X communication. A comprehensive survey by [10] offers a broad overview of IRS applications in vehicular communications, both ground-based and aerial, highlighting the potential and challenges of IRS-aided 6G vehicular communications. Additionally, ref. [11] discusses various use case scenarios for IRS-enabled V2X communications, particularly focusing on vehicular edge computing and IRS-enabled drone communications to minimize vehicle computational time.

A machine learning-based approach proposed in [12] includes two phases: situational and mobility awareness. In the situational awareness phase, a model at the BS predicts the optimal IRS setup for a UE's location. The mobility awareness phase, at the UE, predicts UE mobility. This allows for predicting the optimal IRS configuration, and significantly reduces system overhead during initial access. In highly dynamic scenarios, it achieves higher spectral efficiency by using predicted UE locations. Leveraging location information, ref. [13] examines the impact of prior location data on channel parameter estimation in IRS-aided mmWave MIMO (Multiple-Input Multiple-Output) systems. It presents directional training beams based on rough location data and employs atomic norm minimization for channel parameter extraction. The authors of [14] introduce real-time software-controlled IRSs and examine their optimal placement in wireless networks, addressing reliability requirements for V2X communications. Furthermore, ref. [15] introduces a deep reinforcement learning algorithm for beamforming optimization in mmWave vehicle-to-infrastructure (V2I) communication systems. This novel algorithm tackles the non-convex and time-varying nature of joint beamforming matrix optimization for the BS and the IRS. The results demonstrate substantial network performance improvement, especially in dense V2I network scenarios.

One approach, as proposed in [16], introduces multi-beam training methods for IRS-assisted multiuser communication, designed to reduce training time without compromising passive beamforming performance. It achieves this by partitioning IRS elements into sub-arrays and implementing a multi-beam codebook, thus enabling efficient beam direction control for different users. Ref. [17] explores the use of IRSs in a millimeter-wave vehicular communication network, introducing schemes to reduce channel estimation overhead and considering rate maximization in the uplink. Lastly, the study in [18] employs IRSs to mitigate Inter-Cell Interference (ICI) by using graph theory to minimize beam collisions.

Our Contributions

These research findings collectively underscore the evolving landscape of IRS-assisted mmWave V2X communication, emphasizing innovation in beam management and selection as a key enabler for its widespread adoption and the improvement in next-generation vehicular communication systems. In addition, these traditional beam selection strategies in vehicular mmWave networks often rely on instantaneous signal strength or static location mapping, which do not reflect future link duration or vehicle mobility patterns. As a result, they tend to cause frequent handovers and unstable links. In contrast, our proposed context-aware framework predicts longer-lasting beam assignments based on high-level contextual factors, reducing switching events and improving average link duration. In this paper, we propose a Context-aware Multi-Arm Bandit (C-MAB)-based beam selection for IRS-assisted vehicular communication by incorporating vehicle mobility context under a realistic vehicular mobility environment. The aim is to establish long-lasting connections, unlike signal strength-based approaches for end-to-end V2I communication. Note that, unlike standard C-MAB models, our approach considers a spatiotemporal environment where vehicle mobility and shared beam resources create interaction between decisions. While most C-MABs assume static user contexts and independent arm outcomes, our setting involves overlapping beam coverage, prioritized scheduling, and varying mobility patterns. Therefore, we define reward as the total connection time achieved under a beam, conditioned on the vehicle's context and future trajectory. This formulation aligns more closely with real-world beam stability concerns and differentiates our work from typical bandit designs that focus only on immediate signal gain or throughput. The following describes the main contributions of this paper.

- *Context-driven beam and IRS selection:* We introduce a novel approach that leverages a range of contextual factors to enhance the beam and IRS selection process. By considering variables such as vehicle mobility information (e.g., travel directions and relative locations) and traffic light information, we improve the beam and IRS selection process to enable longer-lasting connections. This context-aware selection approach not only improves the overall connectivity quality but also extends the connection time, a critical factor in V2X communication, especially for mission-critical applications.
- *Predictive testing tool for IRS locations:* Recognizing the importance of IRS placement in achieving optimal signal reflections, we offer a practical tool for assessing the effectiveness of IRS locations. This tool empowers network planners and engineers to simulate IRS deployments and assess their impact on signal propagation before actual implementation. By doing so, we mitigate the risk of suboptimal IRS placement, thereby enhancing the overall performance and robustness of mmWave V2X communication systems.
- *Utilization of traffic light waiting times:* This is the first work to incorporate waiting times at traffic lights as contextual data for predicting beam sojourn times in IRS-assisted V2I communication. This predictive insight allows for the development of proactive mobility and resource management strategies. By factoring in traffic light cycles, we can optimize beam switching, resource allocation, and mobility planning, ensuring that the V2I communication system remains seamless and uninterrupted, even in stop-and-go traffic scenarios. Note that the traffic light status can be acquired through standard V2X interfaces supported by modern infrastructure. Traffic controllers broadcast phase and timing information (e.g., current light status, remaining time, next transition) using SPaT messages [19]. These data can be received directly by the base station via wired connectivity or wirelessly via Roadside Units (RSUs) that act as V2I relays. In addition, vehicles can share their perception of the traffic light status through Cooperative Awareness Messages (CAMs) or Basic Safety Messages

(BSMs), which include location, speed, heading, and, in some cases, braking events that suggest deceleration due to traffic light stops. These standardized messages allow the base station to build a synchronized and real-time environmental context, which includes the dynamic state of the traffic signal affecting each vehicle's route.

In summary, our contributions encompass not only context-aware beam and IRS selection for prolonged connection times but also a practical testing tool to optimize IRS deployment strategies. Moreover, by harnessing waiting times at traffic lights as predictive context, we enable the development of proactive strategies to enhance mobility management and resource allocation in mmWave V2X communication. These innovations collectively pave the way for more robust, reliable, and efficient vehicular communication systems, crucial for the advancement of smart and safe transportation networks. The remainder of this paper is organized as follows: Section 2 describes the considered scenario and formulation of the beam selection problem. In Section 3, the C-MAB based beam selection algorithm is elaborated. The experimental results are explained in Section 4 to show the effectiveness of our proposed algorithm. Finally, we draw important conclusions in Section 5.

3. Scenario Setup and Problem Formulation

We examine a mmWave small cell designed to improve throughput performance and network capacity for downlink transmission. The small cell BS under consideration is equipped with an antenna array directed along predefined orientations, though the number of radio frequency (RF) chains is fewer than the number of antennas. Each antenna may consist of multiple elements that form a beam directed towards a specific region. As described in [20], the coverage area of a beam, or beam sector, is ideally spatially separated. Given the limited RF chains, only a subset of antennas can be activated for downlink at any time.

The scenario is depicted in Figure 1. The mmWave base station (mmBS), located in central London, UK, is equipped with 12 antennas oriented to ensure broad coverage. Due to a limited number of RF chains, the mmBS can activate up to four beams simultaneously, supporting a maximum of four vehicles at once. Beam coverage is determined by the vehicle's presence within a beam's geometric footprint and the achieved SNR meeting the threshold for link activation. We adopt a two-slope path loss model to distinguish line-of-sight (LOS) and non-line-of-sight (NLOS) propagation. If the estimated SNR falls below a threshold (e.g., 2 dB), coverage is not considered valid. The data rate is modeled as a constant value during beam association, consistent with baseline mmWave assumptions and to isolate the impact of connection time from PHY-layer variations. Resource utilization reflects the number of active beams simultaneously serving distinct vehicles, constrained by beamforming hardware limits (i.e., max four beams). A higher beam sojourn time implies better utilization per beam, as it reduces frequent reconfigurations and idle slots.

To further enhance signal coverage and alleviate potential blockages in the urban environment, several IRS nodes are strategically deployed within the area. These IRS nodes dynamically reconfigure signal propagation paths, improving beam delivery in NLOS conditions. Note that the illustrated sector areas are approximate; actual coverage depends on the channel model, antenna configuration, and IRS-assisted reflections.

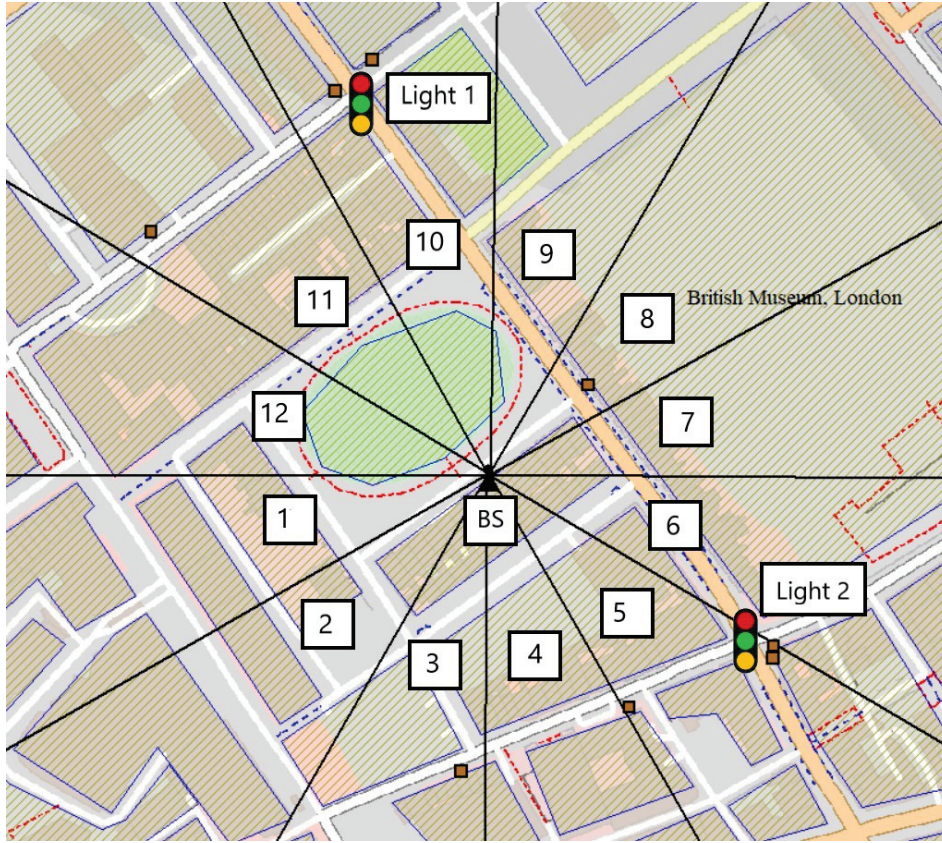


Figure 1. A scenario of a network which consists of a mmBS with 12 beam sectors and multiple IRS nodes shown on a real-world map of around Bedford Square in London.

This work assumes a carrier frequency of 28 GHz, which has been widely adopted in 5G NR-based mmWave V2X studies [21–23]. While traditional DSRC (Dedicated Short-Range Communications) and C-V2X (Cellular V2X) systems operate in the 5.9 GHz band, the 3GPP n257 (26.5–29.5 GHz) is considered a promising band for ultra-high data rate and low-latency vehicular applications, particularly in urban environments with managed infrastructure and directional communication. The adjacent 26 GHz band (n258) is already approved for V2X use in Europe [22]. Field trials such as autonomous beam-steered measurement campaigns [23] further validate the viability of the 28 GHz bands for vehicular communications. Thus, we consider 28 GHz mmWave communication with its path loss model [21] expressed as:

$$PL(d) = PL(d_0) + 10n \log_{10}(d/d_0) + X_g. \quad (1)$$

Here, d refers to the distance between the mmBS and vehicle antennas in meters while X_g represents the channel fading (excluded in our analysis), and $PL(d_0)$ denotes the free space path loss (FSPL) in decibels (dBs). The FSPL depends on the carrier frequency f_c , and is given by $10 \log_{10}((4\pi d_0 f_c / c)^2)$, with $d_0 = 1$ m. We assume a 5 m height difference between the BS and the vehicle, such that the 2D distance $\hat{d}$ between two points is related to d as $d = \sqrt{\hat{d}^2 + 5^2}$.

Assuming that vehicles have steerable beam antennas capable of tracking the BS, the signal-to-noise ratio (SNR) for a vehicle being served is given by:

$$SNR = p_0 - PL(d) + G_{tx} + G_{rx} - N, \quad (2)$$

where p_0 represents the reference power which is the received power at a reference distance of $d_0 = 1$ m. G_{tx} and G_{rx} are the transmitter and receiver antenna gains, respectively, N is the noise, including thermal noise and the receiver noise figure.

As aforementioned, the deployment of mmWave small cells for V2X communication introduces a distinct challenge, as narrow beams are used. The short time a vehicle is present within a beam's coverage results in frequent handovers and increased signal overhead. In addition, due to the fluctuating nature of the channel caused by vehicle mobility, a more conservative approach—using a fixed, robust modulation scheme—may be more suitable than an adaptive modulation scheme. Thus, we assume the BS uses a constant, resilient modulation scheme, regardless of the reported SNR. To maximize data transmission in downlink, the radio resource allocation strategy should prioritize vehicles that stay within the beam for the longest periods.

Let R denote the fixed data rate assigned to each beam when serving a vehicle within its beam sector. Vehicles that leave the beam sector experience a zero data rate. Consider a set $\mathcal{B}(t)$ of available beams from the mmWave small cell BS at time t , where $|\mathcal{B}(t)|$ can be at most 4, representing the maximum number of simultaneously active beams. Let $R_i(t)$ denote the transmission rate of beam i at time t , where $R_i(t) = R$ if a vehicle is being served, and $R_i(t) = 0$ during handovers, when waiting data from the macro BS, or when no vehicle is within the beam's coverage. The objective of the proposed resource allocation strategy is to maximize the total data transmission across all beams over a given time period T . Minimizing handovers is crucial to achieving this goal, as each handover introduces a data transmission interruption. Let S_i represent the number of serving vehicle switches for beam i during T and let T_{ij} denote the sojourn time of j th vehicle served by beam i . By maximizing the vehicle sojourn time within a beam sector, the frequency of handovers is reduced, thereby increasing overall transmission time. Consequently, the resource allocation problem can be formulated as:

$$\max \left(\sum_{i \in \mathcal{B}(t)} \int_0^T R_i(t) dt \right) \text{ or } \max \left(\sum_{i \in \mathcal{B}(t)} \sum_{j=1}^{S_i} T_{ij} \right). \quad (3)$$

To maximize data transmission, the mmBS utilizes all available beams. When a vehicle receives service within a beam sector, the corresponding beam becomes unavailable for other vehicles, and the BS selects another beam from the available set to serve a new vehicle. The BS can either select a vehicle at random or prioritize the vehicle with the highest SNR. However, as we will demonstrate, the conventional strategy of selecting the vehicle with the highest SNR may not be optimal for mmWave vehicular networks due to the narrow beamwidth and frequent handovers. Given the restricted mobility of vehicles within urban street layouts, we propose leveraging vehicle mobility information as an indirect indicator of the street's topology. By constructing vehicle profiles based on their mobility patterns, this contextual information can be utilized for beam selection. Mobility data, such as vehicle orientation, position, and distance from the BS (obtained through local measurements or timing advance), can be employed to enhance the efficiency of beam assignment.

4. Proposed Contextual Multi-Armed Bandit Learning Design

In our previous work [24,25], we demonstrated the effectiveness of the Contextual Multi-Armed Bandit (C-MAB) algorithm in optimizing beam-level communication in mmWave vehicular networks. Based on this, our subsequent study [26] extended C-MAB to relay-assisted V2V-based V2X communication strategies, emphasizing the importance of contextual information in dynamic environments. These studies highlighted how factors such as road layout and blockages significantly impact connectivity, which requires careful

selection of contextual data to improve decision making [24,26]. In this paper, we study a new environment by incorporating IRSs into vehicular communication, addressing the challenges posed by dynamic blockages and harsh propagation conditions in mmWave networks. Unlike relays, IRS extends coverage passively and efficiently but faces key challenges: it lacks signal amplification, limiting effectiveness in weak links; requires continuous phase shift adaptation to vehicle mobility; and demands precise beam alignment, making real-time coordination complex. Additionally, IRS placement is constrained by existing infrastructure, requiring strategic positioning for optimal coverage. To tackle these challenges, we incorporate diverse contextual factors—such as vehicle travel direction, mmBS location, and traffic light information—into contextual learning algorithm for optimized beam selection, ensuring stable and prolonged connections. Our approach strategically leverages IRSs to enhance vehicular communication while mitigating their limitations, enabling robust and adaptive connectivity in dynamic urban environments. The following seven contexts are incorporated into the proposed C-MAB framework to enhance beam selection and connectivity:

- Vehicle Travel Direction: {North, East, South, West};
- Vehicle Travel Direction with respect to mmBS: {Towards, Away};
- Vehicle distance to mmBS: {Close, Mid, Far};
- Traffic light status: {Red, Green};
- Remaining time for traffic light color change: {Small, Mid, Big};
- Vehicle Travel Direction with respect to traffic light: {Towards, Away};
- Vehicle distance to traffic light: {Close, Mid, Far}.

The algorithm utilizes seven distinct contextual features, each represented by a single character. These characters are concatenated into a seven-character string that encodes the complete context. Each character corresponds to one specific context, capturing its current discrete state. For instance, a context string “SFTRSFT” represents a vehicle that is traveling South (S), is Far (F) from the mmBS, and moving Towards (T) it; while the nearest traffic light is Red (R), has a Small (S) remaining time before switching, and the vehicle is Far (F) and moving Towards (T) the light. This context representation balances environmental richness with computational efficiency, enabling intelligent and adaptive beam–IRS decision-making in highly dynamic vehicular environments.

The proposed C-MAB algorithm is designed to integrate IRS-assisted links, ensuring improved coverage and seamless beam allocation in obstructed environments. To mitigate blockages, IRSs are employed to extend coverage to vehicles experiencing connectivity disruptions, leveraging V2I communication when needed, as illustrated in Figure 2. By intelligently selecting direct or IRS-assisted transmission, the algorithm enhances overall network reliability and connectivity in dense urban settings. It is important to note that, unlike active antenna arrays or smart antennas, IRS units do not require RF chains or power-hungry circuitry, which makes them highly energy-efficient and cost-effective. However, this also introduces certain limitations. First, since IRS elements only reflect incoming signals, they depend heavily on favorable deployment geometry—that is, they must be positioned where they can effectively “see” both the BS and the vehicle. In addition, due to path loss from two-hop reflection (BS→IRS→UE), IRS-assisted links often require stronger incident signals or tighter phase alignment to be beneficial compared to direct line-of-sight links. In this work, we adopt the widely used assumption of quasi-static phase configurations for the IRS, consistent with prior works [27]. The IRS is pre-configured to reflect signals toward designated zones based on its physical orientation and deployment geometry. The current algorithm does not perform real-time phase shift optimization but rather selects among available beam–IRS combinations based on learned context-specific rewards. This abstraction allows us to focus on the high-level contextual beam

selection problem without entangling it with the physical-layer phase design. However, in practice, the integration of dynamic IRS control—particularly under high-mobility vehicular scenarios—is an open challenge and represents a valuable direction for future work.

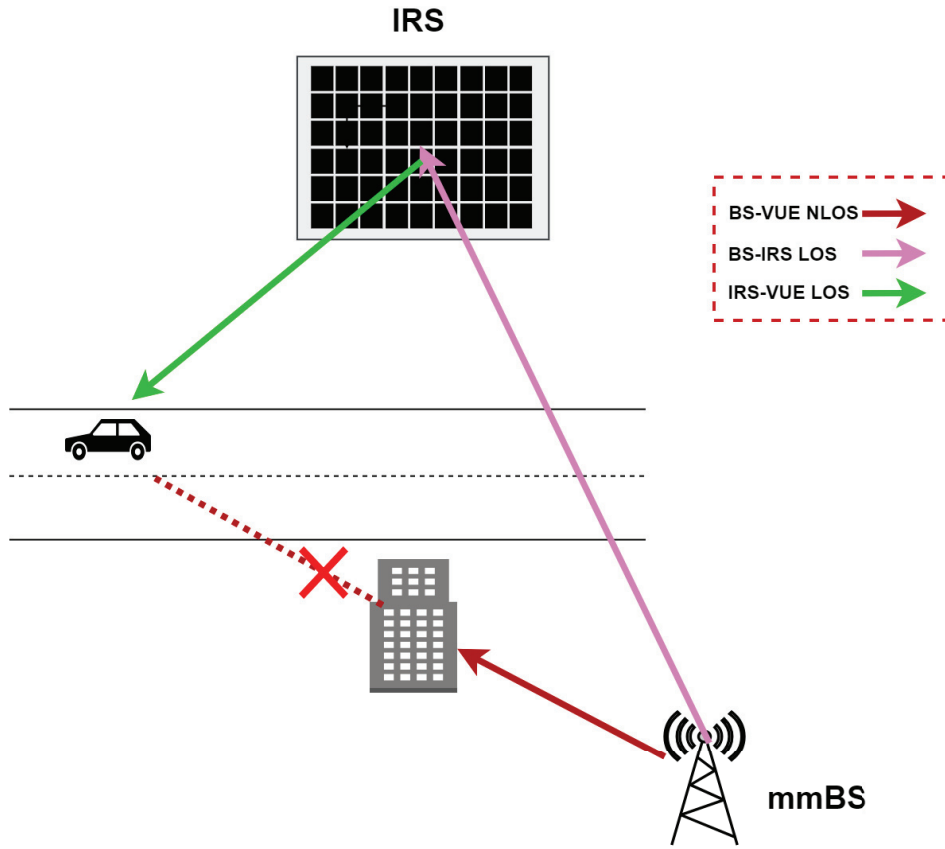


Figure 2. IRS-assisted beam communication under blockages. The dotted red line with a cross symbol represents a blocked (NLOS) path between the mmBS and the vehicle UE.

Algorithm 1 presents the C-MAB algorithm designed for a communication scenario where a BS with B beams and S distributed IRSs serve vehicles in a dynamic environment. The objective is to optimize beam and IRS selection by leveraging contextual information (c), as detailed in this section, to enhance decision-making. To give an example of operation of the C-MAB algorithm, let $\mathcal{C}$ represent the set encompassing all possible contextual information, and $\langle b, s, c \rangle$ denote an ordered triple consisting of beam b , IRS s , and context c . The system records rewards for all combinations of beams, IRSs, and contexts, forming the set $\mathcal{Q}_{\mathcal{C}} = \{q_{\langle b, s, c \rangle} | b \in \mathcal{B}, s \in \mathcal{S}, c \in \mathcal{C}\}$. When a beam b_i and an IRS s_i are selected for transmission at time t , an instantaneous context $c_i(t)$ is observed. Subsequently, after a vehicle service duration of Δt , the reward $q_i(t + \Delta t)$ associated with the current context is measured. This reward is then utilized to update the average reward value of context $c_i(t)$ using the formula:

$$\bar{q}_{c_i}(t + \Delta t) = \frac{\bar{q}_{c_i}(t) \cdot n_{c_i}(t) + q_i(t + \Delta t)}{n_{c_i}(t) + 1}, \quad (4)$$

where $n_{c_i}(t)$ represents the number of times the current context has been observed in the past.

The learning strategy adopts an exploration-first approach. The algorithm initializes parameters, including expected rewards ($Q_b^s(c)$) and observation counts ($N_b^s(c)$), for all beam–IRS configurations. In the exploration phase, a subset of beams is activated randomly,

and vehicles are also selected randomly based on observed contexts. Depending on beam availability, transmission occurs directly or via an IRS. The system updates expected rewards and observation counts through a weighted update rule to adapt to changing conditions. Transitioning to the exploitation phase after a specified number of exploration rounds, the algorithm intelligently selects the top K beams with the highest expected rewards. For each active beam, the algorithm determines the optimal transmission mode (direct or IRS-assisted) to maximize rewards. Upon connection loss due to blockage or vehicle movement, the reward values for each context–beam–IRS triplet are updated to optimize long-term performance.

Algorithm 1 Contextual MAB for Beam and IRS Selection

```

1: Input: Number of beams  $B$ , Number of IRSs  $S$ , Contexts  $C$ , Exploration rounds  $T_e$ ,
   Maximum active beams  $K$ , Counter for observed (beam,context) pair  $N$ 
2: Initialize:  $Q_b^s(c) \leftarrow 0$  and  $N_b^s(c) \leftarrow 0$  for all  $b \in [1, B]$ ,  $s \in [1, S]$ , and  $c \in C$ 
3: procedure EXPLORATION
4:   while  $t \leq T_e$  do
5:     Observe context  $c_t$ 
6:     Randomly activate a set of beams  $K_t$  from  $B$ 
7:     Randomly select vehicle  $v_t$  based on  $c_t$ 
8:     if There exists  $b_t \in K_t$  such that Beam  $b_t$  can reach  $v_t$  then
9:       Transmit directly via Beam  $b_t$ 
10:      Record reward  $r_t$ 
11:     else
12:       Joint transmission via IRS
13:       Record reward  $r_t$ 
14:     end if
15:     Update  $N_{b_t}^{s_t}(c_t) \leftarrow N_{b_t}^{s_t}(c_t) + 1$ 
16:     Update  $Q_{b_t}^{s_t}(c_t) \leftarrow Q_{b_t}^{s_t}(c_t) + \frac{1}{N_{b_t}^{s_t}(c_t)} \cdot r_t$ 
17:   end while
18: end procedure
19: procedure EXPLOITATION
20:   while  $t > T_e$  do
21:     Observe context  $c_t$ 
22:     Select a beam  $\mathbf{b}$  with highest  $Q_b^s(c_t)$  values for active transmission
23:     for Active Beam  $b_t$  do
24:       if Beam  $b_t$  can reach  $v_t$  then
25:         Transmit directly via Beam  $b_t$ 
26:         Record reward  $r_t$ 
27:       else
28:         Joint transmission via IRS
29:         Record reward  $r_t$ 
30:       end if
31:       Update  $N_{b_t}^{s_t}(c_t) \leftarrow N_{b_t}^{s_t}(c_t) + 1$ 
32:       Update  $Q_{b_t}^{s_t}(c_t) \leftarrow Q_{b_t}^{s_t}(c_t) + \frac{1}{N_{b_t}^{s_t}(c_t)} \cdot r_t$ 
33:     end for
34:   end while
35: end procedure

```

5. Results and Discussion

In the section, we present experimental results for the proposed C-MAB machine learning based beam selection. The scenario considered for experiment is presented in Figure 1, with the system parameters outlined in Table 1. In this scenario, we consider a standalone single mmBS with 12 beams. At any given time, the BS can only activate four beams for service. Vehicles are created and absorbed at certain locations on the map. Those

locations are the main entrance point to the city and the exit point from the city. Moreover, we also included several main parking spots in the city as vehicle creation and absorption locations. A pair of vehicle creation and absorption locations is used to create a route simulating a vehicle either passing through the city, entering to or exiting from the city. We use A-STAR path finding algorithm [28] to establish the route for vehicles. We assume that each of these vehicles requires downlink data service when entering the small cell. In our simulation, there are 50 vehicles which continuously travel on the map. We simulate three hours of operation, during which the base station begins with full exploration for learning, and then it switches to full exploitation after the first 50% of the simulation time. As the Explore-First strategy stops triggering exploration after the learning phase, this enables us to focus on the study of learning effectiveness acquired during the learning phase. The key simulation parameters used in our experiments are summarized in Table 1. Experiments are conducted via our own developed Python Mobility Simulation Platform (PyMoSim, version 0.8.10). The detailed documentation of PyMoSim, along with example usage scenarios, is publicly accessible at <https://github.com/cfoh/pymosim-doc>, accessed on 22 June 2025. We plan to release the full source code of PyMoSim and our scenario setup code in the near future).

The proposed C-MAB algorithm is selected for its lightweight, model-free nature, which makes it well-suited for real-time decision-making at the network edge. Unlike deep reinforcement learning (DRL) methods, which require large training datasets, extensive tuning, and high computational resources, C-MAB learns directly from interactions with the environment. This allows it to rapidly adapt to evolving contexts (e.g., vehicle movement, traffic light transitions) without prior training or simulation. C-MAB is also advantageous over location prediction or trajectory estimation methods, which typically rely on accurate positioning data and historical mobility patterns. Instead, our approach uses high-level context information—such as proximity and direction relative to mmBS or traffic lights—making it robust even in GPS-challenged environments. However, C-MAB also has limitations. Its performance can degrade in large or continuous context spaces due to the need to estimate a reward for every context–action pair. Additionally, C-MAB does not capture long-term or delayed rewards, which could limit its performance in sequential decision tasks with temporal dependencies. To address such scenarios, future work may explore hierarchical or hybrid learning frameworks capable of modeling temporally extended behavior.

The proposed algorithm does not directly estimate or rely on channel coherence time, but rather focuses on link-level stability through the observed beam sojourn time. At mmWave frequencies, where highly directional transmissions are used, the concept of beam coherence time becomes more critical than traditional channel coherence time. Beam coherence time denotes the interval during which a beam remains aligned with a moving user before misalignment occurs and is typically an order of magnitude longer than the channel coherence time. As argued in [29], frequent realignment based on channel coherence intervals introduces excessive overhead; instead, beam realignment should occur on the timescale of the beam coherence time to maintain efficient operation. In our model, beams are fixed with a 30° beamwidth, which inherently allows for longer beam coherence times. Beam reselection is triggered only when a user exits the beam sector or when the SNR falls below a defined threshold. This design choice reduces the need for frequent physical layer realignment and is well suited for urban vehicular speeds between 30 and 50 km/h (approximately 8–13.8 m/s). Under such conditions, beam coherence times are typically on the order of seconds, allowing our algorithm sufficient temporal margin to learn and select stable beam associations. Even under higher mobility, the C-MAB framework successfully

identifies long-lasting beam–IRS paths, improving robustness to short beam coherence intervals and reducing control overhead associated with frequent beam association.

Table 1. Simulation parameters.

Parameter	Value
Number of beams per mmBS (N_b)	12
Carrier Frequency (f_c)	28 GHz
Transmit power	30 dBm for BS/23 dBm for vehicles
Path loss exponent (LOS/NLOS)	1.9/3.8
Noise power	−90dBm
Vehicle Speed	30–50 km/h
Traffic light duration (Light1/Light2)	22/18 s
Simulation time	3 h
Sampling time slot	0.1 s
Simulation area	800 m × 450 m

5.1. Effectiveness of C-MAB on Connection Time

We conducted a comparative analysis of the C-MAB algorithm against classical MAB, Best-SNR, and random beam selection schemes, focusing on the average beam sojourn time. In the Best-SNR approach, the base station opts for a vehicle reporting the strongest received signal strength in a greedy manner. In our C-MAB, we leverage our contextual framework stated in Section 4. The beam sojourn time performance among the considered schemes is compared in Figure 3. After the exploration phase, C-MAB exhibits a notable increase, while MAB demonstrates only a slight uptick in mean beam sojourn time. This disparity is attributed to C-MAB’s proficiency in leveraging contextual information for decision-making, enabling it to outperform MAB by a factor of two, and significantly more when contrasted with Best-SNR and random selection schemes.

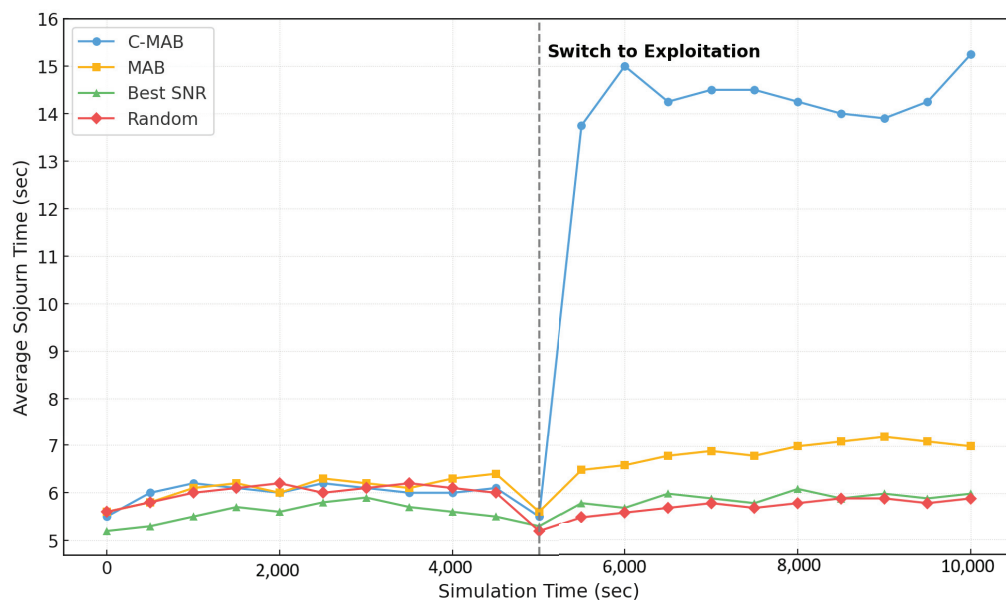


Figure 3. Average beam sojourn time performance of different beam selection schemes.

In Figure 4, the beam sojourn time performance of the proposed C-MAB is illustrated for all beams. Following the exploration phase, there is a notable increase in the mean beam sojourn time for most beams as the algorithm transitions into exploitation. This phenomenon is attributed to the increased utilization of certain beams in specific contexts, leading to longer beam sojourn times during the learning phase. Notably, Beam-8, Beam-5,

and Beam-3, along with Beam-10 and Beam-6, exhibit enhanced mean beam sojourn times. The scenario portrayed in Figure 1 is in line with this observation, where the shaded areas for these beams predominantly represent extended road coverages due to IRSs and higher line-of-sight (LOS) instances owing to fewer obstructions.

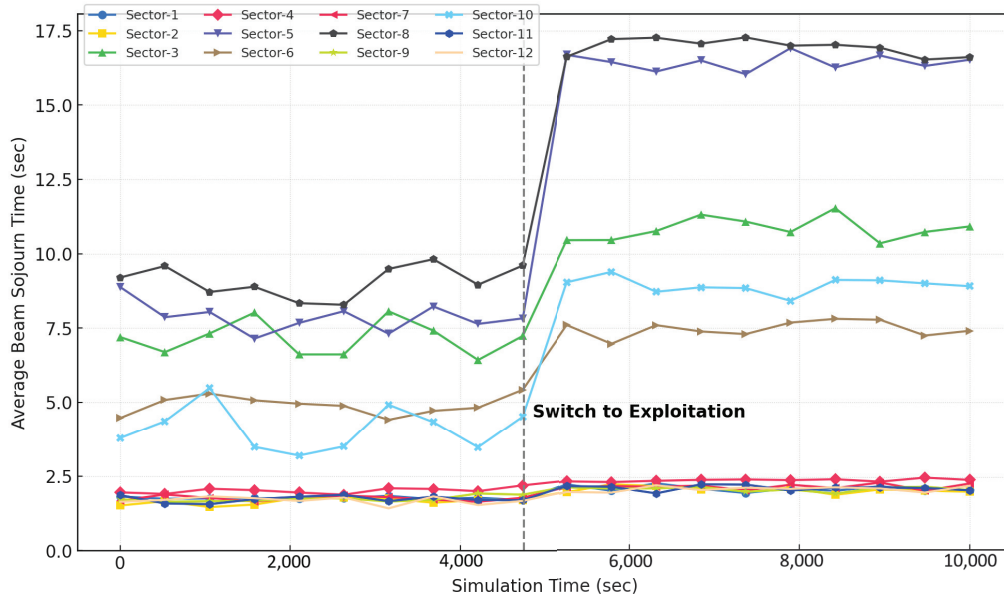


Figure 4. Average beam sojourn time of twelve beams with C-MAB.

Similar observations can be deduced from Figure 5, which illustrates the total utilization ratio of each beam throughout the exploration and exploitation phases. During the exploration phase, initial beam–sector selections are randomized across sectors. Sectors 2, 12, 11, and 1 exhibit an extremely low selection frequency due to inherently minimal vehicle traffic and shorter connection range due to weak received signal strength. Sector 4 similarly experiences limited selection since it lacks IRS deployment, compelling vehicles traveling north to primarily rely on extended coverage from neighboring Sector 3, which benefits significantly from IRS support. Sectors 7 and 9 display higher selection rates during exploration due to heavy traffic on the main road; however, their selections drastically shrink during exploitation due to short sojourn times. In contrast, Sector 3 experiences a substantial increase in connection frequency during the exploitation phase, driven by the synergistic effect of extended signal coverage from IRS deployment and prolonged connectivity opportunities resulting from vehicles frequently stopping and waiting at adjacent traffic lights. The C-MAB algorithm dynamically adjusts beam–sector selection by prioritizing sectors that offer longer and more stable connection times during the exploitation phase, hence responding effectively to contextual conditions. It intelligently considers IRS presence and infrastructure placement, consistently selecting sectors with durable, high-quality connections while systematically deprioritizing sectors characterized by transient, unstable connections. Consequently, the proposed C-MAB adeptly deal with environmental and mobility dynamics by aligning specific contexts with available beams to maximize rewards.

The C-MAB framework is computationally lightweight, making it amenable to real-time implementation at base stations or edge servers. At each decision epoch, the algorithm performs a lookup in the reward table indexed by the current context and selects the beam–IRS pair with the highest historical average. Assuming a bounded context space (e.g., 36 combinations from 3 distance bins $\times$ 3 directions $\times$ 2 traffic states), the table remains compact. No gradient updates or offline training are required. We empirically observe

rapid convergence: Figure 3 shows that the proposed C-MAB outperforms the classical MAB by a factor of two only after 5000 s of exploration time. To further improve efficiency, future extensions may apply context pruning strategies that remove low-utility or rare contexts from the table, thus focusing learning on high-value decision regions.

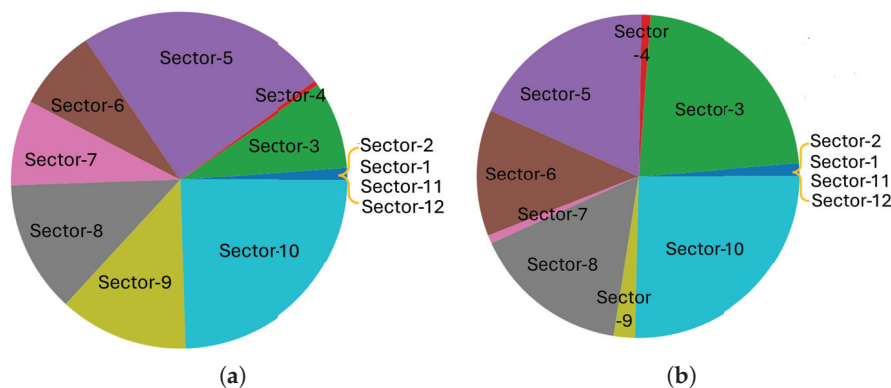


Figure 5. Connection frequency distribution across different beam sectors during (a) the exploration phase and (b) the exploitation phase after decision given by C-MAB algorithm.

5.2. The Impact of IRSs on Connection Time

The performance of C-MAB with and without IRS is also shown in Table 2. We assume that an IRS can reflect one of the beams from a BS in a 180 degree perimeter based on its orientation, which is fixed, providing coverage to one vehicle. Even though the IRS is passive and thus incurs in more path loss, when strategically placed they achieve more than a doubling in average connection time with the traditional greedy approach. This is because it is difficult for a BS to be in LoS with more than one optimum connection area, such as the waiting zone of a traffic light or a blind spot. Therefore, by using the IRS as a “mirror”, these spots become “visible” to the BS. When paired with a C-MAB algorithm, the new connection options can be more intelligently exploited, resulting in more than a quadrupling of connection time. The proposed algorithm achieves this result by “memorizing” the conditions that lead to the longest and shortest connection times. Therefore, we try to avoid connections to vehicles that find themselves in conditions that might lead to short connections and choose instead those vehicles with conditions prone to long connections.

Table 2. The impact of IRS usage on connection time with comparison of average connection time (in seconds) during exploration and exploitation phases with and without IRS assistance.

Case	Exploration	Exploitation
With IRS	5.78	13.51
Without IRS	2.28	3.28

5.3. The Impact of the Traffic Light Context on Connection Time

Our simulation has generated various insightful examples, some of which are straightforward while others are more subtle. Generally, vehicles approaching a red traffic light tend to have longer connections compared to those approaching green lights. However, multiple context variables and numerous connection points are evaluated simultaneously, creating interactions that can amplify or diminish individual effects, thus increasing the complexity of the analysis. Figure 6 specifically evaluates the combined contexts related to traffic lights. To produce this figure, simulation data was processed by calculating the average connection duration for each combination of contexts observed. The range between the shortest and longest bars indicates the discriminatory power of these combined

contexts—wider variations signify stronger predictive capabilities. For instance, the bar with the longest average connection time corresponds to a scenario labeled “Green, Small, Away, Far”, reaching approximately 24 s, although intuitively a vehicle moving away from a green light that is about to change does not have a relationship directly. This result emerges, in the considered scenario where two traffic lights are at the end main road, when a vehicle moving away from one traffic light is simultaneously approaching another traffic signal at the opposite end of the main road and eventually waits at the second light for an extended period.

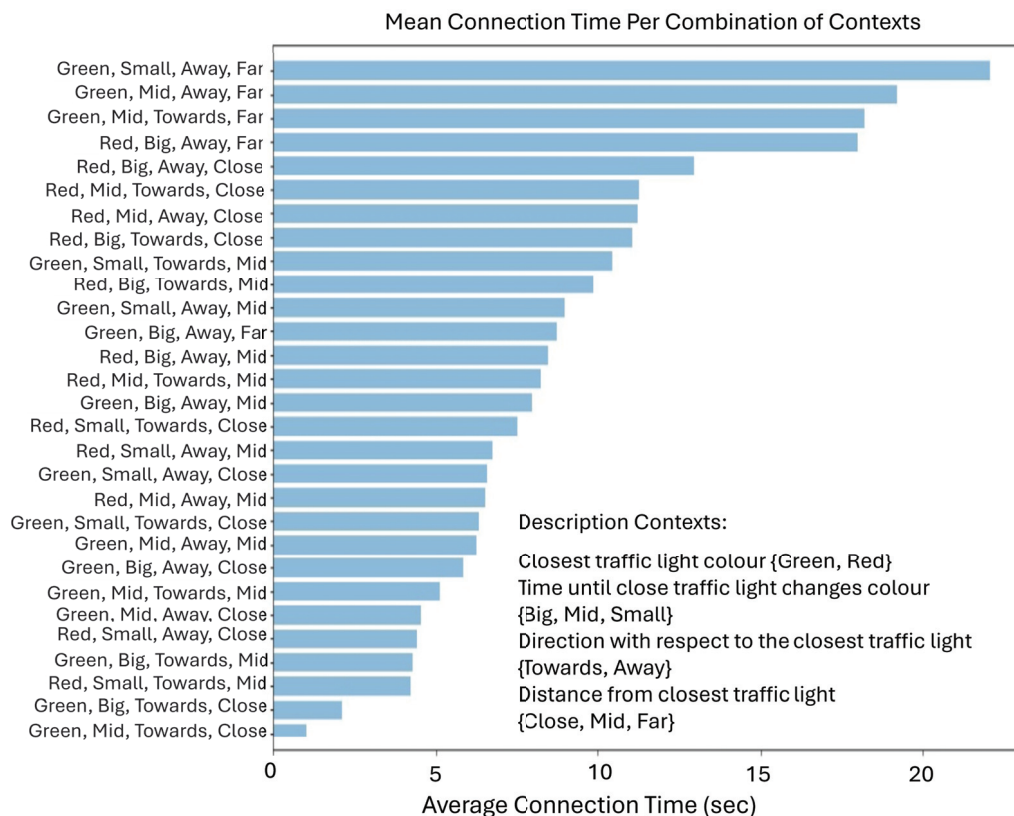


Figure 6. The impact of the traffic light on connection time.

Conversely, scenarios positioned at the shorter end of the chart, such as “Green, Mid, Towards, Close,” yield predictably shorter connection durations due to vehicles swiftly passing through a green light without stopping. A representative middle-range scenario, for example, the eighth-longest bar (“Red, Big, Towards, Close”), illustrates a vehicle approaching a nearby red traffic signal that will remain red for a longer time, thus extending the connection duration slightly beyond the waiting time at the light. These nuanced findings underline the value of leveraging machine learning to determine, through iterative exploration, which combinations of contexts genuinely maximize connectivity performance. Rather than relying solely on intuitive assumptions, this empirical approach ensures that subtle yet critical interactions between multiple contextual variables are effectively identified and utilized.

6. Conclusions

As new technologies constantly emerge to improve the future generation networks, some existing techniques in mobility management may become under-performing which directly challenge the overall network performance. Research interests in mobility management will be renewed to seek for improved solutions or novel solutions to integrate with the emerging technologies. The use of IRSs immediately challenge the beam-based mobility

management by creating more complex multi-cell multi-IRS scenarios where multi-point beam switching will become a major problem and require new techniques and analytical approaches to maintain connectivity across the various network attachment points.

In scenarios with large networks high density of BSs, IRSs, and vehicles, the action space expands to include multi-beam–IRS combinations. This can be accommodated by extending the reward table and leveraging spatial grouping of IRS regions to reduce complexity. The proposed C-MAB framework is also highly scalable and remains computationally feasible under increasing vehicle density. As more vehicles enter the network, the frequency of context–beam interactions rises, which accelerates convergence and enhances learning robustness. The algorithm remains efficient due to its context-conditioned structure, which naturally partitions the learning problem. Exploring more sophisticated coordination among IRS panels, including shared context features, is part of our ongoing and future research.

Author Contributions: Conceptualization, A.K. and H.L.; algorithm design, R.S.d.V. and A.K.; software, R.S.d.V. and A.K.; writing—original draft, A.K. and R.S.d.V.; visualization, R.S.d.V. and A.K.; writing—review and editing, H.L. and A.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Li, S.; Xu, L.L.; Zhao, Y.; Liu, J.; Wei, X. Intelligent reflecting surface (IRS) enabled mmWave communication: Joint beamforming design and IRS phase shift optimization. *arXiv* **2020**, arXiv:2002.11625.
2. Zhao, Y.; Xu, L.L.; Zhang, Y.; Liu, J.; Wei, X. IRS-aided mmWave v2x communication: Joint power allocation and IRS phase shift optimization. *IEEE Trans. Veh. Technol.* **2021**, *70*, 1–14.
3. Jameel, F.; Wyne, S.; Nawaz, S.J.; Chang, Z. Propagation channels for mmWave vehicular communications: State-of-the-art and future research directions. *IEEE Wirel. Commun.* **2019**, *26*, 144–150. [CrossRef]
4. Pan, C.; Ren, H.; Wang, K.; Xu, W.; Elkashlan, M.; Nallanathan, A.; Hanzo, L. Multicell MIMO communications relying on Intelligent Reflecting Surfaces. *IEEE Trans. Wirel. Commun.* **2020**, *19*, 5218–5233. [CrossRef]
5. Sheng, Z.; Pressas, A.; Ocheri, V.; Ali, F.; Rudd, R.; Nekovee, M. Intelligent 5G vehicular networks: An integration of DSRC and mmWave communications. In Proceedings of the 2018 International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Republic of Korea, 17–19 October 2018; pp. 571–576.
6. Perovi, N.S.; Renzo, M.D.; Flanagan, M.F. Channel capacity optimization using reconfigurable intelligent surfaces in indoor mmWave environments. In Proceedings of the ICC 2020—2020 IEEE International Conference on Communications (ICC), Virtual, 7–11 June 2020; pp. 1–7.
7. Cao, X.; Wang, S.; Ren, X. IRS-Enhanced V2X Communication and Computation Systems: Resource Allocation and Performance Optimization. *IEEE Internet Things J.* **2024**, *12*, 9180–9194. [CrossRef]
8. Wang, J.; Zhang, W.H.; Bao, X.; Song, T.; Pan, C. Outage analysis for intelligent reflecting surface assisted vehicular communication networks. *arXiv* **2019**, arXiv:2004.08063.
9. Makarf, A.; Rabie, K.M.; Kaiwartya, O.; Adhikari, K.; Li, X.; Quiroz-Castellanos, M.; Kharel, R. Reconfigurable intelligent surfaces-enabled vehicular networks: A physical layer security perspective. *arXiv* **2020**, arXiv:2004.11288.
10. Zhu, Y.; Mao, B.; Kato, N. Intelligent reflecting surface in 6G vehicular communications: A survey. *IEEE Open J. Veh. Technol.* **2022**, *3*, 266–277. [CrossRef]
11. Khan, W.U.; Mahmood, A.; Bozorgchenani, A.; Jamshed, M.A.; Ranjha, A.; Lagunas, E.; Pervaiz, H.; Chatzinotas, S.; Ottersten, B.; Popovski, P. Opportunities for intelligent reflecting surfaces in 6G-empowered V2X communications. *arXiv* **2022**, arXiv:2210.00494. [CrossRef]

12. Jia, C.; Gao, H.; Chen, N.; He, Y. Machine Learning Empowered Beam Management for Intelligent Reflecting Surface Assisted MmWave Networks. *arXiv* **2020**, arXiv:2003.01306. [CrossRef]
13. He, J.; Wymeersch, H.; Juntti, M. Leveraging location information for RIS-aided mmWave MIMO communications. *IEEE Wirel. Commun. Lett.* **2021**, *10*, 1380–1384. [CrossRef]
14. Ozcan, Y.U.; Ozdemir, O.; Kurt, G.K. Reconfigurable intelligent surfaces for the connectivity of autonomous vehicles. *IEEE Trans. Veh. Technol.* **2021**, *70*, 2508–2513. [CrossRef]
15. Lee, Y.; Lee, J.H.; Ko, Y.C. Beamforming optimization for IRS-assisted mmWave V2I communication systems via reinforcement learning. *IEEE Access* **2022**, *10*, 60521–60533. [CrossRef]
16. You, C.; Zheng, B.; Zhang, R. Fast beam training for IRS-assisted multiuser communications. *IEEE Wirel. Commun. Lett.* **2020**, *9*, 1845–1849. [CrossRef]
17. Dampahalage, D.; Manosha, K.S.; Rajatheva, N.; Latva-aho, M. Intelligent reflecting surface aided vehicular communications. In Proceedings of the 2020 IEEE Globecom Workshops (GC Wkshps. IEEE, 2020), Taipei, Taiwan, 7–11 December 2020; pp. 1–6.
18. Song, Y.; Xu, S. Robust Beamforming for IRS-Aided Multi-cell mmWave Communication Systems. *IEEE Trans. Veh. Technol.* **2023**, *72*, 9189–9205. [CrossRef]
19. ETSI. Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications. Part 3: Specifications of Decentralized Environmental Notification Basic Service. ETSI EN 302 637-3 V1.2.2. 2014. Available online: https://www.etsi.org/deliver/etsi_en/302600_302699/30263703/01.02.02_60/en_30263703v010202p.pdf (accessed on 22 June 2025).
20. Li, D.; Wang, S.; Zhao, H.; Wang, X. Context-and-Social-Aware Online Beam Selection for mmWave Vehicular Communications. *IEEE Internet Things J.* **2021**, *8*, 8603–8615. [CrossRef]
21. Rappaport, T.S.; MacCartney, G.R.; Samimi, M.K.; Sun, S. Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design. *IEEE Trans. Commun.* **2015**, *63*, 3029–3056. [CrossRef]
22. Troch, A.; Peeters, M.; Marquez-Barja, J. Millimeter Wave Cellular Networks for Teleoperated Vehicles: A Simulation Study. In Proceedings of the 2024 IEEE 100th Vehicular Technology Conference (VTC2024-Fall), Washington, DC, USA, 7–10 October 2024; pp. 1–5. [CrossRef]
23. Keshavamurthy, B. Statistical Characterization of 28GHz V2X Channels via Autonomous Beam-Steered Measurements. *IEEE Dataport* **2024**. [CrossRef]
24. Kose, A.; Foh, C.H.; Lee, H.; Moessner, K. Profiling Vehicles for Improved Small Cell Beam-Vehicle Pairing Using Multi-Armed Bandit. In Proceedings of the 2021 IEEE International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Republic of Korea, 20–22 October 2021; pp. 221–226.
25. Kose, A.; Lee, H.; Foh, C.H.; Shojafar, M. Multi-Agent Context Learning Strategy for Interference-Aware Beam Allocation in mmWave Vehicular Communications. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 7477–7493. [CrossRef]
26. Cassillas, A.M.; Kose, A.; Lee, H.; Foh, C.H.; Leow, C.Y. Contextual Multi-Armed Bandit based Beam Allocation in mmWave V2X Communication under Blockage. In Proceedings of the 2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring), Florence, Italy, 20–23 June 2023; pp. 1–6.
27. Wu, Q.; Zhang, R. Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network. *IEEE Commun. Mag.* **2019**, *58*, 106–112. [CrossRef]
28. Duchoň, F.; Babinec, A.; Kajan, M.; Beňo, P.; Florek, M.; Fico, T.; Jurišica, L. Path Planning with Modified a Star Algorithm for a Mobile Robot. *Procedia Eng.* **2014**, *96*, 59–69. [CrossRef]
29. Va, V.; Choi, J.; Heath, R.W. The impact of beamwidth on temporal channel variation in vehicular channels and its implications. *IEEE Trans. Veh. Technol.* **2016**, *66*, 5014–5029. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Incoherent Optical Neural Networks for Passive and Delay-Free Inference in Natural Light

Rui Chen ^{1,2,3}, Yijun Ma ^{1,2,3}, Zhong Wang ^{1,2,3} and Shengli Sun ^{1,3,*}

¹ Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai 200083, China; chenrui@mail.sitp.ac.cn (R.C.); mayijun@mail.sitp.ac.cn (Y.M.); wangzhong@mail.sitp.ac.cn (Z.W.)

² University of Chinese Academy of Sciences, Beijing 100049, China

³ Key Laboratory of Intelligent Infrared Perception, Chinese Academy of Sciences, Shanghai 200083, China

* Correspondence: palm_sun@mail.sitp.ac.cn

Abstract: Optical neural networks are hardware neural networks implemented based on physical optics, and they have demonstrated advantages of high speed, low energy consumption, and resistance to electromagnetic interference in the field of image processing. However, most previous optical neural networks were designed for coherent light inputs, which required the introduction of an electro-optical conversion module before the optical computing device. This significantly hindered the inherent speed and energy efficiency advantages of optical computing. In this paper, we propose a diffraction algorithm for incoherent light based on mutual intensity propagation, and on this basis, we established a model of an incoherent optical neural network. This model is completely passive and directly performs inference calculations on natural light, with the detector directly outputting the results, achieving target classification in an all-optical environment. The proposed model was tested on the MNIST, Fashion-MNIST, and ISDD datasets, achieving classification accuracies of 82.32%, 72.48%, and 93.05%, respectively, with experimental verification showing an accuracy error of less than 5%. This neural network can achieve passive and delay-free inference in a natural light environment, completing target classification and showing good application prospects in the field of remote sensing.

Keywords: natural light processing; mutual intensity propagation; incoherent optical neural network; target classification

1. Introduction

Traditional remote sensing detection methods, particularly electronic computing-based image processing algorithms, face numerous challenges, including slow computation speed and high energy consumption when handling large-scale remote sensing data. These issues make it difficult to meet the demands for real-time performance and low power consumption. As remote sensing technology rapidly advances, the number of remote sensing platforms and sensors continues to increase and is constantly being optimized, leading to a rapid growth in the volume of massive multi-source remote sensing data [1–3]. Therefore, developing an efficient and low-energy image processing technology has become an important research direction in the field of remote sensing detection. In recent years, optical neural networks (ONNs), as hardware neural networks implemented based on physical laws, have shown great potential in the field of deep learning [4–7]. Compared with traditional electronic computing, optical neural networks have advantages such as high speed, low energy consumption, and resistance to electromagnetic interference, making them particularly suitable for handling large-scale parallel computing tasks. They

have already achieved good development in fields such as computer vision [8–20], optical communication [21–25], and multimodal signal processing [26–36]. However, most existing optical neural networks are designed for coherent light inputs, which means that in practical applications, an electro-optical conversion module needs to be introduced before the optical computing device. This conversion not only increases the complexity of the system but also significantly hinders the inherent speed and energy efficiency advantages of optical computing, limiting the application of optical neural networks in actual target detection tasks.

To overcome this limitation, researchers have begun to explore the design methods of incoherent optical neural networks (ICONN). Incoherent optical neural networks aim to directly receive natural light input for inference calculations, avoiding the step of electro-optical conversion, thereby preserving the high speed and low energy consumption characteristics of optical computing [37]. M. Kleiner et al. discussed how the performance of diffractive neural networks is affected by the coherence length and coherence time of the light source, and they proposed optimization schemes for network design with incoherent light inputs [38]. M.S.S. Rahman et al. studied the use of spatially incoherent diffractive processors to achieve universal linear intensity transformations, demonstrating the application potential of incoherent optical neural networks in fields such as imaging and signal processing [39]. In the same year, the team also investigated the use of diffractive optical neural networks for signal restoration in the presence of arbitrary opaque occlusions, showing the application of robust optical communication systems in complex environments [40]. B. Rahmani et al. proposed a method for transmitting full-optical images through multimode fibers using incoherent light, employing diffractive optical neural networks to compensate for distortions caused by multimode fibers, indicating the feasibility of using incoherent light in diffractive neural-network-based optical communication and imaging systems [7]. The method proposed by S. Sun's team, based on partially coherent optical neural networks (PCONN), although theoretically promising, is limited in practical applications due to the complexity of coherence measurements, which restricts its performance [41]. The incoherent all-optical methods mentioned above still have defects such as difficulty in training, large simulation errors, and the need for electrical domain activation.

In response to the aforementioned issues, this paper proposes a diffractive algorithm for incoherent light based on mutual intensity propagation [42–44] and establishes a model of an incoherent optical neural network on this basis. The model is completely passive, directly performing inference calculations on natural light, and it does not require any electro-optical conversion modules before outputting the light intensity at the detection surface, thereby achieving target classification in an all-optical environment. The main contributions of this paper include (1) proposing a new diffractive algorithm for incoherent light that can effectively handle natural light inputs; (2) establishing an incoherent optical neural network model based on this algorithm, achieving passive inference, and conducting simulation tests on the MNIST, Fashion-MNIST, and ISDD datasets, achieving accuracies of 82.32%, 72.48%, and 93.05%, respectively, and analyzing various factors that may affect the model's performance; (3) setting up an optical platform to experimentally verify the proposed model, with an error in simulation accuracy of less than 5%, proving the model's effectiveness and robustness in a natural light environment.

The method proposed in this paper still has room for further development. First, the model has currently only been tested and verified on image classification datasets, and in the future, it can be further tested and optimized for object detection tasks on large-scale remote sensing datasets. Second, because the model directly receives natural light input, it is difficult for the network to generate nonlinear activation, which limits the network's performance on complex tasks. Possible solutions include exploring weak-

light nonlinear mechanisms for natural light or introducing a small amount of electrical domain computation to provide nonlinearity. Future research will focus on solving these problems to further enhance the performance of incoherent optical neural networks in practical applications.

2. Methods

2.1. Subsection

There are two commonly used analytical algorithms for the diffractive transmission of incoherent light. One is based on Fourier optics, which calculates the output through spectral conversion [37,45,46]. This method is usually suitable for simple imaging systems; otherwise, the propagation modes of different spectral components will vary, causing the system to be space-variant and resulting in errors. The other is the random phase superposition method based on statistics [39,47]. This method interprets the input of incoherent light fields as the superposition of multiple coherent lights with random initial phases. These coherent lights are transmitted independently, and then the field strength is calculated separately at the output surface before performing an incoherent average. To obtain a stable output, this method typically requires a high number of superimposed layers, resulting in high computational power demand. Moreover, both of the above algorithms require the input to be completely incoherent light for calculation. In other words, when there are changes in coherence within the system, they cannot perform segmented optimization. In practical applications, especially in the field of remote sensing, long-distance transmission often causes the coherence of the light field to change in various ways. In this paper, we adopted a transmission theory for incoherent light fields based on mutual intensity propagation. This method does not rely on statistical laws, which reduces the computational power demand to a certain extent. Moreover, since the coherence of the light field is fully preserved, the calculation can start at any point in the system. Therefore, the system can be modeled in segments, corresponding to the multi-layer network in neural network algorithms, allowing for layer-by-layer optimization. The following is the derivation of this transmission model.

In nature, targets can be regarded as incoherent light at the initial plane, that is, the degree of partial coherence is 0. Therefore, the input light field can be established as

$$J(P_1, P_2) = I(P_1)\delta(P_1 - P_2) \quad (1)$$

In Equation (1), J represents the mutual intensity, I represents the light field intensity, P_1 and P_2 represent the initial plane of the light field emitted by the target, and the delta function indicates that the value of J is equal to I only when P_1 and P_2 overlap. When P_1 and P_2 do not overlap, $J = 0$, indicating complete incoherence.

The general propagation formula for partial coherence can be obtained from the statistical law of Rayleigh Sommerfeld diffraction formula, that is,

$$\begin{aligned} J(Q_1, Q_2) &= \langle u^*(Q_1)u(Q_2) \rangle \\ &= \frac{1}{4\pi^2} \iint \iint J(P_1, P_2) \exp[jk(r_2 - r_1)] \frac{K(\theta_1)}{r_1} \frac{K(\theta_2)}{r_2} \left(\frac{1}{r_1} + jk\right) \left(\frac{1}{r_2} - jk\right) dS_1 dS_2 \end{aligned} \quad (2)$$

In Equation (2), $J(Q_1, Q_2)$ represents the distribution of mutual intensity at the output plane; r_1 and r_2 represent the distances between P_1, Q_1 and P_2, Q_2 , respectively; and $K(\theta_1)$ and $K(\theta_2)$ represent the tilt factors in diffraction. The above equation represents the mapping relationship between the mutual intensity of two planes in free space, and

there is no approximation, which follows the space-invariant principle. Therefore, it can be established as a convolution formula:

$$\begin{aligned} J(Q_1, Q_2) &= \iint J(P_1, P_2) h(P_1) h^*(P_2) dS_1 dS_2 \\ &= J(P_1, P_2) * h(P_1, P_2, Q_1, Q_2) \end{aligned} \quad (3)$$

In Equation (3), h is the point spread function of mutual intensity in free space. Further solving the Fourier transform of this point spread function allows us to transfer Equation (2) to the frequency domain calculation, which means that we can greatly accelerate our calculation process using the fast Fourier transform (FFT) algorithm. As mentioned earlier, this method can perform segmented optimization on the network.

After the light field has been transmitted, we can process the light field characterized by partial coherence according to the system requirements to obtain the desired information. In our task, what we need is the intensity of the final detection plane. Assuming the output surface coordinates are $Q'(x, y)$, it can be obtained by setting $Q'_1 = Q'_2$:

$$I(Q'_1) = J(Q'_1, Q'_2)|_{Q'_1=Q'_2} \quad (4)$$

In addition, we also note that in remote sensing scenarios, the targets are usually very far from the system, and this distance usually satisfies the Fresnel approximation. Then, we can derive the Van Cittert–Zernike theorem in partial coherent optics:

$$J(x_1, y_1; x_2, y_2) = \frac{\exp(j\psi)}{(\lambda z)^2} \iint I(\alpha, \beta) \exp\left\{-j\frac{2\pi}{\lambda z}(\Delta x \alpha + \Delta y \beta)\right\} d\alpha d\beta \quad (5)$$

In Equation (5), α and β are the input surface coordinates, and x and y are the output surface coordinates. The meaning of this theorem is that when the size of the incoherent target itself and the observation plane are much smaller than the distance between them, the mutual intensity of the observation plane is only related to the relative distance between the two points. Specifically, the mutual intensity is proportional to the Fourier transform of the target intensity. This theorem is also important in our network, as it greatly simplifies the computational complexity of the input light field when mutual intensity is used as the propagation dimension, allowing us to directly establish the neural network model from the first acceptance plane of the system.

2.2. Principle of ICONN

The ICONN model designed in this paper is mainly divided into two modules: the optical input module and the optical inference module, as shown in Figure 1. In the left block diagram, the optical field input module is depicted, where the first lens serves as the system's optical aperture, primarily used for imaging. According to the resolution requirements, an aperture is set on the imaging plane of this input module to filter out a specified small field of view. The small field of view is then transformed into partially coherent light by the second lens and output to the subsequent inference module. The main functions of this input module are light collection, image plane segmentation, and preserving the spatial coherence distribution of the light field.

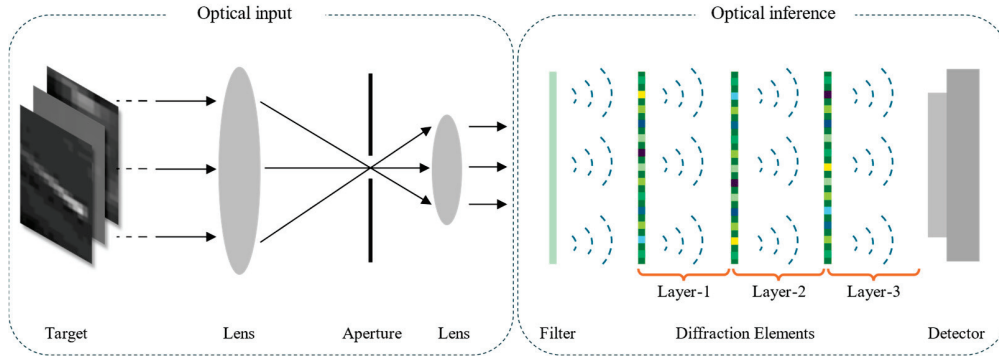


Figure 1. System structure diagram of the incoherent optical neural network.

The inference module mainly consists of multiple layers of incoherent optical neural networks. The optical neural network is composed of discrete diffractive devices. The spatial light field is modulated by the diffractive devices with linear phase modulation, and after diffracting a certain distance in free space, it is modulated again by the next layer of diffractive devices. After multiple layers of modulation and diffraction, the label distribution preset during the algorithm training phase can be obtained. The label is further measured by the output module. The modulation range of the diffractive devices is from 0 to 2π . During the training phase, the data type of the phase modulation quantity is set to float32, and it is re-quantified in the subsequent output manufacturing stage. We will discuss the impact of this quantization on network performance later in the text.

Here, we consider one modulation plus one free-space diffraction as one layer of the optical neural network. Therefore, a single-layer optical neural network can be modeled as

$$w_{ij}^l(P_i, P_j) = \left(\frac{z_l}{\lambda r_i r_j}\right)^2 \exp\left[\frac{j2\pi(r_j - r_i)}{\lambda}\right] \quad (6)$$

$$n_{ij}^l(P_i, P_j) = w_{ij}^l(P_i, P_j) \cdot t_i^l(:, :, x_i, x_i) \cdot t_j^{l*}(x_j, x_j, :, :) \cdot \sum_{m,n} n_{m,n}^{l-1}(P_m, P_n) \quad (7)$$

In Equation (6), $w_{ij}^l(P_i, P_j)$ represents the mutual intensity propagation mode of the four-dimensional sub-wave source (P_i, P_j) , where l denotes the layer number in the network, and i and j represent a pair of points on the P -plane. z_l is a constant representing the center distance between the l -th layer and the next layer. λ represents the center wavelength of the optical field under the quasi-monochromatic approximation, and r_i and r_j represent the distances from the sub-wave source to the output point. In Equation (7), $n_{ij}^l(P_i, P_j)$ represents the output function of the (i, j) -th neuron formed by (P_i, P_j) in the l -th layer of the network, where t_i^l denotes the mutual intensity modulation coefficient of the diffractive element.

There is a filter in front of the diffractive device, whose main function is to improve the temporal coherence of the light field and provide quasi-monochromatic light input for the modulation device. The final output result of the system is received by the detector, and the classification result is determined by the label area set during the training phase. The loss of the network is represented as

$$\text{Loss} = \text{MSE}(I_{gt} \cdot \text{scale}, I) \quad (8)$$

In Equation (8), we adopted the mean squared error (MSE) method to calculate our loss. I_{gt} represents the label distribution, and different classification tasks correspond to different label settings. I_{real} represents the network output result, and scale represents a scaling factor for the output. Since the output of the optical neural network usually does

not perfectly match the label setting, appropriately scaling the label can help the network converge faster.

Optical neural networks typically have a small field of view, and our system also has this characteristic because we placed an aperture on the focal plane of the first lens to perform image plane filtering. However, we can achieve a larger field of view by arraying the system or through line scanning and area scanning methods. That is, our system can perform more complex tasks in a natural light environment, such as object detection, without the need for prior electro-optical conversion, followed by image cutting and using lasers and spatial light modulators to provide input for the subsequent inference model. Our method is expected to achieve real-time processing of massive remote sensing data.

3. Simulation and Experiment Results

3.1. Simulation Strategy

The training and simulations were carried out in a CentOS 7.9.2009 environment, using an Nvidia@4090 graphics card and an AMD EPYC 7282 CPU. The simulation tests included performance evaluations of ICONNs with different numbers of layers, while the experimental tests focused on a single-layer ICONN. We will see in the following text that the simulation and experimental results for the single-layer case are in good agreement.

We conducted simulation tests on the ICONN model proposed earlier, with the detector ground truth settings shown in Figure 2. The ground truth is circularly distributed around the center of the effective detection area. The average light intensity within each square detection region corresponds to the score of a category, and the category with the highest average light intensity is the output category.

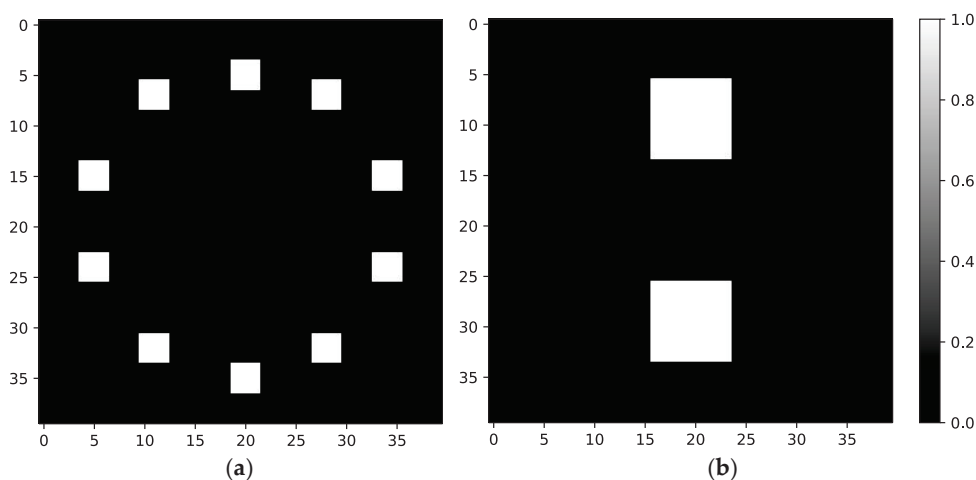


Figure 2. (a) The detector ground truth for a ten-class classification task. The overall effective detection area is 40×40 pixels, the scoring box size is 3×3 pixels, the distribution radius is 15 pixels, and the pixel size is $64 \mu\text{m}$. (b) The detector ground truth for a two-class classification task. The overall effective detection area is 40×40 pixels, the scoring box size is 8×8 pixels, the distribution radius is 10 pixels, and the pixel size is $64 \mu\text{m}$.

During the simulation and experimental testing phases, we used the general image classification datasets MNIST and Fashion MNIST, as well as an infrared ship detection dataset, ISDD, to evaluate the performance of the ICONN model. The first two image classification datasets are ten-class, corresponding to the ground truth in Figure 2a. For the ISDD dataset, we performed preprocessing to conduct binary classification, i.e., ship present and ship absent, corresponding to the ground truth in Figure 2b. The specific processing method is as follows: First, we read the labels (boxes) from the annotation files in the dataset and extracted fixed-width images centered on the label boxes as positive

samples. Here, we set the width to 40 pixels, which has been verified to cover over 95% of the target bounding boxes. Then, we randomly generated images from the large image that do not overlap with the labeled bounding boxes and whose pixel grayscale values are not all less than 10 as negative samples. The ratio of positive to negative samples was approximately 1:1. The extracted small images were used as the new classification dataset for subsequent simulation and experimental testing. All datasets were divided into training and testing sets in a 4:1 ratio, with a batch size of 20. The loss function was the mean squared error (MSE) between the network output and the ground truth. The network operated entirely in the optical domain without the addition of activation functions, and the learning rate was set to 0.001.

3.2. Simulation Result and Analysis

First, we conducted tests on the general image classification datasets MNIST and Fashion MNIST. The training was iterated for 30 epochs, and the performance of networks with 1 to 3 layers was tested. The simulation test accuracies were 79.75%, 82.11%, and 82.32% for MNIST, and 69.76%, 71.67%, and 72.48% for Fashion MNIST. The training results and inference demonstration are shown in Figures 3 and 4. From the test results, we can observe that the performance of the ICONN network increased with the number of layers. However, since the ICONN network is a purely linear light intensity mapping, this increase is not significant compared to fully connected networks in the electrical domain or CONN/PCONN with one layer of nonlinear activation. Nevertheless, our simulation demonstration clearly shows that after increasing the number of layers, the label distribution on the detection surface was closer to the ideal ground truth. In other words, most of the energy was concentrated within the label boxes, and the proportion of light intensity for the correct label was higher. This can avoid misjudgments caused by insufficient detector sensitivity, noise, pixel alignment errors, etc., and is still helpful in practical detection tasks.

We also conducted tests on an open-source infrared remote sensing ship detection dataset. The network was trained for 30 epochs, and the performance of networks with 1 to 3 layers was tested. In line with the application characteristics of optical neural networks, we processed the remote sensing images based on the annotations in the dataset, with the processing method described in Section 3.1. The training parameters for the network were consistent with those in the previous section. The simulation test accuracies for networks with 1 to 3 layers were 92.15%, 92.89%, and 93.05%, respectively. The training results and inference demonstration are shown in Figure 5. Similar conclusions can be drawn from the test results: the performance of ICONN did not significantly improve with an increase in the number of layers, but the label distribution became more ideal and easier to measure. Moreover, the excellent performance of our model on the ISDD dataset demonstrated its great potential in the field of remote sensing detection.

In addition to the number of layers, we conducted simulation tests on other parameters that may affect the performance of ICONN, including modulation plane pixel density, layer spacing, and phase modulation bit depth. The modulation plane pixel density refers to the number of pixels contained in the modulation plane when the area of the modulation layer remains unchanged. This value can be adjusted by zero-padding the input image. The layer spacing refers to the diffraction distance of light in free space after passing through one layer of modulation. The phase modulation bit depth refers to the modulation precision of the modulation plane. In each of the above tests, the method of controlling variables was used, and the size of the input image and the size of the modulation layer remained unchanged. The results are shown in Figure 6.

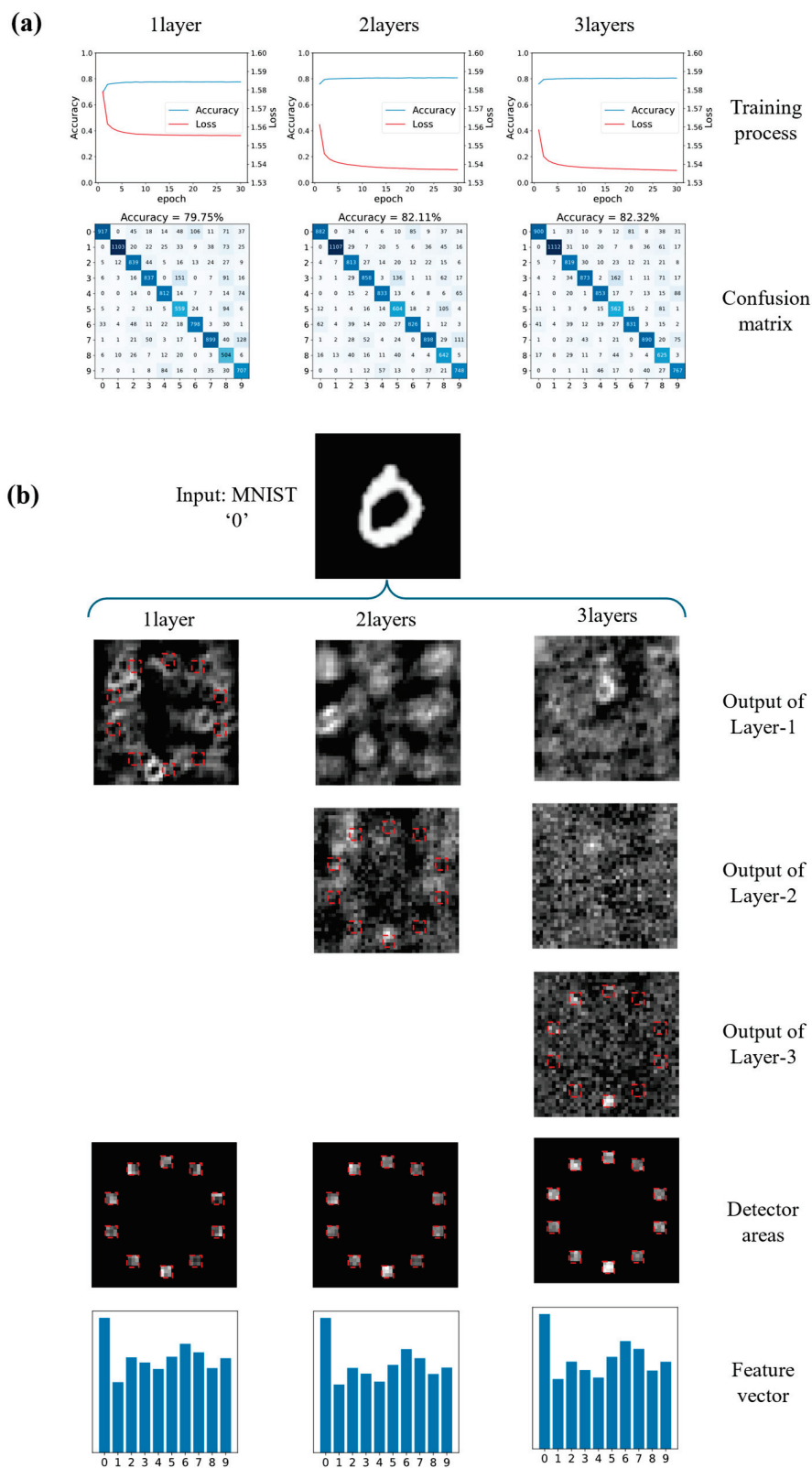


Figure 3. (a) The training process of ICONN on MNIST and the confusion matrix on the test set after training. From left to right are the single-layer, two-layer, and three-layer networks, with test set accuracies of 79.75%, 82.11%, and 82.32%, respectively. (b) When the handwritten digit “0” is input, the output demonstration of each layer of the network occurs, and the ten-dimensional feature vector is calculated according to the set label, as indicated by the red box in the figure, which represents the average light intensity in each of the ten scoring boxes. The first region is the brightest, indicating that the network correctly assigns the input to the label “0”.

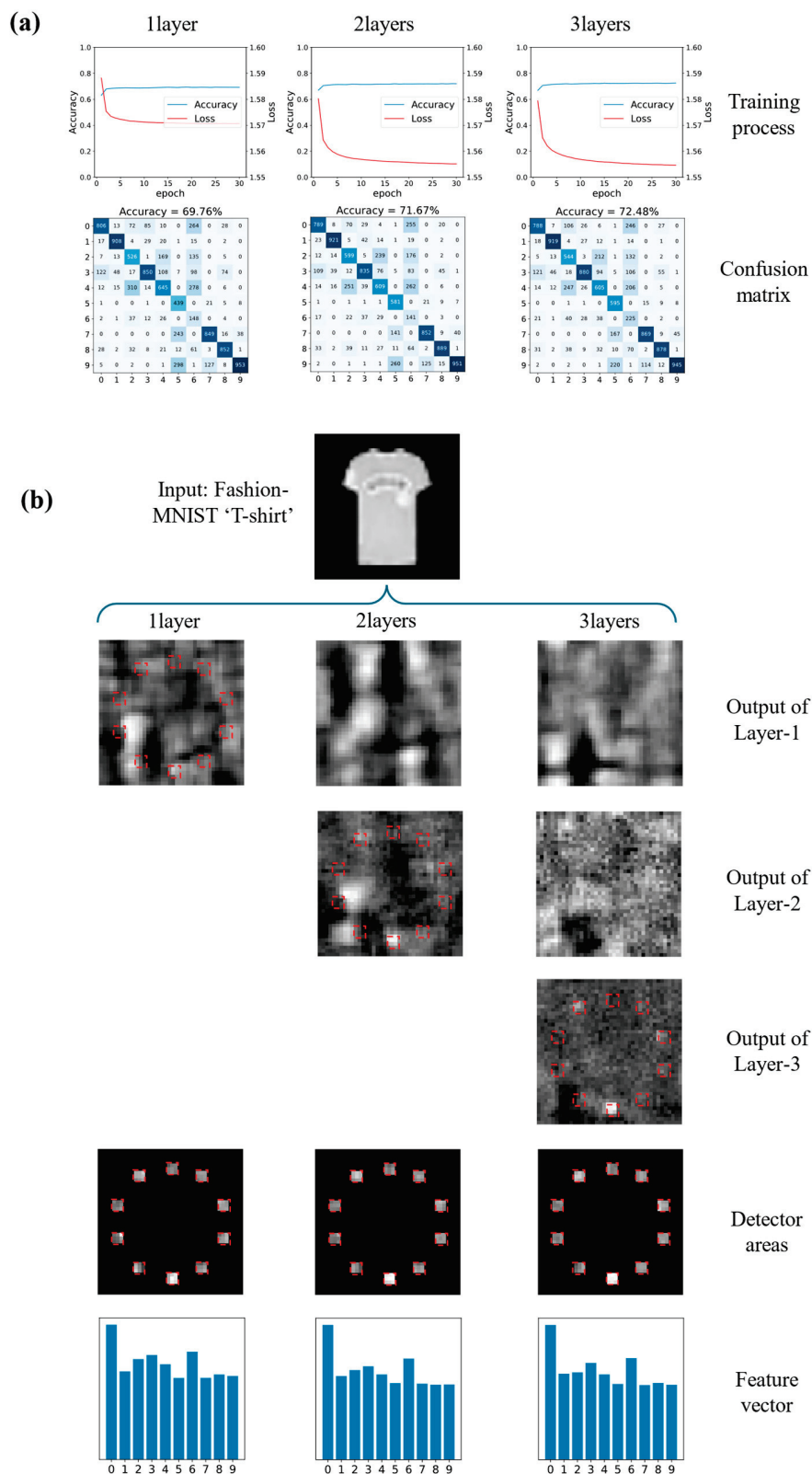


Figure 4. (a) The training process of ICONN on Fashion MNIST and the confusion matrix on the test set after training. From left to right are the single-layer, two-layer, and three-layer networks, with test set accuracies of 69.76%, 71.67%, and 72.48%, respectively. (b) When the fashion item “T-shirt” is input, the output demonstration of each layer of the network occurs, and the ten-dimensional feature vector is calculated according to the set label, as indicated by the red box in the figure, which represents the average light intensity in each of the ten scoring boxes. The first region is the brightest, indicating that the network correctly assigns the input to the label “T-shirt”.

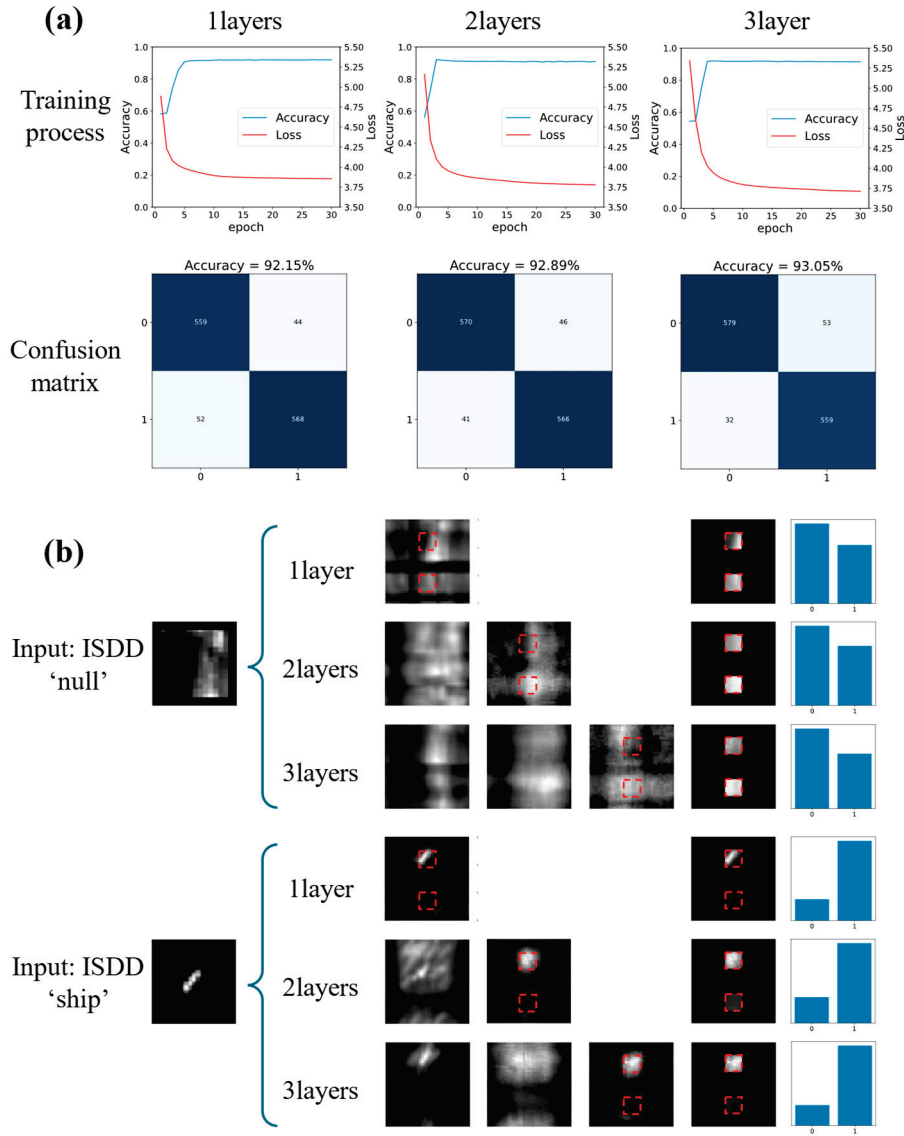


Figure 5. (a) The training process of ICONN on ISDD and the confusion matrix on the test set after training. From top to bottom are the single-layer, two-layer, and three-layer networks, with test set accuracies of 92.15%, 92.89%, and 93.05%, respectively. (b) When the label “ship” is input, the output demonstration of each layer of the network occurs, and the two-dimensional feature vector is calculated according to the set label, as indicated by the red box in the figure, which represents the average light intensity in each of the two scoring boxes. The first region is the brightest, indicating that the network correctly assigns the input to the label “ship”.

The results show that for the three image classification datasets we tested, the original image resolution and the modulation device resolution had little impact on the network’s performance. However, the layer spacing, or more precisely, the Fresnel number F , which is related to the diffraction effect, had a significant impact on the network’s performance. The Fresnel number is defined as follows:

$$F = \frac{d^2}{\lambda L} \quad (9)$$

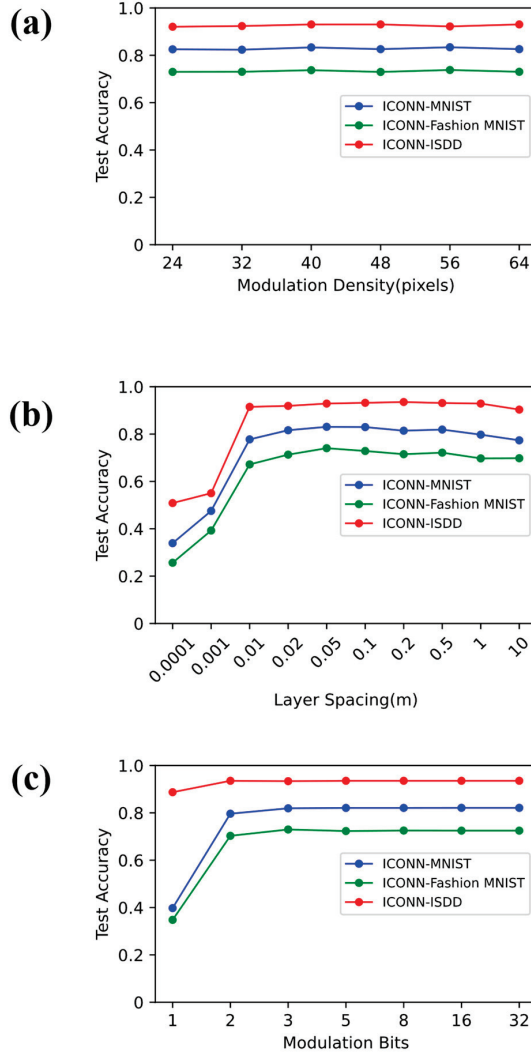


Figure 6. Tests of various parameters affecting the performance of ICONN. In the tests, the input image size and the modulation plane size were fixed, and the network consisted of three layers. (a) Modulation density test: Different pixel densities of the spectrum and modulation layer were obtained by adjusting the zero-padding of the input plane. The layer spacing was fixed at 0.1 m, and the phase modulation bit depth was fixed at 32 bit. (b) Layer spacing test: The number of modulation plane pixels was fixed at 40×40 , and the phase modulation bit depth was fixed at 32 bit. (c) Modulation bits test: The number of modulation plane pixels was fixed at 40×40 , and the layer spacing was fixed at 0.1 m.

In Equation (9), d represents the width of the diffraction plane, i.e., the modulation device, and L represents the diffraction distance, i.e., the layer spacing. When the area of our modulation device is fixed, if the layer spacing is set too small, the diffraction effect between layers is weak, and the light field propagation is more akin to geometric optics. The energy allocation effect brought by phase modulation is very weak, making it difficult for the model to converge. As the layer spacing increases and the diffraction between layers approaches Fresnel diffraction, the phase modulation can have a more pronounced energy allocation effect, making the output results more likely to converge to the preset label region. For this reason, in diffractive neural networks, increasing the pixel density of the modulation plane, or the number of neurons, has a limited impact on ICONN. This is because ICONN is driven by physical computation of diffraction effects, rather than the mathematically driven computation of electrical neural networks based on parameter scale. The physical propagation laws of light will somewhat limit the input–output mapping

relationship. Our simulation results indicate that the network performance is optimal and stable when the Fresnel number is in the range of approximately 25 to 250. Generally, a shorter layer spacing can lead to a more compact and stable system. Therefore, it is sufficient to take the maximum value of the Fresnel number. This conclusion applies to optical neural networks based on diffraction effects.

The impact of phase modulation bit depth on network performance is also limited. Our network uses torch.complex64 type data during training, corresponding to a phase modulation precision of 24 bit. However, when we reduced the modulation precision to 3 bit during the testing phase, the network performance only dropped by about 1 percentage point. When reduced to 2 bit, the network performance dropped by 1–3 percentage points. Three bit means the network can only provide eight modulation phases: $-3\pi/4$, $-\pi/2$, $-\pi/4$, 0 , $\pi/4$, $\pi/2$, $3\pi/4$, π . Similarly, 2 bit can only provide four modulation phases: $-\pi/2$, 0 , $\pi/2$, π . This demonstrates the remarkable robustness of ICONN to the modulation precision of modulation devices.

3.3. Experiment Result and Analysis

We constructed an ICONN experimental platform to verify our model, with the experimental optical layout depicted in Figure 7. The right side represents the optical input module, where we utilized a white light LED (MNWHL4, THORLABS), a polarized beam splitter, and an amplitude SLM (HDSLM80RA Plus, UPOlabs) loaded with the target intensity distribution to simulate incoherent light input from a natural environment. The left side comprises the optical inference module, which consists of a filter (532 nm $\pm$ 20 nm), a 4f system, a beam splitter, a polarizer, a phase SLM (PLUTO-2.1, HOLOEYE), and a CCD. The polarizer is employed to meet the polarization sensitivity requirements of the inference device, while the filter is used to obtain quasi-monochromatic light to satisfy the modeling needs. The 4f system is designed to cut the image plane and map the mutual intensity spectrum plane of the target onto the phase SLM. The phase SLM serves as the modulation layer in the neural network, modulating the input incoherent light field by loading the phase distribution trained in the electrical domain.

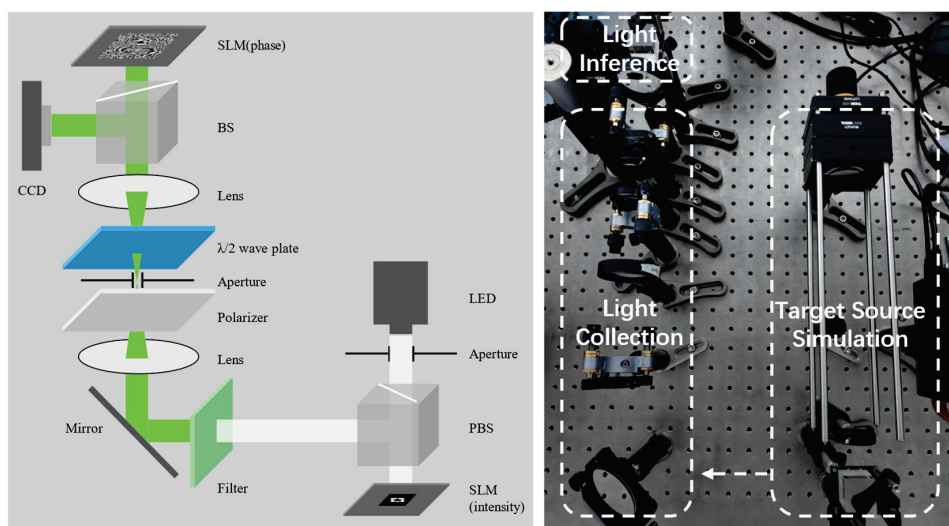


Figure 7. Experimental optical layout. The setup consists of three modules: (1) Target source simulation module (**right**) with an LED, aperture, polarizing beam splitter (PBS), and amplitude-type LCoS-SLM for simulating far-field incoherent targets; (2) Light collection module (**bottom left**) with lenses, aperture, polarizer, and $\lambda/2$ waveplate for collecting and adjusting the polarization of target light; (3) Inference module (**top left**) with a phase-type LCoS-SLM and detector for outputting inference results.

The input image had a pixel density of 80×80 , with a pixel size of $32 \mu\text{m}$. The modulation plane pixel density was set to 40×40 . The distance from the target (amplitude SLM) to the front focal plane of the 4f system was set to 0.308 m, which satisfied the Fresnel approximation. According to the Van Cittert–Zernike theorem, the sampling interval of the mutual intensity spectrum plane is $64 \mu\text{m}$. Since the mutual intensity spectrum is a function of the difference in distance, the length is halved when mapped to a two-dimensional space. Therefore, the sampling width is 40×40 , resulting in an inference device size of $256 \mu\text{m} \times 256 \mu\text{m}$. The magnification of the 4f system was set to 1. The spacing between the phase SLM and the detector was set to 0.1 m, and the phase modulation bit depth was consistent with the training results, set to 32 bit. A comparison of the experimental and simulation results is shown in Figure 8a,b. The single-layer network achieved the best accuracy of 87.57% on the ISDD dataset, demonstrating good agreement between our experimental and simulation results.

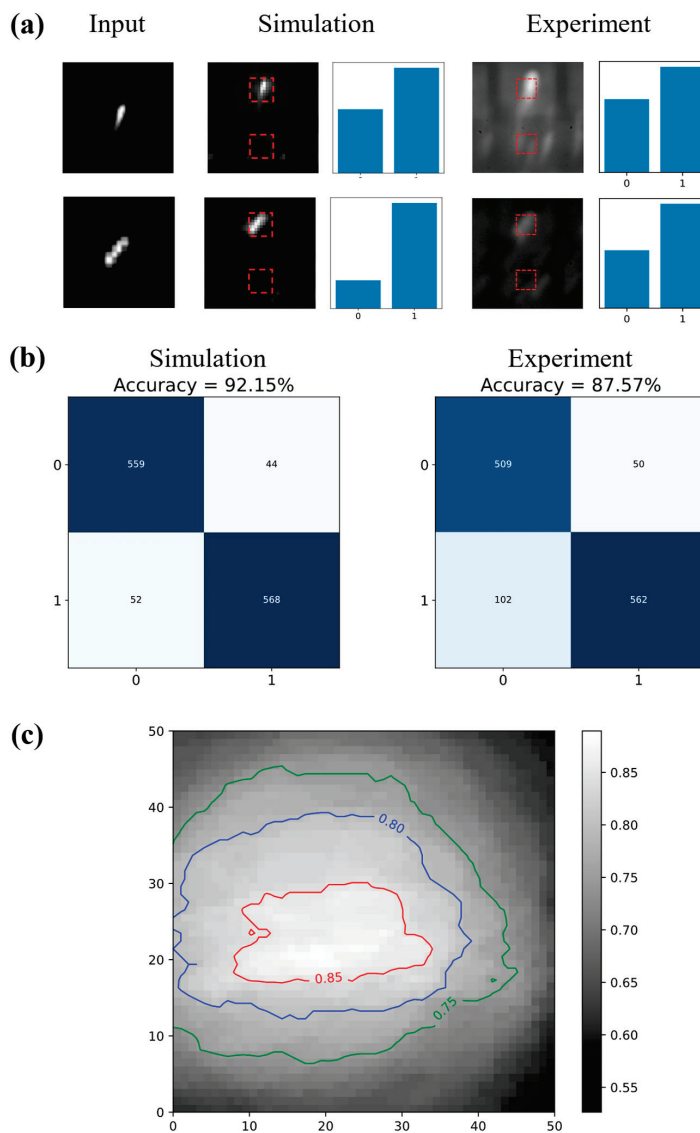


Figure 8. (a) Comparison of simulation and experimental outputs of ICONN on ISDD. The red box represents the label area for generating feature vectors. (b) Confusion matrix of simulation and experimental results of ICONN on ISDD. (c) Impact of measurement area offset on network performance. The modulation pixel size was $64 \mu\text{m}$, and the detector pixel size was $3.45 \mu\text{m}$. The horizontal and vertical coordinates represent the overall offset of the measurement label area, with a step size of 2 detector pixels.

The output of the optical neural network's detection surface also involves a pixel alignment issue. Especially when the number of network layers is small, the convergence effect of the output surface is not good, and a large amount of energy does not fall within the label area. At this time, the accuracy of the label area selection may also have an undeniable impact on the network performance. Here, we conducted tests based on the images obtained from the actual detector, performing a sliding test of the label area on the actual image to observe the impact of measurement area offset on network performance. The results are shown in Figure 8c. It can be seen from the figure that when the label area was selected at its best, the test accuracy was 87.57%. When the offset was about 70 μm , the network performance was still greater than 85%. This offset is already comparable to the modulation pixel size (64 μm), showing that ICONN has good robustness to measurement area offset.

4. Discussion

We have proposed an incoherent optical neural network based on mutual intensity propagation theory, which is more readily convergent. The network operates entirely in the optical domain and can directly perform computational inference on natural light, thus featuring ultra-high real-time capability and truly zero energy consumption. Except for the detector at the final output plane, all other components can operate passively (the SLM used in the experiment can be replaced by a pre-fabricated DOE component). Therefore, in the future, more energy can be devoted to reconfigurability, or a portion of electrical domain computation can be added to provide nonlinearity and enhance performance. Even in that case, our system would still possess significant advantages in terms of speed and energy efficiency.

Our network also achieved excellent classification performance in remote sensing data tests, which is expected to address the data explosion problem currently faced by the remote sensing field. For example, it can expand the field of view through array integration, serve as a feature extraction module to compress data, or directly perform target detection tasks.

We also conducted a series of simulations and experimental tests to investigate factors that may affect the performance of ICONN. The results show that within a certain range, the pixel density of the modulator has a negligible impact on network performance. In contrast, the Fresnel number, which is controlled by the layer spacing, has a significant impact on the network. Specifically, the network performance is optimal and stable when the Fresnel number is in the range of approximately 25 to 250. Additionally, the modulation depth of the modulator has a limited impact on network performance, and the performance stabilizes when the modulation bit depth is greater than 3 bit. Pixel alignment at the detection plane also affects network performance. When the deviation is less than one pixel width, the simulation and experimental error is approximately 2–3 percentage points.

In summary, the ICONN network proposed in this paper is capable of directly performing inference using natural light. It features all-optical passive operation, ease of convergence, zero energy consumption, and zero latency. Moreover, it has demonstrated excellent robustness and consistency in both simulations and experimental tests. This paves the way for low-cost and high-efficiency optical computing systems. The network holds great potential for applications in portable devices, edge computing, infrared remote sensing, and onboard data processing for satellites.

Author Contributions: Conceptualization, R.C. and S.S.; methodology, R.C.; software, R.C.; validation, R.C.; formal analysis, R.C., Y.M. and Z.W.; investigation, R.C. and Y.M.; resources, S.S.; data curation, R.C. and Y.M.; writing—original draft preparation, R.C.; writing—review and editing, Z.W., Y.M. and S.S.; visualization, R.C., Y.M. and Z.W.; supervision, S.S.; project administration, S.S.; funding acquisition, S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We would like to thank Yaqi Han et al. from the University of Electronic Science and Technology of China for collecting and producing the open-source Infrared Ship Detection Dataset (ISDD).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Amani, M.; Ghorbanian, A.; Ahmadi, S.A.; Kakooei, M.; Moghimi, A.; Mirmazloumi, S.M.; Moghaddam, S.H.A.; Mahdavi, S.; Ghahremanloo, M.; Parsian, S. Google earth engine cloud computing platform for remote sensing big data applications: A comprehensive review. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 5326–5350. [CrossRef]
2. Ye, B.; Tian, S.; Ge, J.; Sun, Y. Assessment of WorldView-3 Data for Lithological Mapping. *Remote Sens.* **2017**, *9*, 1132. [CrossRef]
3. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [CrossRef]
4. Chen, H.; Lou, S.; Wang, Q.; Huang, P.; Duan, H.; Hu, Y. Diffractive deep neural networks: Theories, optimization, and applications. *Appl. Phys. Rev.* **2024**, *11*, 021332. [CrossRef]
5. Fu, T.; Zhang, J.; Sun, R.; Huang, Y.; Xu, W.; Yang, S.; Zhu, Z.; Chen, H. Optical neural networks: Progress and challenges. *Light Sci. Appl.* **2024**, *13*, 263. [CrossRef]
6. Hu, J.; Mengü, D.; Tzarouchis, D.C.; Edwards, B.; Engheta, N.; Ozcan, A. Diffractive optical computing in free space. *Nat. Commun.* **2024**, *15*, 1525. [CrossRef]
7. Yu, H.; Huang, Z.; Lamon, S.; Wang, B.; Ding, H.; Lin, J.; Wang, Q.; Luan, H.; Gu, M.; Zhang, Q. All-optical image transportation through a multimode fibre using a miniaturized diffractive neural network on the distal facet. *Nat. Photonics* **2025**. [CrossRef]
8. Chang, J.; Sitzmann, V.; Dun, X.; Heidrich, W.; Wetzstein, G. Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification. *Sci. Rep.* **2018**, *8*, 12324. [CrossRef]
9. Lin, X.; Rivenson, Y.; Yardimci, N.T.; Veli, M.; Luo, Y.; Jarrahi, M.; Ozcan, A. All-optical machine learning using diffractive deep neural networks. *Science* **2018**, *361*, 1004–1008. [CrossRef]
10. Dou, H.; Deng, Y.; Yan, T.; Wu, H.; Lin, X.; Dai, Q. Residual D2NN: Training diffractive deep neural networks via learnable light shortcuts. *Opt. Lett.* **2020**, *45*, 2688–2691. [CrossRef]
11. Li, W.; Liu, X.; Zhang, W.; Ruan, N. The Application of Deep Learning in Space-Based Intelligent Optical Remote Sensing. *Spacecr. Recovery Remote Sens.* **2020**, *41*, 56–65.
12. Liu, C.; Ma, Q.; Luo, Z.J.; Hong, Q.R.; Xiao, Q.; Zhang, H.C.; Miao, L.; Yu, W.M.; Cheng, Q.; Li, L.; et al. A programmable diffractive deep neural network based on a digital-coding metasurface array. *Nat. Electron.* **2022**, *5*, 113–122. [CrossRef]
13. Wang, T.; Sohoni, M.M.; Wright, L.G.; Stein, M.M.; Ma, S.-Y.; Onodera, T.; Anderson, M.G.; McMahon, P.L. Image sensing with multilayer nonlinear optical neural networks. *Nat. Photonics* **2023**, *17*, 408–415. [CrossRef]
14. Bai, B.; Yang, X.; Gan, T.; Li, J.; Mengü, D.; Jarrahi, M.; Ozcan, A. Pyramid diffractive optical networks for unidirectional image magnification and demagnification. *Light Sci. Appl.* **2024**, *13*, 178. [CrossRef] [PubMed]
15. Huang, Z.; Shi, W.; Wu, S.; Wang, Y.; Yang, S.; Chen, H. Pre-sensor computing with compact multilayer optical neural network. *Sci. Adv.* **2024**, *10*, eado8516. [CrossRef]
16. Li, J.; Li, Y.; Gan, T.; Shen, C.-Y.; Jarrahi, M.; Ozcan, A. All-optical complex field imaging using diffractive processors. *Light Sci. Appl.* **2024**, *13*, 120. [CrossRef]
17. Li, Y.; Li, J.; Ozcan, A. Nonlinear encoding in diffractive information processing using linear optical materials. *Light Sci. Appl.* **2024**, *13*, 173. [CrossRef]
18. Xue, Z.; Zhou, T.; Xu, Z.; Yu, S.; Dai, Q.; Fang, L. Fully forward mode training for optical neural networks. *Nature* **2024**, *632*, 280–286. [CrossRef]
19. Hamerly, R.; Bernstein, L.; Sludds, A.; Soljacic, M.; Englund, D. Large-Scale Optical Neural Networks Based on Photoelectric Multiplication. *Phys. Rev. X* **2019**, *9*, 021032. [CrossRef]
20. Wang, Q.; Yu, H.; Huang, Z.; Gu, M.; Zhang, Q. Two-photon nanolithography of micrometer scale diffractive neural network with cubical diffraction neurons at the visible wavelength. *Chin. Opt. Lett.* **2024**, *22*, 102201. [CrossRef]
21. Plöschner, M.; Tyc, T.; Čížmár, T. Seeing through chaos in multimode fibres. *Nat. Photonics* **2015**, *9*, 529–535. [CrossRef]

22. Yildirim, M.; Dinc, N.U.; Oguz, I.; Psaltis, D.; Moser, C. Nonlinear processing with linear optics. *Nat. Photonics* **2024**, *18*, 1076–1082. [CrossRef]
23. Li, Y.; Luo, Y.; Mengu, D.; Bai, B.; Ozcan, A. Quantitative phase imaging (QPI) through random diffusers using a diffractive optical network. *arXiv* **2023**, arXiv:2301.07908.
24. Liu, Z.; Wang, L.; Meng, Y.; He, T.; He, S.; Yang, Y.; Wang, L.; Tian, J.; Li, D.; Yan, P. All-fiber high-speed image detection enabled by deep learning. *Nat. Commun.* **2022**, *13*, 1433. [CrossRef] [PubMed]
25. Lu, K.; Chen, Z.; Chen, H.; Zhou, W.; Zhang, Z.; Tsang, H.K.; Tong, Y. Empowering high-dimensional optical fiber communications with integrated photonic processors. *Nat. Commun.* **2024**, *15*, 3515. [CrossRef]
26. Xu, X.; Tan, M.; Corcoran, B.; Wu, J.; Boes, A.; Nguyen, T.G.; Chu, S.T.; Little, B.E.; Hicks, D.G.; Morandotti, R.; et al. 11 TOPS photonic convolutional accelerator for optical neural networks. *Nature* **2021**, *589*, 44–51. [CrossRef]
27. Wang, T.; Ma, S.-Y.; Wright, L.G.; Onodera, T.; Richard, B.C.; McMahon, P.L. An optical neural network using less than 1 photon per multiplication. *Nat. Commun.* **2022**, *13*, 123. [CrossRef]
28. Zheng, M.; Shi, L.; Zi, J. Optimize performance of a diffractive neural network by controlling the Fresnel number. *Photonics Res.* **2022**, *10*, 2667–2676. [CrossRef]
29. Zhu, H.H.; Zou, J.; Zhang, H.; Shi, Y.Z.; Luo, S.B.; Wang, N.; Cai, H.; Wan, L.X.; Wang, B.; Jiang, X.D.; et al. Space-efficient optical computing with an integrated chip diffractive neural network. *Nat. Commun.* **2022**, *13*, 123. [CrossRef]
30. Chen, Y.; Nazhamaiti, M.; Xu, H.; Meng, Y.; Zhou, T.; Li, G.; Fan, J.; Wei, Q.; Wu, J.; Qiao, F.; et al. All-analog photoelectronic chip for high-speed vision tasks. *Nature* **2023**, *623*, 48–57. [CrossRef]
31. Cheng, J.; Huang, C.; Zhang, J.; Wu, B.; Zhang, W.; Liu, X.; Zhang, J.; Tang, Y.; Zhou, H.; Zhang, Q. Multimodal deep learning using on-chip diffractive optics with in situ training capability. *Nat. Commun.* **2024**, *15*, 6189. [CrossRef]
32. Dai, T.; Ma, A.; Mao, J.; Ao, Y.; Jia, X.; Zheng, Y.; Zhai, C.; Yang, Y.; Li, Z.; Tang, B. A programmable topological photonic chip. *Nat. Mater.* **2024**, *23*, 928–936. [CrossRef] [PubMed]
33. Liu, L.; Liu, W.; Wang, F.; Peng, X.; Choi, D.-Y.; Cheng, H.; Cai, Y.; Chen, S. Ultra-robust informational metasurfaces based on spatial coherence structures engineering. *Light Sci. Appl.* **2024**, *13*, 131. [CrossRef] [PubMed]
34. Wang, H.; Hu, J.; Morandi, A.; Nardi, A.; Xia, F.; Li, X.; Savo, R.; Liu, Q.; Grange, R.; Gigan, S. Large-scale photonic computing with nonlinear disordered media. *Nat. Comput. Sci.* **2024**, *4*, 429–439. [CrossRef] [PubMed]
35. Zhan, Z.; Wang, H.; Liu, Q.; Fu, X. Photonic diffractive generators through sampling noises from scattering media. *Nat. Commun.* **2024**, *15*, 10643. [CrossRef]
36. Cui, K.; Rao, S.; Xu, S.; Huang, Y.; Cai, X.; Huang, Z.; Wang, Y.; Feng, X.; Liu, F.; Zhang, W. Spectral convolutional neural network chip for in-sensor edge computing of incoherent natural light. *Nat. Commun.* **2025**, *16*, 81. [CrossRef]
37. Fei, Y.; Sui, X.; Gu, G.; Chen, Q. Zero-power optical convolutional neural network using incoherent light. *Opt. Lasers Eng.* **2023**, *162*, 107410. [CrossRef]
38. Kleiner, M.; Michaeli, L.; Michaeli, T. Coherence Awareness in Diffractive Neural Networks. In Proceedings of the CLEO 2024, Charlotte, NC, USA, 5 May 2024; p. FW4Q.5.
39. Rahman, M.S.S.; Yang, X.; Li, J.; Bai, B.; Ozcan, A. Universal linear intensity transformations using spatially incoherent diffractive processors. *Light Sci. Appl.* **2023**, *12*, 195. [CrossRef]
40. Rahman, M.S.S.; Gan, T.; Deger, E.A.; Işıl, Ç.; Jarrahi, M.; Ozcan, A. Learning diffractive optical communication around arbitrary opaque occlusions. *Nat. Commun.* **2023**, *14*, 6830. [CrossRef]
41. Chen, R.; Ma, Y.; Zhang, C.; Xu, W.; Wang, Z.; Sun, S. All-optical perception based on partially coherent optical neural networks. *Opt. Express* **2025**, *33*, 1609–1624. [CrossRef]
42. van Cittert, P.H.J.P. Die wahrscheinliche Schwingungsverteilung in einer von einer Lichtquelle direkt oder mittels einer Linse beleuchteten Ebene. *Physica* **1934**, *1*, 201–210. [CrossRef]
43. Zernike, F. The concept of degree of coherence and its application to optical problems. *Physica* **1938**, *5*, 785–795. [CrossRef]
44. Wolf, E. New theory of partial coherence in the space-frequency domain. Part II: Steady-state fields and higher-order correlations. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **1986**, *3*, 76–85. [CrossRef]
45. Yan, T.; Wu, J.; Zhou, T.; Xie, H.; Xu, F.; Fan, J.; Fang, L.; Lin, X.; Dai, Q. Fourier-space Diffractive Deep Neural Network. *Phys. Rev. Lett.* **2019**, *123*, 023901. [CrossRef] [PubMed]
46. Liao, K.; Chen, Y.; Yu, Z.; Hu, X.; Wang, X.; Lu, C.; Lin, H.; Du, Q.; Hu, J.; Gong, Q. All-optical computing based on convolutional neural networks. *Opto-Electron. Adv.* **2021**, *4*, 200060. [CrossRef]
47. Wolf, E. Unified theory of coherence and polarization of random electromagnetic beams. *Phys. Lett. A* **2003**, *312*, 263–267. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

An Efficient Frequency Encoding Scheme for Optical Convolution Accelerator

Gongyu Xia ¹, Jiacheng Liu ¹, Qilin Hong ¹, Pingyu Zhu ², Ping Xu ^{2,*} and Zhihong Zhu ^{1,*}

¹ College of Advanced Interdisciplinary Studies & Hunan Provincial Key Laboratory of Novel Nano Optoelectronic Information Materials and Devices, National University of Defense Technology, Changsha 410073, China; zxc274261@163.com (G.X.); 18464268894@163.com (J.L.); qlhong_95@163.com (Q.H.)

² Institute for Quantum Information and State Key Laboratory of High-Performance Computing, College of Computer Science and Technology, National University of Defense Technology, Changsha 410073, China; pingyuzhu@yeah.net

* Correspondence: pingxu520@nju.edu.cn (P.X.); zzhwxcx@163.com (Z.Z.)

Abstract: In today's era where the demand for computational resources by large models is increasingly high, optical computing offers an alternative physical platform for computation. With its high parallelism and the maturation of integrated photonic technologies, optical computing is expected to further support the computational resources required by large models. For one-dimensional optical convolution accelerators, existing methods can fully utilize the working bandwidth of electro-optic modulators and the frequency resources of light sources. However, most convolution computations currently require the execution of two-dimensional matrix convolutions, and existing encoding schemes suffer a drop in terms of effective computations per second when performing two-dimensional matrix convolutions. In response to this, we propose a new encoding scheme that can fully utilize the computational resources of optical convolution accelerators. For convolution operations with a kernel size of $M \times M$, compared to existing encoding schemes, it can achieve an M -fold increase in effective computation rate. This implies that under the constraints of essentially the same hardware physical performance, adopting our newly proposed encoding scheme can achieve a significant improvement in computational performance. We construct an optical convolution acceleration system to demonstrate the principles of the new encoding.

Keywords: convolution neural network; optical computing; matrix convolution

1. Introduction

The concept of the artificial neural network (ANN) originates from the study of the mechanisms underlying the biological neural processing of signals. With the increasing power of electronic integrated circuits, artificial neural networks and large language models have demonstrated remarkable capabilities across an expanding array of fields, including image recognition, natural language processing, and autonomous driving. However, the evolution of traditional electronic computing chips has encountered two significant limitations: the power consumption wall and the computation rate wall. These constraints have impeded further enhancements in computational performance, urgently necessitating the development of new physical computing platforms to surmount these obstacles.

In response to these challenges, a variety of processing chips aimed at intelligent computing have emerged, such as IBM TrueNorth [1] and Google TPU [2]. These chips have been designed to accelerate specific types of computations, particularly those involved

in deep learning and neural network processing. However, they still operate within the constraints of electronic computing, which limits their scalability and energy efficiency. For instance, electronic chips face challenges in handling the massive quantity of data and complex computations required for training and inference in large-scale neural networks, leading to increased power consumption and longer processing times.

Optical computing, leveraging the high-speed transmission and high parallelism characteristics of light, stands out as a promising candidate for a high-performance computation platform. The inherent advantages of optical computing include its ability to perform computations in parallel, its low power consumption, and its potential for high-speed data processing. These features make it particularly suitable for handling the computationally intensive tasks required by large-scale artificial neural networks and other advanced AI applications. For instance, optical computing can process multiple data streams simultaneously, reducing the need for sequential processing and enabling faster computation of complex operations such as matrix multiplications and convolutions.

Currently, research is being conducted into several areas of optical computing, including coherent optical computing [3], reservoir computing [4–6], and diffractive network computing [7–9]. Coherent optical computing utilizes the phase and amplitude of light waves to perform computations, offering high precision and efficiency. Reservoir computing, on the other hand, uses the natural dynamics of optical systems to process information, providing a novel approach to machine learning and signal processing. Diffractive network computing employs diffractive optical elements to create complex neural networks, enabling efficient and compact optical processing systems.

Notable advancements have been made in multi-wavelength computation, as evidenced by the significant progress achieved in this domain [10–13]. Multi-wavelength computation allows for the simultaneous processing of multiple wavelengths of light, increasing the computational capacity and efficiency of optical systems. This is particularly useful in applications such as optical communication and data processing, where large quantities of data need to be transmitted and processed quickly.

Among these advancements, the employment of high-speed electro-optical modulators facilitates the multiplexing of frequency- and time-domain signals, thereby enabling optically accelerated matrix convolution operations. Electro-optical modulators are crucial components in optical computing systems, as they convert electrical signals into optical signals, allowing for the manipulation and processing of light. The latest work utilizes a Bragg grating array to optimize electro-optic modulators for the realization of convolution accelerators [14]. Bragg grating arrays can precisely control the phase and amplitude of light, enhancing the performance of electro-optical modulators and improving the overall efficiency of optical convolution accelerators.

For one-dimensional convolution operations, existing computational acceleration schemes [13] can fully utilize the frequency resources of optical signals and the bandwidth of electro-optic modulators for the optical acceleration of computations. However, considering practical applications, most convolution operations in artificial intelligence computations are two-dimensional matrix convolutions. Existing encoding schemes may compromise actual computational efficiency when dealing with two-dimensional matrix convolutions. When the size of a single convolution kernel is $M \times M$, the number of effective multiplication and addition operations per second is only $\frac{1}{M}$ of that for one-dimensional vector convolutions. This limitation significantly impacts the performance of optical convolution accelerators in handling complex AI tasks that involve large-scale two-dimensional data, such as image and video processing.

To address these limitations, we propose a novel matrix–vector mapping encoding scheme. This scheme can fully utilize the frequency resources of optical signals and the

bandwidth of electro-optic modulators, achieving the same number of effective multiplication and addition operations per second as one-dimensional vector convolutions. Compared to existing two-dimensional matrix encoding schemes, it offers an M -fold increase in computational operations. This enhancement not only improves the efficiency of optical convolution accelerators but also paves the way for more complex and larger-scale neural network applications to be efficiently processed using optical computing platforms. By overcoming the challenges associated with two-dimensional matrix convolutions, our proposed scheme enables optical computing to better meet the demands of modern AI applications, offering a significant step forward in the development of high-performance computing systems.

Moreover, the proposed encoding scheme has the potential to significantly impact various domains that rely heavily on convolutional neural networks (CNNs). In the field of medical imaging, for example, CNNs are extensively used for tasks such as disease diagnosis, anomaly detection, and image segmentation. By accelerating the convolution operations in CNNs, our encoding scheme can enable faster and more accurate analysis of medical images, leading to improved diagnostic outcomes and more efficient healthcare workflows. In the realm of autonomous vehicles, CNNs play a crucial role in perception tasks, such as object detection, scene understanding, and navigation. The enhanced computational efficiency provided by our scheme can contribute to real-time processing of sensor data, enabling autonomous vehicles to make quicker and more reliable decisions, thereby enhancing their overall performance and safety.

Furthermore, the scalability of our encoding scheme makes it suitable for integration with emerging technologies such as quantum computing and neuromorphic computing. Quantum computing holds the promise of solving complex problems that are intractable for classical computers, while neuromorphic computing aims to emulate the human brain's neural structure and functionality. By combining the strengths of optical computing with these advanced technologies, we can potentially unlock new levels of computational power and efficiency, paving the way for breakthroughs in fields such as cryptography, optimization, and artificial intelligence research.

The remainder of this paper is organized as follows. In Section 2, we present the details of the more efficient frequency encoding scheme for optical convolution accelerators. Section 3 describes the experimental setup and results, demonstrating the effectiveness of our proposed scheme. Finally, Section 4 discusses the implications of our findings and suggests directions for future research.

2. More Efficient Frequency Encoding Scheme for Optical Convolution Accelerator

The optical vector convolution accelerator leverages the high bandwidth characteristics of high-speed electro-optic modulators to perform simultaneous multiplication and addition operations on optical signals of multiple frequencies [15–17]. The input vector $\mathbf{X} \in \mathbb{R}^{1 \times n}$ is encoded in the form of a time series within a one-dimensional voltage signal. When the electro-optic intensity modulator is biased at $\pi/4$, its response to the input optical signal can be approximated as the product of the radio-frequency electrical signal and the input optical power. The one-dimensional convolution kernel $\mathbf{W} \in \mathbb{R}^{1 \times m}$, which is the reversed, $\mathbf{W}^r$, where $W_i^r = W_{m-i+1}$, is encoded into the frequency domain of the optical signal. At timestamp i , the vector $\mathbf{W}^r$ composed of optical signals of different frequencies will undergo the same multiplicative factor X_i . The total power of the current optical signal

is given by the sum $\sum_{j=1}^m X_i W_j^r$. The dispersion of the optical fiber is utilized to interleave the frequencies of the optical signal. The following results are obtained.

$$\sum_{j=1}^m X_i W_j^r \xrightarrow[\text{Interleaving}]{\text{Frequency}} \sum_{j=1}^m X_{i+j-1} W_j^r = \sum_{j=1}^m X_{i+j-1} W_{m-j+1} = Y_i \quad (1)$$

Considering the sampling rate of the electro-optic modulator is Ω , which means the time interval between adjacent elements of vector X is $\frac{1}{\Omega}$, and the number of available frequency points of the optical signal is M , then, within a unit time, the vector convolution accelerator can perform $2 \times \Omega \times M$ multiplication and addition operations. With a sampling rate of 172 GHz [18] and input of an optical signal with 1400 frequency components [19], the optical vector convolution accelerator can perform multiplication and addition operations at a rate of $2 \times 172 \times 1400 = 481$ T per second.

Convolution neural networks predominantly require the execution of convolution operations on two-dimensional matrix inputs. For one-dimensional optical vector convolution accelerators, performing two-dimensional matrix convolution necessitates an encoding scheme that transforms the convolution operations between two-dimensional matrices into computations between one-dimensional vectors.

Taking a 3×3 convolution kernel as an example, let the input matrix be $X \in \mathbb{R}^{N \times N}$, which is mapped to $X' \in \mathbb{R}^{1 \times N^2}$. The mapping is denoted as $f : \mathbb{R}^{N \times N} \rightarrow \mathbb{R}^{1 \times N^2}$. An existing mapping scheme, proposed by Xu et al. [13] in 2021, is denoted as $f_0 : \mathbb{R}^{N \times N} \rightarrow \mathbb{R}^{1 \times N^2}$.

$$f_0(X) = X', X'_k = X_{i,j}, k = ((i-1) \bmod 3) + 1 + (j-1) * 3 + \lfloor (i-1)/3 \rfloor * 3N \quad (2)$$

As shown in Figure 1a, the first 3 rows of the input two-dimensional matrix are taken and concatenated into a one-dimensional vector by columns, and then this operation is repeated by taking 3 rows at a time to complete the mapping from a two-dimensional matrix to a one-dimensional vector. After completing the dimensionality transformation, the one-dimensional vector is placed into the optical vector convolution accelerator. As shown in Figure 1b, after frequency interleaving, every three timestamps, one is the value required for the actual convolution computation. This means that compared to one-dimensional vector convolution operations, the effective number of computations for two-dimensional vector convolution after the encoding scheme f_0 is only 1/3 of that of one-dimensional vectors.

We propose another encoding scheme $f_1 : \mathbb{R}^{N \times N} \rightarrow \mathbb{R}^{1 \times N^2}$.

$$f_1(X) = X', X'_k = X_{i,j}, k = i + (j-1)N \quad (3)$$

The matrix at the bottom of Figure 1a concatenates the entire input two-dimensional matrix into a one-dimensional vector by rows. In the computation, as shown in Figure 1c, we take the power sum of every three values in the frequency domain $W_{i,1} \sim W_{i,3}$, and then sum $\sum_j Y_{k,j}$. It is not difficult to find that compared to the original one-dimensional vector convolution operation, after adopting the encoding scheme f_1 , the number of effective computations is essentially the same as the original, and compared to f_0 , for a 3×3 convolution kernel, there is a three-fold increase. As shown in Table 1, given a sampling rate of 172 GHz and an input optical signal with 1400 frequency components, the actual number of effective multiply-accumulate operations per second required for the two-dimensional matrix convolution of a 9×9 convolution kernel is increased from $2 \times 172 \times 1400/9 = 53.5$ T to $2 \times 172 \times 1400 = 481.6$ T.

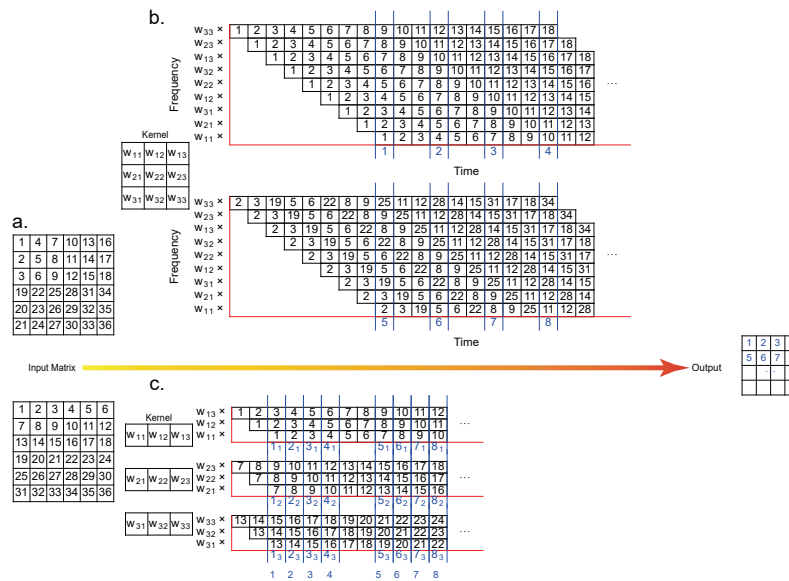


Figure 1. (a) Image signals converted to one-dimensional signal coding scheme: original scheme (top), new scheme (bottom). (b) Calculation process of optical convolution for existing scheme [13]. (c) New optical convolution scheme proposed in this work.

Table 1. The number of multiply–add operations performed per second of original scheme and our scheme for different kernel sizes.

Kernel Size	3×3	5×5	7×7	9×9
Original scheme	160.53 T	96.32 T	68.8 T	53.51 T
Our scheme	481.6 T	481.6 T	481.6 T	481.6 T

We constructed an optical convolutional neural network computing system to demonstrate the actual working principle of the new encoding scheme. Figure 2a shows the structure of the convolutional network, which consists of five layers. From front to back, they are the input layer of size 28×28 , a convolutional layer of size $8 \times 26 \times 26$ with 8 convolution kernels each of size 3×3 , a pooling layer of size $8 \times 13 \times 13$, followed by a flattening operation, and finally a fully connected layer with 10 nodes and the activation function SoftMax for recognizing the handwritten digit dataset of the Mixed National Institute of Standards and Technology database (MNIST), which comprises a total of 50,000 training data entries and 10,000 testing data entries. Out of the 50,000 training data entries, 10,000 were separated as evaluation data during the training process. The overall network was trained by a computer, with the convolutional calculations performed by the optical system shown in Figure 2e, and the subsequent fully connected layers and activation functions completed by an electronic computer.

Eight convolution kernels of size 3×3 were transformed into one-dimensional vectors and encoded onto the frequency domain of the input optical signal. In Figure 2b, for a single convolution kernel, row-by-row concatenation was performed, and the encoding operation was completed by a spectral shaper, ensuring that each element corresponded to the intensity of an optical signal at a specific frequency, with adjacent frequency intervals of $\Delta\omega$.

For the input matrix, we took the handwritten digit “2” as an example. In Figure 2c, according to the encoding scheme f_1 (3), the two-dimensional input matrix was concatenated row-by-row to obtain a one-dimensional vector, which was then encoded into the radio-frequency voltage input of the electro-optic intensity modulator, with adjacent element

time intervals of $1/\Omega$, where Ω is the sampling rate of the arbitrary waveform generator. The bias of the electro-optic intensity modulator was set to $\pi/4$.

$$T(U) = \sin^2\left(\frac{\pi}{2} \frac{U}{U_\pi} + \phi\right) \approx \frac{1}{2} + \frac{\pi}{2} \frac{U}{U_\pi} \quad (\phi = \frac{\pi}{4}, U \ll U_\pi) \quad (4)$$

The transmittance T of the electro-optic intensity modulator as a function of the external field voltage U , when the bias ϕ was set to $\frac{\pi}{4}$, could be approximated as a linear response, and subtracting the DC term completed the multiplication of the input optical intensity X_i and the modulator radio-frequency voltage W_j .

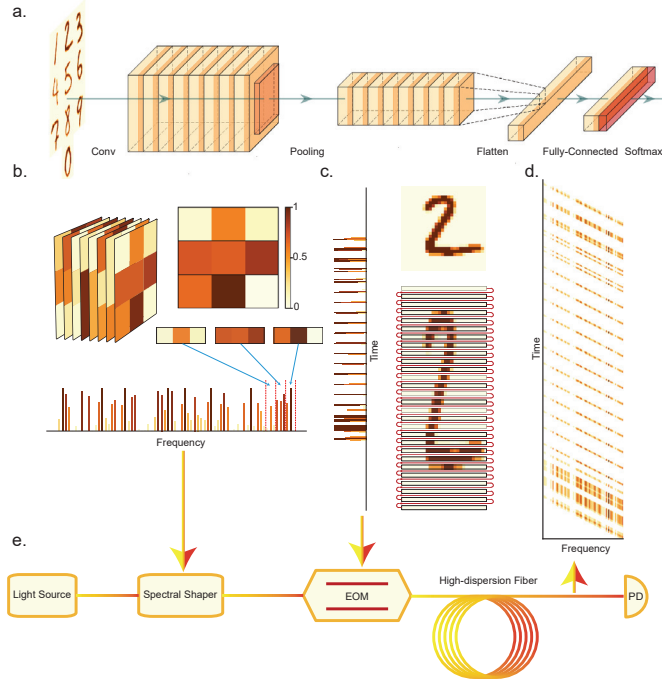


Figure 2. (a) Convolutional network model framework. (b) Convolutional kernel decomposition process. (c) Image decomposition process. (d) Calculation results after dispersive fiber. (e) Experimental framework diagram.

To achieve frequency interleaving, the optical convolution system also required a certain dispersion module to create a time-domain displacement of optical signals at different frequencies. When the sampling rate of the arbitrary waveform generator is Ω , and the frequency interval between adjacent elements of the convolution kernel is $\Delta\omega$, in a system with a dispersion coefficient of D (ps/(km · nm)), such as an optical fiber, the required total length is $L = \frac{1}{\Omega} \frac{\omega}{D\lambda\Delta\omega}$. After passing through the dispersion system, the result shown in Figure 2d was obtained, and the required computational result was obtained using the calculation method of Figure 1c and a power meter. For the multi-frequency computation process, taking a 3×3 convolution kernel as an example, after the optical signal passed through the dispersion fiber, a wavelength division multiplexer (WDM) was used to group every three adjacent frequencies together, and the output was directed to an optical power meter to measure the optical power sum corresponding to the current step.

For the experimental system we constructed to demonstrate the new encoding scheme of the optical convolutional neural network, we could calculate the actual number of effective operations per second and compare it with the existing encoding scheme. As shown in Figure 3a, we built a light source with 9 frequency teeth, and the electro-optic modulator operated at a sampling rate of 1.429 GHz. As shown in Figure 3d, the current system response had a standard deviation of 0.03987, corresponding to a signal-to-noise ratio of 27.98 dB. We conducted simulation calculations using this numerical precision and

obtained the performance of the current demonstration system for the MNIST handwritten digit recognition task, as depicted in Figure 3f. For one-dimensional vector convolution, the actual number of computations per second was $2 \times 9 \times 1.429 = 25.7$ G multiplication and addition operations. After using the encoding scheme f_1 , the actual number of computations per second for two-dimensional convolution operations was the same as for one-dimensional vector convolution. For the encoding scheme f_0 , the actual number of computations per second was reduced to $2 \times 9 \times 1.429/3 = 8.57$ G multiplication and addition operations. It can be seen that when the convolution kernel was a 3×3 matrix, using the new encoding scheme f_1 for computing matrix convolution had three times the improvement compared to the existing scheme. For larger convolution kernels, such as 5×5 , 7×7 , $N \times N$, the improvement would increase with the increase in N , and it would have N times the improvement compared to the existing scheme.

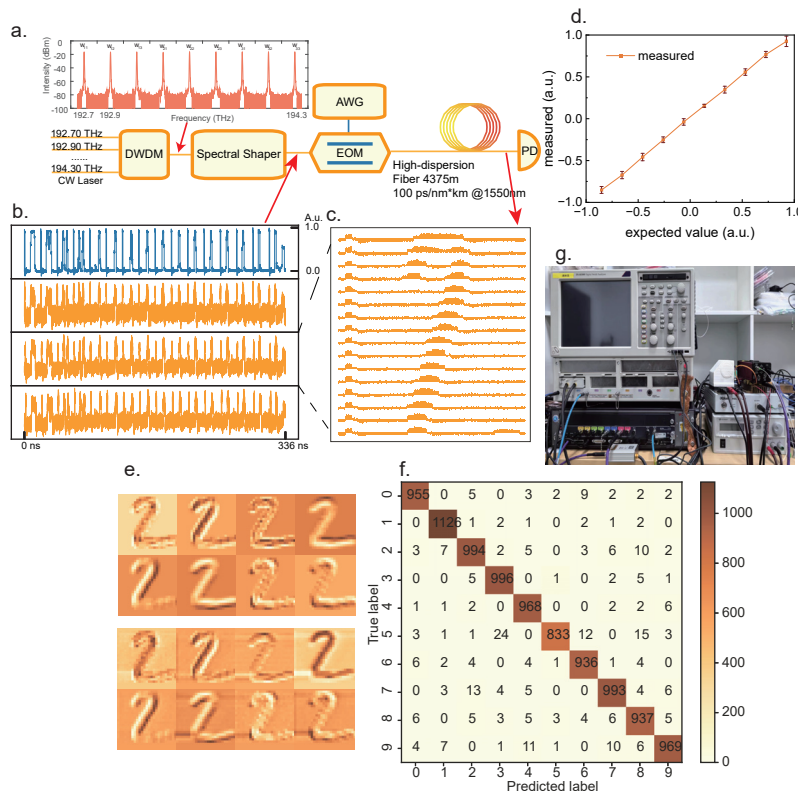


Figure 3. (a) System experimental diagram, Dense Wavelength Division Multiplexer (DWDM), Arbitrary Waveform Generators (AWG), electro-optic modulator (EOM), Photodetector (PD). (b) Time-domain electrical signals loaded onto the modulator and finally detected by the power meter. (c) Results of single, small-kernel calculations. (d) Error analysis of the system. (e) Features obtained by computer calculations of the eight convolution kernels (top) versus features obtained by convolution of the optical system (bottom). (f) Confusion table. (g) Experimental setup.

3. Experimental Setup

The experimental procedure is meticulously illustrated in Figure 3a. The light source employed in this study was a continuous-wave laser, which was characterized by a series of frequency teeth ranging from 192.7 to 194.3 THz, with intervals of 0.2 THz, thereby comprising a total of nine distinct frequency teeth. Prior to the introduction of the light through the spectrum shaper, the initial spectral distribution is depicted in Figure 3a.

The spectrum shaper, specifically the Finisar 16000A model, played a crucial role in encoding the convolution kernel information. That information, initially in a two-dimensional format, was transformed into a one-dimensional vector and subsequently

embedded into the spectral profile of the light source. This encoding process was essential for the subsequent optical signal processing.

The input data for the experiment were derived from the MNIST dataset, which consists of handwritten digit images. Each image is composed of a 28×28 pixel grid, with grayscale values ranging from 0 to 255. To facilitate the encoding process, these grayscale values were normalized to the interval $[0,1]$. The normalized values were then encoded onto the radio-frequency input of the lithium niobate electro-optic intensity modulator (iXblue, MX-LN-20, Saint-germain-en-laye, Ile-de-France, France) using an arbitrary waveform generator (Keysight, M8196A). The arbitrary waveform generator operated at a sampling rate of 1.429 GHz, as shown in Figure 3b. Under the limitation of the current electro-optic modulator's 3 dB bandwidth, we found that at the current modulation frequency of 1.429 GHz, the system could ensure a high signal-to-noise ratio of 27.98 dB. Upon increasing the modulation frequency, it became challenging to maintain a high level of signal-to-noise ratio; hence, we chose the current modulation frequency as a compromise.

Due to the anisotropy of lithium niobate crystals, their electro-optic modulators are polarization-dependent. Before the optical signal entered the electro-optic modulator, it was necessary to use a polarization controller to rotate the polarization of the incident light to the appropriate direction. The electro-optic modulator, which had an approximate half-wave voltage of 5 V, was biased at $\pi/4$ to achieve a more accurate linear approximation of the input signal. The output voltage of the arbitrary waveform generator was modulated to 100 mV to ensure that the response of the optical intensity to the voltage signal could be approximated linearly.

Following the modulation process, the optical signal passed through an optical fiber with a dispersion coefficient of $100 \text{ ps}/(\text{km} \cdot \text{nm})$ and a length of 4375 m. The dispersion introduced by the fiber was critical for achieving the desired frequency interleaving, which was essential for the accurate computation of the convolution operations. After traversing the fiber, the optical signals were measured using an optical power meter (Thorlabs, RXM40AF, Newton, NJ, USA), as indicated by the orange line in Figure 3b. The measured signals were then converted into a two-dimensional matrix to facilitate the observation of the computational results at that stage, as shown in Figure 3c.

For the 83×3 convolution kernels used in this experiment, their positive and negative elements were separated to form 16 convolution kernel matrices, each with uniform sign. The optical system computed the results, which are displayed at the bottom of Figure 3d. These results were compared with the convolution results obtained directly from an electronic computer, as shown at the top of Figure 3d. The current computational framework, as depicted in Figure 2a, enabled the system to achieve an approximate recognition accuracy of 97%, as illustrated in Figure 3e.

4. Discussion

It is a misconception to equate the product of frequency resources and modulation bandwidth with the actual computational power in existing optical convolution systems. As illustrated in Figure 1, for a single convolution kernel of size 3×3 , the actual computational power is merely one-third of the aforementioned product. This efficiency is likely to decrease with an increasing convolution kernel size. In our computational framework, the aforementioned product can be exploited to a high degree of efficiency, albeit at the cost of additional summing operations at the backend. As a trade-off for improving the efficiency of matrix convolution, we require additional Dense Wavelength Division Multiplexing (DWDM) to perform the block summation depicted in Figure 1c. Fortunately, the currently mature DWDM technology allows us not to worry about this additional cost.

Currently, in the case of multiple physical computing systems with limited frequency resources, these can be computed collectively within our framework and subsequently integrated into a larger system with enhanced computational capabilities. In the demonstration experiments, only multiple independent lasers were employed to illustrate that the subsystems can be integrated into a larger system. In actual deployment, thin-film lithium niobate on-chip optical frequency combs coupled with on-chip electro-optical intensity modulators can be utilized to achieve higher integration and more frequency resources, which is anticipated to facilitate the deployment of optical accelerated computing into practical applications in a more expedient manner.

Furthermore, our research reveals the potential of optical computing in handling large-scale datasets. For instance, in image recognition tasks, traditional electronic computing methods require substantial computational resources and time, whereas our optical convolution accelerator can significantly enhance computational speed, thus offering a distinct advantage in real-time image processing and video analysis applications. This advantage is not only reflected in speed but also in energy efficiency. The energy consumption of optical computing systems is far lower than that of traditional electronic computing systems, which is crucial for applications that require long-term operation and processing of large quantities of data.

However, we also recognize that there are challenges in practical applications. First, the stability and reliability of optical computing systems need to be further improved. For example, optical signals may be affected by environmental factors during transmission, such as temperature changes and vibrations, leading to errors in computational results. Therefore, it is necessary to develop more advanced optical signal processing technologies and error correction algorithms to ensure the stability and reliability of the systems. Second, the integration and scalability of optical computing systems also need to be optimized. Although current technologies for on-chip optical frequency combs and electro-optical modulators have made some progress, there are still technical challenges in achieving large-scale integration and diverse functionalities. Future research can explore new materials and manufacturing processes to improve the integration and scalability of optical computing systems.

Moreover, the application scope of optical computing systems needs to be expanded. Currently, optical computing is mainly applied to specific types of computational tasks, such as convolutional neural networks, and the potential of optical computing in other types of tasks, such as graph computing and optimization problems, has not been fully explored. Future research can investigate the possibilities of optical computing in a broader range of application fields, such as bioinformatics, financial analysis, and weather forecasting, developing optical computing algorithms and system architectures suitable for these areas.

In summary, our research provides new insights and methods for the development of the field of optical computing. By proposing a novel frequency encoding scheme and optimizing system architecture, we have not only improved the computational performance of optical convolution accelerators but also laid the foundation for the deployment of optical computing in practical applications. In the future, with the continuous advancement and innovation of related technologies, optical computing is expected to play a significant role in more fields, providing more efficient and energy-saving solutions for solving complex computational problems.

Author Contributions: Conceptualization, G.X.; methodology, G.X. and J.L.; investigation, software, G.X. and P.Z.; writing—original draft preparation, G.X.; writing—review and editing, G.X. and Q.H.; supervision, Z.Z.; project administration, G.X.; funding acquisition, P.X. All authors have read and agreed to the published version of the manuscript.

Funding: National Key R&D Program of China (2022YFF0712800, 2019YFA0308700).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Esser, S.K.; Merolla, P.A.; Arthur, J.V.; Cassidy, A.S.; Appuswamy, R.; Andreopoulos, A.; Berg, D.J.; McKinstry, J.L.; Melano, T.; Barch, D.R.; et al. Convolutional networks for fast, energy-efficient neuromorphic computing. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 11441–11446. [CrossRef]
- Graves, A.; Wayne, G.; Reynolds, M.; Harley, T.; Danihelka, I.; Grabska-Barwińska, A.; Colmenarejo, S.G.; Grefenstette, E.; Ramalho, T.; Agapiou, J.; et al. Hybrid computing using a neural network with dynamic external memory. *Nature* **2016**, *538*, 471–476. [CrossRef]
- Shen, Y.; Harris, N.C.; Skirlo, S.; Prabhu, M.; Baehr-Jones, T.; Hochberg, M.; Sun, X.; Zhao, S.; Larochelle, H.; Englund, D.; et al. Deep learning with coherent nanophotonic circuits. *Nat. Photonics* **2017**, *11*, 441–446. [CrossRef]
- Larger, L.; Soriano, M.C.; Brunner, D.; Appeltant, L.; Gutierrez, J.M.; Pesquera, L.; Mirasso, C.R.; Fischer, I. Photonic information processing beyond Turing: An optoelectronic implementation of reservoir computing. *Opt. Express* **2012**, *20*, 3241. [CrossRef] [PubMed]
- Vandoorne, K.; Mechet, P.; Van Vaerenbergh, T.; Fiers, M.; Morthier, G.; Verstraeten, D.; Schrauwen, B.; Dambre, J.; Bienstman, P. Experimental demonstration of reservoir computing on a silicon photonics chip. *Nat. Commun.* **2014**, *5*, 3541. [CrossRef]
- Van der Sande, G.; Brunner, D.; Soriano, M.C. Advances in photonic reservoir computing. *Nanophotonics* **2017**, *6*, 561–576. [CrossRef]
- Lin, X.; Rivenson, Y.; Yardimci, N.T.; Veli, M.; Luo, Y.; Jarrahi, M.; Ozcan, A. All-optical machine learning using diffractive deep neural networks. *Science* **2018**, *361*, 1004–1008. [CrossRef] [PubMed]
- Mengu, D.; Zhao, Y.; Yardimci, N.T.; Rivenson, Y.; Jarrahi, M.; Ozcan, A. Misalignment resilient diffractive optical networks. *Nanophotonics* **2020**, *9*, 4207–4219. [CrossRef]
- Fu, T.; Zang, Y.; Huang, Y.; Du, Z.; Huang, H.; Hu, C.; Chen, M.; Yang, S.; Chen, H. Photonic machine learning with on-chip diffractive optics. *Nat. Commun.* **2023**, *14*, 70. [CrossRef] [PubMed]
- Tait, A.N.; Nahmias, M.A.; Shastri, B.J.; Prucnal, P.R. Broadcast and Weight: An Integrated Network For Scalable Photonic Spike Processing. *J. Light. Technol.* **2014**, *32*, 4029–4041. [CrossRef]
- Tait, A.N.; de Lima, T.F.; Zhou, E.; Wu, A.X.; Nahmias, M.A.; Shastri, B.J.; Prucnal, P.R. Neuromorphic photonic networks using silicon photonic weight banks. *Sci. Rep.* **2017**, *7*, 7430. [CrossRef] [PubMed]
- Feldmann, J.; Youngblood, N.; Karpov, M.; Gehring, H.; Li, X.; Stappers, M.; Le Gallo, M.; Fu, X.; Lukashchuk, A.; Raja, A.S.; et al. Parallel convolutional processing using an integrated photonic tensor core. *Nature* **2021**, *589*, 52–58. [CrossRef] [PubMed]
- Xu, X.; Tan, M.; Corcoran, B.; Wu, J.; Boes, A.; Nguyen, T.G.; Chu, S.T.; Little, B.E.; Hicks, D.G.; Morandotti, R.; et al. 11 TOPS photonic convolutional accelerator for optical neural networks. *Nature* **2021**, *589*, 44–51. [CrossRef] [PubMed]
- Cai, K.; Chen, L.; Zhang, Y.; Wang, J.; Lin, W.; Duan, S.; Liu, B. On-chip photoelectric hybrid convolutional accelerator based on Bragg grating array. *Results Phys.* **2024**, *65*, 107968. [CrossRef]
- Rizzo, A.; Novick, A.; Gopal, V.; Kim, B.Y.; Ji, X.; Daudlin, S.; Okawachi, Y.; Cheng, Q.; Lipson, M.; Gaeta, A.L.; et al. Massively scalable Kerr comb-driven silicon photonic link. *Nat. Photonics* **2023**, *17*, 781–790. [CrossRef]
- Fu, T.; Zhang, J.; Sun, R.; Huang, Y.; Xu, W.; Yang, S.; Zhu, Z.; Chen, H. Optical neural networks: Progress and challenges. *Light. Sci. Appl.* **2024**, *13*, 263. [CrossRef]
- McMahon, P.L. The physics of optical computing. *Nat. Rev. Phys.* **2023**, *5*, 717–734. [CrossRef]
- Mao, J.; Uemura, F.; Yazdani, S.A.; Yin, Y.; Sato, H.; Lu, G.W.; Yokoyama, S. Ultra-fast perovskite electro-optic modulator and multi-band transmission up to 300 Gbit/s. *Commun. Mater.* **2024**, *5*, 114. [CrossRef]
- Cheng, R.; Yu, M.; Shams-Ansari, A.; Hu, Y.; Reimer, C.; Zhang, M.; Lončar, M. Frequency comb generation via synchronous pumped $\chi(3)$ resonator on thin-film lithium niobate. *Nat. Commun.* **2024**, *15*, 3921. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Hybrid Artificial-Intelligence-Based System for Unmanned Aerial Vehicle Detection, Localization, and Tracking Using Software-Defined Radio and Computer Vision Techniques

Pablo López-Muñoz ¹, Luis Gimeno San Frutos ¹, Christian Abarca ², Francisco José Alegre ², Jose Luis Calle ² and Jose F. Monserrat ^{1,*}

¹ Universitat Politècnica de València-iTEAM, 46022 Valencia, Spain; pabluo@upv.edu.es (P.L.-M.); luigisa1@teleco.upv.es (L.G.S.F.)

² Grupo Simetria, 12100 Castellon, Spain; cabarca@simetriagrupo.com (C.A.); fjalegre@simetriagrupo.com (F.J.A.); jlcalles@simetriagrupo.com (J.L.C.)

* Correspondence: jomondel@upv.es

Abstract: The proliferation of drones in civilian environments has raised growing concerns about their misuse, highlighting the need to develop efficient detection systems to protect public and private spaces. This article presents a hybrid approach for UAV detection that combines two artificial-intelligence-based methods to improve system accuracy. The first method uses a software-defined radio (SDR) to analyze the radio spectrum, employing autoencoders to detect drone control signals and identify the presence of these devices. The second method is a computer vision module consisting of fixed cameras and a PTZ camera, which uses the YOLOv10 object detection algorithm to identify UAVs in real time from video sequences. Additionally, this module integrates a localization and tracking algorithm, allowing the tracking of the intruding UAV's position. Experimental results demonstrate high detection accuracy, a significant reduction in false positives for both methods, and remarkable effectiveness in UAV localization and tracking with the PTZ camera. These findings position the proposed system as a promising solution for security applications.

Keywords: UAV; SDR; autoencoders; YOLOv10; tracking; localization

1. Introduction

In recent years, due to technological development and cost reduction, unmanned aerial vehicles (UAVs) have experienced exponential growth, making them devices that are now accessible for the personal use of any citizen. The increasing popularity of UAVs can also be attributed to their great versatility, as they offer a wide range of applications, from recreational and informal uses to professional use in tasks such as search and rescue operations, fire control, and other natural disaster operations, as described by Shakh-treh et al., who analyzed the growing use of UAVs in various civilian applications and the associated challenges [1]. Also, there are more applications. For instance, Khosravi and Pishro-Nik designed UAVs for Package Delivery and Network Coverage tasks [2]. Others focused on the monitoring and assessment of agriculture, such as Ju and So, who developed a system using multiple UAVs with a distributed control algorithm for system performance in the agricultural environment [3].

However, this rise in the use of UAVs has also raised significant security concerns, as UAVs can be used maliciously. In recent years, the main issues that can arise concerning the use of UAVs have been analyzed. For instance, the use of UAVs for surveillance purposes can pose risks to individual and intellectual privacy. Some of the malicious uses of UAVs are the espionage of essential and famous individuals and industrial espionage against companies, as explained by Lappa et al., who analyzed the perceptions of forty-three security specialists on the risks and threats to privacy associated with the use of

drones for surveillance [4]. Furthermore, there are other threats associated with UAV use, such as smuggling or even terrorist attacks, as indicated by Yaacoub et al., who analyzed the emerging threats of drone use in various attacks and countermeasures to prevent them [5]. For this reason, in the face of this emerging threat, the detection and neutralization of UAVs have become a priority for governments and security organizations worldwide.

The detection of UAVs is a complex challenge due to the diversity in their size, shape, and modes of operation; each UAV can have a different appearance and methods of functioning and communication, making it challenging to develop a valid and effective method for detecting any type of UAV. For this reason, Khan et al. provided a detailed review and critical analysis of existing techniques for the detection of unauthorized UAVs, such as acoustic detection, detection through computer vision techniques, detection based on RF signals, or the use of radar, pointing out their possible advantages and limitations [6]. Each method has its advantages and disadvantages in detection, and most involve using neural networks to identify the presence of UAVs based on the features extracted by each technique.

Regarding acoustic detection, this method is based on the identification of sound characteristics produced by UAVs when they are in operation, mainly due to the propellers' movement and the motors' noise. Shi et al. developed a detection algorithm using two arrays of four microphones arranged in a tetrahedral shape which could detect and locate amateur UAVs utilizing a time delay of arrival (TDOA) algorithm and a support vector machine (SVM) as a neural network [7]. Acoustic methods are instrumental in environments where UAVs are not visible, although their effectiveness can be affected by background noise and weather conditions.

Radar detection involves emitting signals that bounce off the UAV and return to the receiver to estimate the UAV's distance, speed, or shape. In [8], Semkin et al. implemented a detection and classification method based on radar cross-section (RCS) signatures. At the same time, Ezuma et al. developed a system to detect micro-UAVs using a frequency-modulated continuous-wave (FMCW) radar along with a constant false alarm rate (CFAR) detector with adaptive thresholding to detect the presence of UAVs [9].

RF-signal-based detection involves analyzing the communication signals that UAVs emit and receive during their operation by capturing the RF transmissions between the UAV and its controller. Machine learning techniques are typically used to classify and identify whether the signals are coming from a UAV. An example is the study by Kaplan et al., which detected the presence of UAVs communicating using the FHSS technique in the 2.4 GHz band. This was performed through an initial cyclostationarity method that reduces false positives and a variant of the STFT that extracts the necessary parameters to be fed into the classification algorithms [10].

Lastly, UAV detection using computer vision involves using cameras and advanced algorithms to identify UAVs in real-time images and videos. These computer vision systems are mainly based on convolutional neural networks (CNNs) that are trained with large datasets containing images of UAVs under different conditions. Park et al. analyzed six object detection models to evaluate the accuracy of each in detecting UAVs using a PTZ camera [11].

After conducting an in-depth analysis of each of the existing techniques for UAV detection, it was concluded that none of them is perfect, as they all have some limitations. However, computer vision techniques and RF signal analysis techniques offer the best performance, making them the most effective for UAV detection.

Furthermore, it was thought that, since a single technique may have limitations, using both methods together could enhance the system's effectiveness. This is because the limitations of one process can be compensated for by the other, resulting in a final system with high performance capabilities for UAV detection.

Therefore, this article proposes a hybrid approach integrating a radio spectrum analysis module employing autoencoders for UAV detection and a computer vision module based on the YOLOv10 (You Only Look Once) algorithm. Additionally, a localization

algorithm is developed, where, when two fixed cameras detect a UAV, they indicate the exact position to the PTZ mobile camera, along with a tracking algorithm that allows the PTZ camera to follow the UAV continuously.

The rest of the paper is organized as follows: Section 2 reviews the state of the art of the technologies used in this paper. Section 3 delves into the methodology used to develop the UAV detection system, explaining the system's logical and physical architecture and the algorithms designed for the two modules. Section 4 contains the tests conducted and an analysis of the results, and, finally, Section 5 presents the conclusions and future directions.

2. Related Work

The algorithms used for UAV detection, whether through RF signals or computer vision, have evolved over the years due to the specialization and development of UAVs, which necessitate increasingly novel, effective, and robust algorithms and systems to counter the threat of UAVs being used for malicious purposes.

Regarding the algorithms used for detection through computer vision, object detection algorithms are primarily employed, as these techniques use neural networks to locate and classify objects in an image. This is the goal when using cameras: to identify UAVs from real-time video feeds. These algorithms have continuously improved, always building upon CNNs, which are networks capable of recognizing and learning features and patterns in images [12].

Initially, two-stage detector algorithms were used, including the R-CNN, which has three steps: feature extraction by generating different regions to be analyzed in the image, running a CNN on each of these regions, and finally applying an SVM and a linear regressor to classify the object [13]. An evolution of this algorithm is the Faster R-CNN, which is faster because it generates regions more efficiently by introducing the region proposal network (RPN) and unifying the algorithm's stages [14].

Another object detection technique involves one-stage detectors, which are simpler and faster than two-stage detectors while maintaining a high level of effectiveness and accuracy, making them more suitable for real-time applications. Among them is the Single Shot Multibox Detector (SSD), presented by Liu et al., which achieves very good results in terms of precision for small objects by performing localization and classification tasks in a single step [15].

Finally, the most advanced object detection algorithm currently, also a one-stage detector, is the You Only Look Once (YOLO) algorithm, an algorithm introduced by Redmon et al. where the image is divided into different regions, and bounding boxes are predicted in each region along with their probabilities [16].

The YOLO algorithm has continuously advanced with each version, enhancing detection performance, speed, and accuracy. These improvements have proven valuable in UAV detection applications across different scenarios [17–19].

The version used in this paper, YOLOv10, introduces two major advancements compared to previous versions: a Consistent Dual Assignments mechanism for NMS-free training and an Efficiency–Accuracy-Driven Model Design. These innovations include a lightweight classification head, spatial-channel decoupled downsampling, and rank-guided block design, which enhance both accuracy and efficiency, as explained by Wang et al. [20].

YOLOv10 demonstrates superior accuracy and efficiency compared to previous YOLO versions and other advanced models, making it highly suitable for real-time detection tasks. As demonstrated by Wang et al. in his article, YOLOv10 achieves the highest accuracy (COCO AP%) with lower latency, which underscores its advantages in both precision and speed for real-time applications.

When it comes to tracking a detected object, Chen et al. presented an algorithm for the dynamic tracking of UAVs using deep learning techniques. Once the object is identified with YOLOv3, a PTZ camera follows the UAV based on its relative position, the PTZ angle, and the coordinates of its position in the image [21].

On the other hand, Sie et al. performed tracking using motion detection techniques with a Kernelized Correlation Filter (KCF) and Discriminative Correlation Filter (DCF). They also achieved 3D localization of a UAV by triangulating its position from the images captured from different viewpoints, successfully determining the object's spatial coordinates [22].

In terms of identification through radio spectrum analysis techniques, there are several methodologies and tools employed, as well as a variety of neural network algorithms, for UAV detection, each with specific characteristics and limitations. Ezuma et al. developed a system capable of detecting and classifying up to 17 types of UAVs in environments with interfering Wi-Fi and Bluetooth signals using a multi-stage detection approach that distinguishes between UAV and interfering device signals and performs classification using neighborhood component analysis (NCA) [23].

Nemer et al. designed a system that employs four hierarchically organized classifiers. In this system, the first classifier is responsible for identifying the UAV presence, the second classifier determines its type, and the last two classifiers are dedicated to identifying the flight mode of the UAV [24]. For their part, AlKhonaini et al. implemented a similar hierarchical classification approach, but used a reinforcement learning algorithm which applies Monte Carlo sampling to estimate the policy gradient, thus achieving an improvement in detection and classification accuracy compared to the previous study [25].

In the field of radio fingerprinting detection, software-defined radio (SDR) is common and frequently used, as it allows the processing of radio-frequency signals through software, without requiring hardware modifications. Xiang et al. proposed a low-cost solution for UAV detection which combines an SDR with an antenna array. Through an amplitude comparison and least squares search method, they succeeded in estimating the angle of the incident signal, although the system had limitations in urban environments due to signal interference [26]. Flak developed a continuous monitoring system in the 2.400 GHz to 2.483 GHz range using a LimeSDR-USB board, which allows the detection of UAV communication signals by distinguishing specific features [27].

The use of autoencoders has also been presented in the literature for UAV detection, as it is an appropriate method for anomaly detection. Jiang et al. developed a novel detection system based on a spatial-temporal autoencoder using a temporal convolution network (TCN) and CNN along with a dynamic threshold that indicates the presence of anomalies based on the correlations of the flight data [28].

Basak et al. developed an autoencoder-based system called DEFEND, which uses RF signals obtained through an SDR as input data and performs two tasks: classifying signals that are known to the system and detecting those that are not. The novel signals, after being detected, are clustered based on the extracted features to become signals known to the system [29].

Contribution of This Paper

The proposed system in this article presents a hybrid approach for detecting intrusive UAVs, combining two complementary detection techniques to ensure effective and rapid detection. The employed techniques are the detection of radio signals produced between the UAV and its controller using autoencoders that can identify anomalies in the analyzed frequency bands, and visual detection through a camera system with YOLOv10, which can locate the UAV in 3D and track it using an autonomous PTZ camera. In summary, the main contributions of this work are as follows:

1. Development of a detection system using autoencoders based on radio signals that can operate in various frequency bands up to 100 MHz of bandwidth, with these bands being selectable to monitor those most likely to be used by intrusive UAVs. Moreover, the autoencoders do not require a database with UAV signals; they only need samples of the frequency bands to be analyzed when there are no UAVs present for training, which simplifies the creation of the database as it does not require data labeling or samples with an operating UAV.

2. Development of a scalable system with fixed cameras plus an autonomous PTZ camera, where all cameras perform detection with the YOLOv10 algorithm trained on a dataset of over 60,000 images obtained from various online sources.
3. Implementation of a triangulation algorithm with fixed cameras for 3D localization of the UAV to provide the exact position to the PTZ camera, in addition to a tracking algorithm through which the PTZ camera continuously follows the UAV once detected, keeping it always at the center of the camera's image to prevent it from being lost.
4. Implementation of the complete system in which both the SDR module and the camera module are in constant communication to indicate when either of them detects a UAV. There is also communication using the Contact ID and SIA protocol with an external alarm-receiving center, which is responsible for deciding the course of action when an intruding UAV is detected.

3. Methodology

This part of the paper will discuss the algorithms designed for the implementation of a hybrid drone detection system based on radio spectrum analysis, computer vision techniques, and localization and tracking procedures. Initially, the hardware used to construct the integral system will be described.

3.1. Hardware

The developed system is based on two computers that are responsible for controlling and managing each module. One controls the SDR for the analysis of RF signals. At the same time, the other is responsible for the computer vision block from the images of the various cameras. A router is needed to connect the system components via Ethernet for communication. This allows communication between the computers through sockets and between cameras through IP addresses, ensuring all devices are on the same local network.

Figure 1 shows all the devices that make up the system. On the the left side is the detection module that uses the RF signals, consisting of the SDR and a PC inside an IP67 protection box located next to the cameras, and, on the right side of the figure, is the computer vision module. The router is the central element through which all devices are interconnected via IP addressing or being physically connected via Ethernet. Likewise, the router also has Internet access to communicate with the alarm-receiving center (ARC), which is located in the security company's facilities.

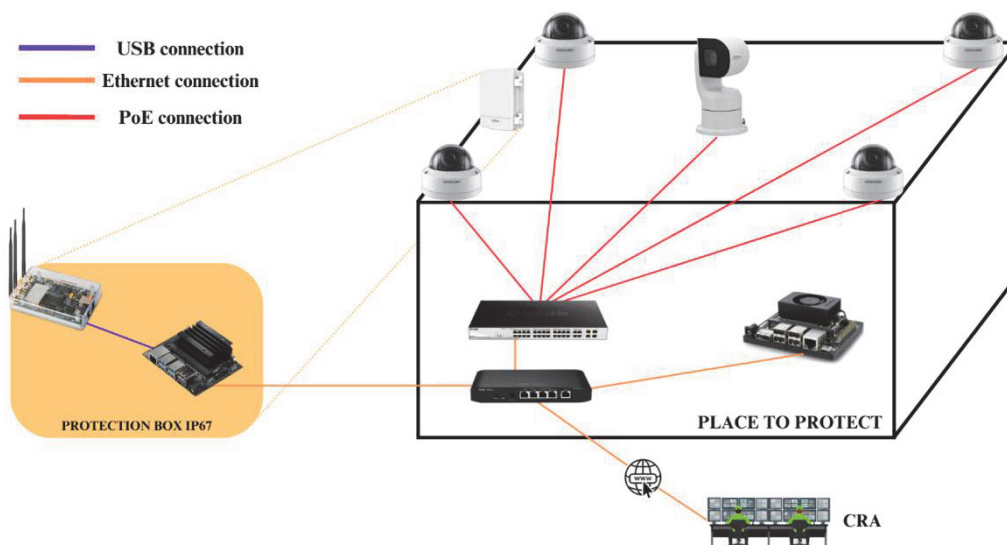


Figure 1. Physical architecture of the system, where the colour of each arrow indicates the type of physical connection between the devices.

The four fixed cameras are the Hikvision DS-2CD2743G0-IZS model (Hikvision Digital Technology Co., Ltd., Hangzhou, China). In contrast, the PTZ camera is the Dahua PTZ1A225-HNR-XA (Zhejiang Dahua Technology Co., Ltd., Hangzhou, China), which is a camera with a 2 MP CMOS sensor capable of 360° horizontal rotation and 180° vertical tilt, featuring a 25× optical zoom. Both camera models have a minimum IP66 rating, making them suitable for outdoor installation, and they support PoE+.

The SDR used for receiving the RF signals is the bladeRF 2.0 micro xA4 (Nuand LLC, Rochester, NY, USA), which offers a frequency range from 47 MHz to 6 GHz and a sampling rate of 61.44 MHz, expandable up to 122.88 MHz, with 2×2 MIMO. It has an integrated Analog Devices AD9361 transceiver (Analog Devices, Inc., Wilmington, MA, USA) and an Altera Cyclone V FPGA (Altera Corporation, ahora parte de Intel Corporation, San Jose, CA, USA), enabling real-time transmission and reception for advanced applications. It is physically connected to the computer via USB 3.0.

The two computers used are from the NVIDIA Jetson family and are both powerful platforms specialized in AI development. The NVIDIA device responsible for the SDR module is the Jetson Nano Developer Kit (NVIDIA Corporation, Santa Clara, CA, USA), while the one for the cameras is the Jetson Orin Nano Developer Kit (NVIDIA Corporation, Santa Clara, CA, USA). The latter is considerably more powerful than the former, as the computational cost in computer vision applications is high, requiring a device capable of managing five cameras using YOLOv10.

Finally, to perform IP addressing for the various components, a RG-EG105G-V2 (Ruijie Networks Co., Ltd., Fuzhou, China) router is used, connecting both computers and the network cable that provides Internet to communicate with the ARC over the network. Additionally, the PFS3218-16GT-135 PoE (Zhejiang Dahua Technology Co., Ltd., Hangzhou, China) switch is used to connect all the cameras as they are all PoE.

3.2. Logic Architecture

Before delving into the development of the algorithms for both modules, it is important to understand the general logic of the system and how both modules are related and communicate with each other.

Initially, the system was designed without fixed cameras, only a PTZ camera, so the SDR module was the first step in detection. When it detected a signal that could be coming from a UAV, it would alert the PTZ camera module, which would then begin sweeping the entire area until the UAV was located.

However, the system has been improved with the addition of fixed cameras. Now, when two fixed cameras simultaneously detect the UAV, they send the exact position to the PTZ camera, and, once it detects the UAV, it begins to track it. The main objective of the system is for the PTZ camera to identify the UAV and continuously follow it, allowing the security company using this system to always know the UAV's location. For this reason, the PTZ camera can detect the UAV through two methods:

1. Through sweeping. This situation occurs either when the SDR module detects the UAV or when it is detected by only one fixed camera. In both cases, the UAV's position is unknown, so the PTZ camera begins to sweep its surroundings to detect it.
2. After being located by two fixed cameras. If two or more fixed cameras detect the UAV, they send its position directly to the PTZ camera, which then moves directly to the UAV's location without needing to sweep.

Furthermore, once the PTZ camera locates the UAV, it tracks it. However, if the UAV disappears from the camera's view at any point because it has flown away or moved too quickly for the camera to follow, the PTZ camera begins a post-loss sweeping period to relocate it. If the UAV is not found after the sweeping period and no alarm is received from another camera or the SDR, the UAV is considered to have left the area.

Finally, when the UAV is detected, an alert is sent to the security company's ARC so they are aware of the presence of a UAV in the protected area. Once the UAV is located and

tracked by the PTZ camera, videos of the UAV will be periodically sent to the ARC, along with a message when the UAV has disappeared.

Communication with the ARC is conducted through the Data Transport Layer Security (DTLS) protocol, which creates a secure and robust communication link between the detection system, acting as the client, and the ARC, acting as the server. The alarms and videos themselves are sent using the SIA DC-09-2021 protocol (Security Industry Association, Silver Spring, MD, USA). On the other hand, communication between the two modules is carried out via sockets programmed in Python, just like the rest of the system.

3.3. SDR Module

In the development of the algorithm for the detection of drones through radio signals using artificial intelligence, it was necessary to consider some conditions that have influenced the development of the algorithm and the selection of the neural networks to be used.

Firstly, the system needed to be a generalist to detect any UAV, regardless of the type, the communication and control protocols it uses, or the frequency bands it operates on. This is because the various UAV detection algorithms in the literature focus on detecting UAVs that use only specific protocols or detect only a single frequency band.

Secondly, the system must be plug and play, meaning its installation should not require testing or prior training at the deployment site. This implies that the selected detection technique should not require an extensive and varied database to be created at the installation site, as would be necessary with supervised methods such as CNNs. These methods require collecting specific spectral data under different environmental conditions and with different UAV models.

For these reasons, autoencoders were chosen. These constitute an unsupervised learning technique that is trained only with unlabeled spectrum samples when no UAVs are present. This means there is no need to manually generate a database at the deployment site, ensuring the system meets the plug-and-play condition. Moreover, autoencoders allow for the creation of a general algorithm, which will be explained later, that is capable of detecting all types of UAVs as required.

The developed algorithm consists of two parts that run in parallel. The first part is responsible for receiving and processing the RF signals obtained with the SDR, and the second part analyzes the samples obtained to identify using the autoencoders whether there is an intruding UAV based on the radio signals.

3.3.1. Reception and Processing of Data from the SDR

The first part is based on GNU Radio, and its purpose is to capture signals from different frequency bands, calculate their power, and store them in text files, which will be used in the second block for detection. Figure 2 shows the structure followed in GNU Radio, consisting of four GNU Radio blocks. The first two blocks are native to the software, while the subsequent two have been created based on the project's requirements.

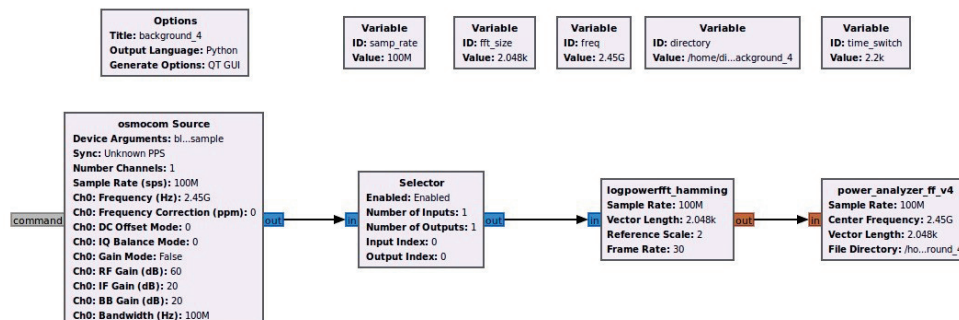


Figure 2. GNU Radio schematic used to receive signals from the bladeRF SDR and process them for input to the detection system.

The first GNU block, Osmocom Source, is responsible for receiving complex signals from the bladeRF SDR, where parameters such as the bandwidth, frequency, and gain of the receiving system are set. The second block, Selector, is used to block data transmission and is only employed when changing the system's frequency. If it is not blocked, there is a risk of data being stored incorrectly.

The third block receives the complex signals from the SDR and calculates the power in dBm by applying a Hamming window. Finally, the fourth block stores the calculated power of that frequency band in a text file, creating a separate file for each analyzed band.

Thus, the system first monitors a series of frequency bands with a bandwidth of 100 MHz, then calculates the power at 2048 points within that band, and, finally, stores those values in a text file that is updated each time values for that frequency band are received.

For the tests conducted, three frequency bands commonly employed in UAV usage are analyzed: 950 MHz, 2.450 GHz, and 5.80 GHz. In this way, three different files are generated, containing various samples obtained from each of the frequencies. Each sample consists of 2048 power values, which, together, represent 100 MHz of that particular frequency band. These samples of 2048 power values are the data input to the autoencoders for detection.

3.3.2. Detecting Algorithm Using Autoencoders

This second part involves analyzing the samples obtained in the first block by inputting them into autoencoders to determine if a UAV is present in any of the analyzed frequency bands.

Autoencoders are used as anomaly detectors within the frequency bands. When the system is launched for the first time after installation, it begins collecting data from the bands to be analyzed, which will be used to train the autoencoders. These samples correspond to the signal power in the absence of a UAV, allowing the training of a specific autoencoder for each frequency band. This data collection and training process is automatic, and, once it is complete, the detection system starts, along with the camera module. If the system shuts down for any reason, it will resume directly with the detection process upon restarting as the training will already have been completed.

Since the training data correspond to the signal power in the absence of a UAV, when inference is performed on the signals constantly received from each band, if the input data do not contain UAV signals, the autoencoders, trained on such data, will reconstruct the signal with low error. However, if the input data contain UAV signals, which were not present during training, the reconstruction error will be high, indicating the possible presence of a UAV.

Figure 3 demonstrates this. In the example on the left, the spectrum of the 2.45 MHz band and the reconstruction of the band after passing through the autoencoders are shown. The reconstruction is similar to the input signal, resulting in a low error, as shown in the black spectrum, indicating no UAV in this band. In contrast, in the example on the right, the autoencoder is unable to reconstruct the input signals, resulting in a high reconstruction error, indicating the potential presence of a UAV. The figure was generated in Python 3.8 using the Matplotlib library, employing the power vectors extracted from GNU Radio and the power vectors obtained after processing them through the autoencoders.

Based on this, the system continuously analyzes each frequency band for anomalies that could indicate the presence of a UAV. An alert is sent to the camera module only if anomalies are detected in the same section of the frequency band in two consecutive readings. An anomaly is considered to be present when the reconstruction error of the signal in that part of the band exceeds a certain threshold.

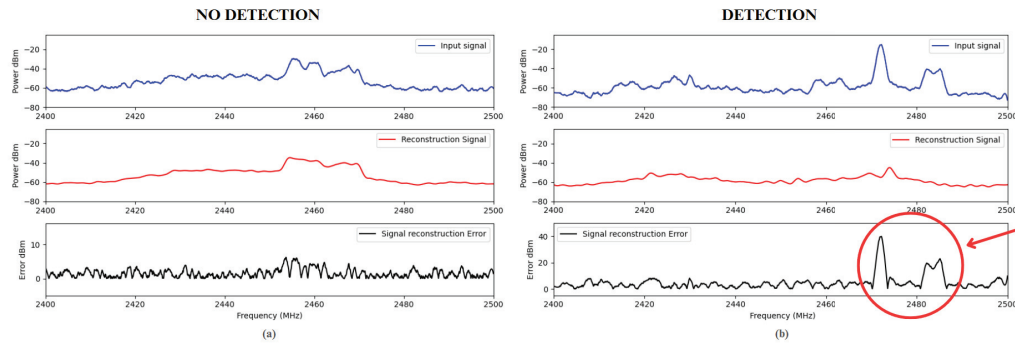


Figure 3. Examples of the performance of the autoencoder-based detection algorithm: (a) shows a situation of no detection as the autoencoder can reconstruct the input signal properly, (b) shows a detection in the part surrounded by the red circle, because the reconstruction error there is very high, as the autoencoder has not been able to reconstruct the input signal, indicating the possible presence of a UAV.

The requirement of two consecutive readings with anomalies is due to the high likelihood of false positives caused by interfering signals in the frequency bands. An anomaly detected in a single reading could result from a momentary interference and may not necessarily indicate the presence of a UAV. Therefore, by requiring anomalies to be recurrent in two consecutive readings, the possibility of false positives is reduced, improving detection accuracy.

The following part explains the steps the algorithm takes to determine if a UAV is present in a frequency band after two consecutive data readings.

1. **Sample Reception:** N samples of a specific frequency are received. These samples are fed into the autoencoder, producing N corresponding reconstructions.
2. **Reconstruction Error Calculation:** For each sample and its reconstruction, the following takes place:
 - Both the sample and its reconstruction are divided into M segments.
 - The reconstruction error is calculated for each segment individually.
 - It is checked whether the reconstruction error in any segment exceeds a predefined threshold.
 - The segments m in sample n where the error exceeds the threshold are recorded.
3. **Storing Results:** This process is repeated for all N samples, storing the segments where the reconstruction error exceeds the threshold for each sample.
4. **Segment Analysis:** For each segment, the percentage of samples that have exceeded the initial reconstruction error threshold is calculated.
5. **Anomaly Detection:** If the percentage in any given segment exceeds a second detection threshold, an anomaly is considered to be present in that segment of the analyzed frequency.
6. **Detection Confirmation:** The process is repeated when N new samples of the same frequency are received. If an anomaly is detected in the same segment of the frequency in this second iteration, the presence of a UAV is confirmed.

Figure 4 graphically illustrates step 2, “Reconstruction Error Calculation”, from the previously described process, while Figure 5 shows the steps of “Storing Results”, “Segment Analysis”, and “Anomaly Detection”.

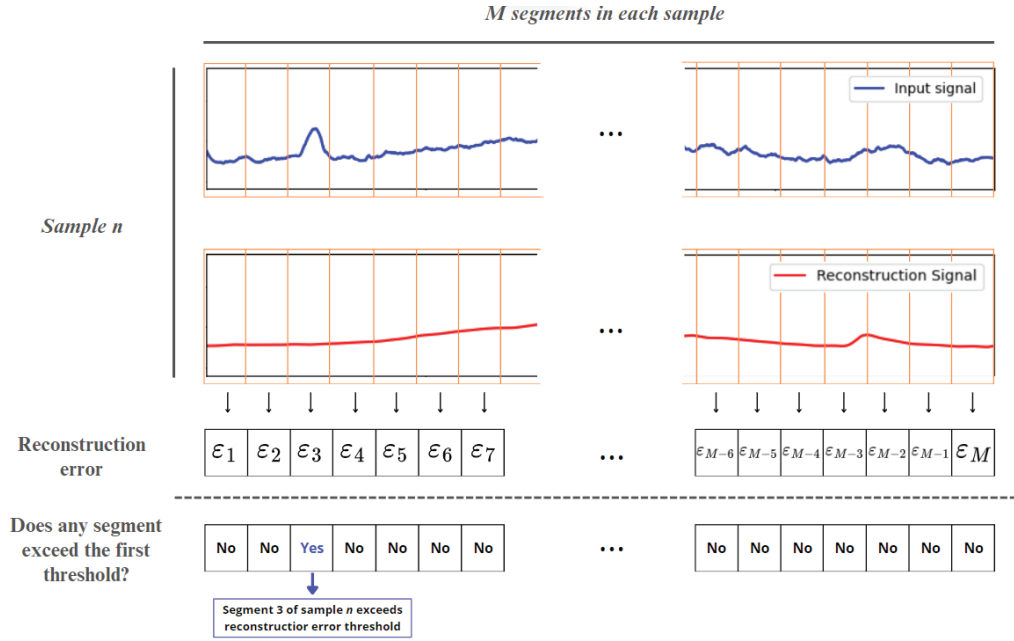


Figure 4. Reconstruction Error Calculation process performed with each sample received.

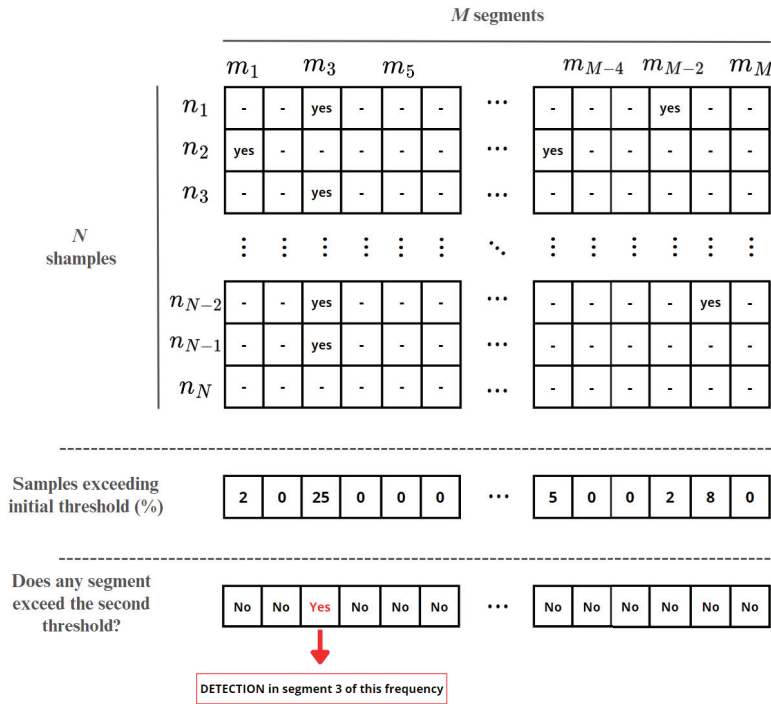


Figure 5. Steps of Storing Results, Segment Analysis, and Anomaly Detection in the autoencoder-based detection process.

Thus, autoencoders are utilized to identify the presence of potential intrusive UAVs by analyzing and comparing the RF signals received from each band with the training data of the band obtained in the absence of UAVs. This approach ensures that the system does not require the manual creation of a complex database, wherein UAVs would need to be flown in the installation area in order for UAV signals to be captured for the database. Moreover, with this approach using autoencoders, the algorithm is capable of identifying all types of UAVs.

The architecture of the developed autoencoders can be seen in Figure 6 and is composed of seven layers, where the input layer has 2048 values, corresponding to the length of the power samples, and the dimensionality is reduced to only 32 neurons. Each layer has a ReLU activation function, except for the output layer, which uses a linear activation function.

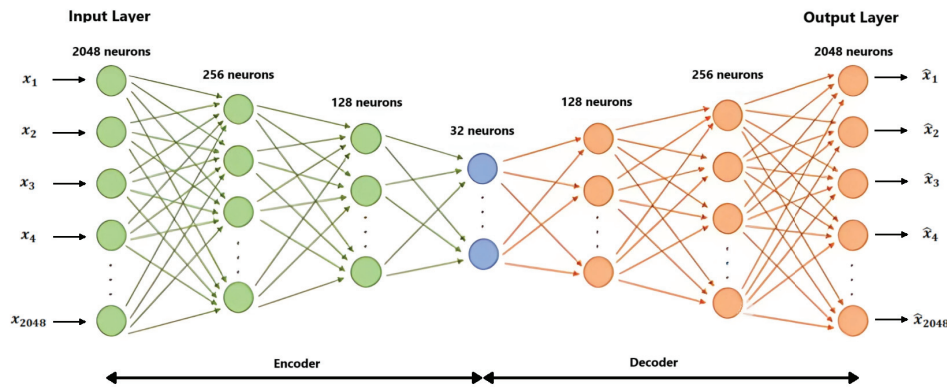


Figure 6. Architecture of autoencoders used for drone detection. The green part, encoder, represents the compression of the input data. The blue is the latent space, which is the representation of compressed data, and the orange, decoder, represents the reconstruction up to the original dimension.

For the training process, the input data are passed through a Savitzky–Golay (Savgol) filter, and the model is trained for 400 epochs using the Adam optimizer with an adaptive learning rate of 0.001 that varies based on the model’s validation performance. The model’s loss function is the Mean Absolute Error (MAE), which evaluates accuracy by measuring the average difference between the autoencoder outputs and the original inputs. The MAE is suitable for this model as it provides a direct measure of the magnitude of the prediction error in the same units as the original data, facilitating the use of a threshold for detection.

3.4. Computer Vision Module

3.4.1. YOLOv10

As previously explained, all the cameras in the system use YOLOv10 for UAV detection, utilizing a custom model trained with a dataset comprising the Birds vs. Drones dataset [30], Maciej and Marek UAV dataset [31], and the GSV-Cities dataset [32]. From these three sources, 60,178 images were used, divided into 48,521 for training and 11,657 for validation.

The reason for using such an extensive dataset is to represent as many real-world scenarios as possible. It includes images of UAVs of different sizes and types, at various distances and under different weather conditions, and in diverse flight environments. There are also background images and elements that could potentially confuse detection, such as birds.

The training was carried out using the YOLOv10-S model, a small version that balances speed and accuracy. It is the most powerful version that could be trained using the created dataset and the selected hyperparameters given that the computer used for training did not have sufficient computational capacity for more complex architectures.

The training was set to 400 epochs with a batch size of 32 and employed data augmentation techniques. The training process lasted approximately three days and yielded very good results when analyzing the metrics extracted by YOLOv10, as shown in Figure 7, achieving very high precision and recall values.

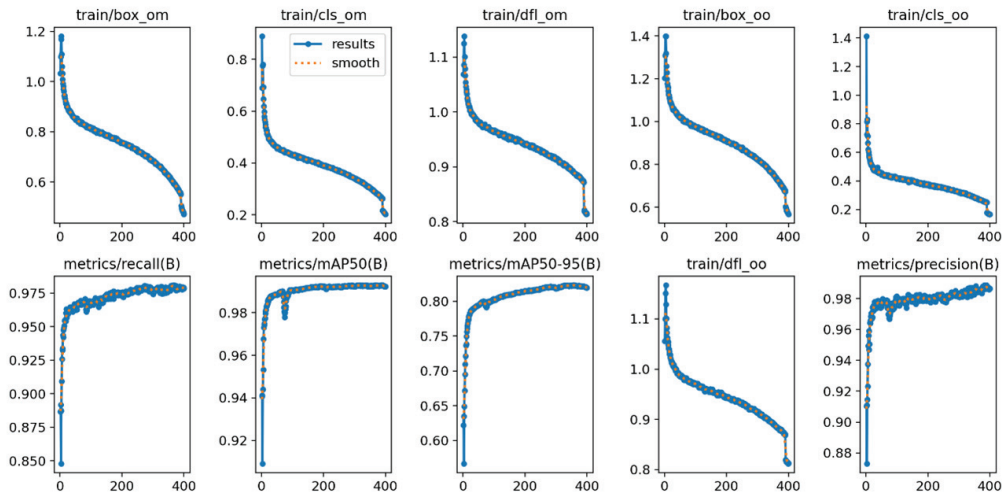


Figure 7. Results of the model trained with YOLOv10.

3.4.2. Tracking Algorithm

The UAV tracking process is conducted using the PTZ camera once the UAV is detected, automatically adjusting the camera's orientation to keep the UAV continuously located, thanks to the developed algorithm, which is based on three main functions.

The first function involves adjusting the camera's zoom to optimize the UAV's size relative to the total image frame. This is achieved by analyzing the percentage of the UAV's area compared to the total area based on the camera's current zoom, as illustrated in Figure 8, and then adjusting the zoom to ensure the UAV has a specified size in the frame.

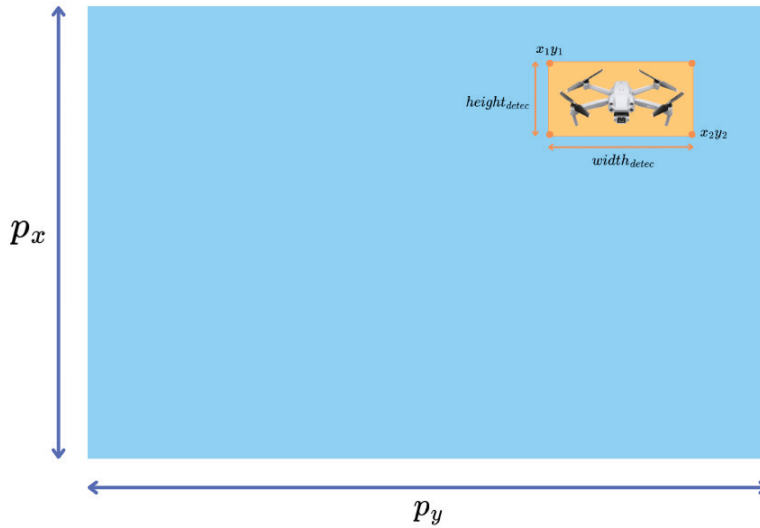


Figure 8. Parameters used for zoom adaptation from UAV detection.

The established value for proper operation is to ensure that the detection area represents between 0.5% and 1.2% of the screen area. If the UAV's area is above or below these threshold values, the camera's zoom is adjusted accordingly.

These limits ensure that the UAV is large enough to be detected by YOLOv10 while also being small enough relative to the screen to prevent it from quickly disappearing from the camera's view due to its rapid movements. If the UAV is too large on the screen, it is more likely to move out of the camera's field of view easily.

The second function is used to calculate the field of view (FoV) of the PTZ camera based on the current zoom level and knowing the camera's specifications. It first establishes a simplified linear relationship between the zoom level (0 to 25) and the focal length (4.8 mm to 120 mm) using Equation (1), which computes the focal length $f(z)$ from the zoom level z :

$$f(z) = f_{min} + (f_{max} - f_{min}) * \frac{z}{zoom_{max}} \quad (1)$$

Then, the field of view (FoV) is calculated using the focal length and the camera sensor's size, as shown in Equation (2):

$$FoV = 2 * \arctan\left(\frac{d}{2 * f(z)}\right) \quad (2)$$

where d is the sensor dimension representing the horizontal FoV width and the vertical FoV height. The sensor size values are assumed to be 5.7 mm $\times$ 2.88 mm, based on the manufacturer's information.

Finally, the third function calculates the relative movement that the camera must make on the pan and tilt axes to keep the UAV at the center of the camera view. This function is based on trigonometry applied to a circle and its chord, where the chord represents the deviation in the image. The calculation is performed independently for both axes of the screen (horizontal and vertical).

First, the camera's FoV is needed from the current zoom level, calculated using the previous function, as the FoV defines the angle of view (θ) that delimits the chord used in the calculations. To simplify the trigonometric calculations, a circle with a radius of $r = 1$ is used, allowing the chord length to depend only on the angle θ . Second, it is necessary to calculate the distance from the center of the circle to the chord c_0 using Equation (3), as illustrated in Figure 9a.

$$c_0 = \cos\left(\frac{\theta}{2}\right) * r \quad (3)$$

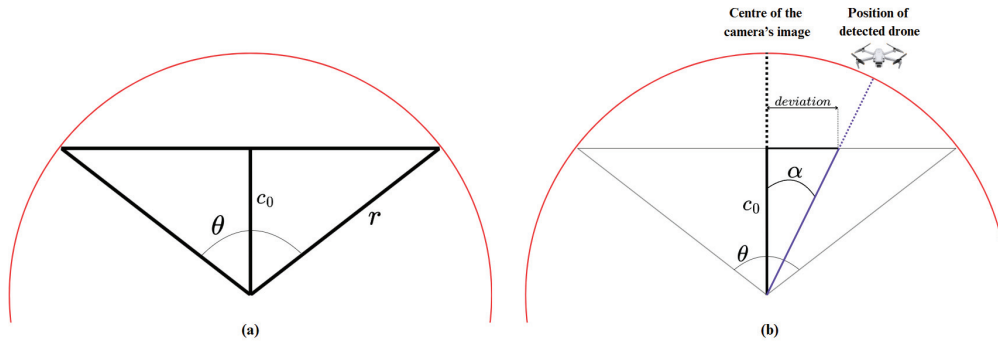


Figure 9. Representation of the trigonometry used to calculate c_0 in (a) and *deviation* in (b), where, the black dotted line refers to the centre of the image over the centre of the circle and the purple dotted line to the position where the UAV is located.

Once this value is known, any deviation of the UAV from the center of the image can be calculated, allowing us to determine α , which represents the angular correction that the camera must apply to center the UAV. The *deviation* is the difference calculated between the midpoint of the image and the center of the detection on the axis. Figure 9b illustrates how the angular correction is derived from the deviation of the UAV detection from the center, and Equation (4) provides the value:

$$\alpha = \arctan\left(\frac{\text{deviation}}{c_0}\right) \quad (4)$$

Compared to existing tracking systems, which often rely on fixed zoom levels or less dynamic orientation adjustments, this algorithm provides a significant improvement in responsiveness and accuracy. Its reliance on real-time data and camera specifications allow it to operate effectively without requiring extensive computational resources, making it a robust and efficient solution for UAV tracking.

The key innovation is the dynamic zoom adjustment, which continuously analyzes the UAV's size in the frame to balance visibility and coverage, preventing the UAV from becoming too small to detect or moving out of view. Additionally, the algorithm calculates the camera's FoV in real time based on the zoom level, leveraging the camera's optical specifications to fine-tune its orientation with precision. Lastly, to ensure smooth and accurate tracking, continuous recalibration is carried out based on trigonometric angular correction.

3.4.3. Localization Algorithm

The localization of the UAV is performed using fixed cameras. If two or more cameras detect the UAV, they send its exact position to the PTZ camera using an approach based on the intersection of the detection vectors. Figure 10 shows how localization is carried out when two cameras simultaneously detect a UAV.

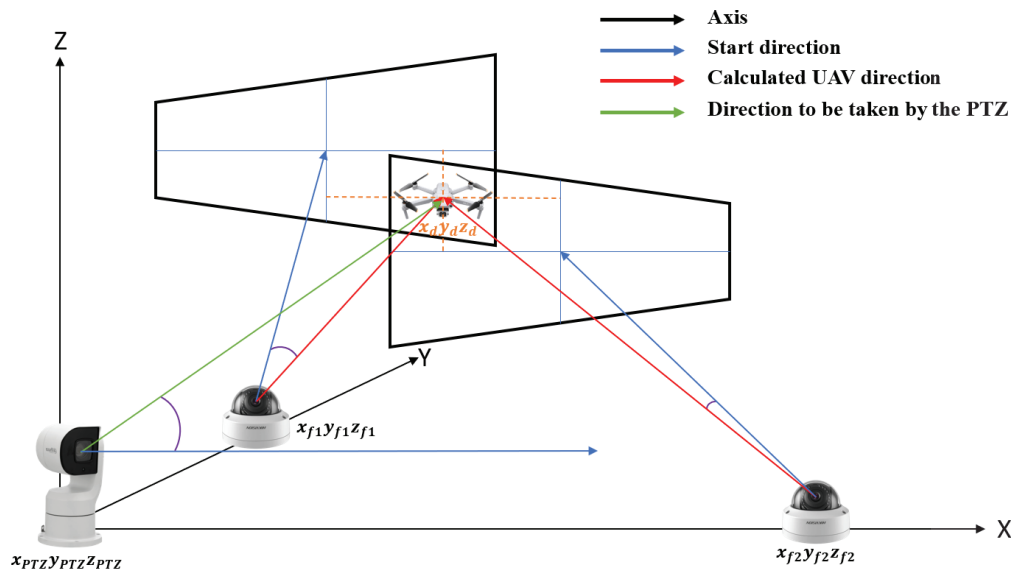


Figure 10. Schematic for UAV position estimation using two fixed cameras.

Each fixed camera has a “start direction”, which is the direction in which the camera is oriented. When a UAV is detected, the “calculated UAV direction” is computed, indicating the precise orientation of the UAV relative to the camera that detected it. These calculated directions are three-dimensional vectors obtained from the start direction of the camera and its deviation to the central point of the detection box, calculating the deviation angle in terms of pan (horizontal) and tilt (vertical) within a range of 0° to 360°.

A non-linear least squares optimization method finds the lines' intersection points representing the fixed cameras' detection directions. The objective function to minimize is the sum of the squared distances from an estimated point in three-dimensional space to each of the lines defined by the detection directions of the cameras. Mathematically, this is represented by Equation (5):

$$D^2 = \left\| \vec{r}_i - \vec{p} - \left(\frac{(\vec{r}_i - \vec{p}) * \vec{d}_i}{\vec{d}_i * \vec{d}_i} \right) * \vec{d}_i \right\|^2 \quad (5)$$

$\vec{r}_i$ is the origin point of camera i , while $\vec{p}$ is the estimated point in space representing the UAV's position and $\vec{d}_i$ is the detection direction of camera i . This calculation is based on

projecting the estimated point onto each of the lines and measuring the Euclidean distance between the point and its projection.

Minimization of this function is performed using the Sequential Least Squares Programming (SLSQP) algorithm from the SciPy library. The goal is to find the values of x, y, z that minimize the following function (6):

$$\min_{\vec{p}} \sum_i D_i^2 \quad (6)$$

Once the UAV's position is estimated, the vector connecting the PTZ camera's position to the UAV's position is calculated. This vector is then transformed into pan and tilt units using Equations (7) and (8), which normalize the calculated angles to an appropriate range for controlling the PTZ camera:

$$pan = \frac{atan2(y, x)}{\pi} \quad (7)$$

$$tilt = \frac{arcsin(z)}{2\pi} \quad (8)$$

The innovation in the localization method using fixed cameras lies in its combination of simplicity and precision in determining the UAV's position. Each camera computes the UAV's orientation based on deviations from its initial direction, resulting in three-dimensional vectors. The intersection of these vectors from multiple cameras is then calculated using a non-linear least squares optimization method. This approach minimizes the Euclidean distance between the estimated point and the vectors, ensuring precise localization.

This method enhances accuracy without requiring complex hardware setups like multiple PTZ cameras. Additionally, it operates in real time and integrates seamlessly with the PTZ camera, directly providing positional data to adjust its focus and tracking. The novelty also lies in its robust mathematical foundation, which employs the SLSQP algorithm, ensuring efficient and accurate computation, even in dynamically changing scenarios.

4. Results Discussion

To evaluate the performance of the developed algorithms and the overall system, various tests were conducted to cover as many scenarios as possible. These tests were performed in different spaces using different types of UAVs and at various times of the day to assess the response of the SDR module and the cameras to the presence of an intrusive UAV, thereby determining the system's effectiveness in terms of detection.

The testing process was also crucial for analyzing the system's limitations and identifying which parameters or situations lead to successful or unsuccessful detection. Factors such as the distance to the UAV, its speed, and the electromagnetic interference in the testing environment were considered.

In most of the tests conducted, the results were analyzed separately, examining the performance of each module individually, as communication between the two modules has never posed a problem. The primary challenge of the project was to ensure that both the SDR module with the autoencoders and the computer vision module accurately detect UAVs, as well as perform localization and tracking functions appropriately.

Each module presented its challenges. For radio-frequency signals, it was necessary to determine whether the algorithm based on autoencoders is effective and how parameters such as the distance of the UAV, which affects the signal strength received by the SDR, impact performance. Additionally, it was important to examine if the algorithm's effectiveness is compromised when operating in environments with numerous interfering signals in the frequency bands being analyzed.

For the camera module, the main challenges were related to assessing the detection performance based on the UAV's distance and speed, determining whether the localization system can correctly triangulate the UAV's position and send the precise location to the PTZ camera, and, finally, evaluating if the PTZ camera can successfully track the UAV.

4.1. SDR Module Results

The first and primary aspect to analyze in the behavior of the SDR module's algorithm was its ability to detect signals from intrusive UAVs. For this purpose, tests were conducted in different environments where a UAV was flown at various distances to determine if the system can indicate the presence of a UAV.

Figure 11 shows the results of all the tests conducted, illustrating the detection success rate as a function of the distance to the UAV. The success rate is simply the percentage of times the UAV was detected by the algorithm out of the total number of times the UAV was flown at that distance.

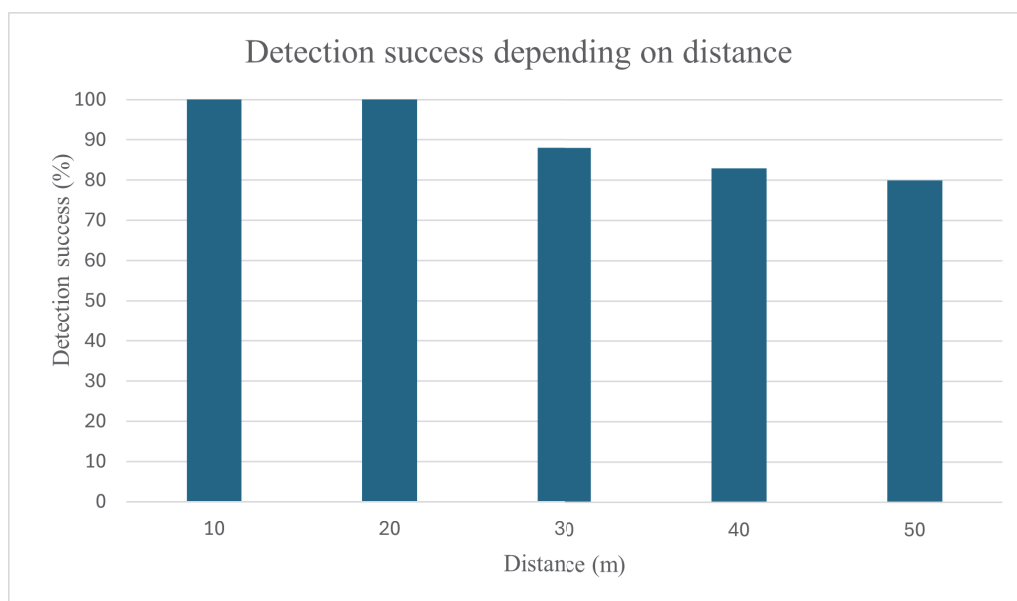


Figure 11. Detection test results using the SDR module as a function of distance from the UAV.

The results show that the UAV detection system achieved a performance of practically 100% at distances of 10 and 20 m, indicating high accuracy and reliability at close ranges. However, as the distance increased, the detection efficiency decreased. This trend can be explained by the signal strength received by the SDR; the further away the UAV is, the lower the received power, making it more challenging for the algorithm to identify anomalies in the received signals.

Nonetheless, the results for these longer distances are still favorable, with successful detection rates of 89% and 83% at distances of 30 and 40 m, respectively, dropping to the lowest value of 80% at a distance of 50 m.

It is important to note that the effectiveness of the algorithm is significantly influenced by the dataset used to train the autoencoders. Generally, the larger the dataset, the better the detection algorithm will perform, as it will have been trained with a greater number of samples. This is particularly crucial in highly saturated bands like those used by Wi-Fi, where there are many interfering signals. A more extensive dataset would include more samples of all these non-UAV signals.

Ideally, when installing the system, the algorithm should collect data from the spectrum of each band for an entire day. This would provide samples of the radio spectrum's state in each band when no UAV is present at different times of the day, which is important as the signals present may vary throughout the day. For these tests, the training was

conducted with data collected over periods of less than two hours, so, with more data, detection could be more effective.

The second aspect analyzed was the average time taken to detect the UAV as a function of distance. As mentioned earlier, greater distances can make detection more challenging due to the weaker signal strength received, which also results in increased detection time.

Figure 12 illustrates this situation, clearly showing that, as the distance increases, the time required for the algorithm to identify the presence of the UAV also increases.

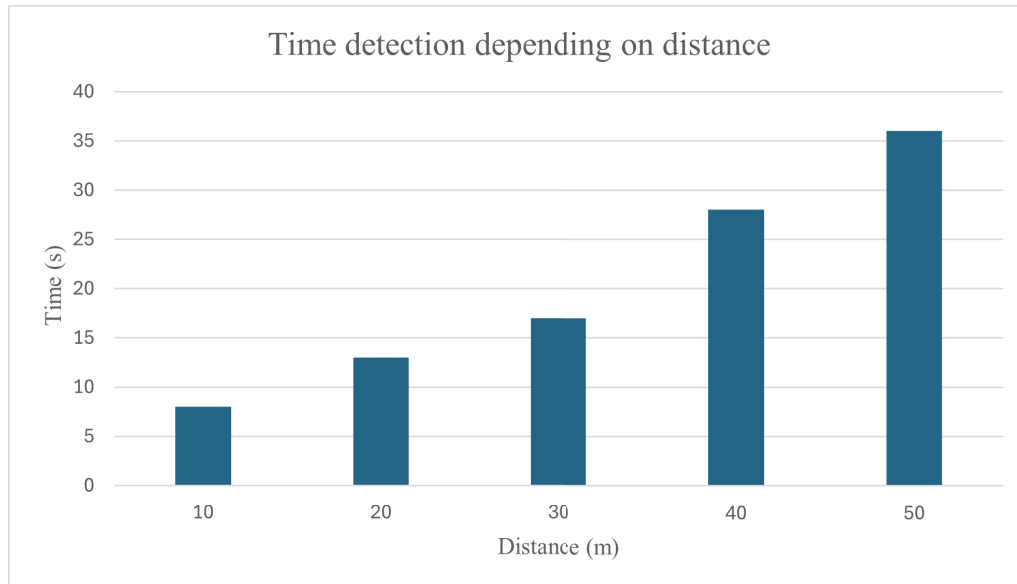


Figure 12. Time required for UAV detection as a function of distance.

At a distance of 10 m, the detection time was minimal because, as previously explained, determining the presence of a UAV requires detecting an anomaly in the same part of the frequency band in two consecutive readings of the same band. Therefore, the time needed to perform two readings of the same band is approximately what was observed for the 10 m distance.

As the distance increases, the detection time also increases because the UAV is not always detected in two consecutive readings. If the UAV is not detected in one reading, at least two more readings are needed to make the detection, thereby increasing the time required. The greater the distance, the weaker the signal received, making it more likely that detecting the UAV will take longer.

This approach is designed to minimize the number of false positives that the system might produce. By requiring two consecutive readings with anomalies in the same part of the band, the system nearly eliminates false positives, though it makes detection more challenging.

During testing, the algorithm's resistance to false positives was also analyzed by running the system to monitor the frequency bands without flying any UAVs, to check if the UAV alarm would trigger when there was none. The results were highly positive, as very few false positives occurred. The few false positives observed happened in environments with a lot of interference, such as an office, where the presence of numerous wireless devices increased the likelihood of false positives. Nevertheless, the system still proved to be quite robust.

4.2. Computer Vision Module Results

The first aspect to analyze was whether the trained model detects UAVs effectively. In the YOLOv10 training graphs, the results appeared highly promising, but these were obtained using static images. To determine if the model truly performs well for the

developed system, it was essential to test it with real-time images in environments where the system could be deployed.

For this reason, Figure 13 illustrates the success rate when a UAV was flown at different distances from the camera. The goal was to verify whether the model detects UAVs properly and to analyze how distance affects detection.

As can be seen from the results, the overall detection performance was very good. The detection rate was always above 75% for distances up to 50 m. At closer distances, the UAV was always detected, with the percentage decreasing as the distance increased, but strong results were maintained throughout.

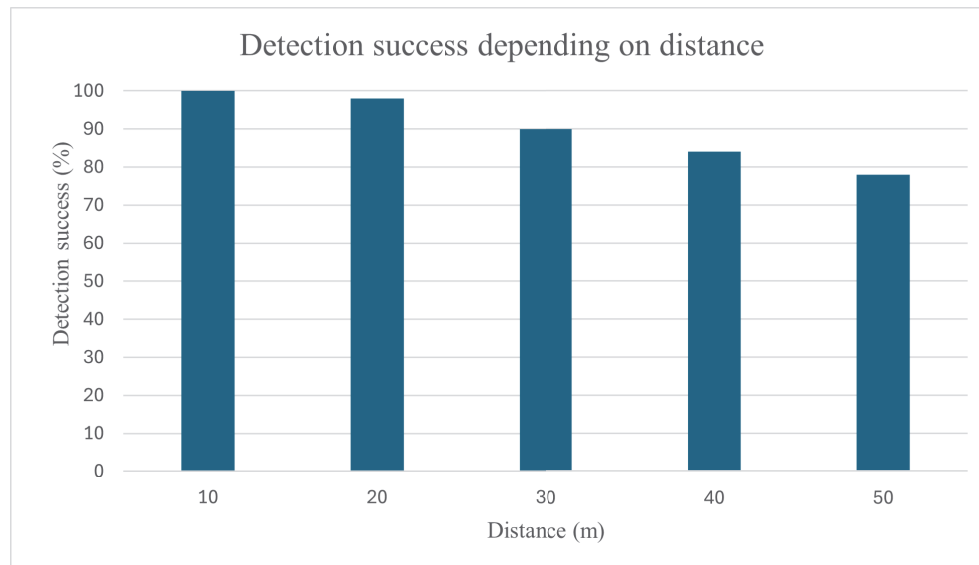


Figure 13. Detection test results using the computer vision module as a function of distance from the UAV.

It is important to note that the results for this parameter, as well as for the others, were obtained from performing different tests in various environments to ensure the reliability of the results. This prevented skewed data that might arise from testing in the same location repeatedly.

It is also important to indicate that the tests were conducted under conditions where the UAV could be detected, avoiding flights in darkness or in scenarios with no visibility where even a human observer would not be able to identify the UAV. Nonetheless, the results were very positive considering the characteristics of the UAVs used, which were small devices that can easily blend in with the environment.

The next aspect to analyze was the effectiveness of the tracking algorithm, evaluating whether the PTZ camera can continuously follow the UAV as it moves through the environment. Once again, the impact of distance on the tracking process was analyzed.

Figure 14 presents the results of the tracking algorithm. As can be observed, the success rate no longer followed the same trend with respect to distance. The best results were obtained at intermediate distances, with the tracking success rate being above 90% for distances of 20 and 30 m and around 83% for 40 m.

The poorest results were obtained at the shortest and longest distances, but both cases have a clear explanation. At shorter distances, the quality of tracking decreases because the angular velocity needed to follow the same movement of the UAV is greater compared to longer distances. A linear displacement of the UAV at close range results in a more significant angular change for the camera, which can make tracking more challenging.

At a distance of 50 m, the tracking effectiveness decreased because the model struggled more to detect the UAV, as seen in Figure 14. At that distance, the detection of the UAV by the camera was not consistent, which affected the tracking algorithm's performance. When

the UAV is not detected, the algorithm lacks the necessary visual reference, reducing the smoothness of the tracking.

Consequently, it can be concluded that the tracking algorithm functions optimally as long as the UAV is detected. The reduction in the tracking rate at greater distances is primarily due to the loss of continuous detection of the UAV, indicating that the smoothness of tracking heavily relies on the model's ability to consistently detect the UAV.

Lastly, in terms of localization, the goal was to verify if the algorithm can accurately send the exact position of the UAV to the PTZ camera. For these tests, the distance was not a factor to consider since, if the algorithm is correctly developed, once the UAV is detected by the fixed cameras, the position will be triangulated independently of the distance.

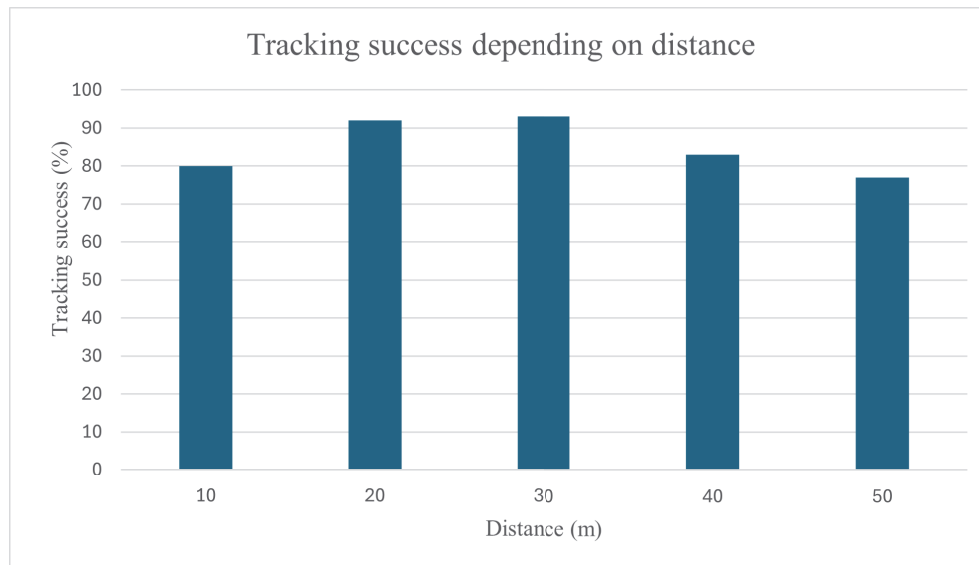


Figure 14. Tracking test results using the computer vision module as a function of distance from the UAV.

The only aspect affected by distance is detection, as the UAV must be detected by at least two cameras to be localized. Thus, at very long distances, it might be more challenging for both cameras to locate the UAV. However, as soon as the UAV is detected simultaneously by both cameras, they will send the position.

In the various tests conducted, it was demonstrated that the localization process works correctly. In all cases where the UAV was detected by two cameras, the PTZ camera moved to the UAV's position and was able to detect it.

Figure 15 provides an example of what was displayed on the NVIDIA Jetson Orin during the localization tests. As shown in Figure 15a, the two fixed cameras simultaneously detected the UAV while the PTZ camera was pointed in a different direction. Figure 15b shows the frame where the PTZ camera rotated to the exact position of the UAV after triangulating its position. The time elapsed between both moments corresponded to the time it took for the camera to complete the rotation, which was less than one second.

It is essential to highlight that, for localization to be effective and accurate, the calibration of the cameras' positions and orientations must be precise. When installing the system, several parameters must be entered into the system before it is launched, including the exact positions of the cameras and the direction in which their lenses are pointing. If these parameters are not accurately input, the localization process will fail because the algorithm will triangulate the UAV's position based on incorrect camera positions.



Figure 15. Vision of the cameras in the localization tests. In (a), it is shown that the fixed cameras simultaneously detected the UAV, while in (b) it is shown that the PTZ camera had already located the UAV after being provided with its exact position.

5. Conclusions

Throughout this article, we have thoroughly explained the two UAV detection methods developed as a potential solution to the growing security concerns stemming from the misuse of UAVs in recent times.

The two developed methods are designed to complement each other, creating a more effective and comprehensive detection system using artificial intelligence algorithms, making this system a novel and intelligent solution. We designed the first detection method based on radio-frequency signals, which are captured and processed by the SDR BladeRF 2.0 micro xa4 (Nuand LLC, Rochester, NY, USA), along with an NVIDIA Jetson Nano Developer Kit (NVIDIA Corporation, Santa Clara, CA, USA) responsible for running the detection algorithm.

The developed algorithm is based on autoencoders, which, as discussed, are unsupervised neural networks frequently used for anomaly detection. In this project, multiple autoencoders were trained, each for a specific frequency band to be analyzed. The input data consisted of 2048 power values that made up the 100 MHz bandwidth of each band.

A validation system was developed based on the autoencoders, in which a set of conditions must be met to confirm the presence of a UAV, aiming to balance detection time with a reduction in false positives.

The second method involves a module comprising four fixed cameras and a PTZ camera, connected via IP addresses to an NVIDIA Jetson Orin Nano responsible for managing them. Each camera detects UAVs using a custom-trained YOLOv10 model with over 60,000 images, showing excellent performance in detecting UAVs even at long distances.

The computer vision module also includes the function of tracking the UAV continuously. An algorithm was developed for tracking, ensuring the PTZ camera keeps the UAV centered in the image. Additionally, a localization system for the UAV was implemented, allowing the PTZ camera to track the UAV once two or more fixed cameras simultaneously detected it, calculating the exact position and sending it to the PTZ for tracking.

Both modules are interconnected and communicate in real time to ensure UAV detection. The developed system also includes a communication protocol using SIA DC-09-2021 (Security Industry Association, Silver Spring, MD, USA), allowing the system to alert a remote monitoring station (CRA) when a UAV is detected, and periodically sending videos that enable them to monitor the UAV's presence and take appropriate actions.

The results indicate that both modules performed well in terms of detection, achieving high success rates even at distances beyond 40 m. The system's primary goals are to detect efficiently while minimizing false positives. This is achieved in the camera module by implementing techniques that filter out a significant portion of false positives. In the SDR module, with an adequate database, the percentage of false positives remains minimal.

Both localization and tracking algorithms functioned appropriately, accurately locating the UAV when it was detected by two cameras, and the tracking algorithm successfully maintained continuous UAV tracking.

Future improvements could mainly focus on expanding the datasets for both modules and upgrading the hardware to more powerful devices. With a more extensive and varied database, YOLOv10 detection could be further improved, enhancing accuracy and reducing false positives. In the SDR module, a more comprehensive initial dataset of non-UAV signals used to train the autoencoders would result in better signal reconstruction when no UAV is present, further improving detection accuracy.

Regarding hardware, this has been the main limitation of the system, as both NVIDIA devices were operating at their maximum capacity. With a more powerful device for the SDR module, more frequency bands could be analyzed, and more samples could be processed per band, significantly improving system performance. In the camera module, a more powerful device would allow for smoother camera operation, better PTZ tracking of the UAV, and the ability to incorporate more fixed cameras.

In this sense, the Research and Development department of the Simetria Group will work on the improvement points indicated above to optimize the performance of the system and broaden the detection spectrum.

Author Contributions: Conceptualization, J.F.M. and J.L.C.; methodology, P.L.-M., F.J.A. and C.A.; software, P.L.-M. and L.G.S.F.; supervision, J.F.M. and C.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Red.Es.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors Christian Abarca, Francisco José Alegre and Jose Luis Calle are employed by Grupo Simetria. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Shakhathreh, H.; Sawalmeh, A.H.; Al-Fuqaha, A.; Dou, Z.; Almaita, E.; Khalil, I. Unmanned Aerial Vehicles (UAVs): A Survey on Civil Applications and Key Research Challenges. *IEEE Access* **2019**, *7*, 48572–48634. [CrossRef]
2. Khosravi, M.; Pishro-Nik, H. Unmanned Aerial Vehicles for Package Delivery and Network Coverage. In Proceedings of the 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring), Antwerp, Belgium, 25–28 May 2020; pp. 1–5.

3. Ju, C.; Son, H.I. Multiple UAV Systems for Agricultural Applications: Control, Implementation, and Evaluation. *Electronics* **2018**, *7*, 162. [CrossRef]
4. Lappa, D.; Fessakis, G.; Karampelas, P. Recognizing the Threats of Drone Surveillance. A Case Study. In Proceedings of the 2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP), Nafplio, Greece, 26–29 June 2022; pp. 1–5.
5. Yaacoub, J.P.; Noura, H.; Salman, O.; Chehab, A. Security analysis of drones systems: Attacks, limitations, and recommendations. *Internet Things* **2020**, *11*, 100218. [CrossRef] [PubMed]
6. Khan, M.A.; Menouar, H.; Eldeeb, A.; Abu-Dayya, A.; Salim, F.D. On the Detection of Unauthorized Drones—Techniques and Future Perspectives: A Review. *IEEE Sens. J.* **2022**, *22*, 11439–11455. [CrossRef]
7. Shi, Z.; Chang, X.; Yang, C.; Wu, Z.; Wu, J. An Acoustic-Based Surveillance System for Amateur Drones Detection and Localization. *IEEE Trans. Veh. Technol.* **2020**, *69*, 2731–2739. [CrossRef]
8. Semkin, V.; Yin, M.; Hu, Y.; Mezzavilla, M.; Rangan, S. Drone Detection and Classification Based on Radar Cross Section Signatures. In Proceedings of the 2020 International Symposium on Antennas and Propagation (ISAP), Osaka, Japan, 25–28 January 2021; pp. 223–224.
9. Ezuma, M.; Ozdemir, O.; Anjinappa, C.K.; Gulzar, W.A.; Guvenc, I. Micro-UAV Detection with a Low-Grazing Angle Millimeter Wave Radar. In Proceedings of the 2019 IEEE Radio and Wireless Symposium (RWS), Orlando, FL, USA, 20–23 January 2019; pp. 1–4.
10. Kaplan, B.; Kahraman, I.; Ekti, A.R.; Yarkan, S.; Görçin, A.; Özdemir, M.K.; Çirpan, H.A. Detection, Identification, and Direction of Arrival Estimation of Drone FHSS Signals with Uniform Linear Antenna Array. *IEEE Access* **2021**, *9*, 152057–152069. [CrossRef]
11. Park, J.; Kim, D.H.; Shin, Y.S.; Lee, S. A comparison of convolutional object detectors for real-time drone tracking using a PTZ camera. In Proceedings of the 2017 17th International Conference on Control, Automation and Systems (ICCAS), Jeju, Republic of Korea, 18–21 October 2017; pp. 696–699.
12. Zou, Z.; Chen, K.; Shi, Z.; Guo, Y.; Ye, J. Object Detection in 20 Years: A Survey. *Proc. IEEE* **2023**, *111*, 257–276. [CrossRef]
13. Bharati, P.; Pramanik, A. Deep learning techniques—R-CNN to mask R-CNN: A survey. *Comput. Intell. Pattern Recognition. Adv. Intell. Syst. Comput.* **2020**, *999*, 657–668.
14. Zhang, C.; Xu, X.; Tu, D. Face Detection Using Improved Faster RCNN. *arXiv* **2018**, arXiv:1802.02142.
15. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In Proceedings of the European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
16. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
17. Aydin, B.; Singha, S. Drone Detection Using YOLOv5. *Eng* **2023**, *4*, 416–433. [CrossRef]
18. Kim, J.-H.; Kim, N.; Won, C.S. High-Speed Drone Detection Based On Yolo-V8. In Proceedings of the ICASSP 2023–2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 4–10 June 2023; pp. 1–2.
19. Zhai, X.; Huang, Z.; Li, T.; Liu, H.; Wang, S. YOLO-Drone: An Optimized YOLOv8 Network for Tiny UAV Object Detection. *Electronics* **2023**, *12*, 3664. [CrossRef]
20. Wang, A.; Chen, H.; Liu, L.; Chen, K.; Lin, Z.; Han, J.; Ding, G. YOLOv10: Real-Time End-to-End Object Detection. *arXiv* **2024**, arXiv:2405.14458.
21. Chen, S.; Yang, H.; Zhang, A.; Chen, B.; Shu, P.; Xiang, J.; Lin, C. UAV Dynamic Tracking Algorithm Based on Deep Learning. In Proceedings of the 2021 3rd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI), Taiyuan, China, 3–5 December 2021; pp. 482–485.
22. Sie, N.J.; Srigrarom, S.; Huang, S. Field Test Validations of Vision-Based Multi-Camera Multi-Drone Tracking and 3D Localizing with Concurrent Camera Pose Estimation. In Proceedings of the 2021 6th International Conference on Control and Robotics Engineering (ICCRE), Beijing, China, 16–18 April 2024; pp. 139–144.
23. Ezuma, M.; Erden, F.; Kumar Anjinappa, C.; Ozdemir, O.; Guvenc, I. Detection and Classification of UAVs Using RF Fingerprints in the Presence of Wi-Fi and Bluetooth Interference. *IEEE Open J. Commun. Soc.* **2020**, *1*, 60–76. [CrossRef]
24. Nemer, I.; Sheltami, T.; Ahmad, I.; Yasar, A.U.-H.; Abdeen, M.A.R. RF-Based UAV Detection and Identification Using Hierarchical Learning Approach. *Sensors* **2021**, *21*, 1947. [CrossRef] [PubMed]
25. AlKhonaini, A.; Sheltami, T.; Mahmoud, A.; Imam, M. UAV Detection Using Reinforcement Learning. *Sensors* **2024**, *24*, 1870. [CrossRef] [PubMed]
26. Xiang, B.; Meng, H.; Zhou, J.; Liu, Y.; He, L.; Luo, Z.; Zeng, C. UAV Detection Research Based on A Low-Cost SDR. In Proceedings of the 2020 International Conference on Microwave and Millimeter Wave Technology (ICMMT), Shanghai, China, 20–23 September 2020; pp. 1–3.
27. Flak, P. Drone Detection Sensor with Continuous 2.4 GHz ISM Band Coverage Based on Cost-Effective SDR Platform. *IEEE Access* **2021**, *9*, 114574–114586. [CrossRef]
28. Jiang, G.; Nan, P.; Zhang, J.; Li, Y.; Li, X. Robust Spatial-Temporal Autoencoder for Unsupervised Anomaly Detection of Unmanned Aerial Vehicle with Flight Data. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 1–14. [CrossRef]
29. Basak, S.; Rajendran, S.; Pollin, S.; Scheers, B. Autoencoder Based Framework for Drone RF Signal Classification and Novelty Detection. In Proceedings of the 2023 25th International Conference on Advanced Communication Technology (ICACT), Pyeongchang, Republic of Korea, 19–22 February 2023; pp. 218–225.

30. Walia, H. Birds vs. Drone Dataset. 2021. Available online: <https://www.kaggle.com/datasets/harshwalia/birds-vs-drone-dataset> (accessed on 9 September 2024).
31. Pawełczyk, M.; Wojtyra, M. Real World Object Detection Dataset for Quadcopter Unmanned Aerial Vehicle (UAV). *IEEE Access* **2020**, *8*, 174394–174409. [CrossRef]
32. Ali-bey, A.; Chaib-draa, B.; Giguère, P. GSV-Cities: Toward appropriate supervised visual place recognition. *Neurocomputing* **2022**, *513*, 194–203. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Machine Learning-Based Structural Health Monitoring Technique for Crack Detection and Localisation Using Bluetooth Strain Gauge Sensor Network

Tahereh Shah Mansouri ^{1,*}, Gennady Lubarsky ², Dewar Finlay ² and James McLaughlin ^{2,*}

¹ School of Engineering and Materials Science, Queen Mary University, Mile End, London E1 4NS, UK

² NIBEC Engineering, University of Ulster, Belfast BT15 1ED, UK; g.lubarsky@ulster.ac.uk (G.L.); d.finlay@ulster.ac.uk (D.F.)

* Correspondence: t.s.mansouri@qmul.ac.uk (T.S.M.); jad.mclaughlin@ulster.ac.uk (J.M.)

Abstract: Within the domain of Structural Health Monitoring (SHM), conventional approaches generally are complicated, destructive, and time-consuming. It also necessitates an extensive array of sensors to effectively evaluate and monitor the structural integrity. In this research work, we present a novel, non-destructive SHM framework based on machine learning (ML) for the accurate detection and localisation of structural cracks. This approach leverages a minimal number of strain gauge sensors linked via Bluetooth Low Energy (BLE) communication. The framework is validated through empirical data collected from 3D carbon fibre-reinforced composites, including three distinct specimens, ranging from crack-free samples to specimens with up to ten cracks of varying lengths and depths. The methodology integrates an analytical examination of the Shewhart chart, Grubbs' test (GT), and hierarchical clustering (HC) algorithm, tailored towards the metrics of fracture measurement and classification. Our novel ML framework allows one to replace exhausting laboratory procedures with a modern and quick mechanism for the material, with unprecedented properties that could provide potential applications in the composites industry.

Keywords: structural health monitoring; machine learning; BLE sensor; Shewhart chart; Grubbs' test; hierarchical clustering; 3D composite

1. Introduction

Structural damage and material ageing in carbon fibre composites significantly reduce the durability, maintainability, and safety of engineering structures such as wind turbine blades, aircraft wings, and boat hydrofoils. Well-timed and appropriate evaluation of structures may prevent catastrophic consequences. Moreover, early detection of deterioration can alleviate the financial burdens associated with maintenance. Therefore, traditional laboratory-based SHM methods have been developed to evaluate damage-sensitive attributes such as stress and strain at critical points, alongside electromechanical impedance techniques and vibration analysis. SHM strategies typically focus on five key aspects: damage detection, localisation, classification, evaluation, and remaining functional life prediction [1].

Regarding the composite material, structural monitoring via laboratory equipment, and workshop techniques, the process can be tedious and non-climate friendly. However, this procedure can be performed via machine learning (ML) with reasonable accuracy. Generally, using ML to predict material properties, is faster and more cost-effective than the traditional laboratory material tests. Moreover, statistical modelling and simulation can provide diverse and precise information on material performance in different situations that would not be effortless to be observed using laboratory experiments. Consequently, ML has gained increasing attention in many SHM applications from fundamental properties

prediction [2,3] to experimental design instruction [4]. Using an algorithmic approach in this field will provide high-throughput recognition and quantifications of material damage.

Numerous studies have shown a successful trend in dealing with material science using machine learning. Kabbani et al. [5] proposed an Artificial Neural Network (ANN) with 90% accuracy in testing unidirectional fibreglass polypropylene material. Liu et al. [6] optimised the same algorithm for the failure detection of woven composite. Zhang et al. [7] offered a high-accurate genetic algorithm to find a link between the mechanical properties and material concentration. Davidson et al. [8] suggested a Support Vector Machine (SVM) algorithm to examine fibre-reinforced composite (FRC) properties and their defections. Chapetti [9,10] applied a short crack and long crack threshold to predict the fatigue propagation of the welded joints. Bang et al. [11] propose a computational modelling framework to predict titanium alloys and aluminium alloy crack growth. The same author introduces a UniGrow model that predicts the fatigue crack growth rate for a given driving force regardless of various R-ratios [12]. Many other ML approaches have been conducted based on crack growth using the stress intensity and strain interval [13–16]. Despite notable advances in machine learning-based SHM, several computational and statistical challenges remain, such as scalability and measurement, imbalance data, unequal size of the sample, incidental endogeneity, and high dimensionality. Moreover, material science databases usually are noticeably below average in size due to tedious and expensive experimental data generation procedures. Scalability issues, for example, affect the framework's application in large-scale, real-time monitoring where high-speed data processing is crucial. Data imbalance, especially given the limited number of crack samples in real-world datasets, may impact model performance, resulting in higher false negatives for minor cracks. Addressing high dimensionality and small sample sizes is essential, as these issues affect the robustness and generalisability of ML models. In this work, feature selection and dimensionality reduction techniques have been applied to mitigate these challenges, allowing the model to maintain a high level of accuracy without becoming computationally burdensome.

The current study proposes a method for accurate crack detection and localisation using a small number of strain gauge sensors and confirms this method's feasibility based on experimental data. Bluetooth strain gauge transmitters have been used to form a sensor network that covers the surface under study. Each transmitter allows the delivery of a single data stream to multiple devices such as tablets and computers as well as the remote configuration of sensors. We demonstrate the potential of the proposed technique via a two- and four-point bending test performed on a composite material specimen without cracks and multiple specimens with different lengths and depths of cracks. The focus will be on the 3D woven composite fracture detection under specific load using ML algorithms. A dataset containing local strains was measured with five strain gauge sensors. The study presents a framework using ML anomaly detection techniques and classification. To achieve this, we review the Shewhart chart, Grubbs' test (GT), and hierarchical clustering (HC) algorithm, in the context of fracture measurement and classification. The proposed method predicts and classifies the strain field and related cracks for all specimens considered in the study. This framework has wide-ranging applications in industries reliant on composite materials, including the aerospace, automotive, and maritime sectors. Its ability to detect cracks with minimal sensor requirements reduces the costs associated with extensive instrumentation. Compared to traditional SHM technologies like non-destructive testing with ultrasound (e.g., Lamb waves), this machine learning-based approach offers self-acting and adaptability in harsh environments, making it particularly valuable for remote or submerged structures, such as aircraft wings, wind turbine blades, and boat hydrofoils. By facilitating early detection, this framework could significantly reduce maintenance costs and mitigate the risks associated with catastrophic failures.

2. Materials and Methods

The fabric structure evaluated in this study is layer-to-layer 3D woven [17] that contains four wefts, six warps, and six binders in the unit cell. To transform this fabric

into a composite material, two resin systems (R1 and R2) are used within three different shapes of 3D preform (beam, square, and rectangle). R1 indicates the Prime 37 epoxy system, whereas R2 is a Pro-set epoxy system. Consolidation via resin transfer moulding (RTM) [18] builds a solid carbon fibre-reinforced matrix. Carbon fibre-reinforced composite was used in the vessel's interior. The material proved to have a better resistance in marine environments [19]; the smooth external surface of the components is within the limit of dirt and subsequently, they are less susceptible. The material is also expected to be used as a composite boat hydrofoil that may experience different forces due to velocity changes, pressure, and the fluid's viscosity. Generally, lift and drag forces could be exerted on the hydrofoil submerged in a stream flow. The hydrofoil surface is subjected to drag force by the fluid due to uneven pressures and viscous friction [20]. Normally, fluid movement over the hydrofoil creates two different zones of fluid velocity due to the angle of attack and the camber. The differences between the velocity of the fluid particles at the higher and lower surface of the hydrofoil generates a low-pressure zone at the upper surface and a high-pressure zone at the lower surface. Uneven pressure distribution between the upper and lower surface of the hydrofoil generates the lift force [21].

In this work, three samples cut from the 3D layer-to-layer composite material discussed above (Figure 1a) were used. Specimen one was a beam (Bs); a two-point bending test was performed using a uniaxial experimental setting to determine the strain of the specimens (Figure 1b). When a beam experiences a bending moment, it will change its shape, and internal stresses will be developed (Figure 1c). The second specimen was a rectangle sample (Rs) of a 3D composite with a dimension of 160×340 mm that was instrumented with five strain gauge sensors and subjected to controlled loads. The strain samples were recorded as a function of fracture development when eleven cracks developed between 3, 4, and 5. The third specimen was a square sample (Ss) with a dimension of 160×160 mm that was equipped with another five strain gauges when ten surface cracks developed between sensors three and four.

In previous discussions, it was noted that various forces from water flow cause dynamic stress on hydrofoil structures. BLE technology [22] represents a new generation of Bluetooth-based strain gauge sensors that offer several advantages for monitoring hydrofoil structures. BLE sensors can be easily mounted on the inner surfaces of the hydrofoil without the need for complex wiring, which makes it a suitable choice for underwater scenarios. In the current study, the sensor system is a compact module that houses the BLE in place while in use (Figure 2). The module is equipped with a bridging board, a BLE transmitter board B24-SSBX-A, and a strain gauge sensor. Additionally, it is designed that the coin cell battery is to be replaced without affecting the system's functionality.

The BLE sensor network was chosen due to its low power consumption and reliability in transmitting strain data over extended periods, making it suitable for continuous SHM. Each sensor module, along with its coin cell battery, provides up to 5 years of operation under regular load cycles. To prolong battery life, an adaptive power-saving algorithm is implemented that minimises energy usage when strain changes are within nominal ranges. This feature enhances the feasibility of deploying BLE sensors in remote or hard-to-access structural locations where frequent maintenance is challenging.

A compact and low-profile enclosure was designed to house the electronics module and a battery. The enclosure aligns and supports the spring-loaded connector, ensuring consistent contact with the strain gauge sensors during use. The modular configuration and its enclosure allow for easy disassembly and the replacement of parts, maintenance, battery changes, and adaptation to various locations.

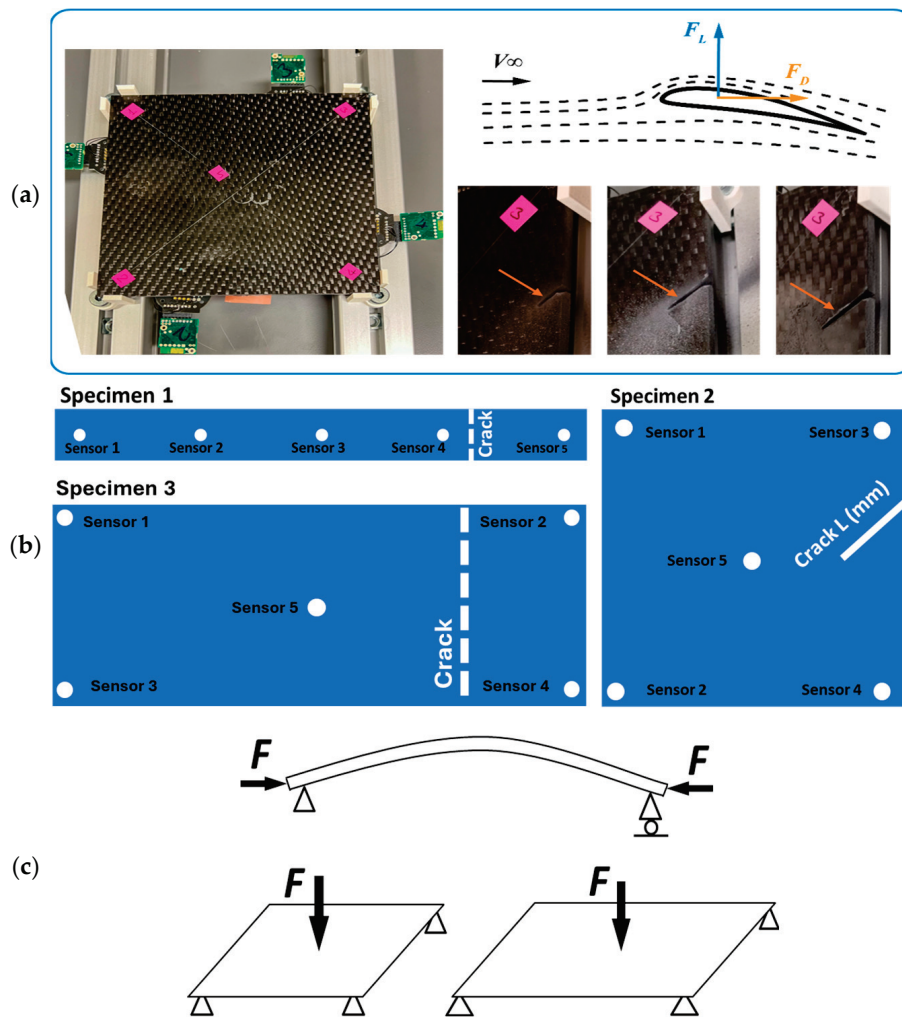


Figure 1. (a) Ss sample cut from 3D layer-to-layer composite material and its developed length crack (red arrows show increasing cracks)—lift and drag force's position on hydrofoil is shown in upper-right. (b) Schematic of the three-specimen geometry and a crack position. (c) Load configuration during the experiment. Please note the crack location identified via defined algorithms in framework in next sections.

The transmitting board and the strain gauge sensor are connected via a bridging board. To provide the highest sensitivity to strain changes, strain gauges were integrated in a half-bridge Wheatstone configuration where two active strain gauges are utilised in lieu of two of the bridge's four resistors. The configuration is able to measure any variations in resistance, which can be brought on by any small strain changes or gauge deformation. The sensor outputs of Bs with various lengths and surface cracks developed are shown in Figure 3, where the average strain value for each sensor was determined.

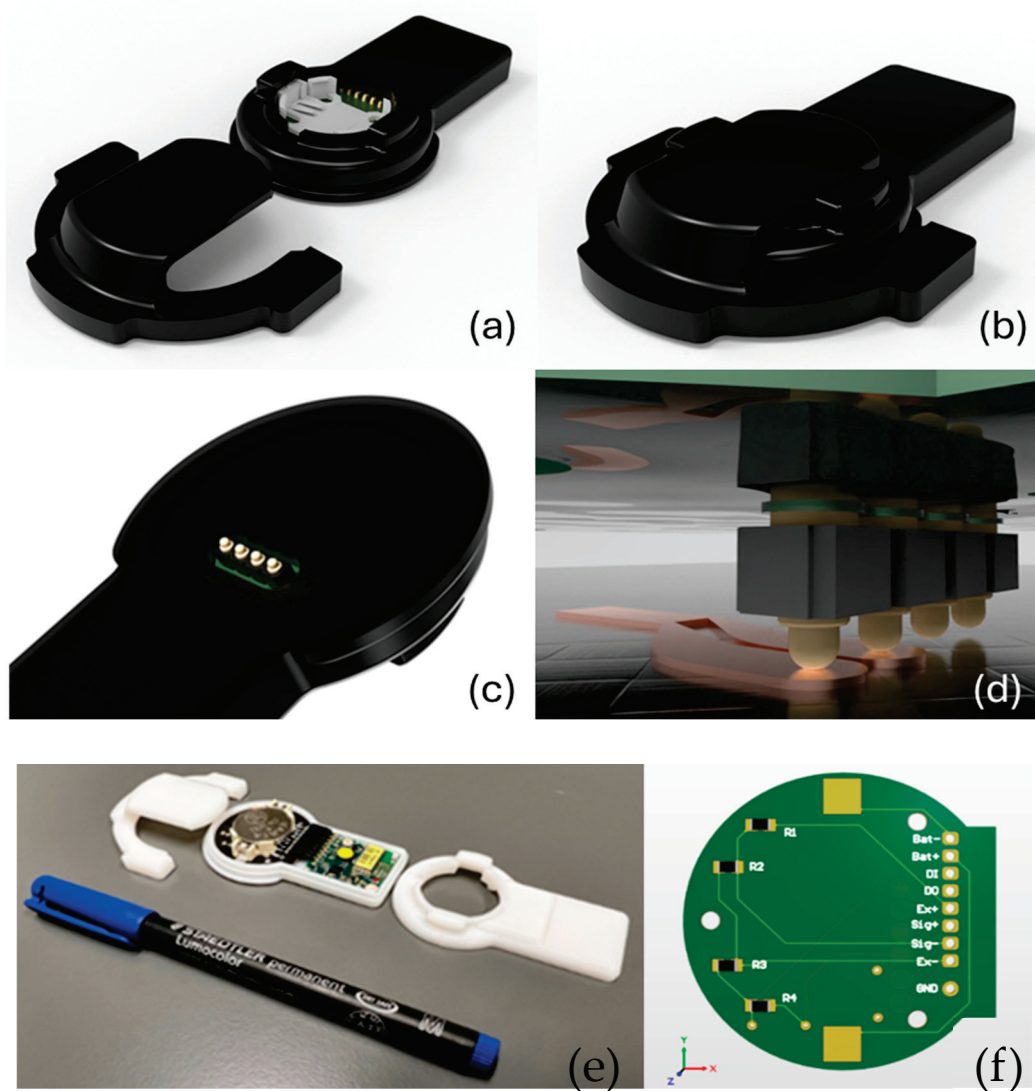


Figure 2. (a) Removable electronic enclosure attaches to the strain gauge electrodes using a holder permanently fixed on the sample surface. The node can be moved to a new location. (b,c) The enclosure accommodates a wireless module (BLE), a coin cell battery, supporting electronics, and a spring-loaded connector. (d) Spring-loaded interconnecting header in contact with copper pads deposited on the sample surface and connected to the strain gauge sensor. (e) Electronic module that accommodates BLE sensors, compared with a pen to evaluate the size. (f) The front side of the bridging board has four resistors. It is a connector between the strain gauge sensors and the transmitting board.

The BLE sensor value deviates from its mean around the crack zone when the fracture increments. The initial laboratory experimental results show that the sensors with a continuous increase in the strain indicate the presence of a crack. Those sensors with a greater distance may show a fluctuation in strain outputs (Figure 3b). In this study, the two above concepts are the basis for crack localisation, which will be discussed in the next sections.

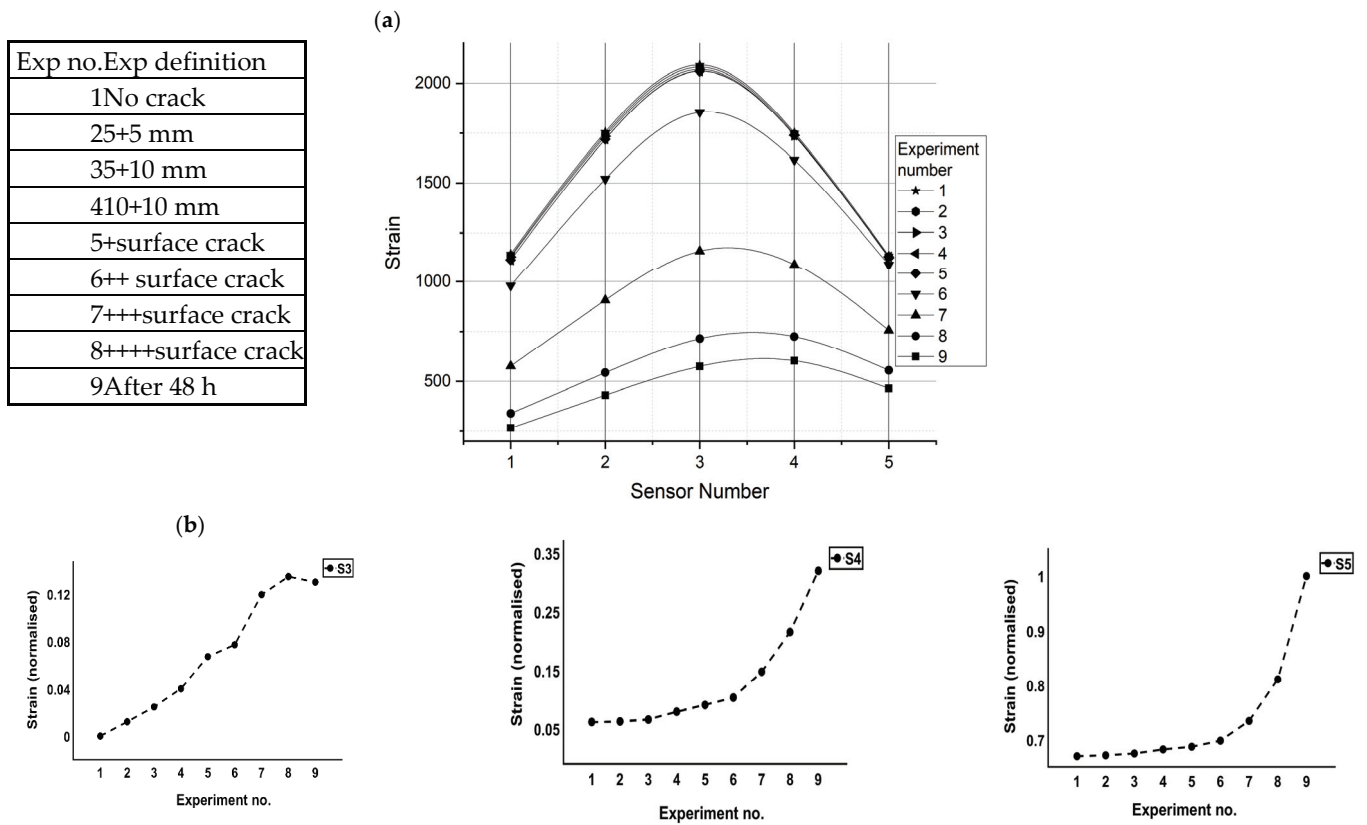


Figure 3. (a) Measured strain vs. sensor number when crack developed through Bs—Figure illustrates the sensor outputs for various experimental configurations labelled as Exp 1–9. These configurations correspond to specific load conditions and crack positions as follows: Exp 1 represents a scenario with minimal strain applied, Exp 2–4 involve moderate strain with an incipient crack in both sides, Exp 5–8 indicate higher strain with multiple surface cracks, and Exp 9 denotes maximum load conditions under which significant crack propagation was observed after 48 h. Each experiment’s unique setup is detailed to provide insights into the strain responses across varying crack severities. (b) The average value of each BLE sensor when cracks developed through Bs [equivalent experiment shown in (a)].

3. Computational Technique and Framework

The study framework proposes three different algorithmic and statistical techniques to identify cracks and classify them. All procedures are time independent. Figure 4 provides a visual summary of the SHM framework, showing the data flow from initial sensor readings to final crack localisation and classification cycle. This flowchart clarifies the sequence of algorithmic steps, demonstrating how each component contributes to the overall accuracy of crack detection.

This section is organised into three units. Initially, the Shewhart control chart [23] was employed to detect anomalous sensors with a non-random condition [24], where healthy sample behaviour is also exhibited. Section two defines the hierarchical clustering algorithm to classify and train a model based on observations of incongruous sensor output. Furthermore, Grubbs’ test [25] with the corresponding threshold values receives the classified cracks at the next stage. The threshold values then define a boundary between the tolerance interval [26] and destructive fracture population.

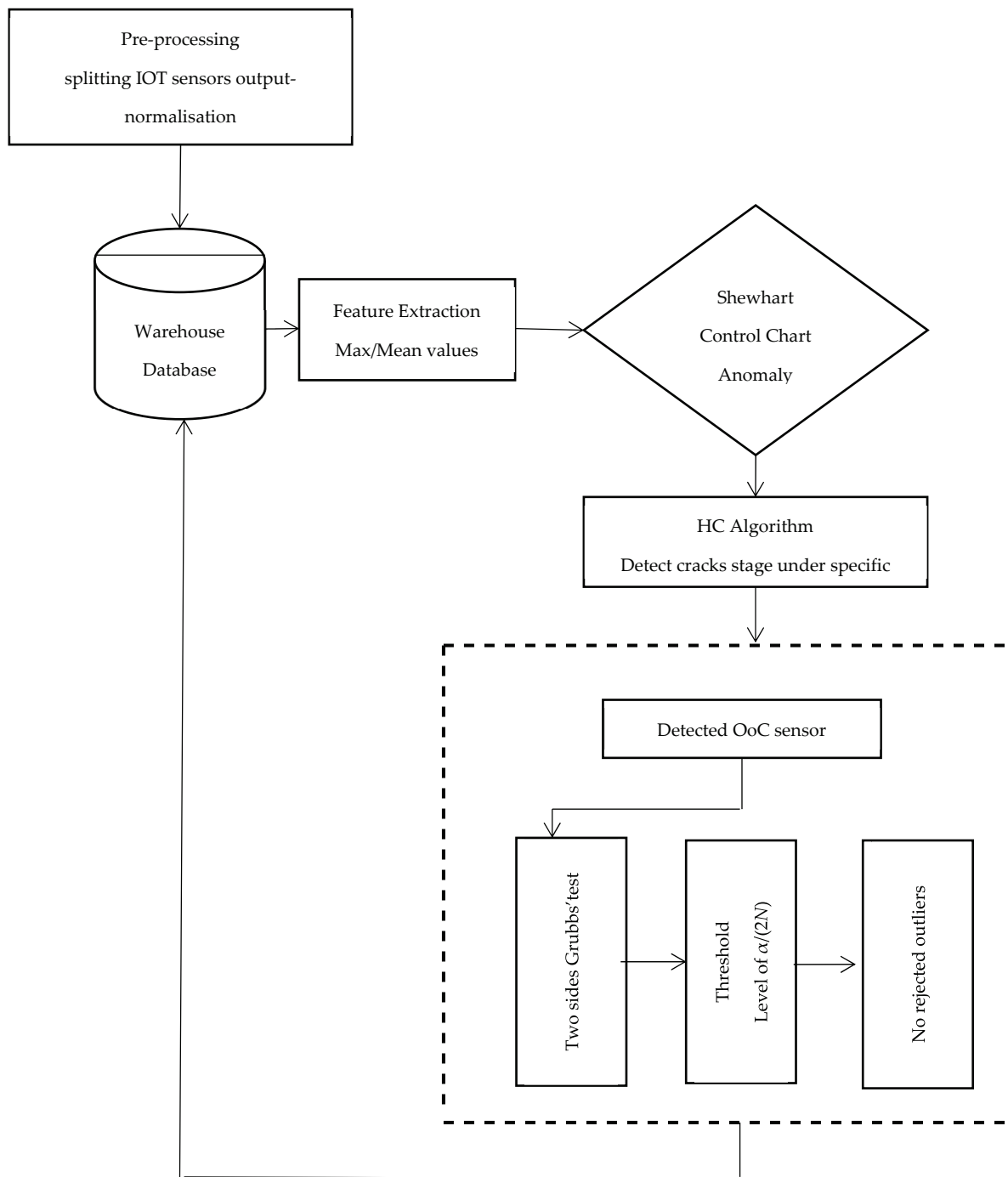


Figure 4. Machine learning-based SHM framework.

3.1. Sensor Control Limit

Shewhart control charts are well-known tools for identifying potential Out-of-Control (OoC) states [27]. The basic principle of the chart is to detect instances where there is an unnatural deviation between parameters [28]. In general, the chart retains a centreline as the mean of the process, which can be controlled with an upper control limit (UCL) and a lower control limit (LCL) set at $\pm 3\sigma$, respectively [29]. The conventional sigma (σ) value can be a moving average, range measurement, standard deviation (SD), or other attributes as required by the system to measure [24,30]. The process is considered in a control or a stable phase if the measurement value fluctuates between the LCL and UCL, and out of control when a single observation is beyond the designated boundary [31]. In the current

study, the standard deviation has been used to determine how far each BLE sensor's output has been located from its mean value when the cracks are developed [32]. Based on our observation of multiple datasets, the sensor's standard deviation closer to the crack exceeds the boundary of the UCL/LCL. This feature helps to localise fractures. Figure 5 compares five sensors with loads (healthy sample) and the same sensors when cracks developed through the specimen (damaged sample).

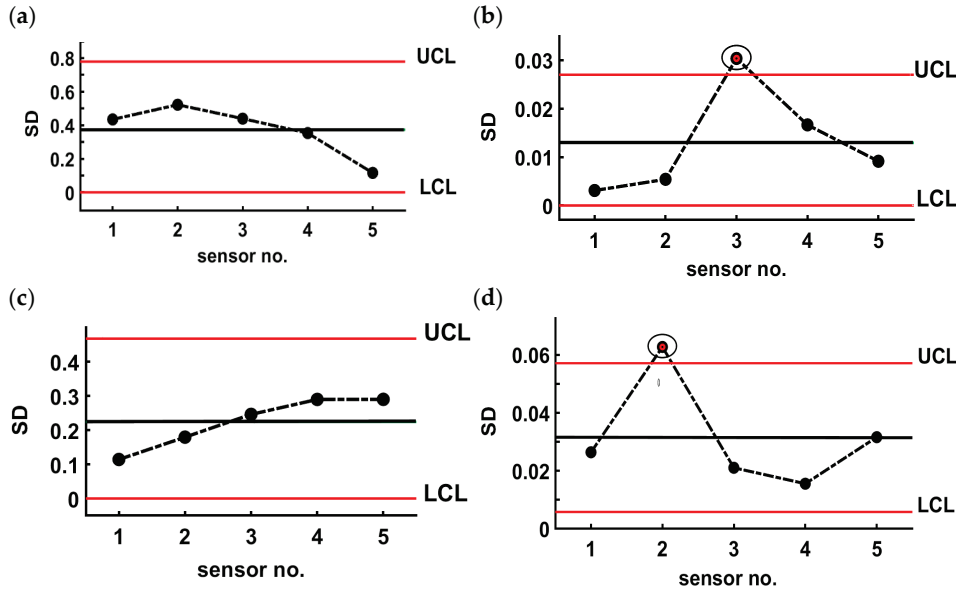


Figure 5. (a) Comparing five sensors' outputs for Ss before crack development (sensors under load but in normal phase). (b) Anomalous sensors after crack development through Ss. (c,d) Comparing five sensor outputs for Rs before and after crack development, respectively. As discussed above, a sensor close to cracks usually appears out of UCL/LCL interval.

The Shewhart charts can be implemented on all the sensor channels simultaneously. If the output of each sensor is x , and it is normally distributed with a mean of m and variance of SD , the confidence interval will be as follows:

$$C = \frac{\bar{x} - m}{SD_{\bar{x}}} \quad (1)$$

where $SD_{\bar{x}} = SD/\sqrt{n}$, n = number of fractures, and $\bar{x}$ is the average value of x .

3.2. Agglomerative Hierarchical Cluster Algorithm

The primary purpose of using unsupervised learning algorithms is to identify and categorise fractures to assess crack levels. This method generates significant clusters at different levels of a hierarchy. Data obtained from deviant BLE sensors will be fed into the HC algorithm for further classification. To group similar BLE outputs, a measurement is required. This measurement can be the distances relative to each other [33–36], meaning that pairs of sensor observations with closer distances could be grouped in similar clusters. In this study, the Chebyshev distance [37] between data points is calculated. This process continues until the intervals between all the observations are calculated and the entire BLE output is organised into various clusters. The visual representation of hierarchical clusters is commonly referred to as a dendrogram [38], as shown in Figure 6. The ultrametric topology illustrates how the fracture observations are developed. All clusters eventually link to each other to generate larger clusters until all observations in the original dataset are linked together [34,35]. The height of each link indicates the distance between the two points that are joined, where a shorter link between two objects represents a closer distance (and similarity), and a higher link shows dissimilarity between objects [33].

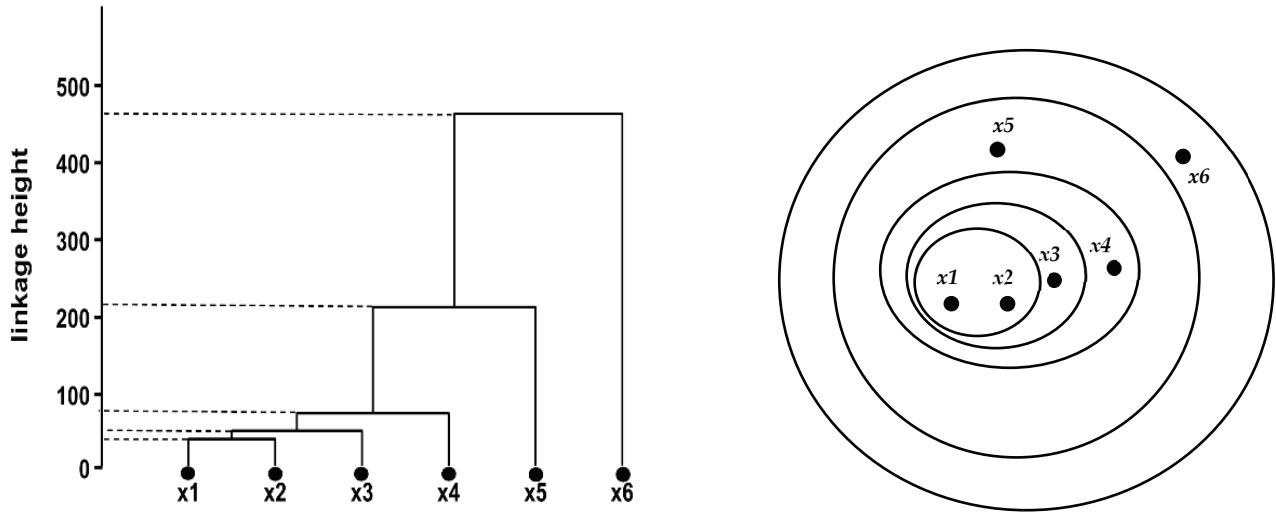


Figure 6. A dendrogram of six initial surface cracks (**left**), and equivalent Venn diagram (**right**). The dendrogram shows classificatory relationships between samples. The height of links increases when there is a progression in crack depth.

3.3. Grubbs' Test

It was perceived that the sensor close to fracture shows an increase in strain measurements when cracks start to form. To prevent deeper cracks, a univariate statistical test [39] is used to identify outliers, such as crevasses, by setting a threshold condition. Grubbs' test (G-test) is employed to detect outliers from the previously categorised high-consequence HC observations by setting a threshold α [40]. The test iterates and rejects outliers in each cycle until no exceptional data is detected (Figure 7). The underlying theory of Grubbs' test is that the observations are normally distributed [41]. The G-statistic test is determined for OoC-BLE sensor output as follows:

$$G - test = \frac{\max_{(i=1,\dots,N)} |x - \bar{x}|}{SD} \quad (2)$$

where x is the sensor outputs include healthy (without damage) observation and various depth crack samples, $\bar{x}$ is the mean value of the observation, and SD is the standard deviation of the data. Any result obtained from the G-test and greater than the G critical value will be eliminated in each cycle. The critical value can be determined as follows:

$$G_{crit} = \frac{(N - 1)t_{crit}}{\sqrt{N(N - 2 + t_{crit}^2)}} \quad (3)$$

The t_{crit} is a significant level that can be defined via $\frac{\alpha}{(2N)}$ where α is a condition threshold. The conventional value for α is 0.5, i.e., 95% of observations are at a confident level and consider non-outliers [42]. However, in the current study, to avoid potential hazards at an early stage, another threshold (10%) is evaluated. A reasonable concern is that a composite material test could contain more than one outlier that can be called a median hazard observation. GT threshold increments could assist in avoiding system failure in the recognition of unusual observations prior to main fractures.

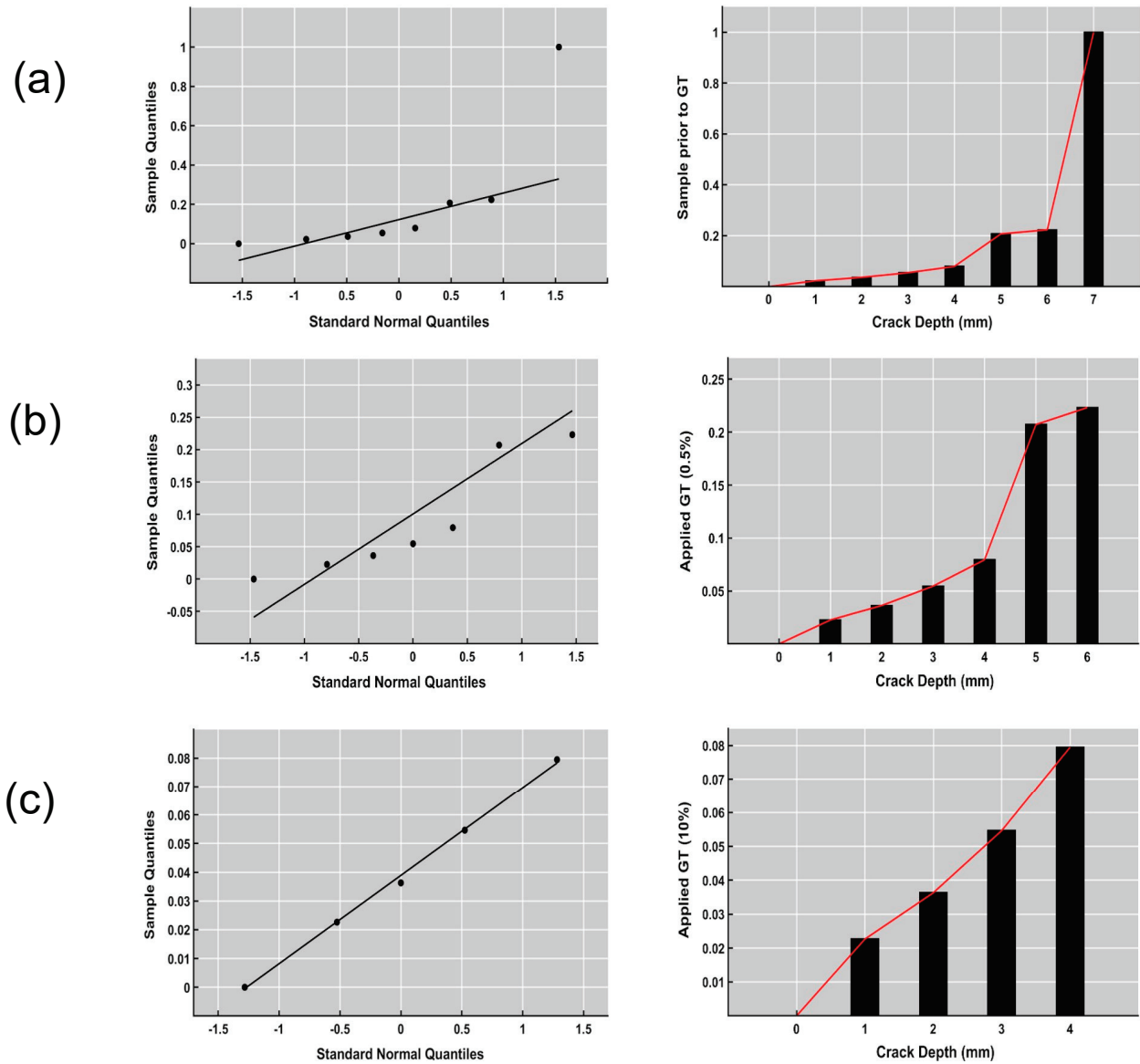


Figure 7. (a) Normal probability plot of developed cracks for Rs before GT application. (b,c) The same samples after applying GT threshold (α); (0.05 % and 10%) with 90% and 99.5 % confidence, respectively. Please note: GT applied on two other forms of samples and similar results achieved.

4. Discussion and Results

In Section 3.1, we discussed the use of the Shewhart control chart to pinpoint damage. This chart utilises average values and standard deviation to detect deviations in sensor output from the mean or change point. Figure 8 illustrates the comparison between the healthy and damaged sample mean intervals, which are determined by the lower and upper control limits (LCL and UCL). The Shewhart chart is used to monitor the sensor output within these boundaries, and any deviation beyond this interval could indicate an abnormal situation. Therefore, if a sensor's distribution value shifts from m to $m + \Delta\sigma$, as shown in Figure 8, it would be considered irregular.

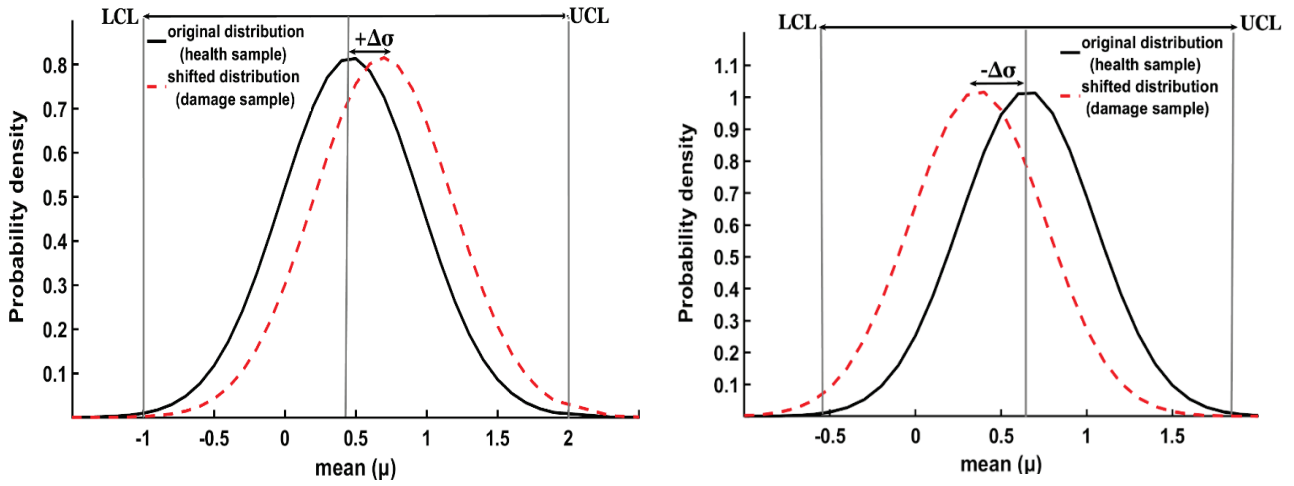


Figure 8. Probability density belonging to sensor 3 (square) on the left, and sensor 2 (rectangle) on the right, before and after crack development. For square, the process shift (Δ) is positive, while the rectangle shows negative variance. This is because the IoT sensors were located on the surface of the square, while the same sensors were located beneath the surface for the rectangle specimen, resulting in negative recordings.

The LCL and UCL are completely flexible and obey the sensor outputs. Each sensor provides various mean and SD values, which could result in non-identical Δ and control limit assumptions. However, studies [31,43] proved that neither the number of observations nor a stringent control limit is effective on the OoC finding.

The Shewhart control chart's capability to detect the crack location is verified by comparing PDF coverage against crack growths for the three experimental samples (Figure 9). Cracks with a similar size will be grouped into logical ranges or bins. The predicted PDF curve outcomes show that the curve fit correlated well with the crack growth at different levels for surface and length cracks. Based on the bin coverage shown in Figure 9, the prediction accuracy varied between 93% and 97% for Cr, Cs, and Cb, respectively.

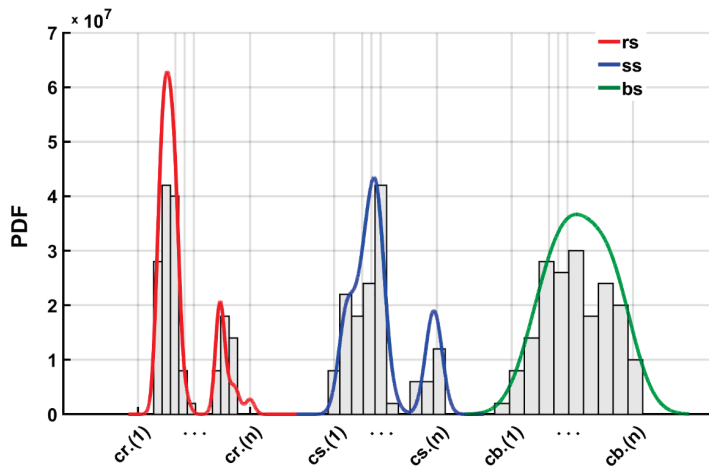


Figure 9. Probability Density Function (PDF) of the prediction error for R, S, and B samples. The bars represent cracks within each sample, and the curve shows PDF prediction. A better Shewhart accuracy in predicting the crack location is indicated by greater coverage of the bars by the curve.

The crack information from irregular sensors gains access to the H-cluster algorithm for classification. The H-clustering algorithm discovered strongly related fractures at a local level to determine similarity (discussed in Section 3.2). Following this, the algorithm provides an insight into the data by splitting the cracks into classes of cracks, such that

cracks in a group are more identical to each other than to cracks in other groups in terms of depth and length.

The number of clusters is unknown in advance by the algorithm, which makes it a suitable solution for partitioning our objects via a non-heuristic search [44]. As a result, the border of fractures could be evaluated, as shown in Figure 10a,b. As the height of the link increased with the developing cracks, the measurement for the HC algorithm was $L(\mu)$, which originated from σ . Therefore, the different σ values shown in the figure are derived from $\Delta\sigma = L(\mu)_{n+1} - L(\mu)_n$. The technique is applied successfully on a variety of datasets [45–48].

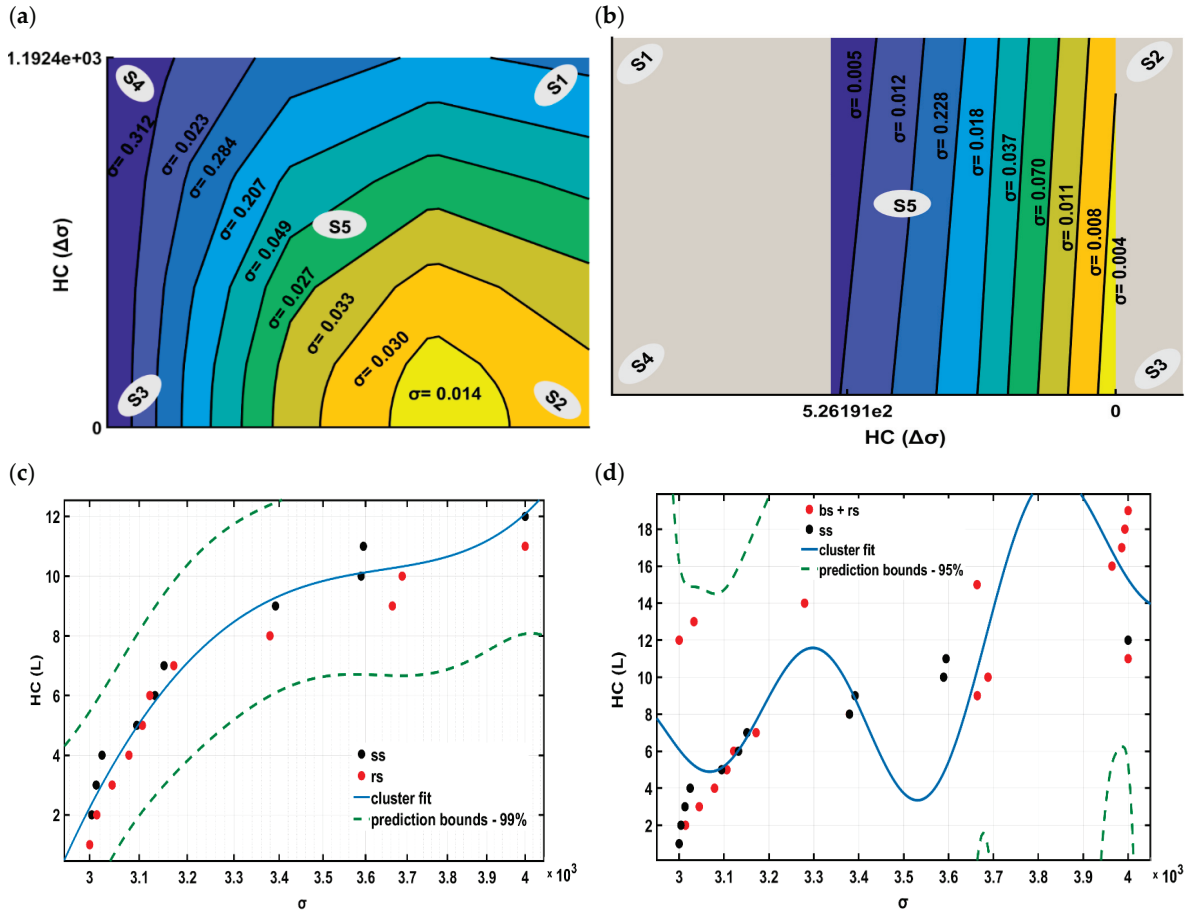


Figure 10. (a) The difference between link height for 'ss' using HC algorithm during crack growth. (b) Repetition of exp a. for 'rs'. (c) Blue line shows a polynomial model adjusted to the cluster link using 'ss' with $R^2 = \%97$ to evaluate the prediction accuracy. Each dot represents a link between developed cracks, and all can be located within 99% of prediction bounds. (d) Repetition of exp c. for 'bs' and 'rs'. $R^2 = \%91$ for prediction bounds of 95%.

To verify the HC model prediction, a polynomial function from the order of two was adjusted to cluster links. For the model fit to 'ss', $R^2 = \%97$, and for the one fit to 'bs+rs', an $R^2 = \%91$ can be achieved. The complete understanding of 'bs' behaviour is slightly complex as compared to the 'ss' and 'rs' sets. It seems the cracks are propagated unevenly and broad.

However, the different sizes of specimens and the shorter development of cracks could be calculated as a potential basis for irregularity in the conduct of crack growth, which is outside the scope of this study.

In terms of Grubbs' test and predicting deeper cracks, three sets of specimens with various levels of damage are compared. For a fair comparison, the same values of $\alpha = 0.5$ and $\alpha = 10$ that were discussed above are applied (Figure 11). The G-test forms a probability

that assumes samples belong to the core population, and any sample outside of this interval will be rejected by the test. The probability distribution will be based on the differences in sample means, where the boundary could be the sample's SD [49].

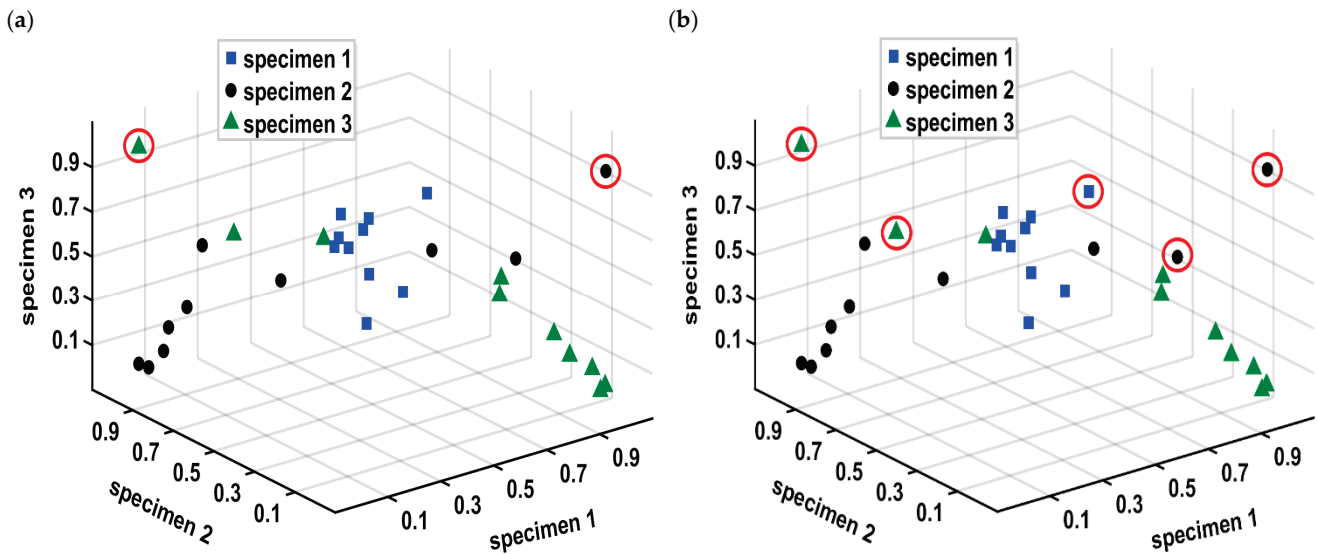


Figure 11. (a) Applying G-test with $\alpha = 0.5$. on anomalous sensors of three different shape specimens (specimen 1: square, specimen 2: rectangle, and specimen 3: beam). Selected α will define one outlier for specimens 2 and 3 and null for specimen 1. (b) Equivalent experiment with $\alpha = 10$; here, five outliers (shown inside the red circle) were selected via G-test from various specimens.

To demonstrate the effectiveness of the proposed ML-based SHM framework, its accuracy was compared with conventional laboratory-based SHM procedures [50,51], which typically rely on extensive instrumentation and time-intensive manual assessments. Traditional methods require numerous sensors placed at critical structural points and are often destructive or semi-destructive. In contrast, the current framework achieves comparable accuracy with a reduced sensor count, enabling quicker, non-destructive assessments. It achieves an overall 95% accuracy, as indicated by the result of three different algorithms, with only five sensors, confirming a reduction in complexity and cost while maintaining reliable crack detection.

5. Conclusions

In this study, we proposed an ML framework to localise and classify the cracks, and furthermore identify the final cracks. We used measured strain as the sole attribute for this framework. For this reason, three different forms of specimens were monitored with a Bluetooth-based strain gauge sensor to measure the strain. All three experiments show that crack development will result in sensor output deviating from its mean value. The sensors with constant strain increments are closer to the fracture location. We use this feature to discover the crack location between the sensors using the Shewhart control chart. The use of this protocol has proved that those sensors close to fracture show SD violation from the LCL and UCL. The capability of the Shewhart control chart was verified using a PDF curve with accuracy values $>93\%$. The hierarchical clustering algorithm also assists one in detecting crack development using its links. Subsequently, a polynomial function was employed to evaluate HC patterns swapping samples from three specimens. To this extent, prediction accuracy $>91\%$ for each model was achieved, which will enhance the speed and accuracy of material discovery. In addition, a G-test was applied to avoid meeting destructive fracture population and predict fatigue cracks. With limited data in hand, two thresholds were defined to distinguish the confidence interval.

Currently, this framework shows the increment in the surface crack depth and length but is not very specific in the dimensions. Future work will focus on improving model

precision in classifying the crack depth and length, as current estimates are approximate. Moreover, incorporating other sensor types (e.g., ultrasonic or thermographic sensors) could broaden the framework's applicability to diverse structural materials.

Author Contributions: Conceptualisation, T.S.M. and G.L.; methodology, T.S.M. and G.L.; validation, T.S.M. and G.L.; formal analysis, T.S.M. and G.L.; investigation, T.S.M. and G.L.; data curation, T.S.M.; writing—original draft preparation, T.S.M.; writing—review and editing, T.S.M. and G.L.; supervision, J.M. and D.F.; project administration, J.M. and D.F.; funding acquisition, J.M. and D.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the UK Research and Innovation's Strength in Places Fund to the Belfast Maritime Consortium and led by Artemis Technologies. This research (including laboratory work and data analysis) was completed at Ulster University, NIBEC.

Data Availability Statement: The data cannot be made publicly available upon publication because they are not available in a format that is sufficiently accessible by other researchers. The data that support the findings of this study are available upon reasonable requests from the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

Nomenclature

3D Composite	Three dimensional Composite
ANN	Artificial Neural Network
BLE	Bluetooth Low Energy
Bs	Beam sample
Cr, Cs, Cb	Crack levels for different sample types (Rectangle, Square, Beam)
G-test	Grubb's test
HC	Hierarchical Cluster
L	link (cluster)
LCL	Lower Control Limit
ML	Machine Learning
OoC	Out of Control
PDF	Probability Density Function
Rs	Rectangle sample
RTM	Resin Transfer Moulding
SD	Standard Deviation
Ss	Square sample
SVM	Support Vector Machine
UCL	Upper Control Limit
$\Delta\sigma$	strain range
μ	average value (mean)
σ	strain

References

- Chen, H.-P.; Ni, Y.-Q. Introduction to Structural Health Monitoring. In *Structural Health Monitoring of Large Civil Engineering Structures*; John Wiley & Sons Ltd.: Hoboken, NJ, USA, 2018; pp. 1–14. [CrossRef]
- Gisario, A.; Mehrpouya, M.; Rahimzadeh, A.; De Bartolomeis, A.; Barletta, M. Prediction Model for Determining the Optimum Operational Parameters in Laser Forming of Fiber Reinforced Composites. *Adv. Manuf.* **2020**, *8*, 242–251. [CrossRef]
- Legrain, F.; Carrete, J.; van Rookeghem, A.; Curtarolo, S.; Mingo, N. How Chemical Composition Alone Can Predict Vibrational Free Energies and Entropies of Solids. *Chem. Mater.* **2017**, *29*, 6220–6227. [CrossRef]
- Raccuglia, P.; Elbert, K.C.; Adler, P.D.F.; Falk, C.; Wenny, M.B.; Mollo, A.; Zeller, M.; Friedler, S.A.; Schrier, J.; Norquist, A.J. Machine-Learning-Assisted Materials Discovery Using Failed Experiments. *Nature* **2016**, *533*, 73–76. [CrossRef] [PubMed]
- Kabbani, M.S.; El Kadi, H.A. Predicting the Effect of Cooling Rate on the Mechanical Properties of Glass Fiber–Polypropylene Composites Using Artificial Neural Networks. *J. Thermoplast. Compos. Mater.* **2018**, *32*, 1268–1281. [CrossRef]
- Liu, X.; Gasco, F.; Goodsell, J.; Yu, W. Initial Failure Strength Prediction of Woven Composites Using a New Yarn Failure Criterion Constructed by Deep Learning. *Compos. Struct.* **2019**, *230*, 111505. [CrossRef]
- Zhang, S.L.; Zhang, Z.X.; Xin, Z.X.; Pal, K.; Kim, J.K. Prediction of Mechanical Properties of Polypropylene/Waste Ground Rubber Tire Powder Treated by Bitumen Composites via Uniform Design and Artificial Neural Networks. *Mater. Des.* **2010**, *31*, 1900–1905. [CrossRef]

8. Davidson, P.; Waas, A.M. Probabilistic Defect Analysis of Fiber Reinforced Composites Using Kriging and Support Vector Machine Based Surrogates. *Compos. Struct.* **2018**, *195*, 186–198. [CrossRef]
9. Chapetti, M.D. Fatigue Propagation Threshold of Short Cracks Under Constant Amplitude Loading. *Int. J. Fatigue* **2003**, *25*, 1319–1326. [CrossRef]
10. Chapetti, M.D.; Jaureguizar, L.F. Fatigue Behavior Prediction of Welded Joints by Using an Integrated Fracture Mechanics Approach. *Int. J. Fatigue* **2012**, *43*, 43–53. [CrossRef]
11. Bang, D.; Ince, A.; Noban, M. Modeling Approach for a Unified Crack Growth Model in Short and Long Fatigue Crack Regimes. *Int. J. Fatigue* **2019**, *128*, 105182. [CrossRef]
12. Bang, D.; Ince, A.; Tang, L. A modification of Unigrow 2-parameter Driving Force Model for Short Fatigue Crack Growth. *Fatigue Fract. Eng. Mater. Struct.* **2018**, *42*, 45–60. [CrossRef]
13. Newman, J.C.; Annigeri, B.S. Fatigue-Life Prediction Method Based on Small-Crack Theory in an Engine Material. *J. Eng. Gas Turbines Power* **2011**, *134*, 032501. [CrossRef]
14. Lam, T.; Topper, T.; Conle, F. Derivation of Crack Closure and Crack Growth Rate Data from Effective-Strain Fatigue Life Data for Fracture Mechanics Fatigue Life Predictions. *Int. J. Fatigue* **1998**, *20*, 703–710. [CrossRef]
15. Noroozi, A.; Glinka, G.; Lambert, S. A Study of the Stress Ratio Effects on Fatigue Crack Growth Using the Unified Two-Parameter Fatigue Crack Growth Driving Force. *Int. J. Fatigue* **2007**, *29*, 1616–1633. [CrossRef]
16. Santus, C.; Taylor, D. Physically Short Crack Propagation in Metals during High Cycle Fatigue. *Int. J. Fatigue* **2009**, *31*, 1356–1365. [CrossRef]
17. Perera, Y.S.; Muwanwella, R.M.H.W.; Fernando, P.R.; Fernando, S.K.; Jayawardana, T.S.S. Evolution of 3D Weaving and 3D Woven Fabric Structures. *Fash. Text.* **2021**, *8*, 11. [CrossRef]
18. Dai, D.; Fan, M. Wood Fibres as Reinforcements in Natural Fibre Composites: Structure, Properties, Processing and Applications. In *Natural Fibre Composites*; Woodhead Publishing: Sawston, UK, 2014; pp. 3–65. [CrossRef]
19. Rubino, F.; Nisticò, A.; Tucci, F.; Carlone, P. Marine Application of Fiber Reinforced Composites: A Review. *J. Mar. Sci. Eng.* **2020**, *8*, 26. [CrossRef]
20. Manwell, J.F.; McGowan, J.G.; Rogers, A.L. *Wind Energy Explained: Theory, Design and Application*; John Wiley & Sons: Chichester, UK, 2006; Volume 30, pp. 169–170. [CrossRef]
21. Muratoglu, A.; Yuce, M.I. Performance Analysis of Hydrokinetic Turbine Blade Sections. *J. Int. Soc. Sports Nutr.* **2015**, *2*, 1–10.
22. Furst, J.; Chen, K.; Kim, H.-S.; Bonnet, P. Evaluating Bluetooth Low Energy for IoT. In Proceedings of the 2018 IEEE Workshop on Benchmarking Cyber-Physical Networks and Systems (CPSBench), Porto, Portugal, 10–13 April 2018; pp. 1–6.
23. Vermaat, M.B.; Ion, R.A.; Does, R.J.M.M.; Klaassen, C.A.J. A Comparison of Shewhart Individuals Control Charts Based on Normal, Non-parametric, and Extreme-value Theory. *Qual. Reliab. Eng. Int.* **2003**, *19*, 337–353. [CrossRef]
24. Roes, K.C.B.; Does, R.J.M.M. Shewhart-Type Charts in Nonstandard Situations. *Technometrics* **1995**, *37*, 15–24. [CrossRef]
25. Aslam, M. Introducing Grubbs’s Test for Detecting Outliers Under Neutrosophic Statistics—An Application to Medical Data. *J. King Saud Univ. Sci.* **2020**, *32*, 2696–2700. [CrossRef]
26. Young, D.S. Statistical Tolerance Regions: Theory, Applications, and Computation. *Technometrics* **2010**, *52*, 143–144.
27. Zhou, C.; Zou, C.; Zhang, Y.; Wang, Z. Nonparametric Control Chart Based on Change-Point Model. *Stat. Pap.* **2007**, *50*, 13–28. [CrossRef]
28. Raji, I.A.; Lee, M.H.; Riaz, M.; Abujija, M.R.; Abbas, N. Outliers Detection Models in Shewhart Control Charts; An Application in Photolithography: A Semiconductor Manufacturing Industry. *Mathematics* **2020**, *8*, 857. [CrossRef]
29. Koutras, M.V.; Bersimis, S.; Maravelakis, P.E. Statistical Process Control using Shewhart Control Charts with Supplementary Runs Rules. *Methodol. Comput. Appl. Probab.* **2007**, *9*, 207–224. [CrossRef]
30. Ryan, T.P. *Statistical Methods for Quality Improvement*; John Wiley & Sons: New York, NY, USA, 2011.
31. Albers, W.; Kallenberg, W.C. Estimation in Shewhart Control Charts: Effects and Corrections. *Metrika* **2004**, *59*, 207–234. [CrossRef]
32. Steiner, S.H.; Geyer, P.L.; Wesolowsky, G.O. Shewhart Control Charts to Detect Mean and Standard Deviation Shifts Based on Grouped Data. *Qual. Reliab. Eng. Int.* **1996**, *12*, 345–353. [CrossRef]
33. Murtagh, F.; Contreras, P. Algorithms for Hierarchical Clustering: An Overview. *WIREs Data Min. Knowl. Discov.* **2011**, *2*, 86–97. [CrossRef]
34. Nielsen, F. Hierarchical Clustering. In *Introduction to HPC with MPI for Data Science*; Springer: Cham, Switzerland, 2016; pp. 195–211. [CrossRef]
35. Murtagh, F. Hierarchical Clustering. In *International Encyclopedia of Statistical Science*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 633–635. [CrossRef]
36. Zhang, T.; Ramakrishnan, R.; Livny, M. Birch: An Efficient Data Clustering Method for Very Large Databases. In Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data, Montreal, QC, Canada, 4–6 June 1996; pp. 103–114. [CrossRef]
37. Zhu, Y.; Zhu, L.; Xiao, Y.; Zha, M.; Hu, T.; Xiang, M. Research on Hierarchical Clustering Leach Protocol Optimization Algorithm Based on Chebyshev Distance. *J. Phys. Conf. Ser.* **2023**, *2456*, 012040. [CrossRef]
38. Tibshirani, R.; Walther, G.; Hastie, T. Estimating the Number of Clusters in a Data Set via the Gap Statistic. *J. R. Stat. Soc. Ser. B* **2001**, *63*, 411–423. [CrossRef]

39. Kachigan, S.K. *Statistical Analysis: An Interdisciplinary Introduction to Univariate & Multivariate Methods*; Radius Press: New York, NY, USA, 1986. [CrossRef]
40. Grubbs, F.E. Sample Criteria for Testing Outlying Observations. *Ann. Math. Stat.* **1950**, *21*, 27–58. [CrossRef]
41. Stefansky, W. Rejecting Outliers in Factorial Designs. *Technometrics* **1972**, *14*, 469–479. [CrossRef]
42. Cohn, T.A.; England, J.F.; Berenbrock, C.E.; Mason, R.R.; Stedinger, J.R.; Lamontagne, J.R. A Generalized Grubbs-Beck Test Statistic for Detecting Multiple Potentially Influential Low Outliers in Flood Series. *Water Resour. Res.* **2013**, *49*, 5047–5058. [CrossRef]
43. Tsai, T.-R.; Lin, J.-J.; Wu, S.-J.; Lin, H.-C. On Estimating Control Limits of $\bar{X}$ Chart When the Number of Subgroups is Small. *Int. J. Adv. Manuf. Technol.* **2005**, *26*, 1312–1316. [CrossRef]
44. Blumenfeld, D. *Operations Research Calculations Handbook*; Taylor & Francis: London, UK, 2009.
45. Fung, B.C.; Wang, K.; Ester, M. Hierarchical Document Clustering Using Frequent Itemsets. In Proceedings of the 2003 SIAM International Conference on Data Mining, San Francisco, CA, USA, 1–3 May 2003.
46. Malik, H.H.; Kender, J.R. High Quality, Efficient Hierarchical Document Clustering Using Closed Interesting Itemsets. In Proceedings of the Sixth International Conference on Data Mining (ICDM'06), Hong Kong, China, 8–22 December 2006; pp. 991–996.
47. Xiong, H.; Steinbach, M.; Tan, P.-N.; Kumar, V. HICAP: Hierarchical Clustering with Pattern Preservation. In Proceedings of the 2004 SIAM International Conference on Data Mining, Lake Buena Vista, FL, USA, 22–24 April 2004.
48. Beil, F.; Ester, M.; Xu, X. Frequent Term-Based Text Clustering. In Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, AB, Canada, 23–26 July 2002; pp. 436–442.
49. Grubbs, F.E. Procedures for Detecting Outlying Observations in Samples. *Technometrics* **1969**, *11*, 1–21. [CrossRef]
50. Anand, Y.; Dutta, V. Testing of Composites: A Review. *Int. J. Adv. Mater. Manuf. Charact.* **2013**, *3*, 359–363. [CrossRef]
51. Saba, N.; Jawaaid, M.; Sultan, M.T.H. An overview of mechanical and physical testing of composite materials. In *Mechanical and Physical Testing of Biocomposites, Fibre-Reinforced Composites and Hybrid Composites*; Woodhead Publishing: Sawston, UK, 2019; pp. 1–12. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Energy-Efficient Online Path Planning for Internet of Drones Using Reinforcement Learning

Zainab AlMania ^{1,2}, Tarek Sheltami ^{1,*}, Gamil Ahmed ¹, Ashraf Mahmoud ¹ and Abdulaziz Barnawi ³

¹ Computer Engineering Department, Interdisciplinary Research Center of Smart Mobility and Logistics, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia; g202102670@kfupm.edu.sa (Z.A.); gamil.ahmed@kfupm.edu.sa (G.A.); ashraf@kfupm.edu.sa (A.M.)

² Computing Department, Applied College, Imam Abdulrahman Bin Faisal University, Dammam 34212, Saudi Arabia

³ Computer Engineering Department, Interdisciplinary Research Center for Intelligent Secure Systems, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia; barnawi@kfupm.edu.sa

* Correspondence: tarek@kfupm.edu.sa; Tel.: +966-13-860-4678

Abstract: Unmanned aerial vehicles (UAVs) have recently been applied in several contexts due to their flexibility, mobility, and fast deployment. One of the essential aspects of multi-UAV systems is path planning, which autonomously determines paths for drones from starting points to destination points. However, UAVs face many obstacles in their routes, potentially causing loss or damage. Several heuristic approaches have been investigated to address collision avoidance. These approaches are generally applied in static environments where the environment is known in advance and paths are generated offline, making them unsuitable for unknown or dynamic environments. Additionally, limited flight times due to battery constraints pose another challenge in multi-UAV path planning. Reinforcement learning (RL) emerges as a promising candidate to generate collision-free paths for drones in dynamic environments due to its adaptability and generalization capabilities. In this study, we propose a framework to provide a novel solution for multi-UAV path planning in a 3D dynamic environment. The improved particle swarm optimization with reinforcement learning (IPSO-RL) framework is designed to tackle the multi-UAV path planning problem in a fully distributed and reactive manner. The framework integrates IPSO with deep RL to provide the drone with additional feedback and guidance to operate more sustainably. This integration incorporates a unique reward system that can adapt to various environments. Simulations demonstrate the effectiveness of the IPSO-RL approach, showing superior results in terms of collision avoidance, path length, and energy efficiency compared to other benchmarks. The results also illustrate that the proposed IPSO-RL framework can acquire a feasible and effective route successfully with minimum energy consumption in complicated environments.

Keywords: IoDs; IPSO; reinforcement learning; Q-learning; actor-critic

1. Introduction

Recently, the use of unmanned aerial vehicles (UAVs), known as drones, has been receiving extensive attention for several applications. The flexibility, agility, and mobility features of UAVs allow them to achieve tasks that are difficult for humans such as monitoring, surveillance, medical supplies, military, telecommunications, rescue operations, and delivery [1–3]. Furthermore, as technology advances, drones gain new capabilities and become increasingly autonomous. Due to the goals of heterogeneity, communication technology, and hardware/software constraints, the internet of drones (IoDs) needs to collaborate to accomplish tasks that exceed their individual capabilities [4].

The efficient and safe navigation of drones to their intended destinations is hindered by various types of obstacles, both static and dynamic. The presence of these obstacles has a significant impact on the performance of drones, particularly in densely populated

areas. Therefore, successfully completing missions in such environments poses a critical challenge, imposing the development of an intelligent and efficient technique that allows UAVs to modify their flight paths in real-time to ensure collision-free journeys towards their destinations. This is considered as a challenging task, especially with dense swarms and various environment constraints [5,6].

Numerous approaches have been employed to address UAV path planning, including meta-heuristic algorithms such as the genetic algorithm (GA) [7], artificial bee colony (ABC) [8], and particle swarm optimization (PSO) [9]. However, these approaches often suffer from slow convergence and the problem of getting stuck in local optima. In response to these limitations, researchers have made efforts to enhance the optimality and convergence speed of some meta-heuristics, such as PSO, resulting in the development of improved particle swarm optimization (IPSO) [10]. Nonetheless, these approaches are primarily effective in static environments and are not suitable for unknown or dynamic environments.

Recently, the use of machine learning, including reinforcement learning, alongside other methods has enhanced the automation of drones and transferred them qualitatively to achieve different goals and objectives. Learning through interactions with the environment forms the core concept on which reinforcement learning is based, and this is precisely what drones require for autonomous navigation [11–13].

Despite the massive number of studies addressing IoDs path planning in the literature, most of these studies focus on addressing path planning in static environments where the information of the environment is known in advance and the paths are generated offline. In dynamic environments, information about the environment is not known in advance, and the path is generated online in a drone field of view. To complete these tasks, IoDs require a sophisticated technique for autonomous flight in the area filled with static and dynamic obstacles, which is still challenging [14]. Additionally, in increasingly complex mission situations with high drone densities, drones face new challenges as they become closer to obstacles or each other. In such situations, IoDs path planning becomes a significant aspect of autonomous navigation [15,16].

Another challenge is the inherent constraints that limit the complete realization of drones' capabilities [17]. The primary limitation is the lifespan of onboard batteries, which restricts the duration of their flights. Consequently, many applications fall short of achieving their maximum potential. Furthermore, in scenarios involving extensive flight paths across large areas, minimizing the energy consumption of UAVs is essential to establish a safe route and successfully reach their destinations.

To address these challenges, we propose an energy-efficient online static and dynamic collision avoidance framework for IoD formation to tackle the multi-UAV path planning problem in a fully distributed reactive manner.

The proposed approach incorporates environmental information and introduces a novel reward structure that can be applied to various environments. The framework combines IPSO guidance obtained through an IPSO path planning algorithm with local RL-based planning. The local RL-based planner considers the surrounding environmental information, including static and dynamic obstacles, as well as other UAVs, to generate appropriate actions. The objective is to avoid potential collisions while following the fixed IPSO guidance. By integrating IPSO guidance and local RL-based planning, the framework facilitates end-to-end learning in dynamic environments. The local planner utilizes spatial and temporal information within a local area, such as the UAV's field of view, to make real-time decisions and navigate effectively.

The proposed IPSO guidance approach ensures scalability by enabling the UAV to learn and navigate using a fixed-sized learning model, even in large-scale environments. It provides a consistent reference path for the UAV while allowing the local RL-based planner to adapt and make responsive decisions based on the current environmental conditions. To achieve this, a reward function will be designed to combine and consider the IPSO guidance in calculating a reward.

The rest of this work is organized as follows: in Section 2, previous related works will be reviewed, covering different techniques that are used in path planning and collision avoidance. Section 3 explains the proposed energy-efficient online reinforcement learning for IoD path-planning description, thoroughly explaining the methods used and the proposed solution. Section 4 explains the simulation results and discussion. Finally, the conclusion are presented in Section 5.

2. Related Work

Drones can be used in path planning, monitoring and coordinating remote objects such as ground vehicles using different algorithms and techniques. These methods can be classified into non-reinforcement learning methods and reinforcement learning methods.

2.1. Non-Reinforcement Learning Methods

Non-reinforcement learning methods include different algorithms, some of them inspired by biological systems and natural processes.

In ref. [5], an algorithm that helps the swarm avoid obstacles was proposed. These obstacles can be dynamic or static, with altered sizes in a limited detection area in order to reduce energy consumption. The algorithm utilizes a gradient-based approach for quick and rapid convergence. Also, the proposed solution allows the drones to fly vertically while hovering or turning around. The results demonstrate the accuracy and efficiency of the proposed algorithm in high-collision-risk scenarios within dense environments with obstacle speeds of up to 10 m/sec.

The authors of ref. [10] generated a 3D path to assist a swarm of drones in reaching their destination without colliding with other drones or encountering terrain obstacles. They used chaos map logic to start the PSO. Furthermore, optimized adaptive mutation is used to balance the global and local search. Dormant particles are replaced by new active particles to push the arrangement toward a global ideal. Monte Carlo simulations are performed, and the outcomes are compared with standard PSO and recent work.

In [18], A path planning scheme is proposed for guiding the UAVs from the start position to their final position. Offline planning creates a predetermined path that the drone traverses while avoiding obstacles. Combining path control and collision avoidance control results in this optimality. Additionally, each of these tasks can be completed on its own, and the movement approach was created by merging them all together. To construct the track control rule, they calculate the path errors from the geometric relationship between the predetermined path and the UAVs. The optimal avoidance maneuver is determined by utilizing angles and zone risk in the collision avoidance strategy.

Population-based evolutionary algorithms have made substantial strides recently [19,20]. To avoid collisions, these strategies rely on computational techniques. Numerous optimization strategies are therefore suggested to deal with this complexity of computation. Many researchers have been drawn to using these techniques for drone path planning like PSO, ABC, GA, and ant colony optimization (ACO) [21]. Based on kinematics, limitations of navigation, and collision probability, an innovative approach, the fast geometric avoidance algorithm (FGA), is suggested in ref. [22] by joining the start time from geometric and critical avoidance. Diverse threat levels were assigned to avoid obstacles sequentially based on time, reducing the computational time compared to other works.

2.2. Reinforcement Learning (RL) Methods

Reinforcement learning (RL) is also one way to make the drones reach their destination using optimal paths while avoiding obstacles and conserving energy [11,12,23,24]. However, one of the recent twin-delayed deep deterministic policy gradients (TD3) [11] suffers from energy issues.

To solve the problem of multiple drone collaborative path planning in complicated situations with various constraints, the authors of ref. [25] proposed a multi-mode collaboration based on RL with a multi-objective PSO algorithm (MCMOPSO-RL) to find the

best paths and deal with constraints simultaneously. Another study proposed a hybrid path planning algorithm based on RL and PSO to address the path planning problem of intelligent driving vehicles [26]. The authors of ref. [27] used a modified Q-learning algorithm and a state action reward state action algorithm (SARSA) to plan the trajectories of drones. The modified algorithm showed better results, reducing the path length and computation time compared with SARSA. In terms of obstacle avoidance, the proposed method uses a deep Q-learning network in a 3D environment using AirSim. Another study suggests using reinforcement learning to address the collision avoidance problem and find an optimal path (forward and backward) to solve the no-return problem [28].

A hybrid path planning algorithm that uses deep RL for local planning and an anytime graph-based path planning algorithm was applied to real-time planning for an autonomous UAV [29]. The authors of ref. [30] suggested an independent policy algorithm with C51 dueling using deep Q-learning (C51-Duel-IP). A swarm of UAVs can use this algorithm to find the best path in a communication-denial environment.

In ref. [31], a deep reinforcement learning algorithm consisting of hindsight experience replay (HER) and soft actor–critic (SAC) was proposed, called SACHER. This algorithm works to find an optimal trajectory with obstacle avoidance. HER was used to enhance learning and the sample efficiency of SAC. The proposed solution showed a better result compared with DDPG and SAC, with a 66.08% success rate.

A system for path planning of UAV swarms using Q-learning with artificial neural networks (ANNs) was proposed [32]. The goal is to provide full coverage of an area with static obstacles without requiring prior information or goals apart from the provided map. The experiment was conducted using different sizes of maps, varying numbers of UAVs, and different obstacle configurations. The results indicated that as the number of UAVs increased, the system achieved solutions with fewer movements.

Another approach [33] utilizes a laser to scan the data around the UAV, providing a simulation closer to the real-world environment. This approach uses the soft actor–critic (SAC) algorithm to plan the path and navigate around obstacles to reach the destination. Then, a sampling-based mechanism is used to create a reference point that the UAVs can follow. These points are updated continuously. The results show how effective the technique is in the training process and task performance, with an 80% success rate in achieving the goal points.

RL can also be a great solution in hazardous weather conditions. A challenge faced by UAVs while navigating in complicated environments, particularly urban areas with dynamic factors such as wind zones, was addressed by proposing a multi-objective navigation reinforcement learning algorithm (MONRL) [34]. This algorithm enables UAVs to navigate and avoid obstacles in unknown environments with varying wind conditions. The algorithm uses DRL with a memory architecture to build the navigation policies that improve the UAV's path while decreasing the impact of winds.

An algorithm was proposed to tackle the time-constrained path planning challenge for UAVs in complex and unknown environments [35]. It involves utilizing off-policy RL along with an enhanced exploration mechanism known as the improved exploration mechanism (IEM), which also incorporates prioritized experience replay (PER) and curiosity-driven exploration to optimize the path planning process for UAVs. The research aims to enhance the effectiveness and efficiency of path planning for UAVs in complicated environments. The simulations showed that the proposed method, which included the improved exploration mechanism (IEM), was superior to the original off-policy RL algorithms. The algorithm was able to decrease the time needed for path planning and successfully rescue all the individuals who were trapped.

A study [36] was proposed to dynamically choose the number of iterations for Q-Learning to overcome the problems that occur due to poorly chosen iteration numbers, such as taking a very long time or not giving an optimal path. This approach was compared to Q-Learning with unchanged iteration numbers, as well as other algorithms like A*, rapid-exploring random tree, and PSO. The proposed method demonstrated effectiveness

and reliability in online path planning for unknown complex environments. Experiments were conducted in various environments to evaluate different factors like memory usage, time, path length, CPU utilization, and completeness. The results showed a remarkable performance for the proposed algorithm, comparable to or better than A*, while keeping computational requirements low and ensuring maximum completeness.

Another study [37] combined deep Q-network (DQN) with artificial potential field (APF) in a method called B-APFDQN to improve the efficiency and effectiveness of path planning. They used B-spline to make the path smoother. Q-Learning with a shortest distance prioritization policy was suggested for efficiently planning the paths for drones while avoiding both static and dynamic obstacles [38].

The performance of the proposed algorithm was compared to other path planning algorithms such as Dijkstra, A*, and SARSA in terms of path length and learning time. The proposed algorithm demonstrated improved performance, even when more obstacles were added. An enhanced dueling double deep Q network (D3QN) with a prioritized experience replay mechanism was utilized in ref. [39], and the algorithm's network structure was designed. This work aimed to plan the path of UAVs in dynamic scenarios. They trained the drone in the static scenario and then retrained it in the dynamic scenario. Table 1 shows a summary of the reviewed RL works.

The previously reviewed works suggest solving some aspects of the problems that drones encounter during their flight while disregarding others. Some of the simulations were conducted in known environments or with static objects, but UAVs usually face more challenging situations. Some research has been conducted in two-dimensional environments, neglecting 3D environments. Other approaches proposed RL for UAV path planning but neglected dynamic obstacles and IoD formation. More importantly, most of the literature neglects the energy consumption of drones during missions. In this work, we will address these issues.

Table 1. Summary of RL works reviewed.

Ref.	Date	Problem Addressed	Algorithm Used	Limitations
[24]	2020	Collision avoidance	Two-stage RL	Not accurate when scenario is changed
[11]	2021	Path planning and energy consumption	Advanced TD3	Static environment
[23]	2021	Path planning and data harvesting	DDQNs	Static environment
[29]	2021	Path planning and obstacle avoidance	iADA*, DQN, and DDPG	Obstacles move in constant directions/speeds
[12]	2022	Energy consumption, coverage, and connectivity	SBG-AC	Did not address path planning
[25]	2022	Path planning	MCMOPSO-R L	Static environment
[28]	2022	Path planning and collision avoidance	convex-TSP	Static environment
[27]	2023	Path planning and obstacle avoidance	SARSA and Q-learning	Avoiding dynamic obstacles if moving at very low speeds

Table 1. Cont.

Ref.	Date	Problem Addressed	Algorithm Used	Limitations
[30]	2023	Path planning and obstacle avoidance	C51-Duel-IP	Did not address power consumption
[31]	2023	Path planning and obstacle avoidance	SACHER	Static environment
[37]	2023	Path planning	DQN and APF	Deals with static obstacles
[38]	2023	Path planning and obstacle avoidance	Q-learning and SDP	Considered only 2D environments
[32]	2024	Path planning	Q-learning and ANN	UAVs flying at the same height
[33]	2024	Path planning	Soft actor–critic (SAC) algorithm	-
[34]	2024	Path planning in urban wind zones	(MONRL)	Works only with specific conditions
[35]	2024	Path planning	RAPER, Q-learning and IEM.	-
[36]	2024	Path planning and obstacle avoidance	Q-learning	Deals with static obstacles.
[39]	2024	Path planning and obstacle avoidance	D3QN	It is work on 2D

3. Energy-Efficient Online Reinforcement Learning for IoD Path-Planning Description

This section provides a comprehensive overview of the system components and their respective roles. Thus, IPSO, Q-learning and actor–critic DRL will be presented. Additionally, the integration between these components within our proposed solution will be detailed.

3.1. Energy Model for IoD Path Planning

The energy model is essential in the path planning process. Despite their increasing popularity and numerous advantages, drones still face inherent limitations that restrict their full potential. The most significant limitation is the short lifespan of their on-board batteries, which is regarded as their primary drawback [40]. When we incorporate the energy model into the path planning and navigation algorithms, the framework can optimize the drone's trajectory to minimize overall energy consumption. By encouraging the drone to follow the IPSO path and providing higher rewards as it approaches the goal, the process enhances energy efficiency. The reward structure motivates the drone to take the most efficient route, minimizing unnecessary movements and thus conserving energy. Thus, when the drone navigates the path, at each step, the energy is calculated, and an amount of reward will be added depending on that energy. To do so, we use the following formula:

$$R_t = r_t + (0.5(1/(1 + E))), \quad (1)$$

where R_t is the cumulative reward, r_t is the current reward, and E is the energy required to reach the goal or the destination. In this work, we apply the energy model applied in ref. [41] to accurately determine the energy consumption during a drone's flight mission. The model explains various maneuvering activities described by the energy model. The

energy consumption for vertical flight over a distance Δh and velocity v_{climb} is calculated as shown in Equation (2):

$$E_{climb}(\Delta h) = P_{climb} \frac{\Delta h}{v_{climb}}. \quad (2)$$

The energy required for vertical descent over a distance Δh and velocity v_{desc} is given by Equation (3):

$$E_{desc}(\Delta h) = P_{desc} \frac{\Delta h}{v_{desc}}. \quad (3)$$

The energy required to hover in place from time t_1 to time t_2 is given by Equation (4):

$$E_{hover} = P_{hover}(t_2 - t_1). \quad (4)$$

A horizontal flight as a function of velocity is given by Equation (5):

$$E_{horiz} = P_v \frac{d}{v}, \quad (5)$$

where d is the horizontal distance and v is the horizontal speed. Measurements were conducted to determine the power and time required for the rotations, where the angular speed was assumed to be ω_{turn} (2.1 rad/s) and a constant power of P_{turn} (225 W/s) was maintained throughout the rotation. Based on these assumptions, the energy needed to cover an angle θ were calculated as shown in Equation (6):

$$E_{turn} = P_{turn} \frac{\Delta \theta}{\omega_{turn}} \quad (6)$$

Therefore, the overall energy can be written in Equation (7) as follows:

$$E = E_{climb} + E_{desc} + E_{hover} + E_{horiz} + E_{turn} \quad (7)$$

3.2. Planning the Path Using Improved Particle Swarm Optimization

IPSO is a modified version of the standard PSO meta-heuristic approach. The aim is to generate the UAVs' paths that ensure the shortest and obstacle-free path, which has a significant effect on the mission of the drones in terms of time and energy consumption. It should be noted that in our proposed solution, the IPSO algorithm is applied offline in static environments. Thus, the paths generated by IPSO avoid only static obstacles during the mission. The IPSO incorporates three main enhancements to standard PSO that will be explained along with other components in the next subsections. The main enhancements include improvements in the initialization stage of swarm particles using a chaos map, improvements in the updating strategy using an adaptive mutation strategy, and inactive particle replacement.

3.2.1. Initialization Using the Chaos Map

To improve the initial distribution of particles in the searching space, the IPSO algorithm operates a chaos map logic for the initialization of the particle population. This initialization technique helps improve the convergence speed and quality of the solutions. Chaos-based particles utilize the logistic map to obtain the initialization formation. The simplest logistic map presented in [10] is applied to IPSO using Equation (8):

$$X_{(n+1)} = f(\mu X_n) = \mu X_n(1 - X_n) \quad (8)$$

where $n = 0, 1, 2, 3, \dots$ and X represent the chaos variable, while μ is a predetermined constant called the bifurcation coefficient.

To effectively demonstrate the superiority of logistic map initialization over random initialization, visual representations of 10,000-time iterations in MATLAB are presented in Figure 1 [42]. The graphical analysis clearly indicates that the distribution of the logistic

chaos map exhibits greater uniformity compared to that of the rand function. This enhanced uniformity contributes to a broader spectrum of potential flight paths for UAVs.

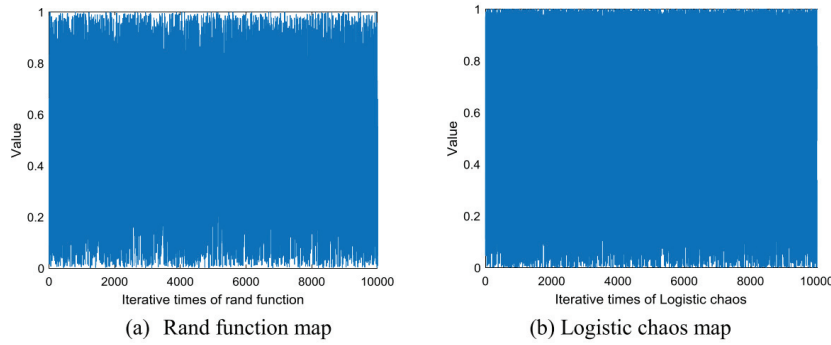


Figure 1. Random function compared with logistic map [42].

3.2.2. Adaptive Mutation Strategy

The primary goals of particles in the optimization process are searching and convergence to achieve the optimal solution. Therefore, the particle needs to initially search (explore) to enhance the diversity, then convergence will be obtained in the second stage. To address issues like exploration efficiency and premature convergence, this strategy is proposed. The strategy adjusts the mutation rate based on the particles' fitness values to balance the exploitation and exploration of the searching space. To achieve that, the adaptive mutation strategy uses Equation (9):

$$x(1+t) = x(t) + \varepsilon v(1+t), \quad (9)$$

where t is the current simulation time, v is the velocity, and ζ selects the particle's movement speed. A high value of ε helps in rapid exploration of the area but may affect the fine-tuned optimization. On the other hand, a low value of ε increases the convergence and adjusts the solution. Accordingly, for an effective optimization process, the particles must initially explore extensively and make significant leaps towards searching regions. As the iterations progress, the speed of the particles should decrease to facilitate rapid convergence. Consequently, ε should dynamically change with each iteration, as indicated by the following Equation (10):

$$\varepsilon = \varepsilon_{max} - \frac{(\varepsilon_{max} - \varepsilon_{min}) * t}{MaxIt}, \quad (10)$$

where $\varepsilon_{min} < \varepsilon_{max}$, and they are constant values. The current time and the total simulation time are t and $MaxIt$, respectively.

3.2.3. Inactive Particle Replacement

One main enhancement of the IPSO algorithm is replacing inactive particles with new, fresh particles. Inactive particles are those that have not been able to participate effectively in the searching process. This occurs when the particles lose the ability to search globally or even locally, such as when they fail to discover a better position, leading to premature convergence. Therefore, by replacing them with fresh ones, the searching process will converge towards the global optimum rather than getting stuck in the local optimum. Additionally, the algorithm incorporates other crucial components that contribute to its overall functionality and efficacy. These components will be highlighted to provide a comprehensive understanding of the IPSO algorithm.

3.2.4. Social Acceleration Coefficient

A crucial role is played by the acceleration coefficients in the IPSO algorithm, indicating the weight of the stochastic acceleration. By multiplying these coefficients, $c1$ and $c2$, with

random vectors, r_1 and r_2 , they introduce controllable stochastic effects on the velocity of the formation of the IoDs. Furthermore, the coefficients represent the weights that are assigned to share the information between particles. For example, if $c_1 = c_2 = 0$, the particle will solely rely on its own knowledge. Conversely, if $c_1 > c_2$, the particles will manage to gravitate towards local attractors (pBest), while if $c_2 > c_1$, it will cause the particles to lean towards the global attractor (gBest). Thus, the choice of c_1 and c_2 determines the balance between exploration and exploitation within the searching space.

3.2.5. Inertia Weight (ω)

The optimization algorithm's performance depends on balancing global and local searching. To accomplish this, the concept of inertia weight is employed to effectively manage the exploitation and exploration in the searching process. To provide further clarity, a large value of ω promotes exploration by encouraging the algorithm to explore a wider space, while a small value enables exploitation by focusing on exploiting promising regions. As a result, some studies have proposed adapting ω linearly, as demonstrated in [10]. The linear adaptation of ω is presented as shown in Equation (11):

$$\omega(t) = -\omega_{max} \frac{(MaxIt - t)}{MaxIt} + \omega_{min} \quad (11)$$

here, t represents the current time, while $MaxIt$ is the higher simulation time. ω_{min} and ω_{max} correspond to the minimum and maximum values of the inertia weight. By utilizing this linear adaptation, the ω value will change dynamically over time, gradually transitioning from ω_{max} to ω_{min} as the simulation progresses. This adaptive approach allows the algorithm to make a necessary balance between exploitation and exploration throughout the optimization process. In light of the aforementioned considerations, it can be generally stated that the IPSO algorithm consists of many parts. It starts by initializing all parameters, creating the initial position, velocity, and solution of all the particles. After that, the process of planning the path will be carried out by obtaining the parameters and environment constraints to return feasible paths to the destination with corresponding average fitness. The improvements mentioned for the PSO are added to increase the search efficiency. Furthermore, as discussed previously, adaptive adjustment of ϵ and ω_{max} occurs at each iteration. At the end of each iteration, the trail of each particle is evaluated. If the trail exceeds the threshold, the inactive particle will be replaced with a refreshed one. Then, the algorithm will be repeated until an optimal solution is attained or the time limit is reached.

3.3. Q-Learning Algorithm

The Q-learning algorithm is an RL algorithm that focuses on how an agent can learn from the environment by making sequential decisions that maximize the cumulative rewards. RL has undergone significant advancements and refinements. The Q-learning algorithm is a model-free and off-policy RL method. The agent of Q-learning is a value-based agent trained to estimate the return or future rewards during the learning process. Thus, the agent will select and output the action for which the greatest return will be obtained using the concept of a Q-value. The Q-value is the content of the Q-table; the rows and columns of the Q-table represent the states and actions. Initially, all the Q-values are random or zeros. Then, the Bellman equation is employed to obtain the Q-value. The max Q-value representing the best action should be selected for a particular state. Thus, Q-learning as an RL algorithm needs to use the sequential behavior decision problem, which is defined by the Markov decision process (MDP). The process of the Bellman equation is formulated by using the MDP and the value function. In the next subsection, a detailed explanation will be provided for the Markov decision process, the value function, and the Bellman equation, which form the foundation of RL algorithms.

3.3.1. Markov Decision Process (MDP)

The MDP captures the idea of the agent interacting with the environment over several time steps. At each time step, the agent takes an action A after observing the current state of the environment S and accordingly will obtain reward R [43,44]. The environment of the agent is probabilistic, meaning that after the action is taken, the transition of the state and the compensation are random. Policies are selections of actions that should be performed in a specific state, which can be formulated using MDP [45]. The MDP consists of the following components:

State: The environment's state at time step t denoted by S_t , which represents the state of the environment at that moment. The agent has access to a set of observable states, denoted as S , which refers to the observations made by the agent regarding its own situation [46].

Action: The action refers to a set of possible actions denoted as A within a given state S . The action A_t is the action that occurs at time step t . Typically, the agent's available actions remain the same across all states. Thus, a single representation can be used to indicate the set of actions A [47].

Policy: Policy in RL can be defined as the behavior of the agent at a given time. In other words, the policy converts the perceived states to actions to be taken when in those states. Policy can be as complicated searching process that needs significant computational resources, or it can be as simple as a lookup table or a small function. The policy represents an essential part of RL, in the sense that the agent cannot determine the behavior without it. In general, the policies specify the probabilities for each action [12]. The π in Equation (12) is the policy probability that the agent may choose an action a in a state s at time step t .

$$\pi(a|s) = P[A_t = a | S_t = s] \quad (12)$$

Reward: The goal of an RL problem can be defined by the reward signal. The environment will give the RL agent numerical value called the reward. Therefore, the agent's role is to try to maximize the total reward that receives over running time. Thus, the reward signal R defines what are the good and bad actions the agent can make.

Discount factor: The discount factor or discount rate is denoted by γ . Its value is between 0 and 1, determining the importance of future rewards compared to immediate rewards. In each state, when the agent takes an action, it will receive a reward as compensation [12]. The discount factor allows the agent to balance between immediate rewards and delayed rewards by enabling it to make decisions that align with either short-term or long-term objectives based on the chosen value of γ . The choice of discount rate depends on the specific problem and the agent's goals. A lower discount rate may be suitable for situations where immediate rewards are of importance, while a higher discount rate may be more appropriate when long-term cumulative rewards are the primary focus [48].

Value Function: The value function is a key concept in RL. It provides the prediction of the expected future reward that can be achieved if the agent follows a particular policy. It measures how each state or state-action pair performs [49]. The value function is linked to the Bellman equation, describing the relationship between the state's value and the values of its neighboring states [50].

3.3.2. Bellman Equation

Q-learning algorithms use the Bellman equation as an exploitation–exploration strategy to update the Q-values and to improve the policy over time. Q-learning iteratively updates the Q-values, which represent the expected cumulative reward for taking a particular action in a given state. After using the Bellman equation to derive the Q-values, the best action is selected based on the max Q-value for a particular state [50]. Equation (13) is simplified to obtain the Q-value [45], while Equation (14) uses the temporal difference:

$$Q^{new}(s, a) = r_t + \gamma \max Q(s_{(t+1)}, a) \quad (13)$$

$$Q^{new}(s, a) = Q(s_{(t-1)}, a_{(t-1)}) + \alpha \cdot (r_t + \gamma \max Q(s_{(t+1)}, a) - Q(s_{(t-1)}, a_{(t-1)})) \quad (14)$$

Table 2 below provides more clarity about the equations.

Table 2. Clarification of parameters in Equations (13) and (14).

Parameter Name	Description
s	Current state
a	Action taken in the current state
r_t	Reward at moment t
α	Learning rate, value between 0 and 1
γ	Discount factor, value between 0 and 1
$Q^{new}(s, a)$	New Q-value for action a after taking state s
$Q(s_{(t-1)}, a_{(t-1)})$	Old Q-value for action a after taking state s
$\max Q(s_{(t+1)}, a)$	Maximum expected value among the policies that the agents can receive
$(r_t + \gamma \max Q(s_{(t+1)}, a) - Q(s_{(t-1)}, a_{(t-1)}))$	Temporal difference

3.4. Deep Reinforcement Learning (DRL)

After providing an overview of Q-learning, which is one of the traditional RL algorithms, we explore the advancements made in the field of deep reinforcement learning (DRL). DRL combines the principles of RL with the powerful function approximation capabilities of deep neural networks, allowing for the efficient learning of complex, high-dimensional environments. Thus, the key advantage of DRL is its ability to handle large state and action spaces like what we have in path planning in a 3D environment, which are often present in real-world problems. Deep neural networks can effectively learn to represent the value function or the policy, enabling the agent to make informed decisions in complex environments [51]. An alternative approach in DRL is the actor–critic model, which combines two methods, policy-based and value-based methods, to utilize the strengths of both. The next subsection provides a detailed explanation of the actor–critic model.

Actor–Critic Model

The actor–critic model is a method that combines the strengths of value-based methods like Q-learning and policy-based methods like policy gradients to help the agent learn both the value function and the policy. This approach can often lead to more stable and efficient learning, as the critic provides valuable guidance to the actor. The actor can focus on exploration and policy optimization, while the critic provides a stable and informative value function to guide the actor's updates [12,51].

Neural Networks: When dealing with complex problems, the Q-table used in reinforcement learning needs to store a very large number of states and actions. This can result in a significant slowdown in the training speed and reduce the overall effectiveness of the approach. Path planning and obstacle avoidance are examples of complex problems of this nature. To address this, a neural network is used to calculate values rather than relying solely on the Q-table. By using the DRL, the system can handle the large number of states and actions required for the complex obstacle avoidance problem without the performance issues associated with a Q-table of that scale [27]. In DRL, the neural networks are used as the function approximators for both the policy and value functions. The specific neural network architectures can vary depending on the problem domain [52]. The actor–critic model involves two main components:

1. **Actor network:** This network is responsible for mapping the current state to actions and effectively learning the policy that determines the agent's behavior. It aims to maximize the expected cumulative reward.

2. Critic network: This network evaluates the actions taken by the actor network and provides feedback about the quality and precision of those actions. It learns to estimate the value function, which represents the expected future reward [51,53].

Experience replay memory: Experience replay memory is a popular technique used in deep Q-learning (DQL). It helps to reduce the correlation between samples, which improves the efficiency of the learning process [52]. There are some limitations of DQN that experience replay memory addresses:

1. The optimal policy is deterministic, which limits its use in adversarial domains where more flexibility is needed.
2. Finding the best action with respect to the Q-function can be computationally expensive, especially when there are many possible actions [54].

Experience replay memory refers to the storage of the agent's past experiences (state, actions, rewards, next states) in a replay buffer. During training, the agent periodically samples a batch of experiences from this buffer and uses them to update the neural network parameters rather than just using the most recent experience. The main benefit of using such memory is that by sampling the experiences randomly from the buffer, it breaks the correlation between consecutive samples that occurs during learning. This can lead to more stable and efficient learning. Also, the reuse of past experiences allows the agent to learn more from the number of interactions with the environment. Additionally, replaying the past experiences helps the agent retain knowledge about earlier parts of the task, preventing it from forgetting important information. Therefore, the combination of the actor–critic method and experience replay can lead to powerful and data-efficient training of the agent [51,52].

3.5. Proposed IPSO-DRL System Model

In our proposed solution, IPSO and RL-based methods are integrated to formulate the IPSO-RL framework. The system model of the proposed framework is illustrated in Figure 2. The IPSO and DRL components are described above. The construction of this framework considers the surrounding environmental information, including static and dynamic obstacles, as well as other UAVs, to generate appropriate actions while considering energy consumption. One key feature of our proposed framework is scalability, even in large-scale environments, by utilizing the IPSO guidance, enabling the UAV to learn and navigate using a fixed-sized learning model. By incorporating IPSO guidance, the framework maintains its efficiency and effectiveness in complex environments. The RL-based planner utilizes spatial and temporal information within a local area, such as the UAV's field of view (FoV), as shown in Figure 2, to make real-time decisions and navigate effectively. To drive the learning-based planning process, a reward function is designed to combine and consider the IPSO guidance in calculating the reward, motivating the UAVs to explore various potential solutions while encouraging convergence in the learning-based planning process. The FoV observation, as well as IPSO guidance, are utilized as inputs to the CNN that learns from current observation, IPSO guidance, and historical information, predicting an appropriate action.

The process begins by initializing the environment, which involves creating a random map containing static obstacles, dynamic obstacles, and drones, which represents the RL agents. Next, the start and end coordinates are established. The energy-efficient path planning conducted using the IPSO algorithm ensures minimum energy consumption and avoids static obstacles. The IPSO path serves as a consistent reference for the drone, allowing the RL-based planner to adapt and make responsive decisions based on real-time environmental conditions. After the establishment of the path, the drone navigates the environment using an RL-based planner—in our case, Q-learning or DRL—to reach its goal by generating appropriate actions. The drone receives positive rewards for following the IPSO path, making this path preferable to others. Conversely, if the drone collides with any obstacles or goes outside the boundary, it incurs a negative reward or penalty. Additionally, the drone earns positive rewards as it follows IPSO path or gets closer to

the goal, encouraging its learning process. The highest reward is given when the drone successfully reaches the goal, to reinforce the correct completion of the task. Energy consumption is a critical factor in drone path planning; therefore, we provide higher rewards when the drone follows the IPSO path to encourage the drone to use it. Also, for each step, the drone will receive a reward that corresponds with its approach to the goal. This process will lead to enhanced energy efficiency. Figure 3 shows the flowchart of the IPSO-RL framework.

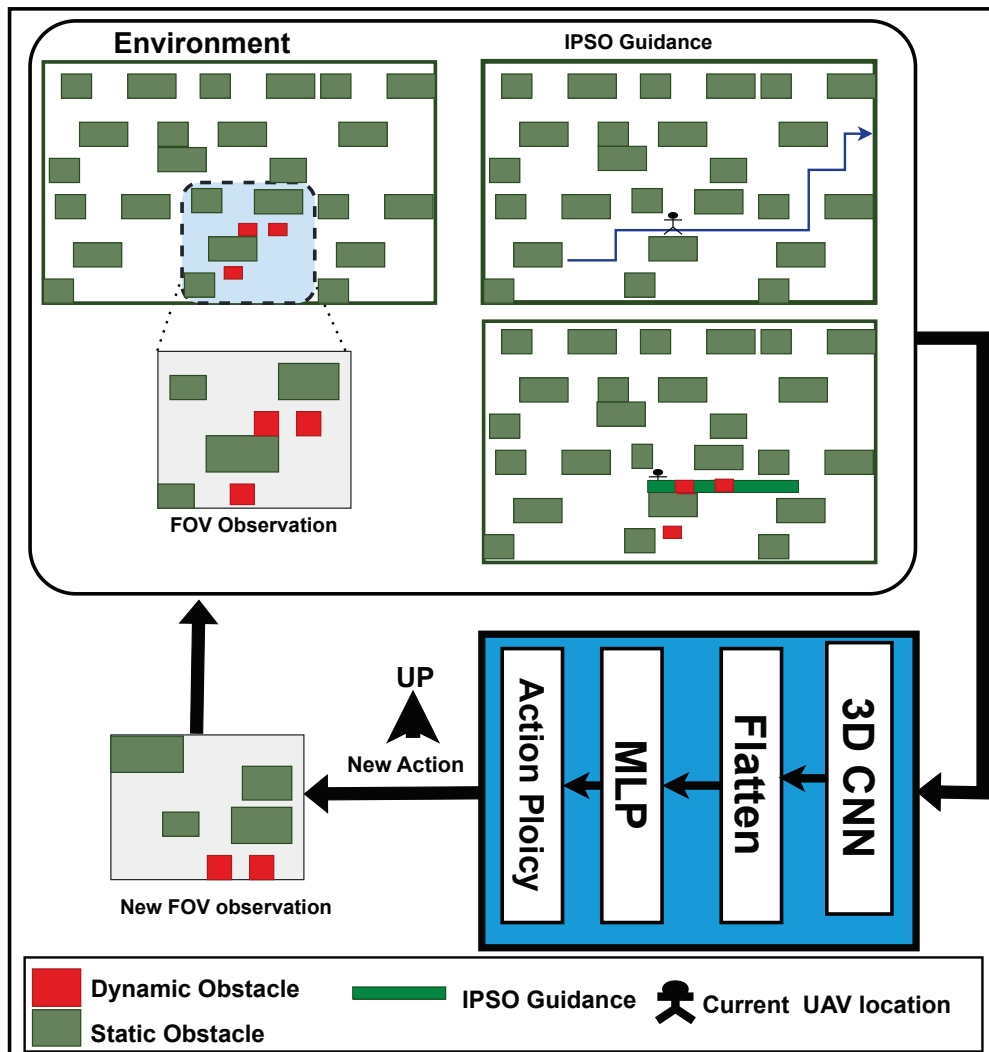


Figure 2. Proposed IPSO-DRL system model.

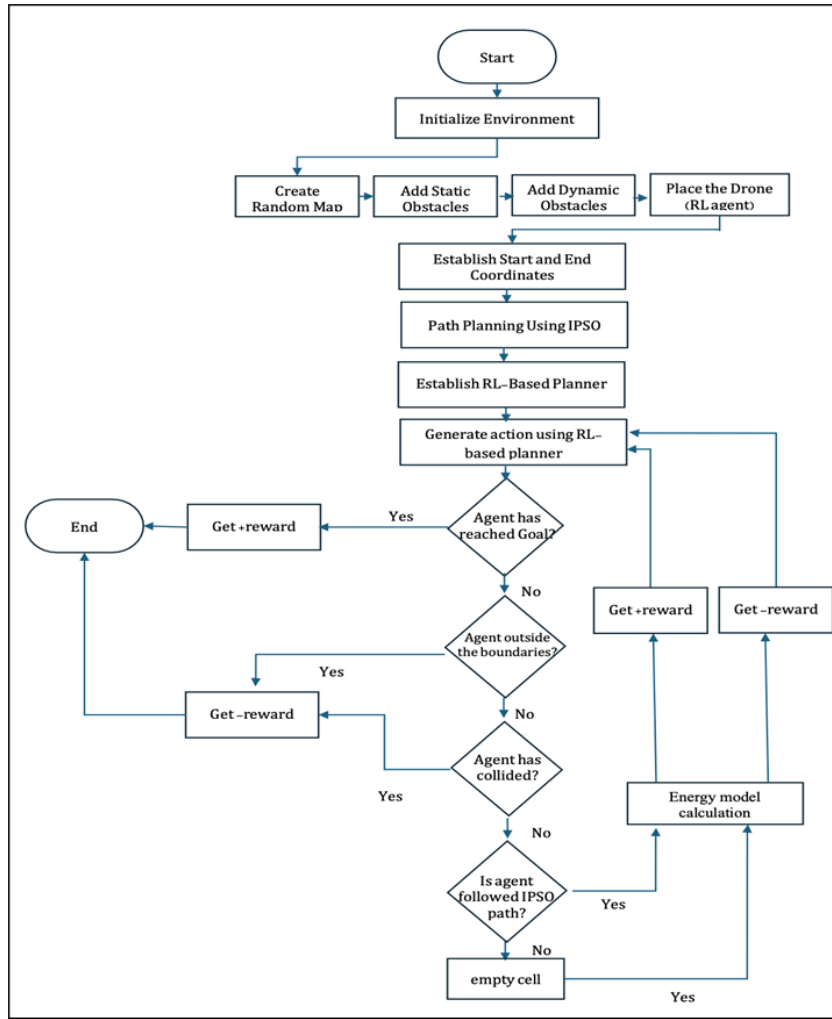


Figure 3. Flowchart of IPSO-RL framework.

3.5.1. Reward Function Structure

The reward function of the proposed approach is designed to manage the movement of the drones within the environment by calculating the reward that helps the agent navigate around obstacles effectively to reach the destination. To accomplish this, this part of the proposed solution is organized into functions and helper methods that collectively constitute the goal of the reward operation. The helper methods include methods for calculating the distance and another method for the dictionary to return rewards based on different cases. The drone will face several cases while navigating toward the goal, like obstacles or the IPSO path. These cases are carefully identified and processed to ensure the best learning for the drone. These cases are treated in this sequence with related rewards and processes:

- If the drone reaches the goal, a large positive reward is defined.
- If the drone collides with obstacles or goes out of bounds, a large negative reward is defined.
- If the drone reaches one of the locations on the IPSO path, a large positive reward is defined, as well as more rewards for being near the goal. This reinforces the drone to follow the IPSO guidance.
- If the drone reaches a free cell that is not part of the IPSO guidance, a small negative reward is defined. This discourages deviation from the reference path. Also, this small negative reward is subjected to change depending on the distance from that point to the goal.

After setting up the environment and implementing the necessary functions, the training process commences using two distinct approaches: Q-learning and actor–critic DRL.

3.5.2. Q-Learning Approach

For the Q-learning approach, we trained the drone by iteratively updating the Q-values based on the agent's interactions with the environment. This required the agent to observe the current state, select an action, receive a reward, and then update the Q-value estimate accordingly. The goal of this training was to enable the agent to learn the optimal action–value function, which would guide the drone's decision-making process. Algorithm 1 presents the pseudocode that we use in our approach.

In contrast, we use the actor–critic model as our DRL approach, utilizing a more sophisticated neural network architecture. This method consists of two main components: the actor network and the critic network. The actor network is responsible for selecting the most appropriate actions based on the current state, while the critic network evaluates the quality of those actions and provides feedback to the actor network. Through this iterative process of action selection and value estimation, the drone is able to learn a more complex and effective policy for navigating the environment. Algorithm 2 is the pseudocode that explains the general structure of the training process of the IPSO-DRL framework.

Algorithm 1: Pseudocode of Q-learning of the IPSO-RL framework

```

Initialization: Initialize the environment boundary;
                  -Initialize Q-table q-table with zeros;
                  -Set hyperparameters: learning rate ( $\alpha$ ), discount factor ( $\gamma$ ),
                    exploration rate ( $\epsilon$ ) and the  $\epsilon$  decay rate;
                  -Initialize variables: Episodes, Epochs, reward, all rewards ;

for  $i \leftarrow 1$  to  $Episodes$  do
  Reset the environment to initial state;
  for  $j \leftarrow 1$  to  $Epochs$  do
    Choose an action using  $\epsilon$ -greedy policy based on Q-table ;
    Take the chosen action;
    Calculate the reward based on action and IPSO guidance information;
    Obtain next state;

    if  $numEpochs$  is greater than the inverse of  $\epsilon\_Decay$  then
      | set  $\alpha$  to 0.01;
    end
    while  $episode$  is not done do
      if  $RandomNum$  is less than  $\epsilon$  then
        | The action is selected randomly from the actions set;
      else
        | The action is selected as argmax of the Q-table;
      end
      [next state, reward, done]  $\leftarrow$  Apply action;
      Update the cumulative reward;
      Update the Q-table with the new value;
      Update the state to the next state;
    end
  end
end

```

Algorithm 2: Pseudocode of actor–critic of the IPSO-DRL framework

```

Initialization:
Initialize the environment boundary, and obstacles.;
for Each UAVi in IoDs do
    Initialize starting and destination point ;
    Initialize the actor network;
    Initialize the critic network;
    Initialize the target actor network;
    Initialize the target critic network, Initialize experience replay buffer  $\kappa$ ;
end
Define store_experiences, target_model, actor;
Set hyperparameters:  $\epsilon$ ,  $\epsilon_{\text{decay}}$ ,  $\epsilon_{\text{min}}$ , batch_size, timesteps_per_episode, and numEpisodes.
for Each episode in Episodes do
    for Each time t in timesteps do
        for Each UAVi in IoDs do
            Action  $\leftarrow$  agent.actor(state,  $\epsilon$ , IPSO Guidance);
            next state, reward, done  $\leftarrow$  Apply action;
            Store experience (state, action, reward, next_state, done) in  $\kappa$ ;
            Update state to next_state;
            if length( $\kappa$ ) > batch_size then
                | Retrains the Q-network using a minibatch from the replay buffer;
            end
            if episode done then
                | Align target model and break;
            end
        end
    end
end

```

4. Simulation Results and Discussion

To evaluate our proposed framework for planning the path and avoiding obstacles with minimum energy consumption, we carried out a simulation in a dynamic environment.

In the following subsections, we explain the relevant parameters and discuss the results.

4.1. Parameter Settings

The hardware and software specifications are illustrated in Table 3.

Table 3. Hardware and software specifications.

Hardware/Software	Description
Operating system	macOS Sonoma
Processor	Apple M2/Intel i5
Memory	8GB/16GB
Python	3.10.10
TensorFlow	2.15.0

We use the epsilon decay technique to balance the agent exploration and exploitation. The parameters setting are listed in Table 4.

Table 4. Parameters used in Q-learning.

Parameter Name	Value
Learning rate (α)	0.01
Discount factor (γ)	0.99
Exploration rate (ϵ)	1.0, decaying to 0.02 over time

The environment is simulated with static obstacles and dynamic objects, where drones navigate to plan the path. Our environment is $8 \times 8 \times 4$. It contains three static obstacles and two dynamic objects, in addition to our drones, which are the RL agents. The speed is assumed to be 10 m/s for the drones and 1 m/s for dynamic obstacles.

Finally, Table 5 clarifies the parameters used in training the DRL approach.

Table 5. Parameters used in DRL.

Parameter Name	Value
Learning rate (α)	0.0001
Discount factor (γ)	0.95
Exploration rate (ϵ)	1.0, decaying to 0.01 over time
Batch size	64
Replay memory size	1000

4.2. Simulation Results

In this section, we present the simulation results for our proposed IPSO-RL framework. We conducted several comparative analyses to demonstrate the validity and effectiveness of the proposed method. First, we examined the outcomes of integrating the IPSO algorithm with the Q-learning method and analyzed the effects of incorporating the energy model into the framework by comparing the results with and without the energy model. Second, we further expanded our analysis by comparing the performance of our IPSO-RL framework against the DRL approach. This comparison helps to evaluate the advantages and trade-offs of using the IPSO-RL framework versus a more advanced DRL technique. Finally, we compared our findings from the IPSO-DRL framework against a recent study in the field. This comparative analysis provides valuable insights into how our proposed IPSO-DRL framework fares against state-of-the-art techniques in the literature.

4.2.1. IPSO-RL results

To achieve optimal performance, we examined various sets of parameters across 50,000 episodes. When we identified promising results, we extended the range and conducted additional training with those parameters to confirm their efficacy. Figure 4 depicts the drone's performance during the training process. We experimented with different ϵ values of 0.99, 0.5, and 0.1, in addition to the decaying technique. Moreover, we tested α values of 0.0001, 0.001, 0.01, and 0.1 and γ values of 0.99 and 0.

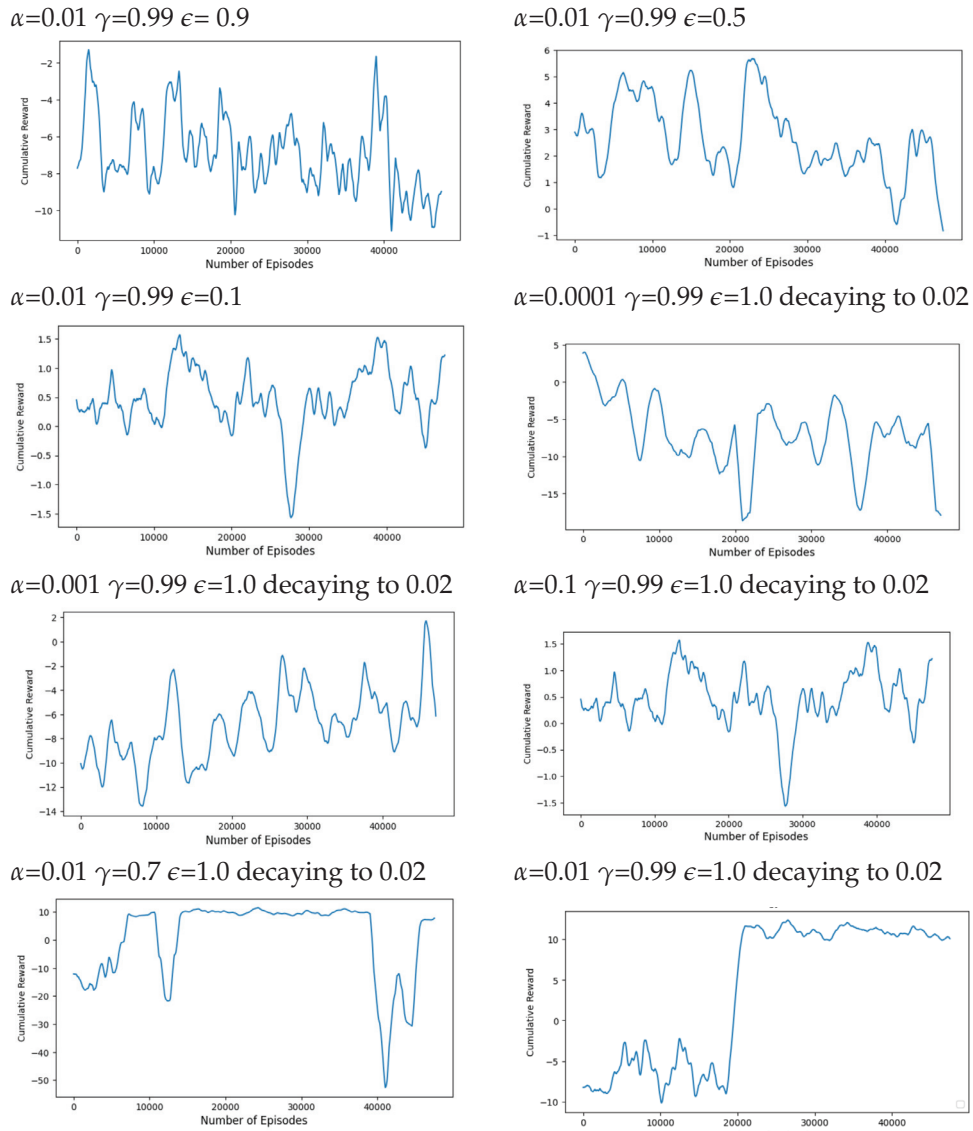


Figure 4. Cumulative reward with respect to the number of episodes for different values of α , γ , and ϵ .

4.2.2. DRL Results

First, we used our proposed framework to train the drone for 250,000 episodes using Q-learning under two conditions, one with the energy model included and one without. By monitoring the reward amount during training, we are able to observe the learning progress and the differences between the two approaches. When the energy model was included, we observed a significant improvement in the drone's performance after approximately 15,000 episodes. The cumulative reward steadily increased, indicating continuous learning and enhancement, and it became relatively stable after approximately 30,000 episodes, indicating that the drone had reached an optimal level of learning and performance. Figure 5 shows a smaller scale of the learning process to visualize the interval where the gradual increase of the reward occurs, while Figure 6 shows the cumulative reward obtained by the drone throughout the whole learning process. The reward amount shown in the figures indicates that the drone was able to learn to operate in a more energy-efficient manner. Figure 7 shows the energy consumption during the learning process and how it decreases significantly as the number of episodes increases, indicating that the drone is able to learn and optimize its energy usage over time.

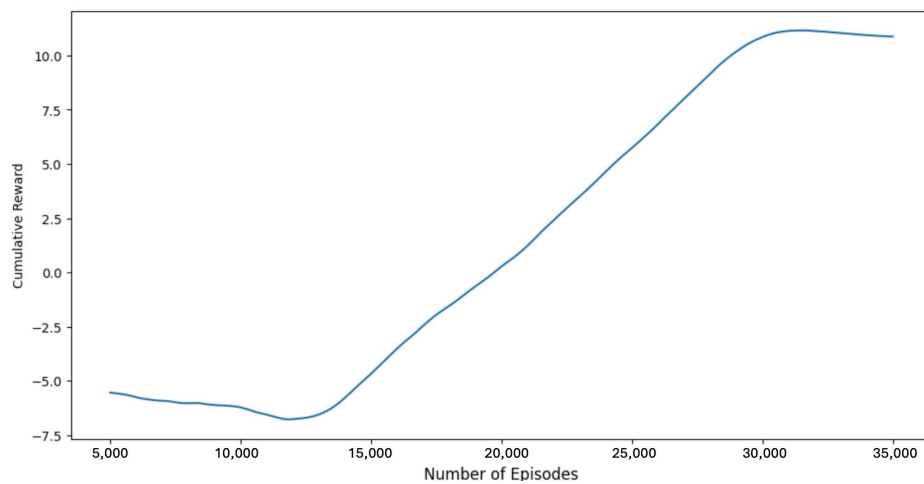


Figure 5. Scale where the improvement in the learning process occurs for IPSO-DRL with the energy model.

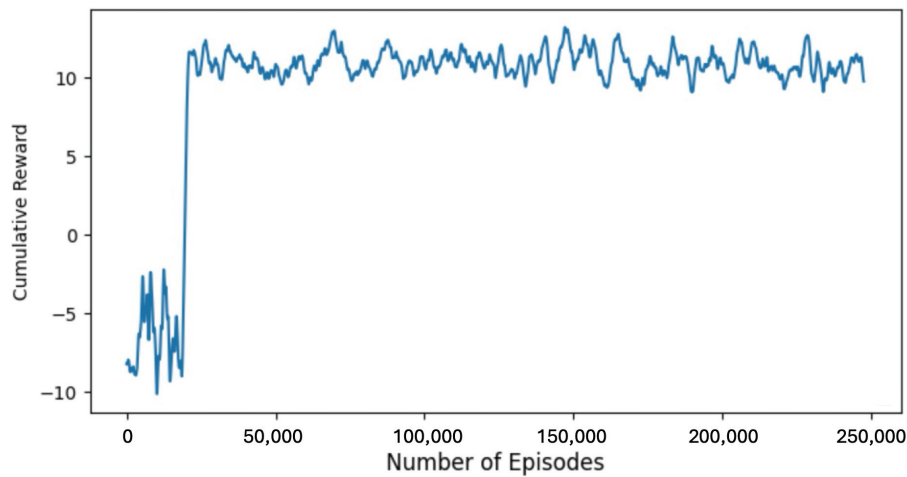


Figure 6. Reward curve for IPSO-RL with the energy model.

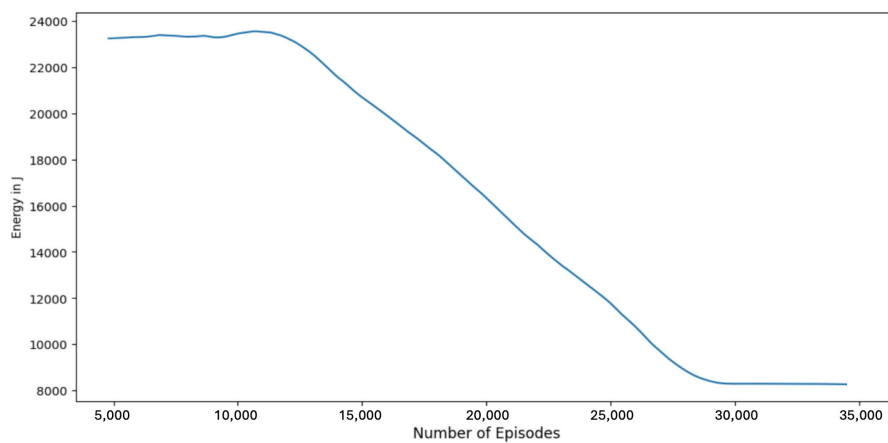


Figure 7. Energy consumption during the learning process.

In contrast, without the energy model, the improvement in performance is much slower. The improvement was noticed after approximately 150,000 episodes. Figure 8 shows where the exponent of the learning process starts. It begins after approximately 135,000 episodes and reaches its optimal after 155,000 episodes. Figure 9 shows the cumulative reward curve without using the energy model for all episodes.

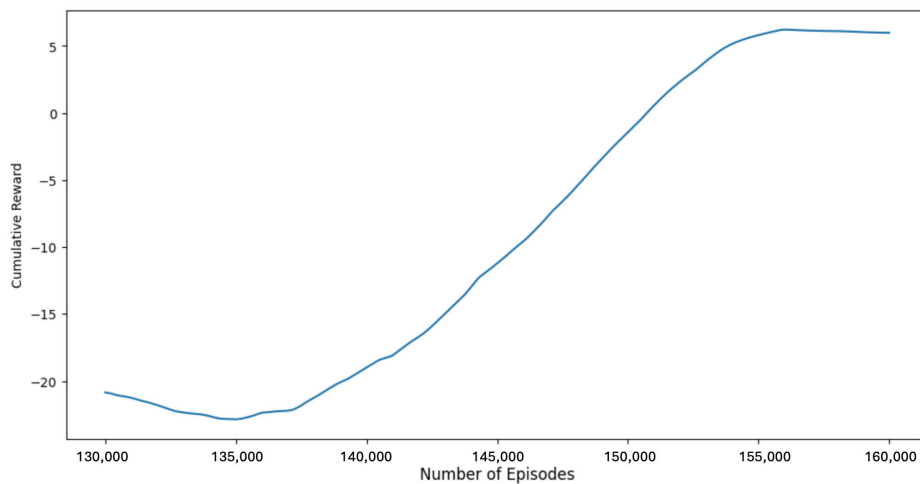


Figure 8. Scale where the improvement in the learning process occurs for IPSO-RL without the energy model.

We also analyzed the average reward curve, which provides an overall measure of the drone’s learning progress. With the energy model included, the moving average reward curve showed clear improvements over time, as depicted in Figure 10, demonstrating that the drone is able to learn more effective and comprehensive behaviors.

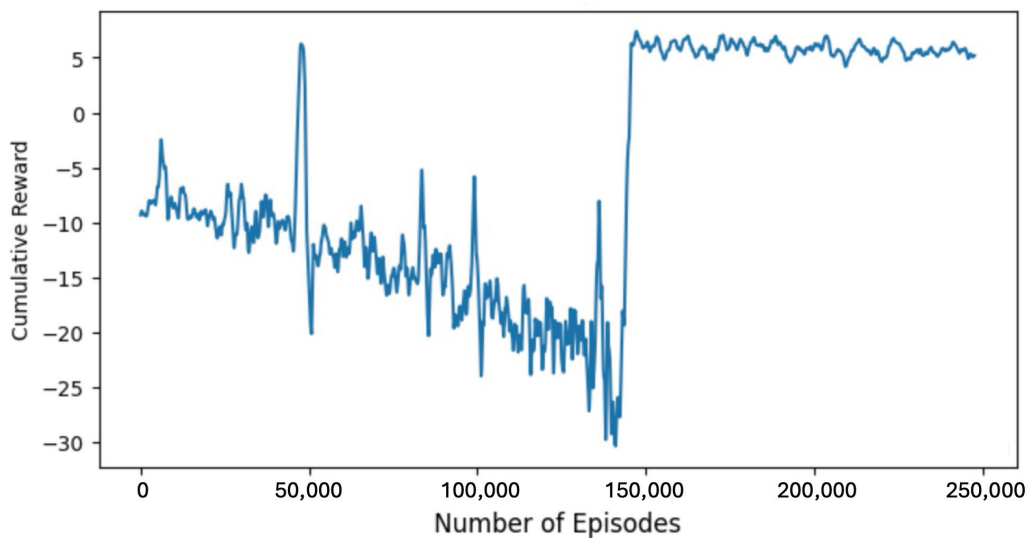


Figure 9. Reward curve for IPSO-RL without the energy model.

Without the energy model, the improvement in the moving average reward curve was more gradual and less apparent in comparison, as illustrated in Figure 11.

Next, we compared the performance of both the Q-learning RL (IPSO-RL) with the actor–critic DRL (IPSO-DRL). We can notice the rapid improvement of the IPSO-DRL model, shown in Figure 12, compared with the IPSO-RL model. Also, the upward trend in the average reward over the training indicates that the DRL model can learn more effectively within fewer episodes.

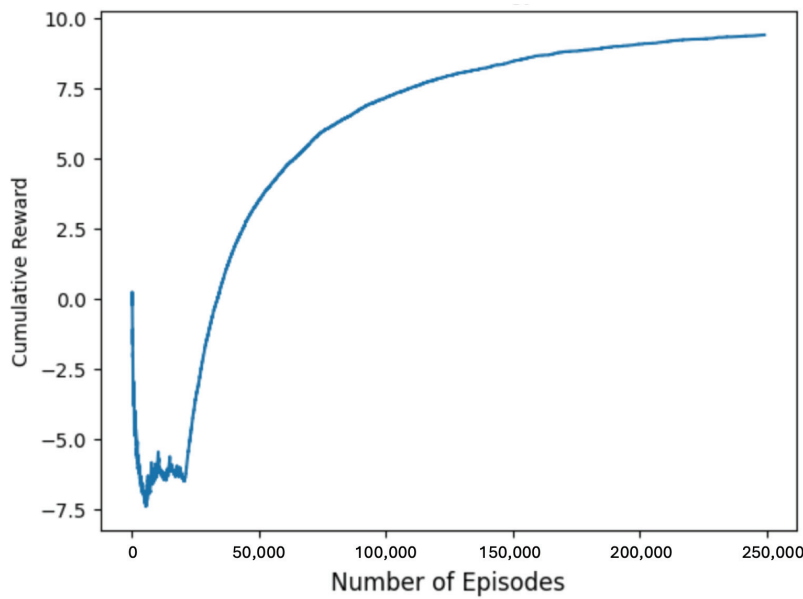


Figure 10. Moving average curve for IPSO with the energy model.

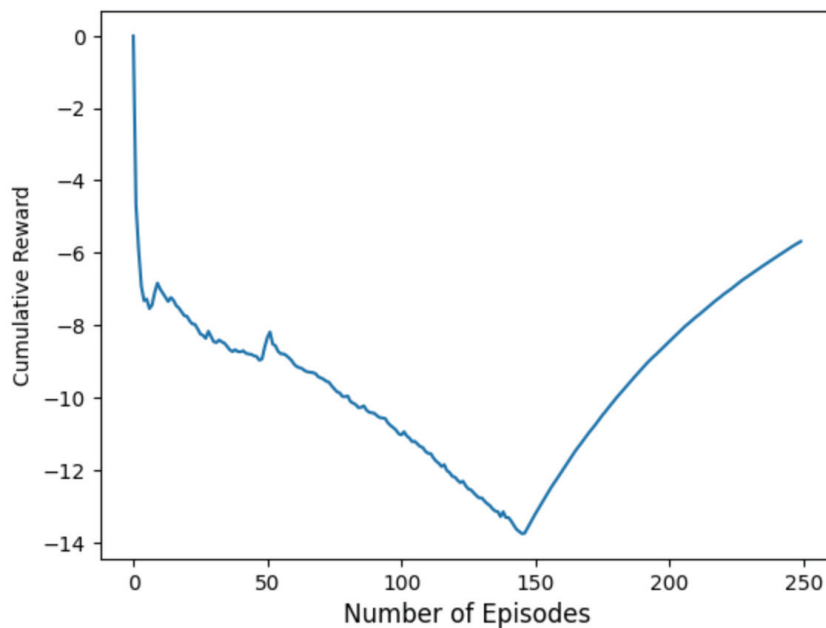


Figure 11. Moving average curve for IPSO with the energy model.

In contrast, the IPSO-RL plot (Figure 6) displays a more gradual improvement in the reward curve. While the IPSO-RL approach still demonstrated a positive learning trend, the magnitude of the rewards achieved was generally lower compared to the IPSO-DRL.

This difference in performance can be attributed to the inherent strengths of DRL techniques, such as their ability to effectively capture complex relationships and patterns in high-dimensional state spaces. The use of deep learning in the IPSO-DRL approach allows the agent to learn more expressive and efficient representations of the problem, leading to faster convergence and higher-quality solutions.

Finally, we compared our result of IPSO-DRL with a recent study. The study suggests using an improved DRL technique to solve the path planning problem in dynamic environments, employing the Q-function approximation of the prioritized experience replay D3QN to estimate the agent's action Q-values. The algorithm's network structure combines a competitive neural network architecture with double Q-learning and prioritized experience

replay. The simulation of the study was implemented using the TensorFlow platform. The environment includes static obstacles that the drone is trained to avoid; after that, the same model is used to train the drone with the dynamic obstacles. The dynamic obstacles are added randomly to the environment, and their movements are randomly generated as well. The reward function penalizes collisions with obstacles and rewards progress towards the goal. The start and goal positions for the drone are randomly generated within the environment. Figure 13 depicts the reward curve for the study. We can notice the improvement in the DRL technique for dynamic path planning. However, the reward curve depicted in these results, although improving over time, stabilizes at a lower cumulative reward compared to our framework. We can notice that our framework is able to reach a much higher level of cumulative reward compared to the previous study, highlighting a superior learning capability of the IPSO-DRL framework.

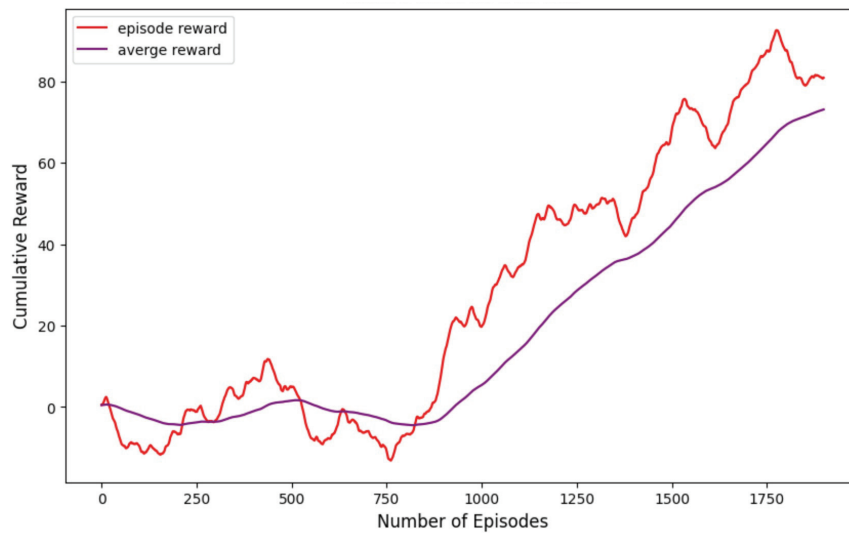


Figure 12. Reward curves for IPSO-DRL.

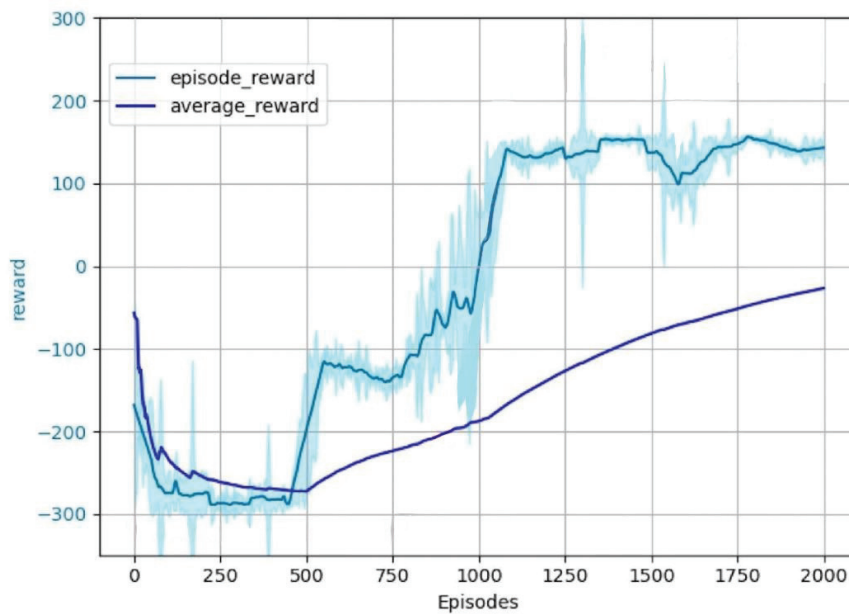


Figure 13. Reward curve for the recent study.

Through these comprehensive comparisons, we aimed to thoroughly assess the merits of the IPSO-RL framework and its ability to address the complexities involved in the problem at hand.

4.3. Discussion

This study highlights that the proposed IPSO-DRL framework significantly generates online energy-efficient paths for IoDs in dynamic environments compared to other existing methods. Furthermore, the proposed IPSO-RL framework acquires a feasible and effective route successfully with minimum energy consumption. The reason behind this can be attributed to the adaptive learning capabilities of the solution that integrates IPSO with RL. A significant improvement in the drone's performance, with the cumulative reward steadily increasing when the path is optimized for energy consumption minimization, indicated continuous learning and enhancement. This suggests that the drone had reached an optimal level of learning and performance. The energy consumption during the learning process significantly decreased as the number of episodes increased, indicating that the drone learned to optimize its energy usage over time. Moreover, the performance comparison of Q-learning with the actor–critic DRL models showed a much earlier improvement in the IPSO-DRL. The superior performance of the IP-SO-DRL model is attributed to its ability to capture complex relationships in high-dimensional state spaces through deep learning, leading to faster convergence and higher-quality solutions. Furthermore, our IPSO-DRL framework outperforms the recent study on dynamic path planning, demonstrating superior learning capabilities and achieving higher results. Unlike previous approaches that rely on static environments, our method dynamically adjusts paths in response to surrounding environmental changes, resulting in a more efficient solution. These findings suggest that implementing IPSO-RL in IoD systems leads to more sustainable and cost-effective operations, especially in energy-constrained environments.

The main limitation of our study is that it assumes all UAVs are flying at a fixed altitude. This choice was made to ensure a consistent resolution in applications that require the collection of useful information from the environment. Thus, this framework can not address different altitudes of IoD formations.

5. Conclusions

In this work, we proposed a novel approach for planning collision-free paths for UAVs operating in unknown, dynamic, 3D environments. By integrating the improved particle swarm optimization (IPSO) algorithm with reinforcement learning (RL), the developed IPSO-RL framework could generate energy-efficient and collision-free paths for drones in real-time. The proposed framework offered a flexible and adaptive solution for UAV path planning, capable of responding to dynamic environmental changes, thus addressing the limitations of traditional heuristic methods designed for static environments. Additionally, incorporating an energy model into the RL reward structure allowed the drone to optimize the energy of paths, leading to more sustainable and enduring drone operations. We conducted simulations to evaluate our proposed solution, training the drone for 250,000 episodes using Q-learning and actor–critic algorithms within the IPSO-RL framework. We tested the framework under two conditions: with and without the energy model included. The extensive simulations highlighted the superior performance of the IPSO-DRL approach compared to other benchmarks in terms of collision avoidance, path length, and energy consumption. By addressing the critical challenges of collision avoidance and energy efficiency, this research advances the state-of-the-art in autonomous UAV navigation. The proposed framework represents a significant step towards enabling safe and sustainable UAV operations in complex, dynamic environments.

In future work, we will consider variable altitudes of IoDs, requiring a sophisticated reward function. We will also optimize the UAV system to incorporate weather conditions and other constraints.

Author Contributions: Z.A. led the development of the work, performed most of the implementation tasks, and wrote the first draft. T.S., mentored the research, provided key ideas, guided the overall direction of the project, contributed to the validation of the work, and reviewed and correct the first draft of the paper. G.A. played a key role in writing the simulation codes, significantly contributed to the development of the IPSO-DRL model, and was involved in the conceptualization and validation of the research. A.M. provided critical feedback on the developed work and reviewed the final version of the paper. A.B. offered valuable feedback on the obtained results and reviewed the final version of the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the interdisciplinary center of smart mobility and logistics at King Fahd University of Petroleum and Minerals (Grant number [INML2400]).

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Shakhathreh, H.; Sawalmeh, A.H.; Al-Fuqaha, A.; Dou, Z.; Almaita, E.; Khalil, I.; Othman, N.S.; Khreishah, A.; Guizani, M. Unmanned aerial vehicles (UAVs): A survey on civil applications and key research challenges. *IEEE Access* **2019**, *7*, 48572–48634. [CrossRef]
- Ahmed, G.A.; Sheltami, T.R.O.; Mahmoud, A.S.; Yasar, A. 3D Simulation Model for IoD-to-Vehicles Communication in IoD-assisted VANET. *Front. Built Environ.* **2023**, *9*, 1287373. [CrossRef]
- Vo, T.M.N.; Wang, C.N.; Yang, F.C.; Singh, M. Internet of Things (IoT): Wireless Communications for Unmanned Aircraft System. *Eurasia Proc. Sci. Technol. Eng. Math.* **2023**, *23*, 388–399. [CrossRef]
- Ahmed, G.; Sheltami, T.; Mahmoud, A.; Imam, M. Performance Evaluation of Three Routing Protocols for Drone Communication Networks. *Arab. J. Sci. Eng.* **2024**, *49*, 13149–13161. [CrossRef]
- Ahmed, G.; Sheltami, T.; Deriche, M.; Yasar, A. An energy efficient IoD static and dynamic collision avoidance approach based on gradient optimization. *Hoc Netw.* **2021**, *118*, 102519. [CrossRef]
- Sheltami, T.; Ahmed, G.; Yasar, A. *An Optimization Approach of IoD Deployment for Optimal Coverage Based on Radio Frequency Model*; TECH SCIENCE PRESS: Henderson, NV, USA, 2024.
- Hu, Y.; Yang, S.X. A knowledge based genetic algorithm for path planning of a mobile robot. In Proceedings of the IEEE International Conference on Robotics and Automation, ICRA'04, New Orleans, LA, USA, 26 April–1 May 2004; Volume 5, pp. 4350–4355.
- Lin, S.; Li, F.; Li, X.; Jia, K.; Zhang, X. Improved artificial bee colony algorithm based on multi-strategy synthesis for UAV path planning. *IEEE Access* **2022**, *10*, 119269–119282. [CrossRef]
- Lin, A.; Sun, W.; Yu, H.; Wu, G.; Tang, H. Global genetic learning particle swarm optimization with diversity enhancement by ring topology. *Swarm Evol. Comput.* **2019**, *44*, 571–583. [CrossRef]
- Ahmed, G.; Sheltami, T.; Mahmoud, A.; Yasar, A. IoD swarms collision avoidance via improved particle swarm optimization. *Transp. Res. Part Policy Pract.* **2020**, *142*, 260–278. [CrossRef]
- Hong, D.; Lee, S.; Cho, Y.H.; Baek, D.; Kim, J.; Chang, N. Energy-efficient online path planning of multiple drones using reinforcement learning. *IEEE Trans. Veh. Technol.* **2021**, *70*, 9725–9740. [CrossRef]
- Nemer, I.A.; Sheltami, T.R.; Belhaiza, S.; Mahmoud, A.S. Energy-efficient UAV movement control for fair communication coverage: A deep reinforcement learning approach. *Sensors* **2022**, *22*, 1919. [CrossRef]
- Sutton, R.S. *Reinforcement Learning: An Introduction*; A Bradford Book; The MIT Press: Cambridge, MA, USA; London, UK, 2018.
- Al-Hourani, A.; Kandeepan, S.; Lardner, S. Optimal LAP altitude for maximum coverage. *IEEE Wirel. Commun. Lett.* **2014**, *3*, 569–572. [CrossRef]
- Ahmed, G.A.; Sheltami, T.R.; Mahmoud, A.S.; Imran, M.; Shoaib, M. A novel collaborative IoD-assisted VANET approach for coverage area maximization. *IEEE Access* **2021**, *9*, 61211–61223. [CrossRef]
- Xia, X.; Xing, Y.; Wei, B.; Zhang, Y.; Li, X.; Deng, X.; Gui, L. A fitness-based multi-role particle swarm optimization. *Swarm Evol. Comput.* **2019**, *44*, 349–364. [CrossRef]
- Ahmed, G.; Sheltami, T. A safety system for maximizing operated uavs capacity under regulation constraints. *IEEE Access* **2023**, *11*, 139069–139081. [CrossRef]
- Ha, L.N.N.T.; Bui, D.H.P.; Hong, S.K. Nonlinear control for autonomous trajectory tracking while considering collision avoidance of UAVs based on geometric relations. *Energies* **2019**, *12*, 1551. [CrossRef]
- Ma, H.; Shen, S.; Yu, M.; Yang, Z.; Fei, M.; Zhou, H. Multi-population techniques in nature inspired optimization algorithms: A comprehensive survey. *Swarm Evol. Comput.* **2019**, *44*, 365–387. [CrossRef]
- Zhao, Y.; Zheng, Z.; Liu, Y. Survey on computational-intelligence-based UAV path planning. *Knowl.-Based Syst.* **2018**, *158*, 54–64. [CrossRef]
- Zhang, C.; Zhen, Z.; Wang, D.; Li, M. UAV path planning method based on ant colony optimization. In Proceedings of the 2010 Chinese Control and Decision Conference, Xuzhou, China, 26–28 May 2010; pp. 3790–3792.

22. Lin, Z.; Castano, L.; Mortimer, E.; Xu, H. Fast 3D collision avoidance algorithm for fixed wing UAS. *J. Intell. Robot. Syst.* **2020**, *97*, 577–604. [CrossRef]
23. Theile, M.; Bayerlein, H.; Nai, R.; Gesbert, D.; Caccamo, M. UAV path planning using global and local map information with deep reinforcement learning. In Proceedings of the 2021 20th International Conference on Advanced Robotics (ICAR), Ljubljana, Slovenia, 6–10 December 2021; pp. 539–546.
24. Wang, D.; Fan, T.; Han, T.; Pan, J. A two-stage reinforcement learning approach for multi-UAV collision avoidance under imperfect sensing. *IEEE Robot. Autom. Lett.* **2020**, *5*, 3098–3105. [CrossRef]
25. Zhang, X.; Xia, S.; Li, X.; Zhang, T. Multi-objective particle swarm optimization with multi-mode collaboration based on reinforcement learning for path planning of unmanned air vehicles. *Knowl.-Based Syst.* **2022**, *250*, 109075. [CrossRef]
26. Liu, X.; Zhang, D.; Zhang, T.; Zhang, J.; Wang, J. A new path plan method based on hybrid algorithm of reinforcement learning and particle swarm optimization. *Eng. Comput.* **2022**, *39*, 993–1019. [CrossRef]
27. Tu, G.T.; Juang, J.G. UAV path planning and obstacle avoidance based on reinforcement learning in 3d environments. *Actuators* **2023**, *12*, 57. [CrossRef]
28. Hsu, Y.H.; Gau, R.H. Reinforcement learning-based collision avoidance and optimal trajectory planning in UAV communication networks. *IEEE Trans. Mob. Comput.* **2020**, *21*, 306–320. [CrossRef]
29. Maw, A.A.; Tyan, M.; Nguyen, T.A.; Lee, J.W. iADA*-RL: Anytime graph-based path planning with deep reinforcement learning for an autonomous UAV. *Appl. Sci.* **2021**, *11*, 3948. [CrossRef]
30. Xu, Y.; Wei, Y.; Jiang, K.; Wang, D.; Deng, H. Multiple UAVs path planning based on deep reinforcement learning in communication denial environment. *Mathematics* **2023**, *11*, 405. [CrossRef]
31. Lee, M.H.; Moon, J. Deep reinforcement learning-based model-free path planning and collision avoidance for UAVs: A soft actor-critic with hindsight experience replay approach. *ICT Express* **2023**, *9*, 403–408. [CrossRef]
32. Puente-Castro, A.; Rivero, D.; Pedrosa, E.; Pereira, A.; Lau, N.; Fernandez-Blanco, E. Q-learning based system for path planning with unmanned aerial vehicles swarms in obstacle environments. *Expert Syst. Appl.* **2024**, *235*, 121240. [CrossRef]
33. Zhao, X.; Yang, R.; Zhong, L.; Hou, Z. Multi-UAV Path Planning and Following Based on Multi-Agent Reinforcement Learning. *Drones* **2024**, *8*, 18. [CrossRef]
34. Wu, J.; Ye, Y.; Du, J. Multi-objective reinforcement learning for autonomous drone navigation in urban areas with wind zones. *Autom. Constr.* **2024**, *158*, 105253. [CrossRef]
35. Wang, Z.; Gao, W.; Li, G.; Wang, Z.; Gong, M. Path Planning for Unmanned Aerial Vehicle via Off-Policy Reinforcement Learning With Enhanced Exploration. *IEEE Trans. Emerg. Top. Comput. Intell.* **2024**, *8*, 2625–2639. [CrossRef]
36. da Rocha, L.G.S.; Caldas, K.A.Q.; Terra, M.H.; Ramos, F.; Vivaldini, K.C.T. Dynamic Q-planning for Online UAV Path Planning in Unknown and Complex Environments. *arXiv* **2024**, arXiv:2402.06297.
37. Kong, F.; Wang, Q.; Gao, S.; Yu, H. B-APFDQN: A UAV path planning algorithm based on deep Q-network and artificial potential field. *IEEE Access* **2023**, *11*, 44051–44064. [CrossRef]
38. Sonny, A.; Yeduri, S.R.; Cenkeramaddi, L.R. Q-learning-based unmanned aerial vehicle path planning with dynamic obstacle avoidance. *Appl. Soft Comput.* **2023**, *147*, 110773. [CrossRef]
39. Tang, J.; Liang, Y.; Li, K. Dynamic Scene Path Planning of UAVs Based on Deep Reinforcement Learning. *Drones* **2024**, *8*, 60. [CrossRef]
40. Li, M.; Cheng, N.; Gao, J.; Wang, Y.; Zhao, L.; Shen, X. Energy-efficient UAV-assisted mobile edge computing: Resource allocation and trajectory optimization. *IEEE Trans. Veh. Technol.* **2020**, *69*, 3424–3438. [CrossRef]
41. Ahmed, G.; Sheltami, T.; Mahmoud, A. Energy-Efficient Multi-UAV Multi-Region Coverage Path Planning Approach. *Arab. J. Sci. Eng.* **2024**, *49*, 13185–13202. [CrossRef]
42. Shao, S.; Peng, Y.; He, C.; Du, Y. Efficient path planning for UAV formation via comprehensively improved particle swarm optimization. *ISA Trans.* **2020**, *97*, 415–430. [CrossRef]
43. Agarwal, A.; Kakade, S.M.; Lee, J.D.; Mahajan, G. On the theory of policy gradient methods: Optimality, approximation, and distribution shift. *J. Mach. Learn. Res.* **2021**, *22*, 1–76.
44. Liu, Q.; Zhuang, Y.; Bi, H.; Huang, Z.; Huang, W.; Li, J.; Yu, J.; Hu, Z.; Hong, Y.; et al. Survey of Computerized Adaptive Testing: A Machine Learning Perspective. *arXiv* **2024**, arXiv:2404.00712.
45. Jang, B.; Kim, M.; Harerimana, G.; Kim, J.W. Q-learning algorithms: A comprehensive classification and applications. *IEEE Access* **2019**, *7*, 133653–133667. [CrossRef]
46. Cassandra, A.R. *Exact and Approximate Algorithms for Partially Observable Markov Decision Processes*; Brown University: Providence, RI, USA, 1998.
47. Littman, M.L. Value-function reinforcement learning in Markov games. *Cogn. Syst. Res.* **2001**, *2*, 55–66. [CrossRef]
48. Miller, S.L.; Childers, D. CHAPTER 9—Markov Processes; Academic Press: Cambridge, MA, USA, 2012; pp. 383–428. [CrossRef]
49. Pan, Y.; Zhang, J.; Yuan, C.; Yang, H. Supervised Reinforcement Learning via Value Function. *Symmetry* **2019**, *11*, 590. [CrossRef]
50. Sivamayil, K.; Rajasekar, E.; Aljafari, B.; Nikolovski, S.; Vairavasundaram, S.; Vairavasundaram, I. A systematic study on reinforcement learning based applications. *Energies* **2023**, *16*, 1512. [CrossRef]
51. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.; Fidjeland, A.K.; Ostrovski, G.; et al. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533. [CrossRef] [PubMed]

52. François-Lavet, V.; Henderson, P.; Islam, R.; Bellemare, M.G.; Pineau, J. An introduction to deep reinforcement learning. *Found. Trends® Mach. Learn.* **2018**, *11*, 219–354. [CrossRef]
53. Lillicrap, T.P.; Hunt, J.J.; Pritzel, A.; Heess, N.; Erez, T.; Tassa, Y.; Silver, D.; Wierstra, D. Continuous control with deep reinforcement learning. *arXiv* **2015**, arXiv:1509.02971.
54. Schaul, T.; Quan, J.; Antonoglou, I.; Silver, D. Prioritized experience replay. *arXiv* **2015**, arXiv:1511.05952.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Tackling Few-Shot Challenges in Automatic Modulation Recognition: A Multi-Level Comparative Relation Network Combining Class Reconstruction Strategy

Zhao Ma ¹, Shengliang Fang ^{2,*}, Youchen Fan ², Shunhu Hou ¹ and Zhaojing Xu ¹

¹ Graduate School, Space Engineering University, Beijing 101416, China; nuaamazhao@163.com (Z.M.); hshhsc2022@163.com (S.H.); zhaojing3973@163.com (Z.X.)

² School of Aerospace Information, Space Engineering University, Beijing 101400, China; love193777@sina.com

* Correspondence: eeifslyl@163.com

Abstract: Automatic Modulation Recognition (AMR) is a key technology in the field of cognitive communication, playing a core role in many applications, especially in wireless security issues. Currently, deep learning (DL)-based AMR technology has achieved many research results, greatly promoting the development of AMR technology. However, the few-shot dilemma faced by DL-based AMR methods greatly limits their application in practical scenarios. Therefore, this paper endeavored to address the challenge of AMR with limited data and proposed a novel meta-learning method, the Multi-Level Comparison Relation Network with Class Reconstruction (MCRN-CR). Firstly, the method designs a structure of a multi-level comparison relation network, which involves embedding functions to output their feature maps hierarchically, comprehensively calculating the relation scores between query samples and support samples to determine the modulation category. Secondly, the embedding function integrates a reconstruction module, leveraging an autoencoder for support sample reconstruction, wherein the encoder serves dual purposes as the embedding mechanism. The training regimen incorporates a meta-learning paradigm, harmoniously combining classification and reconstruction losses to refine the model's performance. The experimental results on the RadioML2018 dataset show that our designed method can greatly alleviate the small sample problem in AMR and is superior to existing methods.

Keywords: few-shot learning; automatic modulation recognition; relation network; deep learning

1. Introduction

AMR is an important step between signal detection and demodulation [1]. As a key technology in the field of cognitive communication, AMR is a prerequisite for achieving efficient spectrum sensing and understanding, and it is widely used in wireless monitoring, signal interception and interference problems [2].

For a long time in the past, maximum likelihood theory and expert feature extraction were the main methods for solving AMR problems. Based on the maximum likelihood theory method, it is necessary to first obtain the statistical characteristics of the modulated signal, then construct a decision criterion and finally construct a maximum likelihood classifier. In theory, this type of method can bring high recognition accuracy, but its high computational complexity and strong dependence on prior knowledge limit its application in practical problems. The method based on expert features relies on expert knowledge in the field of signal processing, designing feature-extraction models that can distinguish different modulation signals manually, and then designing classifiers for recognition. The recognition accuracy of this type of method depends on the extracted statistical features and is limited by the weak learning ability of traditional classifiers [3].

In recent years, deep learning technology has shown excellent feature-extraction and data-analysis capabilities in fields such as image processing and natural language

processing, and it has been widely applied in wireless communication [4]. At the same time, it also provides a data-driven research approach for AMR. The AMR method based on deep learning utilizes a deep neural network model to learn data distribution from a large number of historical signal samples, extract more robust and effective signal features and use them to distinguish different modulation categories. Recent research has shown that signal modulation-recognition methods based on deep learning have higher recognition performance compared to traditional methods [5].

However, most DL-based AMR studies assume sufficient historical signal samples for training deep neural network models and assume that the sample categories are the same when training offline and deploying applications in practice. In practical scenarios, fulfilling this assumption is inherently difficult. Firstly, this is due to the dynamically complex electromagnetic environment, which constantly faces interference from a multitude of sources: various wireless communication devices, different electronic instruments and the ever-changing influences of natural phenomena, including weather fluctuations and terrain characteristics. This continuous change leads to rapid fluctuations in signal characteristics, making it difficult to obtain widely applicable and high-quality signal samples as the samples collected at one time may not represent the situation at other times. Secondly, the collection of high-quality signal samples often requires high-performance measurement equipment and professional data-collection techniques, which brings significant economic and operational burdens. The signal samples also require precise labels for model training, but obtaining accurate labels is also a challenging task. This not only requires a high level of professional knowledge but may also involve a lot of manual work, especially when facing diverse signal types generated by complex electromagnetic environments [6]. Therefore, in practical scenarios, AMR often faces the few-shot challenge, which greatly limits the implementation and widespread application of deep learning-based signal modulation-recognition methods in practical scenarios.

In the field of image recognition, the few-shot problem has also received sufficient attention, and methods such as data augmentation and meta-learning have been proposed to solve the few-shot dilemma. Among them, meta-learning is considered the most promising method for solving small sample problems [7]. Its idea is to continuously construct small sample training tasks on the source domain dataset so that the model can distinguish similarities and differences between different categories. Even when facing new small sample tasks in the target domain, it can still have good recognition performance. Metrics-based meta-learning methods, such as prototype networks (PNs) [8] and relation networks (RNs) [9], have been widely studied and applied due to their simple structure, ease of training and good performance.

Therefore, this paper proposes a few-shot AMR method based on a multi-level comparison class reconstruction relation network (MCRN-CR) based on the meta-learning paradigm, aiming to solve the few-shot problem faced in AMR.

The main contributions of this article are as follows:

- (1) A meta-learning paradigm is proposed for the FS-AMR problem. By utilizing existing source domain datasets and constructing meta-task training methods, the model can distinguish differences between different categories and can be well generalized to target domain classification tasks. Namely, it can attain satisfactory classification performance on the target domain with either no or minimal training samples from the target domain.
- (2) A multi-level embedding comparison relationship network has been proposed. Based on the relation network architecture in meta-learning, we propose an improved version, which involves multi-level measurement of the deep and shallow features output by the embedding function, synthesizing the relationship score of the test sample relative to the support category and using this to determine the category of the test sample. Compared to general relational networks, the comprehensive comparison of deep and shallow features can minimize the confusion of different modulation categories in the metric space as much as possible.

- (3) A class reconstruction meta-learning framework was proposed. The class reconstruction network consists of an encoder and a decoder, where the encoder serves as an embedding function and the decoder reconstructs the potential representation of the sample into the original signal. Furthermore, by combining class reconstruction loss and classification loss to jointly train the model, more robust feature representations can be learned, reducing the impact of noise on recognition performance.

The rest of this article is organized as follows. In the second section, relevant works on AMR are introduced, including traditional AMR methods, DL-based AMR methods and FS-AMR methods. In the third section, the relevant theories and methods of this article are introduced. The fourth section provides a detailed introduction to the multi-level comparison relation network proposed in this article, which combines class reconstruction. The fifth section provides detailed experimental results and discussion. Finally, a summary of the entire article was provided.

2. Related Works

This section focuses on the problem of AMR, introducing traditional AMR methods, DL-based methods and existing FS-AMR methods.

2.1. Traditional AMR Methods

2.1.1. Maximum Likelihood Theory-Based AMR Method

The modulation-recognition method based on maximum likelihood theory analyzes signal characteristics through probability theory, and its modulation-recognition process is a multi-hypothesis testing problem. This method is based on the likelihood function of electromagnetic signals, using the comparison results of the likelihood ratio and set threshold as the basis for judgment. Common methods include average likelihood ratio test (ALRT) [10], generalized likelihood ratio test (GLRT) [11] and mixed likelihood ratio test (HLRT) [12]. This type of method has the best signal feature performance in ideal channels, but it has high computational complexity and requires complete signal prior knowledge, making it susceptible to environmental interference factors.

2.1.2. Expert Features-Based AMR Method

Feature extraction is a process of data mapping, which involves mapping the original signal data to a specific feature space. The purpose of feature mapping is to generalize the differences between modulation categories, reduce data dimensionality, and improve algorithm efficiency. Common signal features include the following: (1) instantaneous features; (2) statistical characteristics; (3) transform domain features [6].

Instantaneous features describe the instantaneous properties of a signal in the time domain. It mainly considers the instantaneous change characteristics of the signal at a certain moment or period [13]. Common instantaneous characteristics include instantaneous amplitude, instantaneous frequency and phase. Statistical features are obtained through statistical analysis of signals. They describe the probability distribution, average level, dispersion degree and other characteristics of signals, mainly including high-order moments, high-order cumulants and cyclic cumulants [14,15]. The transformation domain feature is acquired through the mathematical transformation of a signal, such as using methods like the short-time Fourier transform or wavelet transform. These methods convert the signal from the time domain to the frequency domain or other transformation domains, revealing valuable information such as the frequency, phase and energy of the signal. The commonly used transform domain features mainly include wavelet transform features, spectral analysis features, time-frequency maps, constellation maps [16], etc.

Feature extraction is fundamental in expert feature-based methods and requires a high level of expertise. The effectiveness of features directly impacts the final classification performance. However, this method is limited by its reliance on a single feature, which constrains its data representation capability and, consequently, caps its performance potential.

Additionally, feature extraction and classifier design are conducted separately, resulting in low automation and high time costs.

2.2. DL-Based AMR Methods

The deep learning-based AMR method uses deep neural network models for feature extraction and classification. Compared with traditional methods, the deep learning-based method achieves higher recognition accuracy, stronger generalization ability and wider applicability, thus gaining widespread attention. From the perspective of network structure, currently, the types of deep neural networks (DNNs) used for designing modulation recognition include Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs) and their hybrid networks.

2.2.1. CNN-Based AMR Methods

O'SHEA et al. first proposed a modulation-recognition neural network model for directly processing raw In phase and Quadrature (I/Q) signals in 2016. Although this model only uses a simple CNN structure, its recognition performance far exceeds traditional artificial feature methods. At the same time, the author also opened up the RadioML2016.10a and RadioML2016.10b datasets, greatly promoting the development of this field. Furthermore, classic deep learning architectures such as VGG [17] and ResNet [18], which are widely used in the image field, have also been applied to AMR problems. O'SHEA released the RadioML2018.01A dataset in 2018 and compared the classification performance of two types of networks for AMR. The experiment proved that the ResNet structure with skip connections could bring better recognition accuracy under a high signal-to-noise ratio. In addition, the author also compared the recognition performance of the model with different convolutional layers in VGG and different residual block numbers in ResNet and pointed out that a too-deep network architecture not only cannot bring higher recognition performance [19] but also can bring higher complexity and computational complexity, leading to overfitting of the model during training.

In addition to processing I/Q signals directly, the original I/Q data is typically pre-processed and transformed into other forms of features. Corresponding DNN models are then developed for deep feature extraction and recognition. Constellations [20], spectrograms [21], as well as cyclic spectrograms [22], eye charts [23], amplitude histograms [24], etc., are all used as data preprocessing forms in AMR problems. In contrast, the DL-AMR method with preprocessed form has stronger comprehensibility, but there is additional computational complexity in the process of converting I/Q signals into images. The DL-AMR method, which directly processes I/Q signals, does not require additional preprocessing processes, requires less computation and has a higher degree of automation.

2.2.2. RNN-Based AMR Methods

Note that the modulated signal is a time series, and CNNs typically struggle to capture temporal features well, so RNN-based AMR methods are gradually being proposed. A new AMR method based on RNN is proposed in [25], which uses gated recurrent units (GRUs) to achieve better recognition accuracy than some CNN models. The author of reference [26] converted the I/Q signal into amplitude and phase (A/P) and input it into LSTM, achieving high recognition accuracy.

2.2.3. Hybrid AMR Method of CNN and RNN

Based on a single CNN or RNN network architecture, it is difficult to simultaneously focus on the spatial or temporal characteristics of wireless signals. Using only one type may not achieve optimal performance. Therefore, some scholars have proposed an AMR model that combines the two network architectures. A Convolutional Long Short Term Deep Neural Network (CLDNN) model is proposed in [26], consisting of one LSTM and three CNN layers. This model has a skip connection before LSTM, bypassing two CNN layers, aiming to provide longer temporal correlation information for the extracted features.

Reference [27] proposes another CLDNN2 model without bypass layer connections, which can achieve higher recognition accuracy than CLDNN at the cost of increasing the number of layers and parameters. Inspired by the excellent feature-extraction characteristics of hybrid models and the complementary information present in individual channels, reference [28] proposes a new multi-channel deep learning model (MCLDNN), which utilizes single and combined I/Q symbols of received data to extract features from a spatiotemporal perspective, achieving high recognition performance.

2.2.4. Other Network Architecture-Based AMR Method

In addition to CNN and RNN, popular network structures in deep learning such as Complex Value Neural Networks (CVNN), Graph Neural Networks (GNN) and Transformer have also been used in the study of AMR problems. Compared with traditional real neural networks, CVNN can simultaneously process data containing both real and imaginary parts, providing richer and more diverse feature-expression capabilities than real neural networks [29]. Therefore, CVNN is very suitable for processing periodic signals. Reference [30] proposed a novel AMR model based on CVNN and found that complex-valued operations are more effective in helping neural networks extract correct statistical information than real-valued operations at low signal-to-noise ratios. To reduce the additional computational cost of complex convolution, reference [31] proposes an AMC model based on complex depthwise separable convolution (CDSCNN) to achieve a balance between classification accuracy and model complexity. When the essence of a problem can be represented by a graph structure, a graph neural network (GNN) can be used to solve the problem [32]. Reference [33] first proposed an automatic modulation-recognition model based on Graph Convolutional Neural Network (GCN). Transformer [34] has also been proven to be useful for solving AMR tasks. For example, the Mcformer model in reference [35] achieved good classification performance by extracting global information from time-domain signals through the Transformer.

2.3. FS-AMR Methods

Based on different research approaches to solving few-shot problems, we divide them into data-driven few-shot learning methods and model-driven few-shot learning methods.

2.3.1. Data-Driven FS-AMR Methods

The data-driven approach mainly focuses on how to enhance or expand existing small training samples and improve the learning performance of the model under few-shot conditions. When the labeled samples are insufficient, on the one hand, simple data transformations such as cropping and concatenation, and adding noise can be used to expand the training samples. On the other hand, Generative Adversarial Networks (GAN) can be used to generate virtual data. Based on the characteristics of electromagnetic signal data, reference [36] analyzed and applied several simple and effective time-series enhancement methods that conform to the characteristics of electromagnetic signals, including noise disturbance, amplitude and time delay transformation, frequency shift and phase shift. Finally, the effectiveness of the proposed simple data-augmentation method in small sample radiation signal-recognition tasks was verified through experiments. The author of reference [37] used GAN to generate pseudo data similar to the original signal and reversed the original signal and false signal to further increase sample diversity. Finally, the effectiveness of data augmentation based on the GAN method was verified in a CNN network model.

2.3.2. Model-Driven FS-AMR Methods

This type of method focuses on designing and optimizing model structures, learning strategies or optimization algorithms, enabling them to have inherent few-shot learning capabilities and efficiently learn and generalize under limited data. Meta-learning is a commonly used learning framework for such methods, which enables machines to learn

to learn and distinguish similarities and differences between things. Metrics-based meta-learning, such as prototype networks [8] and relational networks [9], is a simple and efficient method for solving few-shot problems. Reference [38] proposes a novel small sample learning framework called Attention Relationship Network (AMCRN), which introduces channel and spatial attention to learn more effective feature representations of supporting samples and uses the relationship network as the learning architecture. The experimental results show that even with only one supporting sample, this method can achieve excellent performance in fine-grained signal modulation recognition and is robust to low signal-to-noise ratio conditions. References [39,40] also proposed corresponding small sample modulation-recognition frameworks based on relational networks. Both use one-dimensional convolution and denoising autoencoder as embedding functions for feature extraction. The experimental results show that compared with traditional FSL algorithms, the method based on relational networks can achieve higher classification accuracy in signal modulation recognition. In addition, reference [41] also proposed the AMR-CapsNet method based on capsule networks to solve the problem of few-shot AMR and achieved good classification results.

3. Signal Model and Basic Theory

3.1. Signal Model

AMR is a prerequisite for signal demodulation. The module for processing modulation classification is usually deployed at the receiving end of the entire communication system, sending the received modulation information to the demodulator, which then demodulates it.

In wireless communication systems, the received signal can be represented as [42]:

$$X(t) = h(t) \times s(t) + n(t), \quad (1)$$

where $h(t)$ represents the wireless channel pulse response, $s(t)$ represents the modulated signal generated by the transmitter, and $n(t)$ represents the most basic additive Gaussian white noise (AWGN).

After orthogonal sampling by the receiver, the same I/Q signal $X(n)$ is obtained:

$$X(N) = \left[\left(x_1^I, x_1^Q \right), \left(x_2^I, x_2^Q \right), \dots, \left(x_N^I, x_N^Q \right) \right], \quad (2)$$

where N is the sampling length. Therefore, when I/Q signals are used as network inputs, their data format can be written as:

$$X_{\text{input}} = \begin{bmatrix} x_1^I & x_2^I & \dots & x_N^I \\ x_1^Q & x_2^Q & \dots & x_N^Q \end{bmatrix}. \quad (3)$$

AMR can be described as determining the modulation method of $s(t)$ by observing the signal $X(t)$.

3.2. Description of DL-Based AMR Methods

In DL-based AMR tasks, the collected historical signals are usually used as training data, and known modulation methods are used as labels to train deep neural networks and obtain recognition models.

Given the training sample set

$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}, m = 1, 2, \dots, M, \quad (4)$$

Among them, x_m is the m -th sample data, y_m is the true label of the m -th sample, with a value range of label number, $y_m \in (0, 1, 2, \dots, L - 1)$. The training task is to obtain the

optimal parameters of the given model by learning the training sample set under the given optimization objective conditions:

$$\theta = \operatorname{argmin}_{(x_m, y_m) \in D_{m=1}} \sum_{m=1}^M \operatorname{loss}(\hat{y}_m, y_m) = \operatorname{argmin}_{(x_m, y_m) \in D_{m=1}} \sum_{m=1}^M \operatorname{loss}(\operatorname{Model}_{\theta}(x_m), y_m). \quad (5)$$

Then, by applying the trained model parameters θ , substituting them into the Model and testing the test sample set, we can infer the category of the test samples and obtain a measure of the Model's generalization ability.

However, the actual signal environment is very complex, and various electromagnetic activities interact with each other, making it difficult to obtain sufficient high-quality labeled samples. Therefore, the training samples cannot accurately represent the data distribution, which makes it difficult for deep neural network models to learn appropriate parameters. This results in significant classification model errors and poor generalization ability. This is the few-shot dilemma faced by AMR.

3.3. Metrics-Based Meta-Learning

Meta-learning is an important method for solving few-shot problems. Unlike traditional deep learning methods, its purpose is to enable machines to learn to learn and have the ability to distinguish similarities and differences between things, rather than learning a classification model specific to a specific task. We assume that the training set $D_{m\text{-train}}$ in the source domain and the testing set $D_{m\text{-test}}$ in the target domain have different categories. The basic idea of meta-learning is to construct meta-tasks on the source domain training set, and through training on different meta-tasks, make the machine have good generalization ability when facing different classification tasks, and thus have good recognition performance on target domain testing tasks.

Metrics-based meta-learning consists of an embedding function and a metric function defined in the representation space. As shown in Figure 1, the embedding function is used to extract prototype features of different categories, and the measurement function is used to measure the similarity between the test sample and different categories. The higher the similarity, the closer it is to that category. In meta-learning, one training iteration is called an Episodic, which is a C -Way K -shot classification task. C -Way K -shot task extraction refers to randomly selecting C categories from the training set, with K samples taken from each category to form a support set, and then extracting Q samples as a query set. The category labels of the support set are known and are used to construct the embedding benchmark of category C in the metric space, while the query set is the sample to be predicted.

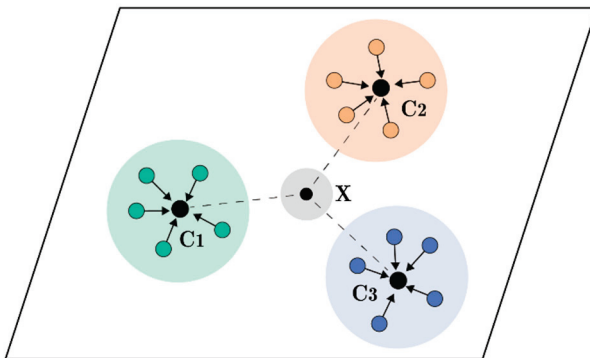


Figure 1. Schematic diagram of measurement space.

Figure 2 is a schematic diagram of signal modulation recognition using a relational network in a 5-way 1-shot scenario. The embedding function and relation metric function of the relational network are f_{θ} and g_{ϕ} , respectively. Assuming the support sample is x_i and

the query sample is x_j , the embedding function calculates the embeddings $f_\theta(x_i)$ and $f_\theta(x_j)$, and then concatenates the two feature maps:

$$Z(f_\theta(x_i), f_\theta(x_j)), \quad (6)$$

and feed the results into the relationship metric function to calculate the relation score, which is the similarity score:

$$r_{i,j} = g_\phi(Z(f_\theta(x_i), f_\theta(x_j))). \quad (7)$$

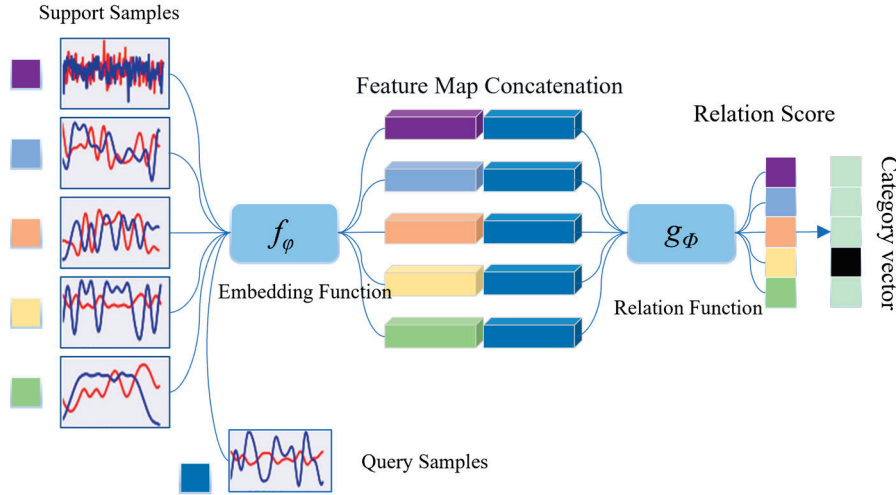


Figure 2. AMR method based on relational networks.

Finally, the category of the query sample will be determined based on the relation score. During the training process, minimize the objective function to seek the optimal solution θ related to ϕ , which enables the network to optimize task modeling, and quickly converge when accepting new small sample tasks.

4. The Proposed Method

This section will introduce the AMR method based on MCRN-CR in three parts: (1) the framework of MCRN-CR, (2) the details of MCRN-CR and (3) the training steps.

4.1. The Framework of MCRN-CR

To solve the few-shot dilemma in AMR, this paper proposes a class reconstruction relation network based on multi-level comparison (MCRN-CR), whose framework structure is shown in Figure 3. MCRN-CR consists of two parts: class reconstruction and multi-level comparison relation network. The class reconstruction part is used to generate low-dimensional latent representations of input samples, which is the embedding of support samples and query samples. The multi-level comparison relation network is an improved version of the relation network, used to achieve recognition tasks under few-shot conditions. It should be noted that Figure 3 is a schematic diagram. Although there are only 4-levels of MCRN-CR in Figure 3, we can construct models at any level. In this article, we construct a 5-level MCRN-CR.

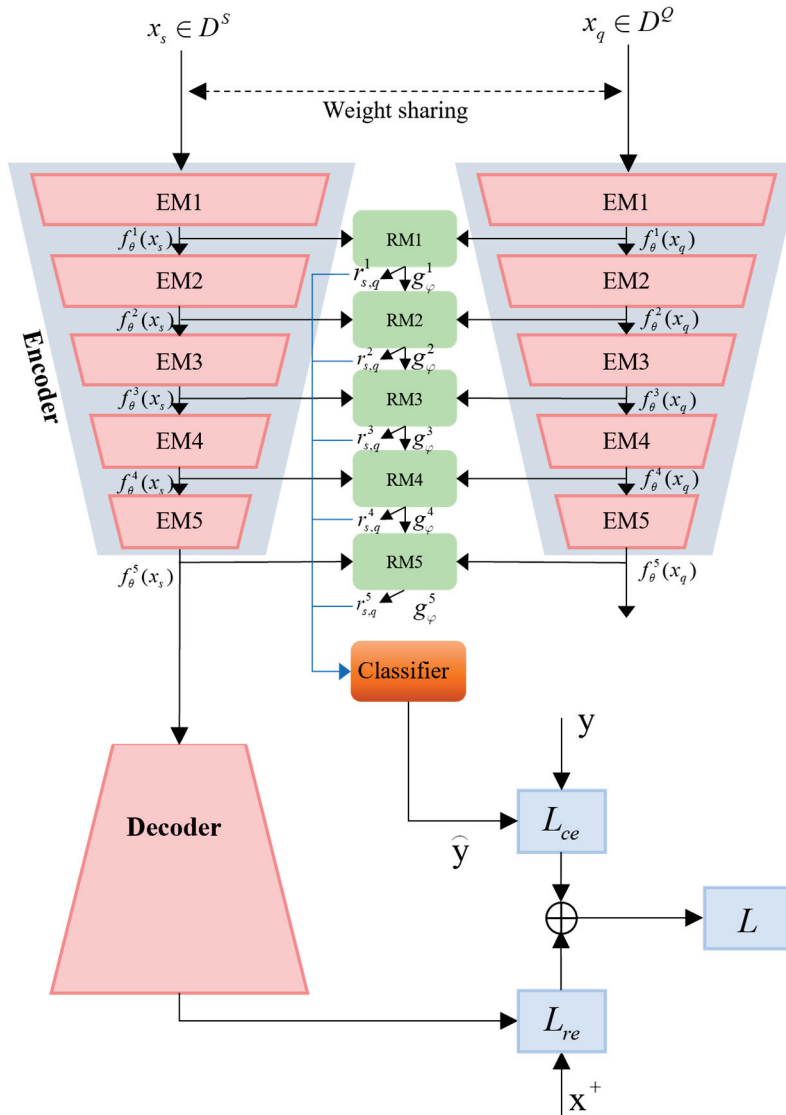


Figure 3. The overall framework diagram of the proposed MCRN-CR.

During the training process, the class reconstruction part first generates a low dimensional latent representation z_s of the support sample x_s . On the one hand, z_s participates in the similarity calculation between query samples. On the other hand, it needs to be further fed into the decoder to generate the reconstruction support sample $\bar{x}_s$. To eliminate the influence of noise, when further calculating the reconstruction loss L_{re} , a high signal-to-noise ratio sample x_s^+ with the same category as the supporting sample is used as the reconstruction target,

$$L_{re} = \mathbb{E} \left(\|\bar{x}_s - x_s^+\|_2^2 \right). \quad (8)$$

Then, the classifier in the MCRN part predicts the class labels of the query samples and calculates the cross-entropy classification loss with the real labels:

$$L_{ce} = -\mathbb{E}(y \log \hat{y}). \quad (9)$$

The class reconstruction part helps the encoder as an embedding function extract more recognizable sample features, increase the differences between different class embeddings and thus improve signal-recognition performance under few-shot conditions. In MCRN-CR, L_{re} and L_{ce} together constitute the model's loss, with coefficients of λ_{ce} and λ_{re} . We

believe that classification loss is more important for model training in the classification task of this article; therefore, λ_{ce} and λ_{re} is set to 0.8 and 0.2, respectively.

$$L = \lambda_{ce}L_{ce} + \lambda_{re}L_{re}. \quad (10)$$

4.2. The Details of MCRN-CR

4.2.1. The Structure of Improved Relational Networks MCRN

We believe that in order to match query samples and support categories accurately, it is necessary to learn nonlinear distances at different feature levels on both deep and shallow abstract features and comprehensively calculate relation scores. Therefore, in response to AMR, this paper proposes a multi-level comparison relation network, as shown in Figure 3. The multi-level comparison relation network includes an embedding module and a relation-measurement module, where the embedding module is composed of the encoder of the class reconstruction part.

Specifically, the support sample and query sample are, respectively, embedded to generate the v -th level feature map $z_i^v = f_\theta^v(x_i)$ and $z_j^v = f_\theta^v(x_j)$, then concatenate it into $[z_i^v, z_j^v]$ and send it to the corresponding v -level relation module for comparison.

At the $v - 1$ level, the relation module outputs the similarity feature maps of samples x_i and x_j . The v -th level relation module takes both the embedding outputs of the v -th level support and query samples as inputs and the similarity feature maps of the $v - 1$ level relation module as inputs:

$$g_\varphi^v = g\left([z_i^v, z_j^v, g_\varphi^{v-1}]\right). \quad (11)$$

For the first level relationship module, since it does not have the input of the previous one, it is assumed that

$$g_\varphi^1 = g\left([z_i^1, z_j^1]\right). \quad (12)$$

Assuming $q(\cdot)$ represents average pooling and fully connected operations, the similarity (relationship) score between the support samples and query samples output by each relationship module in the feature map at level v is:

$$r_{i,j}^v = q\left(g_\varphi^v\right). \quad (13)$$

The similarity score between the further query sample x_j and each supporting category y_c is:

$$r_{c,j} = \sum_{v=1}^V w_j^v r_{c,j}^v, \quad c = 1, 2, \dots, C. \quad (14)$$

Among them, $w_{c,j}^v = \alpha^v(g_{c,j}^v)$ represents the scalar attention weight of the relation scores at all levels, α^v is a fully connected layer, the activation function is sigmoid and the weight parameters of α^v are included in the relationship module.

4.2.2. The Class Refactoring in MCRN-CR

The class reconstruction part includes two parts: the encoder and the decoder. The encoder generates the potential feature-expression z_s of the supporting sample x_s , and the decoder further reconstructs the supporting sample $\bar{x}_s$. After training, the smaller the reconstruction error, the better z_s can represent the sample features. Since complex neural networks can handle data containing both real and imaginary parts, providing richer and more diverse feature-expression capabilities than real neural networks, this paper adopts complex convolution as the basic unit of encoders and decoders. As shown in Figure 4, the structure of the encoder and decoder includes five layers of complex convolution.

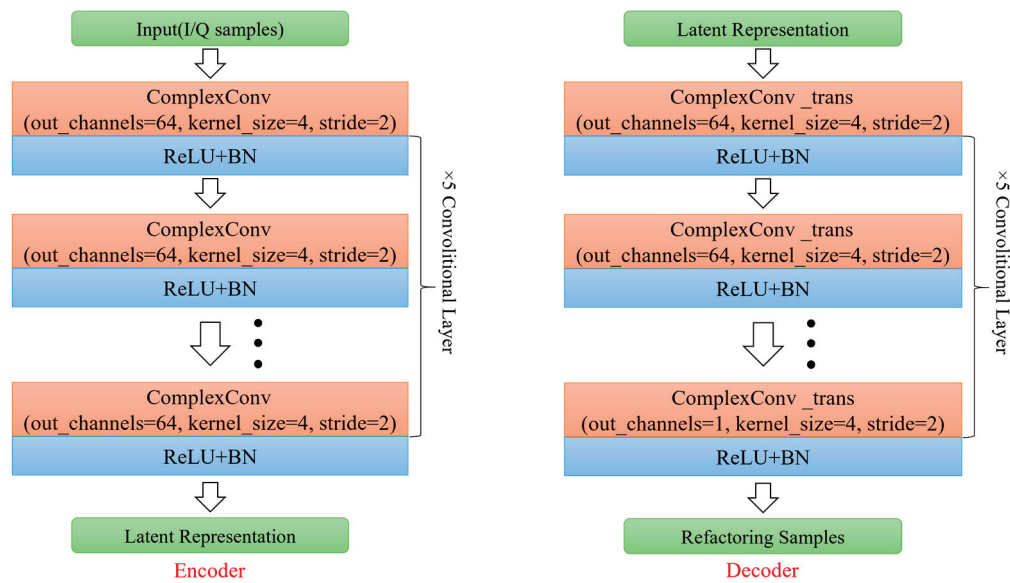


Figure 4. Structure of encoder and decoder.

4.2.3. Relation Metric Module

The relation metric module calculates the nonlinear distance between query samples and support categories, obtains the relationship score between each sample and different category families and further determines the category of query samples. Table 1 shows the specific structure of the relationship-measurement module.

Table 1. Structure of relationship module.

Layer	Structure
RM1	Conv1d (out_channels = 64, kernel_size = 4, stride = 2) ReLU + BN
RM2	Conv1d (out_channels = 64, kernel_size = 4, stride = 2) ReLU + BN
RM3	Conv1d (out_channels = 64, kernel_size = 4, stride = 2) ReLU + BN
RM4	Conv1d (out_channels = 64, kernel_size = 4, stride = 2) ReLU + BN
RM5	Conv1d (out_channels = 64, kernel_size = 4, stride = 2) ReLU + BN

4.3. Training Procedure

During training, C-way K-shot tasks will be extracted from the training set for each iteration, obtaining the support set $D_S = \{(x_i, y_i)\}_{i=1}^m$ and query set $D_Q = \{(x_j, y_j)\}_{j=1}^n$ for each training task, where $m = K \times C$ and $n = Q \times C$.

Subsequently, according to Section 3.2, calculate the relationship score $r_{c,j}$ between each query sample x_j and each support category y_c , and further determine the predicted label $\hat{y}$ for the query sample. Secondly, the hidden layer feature $\mathbf{z}_s$ of the support sample calculated by the encoder is further fed into the decoder to reconstruct the support sample $\bar{x}_s$. Then, based on the predicted labels and reconstructed samples, the classification error and reconstruction error are settled separately. Finally, the error is backpropagated, and the encoder and decoder parameters of the class reconstruction part and the relationship module parameters are updated. Algorithm 1 describes the complete training process of MCRN-CR.

Algorithm 1: Training process of AMR model based on MCRN-CR.**Symbol Description:**

- E: The number of training iterations, the size of Episodic;
- $\Theta, \theta_{de}, \varphi$: Parameters are encoder, decoder and relationship module, respectively;
- $lr_{en}, lr_{de}, lr_{rm}$: Learning rates of the encoder, decoder and relation module, respectively;
- $\lambda_{ce}, \lambda_{re}$: a scalar used to balance the overall training loss.

Training iteration:

1: **For** Episodic = 0 to E − 1:

2: Task Extraction from $D_{m-train}$, get Support set D_S and Query set D_Q ;

3: $x^+ \leftarrow$ The mean value of the samples corresponding to the category of all supporting samples in the current task at the highest SNR;

Forward propagation:

4: Calculate the levels of embeddings z_i^v and z_j^v for support samples and query samples:

$$z_i^v = f_\theta^v(x_i), \quad z_j^v = f_\theta^v(x_j);$$

5: Calculate support category feature embeddings:

$z_c^v \leftarrow$ The mean embedding of K supporting samples in each category;

6: Calculate the relationship score between query samples and support categories:

$$r_{c,j} = \sum_{v=1}^V w_j^v r_{c,j}^v, \quad c = 1, 2, \dots, C;$$

7: Category labels for predicted query samples:

$$\hat{y} = \operatorname{argmax}(r_{c,j});$$

8: Calculate the output of the decoder:

$$\bar{x} = f(\theta_{de}, x_s);$$

9: Calculate losses:

$$\begin{aligned} L_{ce} &= L_{ce}(\hat{y}, y), \\ L_{re} &= L_{re}(\bar{x}, x^+), \\ L &= \lambda_{ce} L_{ce} + \lambda_{re} L_{re}; \end{aligned}$$

10: Error backpropagation, updating parameters:

$$\begin{aligned} \theta &\leftarrow \operatorname{Adam}(\nabla_\theta, L, lr_{en}, \theta) \\ \theta_{de} &\leftarrow \operatorname{Adam}(\nabla_{\theta_{de}}, L, lr_{de}, \theta_{de}); \\ \varphi &\leftarrow \operatorname{Adam}(\nabla_\varphi, L, lr_{rm}, \varphi) \end{aligned}$$

11: **End for.**

5. Experiment and Discussion

5.1. Dataset and Experimental Setup

To verify and evaluate the effectiveness of the proposed method, experiments were conducted on the RML2018.01A modulated signal dataset in this paper. This dataset consists of 24 types of modulation signals, with each signal sample containing two channels of I/Q data in the format of $[1024, 2]1024 \times 2$, with a signal-to-noise ratio range of -20 dB~ 30 dB and an interval of 2 dB. A single category has 4096 samples at each signal-to-noise ratio, totaling 2,555,904 pieces of data.

Firstly, divide the dataset into 14 modulation categories as known source domain training sets and the other 10 as target domain datasets. Due to the large size of the original dataset, it is redundant for the experiment in this article. Therefore, only 1000 samples were taken for each category under a single signal-to-noise ratio. The existing training samples in the target domain are usually scarce in few-shot scenarios. Therefore, to comply with this setting, a subset of the target domain dataset will be divided as a known small sample dataset, with a quantity of 2% of the target domain samples. For each signal-to-noise ratio, 20 samples will be taken for a single modulation method. During each testing process, the support set is extracted from that subset. The remaining part is the test set, from which query samples are extracted during each test. The specific division is shown in Table 2.

Table 2. Dataset partitioning.

Dataset Name		Modulation Category	Sample Quantity
Source domain	Training set	128QAM, 32PSK, 16APSK, 32QAM, FM, GMSK, 32APSK, 64QAM, BPSK, 8PSK, AM-SSB-SC, 4ASK, 16PSK, 64APSK	$14 \times 24 \times 1000$
	Support subset	128APSK, AM-DSB-SC, AM-SSB-WC, QPSK, 256QAM, AM-DSB-WC, OOK, 16QAM, OQPSK, 8ASK	$10 \times 24 \times 20$
Target domain	Test set	128APSK, AM-DSB-SC, AM-SSB-WC, QPSK, 256QAM, AM-DSB-WC, OOK, 16QAM, OQPSK, 8ASK	$10 \times 24 \times 980$

The hardware environment for the experiment was as follows: Intel (R) Core (TM) i7-10700k CPU@3.8GHz NVIDIA GeForce GTX3090 GPU. The software environment was as follows: Python 3.8, PyTorch deep learning framework.

5.2. Comparison of Recognition Performance with Other Models

To verify the effectiveness and superiority of the proposed method under few-shot conditions, we compared the recognition performance of MCRN-CR with other networks under the 5-way 5-shot setting. The compared models include the most recent FS-AMR methods, AMCRN [40], AMR-CapsNet [42], Prototype Networks (PNs) [8] and Relation Networks (RNs) [9] for few-shot learning methods in the image domain. At the same time, in order to demonstrate that the proposed method can solve the small sample dilemma faced by traditional deep learning methods, a complex neural network-based automatic modulation-recognition model AMR-CVNN is formed by adding a classifier to the structure of the encoder mentioned above, and it is used as a comparison term. To ensure the fairness of the experiment, AMR-CapsNet and AMCRN maintain the same settings as the original text. The laboratory settings for PNs and RNs are consistent with MCRN-CR. AMR-CVNN adopts the idea of transfer learning, first pre-training on the source domain dataset and then fine-tuning the weight parameters on the target domain dataset.

Figure 5 shows the recognition accuracy of all comparison models on a given dataset under all SNRs. It shows that when the SNR is greater than 0 dB, the proposed method MCRN-CR has the best recognition performance. When the SNR is below 0 dB, the recognition performance of all metric-based meta-learning methods is not significantly different, but the recognition performance of MCRN-CR is much higher than that of AMR CapsNet and AMR CVNN.

Table 3 presents more specific identification performance statistics for each comparative model. The highest recognition accuracy of MCRN-CR reached 89.25%. Under all SNRs, the average recognition accuracy of MCRN-CR reached 65.98%. When SNR > 0 dB, the average recognition accuracy reached 87.41%. Compared to AMR-CVNN, the average recognition accuracy of MCRN-CR has increased by nearly 10% below 0 dB, 5.86% above 0 dB and 8.04% above all signal-to-noise ratios. Therefore, it can be proven that the MCRN-CR proposed in this article can effectively solve the small sample dilemma faced by traditional deep learning methods in AMR problems.

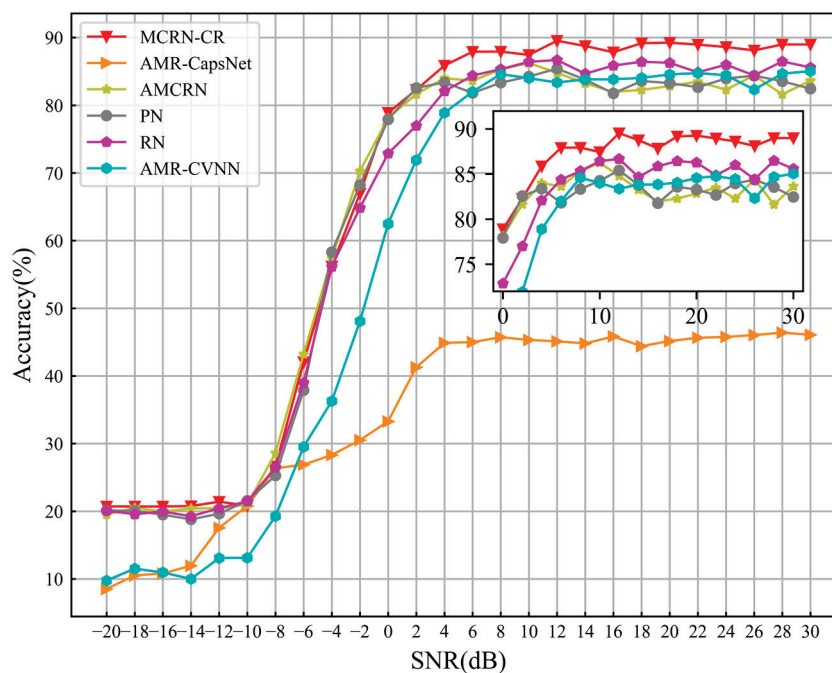


Figure 5. The recognition accuracy of different models at all SNRs.

Table 3. Comparison of average recognition accuracy of different methods.

Model	−20:2:−2 dB	0:2:30 dB	−20:2:30 dB	Highest Accuracy
MCRN-CR	31.69%	87.41%	65.98%	89.25%
AMR-CapsNet [42]	19.22%	44.42%	34.73%	46.40%
AMCRN [40]	32.13%	83.09%	63.49%	86.25%
PN [8]	30.93%	83.00%	62.98%	85.41%
RN [9]	30.72%	84.08%	63.56%	86.68%
AMR-CVNN	20.16%	81.55%	57.94%	85.04%

Figure 6 shows the confusion matrix of all models at the 12 dB SNR. The horizontal axis of the confusion matrix represents the actual modulation category, and the vertical axis represents the modulation category predicted by the model. It can be seen that when distinguishing between high-order digital modulation signals 16QAM and 256QAM, as well as analog signals AM-DSB-SC and AM-DSB-WC, there is still some confusion in MCRN-CR, which is the main reason for the bottleneck in model-recognition performance. The same applies to models other than MCRN-CR. Therefore, we believe that the proposed AMR method based on the meta-learning paradigm can solve the problem of small samples, but for some easily confused signals, it is necessary to design a feature-extraction network with stronger feature-extraction capabilities to extract finer features and achieve the goal of distinguishing confused signals.

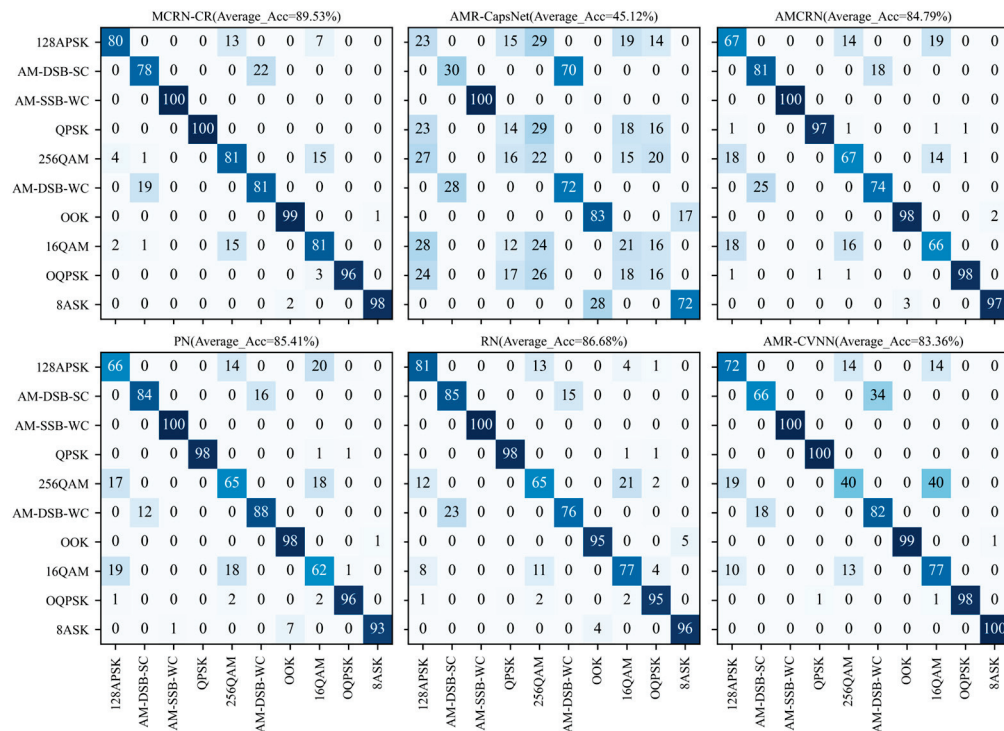


Figure 6. Confusion matrix diagram at the 12 dB SNR.

5.3. Structural Effectiveness Analysis

This section analyzes the impact of multi-level comparison operations and the introduction of class reconstruction modules in the proposed framework. Therefore, four control groups were set up: MCRN-CR, MCRN, RN-CR and RN. Among them, MCRN, RN-CR and RN, respectively, remove the class reconstruction part, the multi-level comparison part and the class reconstruction part and multi-level comparison part.

Figure 7 shows the recognition performance curves of four control groups at all SNRs, while Table 4 presents further experimental results and statistical data. It can be seen that with the addition of class reconstruction modules and multi-level comparison operations on the basis of the original RN model, the recognition performance has also been correspondingly improved. Figure 8 shows the confusion matrix of four control models at the 12 dB SNR. The figure shows that the basic model RN has more severe confusion for modulation signals 16QAM and 256QAM, as well as modulation signals AM-DSB-SC and AM-DSB-WC. The addition of class reconstruction modules and multi-level comparison operations has somewhat alleviated their confusion. This proves that our design concept is correct and effective, the addition of class reconstruction modules enhances the ability to accurately express potential features of samples, while multi-level comparison operations can make comprehensive decisions on the abstract expression of samples at different levels, thereby improving the decision-classification ability of the model.

Table 4. Comparison of average recognition accuracy of different control models.

Model	−20:2:−2 dB	0:2:30 dB	−20:2:30 dB	Highest Accuracy
MCRN-CR	31.69%	87.41%	65.98%	89.25%
MCRN	31.42%	86.24%	65.15%	88.15%
RN-CR	31.02%	85.16%	64.34%	87.80%
RN	30.72%	84.08%	63.56%	86.43%

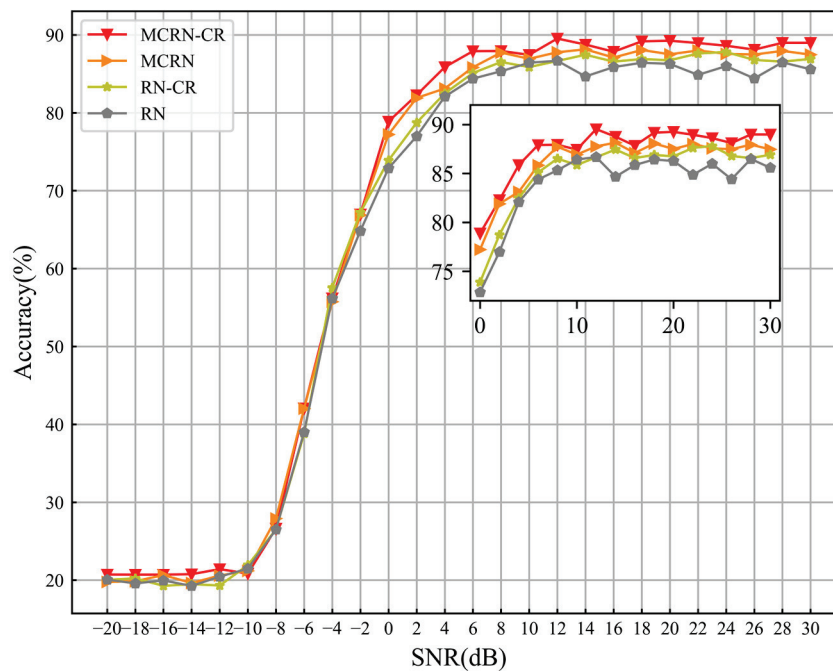


Figure 7. The recognition accuracy of the comparison model at all SNRs.

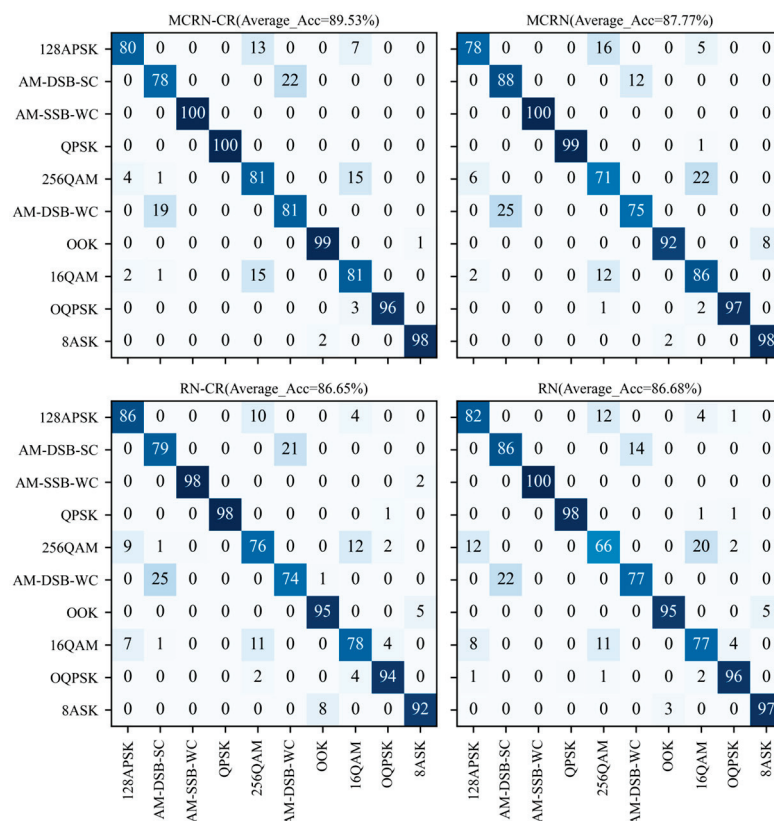


Figure 8. Confusion matrix diagram of the control model at the 12 dB SNR.

5.4. Comparison of Recognition Performance under Different C-Way K-Shot Settings

This section conducted experimental exploration of different values of C and K to further verify the impact of the number of categories C extracted during the construction of the meta task and the number of samples K extracted for each category on the recognition accuracy of the algorithm proposed in this paper.

Firstly, we fix $C = 5$ and then take $K = 1, 5, 10, 15$ and 20 for experiments. As shown in Figure 9, the recognition performance of the model varies with the SNR under different sample sizes. From the graph, it can be seen that as the K value increases, the recognition performance of the model also improves further. According to the aforementioned theory, the method proposed in this article determines its category by comparing the relation scores between query samples and category features in the metric space. Therefore, changing the K value for training and testing will definitely impact recognition performance. More specifically, the metric-based theory suggests that each category has a feature embedding in the metric space that is far from each other. The mean of each category's support sample feature vectors is used in the relation network as the class feature embedding. As the K value increases, the estimation of class feature embedding becomes more accurate, resulting in smaller recognition errors when further determining the category of the test sample, thereby improving the overall recognition performance. When the value of K increases from 1 to 5, there is a significant improvement in recognition performance. However, when $K > 5$, further increasing the value of K has a more gradual effect on recognition performance. Specifically, when K increased from 1 to 5, the average recognition accuracy of the model increased from 61.54% to 65.98% and the performance improved by 4.44%. When K increased from 5 to 20, the average recognition accuracy of the model increased from 65.98% to 66.86%, and the recognition performance only improved by 0.88%. This also indicates that the algorithm proposed in this article does not require many supporting samples, but it can still achieve excellent recognition performance.

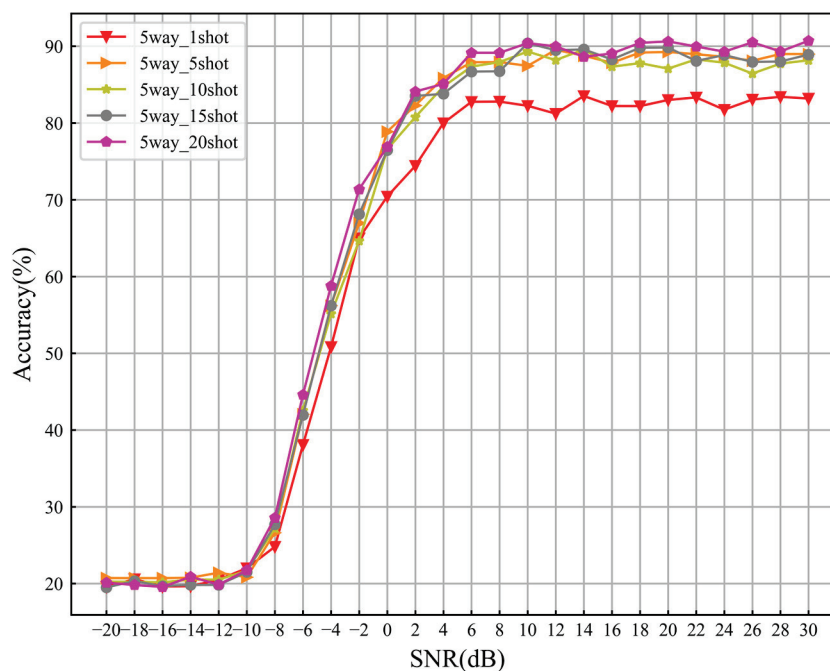


Figure 9. Recognition accuracy curves of models under different K values.

Secondly, fix $K = 5$ and conduct experiments with $C = 3, 5$ and 10 to verify the impact of the number of supporting sample categories C on recognition performance. Figure 10 shows the model-recognition accuracy variation curve with SNRs for different C values. It can be seen from the figure that as the C value increases, the overall recognition performance of the model decreases. When C increased from 5 to 10, the overall recognition accuracy of the model decreased from 65.98% to 50.70%, and the recognition performance decreased by 15.28%. We believe that as the C value increases, i.e., the number of supporting categories increases, the decision space for model category inference will become larger. On the one hand, this increases the difficulty of learning the model. On the other hand, it increases the difficulty of judging the relation between query samples and supporting categories in the metric space, leading to a decrease in the overall recognition performance of the model.

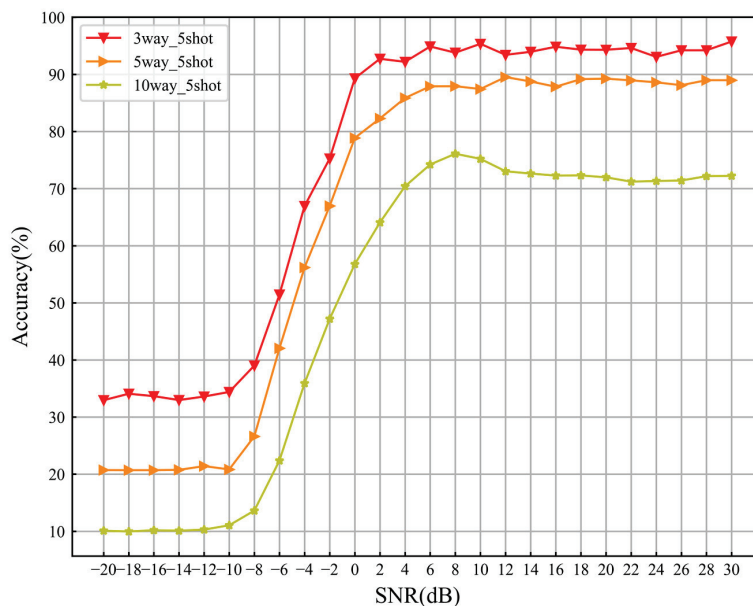


Figure 10. Recognition accuracy curves of models under different C values.

6. Conclusions

This article proposes a novel meta-learning method called Multi-level Comparison Relationship Network (MCRN-CR) with class reconstruction to address the few-shot dilemma in AMR. In this method, a multi-level comparison relation network structure is designed, which outputs feature maps through embedding functions, comprehensively calculates the relationship scores between query samples and support samples and determines the modulation category. At the same time, a class reconstruction module is introduced for the embedding function in the network. An autoencoder is used to reconstruct the supporting samples, and its encoder is used as the embedding function. The meta-learning training paradigm is used to train the model, combining classification error and reconstruction error. The experimental results on the RadioML2018 dataset show that MCRN-CR can significantly alleviate the small sample problem in AMR, and its performance is superior to existing methods. This study provides new ideas for solving AMR problems under few-shot conditions and has important application value in related fields such as wireless security. However, it should be acknowledged that in the experiment of this article, the category spaces of the source domain and the target domain are both small. Therefore, in future work, we will collect more diverse training samples from the source domain to support the model's training and achieve good recognition performance on the target domain test set with a larger category space.

Author Contributions: Conceptualization, Z.M. and Y.F.; methodology, S.F.; software, Z.M. and S.H.; validation, Z.M. and Z.X.; formal analysis, Z.X. and S.H.; resources, S.F.; data curation, Z.M. and S.H.; writing—original draft preparation, Z.M.; writing—review and editing, Z.M. and S.H.; supervision, Z.X. and Y.F.; funding acquisition, S.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Key Basic Research Projects of the Basic Strengthening Program, grant number 2020-JCJQ-ZD-071.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Wang, Y.; Gui, G.; Gacanin, H.; Ohtsuki, T.; Sari, H.; Adachi, F. Transfer Learning for Semi-Supervised Automatic Modulation Classification in ZF-MIMO Systems. *IEEE J. Emerg. Sel. Top. Circuits Syst.* **2020**, *10*, 231–239. [CrossRef]
- Meng, F.; Chen, P.; Wu, L.; Wang, X. Automatic Modulation Classification: A Deep Learning Enabled Approach. *IEEE Trans. Veh. Technol.* **2018**, *67*, 10760–10772. [CrossRef]
- Shi, J.; Hong, S.; Cai, C.; Wang, Y.; Huang, H.; Gui, G. Deep Learning-Based Automatic Modulation Recognition Method in the Presence of Phase Offset. *IEEE Access* **2020**, *8*, 42841–42847. [CrossRef]
- Rajendran, S.; Meert, W.; Giustiniano, D.; Lenders, V.; Pollin, S. Deep Learning Models for Wireless Signal Classification with Distributed Low-Cost Spectrum Sensors. *IEEE Trans. Cogn. Commun. Netw.* **2018**, *4*, 433–445. [CrossRef]
- O’Shea, T.J.; Corgan, J.; Clancy, T.C. Convolutional radio modulation recognition networks. In Proceedings of the Engineering Applications of Neural Networks: 17th International Conference, EANN 2016, Aberdeen, UK, 2–5 September 2016; Proceedings 17; pp. 213–226.
- Xiao, W.; Luo, Z.; Hu, Q. A Review of Research on Signal Modulation Recognition Based on Deep Learning. *Electronics* **2022**, *11*, 2764. [CrossRef]
- Hospedales, T.; Antoniou, A.; Micaelli, P.; Storkey, A. Meta-Learning in Neural Networks: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 5149–5169. [CrossRef]
- Snell, J.; Swersky, K.; Zemel, R.S. Prototypical Networks for Few-shot Learning. In Proceedings of the Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017.
- Sung, F.; Yang, Y.; Zhang, L.; Xiang, T.; Torr, P.H.S.; Hospedales, T.M. Learning to Compare: Relation Network for Few-Shot Learning. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 1199–1208.
- Wen, W.; Mendel, J.M. Maximum-likelihood classification for digital amplitude-phase modulations. *IEEE Trans. Commun.* **2000**, *48*, 189–193. [CrossRef]
- Panagiotou, P.; Anastasopoulos, A.; Polydoros, A. Likelihood ratio tests for modulation classification. In *MILCOM 2000 Proceedings, Proceedings of the 21st Century Military Communications. Architectures and Technologies for Information Superiority (Cat. No.00CH37155)*, Los Angeles, CA, USA, 22–25 October 2000; IEEE: New York, NY, USA, 2000; Volume 672, pp. 670–674.
- Hameed, F.; Dobre, O.A.; Popescu, D.C. On the likelihood-based approach to modulation classification. *IEEE Trans. Wirel. Commun.* **2009**, *8*, 5884–5892. [CrossRef]
- Nandi, A.K.; Azzouz, E.E. Algorithms for automatic modulation recognition of communication signals. *IEEE Trans. Commun.* **1998**, *46*, 431–436. [CrossRef]
- Swami, A.; Sadler, B.M. Hierarchical digital modulation classification using cumulants. *IEEE Trans. Commun.* **2000**, *48*, 416–429. [CrossRef]
- Spooner, C.M.; Mody, A.N.; Chuang, J.; Petersen, J. Modulation recognition using second- and higher-order cyclostationarity. In Proceedings of the 2017 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN), Baltimore, MD, USA, 6–9 March 2017; pp. 1–3.
- Hassan, K.; Dayoub, I.; Hamouda, W.; Berbineau, M. Automatic modulation recognition using wavelet transform and neural network. In Proceedings of the 2009 9th International Conference on Intelligent Transport Systems Telecommunications, (ITST), Lille, France, 20–22 October 2009; pp. 234–238.
- Simonyan, K.; Zisserman, A.J.C. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
- He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
- O’Shea, T.J.; Roy, T.; Clancy, T.C. Over-the-Air Deep Learning Based Radio Signal Classification. *IEEE J. Sel. Top. Signal Process.* **2018**, *12*, 168–179. [CrossRef]
- Liu, K.; Gao, W.; Huang, Q. Automatic Modulation Recognition Based on a DCN-BiLSTM Network. *Sensors* **2021**, *21*, 1577. [CrossRef]
- Zeng, Y.; Zhang, M.; Han, F.; Gong, Y.; Zhang, J. Spectrum Analysis and Convolutional Neural Network for Automatic Modulation Recognition. *IEEE Wirel. Commun. Lett.* **2019**, *8*, 929–932. [CrossRef]
- Jingpeng, G.; Fan, L.; Yang, L. Modulation recognition algorithm using innovative CNN and cyclic-spectrum graph. In Proceedings of the 2019 14th IEEE International Conference on Electronic Measurement & Instruments (ICEMI), Changsha, China, 1–3 November 2019; pp. 466–473.
- Zha, X.; Peng, H.; Qin, X.; Li, G.; Yang, S. A Deep Learning Framework for Signal Detection and Modulation Classification. *Sensors* **2019**, *19*, 4042. [CrossRef]
- Zhang, J.; Li, Y.; Hu, S.; Zhang, W.; Wan, Z.; Yu, Z.; Qiu, K. Joint Modulation Format Identification and OSNR Monitoring Using Cascaded Neural Network with Transfer Learning. *IEEE Photonics J.* **2021**, *13*, 7200910. [CrossRef]
- Hong, D.; Zhang, Z.; Xu, X. Automatic modulation classification using recurrent neural networks. In Proceedings of the 2017 3rd IEEE International Conference on Computer and Communications (ICCC), Chengdu, China, 13–16 December 2017; pp. 695–700.
- West, N.E.; Shea, T.O. Deep architectures for modulation recognition. In Proceedings of the 2017 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN), Baltimore, MD, USA, 6–9 March 2017; pp. 1–6.

27. Liu, X.; Yang, D.; Gamal, A.E. Deep neural network architectures for modulation classification. In Proceedings of the 2017 51st Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, 29 October–1 November 2017; pp. 915–919.
28. Xu, J.; Luo, C.; Parr, G.; Luo, Y. A Spatiotemporal Multi-Channel Learning Framework for Automatic Modulation Recognition. *IEEE Wirel. Commun. Lett.* **2020**, *9*, 1629–1632. [CrossRef]
29. Trabelsi, C.; Bilaniuk, O.; Serdyuk, D.; Subramanian, S.; Santos, J.F.; Mehri, S.; Rostamzadeh, N.; Bengio, Y.; Pal, C.J.J.A. Deep Complex Networks. *arXiv* **2017**, arXiv:1705.09792.
30. Tu, Y.; Lin, Y.; Hou, C.; Mao, S. Complex-Valued Networks for Automatic Modulation Classification. *IEEE Trans. Veh. Technol.* **2020**, *69*, 10085–10089. [CrossRef]
31. Xiao, C.; Yang, S.; Feng, Z. Complex-Valued Depthwise Separable Convolutional Neural Network for Automatic Modulation Classification. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 2522310. [CrossRef]
32. Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; Yu, P.S. A Comprehensive Survey on Graph Neural Networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 4–24. [CrossRef]
33. Liu, Y.; Liu, Y.; Yang, C. Modulation Recognition with Graph Convolutional Network. *IEEE Wirel. Commun. Lett.* **2020**, *9*, 624–627. [CrossRef]
34. Vaswani, A.; Shazeer, N.M.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is All you Need. In Proceedings of the Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017.
35. Hamidi-Rad, S.; Jain, S. MCformer: A Transformer Based Deep Neural Network for Automatic Modulation Classification. In Proceedings of the 2021 IEEE Global Communications Conference (GLOBECOM), Madrid, Spain, 7–11 December 2021; pp. 1–6.
36. Cai, Z.; Ma, W.; Wang, X.; Wang, H.; Feng, Z. The Performance Analysis of Time Series Data Augmentation Technology for Small Sample Communication Device Recognition. *IEEE Trans. Reliab.* **2023**, *72*, 574–585. [CrossRef]
37. Zhou, H.; Bai, J.; Wang, Y.; Jiao, L.; Zheng, S.; Shen, W.; Xu, J.; Yang, X. Few-shot electromagnetic signal classification: A data union augmentation method. *Chin. J. Aeronaut.* **2021**, *35*, 49–57. [CrossRef]
38. Zhang, Z.; Li, Y.; Gao, M. Few-Shot Learning of Signal Modulation Recognition based on Attention Relation Network. In Proceedings of the 2020 28th European Signal Processing Conference (EUSIPCO), Amsterdam, The Netherlands, 18–21 January 2021; pp. 1372–1376.
39. Zhou, Q.; Zhang, R.; Mu, J.; Zhang, H.; Zhang, F.; Jing, X. AMCRN: Few-Shot Learning for Automatic Modulation Classification. *IEEE Commun. Lett.* **2022**, *26*, 542–546. [CrossRef]
40. Kim, H.; Shahid, A.; Fontaine, J.; Poorter, E.D.; Moerman, I.; Nam, H. Automatic Modulation Classification using Relation Network with Denoising Autoencoder. In Proceedings of the 2022 13th International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Republic of Korea, 19–21 October 2022; pp. 485–488.
41. Li, L.; Huang, J.; Cheng, Q.; Meng, H.; Han, Z. Automatic Modulation Recognition: A Few-Shot Learning Method Based on the Capsule Network. *IEEE Wirel. Commun. Lett.* **2021**, *10*, 474–477. [CrossRef]
42. O’Shea, T.; West, N.E. Radio Machine Learning Dataset Generation with GNU Radio. In Proceedings of the GNU Radio Conference 2016, Boulder, CO, USA, 12–16 September 2016.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

VLCMnet-Based Modulation Format Recognition for Indoor Visible Light Communication Systems

Xin Zheng, Ying He *, Chong Zhang and Pu Miao

College of Electronics and Information, Qingdao University, Qingdao 266071, China;
2021020649@qdu.edu.cn (X.Z.); zhangchong@qdu.edu.cn (C.Z.); mpvae@qdu.edu.cn (P.M.)

* Correspondence: yinghe@qdu.edu.cn

Abstract: In indoor visible light communication (VLC), the received signals are subject to severe interference due to factors such as high-brightness backgrounds, long-distance transmissions, and indoor obstructions. This results in an increase in misclassification for modulation format recognition. We propose a novel model called VLCMnet. Within this model, a temporal convolutional network and a long short-term memory (TCN-LSTM) module are utilized for direct channel equalization, effectively enhancing the quality of the constellation diagrams for modulated signals. A multi-mixed attention network (MMAnet) module integrates single- and mixed-attention mechanisms within a convolutional neural network (CNN) framework specifically for constellation image classification. This allows the model to capture fine-grained spatial structure features and channel features within constellation diagrams, particularly those associated with high-order modulation signals. Experimental results obtained demonstrate that, compared to a CNN model without attention mechanisms, the proposed model increases the recognition accuracy by 19.2%. Under severe channel distortion conditions, our proposed model exhibits robustness and maintains a high level of accuracy.

Keywords: modulation format recognition; convolutional neural network; visible light communication; attention mechanism; channel equalization

1. Introduction

Direct current biased optical orthogonal frequency division multiplexing (DCO-OFDM) is a mature modulation technique widely adopted in indoor visible light communication (VLC) systems. To ensure optimal transmission performance and reliability, adaptive modulation should be carried out according to different channel conditions; hence, the receiver must autonomously detect and identify the modulation format of the received signals. However, challenges, such as high-brightness environments, long-distance transmission, and indoor obstructions, significantly reduce the signal-to-noise ratio (SNR), which, in turn, affects modulation format recognition (MFR) performance. Therefore, automatic modulation recognition methods under low SNR conditions have gained increasing attention.

Modulation format recognition techniques are generally classified into two categories: likelihood-based methods and feature-based methods [1]. Likelihood-based approaches, although typically characterized by higher accuracy, often require substantial amounts of prior information and higher computational complexity. Feature-based methods rely on extracting key features from the signals, such as higher-order accumulations [2], sequence features [3], and image features [4]. These methods generally exhibit higher computational efficiency, and their recognition results tend to be more interpretable and analyzable. However, improper feature selection can lead to decreased recognition performance.

Initially, the modulation format recognition technique rapidly developed in the field of radio communications [5–7]. In recent years, numerous machine learning algorithms have also made strides in the field of MFR in optical communications [8–13]. While inputting time-series signals can capture the complete dynamic characteristics of signals

over time, severe noise interference may cause modulation features to become blurred, thus increasing the difficulty of recognition. Constellation diagrams, by mapping signals onto discrete points, present unique patterns for different modulation formats; hence, utilizing constellation diagrams for modulation format recognition can mitigate the impact of noise.

Zhen. Z et al. employed an improved AlexNet [14] to classify constellation diagrams of the same six modulation formats, achieving a recognition accuracy of 77.45% across the SNR range of 0 dB to 15 dB. The incorporation of data augmentation (DA) operations was able to further improve the recognition accuracy to 88%, showcasing the potential of advanced image processing techniques in optical communication. Ma et al. [15] investigated a blind modulation format recognition method based on constellation diagrams derived from channel estimation in an OFDM system. They employed a combination of a peak density clustering algorithm and K-nearest neighbor (KNN) regression. At an OSNR of 30 dB and above, high-order modulation signals achieved a recognition accuracy of 80%. However, studies regarding low OSNR environments were not conducted.

Some research efforts are focusing on the preprocessing of constellation diagram features. W. Liu [16] proposed a novel modulation classification scheme for optical communications. Within an SNR range of 0 dB to 15 dB, the original constellation diagrams are preprocessed into density constellation maps, which are used for classifying six different modulation formats. Due to the reduced Euclidean distance in normalized density constellation diagrams, the accuracy rate at an SNR of 0 dB is merely 54%. In addition, a fan-beam projection algorithm to handle constellation diagrams was presented in [17]. Wei et al. [18] utilized probabilistic constellation shaping to modify the distribution of signal constellation points, aiming to enhance the quality of constellation diagrams and enhance recognition accuracy.

The complexity and nonlinearity of VLC channels result in the received signal's constellation diagram deviating from a standard lattice structure with each constellation point appearing stretched inward toward the center [19]. Pre-distortion [20–23] is a valid equalization scheme for compensating the memory nonlinearity. Nevertheless, some additional physical circuits must be devised to assist the pre-distortion at the transmitter. Unlike the former method, nonlinear post-equalization (NPE) is a superior and cost-effective technique as it can address the comprehensive nonlinearity of the LED and optical channel cascade. H. Chen et al. [24] introduced a collaborative time-frequency deep neural network (TFDNet) to mitigate nonlinear distortions in VLC systems. Miao et al. [25] also presented a model-driven deep learning (DL) equalization scheme aimed at resolving severe channel impairment issues. In [26], a CRNN-based equalizer employing a combined method of CNN and LSTM was proposed, which achieved an accuracy rate of 90% for identifying quadrature phase shift keying (QPSK) signals when the SNR exceeded 10 dB. However, the symbol error rate rose above 30% when the SNR fell below 0 dB. To simplify the model, Costa et al. [27] proposed a channel equalization scheme based on a one-dimensional CNN, which eliminated the need for channel estimation interpolation operations. But the bit error rate (BER) increased to 1% at 4 dB. A novel equalizer based on a gated recurrent unit (GRU) neural network was introduced for the first time to mitigate both linear and nonlinear distortions in carrier-less amplitude phase (CAP) band-limited VLC systems [28]. A hybrid temporal CNN with frequency domain assistance was employed to enhance the nonlinear compensation efficiency of the CAP underwater VLC system [29]. Nonetheless, the training datasets utilized in these research works consisted of real-numbered signals that were not down-sampled, thus lacking direct relevance to the efficiency of complex-valued VLC systems, leading to limited generalization capacity. Moreover, the up-sampling factor further adds to the computational complexity.

In summary, we choose constellation diagrams as the primary feature for addressing the MFR problem in indoor visible light communication, and employ channel equalization to restore the constellation diagrams. This paper introduces a novel model, termed the visible light communication modulation recognition network (VLCMnet), which consists of TCN-LSTM and MMAnet modules. It leads to increased robustness of the modulation

format recognition model, enabling accurate signal identification even under harsh channel conditions. The major contributions of this paper are as follows:

- (1) A novel channel equalization model is proposed, named TCN-LSTM. Firstly, a temporal convolutional network (TCN) [30] is used to extract deep local features from the signal, followed by a two-layer LSTM network that screens critical information and suppresses noise. The model significantly improves the quality of constellation diagrams at low SNR.
- (2) At the recognition stage, a constellation diagram classifier based on a multi-mixed attention network (MMAnet) is proposed. It integrates shallow feature extraction, single attention, and hybrid attention mechanisms, possesses multi-scale feature fusion capabilities, and effectively captures crucial spatial structure and channel features within constellation diagrams, thereby significantly boosting the model's recognition performance.

The rest of this paper is organized as follows: Section 2 introduces the indoor visible light communication system based on DCO-OFDM. Section 3 provides a detailed description of the principles behind the TCN-LSTM and MMAnet models. Section 4 discusses the performance advantages and disadvantages of our proposed MFR model from multiple perspectives. Finally, we conclude the full article in Section 5.

2. DCO-OFDM System

The DCO-OFDM system comprises a transmitter, intensity modulation, a direct detection (IM/DD) channel, and a receiver. A common modulation scheme is M-ary quadrature amplitude modulation (M-QAM), as shown in Figure 1. The VLCMnet model proposed in this paper serves as an intelligent channel equalization and modulation format identification module within the receiver. The modulation formats studied in this paper include M-QAM (where $M = 4, 8, 16, 32, 64$).

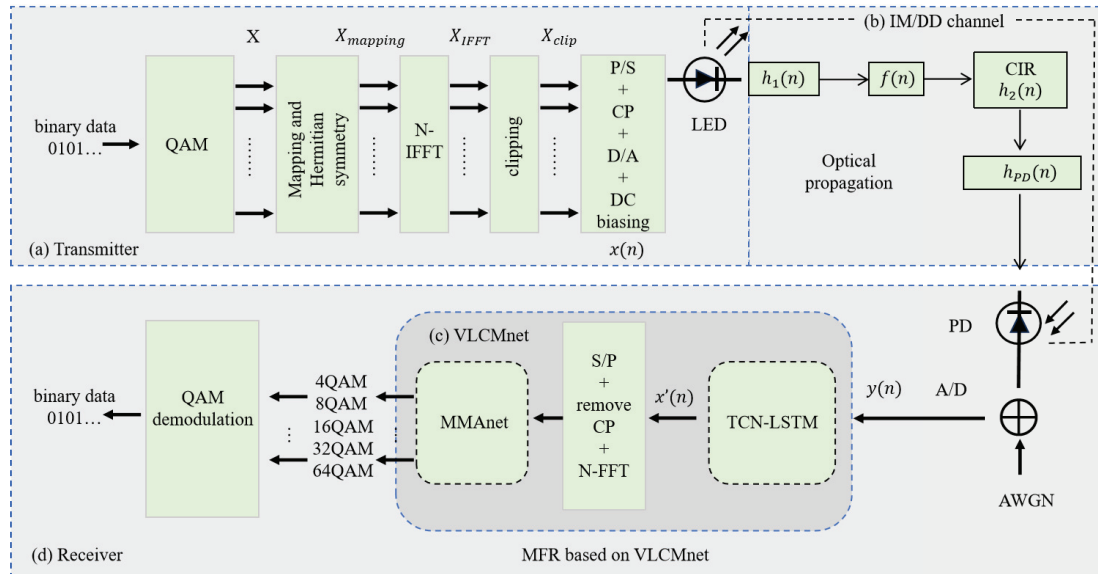


Figure 1. VLCMnet-based modulation format recognition framework for DCO-OFDM. (a) Transmitter side of the visible system. (b) IM/DD channel. (c) MFR based on VLCMnet. (d) Receiver.

At the transmitter end, firstly, M-QAM and constellation mapping are adopted to generate OFDM symbols. The information bits are mapped into real-valued symbols and the inserted cyclic prefix. Then, the bipolar baseband signal $x(n)$ is transformed into a unipolar signal by adding a direct current (DC) bias. Finally, the light-emitting diodes (LEDs) are used to generate modulated unipolar light signals. The IM/DD channel consists of a modulator at the transmitting end, a transmission medium, and a photodetector at the

receiving end. During transmission along the optical link, the signal is subject to additive white Gaussian noise (AWGN) interference. Taking all these factors into account, the output $y(n)$ of the channel can be expressed as:

$$y(n) = R_{PD} \times f\{(x(n) + I_{DC}) * h_1(n)\} * h_2(n) + \varepsilon(n), \quad (1)$$

where $h_1(n)$ represents the memory effect nonlinearity compensation of the LED, and $h_2(n)$ denotes the channel impulse response (CIR) function of the optical propagation link. R_{PD} stands for the photoelectric conversion coefficient of the photodiode (PD), and I_{DC} signifies the DC component. $\varepsilon(n)$ represents the Gaussian distribution channel noise. The $*$ indicates convolution. A range of diverse input current values were chosen to measure the real output brightness of the LEDs corresponding to each input current value. The errors between the measured data and the curves were reduced to a minimum using the least squares method. This enabled the identification of the polynomial function that most accurately matched the data points. The coefficients of the third-order polynomial were established as 0.2855–1.0886, 2.0565, and -0.0003 . The function $f(\cdot)$ adopted to describe the memoryless nonlinear response of the LED can be calculated by:

$$f(x) = 0.2855x^3 - 1.0886x^2 + 2.0565x - 0.0003. \quad (2)$$

As the modulation frequency increases, the modulation efficiency of the LED gradually decreases, exhibiting a low-pass filter effect. The pulse response $h_1(n)$ can be represented as:

$$h_1(n) = \exp(-2\pi n f_0), \quad (3)$$

where f_0 denotes the 3 dB cut-off frequency. Considering that light travels through multiple reflections and scatterings before reaching the receiver, a multi-path propagation model for VLC channels is established using ray tracing methods. The channel impulse response $h_2(n)$ is related to the path length and number of reflections for each individual ray and can be calculated as follows:

$$h_2(n) = \sum_{i=1}^{N_r} p_i \delta(n - \tau_i), \quad (4)$$

where N_r represents the number of rays received by the detector, p_i denotes the optical power of the i th ray, and τ_i signifies the delay of the i th ray. The impulse response of a photodetector can be approximately modeled as the product of a Dirac delta function and the photoelectric conversion coefficient R_{PD} , given by the following formula:

$$h_{PD}(n) = R_{PD} \delta(n). \quad (5)$$

The output signal $y(n)$ is affected by the memoryless nonlinearity of the light-emitting diodes and multipath effects in the optical transmission link, causing the constellation points to deviate from their intended grid locations, resulting in degraded constellation diagram quality. Therefore, channel equalization techniques are needed to improve the distribution of points on the constellation diagram, bringing it closer to the ideal state.

In this paper, the VLCNet model first employs an integrated neural network model to accomplish the task of channel equalization, following which it utilizes constellation diagrams along with the neural network model for modulation format recognition. The system unmaps the signal based on the category information outputted by the model to recover the original information bits.

3. Proposed Method

This section outlines the architecture and components of VLCNet, a novel approach for modulation format recognition in VLC systems. We detail the overarching structure of VLCNet, followed by specific discussions on the roles and functionalities of the TCN-LSTM and MMA-net modules within our system.

3.1. Overall Architecture of VLCMnet

As illustrated in Figure 1c, VLCMnet's architecture integrates three primary components: a TCN-LSTM module signal equalization module, a signal preprocessing module, and a MMA-net module. Firstly, the TCN-LSTM module reconstructs $y(n)$ into $x'(n)$. Subsequently, preprocessing operations, such as removing the CP and generating constellation diagrams, are performed. Finally, through the extraction of constellation diagram features by MMA-net, the modulation format is recognized.

3.2. TCN-LSTM for Channel Equalization

In the channel, noise and multipath fading cause diffusion and displacement of points on the constellation diagram. To mitigate these adverse effects, we propose the TCN-LSTM model to achieve channel equalization, as shown in Figure 2. This is because LSTM, although powerful, is not particularly adept at capturing local features in both space and time for transient distortions in communication channels. The TCN network complements LSTM's limitations in this regard.

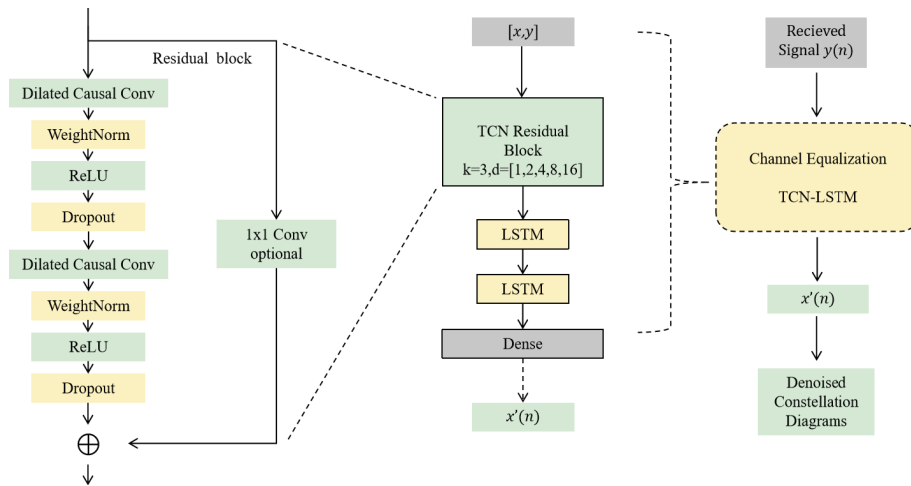


Figure 2. The hierarchical structures of TCN-LSTM.

The model training process is as follows: Using a sliding window to segment the input signal, the sequence $y(n)$ is fed into five residual modules in TCN to be computed. Each module consists of one-dimensional dilated convolutions, causal convolutions, weighted normalization, and one-dimensional convolutional residual connections. By expanding the receptive field and alleviating gradient vanishing, noise-induced point spread and constellation map distortion can be recognized and mitigated. The output from the TCN network further undergoes processing by two layers of LSTMs. The first LSTM layer captures short-term and local dependencies within the signal sequence, removing easily discernible high-frequency noise or periodic disturbances. The second LSTM layer more effectively captures long-range temporal dependencies, demonstrating increased robustness when dealing with non-stationary noise and abrupt signal changes. The model employs the mean squared error (MMSE) as its loss function. Through training, the fully connected layer outputs predict results $x'(n)$, aiming to restore the original transmitted data sequence $x(n)$ as closely as possible.

3.3. MMA-net for Modulation Format Recognition

Traditional convolutional neural networks, relying solely on local convolution operations, are unable to fully capture the global spatial layout and structural relationships in constellation diagrams. A single attention mechanism can only focus on one aspect of the features, whereas multiple attention mechanisms can extract features from multiple perspectives, constructing a more complex nonlinear model that enhances the model's expressive capability. We propose the MMA-net model, aimed at enhancing the accuracy

of MFR in severely distorted channel environments. It comprises three main modules: a shallow feature extraction (SFE) module, a multi-channel mixed attention (MCMA) module, and a feature fusion classification (FFC) module, as shown in Figure 3.

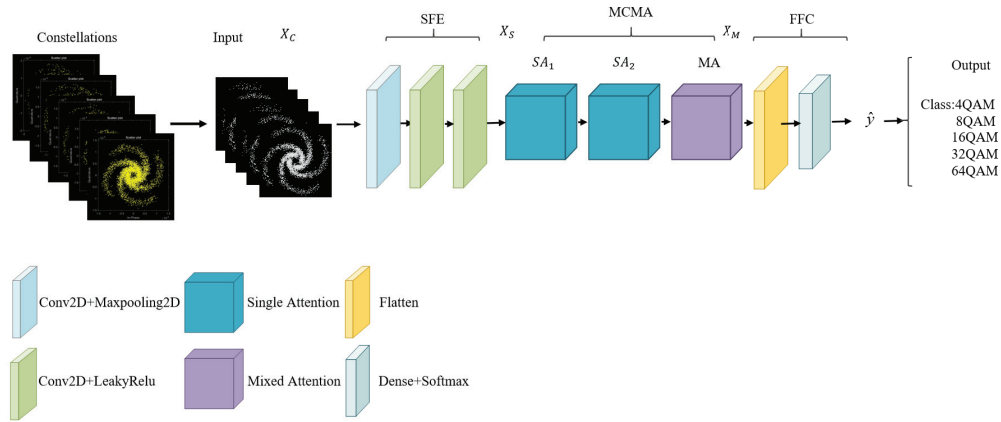


Figure 3. The overall network framework of our proposed MMAnet.

The input data consist of constellation diagrams after preprocessing steps such as channel equalization and CP removal. The training samples are denoted as X_C , which are composed of vectors $[x, y]$, where x represents a grayscale matrix of the constellation diagram with dimensions 64×64 , and y indicates the corresponding modulation format type.

The SFE module is composed of three convolutional layers. Through this module, the raw constellation diagrams undergo a progressive extraction and transformation of features from lower to higher levels, culminating in the output vector X_S .

The MCMA submodule is the key component of the MMAnet module. It consists of two single attention (SA) submodules and one mixed attention (MA) submodule, as shown in Figure 4. After the input image X_S undergoes processing through the SA modules, crucial features X_{SA2} are extracted from critical regions on the constellation diagram. Subsequently, the MA submodule focuses channel attention and spatial attention onto feature responses across different channels, as well as the positions and adjacency relationships among the constellation points. Ultimately, this enables the classification model to enhance its understanding of the modulation signal characteristics, outputting effective features X_M .

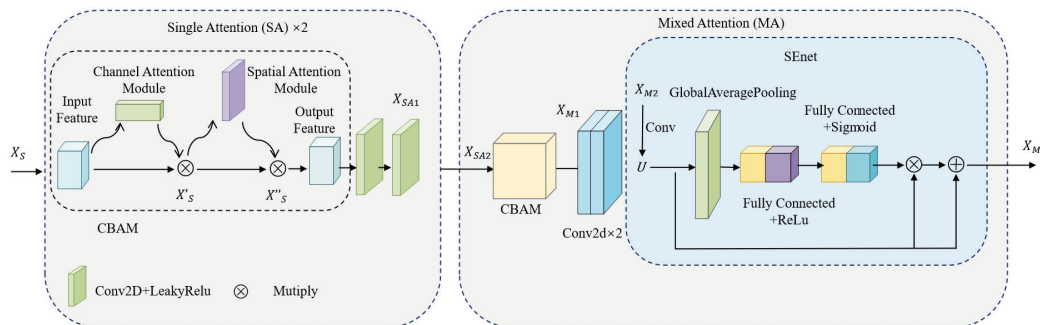


Figure 4. The architecture of the proposed MCMA module.

The SA module employs a single attention mechanism to process constellation diagrams. It consists of a convolutional block attention module (CBAM) and two convolutional layers. To enhance the model's ability to learn complex and subtle features in constellation charts, here, we use two consecutive single-attention modules in series with the aim to realize multi-level feature extraction and dynamically adjust the focus on different features. Input feature X_S first passes through the channel attention module (CAM) in CBAM, using average pooling and max pooling techniques to compress the spatial dimensions of the

input features. Subsequently, the compressed feature vectors are processed through two fully connected layers. Finally, the generated channel attention coefficients vector $M_c(X_S)$ is element-wise multiplied with the original input feature map X_S to produce a weighted feature map X'_S . The computation formula is as follows:

$$M_c(X_S) = \sigma[MLP(AvgPool(X_S))] + MLP[MaxPool(X_S)] \\ = \sigma(W_1(W_0(X_{S_{avg}}^c))) + W_1(W_0(X_{S_{max}}^c)), \quad (6)$$

$$X'_S = M_c(X_S) \otimes X_S, \quad (7)$$

where W_0 and W_1 are the weight parameters of the two fully connected layers, respectively.

After filtering out redundant information, the spatial attention module (SAM) focuses on the spatial layout of feature maps pertaining to different modulation formats. Firstly, by performing global average pooling and global maximum pooling on the input feature maps, spatial feature statistics information at different scales is obtained. Finally, two fully connected layers are employed to generate the attention weight map. The formula for the output X''_S is:

$$M_s(X'_S) = \sigma(f^{7 \times 7}([AvgPool(X'_S); MaxPool(X'_S)])) \quad (8)$$

$$X''_S = M_s(X'_S) \otimes X'_S. \quad (9)$$

After obtaining the attention weight maps, further processing is carried out through two subsequent convolutional layers to maintain the depth of the entire network and refine and transform the features guided by attention. The output vector from the first-stage SA module is denoted as X_{SA1} . This process is then repeated, culminating in the final output referred to as X_{SA2} .

The feature maps are fed into the MA module, which is capable of integrating feature information across different scales, thereby enabling the model to adapt to variations in constellation diagrams under varying channel conditions. The MA module consists of a CBAM as well as a squeeze-and-excitation networks (SE) submodule. The input vector X_{SA2} undergoes processing by the CBAM to generate an attention weight map X_{M1} . Following this, the output vector X_{M2} is produced through two convolutional layers. The SE module enhances the dynamic selection and amplification of features within the constellation diagram of the input modulated signal. Initially, after the convolutional operations, an output vector U is obtained, consisting of C feature maps, each with dimensions $H \times W$. Each u_c represents the c th two-dimensional matrix within U . Subsequently, the squeeze operation, achieved through global average pooling, maps the output dimension from $H \times W \times C$ to a $1 \times 1 \times C$ matrix, which is the statistical vector of global information for each channel. Finally, the excitation operation carried out by the two fully connected layers performs nonlinear transformations to model and reweight the importance of each channel. The formula for calculating each channel attention weight vector s is:

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)), \quad (10)$$

where F_{ex} represents the dynamic weighting function for channel features. z denotes the input feature map, W_1 signifies the weight matrix of the reduction layer, and W_2 represents the weight matrix of the expansion layer. δ denotes the rectified linear unit (ReLU) activation function, and σ represents the sigmoid activation function. F_{scale} denotes a scale adjustment operation function. s_c refers to the c th element of the attention weight vector, while u_c represents the feature of the c th channel in the input feature map. The final output X_M is derived as:

$$X_M = F_{scale}(u_c, s_c) = s_c u_c. \quad (11)$$

The FFC module consists of a flattening layer and a dense layer. The constellation map features X_M extracted by the preceding module undergo nonlinear transformations

within the FFC, followed by the computation of matching probabilities through the Softmax activation function to generate the output $\hat{y}$. The computation formula is as follows:

$$\hat{y}_i = \text{softmax}(o) = \frac{\exp(o_i)}{\sum_k \exp(o_k)}. \quad (12)$$

The output of training samples is represented using one-hot encoding, denoted as $y = [y_1, y_2, \dots, y_5]^T$. The highest predicted value in $\hat{y}_i$ is taken as the output category y_p , which represents the category of the modulation format.

4. Results and Discussion

4.1. Experiment Details

The VLC experimental setup encompasses an array of equipment, including a source computer, an arbitrary function generator, an amplifier, a bias-T, an LED, a photodiode, and a mixed-domain oscilloscope. The parameters for these components are meticulously outlined in Table 1.

Table 1. Devices and parameters of the VLC system prototype.

Device/Parameter	Value
Arbitrary function generator	Tektronix AFG3152C
Amplifier	Mini-Circuits ZHL-6A-S+
Bias-T	SHWBT-006000-SFFF
PD	PDA10A-EC
Responsivity of PD	0.44 A/W at 750 nm
Mixed-domain oscilloscope	Tektronix MDO
Power of LED	7.35 W

The entire model training involves TCN-LSTM for channel equalization on received signals, as well as MMAnet for classifying the generated constellation diagrams. Table 2 lists the configurations of the training parameters. It also presents the training parameters of the original 9-layer CNN before the algorithmic improvements.

Table 2. Network training parameters.

Parameters	TCN-LSTM	CNN	MMAnet
Optimizer	Adam	Adam	Adam
Learning rate	0.0001	0.00001	0.00001
Epochs	2000	200	200
Datatype	Sequence	Image	Image
Loss function	mse	categorical_crossentropy	categorical_crossentropy
Activation function	tanh	softmax	softmax
Batch size	64	32	32

The training dataset consists of 21,000 constellation diagrams, while the testing dataset includes 5250 constellation diagrams. The data were collected from received signals under 21 different SNR conditions ranging from -10 dB to 30 dB with an interval of 2 dB. Both the original CNN and the MMAnet models have their input data undergo an image preprocessing procedure before training, including cropping, grayscale conversion, and normalization. First, the CV2 library is utilized to resize the images to a dimension of 64×64 . Following this, transformation is applied to obtain grayscale images of the constellation diagrams. Lastly, each pixel value is divided by 255. The experiments were executed using Python 3.9 and Keras 2.6.0 on an NVIDIA GeForce RTX 3060 GPU (NVIDIA, Santa Clara, CA, USA).

4.2. Evaluation Metrics

To quantitatively assess the performance of our model, we employed standard evaluation metrics, including the loss value, accuracy, recall, F1-score, kappa coefficient, and mean average precision (mAP). The kappa coefficient measures the consistency between model predictions and true labels (values range from 0 to 1, with higher values indicating superior consistency). The mAP represents class-wide average precision, which is calculated by a weighted average, combining the recall and precision of each class. The formulas for these metrics are as follows:

$$mAP = \frac{1}{n} \sum_1^n (AP)_i, \quad (13)$$

$$kappa = (p_0 - p_e) / (1 - p_e) = (\sum_{i=1}^n x_{ii} / N - \sum_{i=1}^n x_{i+} x_{+i} / N^2) / (1 - \sum_{i=1}^n x_{i+} x_{+i} / N^2), \quad (14)$$

where x_{ii} denotes diagonal elements in the confusion matrix, x_{i+} , x_{+i} represent the sum of the row and column elements for i , respectively, and N is the total element count. p_0 denotes the agreement observed, while p_e denotes the agreement expected by chance.

4.3. Channel Equalization Effects on Classification Performance

We trained a 9-layer CNN model as our baseline for evaluating VLC modulated signal recognition capabilities. The detailed structural parameters of this benchmark model are shown in Table 3. The classification performance of the benchmark model was compared with classical image classification networks, including Googlenet, Vgg16, Alexnet, and Resnet.

Table 3. Parameter settings of CNN.

Layer	Filters	Kernel Size	Layer	Filters	Kernel Size
Conv2D-1	64	7×7	MaxPooling	\	2×2
MaxPooling	\	2×2	Conv2D-6	256	3×3
Conv2D-2	128	3×3	Conv2D-7	256	3×3
Conv2D-3	128	3×3	MaxPooling	\	2×2
Conv2D-4	64	3×3	Conv2D-8	64	5×5
Conv2D-5	64	3×3	Conv2D-9	64	5×5

The experiments were performed with five prevalent convolutional neural network models for modulation recognition without channel equalization, respectively. Figure 5 presents a comparative analysis of their test accuracies. There is a uniform improvement in recognition accuracy across all models as SNR increases. The CNN model performs best, followed by Googlenet. When the SNR is higher than 26 dB, all models can achieve an accuracy of 100%. The CNN model achieves a modulation recognition accuracy of 80% when SNR is 14 dB, which represents an improvement of about 2 dB to 8 dB over the other four models. But at SNRs below 10 dB, the recognition accuracies are below 60%. Due to the possibility of noise masking the unique time-frequency characteristics inherent in different modulation signals, the challenge of improving recognition accuracy persists. Consequently, we deploy a TCN-LSTM model at the receiver end with the objective of enhancing the quality of the constellation diagrams.

After adding the TCN-LSTM module, both the CNN and Googlenet models achieved an average accuracy improvement of over 6%, as depicted in Figure 6. In the range from 0 dB to 10 dB, the CNN model exhibited significant performance enhancement. At an SNR of 6 dB, the CNN model reached an accuracy close to 100%. The threshold of the signal-to-noise ratio required for accurate signal recognition by the CNN model dropped by 14 dB compared to Googlenet.

The changes in constellation diagrams before and after equalization can be observed in Figure 7. The first row of this figure shows the constellation diagrams of the five modulated signals without channel equalization. It visually illustrates that noise-induced

point spreading diminishes the clarity between adjacent constellation points. After channel equalization by TCN-LSTM, the amplitude and phase distortions introduced by the channel are significantly mitigated. The constellation points are more clustered, the distribution becomes more regular, and the edge clarity improves. Therefore, it becomes easier to distinguish the modulated signals, especially the 8QAM signals.

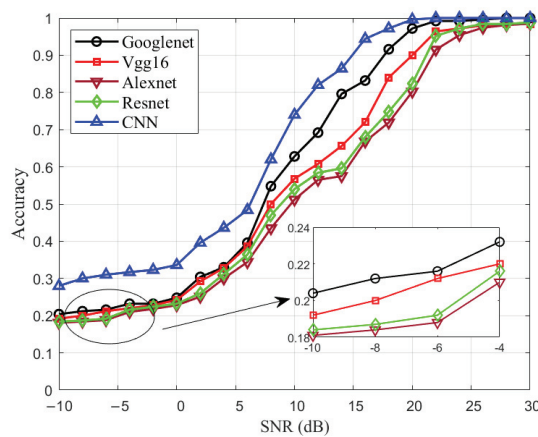


Figure 5. The recognition accuracy of five models on test samples at various SNRs.

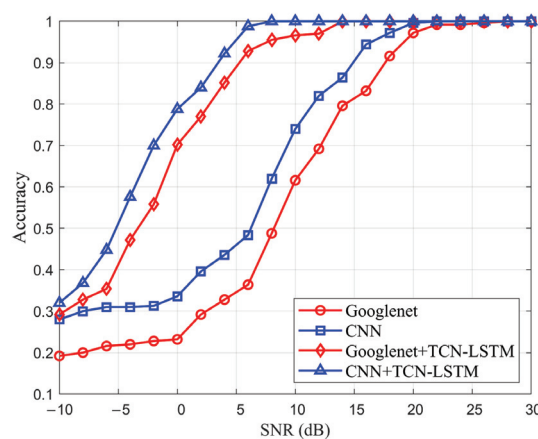


Figure 6. A comparison of modulation recognition accuracy before and after incorporating the TCN-LSTM module into both the CNN and GoogleNet models.

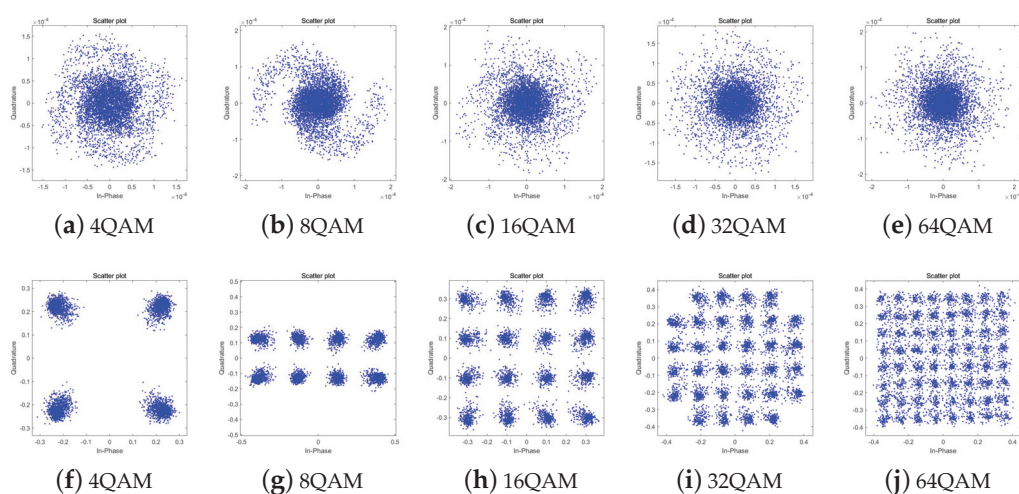


Figure 7. Comparing the constellation diagram quality for modulated signals under two conditions. (a–e) Without TCN-LSTM. (f–j) With TCN-LSTM.

The following experiments illustrate the effect of our proposed channel equalization model on the overall recognition accuracy of the different modulated signals. Figure 8 presents a detailed comparison of the performance differences before and after combining the TCN-LSTM model with the CNN model, leading to the following conclusions:

- (1) After adding a TCN-LSTM module for channel equalization prior to modulation recognition, the modulation recognition accuracy of 4QAM, 8QAM, 16QAM, 32QAM, and 64QAM increased by 4.2%, 1.7%, 8.1%, 6.8%, and 6.1%, respectively. The TCN model efficiently extracts both local and global temporal features, which contributes to the attenuation of high-frequency noise. The LSTM, through its gating mechanism, captures time-contextual information within the signal. Two models are cascaded to achieve complementary advantages.
- (2) There are differences in recognition accuracy among the various signal types. Lower-order QAM modulation formats generally exhibit higher recognition accuracy. Higher-order modulation signals tend to suffer from more misclassifications or missed detections compared to lower-order modulation. The main reason is that lower-order QAM constellation diagrams have simpler distributions and larger point spacings. Due to interference, such as burst noise and multipath fading, the constellation diagrams of 16QAM, 32QAM, and 64QAM become more alike due to blurring.
- (3) There are two noteworthy recognition results. After channel equalization, there is only a marginal improvement in the recognition accuracy for 8QAM. The recognition rate for 32QAM remains higher than that of 16QAM and 64QAM. The underlying reason is that the constellation pattern of 8QAM distinctly differs from those of other signals. Even under poor channel conditions that cause blurriness in the constellation diagrams, 8QAM is relatively easier to discern. For 16QAM and 64QAM, their constellation diagrams are arranged in a square lattice pattern, making them quite similar in shape. The 32QAM scheme is composed of two orthogonally superimposed 16QAM constellation diagrams. It possesses unique boundary shapes at the points of maximum and minimum amplitudes and thereby has relatively higher distinctions.

Although the use of the TCN-LSTM model proposed in this paper to accomplish channel equalization results in more time-consuming training of the entire modulation recognition model, the prediction execution time of the TCN-LSTM + CNN model during testing is only 0.35 s longer than that of the CNN model alone. In summary, while the introduction of the TCN-LSTM model has indeed led to an increase in the overall average accuracy of the CNN model from 0.795 to 0.838, its performance remains unsatisfactory when dealing with the recognition of higher-order modulation formats. Therefore, deeper mining of constellation diagram features for higher-order modulation formats is needed.

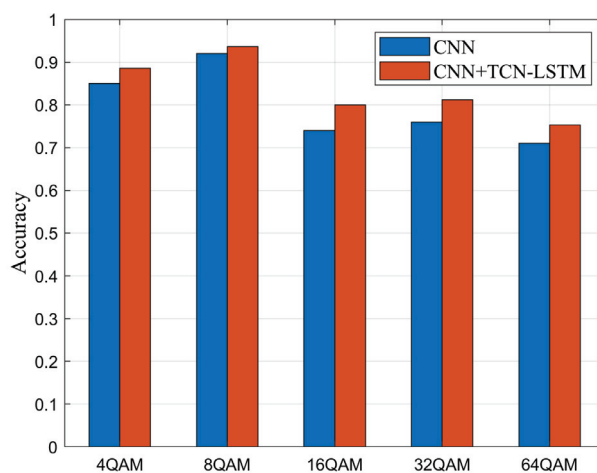


Figure 8. Prediction accuracy for different modulation formats across all SNRs.

4.4. MFR Performance Analysis Based on VLCMnet

In this section, the modulation format recognition model employs the sub-module MMAnet of our proposed VLCMnet model, which introduces a multi-attention mechanism into the original 9-layer CNN model. Based on the experimental results, we initially assessed the models using the loss function, followed by conducting a comparative analysis of the recognition accuracy of different models, and ultimately analyzed the differences in constellation diagram features extracted by the convolutional layers before and after the attention mechanism is integrated. The specific network structure and corresponding output dimensions are presented in Table 4.

Table 4. The structural parameters of MMAnet.

	Layer Name	Output Shape
SFE	Conv2d-1	(64, 64, 64)
	MaxPooling2D	(32, 32, 64)
	Conv2D-2	(32, 32, 128)
	Conv2D-3	(32, 32, 128)
SA1	cbam_block1	(32, 32, 128)
	Conv2D-4	(32, 32, 64)
	Conv2D-5	(32, 32, 64)
SA2	cbam_block2	(32, 32, 64)
	Conv2D-6	(32, 32, 256)
	Conv2D-7	(32, 32, 256)
MA	cbam_block3	(32, 32, 256)
	Conv2D-8	(32, 32, 64)
	Conv2D-9	(32, 32, 64)
	Se_block	(32, 32, 64)
FEC	Flatten	(65, 536)
	Dense	(5)

The following compares the convergence speed and fitting degree of the CNN and MMAnet models. The curves showing the change in training and validation losses are displayed in Figure 9. For both models, the validation losses do not notably exceed the training losses, indicating no signs of overfitting. MMAnet reaches a loss value of 0.05 at the 50th generation and converges after the 100th generation, suggesting a faster convergence rate compared to CNN. The validation loss curve for CNN exhibits slight oscillations, and its validation loss is on average 0.1 higher than that of MMAnet, indicating that CNN is less capable of learning from data with substantial noise compared to MMAnet.

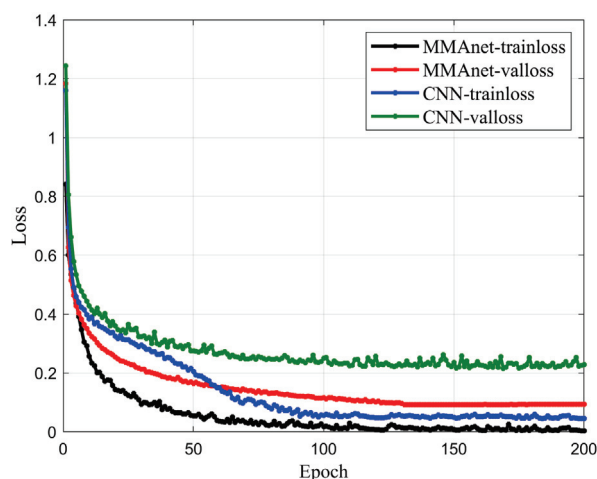


Figure 9. Comparison of CNN and MMAnet loss function curves.

Our proposed algorithm can enhance the recognition accuracy of modulation formats in indoor visible light communication under severe channel distortion. Table 5 presents the accuracy and mAP values of six models across a range of SNRs from 0 dB to 12 dB. The experimental results demonstrate that, under extremely adverse channel conditions (SNRs of 0 dB and 4 dB), the VLCMnet model achieves recognition rates exceeding those of other models by more than 7%, and also shows an improvement of over 6% in mAP for various categories of modulation formats. At an SNR of 12 dB, all six models are virtually capable of correctly identifying all sample categories in the test set.

Table 5. Comparison of test accuracies for six models under different SNRs.

SNR (dB)		0	4	8	12	Average
ACC	TCN-LSTM + Alexnet	0.648	0.724	0.916	0.984	0.782
	TCN-LSTM + Resnet	0.684	0.802	0.968	0.996	0.799
	TCN-LSTM + Vgg16	0.728	0.824	0.985	1	0.814
	TCN-LSTM + Googlenet	0.732	0.902	0.988	1	0.822
	TCN-LSTM + CNN	0.788	0.922	1	1	0.838
	VLCMnet	0.933	0.992	1	1	0.948
mAP	TCN-LSTM + Alexnet	0.674	0.743	0.934	0.994	0.791
	TCN-LSTM + Resnet	0.712	0.825	0.984	1	0.812
	TCN-LSTM + Vgg16	0.752	0.847	1	1	0.826
	TCN-LSTM + Googlenet	0.754	0.914	1	1	0.833
	TCN-LSTM + CNN	0.822	0.941	1	1	0.847
	VLCMnet	0.936	1	1	1	0.955

The TCN-LSTM channel equalization algorithm plays a pivotal role here. Concurrently, it is noted that MMA-net adopts a hybrid attention mechanism, dynamically allocating attention weights to the most critical channels and spatial locations, enabling the model to better focus on key spatial regions that determine the position and distribution of constellation points, thereby ignoring background noise or other non-critical spatial information. So this model improves its fine-grained parsing capability of constellation diagrams, subsequently boosting both recognition accuracy and stability.

The VLCMnet model demonstrates a more accurate ability to identify multi-level QAM modulation formats even under severe channel distortion, indicating that the MMA-net submodule enhances the feature representation of the constellation diagram. We perform feature map visualization on the convolutional layers of both the TCN-LSTM + CNN model and the VLCMnet model. The experimental analysis examines the capacity of both models to capture fine-grained constellation diagram features at an SNR of 4 dB, as shown in Figure 10.

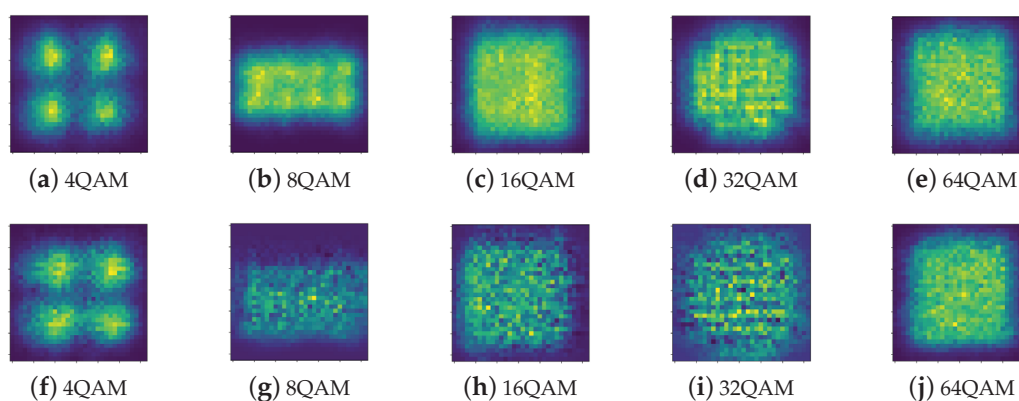


Figure 10. (a–e) Feature maps of the constellation diagrams extracted by CNN. (f–j) Feature maps of the constellation diagrams extracted by VLCMnet.

The feature maps of CNN have distinct features in 4QAM and 8QAM, and the detailed features in 16QAM, 32QAM, and 64QAM are not obvious. In contrast, there are numerous regions with varying color depths in VLCMnet. The internal features of the 16QAM, 32QAM, and 64QAM constellation diagrams are assigned different weights. It indicates that after incorporating multiple attention mechanisms, the convolutional layers focus more intently on the image areas that are most crucial for the recognition task, resulting in clearer and more pronounced feature outlines and structures within the feature maps, which are related to the arrangement patterns and distribution shapes of the constellations for different signals. Consequently, the VLCMnet model successfully captures intricate edge and interior features in the constellation diagrams, thereby enhancing the model's generalization capability.

4.5. The Impact of Multi-Attention Mechanisms on VLCMnet Performance

In this section, we discuss the impact of multiple attention mechanisms on the modulation format recognition performance of the VLCMnet model. Here, the model without any attention mechanism is denoted as (Base), the model using only a single attention mechanism is referred to as (W/O MA), the model employing only a mixed attention mechanism is termed (W/O SA), and the model using multi-attention mechanisms is denoted as (With All).

As shown in Figure 11, when different attention mechanisms are incorporated into the classification model, the accuracy for recognition consistently surpasses that of models with no attention mechanisms. Within the SNR range from -10 dB to 0 dB, the recognition accuracy can be improved by more than 15%.

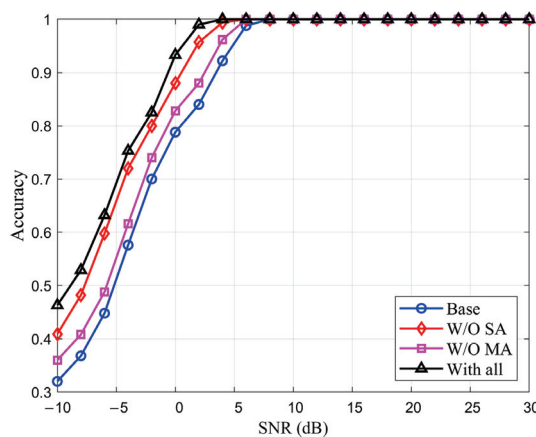


Figure 11. An analysis of ablation experiments on the modulation recognition model with the inclusion of different attention mechanisms.

To evaluate the effect of different attention mechanisms on the recognition performance for each modulation format, an ablation study was conducted on 5250 test samples. The classification confusion matrix is shown in Figure 12.

- (1) As shown in Figure 12a, without attention mechanisms, the features extracted from the convolutional layers tend to be susceptible to noise and irrelevant feature interference. While the model demonstrates relatively high accuracy in recognizing lower-order modulations, particularly 8QAM, there exist numerous errors in recognizing higher-order modulations. For instance, 154 instances of 16QAM are misclassified as 64QAM.
- (2) Figure 12b shows that the recognition performance improves a little if only the SA module is introduced into VLCMnet. The number of misclassifications for higher-order modulations decreases but not significantly. The recognition accuracy of 64QAM is almost unchanged. It is easily confusedly recognized as 16QAM and 32QAM, possibly due to inadequate extraction of spatial structure features within the distribution of constellation points.

- (3) Figure 12c shows that when solely the MA module is incorporated, VLCMnet exhibits a better enhancement in recognition performance compared to when only the SA module is added. The recognition accuracy for all five modulation formats improves. In particular, the recognition accuracy of 64QAM is improved by 11%. The proportion of 64QAM being misclassified as 32QAM notably declines.
- (4) Figure 12d demonstrates that the inclusion of both SA and MA modules in VLCMnet is more helpful in improving the model performance. The misclassification between 16QAM and 64QAM is further reduced. The recognition accuracy for higher-order modulations increases by more than 13% compared to the Base model.

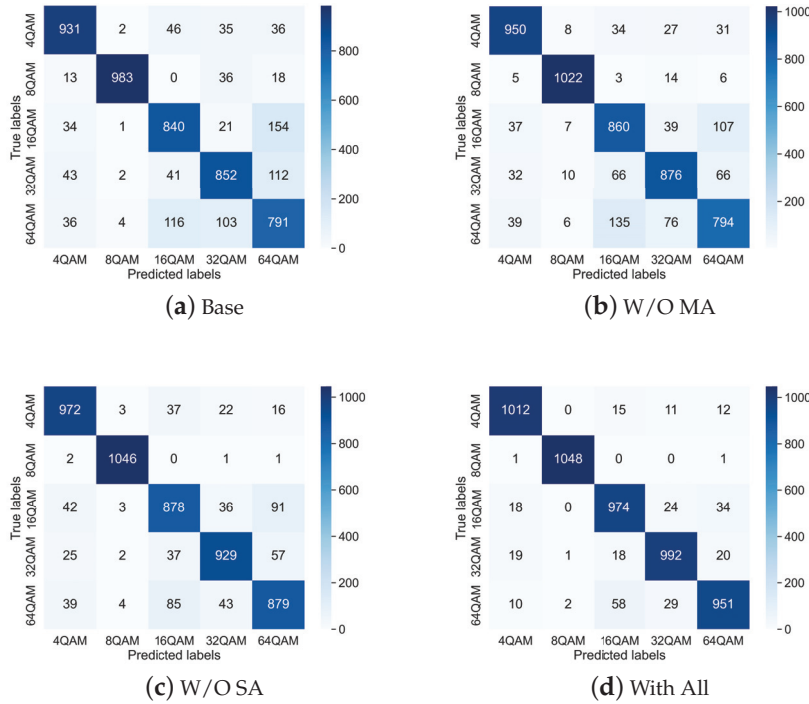


Figure 12. Confusion matrixes at Base, W/O MA, W/O SA, and With All.

In summary, VLCMnet is able to concentrate on critical point-position features via its SA module and simultaneously attend to key spatial structures in the distribution of constellation points through its MA module, thereby enhancing the model's recognition accuracy and robustness, particularly for higher-order QAM modulation signals.

4.6. Comprehensive Evaluation of Model Performance

This section first discusses the impact of system transmission data volume indicators on model performance and subsequently conducts a comparative analysis of the generalization capability and network complexity among six models. The IFFT length and the carrier count represent the number of subcarriers used by the transmitter to generate OFDM symbols and actually carry the data payload.

As shown in Figure 13a, the recognition accuracy of the model increases with growth in the IFFT length. When the carrier count is 256 and the IFFT length is 2048, the model performance outperforms other combinations of carrier counts. However, both parameter values should not be excessively large, as this can lead to oversampling, causing severe point overlapping in the constellation diagram, which, in turn, reduces the recognition accuracy.

The Rx data length refers to the length of the effective data payload at the receiver end. As illustrated in Figure 13b, when the Rx data length ranges from 20,000 to 60,000, the model accuracy generally exhibits an upward trend as the data length increases. The model performs optimally at a length of 41,600. However, excessive increase in the Rx data length

can, conversely, introduce more noise and redundant information, reducing the clarity of the constellation diagram and recognition accuracy.

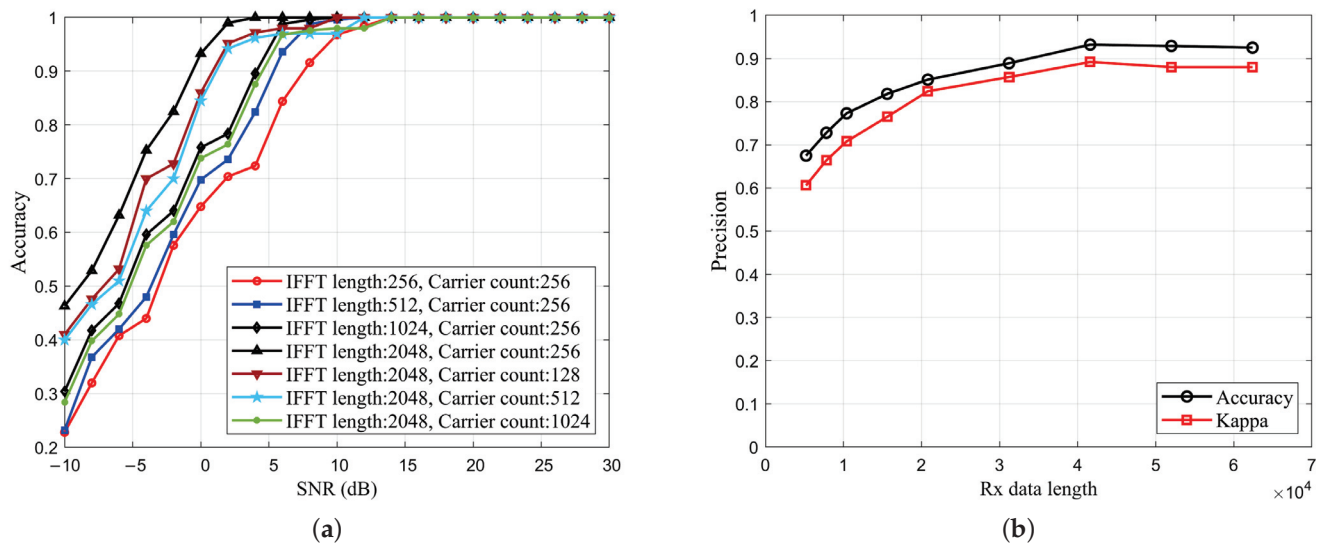


Figure 13. (a) Comparison of recognition accuracy for different channel parameters. (b) Comparison of recognition precision for different Rx data length.

In the following, the MMA-net module within the VLCM-net model is replaced with five alternative convolutional neural networks, and a comprehensive comparison of the classification performance and complexity differences among these models is presented, as shown in Table 6. The model complexity is assessed using million floating-point operations per second (MFLOPs) and the number of parameters. The generalization ability of the model is evaluated using the F1-score, recall, and accuracy metrics.

Table 6. Comparison of the comprehensive performance of the models.

Method	MFLOPs	Parameters (M)	Recall	F1	Accuracy
Alexnet	43.19	21.59	0.783	0.781	0.782
Resnet	42.56	21.28	0.801	0.800	0.799
Vgg16	33.75	16.88	0.816	0.815	0.814
Googlenet	12.01	6.01	0.823	0.822	0.822
CNN	3.2	1.6	0.839	0.838	0.838
MMA-net	3.9	1.9	0.949	0.948	0.948

The experimental results indicate that MMA-net achieves higher precision with relatively lower computational requirements. Although CNN has the lowest complexity, its accuracy is 11% lower than that of MMA-net. The precision and complexity of Googlenet are moderate. AlexNet and Vgg16 are both complex models and show inferior precision. Therefore, the proposed model in this paper possesses good generalization capabilities, fast training speed, and convenient model deployment.

In Section 4.3, it is reported that the average accuracy achieved by a 9-layer CNN model for modulation recognition was only 0.795, whereas our proposed VLCM-net reaches an accuracy of 0.948, representing a 19.2% improvement over the CNN model. This indicates that the model effectively mitigates channel-induced interference and accurately captures and discriminates key features of different modulation formats.

5. Conclusions

In this paper, we propose the VLCM-net model for DCO-OFDM systems, an integrated deep learning-based network model aimed at addressing the problem of low accuracy in

modulation format recognition under severe channel distortion conditions. We analyze the model characteristics from several aspects, including the changes in model performance before and after channel equalization, the effects of different attention mechanisms on model performance, the model complexity, and the effects of the system input parameters. The experimental results show that the TCN-LSTM module effectively extracts temporal and frequency local features of the signals, and performs repair on impaired signals. The MMAnet module introduces a multi-attention mechanism in the CNN, which captures diverse feature information from distinct subspace perspectives, and contributes to improving recognition accuracy. In summary, the modulation format recognition model presented in this paper can effectively reduce the confusion and misclassifications for higher-order modulation formats. It exhibits strong robustness, computational efficiency, and effectiveness. This model plays a significant role in promoting the task of modulation recognition under complex channel environments.

Author Contributions: Conceptualization, X.Z. and Y.H.; methodology, X.Z. and P.M.; software, X.Z. and C.Z.; validation, X.Z., Y.H., and C.Z.; formal analysis, X.Z.; investigation, X.Z. and C.Z.; resources, Y.H. and P.M.; writing—original draft preparation, X.Z.; writing—review and editing, Y.H. and X.Z.; funding acquisition, Y.H. and P.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (NSFC) under grants 61801257 and 61701271, and by the Qingdao University Research Program through project (ID: RH2000004239).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available from the first author upon request.

Acknowledgments: The authors express their gratitude to the editors and the anonymous reviewers for their insightful suggestions and general assistance.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Xiao, W.; Luo, Z.; Hu, Q. A Review of Research on Signal Modulation Recognition Based on Deep Learning. *Electronics* **2022**, *11*, 2764. [CrossRef]
2. Liu, K.; Li, F. Automatic modulation recognition based on a multiscale network with statistical features. *Phys. Commun.* **2023**, *58*, 102052. [CrossRef]
3. Hou, S.; Fan, Y.; Han, B.; Li, Y.; Fang, S. Signal Modulation Recognition Algorithm Based on Improved Spatiotemporal Multi-Channel Network. *Electronics* **2023**, *12*, 422. [CrossRef]
4. Leblebici, M.; Çalhan, A.; Cicioğlu, M. Deep learning-based modulation recognition with constellation diagram: A case study. *Phys. Commun.* **2024**, *63*, 102285. [CrossRef]
5. Peng, Y.; Guo, L.; Yan, J.; Tao, M.; Fu, X.; Lin, Y.; Gui, G. Automatic Modulation Classification Using Deep Residual Neural Network with Masked Modeling for Wireless Communications. *Drones* **2023**, *7*, 390. [CrossRef]
6. Marey, A.; Marey, M.; Mostafa, H. Novel Deep-Learning Modulation Recognition Algorithm Using 2D Histograms over Wireless Communications Channels. *Micromachines* **2022**, *13*, 1533. [CrossRef]
7. Ali, A.K.; Erçelebi, E. Modulation Format Identification Using Supervised Learning and High-Dimensional Features. *Arab. J. Sci. Eng.* **2023**, *48*, 1461–1486. [CrossRef]
8. He, J.; Zhou, Y.; Shi, J.; Tang, Q. Modulation Classification Method Based on Clustering and Gaussian Model Analysis for VLC System. *IEEE Photonics Technol. Lett.* **2020**, *32*, 651–654. [CrossRef]
9. Ağır, T.T.; Sönmez, M. The modulation classification methods in PPM-VLC systems. *Opt. Quantum Electron.* **2023**, *55*, 223. [CrossRef]
10. Zhou, Z.; Guan, W.; Wen, S. Recognition and evaluation of constellation diagram using deep learning based on underwater wireless optical communication. *arXiv* **2020**, arXiv:2007.05890.
11. Gu, Y.; Wu, Z.; Li, X.; Tian, R.; Ma, S.; Jia, T. Modulation Format Identification in a Satellite to Ground Optical Wireless Communication Systems Using a Convolution Neural Network. *Appl. Sci.* **2022**, *12*, 3331. [CrossRef]

12. Mohamed, S.E.D.N.; Mortada, B.; Ali, A.M.; El-Shafai, W.; Khalaf, A.A.M.; Zahran, O.; Dessouky, M.I.; El-Rabaie, E.S.M.; El-Samie, F.E.A. Modulation format recognition using CNN-based transfer learning models. *Opt. Quantum Electron.* **2023**, *55*, 343. [CrossRef]
13. Li, J.; Ma, J.; Liu, J.; Lu, J.; Zeng, X.; Luo, M. Modulation Format Identification and OSNR Monitoring Based on Multi-Feature Fusion Network. *Photonics* **2023**, *10*, 373. [CrossRef]
14. Zhao, Z.; Khan, F.N.; Li, Y.; Wang, Z.; Zhang, Y.; Fu, H. Application and comparison of active and transfer learning approaches for modulation format classification in visible light communication systems. *Opt. Express* **2022**, *30*, 16351–16361. [CrossRef] [PubMed]
15. Ma, Y.; Gao, M.; Zhang, J.; Ye, Y.; Chen, W.; Ren, H.; Yan, Y. Modulation format identification based on constellation diagrams in adaptive optical OFDM systems. *Opt. Commun.* **2019**, *452*, 203–210. [CrossRef]
16. Liu, W.; Li, X.; Yang, C.; Luo, M. Modulation classification based on deep learning for DMT subcarriers in VLC system. In Proceedings of the 2020 Optical Fiber Communications Conference and Exhibition (OFC), San Diego, CA, USA, 8–12 March 2020; pp. 1–3.
17. Mortada, B.; El-Shafai, W.; Zahran, O.; El-Rabaie, E.S.; Abd El-Samie, F. Fan-beam projection based modulation classification for optical systems with phase noise effect. *J. Opt.* **2023**, *52*, 1–15. [CrossRef]
18. Wei, Z.; Zhang, J.; Li, W.; Plant, D.V. Active Learning-Aided CNN-Based Entropy-Tunable Automatic Modulation Identification for Rate-Flexible Coherent Optical System. *J. Light. Technol.* **2023**, *41*, 4598–4608. [CrossRef]
19. Xu, C.; Jin, R.; Gao, W.; Chi, N. Efficient Modulation Classification Based on Complementary Folding Algorithm in UVLC System. *IEEE Photonics J.* **2022**, *14*, 1–6. [CrossRef]
20. Chen, X.; Wang, Y.; Liu, X.; Wang, Z.; Zhang, X.; Zhou, F.; Ng, D.W.K. A Hybrid Active-Passive Single-Order Equalizer for Visible Light Communication Systems. *IEEE Photonics Technol. Lett.* **2023**, *35*, 1395–1398. [CrossRef]
21. Sun, W. Research on the receiving and equalization of visible light communication system based on the light-emitting-diode emission process and mechanism. *Opt. Eng.* **2023**, *62*, 098102. [CrossRef]
22. Bostanoglu, M.; Dalveren, Y.; Catak, F.O.; Kara, A. Modelling and Design of Pre-Equalizers for a Fully Operational Visible Light Communication System. *Sensors* **2023**, *23*, 5584. [CrossRef] [PubMed]
23. Chen, C.; Nie, Y.; Liu, M.; Du, Y.; Liu, R.; Wei, Z.; Fu, H.Y.; Zhu, B. Digital Pre-Equalization for OFDM-Based VLC Systems: Centralized or Distributed? *IEEE Photonics Technol. Lett.* **2021**, *33*, 1081–1084. [CrossRef]
24. Chen, H.; Zhao, Y.; Hu, F.; Chi, N. Nonlinear Resilient Learning Method Based on Joint Time-Frequency Image Analysis in Underwater Visible Light Communication. *IEEE Photonics J.* **2020**, *12*, 1–10. [CrossRef]
25. Miao, P.; Yin, W.; Peng, H.; Yao, Y. Study of the Performance of Deep Learning-Based Channel Equalization for Indoor Visible Light Communication Systems. *Photonics* **2021**, *8*, 453. [CrossRef]
26. Li, Z.; Hu, F.; Li, G.; Zou, P.; Wang, C.; Chi, N. Convolution-Enhanced LSTM Neural Network Post-Equalizer used in Probabilistic Shaped Underwater VLC System. In Proceedings of the 2020 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), Macau, China, 21–24 August 2020; pp. 1–5. [CrossRef]
27. Costa, W.S.; Samatelo, J.L.; Rocha, H.R.; Segatto, M.E.; Silva, J.A. Direct Equalization with Convolutional Neural Networks in OFDM based VLC Systems. In Proceedings of the 2019 IEEE Latin-American Conference on Communications (LATINCOM), Salvador, Brazil, 11–13 November 2019; pp. 1–6. [CrossRef]
28. Li, S.; Zou, Y.; Shi, Z.; Tian, J.; Li, W. Performance enhancement of CAP-VLC system utilizing GRU neural network based equalizer. *Opt. Commun.* **2023**, *528*, 129062. [CrossRef]
29. Chen, H.; Jia, J.; Niu, W.; Zhao, Y.; Chi, N. Hybrid frequency domain aided temporal convolutional neural network with low network complexity utilized in UVLC system. *Opt. Express* **2021**, *29*, 3296–3308. [CrossRef]
30. Bai, S.; Kolter, J.Z.; Koltun, V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv* **2018**, arXiv:1803.01271.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

A Survey of PAPR Techniques Based on Machine Learning

Bianca S. de C. da Silva ^{1,*}, Victoria D. P. Souto ¹, Richard D. Souza ² and Luciano L. Mendes ¹

¹ National Institute of Telecommunications, Santa Rita do Sapucaí 37540-000, Brazil; victoria.souto@inatel.br (V.D.P.S.); lucianol@inatel.br (L.L.M.)

² Department of Electrical and Electronic Engineering, Federal University of Santa Catarina, Florianópolis 88040-900, Brazil; richard.demo@ufsc.br

* Correspondence: bianca.silva@mtel.inatel.br

Abstract: Orthogonal Frequency Division Multiplexing (OFDM) is the modulation technology used in Fourth Generation (4G) and Fifth Generation (5G) wireless communication systems, and it will likely be essential to Sixth Generation (6G) wireless communication systems. However, OFDM introduces a high Peak to Average Power Ratio (PAPR) in the time domain due to constructive interference among multiple subcarriers, increasing the complexity and cost of the amplifiers and, consequently, the cost and complexity of 6G networks. Therefore, the development of new solutions to reduce the PAPR in OFDM systems is crucial to 6G networks. The application of Machine Learning (ML) has emerged as a promising avenue for tackling PAPR issues. Along this line, this paper presents a comprehensive review of PAPR optimization techniques with a focus on ML approaches. From this survey, it becomes clear that ML solutions offer customized optimization, effective search space navigation, and real-time adaptability. In light of the demands of evolving 6G networks, integration of ML is a necessity to propel advancements and meet increasing prerequisites. This integration not only presents possibilities for PAPR reduction but also calls for continued exploration to harness its potential and ensure efficient and reliable communication within 6G networks.

Keywords: 6G networks; PAPR reduction; artificial intelligence; machine learning

1. Introduction

Over the last few decades, mobile communication networks have undergone constant evolution. In the 1980s, a significant milestone in global telecommunications was reached with the emergence of First Generation (1G) mobile communication networks [1]. During this period, various 1G standards were developed around the world. The United States adopted Advanced Mobile Phone System (AMPS) as a prominent standard for 1G, employing Frequency Division Multiple Access (FDMA) to provide voice services [1–4]. Internationally, different 1G standards gained recognition. The Nordic Mobile Telephone (NMT) system, originating in the Nordic countries and operating on the 450 MHz frequency band, became notable [5,6]. The United Kingdom and some other regions embraced the Total Access Communication System (TACS), operating at around 900 MHz [7]. Japan implemented the C-450 standard, an analog cellular system on the 450 MHz band [8,9], while France contributed to 1G with the Radiocom 2000 system [8,10]. Although they are predominantly based on analog technology, these 1G standards played a crucial role in the delivery of essential voice services. However, they faced limitations in terms of capacity, security, and efficiency [11,12].

The progression to digital technologies in subsequent generations, beginning with Second Generation (2G), led to significant improvements in communication quality and security and the introduction of data services in addition to voice communication [13]. Mobile communication networks entered the 2G era in the 1990s, revolutionizing telecommunications through widespread adoption of the Global System for Mobile Communications (GSM) standard. This advancement enabled digital encryption of data and voice signals, introduced features like roaming, and optimized spectrum utilization [14–16]. The

2G networks continued to evolve, resulting in so-called 2.5 Generation (2.5G) systems, which introduced a new packet switching technique and provided enhanced voice and data services to users [1]. The 2.5G networks brought important improvements that not only evolved the data rate but also improved established voice communications. These laid the foundations for high-performance mobile communications, laying solid ground for the next generations of mobile communication systems [17,18].

The constant evolution of wireless technologies led to an exponential necessity for high data rates and support for a large number of devices. Therefore, Third Generation (3G) mobile communication networks introduced mobile broadband services and improved data rates compared with previous generations [19,20]. In addition, 3G networks met the requirements of important vertical applications that require the exchange of high-definition images and videos, such as video calls and mobile Television (TV) [21,22]. The 4G mobile communication networks introduced a paradigm shift by adopting communication based on the Internet Protocol (IP). The 4G networks were meticulously designed to harness augmented bandwidths and elevated transmission data rates, allowing improved mobile broadband services, seamless web browsing, uninterrupted high-definition video streaming, substantially reduced latency, and a seamless handover [23–26]. These advancements were made possible through the development of cutting-edge technologies, such as Multiple-Input Multiple-Output (MIMO) and advanced signal processing techniques. Furthermore, 4G networks introduced OFDM in its physical layer, seamlessly aligning with the intricate demands for flexibility of 4G networks [1]. The use of the OFDM technique allows the system to use the available bandwidth efficiently and consequently makes the system less susceptible to interference, improving its spectral efficiency [27,28].

To meet the demands of an unprecedented number of connected devices, 5G mobile communication systems began to be implemented in 2020. These 5G networks will be able to support ultralow latency (~ 1 ms), ultrahigh data rates (~ 10 Gbps), high connectivity (~ 1 million/km²) and reliability ($\sim 10^{-5}$), and ultralow energy consumption. Furthermore, seamless integration of small cells and massive MIMO antennas has become essential to achieve the desired network capacity and coverage, especially in densely populated areas [29–31]. Additionally, 5G networks support communications at high frequencies (i.e., Millimeter Wave (mmWave) bands), which open up a range of possibilities, particularly in the context of the Internet of Things (IoT) [32]. Despite the ongoing deployment of 5G networks, both academia and industry have focused their attention on the development of 6G mobile communication networks, which are envisioned to present extraordinary advancements in terms of data rates, energy efficiency, latency, and reliability when compared with 5G networks. However, the architecture and performance components of 6G networks remain mostly undefined [33,34]. Despite this, to meet their anticipated requirements, existing communication systems will have to undergo a series of transformations that will allow them to operate under the conditions imposed by future 6G networks. These changes will affect all levels of communication, from the application layer to the physical layer. Within this transformative landscape, it is expected that the strategic integration of ML techniques plays a crucial role [35,36]. At the same time, a wide spectrum of modulation techniques are being investigated or proposed, including OFDM and its variants, to achieve a high level of spectral efficiency [37–40].

OFDM is the modulation technique for 4G and 5G wireless systems and is expected to remain indispensable to meet the rigorous demands of 6G networks [41–43]. Unlike their predecessors, 6G networks operate at higher frequencies and broader bandwidths, presenting distinctive challenges for the deployment of OFDM, particularly concerning issues such as a high PAPR and Inter-Carrier Interference (ICI), as outlined in [44]. In addition, OFDM is highly regarded for its ability to handle frequency selective fading, efficient spectrum usage, and resistance to interference [45]. However, it faces obstacles due to the close spacing of the subcarriers inherent in its design, which leads to a high PAPR. This elevated PAPR has several implications. First, it can result in intermodulation distortion, degrading the quality of the transmitted signal. Moreover, it decreases the efficiency of the amplifiers used in the

system, as they must accommodate the high peaks by operating with a broader dynamic range. Finally, a high PAPR can cause unwanted signals to escape into neighboring frequency bands, causing interference with other wireless systems [46,47].

In scenarios where a high PAPR is not effectively managed, advanced signal processing and amplifier designs are needed, which can introduce additional delays in signal transmission and reception. These delays can lead to increased latency, which is undesirable in 6G networks that strive for ultralow-latency communication [48]. Consequently, optimizing the PAPR is crucial not only to mitigating interference, achieving high spectral efficiency, and reducing distortion, as previously mentioned, but also to obtaining low-latency signals in the context of 6G networks [49].

The literature is filled with a wide range of strategies designed to address the challenge of reducing the PAPR in communication systems [50]. These strategies encompass a spectrum of approaches, including conventional methods tested over time and more advanced techniques [51]. Some of these prominent techniques are clipping and filtering, Partial Transmit Sequence (PTS), Selected Mapping (SLM), interleaving, Tone Injection (TI), Tone Reservation (TR), and Active Constellation Extension (ACE). Clipping and filtering have gained recognition for their simplicity and effectiveness in mitigating the PAPR. However, they come with a notable caveat: the potential introduction of signal distortion, which can compromise the overall transmission quality [52,53]. PTS distinguishes itself through its ability to achieve substantial reductions in the PAPR. However, this method has a trade-off: it is a complex and computationally demanding approach that requires multiple iterations to achieve its objective [54,55]. SLM represents another powerful strategy for reducing the PAPR, with the added benefit of being less susceptible to signal distortion. However, it places significant demands on processing power and is sensitive to nonlinear channel conditions [56,57]. Interleaving exhibits effectiveness in PAPR reduction, particularly when used in conjunction with other techniques. However, its efficacy depends on specific configurations and may not yield optimal results when used alone [58,59]. Moreover, TI has emerged as an effective method for PAPR reduction, and it is less prone to distortion. However, to optimize its performance, it may require access to additional channel information [60]. TR shines in its high effectiveness in PAPR reduction, particularly in systems with nonlinear channels. However, it is a complex method that demands the allocation of additional tones, which can result in a reduction in spectral efficiency [61]. Finally, ACE stands out as another powerful tool in the arsenal for PAPR reduction, offering greater flexibility with different constellation schemes. However, it introduces some overhead and requires careful allocation of constellation points [62].

Motivated by the importance of the previously described PAPR techniques, in [50], the authors presented a review of the literature, recent research findings, simulations, and complexity analyses of these techniques. Organized into three primary categories for PAPR reduction schemes (i.e., signal distortion, multiple signaling, and coding), the paper presents a systematic approach to comprehending various methods that address the high PAPR challenge in OFDM systems. It serves as an invaluable resource for emerging researchers seeking a comprehensive understanding of the PAPR. However, the techniques discussed in [50] for PAPR reduction in wireless communication systems may fall short in addressing the complex and dynamic nature of the problem. Conventional approaches often rely on predefined signal processing strategies and parameter configurations which might not sufficiently adapt to the evolving conditions of wireless channels. While these methods may offer some degree of PAPR reduction, they might lack the adaptability required to consistently maintain optimal signal quality in the face of changing communication environments. Therefore, the authors of [50] acknowledged the significance of classical techniques, but they notably refrained from delving into ML approaches for PAPR reduction.

1.1. Contributions

To overcome the limitations of classical PAPR reduction techniques, recent works have revealed exciting prospects for applying ML to significantly reduce the PAPR while

keeping the computational complexity at a minimum, which we will elaborate upon in more detail in Section 4, demonstrating that the use of ML has great potential for the field of wireless communications. Therefore, motivated by the importance of ML to 6G networks, this work presents a comprehensive review of PAPR reduction techniques based on ML. More specifically, this survey not only outlines various approaches but also presents the results, contributions, and limitations associated with several methods based on ML. In the process, this study endeavors to provide readers with comprehensive insights into the application of ML in this specific domain. In this context, it is worth highlighting that our research occupies a unique and indispensable position within the existing scholarly body of work because, to the best of our knowledge, this is the first work to present a review of ML-based techniques for PAPR reduction. Consequently, the main contributions of this paper can be summarized as follows:

- We present a comprehensive analysis of the classical PAPR solutions, highlighting their main features, advantages, and challenges;
- We present a concise review of the main PAPR reduction approaches based on ML and describe their main features, advantages, and challenges;
- We discuss the role of ML in the PAPR issue for 6G networks, giving directions for future work.

1.2. Research Methodology

To select the previously cited works, we considered the following methodology: (1) We identified relevant keywords that were aligned with the scope of the survey. These keywords included terms like PAPR, ML, and OFDM. (2) We performed comprehensive searches of several recognized academic databases, including Elsevier, IEEE Xplore, and SpringerLink. We used the keywords defined in the first step to find potentially relevant articles. (3) The articles found were initially filtered based on titles and abstracts to ensure that they were directly related to our research scope. Articles that did not meet the criteria were excluded. Special care was taken to avoid articles published in scientific events without peer review or little recognition. (4) Finally, after the initial filtering, we proceeded to the full reading of the selected articles. This allowed us to make an assessment of their contributions and overall relevance. Some references from the selected works were separated for detailed reading and cataloging, feeding back into the process. To finish, Figure 1 illustrates the post-selection distribution of the selected articles across the databases.

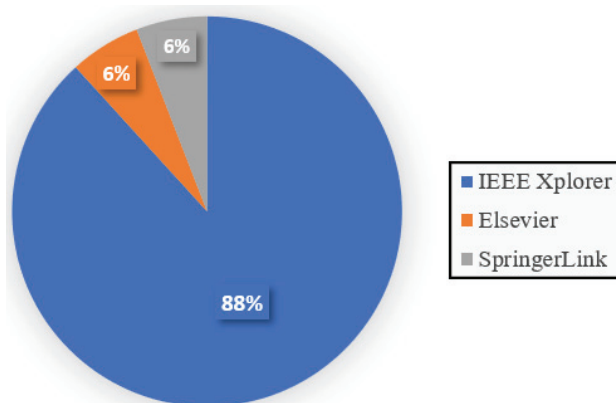


Figure 1. Illustration of the post-selection distribution across the databases.

To improve the readability of this survey, its organization is illustrated in Figure 2. The remainder of this paper is organized as follows. Section 2 presents a review of the main concepts of the OFDM modulation technique. Section 3 reviews the classical PAPR reduction techniques. Section 4 reviews the PAPR solutions based on ML. Section 5

describes the main challenges of applying ML techniques to PAPR optimization. Finally, Section 6 concludes the paper.

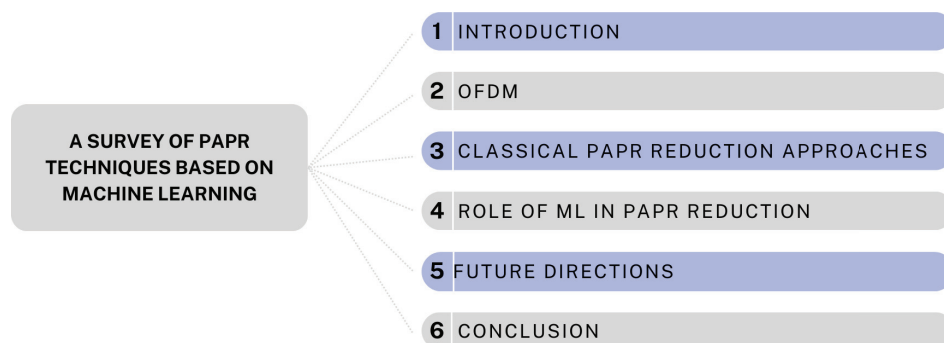


Figure 2. Organization of this survey.

2. OFDM

Frequency Division Multiplexing (FDM) and OFDM are pivotal multiplexing and modulation techniques with deep historical roots and practical significance in contemporary wireless communication systems [63,64]. Their development over time has been driven by the imperative need to overcome a range of challenges associated with wireless data transmission, including issues such as interference, fading, and the constraint of limited bandwidth. These technologies serve as the foundation for numerous communication standards and are expected to continue to play a crucial role in the evolution of wireless communication systems [49,64,65].

More specifically, FDM is a classical multiplexing technique that splits the available frequency spectrum into multiple nonoverlapping bands, with each band allocated to a different channel or user. Each channel operates at a specific frequency, and multiple channels can coexist in the same communication medium without interference [66]. On the other hand, OFDM divides the available frequency spectrum into multiple orthogonal subcarriers, which are closely spaced, and each subcarrier is allocated a portion of the data, efficiently utilizing the spectrum. The orthogonality between subcarriers prevents interference, enhancing reliability [67]. Therefore, as illustrated in Figure 3, FDM modulation is less efficient in terms of spectrum utilization compared with OFDM [68].

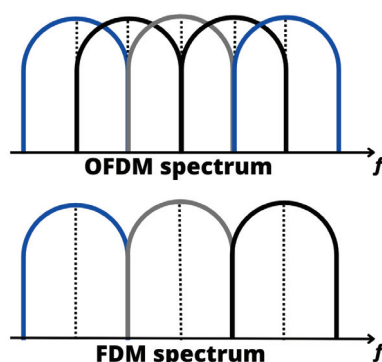


Figure 3. Ideal spectrum illustration for FDM and OFDM techniques.

In broadband communication systems, OFDM enables simultaneous data transmission by splitting the spectrum into multiple orthogonal subcarriers. The OFDM modulation relies on Inverse Fast Fourier Transform (IFFT) for transmission and Fast Fourier Transform (FFT) for reception. More specifically, the IFFT allocates data symbols to subcarriers, with a Cyclic Prefix (CP) added before transmission to avoid multipath effects, resulting in the signal $x(n)$ [69]. The CP provides a guard period for the demodulation process, removing delayed echoes and facilitating accurate symbol detection in the frequency domain [70].

It is imperative to emphasize the importance of meticulously choosing the CP length with due regard to the expected channel delay spread [71]. Prudent selection of the CP length holds the utmost importance in wireless communication systems, especially in digital communication systems such as those used in mobile networks, Wi-Fi, and various wireless technologies [72]. This careful consideration plays a crucial role in minimizing the effects of temporal dispersion, ensuring reliable signal transmission, and enabling precise data reception. A well-considered choice for the CP length enhances the overall resilience and optimal functionality of the communication system [73,74]. For clarity, Figure 4 illustrates the OFDM transmitter/receiver system.

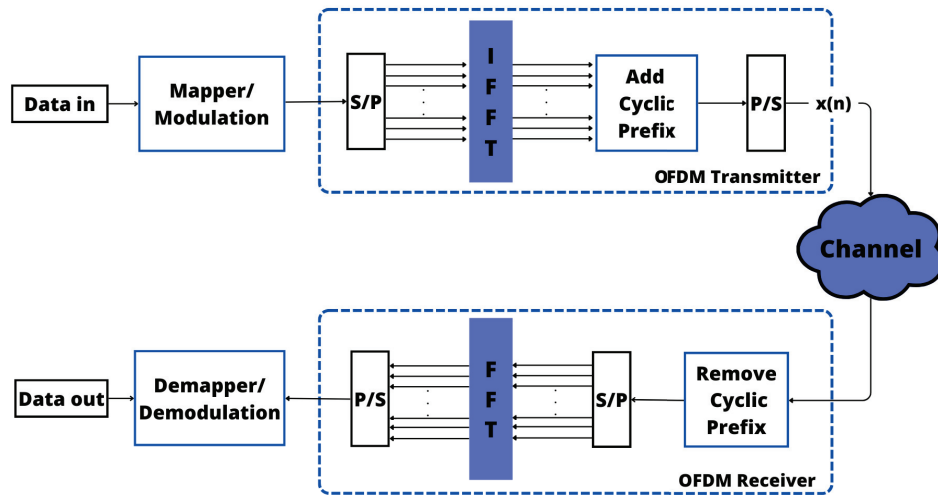


Figure 4. Block diagram of a baseband OFDM system.

One of the main technical challenges inherent to OFDM systems is the emergence of an amplitude peak in the time domain due to constructive interference among multiple subcarriers. This intricate phenomenon is conventionally called the PAPR issue [75]. The PAPR has been a fundamental metric that characterizes the behavior of multicarrier signals, and it can be defined as [50]

$$\text{PAPR} = \frac{\max |x(n)|^2}{\mathbb{E}|x(n)|^2} \quad (1)$$

where $\mathbb{E}|x(n)|$ is the mean value of the OFDM signal $x(n)$ and $\max |x(n)|$ denotes the maximum amplitude peak of the OFDM signal.

As the number of carriers increases, phase accumulation arises among specific carriers during transmission amplification [76]. This results in power peaks (i.e., the amplitude of the signal reaches saturation levels which induce signal clipping and, consequently, result in abrupt discontinuities that add noise due to the superposition of multiple signals). In the context of OFDM, excessive noise can amplify interference between adjacent frequency bands, adversely affecting both the upper and lower bands. Therefore, a high PAPR value indicates a more pronounced deviation of the signal from its average power level, which results in lower amplifier performance and leads to height saturation and noise generation [77,78].

Figure 5 illustrates the Complementary Cumulative Distribution Function (CCDF) of the PAPR across different numbers of subcarriers (32, 128, 512, and 2048) with 16 Quadrature Amplitude Modulation (QAM) and with two of each subcarrier serving as pilot subcarriers, bearing the values -1 and 1 . This arrangement resulted in a batch size of 1000, forming a matrix with 1000 rows and N columns. Notably, the pilot subcarriers were strategically placed in the second and $N-2$ columns. In addition, the CCDF is a valuable tool for assessing the reliability of a signal at various PAPR levels. This helps us understand the likelihood of the signal functioning correctly and maintaining its integrity.

Notably, from Figure 5, an observable trend emerges: higher subcarrier counts correlate with elevated PAPRs. This outcome is a consequence of the cumulative impact of

subcarrier summations during the amplification process, as previously discussed. The graphical representation underscores the direct relationship between the number of subcarriers and the resulting PAPR, providing valuable insights into the system performance under varying configurations.

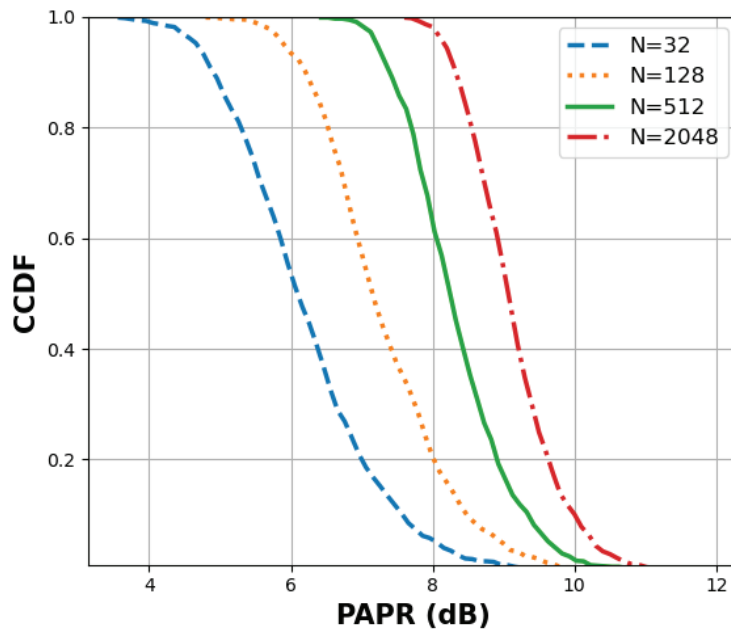


Figure 5. CCDF of the PAPR for 32, 128, 512 and 2048 subcarriers.

3. Classical PAPR Reduction Approaches

In the domain of OFDM systems, numerous techniques have been devised to mitigate the problem of the PAPR, which is a widespread issue in these systems. Such techniques can be classified into signal distortion approaches, signal scrambling methods, and hybrid and coding techniques [79,80]. However, this survey focused specifically on signal distortion approaches and signal scrambling methods. This is because our emphasis is on solutions that strike a harmonious balance between efficacy and simplicity, rendering them more accessible and applicable in practical implementations of OFDM systems [81]. Signal distortion approaches involve making direct modifications to the original signal in an effort to reduce the PAPR while striving to maintain the quality of the data [82]. These strategies include actions such as cutting, filtering, and ACE. In contrast, signal scrambling methods take an indirect approach to reduce the PAPR by carefully altering the signal without causing significant distortion to the data. Techniques falling within this category include SLM, PTS, interleaving methods, TI, and TR. The choice of technique to use depends on the specific requirements, trade-offs, and limitations of the particular communication system under consideration [80,83]. The main classical PAPR reduction techniques previously cited are described next.

3.1. PTS

The PTS technique is utilized in digital communication systems to address the challenge of a high PAPR in Transmit (Tx) signals without adding distortion to the signal. This method involves dividing the input data stream into multiple sub-blocks which are independently modulated. After splitting the data stream, the Inverse Discrete Fourier Transform (IDFT) of each sub-block is obtained and weighted by a phase factor which is optimized to minimize the PAPR of the sum of the signals of all sub-blocks [84]. The main step of the PTS technique is the phase optimization process, which strategically manipulates the phases of the sub-blocks within the signal without altering the original Tx signal [85]. In this technique, the reduction in the PAPR is dependent on the sub-blocks' size and the number of allowed phase factors. Despite its effectiveness in mitigating PAPR issues, the PTS computational complexity can pose challenges, particularly depending on the modulation order [86–88]. In addition, it

is important to highlight that the PTS technique needs some bits of side information to be transmitted, conveying the phase factors, which reduces its spectral efficiency.

Integration of the PTS scheme into OFDM systems is pivotal for reducing the PAPR and enhancing Bit Error Rate (BER) performance. Figure 6 visually depicts this integration, illustrating how the PTS scheme is implemented within an OFDM system [89,90]. In Figure 6, the phase optimization is particularly emphasized after the IDFT operation [91]. By judiciously selecting these phase coefficients, the PTS technique minimizes constructive interference, leading to an overall reduction in the PAPR of the Tx signal. This reduction is highly desirable as it not only enhances BER performance but also improves the reliability and resilience of the wireless communication system [87,88].

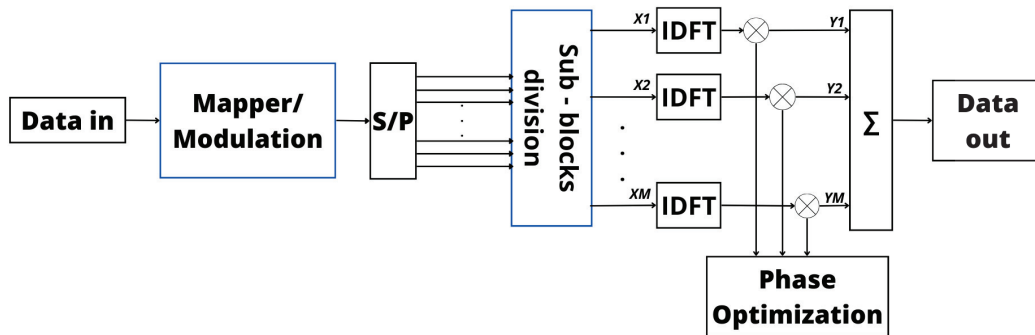


Figure 6. Block diagram of OFDM transmitter with PTS.

3.2. SLM

In SLM, the process of generating alternative signals from a single information signal involves a systematic and adaptive approach. This involves creating a set of potential signals, called “candidate signals”, by introducing variations based on different phase sequences. These candidate signals are then subject to adaptive optimization techniques, and then the signal with the minimum PAPR is transmitted [92]. To elaborate further, think of these candidate signals as alternative versions of the original signal, with each slightly modified in a systematic way. The variations are introduced in a controlled manner to cover a range of possibilities. Then, adaptive optimization techniques come into play to dynamically refine these candidate signals during the transmission process [93,94].

Once generated, the signals undergo a thorough evaluation with careful scrutiny to identify the one with the lowest PAPRs. This comprehensive method ensures both the efficiency and reliability of the transmission. The ultimate goal of SLM is to minimize signal distortion and enhance power efficiency, a critical aspect of the overall performance of the communication system. Furthermore, the selected phase sequence needs to be transmitted as side information so that the receiver can recover the original sequence, leading to a reduction in the spectrum efficiency. Moreover, it is important to note that the computational complexity of SLM increases with the number of candidate signals generated [94,95]. Each signal requires meticulous evaluation and comparison, increasing the computational overhead, especially when dealing with a large number of candidates [96].

In essence, SLM involves a careful and adaptive exploration of potential signal variations to optimize the transmission process, ensuring both efficiency and reliability in communication systems. Figure 7 provides a visual representation of this approach in the Tx, highlighting the strategic incorporation of signals before the IDFT process to optimize power efficiency within the signal. These preemptive signals are meticulously crafted to alleviate the elevated PAPR encountered in the transmitted signal [94,97,98]. Moreover, from Figures 6 and 7, it is possible to verify that the phase variation is applied at different points in the transmitter in PTS and SLM, producing different results but with the same goal.

SLM’s effectiveness is attributed to its capability to substantially decrease the PAPR while causing a smaller impact on the spectral efficiency with respect to PTS. The direct approach of SLM and its proven success in PAPR reduction position it as a preferred method

for striking a balance between reducing the PAPR and preserving spectral efficiency in OFDM communication systems [94,97,98].

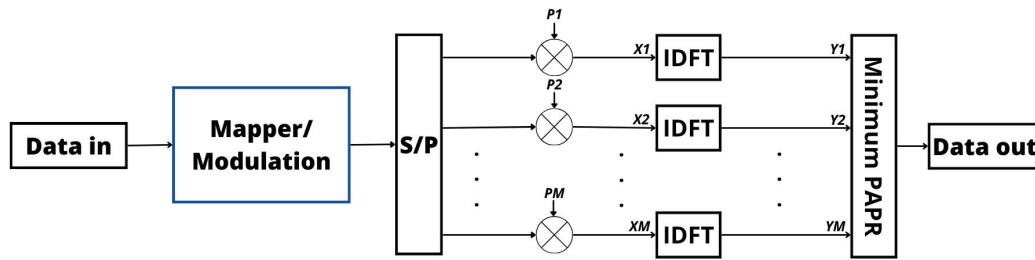


Figure 7. Block diagram of OFDM transmitter with SLM.

3.3. Interleaving

This PAPR reduction technique is similar to the SLM and PTS approaches, but in the interleaving technique, interleaves are considered to generate multiple signals instead of phase sequences. More specifically, the interleaving technique rearranges the order of the transmitted symbols to decrease the chances of encountering high-power peaks. Then, the signal with the smallest PAPR is chosen for transmission. By interleaving the symbols, the high-power peaks are more evenly spread out, reducing the likelihood of their occurrence [99]. This enhances the BER performance by mitigating the impact of significant power variations. Moreover, interleaving does not introduce additional signal distortions or require intricate processing, ensuring a low BER [100]. However, to correctly decode the transmitted signal, it is necessary to transmit some side information bits to inform the receiver which interleaver was considered to generate the selected signal for transmission [99,100].

Regarding complexity, the interleaving process is generally straightforward, typically involving symbol permutation according to a predefined pattern. Its computational complexity is usually low, making it a practical choice for implementation in communication systems [101]. However, it is essential to note that the interleaving technique cannot achieve the same level of PAPR reduction as more advanced techniques like SLM. Therefore, the achieved PAPR reduction through interleaving should be carefully balanced with the desired BER performance. In comparison with SLM and PTS, interleaving introduces a distinct dimension to signal processing, emphasizing error resilience and the ability to recover from channel-induced disruptions [99,102]. Figure 8 illustrates a block diagram of the OFDM transmitter with the interleaving approach. This visual representation provides a clear overview of how the interleaving technique is applied in an OFDM system to reduce the PAPR and improve BER performance [99].

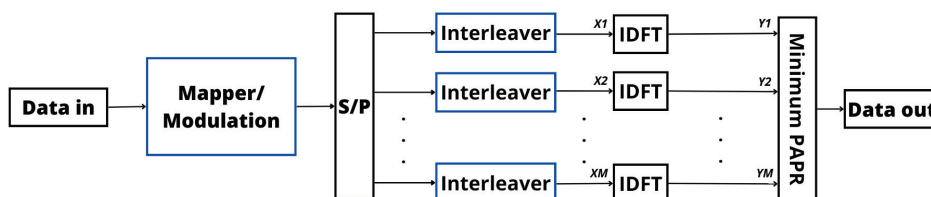


Figure 8. Block diagram of OFDM transmitter with interleaving.

3.4. Clipping and Filtering

Clipping and filtering play a pivotal role in the management of both signal characteristics and system performance. This PAPR reduction approach limits the signal power to a predetermined Clipping Level (CL) by using a clipper (i.e., if the signal exceeds the CL, then the signal is clipped; otherwise, the signal does not suffer any change). Clipping serves as a highly effective approach to governing signal amplitudes and diminishing the PAPR without the need for side information at the receiver [103]. However, although the clipping strategy is performed at the transmitter, the receiver must estimate the location and size of the clip, which is difficult in practice and results in the introduction of nonlinear distortions, thereby bearing

implications for BER performance. The distortions that arise from clipping can lead to an increased BER, particularly when the clipping threshold is aggressively configured [103,104].

In contrast, filtering the signal after clipping is a key solution to reducing these distortions and consequently increasing the overall signal quality [105]. Filtering can eliminate out-of-band distortions, thereby fostering spectral efficiency and curtailing interference with adjacent channels [106]. Nonetheless, it is crucial to strike a balance between mitigating out-of-band distortions and reducing the peak power of the signal, as excessive filtering may give rise to power inefficiencies [107]. Furthermore, it is worth noting that the intricacy of the clipping and filtering processes can impose a high computational demand on the system due to the incorporation of advanced filtering algorithms or iterative procedures. Consequently, system designers must evaluate the trade-offs between computational complexity, BER, and power efficiency during the implementation of these approaches [108,109].

The above considerations underscore the importance of carefully configuring and optimizing the clipping and filtering processes to attain the desired signal quality while reducing the system complexity, BER, and power consumption [106]. Figure 9 depicts the implementation of clipping and filtering at the OFDM transmitter to manage the PAPR, optimize BER performance, and enhance signal quality [110].

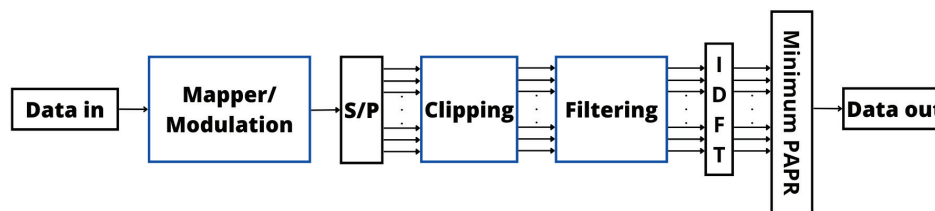


Figure 9. Block diagram of OFDM transmitter with clipping and filtering.

3.5. TI and TR

TI is an approach that involves the injection of specific tones or subcarriers into the transmitted signal to mitigate high-power peaks [111]. The primary goal of TI is to strategically redistribute the power peaks across the signal, resulting in a reduced PAPR [112]. One of the key advantages of the TI method is the ability to improve the BER performance while maintaining relatively low complexity. In addition, the TI technique does not need side information at the receiver. Therefore, the TI approach introduces minimal signal distortions and can be implemented with modest computational resources [113,114]. On the other hand, TR is a distinct technique that focuses on allocating a portion of the available subcarriers in an OFDM system for the inclusion of additional tones. These reserved tones are carefully designed to nullify high-power peaks in the transmitted signal. The TR method aims to enhance the BER performance by mitigating the effects of significant power variations [115,116]. Similar to the TI approach, TR can be implemented with relatively low complexity, as it involves generating and adding reserved tones to the signal [117]. However, different from the TI method, the TR approach requires additional side information at the receiver to correctly decode the received signal [117].

In addition, TI is known for its efficient redistribution of power peaks, which directly contributes to reducing the PAPR and improving the BER. Moreover, TI introduces minimal distortions and is suitable for practical applications [118]. On the other hand, TR offers a dedicated approach by reserving specific subcarriers for additional tones, further reducing the impact of power variations on the transmitted signal. The choice between TI and TR should be made based on the specific requirements and characteristics of the communication system [119,120]. TI and TR are not one-size-fits-all solutions, and their effectiveness can vary depending on factors such as the number of reserved tones, their positioning, and the prevailing channel conditions [121]. Both techniques require careful consideration of the trade-off between PAPR reduction and BER performance [122]. Finally, Figures 10 and 11 illustrate the block diagram of the OFDM transmitter for TR and TI, respectively [123]. From these figures, it is possible to verify that, different from TI, the TR

approach adds nulls in the data subcarrier positions to avoid the high-power peaks in the transmitted signal [50]. The presence of zeros in TR serves to generate interference patterns, aiding in a reduction in the PAPR when integrated with the primary data signal. This forms a crucial component of the design approach aimed at efficiently mitigating or minimizing peaks within the transmitted signal. In contrast, TI accomplishes PAPR reduction by introducing artificial tones directly into the primary data signal. The lack of zeros between data subcarriers streamlines the injection of these tones, and optimization algorithms ensure that these artificial tones interact seamlessly with the data subcarriers [50,120,122,123].

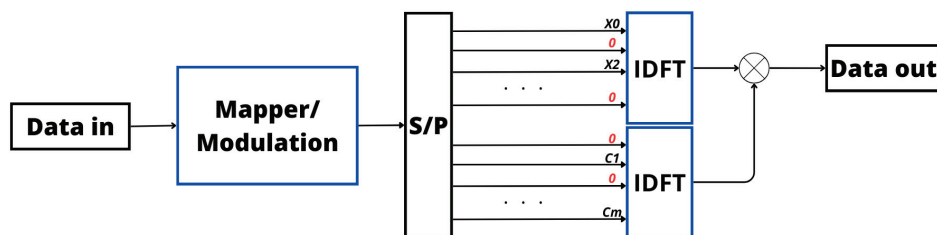


Figure 10. Block diagram of OFDM transmitter for TR.

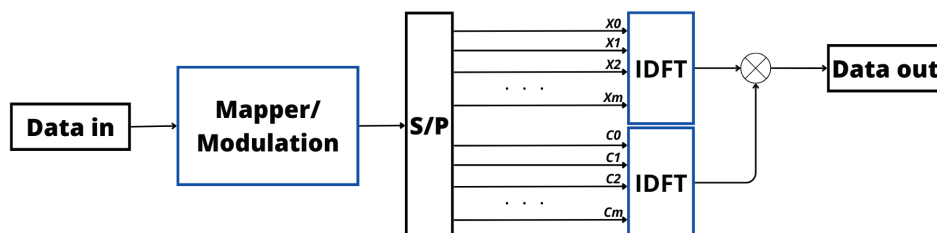


Figure 11. Block diagram of OFDM transmitter for TI.

3.6. ACE

The ACE technique is used to expand the constellation points within a transmitted signal to reduce the impact of high-power peaks [124]. More specifically, the constellation points are extended toward the outside of the original constellation, and regions named “feasible regions” are created as illustrated in Figure 12 [123]. This expansion of constellation points enlarges the signal space, resulting in a more even distribution of signal amplitudes and, consequently, a lower PAPR and BER due to reduction of the influence of significant power fluctuations [125]. Furthermore, the ACE technique does not require any side information at the receiver and does not add distortions in the signal as it is considered a signal scrambling approach. Implementation of the ACE technique presents moderate complexity and requires adjustments to the way constellation points are mapped and processed [126,127]. ACE introduces some additional computational load, as expanding constellation points requires careful design considerations and may require iterative optimization algorithms [128]. Nevertheless, its complexity remains manageable and practical for implementation [129].

It is crucial to understand that the effectiveness of ACE in reducing the PAPR, and its impact on the BER depends on various factors [130]. These factors include the specific constellation design, Signal-to-Noise Ratio (SNR), and channel impairments. When selecting and optimizing extended constellation points, a balance must be struck between reducing the PAPR and maintaining BER performance [131,132]. In some situations, expanding constellation points can introduce additional signal distortions, which could potentially lead to an increased BER. Therefore, a trade-off between PAPR reduction and the BER needs to be carefully considered [133].

We finish with Table 1, which summarizes the main features of the aforementioned PAPR reduction techniques. More specifically, Table 1 compares the classical PAPR techniques in terms of computational complexity, BER escalation, the need for side information, and whether the technique adds distortion to the signal or not. Evaluating these factors is

crucial for determining the efficacy and cost-benefit of a given method [132]. In addition, for more details about the classical PAPR reduction techniques, readers should consult [50].

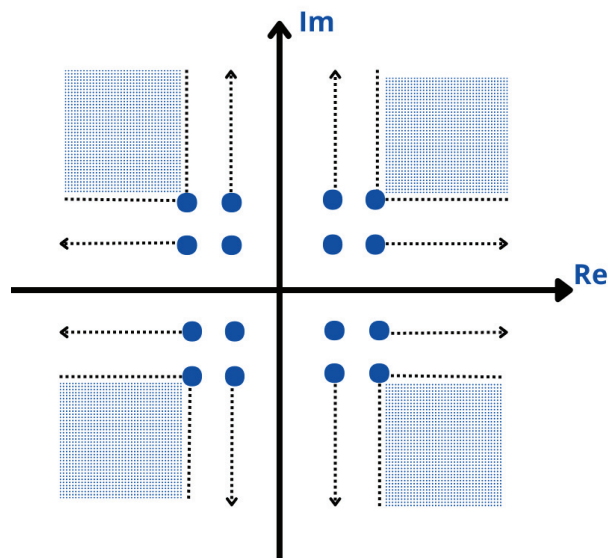


Figure 12. ACE constellation for 16 QAM.

Table 1. Comparison between PAPR reduction techniques.

Technique	BER Increase	Complexity	Side Information	Add Distortion
PTS	no	high	yes	no
SLM	no	high	yes	no
Interleaving	no	high	yes	no
Clipping and Filtering	yes	low	no	yes
TI	no	high	no	no
TR	no	high	yes	no
ACE	no	high	no	no

4. Role of ML in PAPR Reduction

The techniques discussed in Section 3, covering PAPR reduction in wireless communication systems, may fall short in addressing the complex and dynamic nature of the problem. Conventional approaches often rely on predefined signal processing strategies and parameter configurations that might not sufficiently adapt to the evolving conditions of wireless channels. While these methods may offer some degree of PAPR reduction, they might lack the adaptability required to consistently maintain optimal signal quality in the face of changing communication environments. On the contrary, ML plays a key role in overcoming these limitations. It offers a more sophisticated and adaptive approach to PAPR reduction in OFDM systems [134]. ML algorithms demonstrate remarkable capabilities in finding intricate patterns within PAPR levels, making accurate predictions about future PAPR levels [135,136]. The ability to dynamically adjust transmission parameters in real time further distinguishes ML techniques, ensuring a continuous optimization process that minimizes PAPR levels while preserving the optimal signal quality.

Moreover, ML enables exploration in a vast solution space of signal processing strategies and parameter configurations, leading to the identification of high-performance approaches for PAPR reduction [137]. Through algorithmic optimization and feature selection, ML assists in the development of PAPR reduction strategies that strike a balance between performance and complexity. This is crucial for practical implementation in real-world communication systems. The adaptability of ML techniques ensures seamless integration of PAPR reduction strategies within dynamic communication systems [138]. This adaptability adapts to the evolving requirements of communication systems, accommodating changes

in the environment and communication protocols [139]. Therefore, ML not only assists with reducing the PAPR but also contributes to achieving notable enhancements in signal quality by minimizing signal distortion and optimizing power consumption [140].

4.1. ML Technique Review

Motivated by the above cited advantages, several works in the literature explored different ML techniques [134,135,141–155] to reduce the PAPR. Therefore, to improve the understanding of readers not particularly familiar with ML, we next present a brief explanation of the ML techniques considered by the surveyed papers.

4.1.1. Artificial Neural Network (ANN)

An ANN is an ML technique that is inspired by the functioning of the human brain. It is made up of fundamental units called neurons, which are organized in layers. The input layer receives the initial data and passes them to the next layer. Each neuron in this layer represents a specific feature of the input data. Hidden layers are intermediate layers between the input layer and the output layer. Each neuron in a hidden layer processes information from previous neurons and forwards it to the next one. The output layer generates the final result of the neural network. This output can be a rating, a continuous value, or any other form of representation, depending on the task performed by the network. Each connection between neurons has an associated weight, determining the strength and direction of the influence of one neuron on the other. During training, the network adjusts these weights to minimize the difference between the predicted outputs and desired outputs using a process called backpropagation. This involves calculating the error at the network output and adjusting the weights back across layers to improve prediction accuracy. This process is repeated iteratively with training datasets until the network achieves satisfactory performance [156,157].

4.1.2. Deep Neural Network (DNN)

A DNN with encoder and decoder architecture is often used in Deep Learning (DL) tasks, such as autoencoder models, and also in more complex architectures, such as those used in convolutional neural networks. More specifically, the encoder presents five components: (1) an Fully Connected (FC) layer, which is a dense layer that connects each neuron to all neurons in the next layer, making it easier to capture global patterns in the data; (2) Batch Normalization (BatchNorm), a component which normalizes the activation values, reducing internal covariance and consequently accelerating training and improving generalization; (3) Rectified Linear Unit (ReLU), an activation function that introduces nonlinearity to the model, allowing the incorporation of complexity between layers; (4) another FC layer, namely a dense layer which possibly has fewer neurons than the previous layer, helping to compress information; and (5) a sigmoid function, which is a final activation function in the encoder layer. It can be used to generate values between 0 and 1, being useful in binary classification problems or representing probabilities [158].

In addition, the decoder is composed of the same components as the encoder, but in the decoder, the FC layer receives the compressed representation of the encoder data and seeks to reconstruct the original data. Similar to the BatchNorm layer in the encoder, it normalizes the activation values during training to speed up convergence. ReLU adds nonlinearity to the network, enabling the learning of complex relationships in the data. The FC layer generates the final output of the decoder, which must be a reconstruction of the original data. BatchNorm normalizes the activation values before the last ReLU layer in the decoder. ReLU is the last activation layer to introduce nonlinearity in the decoder output [159].

4.1.3. Monte Carlo Tree Search (MCTS)

The MCTS emerges as an innovative approach to deal with complex search problems in vast and little-explored decision spaces. This method is based on stochastic simulation and adaptive evaluation, offering an effective solution for situations where an exhaustive

search of all possibilities is impractical. In simple terms, MCTS represents the problem as a tree, where each node represents a possible state and the edges indicate achievable actions. Starting from a root node that represents the initial state of the problem, MCTS follows an iterative process that involves selection, expansion, simulation, evaluation, and backpropagation. This adaptive approach allows one to explore the decision tree efficiently, focusing efforts on the most promising areas of the search space. MCTS excels in scenarios in which the absence of an explicit model of the environment or an extensive search space makes exhaustive searching impractical. By incorporating stochastic simulations and dynamic evaluations, MCTS proves to be a useful tool for solving complex problems, providing practical and adaptable solutions in diverse applications [160].

4.1.4. Deep Unfolding Network (DUN)

A DUN is a specific type of model-driven Neural Network (NN) that operates in layers. Each layer performs progressive analysis of the data, learning more complex representations over time. It is interesting to think of this as a series of observation steps, starting with simple details and gradually moving to more abstract features. This process allows the network to understand patterns and become effective in performing complex tasks [161].

4.1.5. Random Sample Consensus (RANSAC)

RANSAC is an iterative ML method that is used to estimate the parameters of a mathematical model considering a database that contains outliers. This non-deterministic method is widely used when there is a large amount of noisy data, and the main goal is to design a model that fits the data well. Then, RANSAC can be used to classify the remaining data as either inliers or outliers until a satisfactory model is found [134].

Finally, the next subsection presents a concise review of PAPR reduction solutions based on ML as documented in the literature. This exploration aims to provide insights into the diverse range of ML techniques employed to address the challenges associated with the PAPR in wireless communication systems.

4.2. PAPR Reduction Approaches Based on ML

As previously mentioned, ML is a key to solving the PAPR issue in OFDM systems. Therefore, several works in the literature have studied the application of different ML techniques to reduce the PAPR [134,135,141–155]. Therefore, as the main goal of this paper is to show the role of ML in PAPR reduction, we next present the reviewed papers, which are organized following the considered ML area, and describe their main contributions, features, and limitations.

4.2.1. ANN [135,141–146]

The main ML technique considered in the literature to solve the PAPR reduction problem is the ANN [135,141–146]. These proposed solutions present the same main steps, which are illustrated in Figure 13. More specifically, when an ANN technique is considered, the OFDM signal generated by the IFFT is the input of the ANN, which is composed of multiple FC layers. For each layer, a weight matrix and a bias matrix are applied to find the weighted sum of the inputs, and then an activation function is considered. In a supervised training process, these weights are optimized with the goal of minimizing a cost function, as with the PAPR. Then, in regular operation, the trained ANN is used to generate an OFDM signal with a low PAPR.

After understanding the basic operation of an ANN-based PAPR reduction solution, we can discuss in more detail the main approaches in the literature [135,141–146]. More specifically, in [141], an innovative approach is proposed to mitigate the PAPR in a Visible Light Communication (VLC) Direct Current-biased Optical OFDM (DCO-OFDM) system, employing a deep autoencoder. This method leverages a deep autoencoder to diminish the PAPR by inputting the constellation symbols into the autoencoder. When introducing a novel PAPR reduction scheme for VLC DCO-OFDM, the performance of this approach

was scrutinized in comparison to conventional methods using CCDF. The proposed deep autoencoder demonstrated a notable 6.5 dB reduction in the PAPR within the VLC DCO-OFDM system while preserving an acceptable BER and efficiently curtailing the PAPR for high-speed data transmission. In [142], the authors introduced a novel method called Neural Network ACE (NN-ACE), and this paper aims to mitigate the PAPR in OFDM systems. NN-ACE involves extending the constellation using an autoencoder, where the autoencoder is trained to expand the input symbols in the complex domain, resulting in a reduced PAPR. In diverse channel conditions, NN-ACE demonstrates similar effects in extending Quadrature Phase Shift Keying (QPSK) constellations and reducing the PAPR compared with traditional ACE methods. However, it surpasses conventional schemes. When analyzing the CCDF curves of the PAPR and the symbol power, it is possible to note that while the ACE method decreases the PAPR, it comes at the expense of increased signal power. On the other hand, NN-ACE achieves a lower PAPR but at the cost of higher power consumption. Notably, when NN-ACE is implemented, there is a degradation in the overall BER performance.

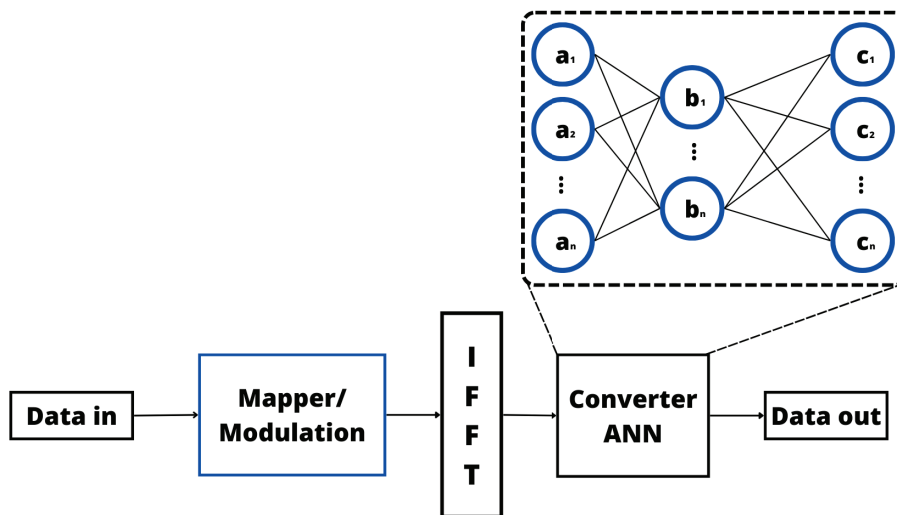


Figure 13. Illustration of the ANN transmitter.

In addition, the authors of [143] presented an innovative approach to improving the efficiency of indoor VLC-OFDM systems. It addresses challenges like a high PAPR and Light Emitting Diode (LED) nonlinearity by using a unique PAPR reduction scheme involving weighted autoencoder and amplitude clipping techniques. DL is applied for adaptive acquisition and optimization of the constellation mapping, de-mapping, and phase factors. The study systematically investigated the impact of hyperparameters, network architecture, and channel types on performance metrics, focusing on the BER and PAPR. The results show that the hybrid autoencoder method achieved a significant PAPR reduction of about 12 dB, demonstrating resilience to LED nonlinearities and improved BER performance compared with conventional methods. Tailored for Layered Asymmetrically Clipped OFDM (LACO-OFDM)-based VLC systems, the hybrid autoencoder network adapts to diverse channel conditions, leading to a substantial reduction in the PAPR. The simulation results suggest seamless integration into typical VLC-OFDM systems, offering favorable PAPR and BER performance even in the presence of Inter-Symbol Interference (ISI) in the diffused optical wireless channel. While acknowledging limitations in real-world implementations and variations in system conditions not fully captured in simulations, the study provides valuable insights into addressing challenges in indoor VLC-OFDM systems.

In [144], the mitigation of the PAPR in a Generalized Frequency Division Multiplexing (GFDM) system was accomplished through the utilization of an encoder-decoder NN, commonly referred to as an autoencoder. The assessment of the proposed PAPR-Reducing Network (PRnet) strategy was carried out by analyzing the CCDF of the PAPR. The

effectiveness of this approach was substantiated through a comparative analysis against conventional methods, illustrating its superior capability in PAPR reduction. Moreover, the authors of [145] introduced a new method for mitigating the high PAPR issue in wireless transmitters using OFDM systems. The proposed method utilizes a real-valued NN, which consists of a PAPR reduction module and a PAPR decompression module. These modules are jointly trained offline to simultaneously reduce the PAPR and minimize the BER. At the receiver, the PAPR decompression module aids in signal reconstruction to minimize the BER without introducing significant distortion. Through extensive simulations, the method demonstrated superior performance in terms of the PAPR and BER compared with existing methods like clipping, the NN model, and PRnet. Notably, there was a reduction in the BER of 3.5 times at a 20 dB SNR. Importantly, the proposed method exhibited lower computational complexity than PRnet, making it a promising solution. It is crucial to acknowledge the limitations and specify scenarios where the proposed method remains applicable, such as in MIMO systems, without impacting the overall system structure.

The work in [135] aimed to tackle the challenge of nonlinear relationships between the intended and actual PAPR in frequency-selective PAPR reduction. The proposed solution, PAPRer, employs a novel ML-based approach to automatically and precisely adjust the optimal PAPR target. This involves utilizing features related to the clipping noise filter and minimizing a defined loss function through supervised learning. In the context of 5G New Radio, PAPRer was demonstrated through numerical evaluations to predict and adjust optimal PAPR targets with high accuracy. The complexity assessment presented indicates that PAPRer strikes a favorable balance between performance and complexity. Moreover, the article proposes an ML-based method specifically addressing Inter-Carrier Waveform (ICWEF). It highlights the capability of this method to accurately predict optimal PAPR targets. Moreover, the authors of [146] emphasized the LSTM-Autoencoder (LSTM-AE) model for its ability to handle variable input sequential data and its flexibility in finding a compromise between the PAPR and BER through hyperparameter tuning. This paper aims to tackle the PAPR issue in OFDM systems specifically within VLC. The proposed solution involves employing an LSTM-AE, a combination of an autoencoder and Long Short Term Memory (LSTM), to efficiently learn a concise representation of the input and accommodate variable-length sequences. The proposed LSTM-AE model presented superior performance in terms of the PAPR compared with various existing strategies, all while maintaining a balance with the BER. The conclusion asserts the model's superiority in simulations over traditional schemes. Nevertheless, the text acknowledges certain limitations, underscoring the importance of further research, particularly in the development of faster training methods for larger OFDM systems with increased subcarriers and modulation orders.

4.2.2. DL [147–154]

Another area of ML largely considered to solve the PAPR optimization problem is DL [147–154]. As for the ANN-based approaches, the DL-based solutions usually consider the same basic operation [147–154]. Therefore, the basic blocks of the main DL technique considered in the literature (i.e., DNN) is illustrated in Figure 14, where it is possible to verify that, different from the ANN, a DNN based on an autoencoder (encoder/decoder) is composed of the cascaded sublayer which contains FC dense layers, normalization algorithms, and activation functions. These sublayers are responsible for learning, through a supervised training process, an efficient abstract representation of the OFDM signal generated by the IFFT and generating a new OFDM signal based on superimposing the output signals. The generated OFDM signal should present a low PAPR.

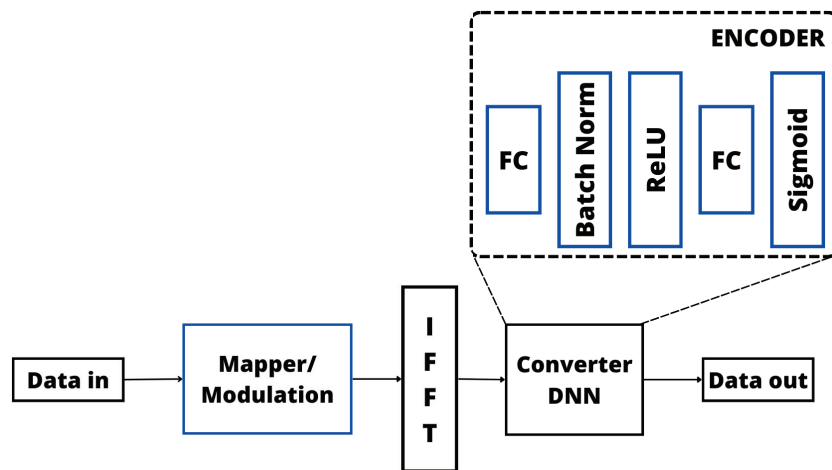


Figure 14. Illustration of the DNN transmitter.

Among some relevant works considering DL techniques to reduce the PAPR in OFDM systems [147–154]. In [147], the authors employed a DL paradigm featuring an autoencoder network that proved to be highly effective in addressing the nonlinearity induced by LED in VLC. The transmitter integrates a DNN employing Discrete Fourier Transform Spread (DFT-S), while the receiver is bifurcated into two subnets. The cost function, encompassing both autocorrelation and the mean square error, contributes to the scheme's superior performance in terms of the BER and training speed. This model-driven DL approach incorporates an autoencoder network with distinct encoder and decoder subnets for symbol mapping and detection. To tackle LED nonlinearity within VLC employing OFDM, an autoencoder-based NN, enriched with expert domain knowledge, was deployed. The architecture and loss function of the autoencoder network were meticulously designed to encapsulate expert knowledge, and an investigation into the impact of the regularization parameter and training SNR on the network was conducted. The proposed NN, infused with expert domain knowledge, adeptly mitigated LED nonlinearity, nonlinear distortions, and ISI while remaining compatible with CP removal. It efficiently addressed interference stemming from multipath channels, thereby enhancing overall performance. Moreover, in [148], the innovative Tone Reservation Network (TRNet) scheme, grounded in DL principles, was devised to effectively diminish the PAPR. TRNet exhibits adaptability to input signals, strategically generating signals that neutralize peaks. Through simulations, it was evident that the proposed TRNet outperformed the other strategies in terms of PAPR reduction, achieving notable results with a reduced number of reserved tones. The efficacy of TRNet in PAPR reduction within OFDM signals was convincingly demonstrated. Notably, TRNet contributed to accelerated training speeds while maintaining commendable BER compensation. This DL-based TRNet scheme showcased its superiority in PAPR reduction with fewer tones, ultimately enhancing bandwidth efficiency in OFDM systems. The utilization of an Feedforward Neural Network (FFNN) within TRNet further solidified its capability to generate peak-canceling signals, further emphasizing its effectiveness in addressing PAPR challenges.

A novel DL algorithm guided by a model was introduced in [149] to effectively mitigate the PAPR in OFDM systems. This algorithm incorporates an iterative peak-canceling signal generation scheme within a DL network. The trainable parameters within this scheme optimize both the clipping threshold and the weights of the time domain kernel function, resulting in comparable PAPR performance to existing approaches. Notably, this algorithm achieves these outcomes with a commendable reduction in complexity and training costs. Furthermore, the proposed tone reservation algorithm, grounded in a model-driven DL approach, is specifically designed for PAPR reduction. The fusion of a physical model with DL enhances its overall performance and provides a more insightful interpretation of the results. This innovative approach represents a significant stride in

addressing PAPR challenges, combining efficiency with a more comprehensive understanding of the underlying processes. Moreover, a groundbreaking adaptive modulation scheme attuned to the intricacies of the PAPR was innovatively introduced in [150] to enhance the energy efficiency of OFDM. The task of maximizing energy efficiency was aptly framed with meticulous considerations for power and modulation constraints. Confronted with the non-convex nature of this optimization challenge, a cutting-edge solution grounded in online DL was proposed. This pioneering protocol demonstrated a substantial increase in energy efficiency, achieving a remarkable gain of up to 3 dB. Notably, in comparison with conventional PAPR-unaware protocols, the PAPR-aware adaptive modulation scheme reduced the PAPR by an additional 3 dB. The methodology hinges on the application of online DL for optimization purposes.

In [151], the authors delved into an in-depth exploration of a DNN-based receiver tailored to LACO-OFDM systems. In this comprehensive analysis, various schemes were meticulously compared, scrutinizing their spectral efficiency and complexity. The study also addressed the determination of the minimum number of FC layers essential for optimal signal detection. Introducing a novel approach termed O-OFDMNet, the paper leveraged DL techniques for encoding and detection in LACO-OFDM, DCO-OFDM, and low-complexity LACO-OFDM systems. Impressively, O-OFDMNet achieved throughput levels akin to Radio Frequency (RF)-OFDM, maintaining a moderate level of complexity. Noteworthy is its superior resilience to practical optical channel conditions when benchmarked against existing counterparts. The DL-enhanced O-OFDMNet not only matches the spectral efficiency of classical RF-OFDM but also stands out as a power-efficient solution. It succeeds in reducing the BER and PAPR, concurrently enhancing the capacity. Furthermore, O-OFDMNet exhibits robustness against noise and distortion, showcasing a shortened runtime for added efficiency. Moreover, the authors of [152] proposed a DUN called the PR-DUN model, which was intricately crafted to address the challenge of the PAPR in OFDM systems, employing a technique known as deep unfolding. This model dynamically adjusts its parameters through backpropagation and minimizes a loss function intricately linked to the PAPR. Notably, the PR-DUN model accommodates any transmit power constraint, providing fine-tuned control over power increases. These distinctive features collectively empower the PR-DUN model to achieve a more substantial reduction in the PAPR alongside a diminished BER. It is worth highlighting the commendable computational efficiency of the PR-DUN model when compared with other similar solutions. The incorporation of trainable parameters and a stochastic optimization algorithm allows the model to optimize its variables for effective PAPR reduction. The utilization of the ReLU activation function within its layers adds to the model's adaptability and performance.

A method leveraging soft clipping was introduced in [153] for Tensor-train Residual Deep Neural Network (TT-RDNN), aiming to deterministically diminish the PAPR. The traditional Residual Deep Neural Network (RDNN) is replaced with a TT-RDNN, resulting in a reduction in the parameters and memory usage. The efficacy of the TT-RDNN was thoroughly assessed across diverse modulation alphabets. A pioneering DL architecture was proposed to strategically lower the PAPR of OFDM systems. This innovative architecture ensures a linear power amplifier response with a deterministic threshold while minimizing performance loss. The incorporation of tensor-train decomposition in the architecture not only reduces the parameters but also accelerates training and lowers memory usage. The outcome was a DL model that significantly lowered the PAPR of OFDM, adhering to spectral mask requirements and minimizing performance loss. In summation, a learning-driven strategy was harnessed to jointly formulate the transmit and receive filters in communication systems [154], with the primary goal of maximizing the achievable information rates while adeptly managing the Adjacent Channel Leakage Ratio (ACLR) and PAPR. In stark contrast to conventional techniques, this approach achieved a substantial reduction in both the ACLR and PAPR, paving the way for the transmission of competitive rates within a 3rd Generation Partnership Project (3GPP) multipath channel. Crucially, this learning-based method seamlessly integrates into the transmitter without introducing

any supplementary complexity, rendering it particularly advantageous for 6G networks. The joint design of transmit and receive filters is facilitated by an NN-based detector for waveform learning. The method not only optimizes waveform design for communication systems, ensuring competitive rates with diminished ACLRs and PAPRs, but it also enables the multiplexing of multiple users through joint optimization. Importantly, it achieves these enhancements without a significant loss in the information rate when compared with conventional baselines. Furthermore, a reduction in the ACLR and PAPR is achieved without adding complexity to the transmitter, underscoring the efficiency and practicality of this proposed approach.

4.2.3. Other ML Techniques [134,155]

In addition to previously considered ML techniques, a few works in the literature explored different ML techniques. For example, in [155], the authors presented a low-complexity method for reducing the PAPR in OFDM signals using MCTS. The MCTS strategy is used to create a tree structure that represents the searching space and searches for the optimal solution by sequentially finding a path from the root node to a leaf node. Figure 15 illustrates the MCTS strategy for PAPR reduction, from which it is possible to observe that the MCTS receives the OFDM signal generated by the IFFT or another classical PAPR reduction method, and after this, the MCTS is used to search for the optimal phase vector and generate a new OFDM signal with a low PAPR. More specifically, the authors outlined proposals for a low-complexity approach aimed at mitigating the PAPR of OFDM signals. The proposed method leverages the MCTS to enhance PAPR reduction, emphasizing a delicate equilibrium between minimizing the PAPR and efficiently managing computational load. In addition, to introducing the proposed low-complexity PTS method, the authors also considered the conventional reduced complexity PTS method and the iterative flipping PTS method. They further introduced the Monte Carlo PTS (M-PTS) method, which integrates the MCTS to reduce the PAPR in OFDM. In addition, this paper underscores the novel combination of the MCTS with dominant PTS methods as an effective strategy for complexity reduction. According to the proposal, this method significantly decreases computational complexity by a factor of 35, claiming a noteworthy 0.2 dB enhancement in the PAPR at half the computation cost. However, it is crucial to note that the proposed method does exhibit a PAPR performance degradation of 1.16 dB.

Finally, the authors of [134] introduced a novel ML technique for the purpose of diminishing the PAPR in massive MIMO OFDM systems. This approach optimizes the PAPR by considering a model fitting method named RANSAC. This optimization process is conducted by generating multiple candidate approximations obtained by randomly subsampling the grid points. Then, RANSAC is considered to detect the number of outliers (larger distance of the considered optimization function) in each generated candidate. Therefore, the determined best model is the one that minimizes the number of outliers. In addition, PAPR reduction is also accomplished through diverse methods, including clipping and filtering, ACE, TR, Selective Tone Reservation (STR), and Unused Beam Reservation (UBR), as summarized in Figure 16. Additionally, the DNN and Self-Organized Map (SOM) techniques are integrated into the process. This ML-based approach proves effective in significantly reducing the PAPR within a massive MIMO system employing OFDM. To alleviate the laborious task of manually tuning the hyperparameters for various scenarios, the introduction of an optimal hyperparameter function computation emerges as a practical solution, streamlining the optimization process.

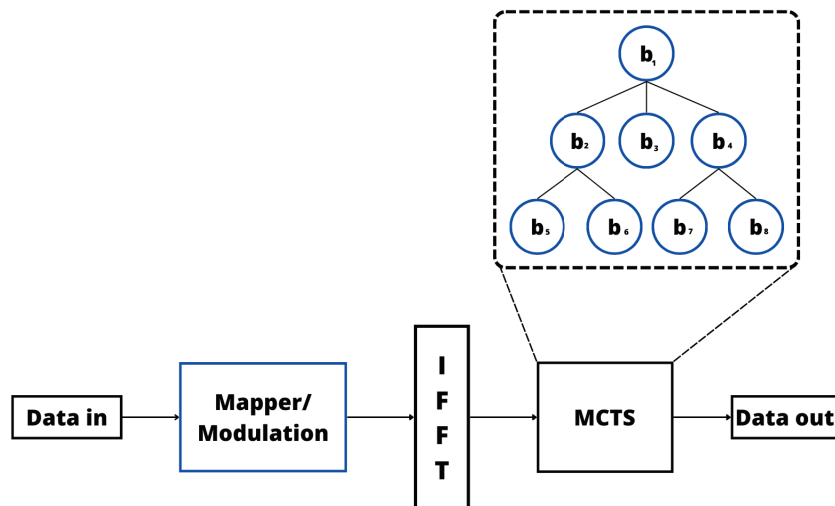


Figure 15. Illustration of the MCTS transmitter.

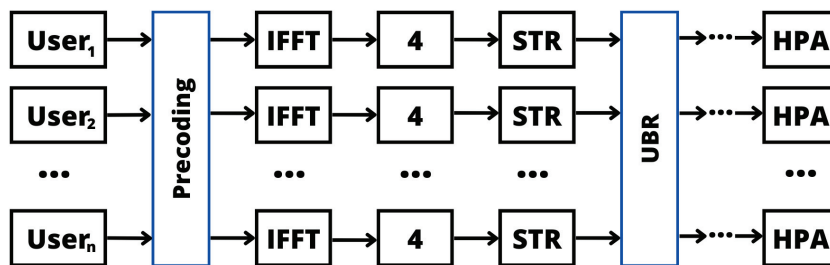


Figure 16. Illustration of the RANSAC transmitter [134].

Finally, from the previous detailed discussion about the ML-based PAPR reduction approaches, we can say that the spectral efficiency, computational complexity, BER, and PAPR are crucial elements, each presenting specific challenges. Finding the ideal ML approach demands a balance between these factors while considering their interactions. Each technique has different characteristics and performance levels in several aspects. Therefore, choosing the most appropriate technique depends on the specifics of the evaluated scenario and its requirements. In some situations, the complexity of the process increases, but the performance in terms of PAPR reduction stands out, ultimately paying off. On the other hand, in different contexts, an increase in complexity cannot be tolerated, or even some latency constraints influence the decision.

The applicability of each technique depends on the specific needs of each implementation. It is essential to consider the trade-offs between the spectral efficiency, computational complexity, BER, and PAPR, seeking to find the ideal balance to meet the objectives of the communication system in question. Moreover, we can also verify that each work addresses a distinct system model, incorporating specific modulation schemes and varied OFDM parameters. Due to these differences, a direct comparison between the amount of PAPR reduction that each one attains would be unfair and even potentially misleading, as each approach has its own particularities, including the considered level of BER degradation.

Therefore, Table 2 summarizes the performance of all reviewed ML-based methods in terms of PAPR reduction when compared with the classical approaches presented in Section 3. From Table 2, it is possible to verify the superiority of the ML-based approaches when compared with classical PAPR methods, showing the importance of studying new ML techniques to solve the emerging PAPR reduction problem. To finish, this table serves as a valuable resource for researchers and practitioners looking to explore and compare different ML-based PAPR reduction methods. However, each technique has different characteristics and performance levels in several aspects. Therefore, choosing the most appropriate technique depends on the specifics of the evaluated scenario and requirements.

In some situations, the complexity of the process increases, but the performance in terms of PAPR reduction stands out, ultimately paying off. On the other hand, in different contexts, an increase in complexity cannot be tolerated, or even some latency constraints influence the decision. Therefore, it is quite hard (if possible) to answer which one is preferable. It is crucial to carefully evaluate the individual needs of each situation before deciding which technique to adopt. Finally, Table 3 summarizes in chronological order the PAPR solutions based on ML cited previously, providing a comprehensive understanding of their proposed methods and goals. In addition, Table 4 describes the main contributions and limitations of the previously described works.

Table 2. Reduction in the PAPR achieved by the ML-based techniques when compared with classical PAPR methods.

Reference	PTS	SLM	Clipping and Filtering	TR	ACE
[141]	4 dB	-	7 dB	-	-
[143]	-	9 dB	6.5 dB	-	-
[147]	3 dB	-	-	-	-
[142]	-	-	-	-	5 dB
[148]	-	-	4.5 dB	7 dB	-
[145]	-	-	3 dB	-	-
[134]	-	-	4.5 dB	-	-
[149]	-	2 dB	-	-	-
[155]	2.5 dB	-	-	-	-
[146]	-	-	2.7 dB	-	-
[152]	-	3 dB	-	-	-
[153]	-	-	4 dB	-	-
[144]	-	7 dB	-	-	-

Table 3. Overview of the state of the art of PAPR reduction approaches based on ML (2019–2023).

Year	Reference	ML Technique	Goal
2019	[141]	ANN	Minimize the PAPR.
2019	[143]	ANN	Enhance indoor VLC-OFDM system performance.
2019	[147]	DNN	Enhance BER performance in VLC.
2019	[142]	ANN	Reduce the PAPR of OFDM symbols.
2020	[148]	DNN	Used to generate peak-canceling signal to reduce PAPR.
2020	[145]	ANN	Reduce PAPR in wireless transmitters.
2020	[134]	RANSAC	PAPR Reduction in Massive MIMO
2021	[149]	DNN	PAPR reduction in OFDM system.
2021	[151]	DNN	PAPR reduction using O-OFDMNet.
2021	[155]	MCTS	PAPR Reduction for OFDM signals.
2022	[146]	ANN	Enhance VLC system performance by reducing PAPR.
2022	[152]	DUN	Address PAPR challenges and power constraints.
2022	[135]	ANN	Frequency-selective PAPR reduction.
2022	[153]	DNN	Mitigate high PAPR in multicarrier transmissions.
2022	[154]	DNN	Enhance spectral efficiency and mitigate PAPR issues.
2023	[150]	DNN	Maximize energy efficiency and reduce PAPR.
2023	[144]	ANN	Reduce PAPR of GFDM.

Table 4. Contributions and Limitations of the PAPR reduction approaches based on ML (2019–2023).

Reference	Contributions	Limitations
[141]	PAPR is reduced to 6.5 dB compared with conventional methods.	Lack of complexity comparison, unclear if reducing PAPR maintains original BER, uncertainty about system performance impact.
[143]	A 12 dB PAPR reduction and improved BER performance.	Real-world implementations, variations in system conditions not fully captured in simulations.
[147]	Effective LED nonlinearity mitigation and improved BER performance.	The randomness and instability of network training indicates that there may be challenges in the neural network training phase, making it a challenge to obtain more reliable and consistent performance.
[142]	Achieve a lower PAPR and reduce the BER in a nonlinear channel model.	NN-ACE raises power consumption. Degradation in overall BER performance when NN-ACE is implemented.
[148]	Provide better PAPR reduction performance with fewer TRs.	Considerable computational effort and time in the training phase.
[145]	PAPR reduction and BER minimization through joint offline training.	Absence of exploration of trade-offs between PAPR reduction and other performance metrics.
[134]	PAPR reduction based on the classical ML approaches.	Time-consuming optimization of hyperparameters. RANSAC requires a high number of iterations.
[149]	Achieve comparable PAPR performance with low complexity and training costs.	Risk of overfitting due to excessive training data and numerous trainable parameters. Subpar performance is observed when the number of peak subcarriers exceeds a specific value.
[151]	O-OFDMNet optimizes BER and PAPR in soft decision channel decoder.	Heightened complexity in LACO-OFDM results in significant nonlinear distortion, primarily due to the high PAPR of the transmitted signal.
[155]	Reduced complexity, PTS and PAPR performance improvement.	As PAPR decreases, finding an optimal phase vector becomes harder. The proposed method presents a PAPR drop of 1.16 dB.
[146]	PAPR reduction method that effectively improves VLC performance while achieving a balanced compromise with the BER.	High time consumption in the training process.
[152]	Significant PAPR reduction, BER, and reduced computational load.	Performance enhancement reaches saturation point due to a constraint on the transmit power. More reserved subcarriers lead to an increase in power requirements, potentially generating additional side lobes.
[135]	PAPRer offers favorable performance complexity for frequency-selective PAPR reduction.	Lack of comparison of computational complexity.
[153]	PAPR reduction while upholding spectral efficiency and minimizing in-band distortion through the TT-RDNN technique.	Does not consider important trade-offs, such as spectral efficiency and potential nonlinear distortion in systems where the PAPR exceeds the IBO.
[154]	PAPR reductions compared with traditional methods.	Difficulty balancing trade-offs in information rate, ACLR, and PAPR. Challenges in multiplexing and better interpretation of learned filters.
[150]	A DL solution for a 3 dB higher energy efficiency and a 3 dB PAPR reduction compared with conventional methods.	Lack of complexity analysis, feasibility for implementation, and potential drawbacks.
[144]	The PAPR and BER of the GFDM system are minimized considerably.	Lack of computational complexity analysis.

4.3. Limitations of ML-Based PAPR Reduction Approaches

Previously, we presented a deep discussion about the main features of the PAPR reduction approaches based on ML in the literature. However, it is essential to discuss the main limitations of the existing solutions to identify possible open research topics. Therefore, we next present a discussion over the main limitations of each previously cited work categorized by the ML technique considered.

4.3.1. ANN [135,141–146]

The work in [141] has a couple of important missing points. First, it does not compare how complex the methods are. Second, to understand the BER better, it is crucial to compare it with the original signal and the methods used in the paper. Unfortunately, the paper does not do this. It is unclear if reducing the PAPR keeps the original BER or if the system loses performance in terms of the BER. This missing information makes it hard to fully grasp how reducing the PAPR affects the system performance. In [142], when analyzing the CCDF curves of the PAPR and the symbol power, it is possible to note that while the ACE method decreased the PAPR, it came at the expense of increased signal power. On the other hand, NN-ACE achieved a lower PAPR but at the cost of higher power consumption. Notably, when NN-ACE was implemented, there was a degradation in the overall BER performance. The simulation results in [143] suggest seamless integration into typical Visible Light Communication OFDM (VLC-OFDM) systems, offering favorable PAPR and BER performance even in the presence of ISI in the diffuse optical wireless channel. While acknowledging limitations in real-world implementations and variations in system conditions not fully captured in the simulations, the study provides valuable insights into addressing challenges in indoor VLC-OFDM systems. The work in [144] did not consider the computational complexity of the system, and this information is crucial for evaluating if reduction of the PAPR is applicable or not.

In addition, the authors of [145] mentioned that the proposed method can be used in MIMO scenarios, but they did not provide specific details about the challenges or limitations it might encounter in such situations. The impact on the actual performance of the system in practical MIMO set-ups was not discussed. While the paper notes an improvement in the BER with a reduced PAPR, it does not delve into the trade-off between reducing the PAPR and its potential impact on other important performance metrics. There could be instances where optimizing for one parameter adversely affects another. The paper also does not address how the proposed method reacts to changes in channel conditions or variations in transmission environments. In real-world applications, the robustness of the method to dynamic channel conditions is crucial, and this aspect was not covered. The work in [135] has two notable shortcomings. First, it does not provide a comparison of computational complexity, a crucial aspect that remains unexplored. Second, although the paper includes a table comparing BER and PAPR loss, there is a lack of a detailed explanation in the text. This omission makes it challenging for readers to comprehend whether the BER after PAPR reduction aligned with the original BER, a vital consideration in evaluating the effectiveness of PAPR reduction methods. The authors of [146] recognized certain limitations and emphasized the need for further research, especially in the area of developing faster training methods. This is particularly important for larger OFDM systems that have increased subcarriers and modulation orders. The acknowledgment highlights the awareness of existing constraints and directs attention toward addressing these challenges through advancements in training methodologies.

4.3.2. DL [147–154]

The suggested NN in [147], incorporating specialized domain knowledge, effectively minimizes issues related to LED nonlinearity, nonlinear distortions, and ISI. Additionally, it remains compatible with CP removal. The NN handles interference arising from multipath channels, contributing to an overall enhancement in performance. TRNet in [148] has a substantial training burden, indicating a considerable amount of computational effort and time needed during the training phase. Despite acknowledging this training burden, it is highlighted that the execution time of TRNet is significantly lower than that of the TR technique when utilizing the trained model. The comparison mainly concentrates on the execution time and time complexity rather than the training time. Although TRNet demonstrates efficiency during execution, the acknowledgment of the training burden underscores its practicality in scenarios involving the processing of a large number of symbols. The proposed algorithm

in [149] demonstrated better learning capabilities when compared with traditional algorithms, even though its PAPR performance fell short of PRnet. The observation that excessive training data and too many trainable parameters can result in redundancy implies a potential risk of overfitting. Notably, the algorithm demonstrated subpar performance when the number of peak subcarriers surpassed a specific value.

The limitation of the scheme in [150] is that it is described as PAPR-unaware, indicating it does not consider the PAPR in its optimization. This could be problematic in situations where the PAPR is crucial, and not factoring it in might lead to suboptimal performance. Furthermore, the paper offers limited details about the proposed PAPR-aware scheme, making it difficult to evaluate its complexity, feasibility for implementation, and potential drawbacks. This lack of information hinders a comprehensive understanding of the proposed scheme's characteristics and its effectiveness in addressing PAPR-related challenges. When employing advanced schemes like LACO-OFDM in [151], the transceiver system experienced heightened complexity. This complexity escalation can result in significant nonlinear distortion of the signal, primarily due to the high PAPR of the transmitted signal. Improving PAPR reduction in [152] was achieved by increasing the number of reserved subcarriers. Nevertheless, this performance enhancement reached a saturation point due to a constraint on the transmit power. As more subcarriers were reserved, there was an associated increase in power requirements, potentially leading to the generation of additional side lobes. In [153], the authors suggested that the proposed TT-RDNN method offers advantages in terms of a deterministic PAPR, performance compared to certain existing methods, and the consideration of spectral mask requirements. However, it is important to consider trade-offs, such as spectral efficiency and potential nonlinear distortion in systems where the PAPR exceeds the IBO. The results in [154] indicate that the proposed method allows fine control over the trade-off between the information rate, ACLR, and PAPR. On the Additive white Gaussian noise (AWGN) channel, the learned waveforms achieved competitive rates with conventional waveforms but exhibited significantly lower ACLRs. Lower PAPRs were also possible, though at the cost of a rate loss. The method appears to offer advantages in terms of fine-tuning trade-offs in communication system design, but there are challenges and areas for further exploration, especially for understanding the learned filters and extending the method to different scenarios and technologies.

4.3.3. Other ML Techniques [134,155]

The limitation in [155] is that as the PAPR decreased, it became more challenging to find a better phase vector for achieving a lower PAPR. Consequently, the marginal contribution of the complexity parameter diminished, leading to flattened curves when the parameter was large. However, it is crucial to note that the proposed method did exhibit PAPR performance degradation of 1.16 dB. Finally, in [134], the authors presented a manual optimization of the hyperparameters for each particular situation, a laborious and time-consuming process. In the case of Ridge models, an analytical solution was employed, while in the case of RANSAC, the number of iterations reached a magnitude of 10^5 .

4.4. Overall ML Technique Limitations

Minimizing the PAPR in wireless communication systems presents a significant challenge, and ML-based approaches have been investigated to address this issue. However, it is crucial to recognize the general limitations associated with these techniques. We highlight below some of these limitations regarding the methods specifically mentioned in this paper:

- **Computational Complexity:** ML methods often consider complex models with numerous parameters. The requirement of substantial computational resources for training and optimizing these models may limit their applicability in hardware-constrained environments [35].
- **Interpretability:** Lack of interpretability is a known characteristic of NNs, including ANNs and DNNs. Understanding how these models make specific decisions

can be challenging, making it difficult to justify the choices made during the PAPR minimization process [162].

- **Sensitivity to Training Data:** ML methods such as the ANN and DNN are sensitive to the quality and quantity of the training data. If data are not representative or comprehensive enough, then models may fail to generalize to new scenarios, resulting in suboptimal performance [39].
- **Limitations of Generalization:** Techniques like MCTS may face difficulties in generalizing to situations outside the training set. Variations in communication channel conditions can compromise the ability of these methods to adapt effectively, leading to less efficient results [163].
- **Extensive Training Requirements:** Models such as the DUN may require extensive training to achieve satisfactory performance, which may be impractical in scenarios with limited availability of labeled data or when training needs to occur in real time [164].
- **Theoretical Limitations:** Certain methods, such as RANSAC, may be based on specific theoretical assumptions that may not be fully valid in all practical situations. Variations in channel conditions can compromise the effectiveness of these methods [134].
- **Trade-off between Accuracy and Computational Time:** PAPR optimization often involves a trade-off between accuracy and processing time. More accurate methods, such as the MCTS and DUN, may require more computational time and are impractical in real-time applications [165].

5. Future Directions

In the foreseeable future, addressing numerous challenges within wireless communication systems is paramount, particularly in the context of ML algorithms. Herein, we discuss the challenges of ML in the context of the PAPR.

5.1. ML Training

The effectiveness of ML models in optimizing the PAPR hinges on high-quality training data. Acquiring and managing such data poses challenges, leading to the need for protocols in data collection and annotation. To address data scarcity, generative models and data augmentation techniques are essential [45,166].

5.2. Generalization across Diverse Scenarios

Generalizing ML models across varied channel conditions, signal types, and evolving wireless standards is a significant challenge. Resilient models that adapt to diverse scenarios require a multi-faceted strategy, including meta-learning, transfer learning, domain adaptation, and the infusion of domain-specific knowledge [45,166].

5.3. Resource Constraints in IoT Devices

Implementing ML techniques in resource-constrained devices, such as those in the IoT, introduces challenges. Techniques like model compression, quantization, and hardware acceleration are crucial to balancing computational demands with device constraints [167–169].

5.4. Real-Time Implementation

Real-time implementation of ML-based PAPR optimization, especially in upcoming wireless standards like 6G, remains a challenge. Low-latency operation demands optimized algorithms and efficient inference techniques, such as model quantization, model parallelism, and hardware accelerators [170–172].

5.5. Computational Complexity

The computational complexity associated with ML techniques, particularly DL models, presents a central challenge. Specialized hardware and streamlined software frameworks

are required to expedite the execution of resource-intensive algorithms, avoiding bottlenecks in real-time wireless communication [173,174].

5.6. Adaptability to Dynamic Conditions

Ensuring the adaptability of ML models to rapidly changing wireless environments is an ongoing concern. Techniques like meta-learning, continuous learning, and Reinforcement Learning (RL) are under exploration to make ML models more agile and responsive to dynamic network conditions [175,176].

5.7. Energy Efficiency

Energy efficiency is critical for ML-enabled wireless systems, particularly in battery-powered devices. Optimizing energy consumption during training and inference involves techniques such as model sparsity, energy-efficient hardware, and dynamic power management [177–179].

5.8. Standardization for Seamless Integration

The seamless integration of ML into PAPR optimization solutions relies on standardized interfaces. Open standards and interoperability protocols are essential for the adoption of ML techniques in the wireless industry, encompassing communication, data formats, model interchangeability, and security protocols [180,181].

5.9. Robust Wireless Communication Infrastructure

The establishment of a robust wireless communication infrastructure has necessitated addressing prevalent issues such as bit errors, heightened power consumption, and signal distortion. One solution involves deploying sophisticated error correction codes, employing modulation techniques, and implementing innovative signal processing methods to enhance the resilience of communication links against data corruption.

5.10. Power Management

Efficient power management is a critical aspect, and it is achieved through the utilization of intelligent algorithms that meticulously balance signal strength and power consumption. Simultaneously, the integration of advanced filters and shaping techniques plays a pivotal role in reducing distortion, ensuring clarity of communication.

6. Conclusions

This paper has provided a comprehensive overview of the techniques used in the literature for reducing the PAPR in OFDM systems. We discussed various traditional approaches, novel schemes, and emerging trends in the field. The literature review revealed that conventional techniques, such as PTS, SLM, and clipping, among other techniques, have been widely studied and utilized for PAPR reduction in OFDM. These methods have shown significant improvements in reducing the PAPR. However, they often come with certain drawbacks, such as increased complexity, BER, and distortion. One of the notable advancements in recent years is the integration of ML techniques for PAPR reduction. ML-based approaches have shown great potential in addressing the challenges associated with the PAPR in OFDM systems. These methods leverage the power of data-driven learning and optimization to adaptively and intelligently reduce the PAPR while preserving system performance.

Author Contributions: Conceptualization, B.S.d.C.d.S., V.D.P.S., R.D.S. and L.L.M.; methodology, B.S.d.C.d.S., V.D.P.S., R.D.S. and L.L.M.; investigation, B.S.d.C.d.S., V.D.P.S., R.D.S. and L.L.M.; resources, B.S.d.C.d.S., V.D.P.S., R.D.S. and L.L.M.; writing—original draft preparation, B.S.d.C.d.S., V.D.P.S., R.D.S. and L.L.M.; writing—review and editing, V.D.P.S. and R.D.S.; Visualization, B.S.d.C.d.S., V.D.P.S., R.D.S. and L.L.M.; supervision, V.D.P.S., R.D.S. and L.L.M.; funding acquisition, L.L.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work received partial funding from project XGM-AFCCT-2024-2-15-1, supported by the xGMobile–EMBRAPII–Inatel Competence Center on 5G and 6G Networks, with financial resources from the PPI IoT/Manufacturing 4.0 program of MCTI (grant number 052/2023) signed with EMBRAPII. Additionally, this work was partially supported by the Ciência por Elas project (APQ-04523-23, funded by Fapemig), the Internacionalização das ICTMGS project (APQ-05305-23, funded by Fapemig), the SEMEAR project (22/09319-9, funded by FAPESP), the Brasil 6G project (01245.010604/2020-14, funded by RNP and MCTI), CNPq-Brasil, and the National Council for Scientific and Technological Development (CNPq) (Grant Nos. 402378/2021-0 and 305021/2021-4).

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

Acronyms and Symbols

1G	First generation
2G	Second generation
2.5G	2.5 generation
3G	Third generation
3GPP	3rd Generation Partnership Project
4G	Fourth generation
5G	Fifth generation
6G	Sixth generation
ACE	Active constellation extension
ACLR	Adjacent channel leakage ratio
AMPS	Advanced Mobile Phone System
ANN	Artificial neural network
AWGN	Additive white Gaussian noise
BatchNorm	Batch normalization
BER	Bit error rate
CCDF	Complementary cumulative distribution function
CP	Cyclic prefix
CL	Clipping level
DL	Deep learning
DCO-OFDM	DC-biased optical OFDM
DFT-S	Discrete Fourier transform spread
DNN	Deep neural network
DUN	Deep unfolding network
FC	Fully connected
FDM	Frequency division multiplexing
FDMA	Frequency division multiple access
FFNN	Feedforward neural network
FFT	Fast Fourier transform
GFDM	Generalized frequency division multiplexing
GSM	Global System for Mobile Communications
IBO	In-band outage
ICI	Inter-carrier interference
ICWEF	Inter-carrier waveform
IDFT	Inverse discrete Fourier transform
IFFT	Inverse fast Fourier transform
IoT	Internet of Things
IP	Internet Protocol
ISI	Inter-symbol interference
LACO-OFDM	Layered asymmetrically clipped OFDM

LED	Light-emitting diode
LSTM	Long short-term memory
LSTM-AE	LSTM autoencoder
MCTS	Monte Carlo tree search
MIMO	Multiple-input multiple-output
ML	Machine learning
mmWave	Millimeter wave
M-PTS	Monte Carlo PTS
NN	Neural network
NN-ACE	Neural network ACE
NMT	Nordic Mobile Telephone
OFDM	Orthogonal frequency division multiplexing
O-OFDM	Optical OFDM
PTS	Partial transmit sequence
PAPR	Peak-to-average-power ratio
PRnet	PAPR-reducing network
QAM	Quadrature amplitude modulation
QPSK	Quadrature phase shift keying
RANSAC	Random sample consensus
RDNN	Residual deep neural network
ReLU	Rectified linear unit
RF	Radio frequency
RL	Reinforcement learning
SLM	Selected mapping
SNR	Signal-to-noise ratio
SOM	Self-organized map
STR	Selective tone reservation
TACS	Total Access Communication System
TI	Tone injection
TR	Tone reservation
Tx	Transmit
TT-RDNN	Tensor-train residual deep neural network
TRNet	Tone reservation network
TV	Television
VLC	Visible light communication
VLC-OFDM	Visible light communication OFDM

References

1. Alsabah, M.; Naser, M.A.; Mahmmoud, B.M.; Abdulhussain, S.H.; Eissa, M.R.; Al-Baidhani, A.; Noordin, N.K.; Sait, S.M.; Al-Utaibi, K.A.; Hashim, F. 6G wireless communications networks: A comprehensive survey. *IEEE Access* **2021**, *9*, 148191–148243. [CrossRef]
2. Stemick, M.; Rohling, H. OFDM-FDMA scheme for the uplink of a mobile communication system. *Wirel. Pers. Commun.* **2007**, *40*, 157–170. [CrossRef]
3. Lyytinen, K.; Fomin, V.V. Achieving high momentum in the evolution of wireless infrastructures: The battle over the 1G solutions. *Telecommun. Policy* **2002**, *26*, 149–170. [CrossRef]
4. Kabalci, Y. 5G mobile communication systems: Fundamentals, challenges, and key technologies. In *Smart Grids and Their Communication Systems*; Springer: Singapore, 2018; pp. 329–359.
5. Lehenkari, J.; Miettinen, R. Standardisation in the construction of a large technological system—The case of the Nordic mobile telephone system. *Telecommun. Policy* **2002**, *26*, 109–127. [CrossRef]
6. Funk, J.L. Competition between regional standards and the success and failure of firms in the world-wide mobile communication market. *Telecommun. Policy* **1998**, *22*, 419–441. [CrossRef]
7. Pandya, R. Emerging mobile and personal communication systems. *IEEE Commun. Mag.* **1995**, *33*, 44–52. [CrossRef]
8. Forde, T.K.; Macaluso, I.; Doyle, L.E. Exclusive sharing and virtualization of the cellular network. In Proceedings of the 2011 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN), Aachen, Germany, 3–6 May 2011; pp. 337–348. [CrossRef]

9. Finol, J.; Mielke, J. Past and future directions in cellular telephony. In Proceedings of the 1998 Second IEEE International Caracas Conference on Devices, Circuits and Systems. ICCDCS 98. On the 70th Anniversary of the MOSFET and 50th of the BJT. (Cat. No.98TH8350), Isla de Margarita, Venezuela, 4 March 1998; pp. 7–13. [CrossRef]
10. Holley, K. Global System for Mobile Communications—What’s in Store? In *Mobile Systems*; Springer: Boston, MA, US, 1997; pp. 67–80.
11. Stoumpos, C.K.; Chryssomallis, M.T. Improvement of Channel Models Performance with Parameters Values from Measurements: A Test Case for Suburban and Rural Environments at 450 MHz. *IEEE Antennas Wirel. Propag. Lett.* **2018**, *17*, 1396–1400. [CrossRef]
12. Mavoungou, S.; Kaddoum, G.; Taha, M.; Matar, G. Survey on threats and attacks on mobile networks. *IEEE Access* **2016**, *4*, 4543–4572. [CrossRef]
13. Neonakis Aggelou, G.; Tafazolli, R. On the relaying capability of next-generation GSM cellular networks. *IEEE Pers. Commun.* **2001**, *8*, 40–47. [CrossRef]
14. Xiuhua, Q.; Chuanhui, C.; Li, W. A study of some key technologies of 4G system. In Proceedings of the 2008 3rd IEEE Conference on Industrial Electronics and Applications, Singapore, 3–5 June 2008; pp. 2292–2295.
15. Toorani, M.; Beheshti, A. Solutions to the GSM security weaknesses. In Proceedings of the 2008 The Second International Conference on Next Generation Mobile Applications, Services, and Technologies, Cardiff, UK, 16–19 September 2008; pp. 576–581.
16. Rahnema, M. Overview of the GSM system and protocol architecture. *IEEE Commun. Mag.* **1993**, *31*, 92–100. [CrossRef]
17. Chen, R.; Long, W.X.; Mao, G.; Li, C. Development trends of mobile communication systems for railways. *IEEE Commun. Surv. Tutor.* **2018**, *20*, 3131–3141. [CrossRef]
18. Drane, C.; Macnaughtan, M.; Scott, C. Positioning GSM telephones. *IEEE Commun. Mag.* **1998**, *36*, 46–54. [CrossRef]
19. Ezhilarasan, E.; Dinakaran, M. A review on mobile technologies: 3G, 4G and 5G. In Proceedings of the 2017 Second International Conference on Recent Trends and Challenges in Computational Models (ICRTCCM), Tindivanam, India, 3–4 February 2017; pp. 369–373.
20. Ekstrom, H.; Furuskar, A.; Karlsson, J.; Meyer, M.; Parkvall, S.; Torsner, J.; Wahlqvist, M. Technical solutions for the 3G long-term evolution. *IEEE Commun. Mag.* **2006**, *44*, 38–45. [CrossRef]
21. Honkasalo, H.; Pehkonen, K.; Niemi, M.; Leino, A. WCDMA and WLAN for 3G and beyond. *IEEE Wirel. Commun.* **2002**, *9*, 14–18. [CrossRef]
22. Jamalipour, A.; Wada, T.; Yamazato, T. A tutorial on multiple access technologies for beyond 3G mobile networks. *IEEE Commun. Mag.* **2005**, *43*, 110–117. [CrossRef]
23. Hui, S.Y.; Yeung, K.H. Challenges in the migration to 4G mobile systems. *IEEE Commun. Mag.* **2003**, *41*, 54–59. [CrossRef]
24. Varshney, U. 4G Wireless Networks. *IT Prof.* **2012**, *14*, 34–39. [CrossRef]
25. Cavalcanti, F.R.P. *Resource Allocation and MIMO for 4G and Beyond*; Springer: New York, NY, US, 2014.
26. Zhang, H.; Wang, Z.; Yu, J.; Huang, J. A compact MIMO antenna for wireless communication. *IEEE Antennas Propag. Mag.* **2008**, *50*, 104–107. [CrossRef]
27. Domathoti, B.; Ch, C.; Madala, S.R.; Berhanu, A.A.; Rao, Y.N. Simulation analysis of 4G/5G OFDM systems by optimal wavelets with BPSK modulator. *J. Sens.* **2022**, *2022*, 8070428. [CrossRef]
28. Sun, S.; Ju, Y.; Yamao, Y. Overlay cognitive radio OFDM system for 4G cellular networks. *IEEE Wirel. Commun.* **2013**, *20*, 68–73. [CrossRef]
29. Giordani, M.; Mezzavilla, M.; Zorzi, M. Initial access in 5G mmWave cellular networks. *IEEE Commun. Mag.* **2016**, *54*, 40–47. [CrossRef]
30. Zhao, J.; Liu, J.; Yang, L.; Ai, B.; Ni, S. Future 5G-oriented system for urban rail transit: Opportunities and challenges. *China Commun.* **2021**, *18*, 1–12. [CrossRef]
31. Gupta, A.; Jha, R.K. A survey of 5G network: Architecture and emerging technologies. *IEEE Access* **2015**, *3*, 1206–1232. [CrossRef]
32. Shariatmadari, H.; Ratasuk, R.; Irabi, S.; Laya, A.; Taleb, T.; Jäntti, R.; Ghosh, A. Machine-type communications: Current status and future perspectives toward 5G systems. *IEEE Commun. Mag.* **2015**, *53*, 10–17. [CrossRef]
33. Merluzzi, M.; Borsos, T.; Rajatheva, N.; Benczúr, A.A.; Farhadi, H.; Yassine, T.; Mueck, M.D.; Barmounakis, S.; Strinati, E.C.; Dampahalage, D.; et al. The Hexa-X Project Vision on Artificial Intelligence and Machine Learning-Driven Communication and Computation Co-Design for 6G. *IEEE Access* **2023**, *11*, 65620–65648. [CrossRef]
34. Long, Q.; Chen, Y.; Zhang, H.; Lei, X. Software defined 5G and 6G networks: A survey. *Mob. Networks Appl.* **2019**, *27*, 1792–1812. [CrossRef]
35. Ali, S.; Saad, W.; Steinbach, D. White Paper on Machine Learning in 6G Wireless Communication Networks. 6G Research Visions, No. 7. University of Oulu. 2020. Available online: <http://urn.fi/urn:isbn:9789526226736> (accessed on 1 February 2024).
36. Noman, H.M.F.; Hanafi, E.; Noordin, K.A.; Dimyati, K.; Hindia, M.N.; Abdrabou, A.; Qamar, F. Machine Learning Empowered Emerging Wireless Networks in 6G: Recent Advancements, Challenges and Future Trends. *IEEE Access* **2023**, *11*, 83017–83051. [CrossRef]
37. Taleb, T.; Aguiar, R.L.; Yahia, I.G.B.; Christensen, G.; Chunduri, U.; Clemm, A.; Costa, X.; Dong, L.; Elmirghani, J.; Yusuf, B.; et al. White Paper on 6G Networking. 2020. Available online: <https://www.6gflagship.com/white-paper-on-6g-networking/> (accessed on 1 February 2024).
38. Katz, M.; Ahmed, I. Opportunities and Challenges for Visible Light Communications in 6G. In Proceedings of the 2020 2nd 6G Wireless Summit (6G SUMMIT), Levi, Finland, 17–20 March 2020; pp. 1–5. [CrossRef]

39. Ylianttila, M.; Kantola, R.; Gurtov, A.; Mucchi, L.; Oppermann, I. 6G White Paper: Research Challenges For Trust, Security And Privacy. 6G Research Visions, No. 9. 2020. Available online: <http://urn.fi/urn:isbn:9789526226804> (accessed on 15 January 2024).
40. Mahmood, N.H.; López, O.; Park, O.S.; Moerman, I.; Mikhaylov, K.; Mercier, E.; Munari, A.; Clazzer, F.; Böcker, S.; Bartz, H. White Paper on Critical and Massive Machine Type Communication Towards 6G. 6G Research Visions, No. 11. University of Oulu. 2020. Available online: <http://urn.fi/urn:isbn:9789526226781> (accessed on 10 January 2024).
41. Uusitalo, M.A.; Ericson, M.; Richerzhagen, B.; Soykan, E.U.; Rugeland, P.; Fettweis, G.; Sabella, D.; Wikström, G.; Boldi, M.; Hamon, M.H.; et al. Hexa-X The European 6G flagship project. In Proceedings of the 2021 Joint European Conference on Networks and Communications and 6G Summit (EuCNC/6G Summit), Porto, Portugal, 8–11 June 2021; pp. 580–585. [CrossRef]
42. Chowdhury, M.Z.; Shahjalal, M.; Hasan, M.K.; Jang, Y.M. The role of optical wireless communication technologies in 5G/6G and IoT solutions: Prospects, directions, and challenges. *Appl. Sci.* **2019**, *9*, 4367. [CrossRef]
43. Saarela, H.; Wittenberg, V. 6G Waves Magazine: Spring 2021. 2021. Available online: <https://www.6gflagship.com/6g-waves-magazine/> (accessed on 1 February 2024).
44. Ozdemir, M.; Arslan, H. Channel estimation for wireless OFDM systems. *IEEE Commun. Surv. Tutor.* **2007**, *9*, 18–48. [CrossRef]
45. Rajatheva, N.; Atzeni, I.; Björnson, E.; Bourdoux, A.; Buzzi, S.; Doré, J.B.; Erkucuk, S.; Fuentes, M.; Guan, K.; Hu, Y.; et al. White Paper on Broadband Connectivity in 6G. 6G Research Visions, No. 10. University of Oulu. 2020. Available online: <http://urn.fi/urn:isbn:9789526226798> (accessed on 10 January 2024).
46. Wang, X.; Tjhung, T.T.; Ng, C.S. Reduction of peak-to-average power ratio of OFDM system using a companding technique. *IEEE Trans. Broadcast.* **1999**, *45*, 303–307. [CrossRef]
47. Jiang, T.; Zhu, G. Complement block coding for reduction in peak-to-average power ratio of OFDM signals. *IEEE Commun. Mag.* **2005**, *43*, S17–S22. [CrossRef]
48. Popovski, P.; Chiariotti, F.; Huang, K.; Kalør, A.E.; Kountouris, M.; Pappas, N.; Soret, B. A Perspective on Time Toward Wireless 6G. *Proc. IEEE* **2022**, *110*, 1116–1146. [CrossRef]
49. Pärssinen, A.; Alouini, M.S.; Berg, M.; Kürner, T.; Kyösti, P.; Leinonen, M.E.; Matinmikko-Blue, M.; McCune, E.; Pfeiffer, U.; Wambacq, P. White Paper on RF Enabling 6G: Opportunities and Challenges from Technology to Spectrum. 6G Research Visions. 2021. Available online: <https://www.6gflagship.com/white-paper-on-rf-enabling-6g-opportunities-and-challenges-from-technology-to-spectrum/> (accessed on 1 February 2024).
50. Rahmatallah, Y.; Mohan, S. Peak-to-average power ratio reduction in OFDM systems: A survey and taxonomy. *IEEE Commun. Surv. Tutor.* **2013**, *15*, 1567–1592. [CrossRef]
51. Arunjeeva, L.; Arunmozhi, S. A novel complexity PAPR reduction scheme for MIMO-OFDM systems. In Proceedings of the IEEE-International Conference on Advances in Engineering, Science and Management (ICAESM -2012), Nagapattinam, India, 30–31 March 2012; pp. 88–93.
52. Song, W.; Du, H.; Xue, J. A Novel Iterative Receiver for Clipping Distortion Recovery in OFDM Systems. In Proceedings of the 2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring), Florence, Italy, 20–23 June 2023; pp. 1–6.
53. Wang, Y.; Wang, L.-h.; Ge, J.-h.; Ai, B. Nonlinear companding transform technique for reducing PAPR of OFDM signals. *IEEE Trans. Consum. Electron.* **2012**, *58*, 752–757. [CrossRef]
54. Ibraheem, Z.T.; Ahmed, K.K.; Fazea, Y.; Madi, M.; Mohammed, F.; Ali, A.Q. Boosted PTS Method with Mu-Law Companding Techniques for PAPR Reduction in OFDM Systems. *Wirel. Pers. Commun.* **2022**, *124*, 423–436. [CrossRef]
55. Yang, L.; Chen, R.; Siu, Y.; Soo, K. PAPR reduction of an OFDM signal by use of PTS with low computational complexity. *IEEE Trans. Broadcast.* **2006**, *52*, 83–86. [CrossRef]
56. Jeemon, B.K.; Shahana, T. PAPR reduction using constrained convex optimization for vector OFDM. *Ann. Telecommun.* **2023**, *78*, 91–99. [CrossRef]
57. Prasad, S.; Jayabalan, R. PAPR reduction in OFDM systems using modified SLM with different phase sequences. *Wirel. Pers. Commun.* **2020**, *110*, 913–929. [CrossRef]
58. Arulkarthick, V.; Srihari, K.; Arvind, C.; Mukunthan, M.; Sundar, R. A Hybrid Precoded-Companding Scheme for PAPR Reduction in OFDM systems. *Natl. Acad. Sci. Lett.* **2023**, *1*–7. [CrossRef]
59. Yang, Z.; Fang, H.; Pan, C. ACE with frame interleaving scheme to reduce peak-to-average power ratio in OFDM systems. *IEEE Trans. Broadcast.* **2005**, *51*, 571–575. [CrossRef]
60. Qasem, Z.A.; Ali, A.; Deng, B.; Li, Q.; Fu, H. Spectral and Energy Efficient Pilot-Assisted PAPR Reduction Technique for Underwater Wireless Optical Communication Systems. *J. Light. Technol.* **2023**, *42*, 841–852. [CrossRef]
61. Tu, Y.P.; Chang, C.C. A Novel Low Complexity Two-Stage Tone Reservation Scheme for PAPR Reduction in OFDM Systems. *Sensors* **2023**, *23*, 950. [CrossRef]
62. Rajkumarsingh, B.; Rajarai, R.R.; Hosany, M. Performance Analysis of Hybrid Filtering Technique for Reduction OF PAPR in Alamouti Coded MIMO-OFDM Systems. *Wirel. Pers. Commun.* **2023**, *130*, 1639–1660. [CrossRef]
63. Kanaras, I.; Chorti, A.; Rodrigues, M.R.D.; Darwazeh, I. Spectrally Efficient FDM Signals: Bandwidth Gain at the Expense of Receiver Complexity. In Proceedings of the 2009 IEEE International Conference on Communications, Dresden, Germany, 14–18 June 2009; pp. 1–6. [CrossRef]
64. Weinstein, S.B. The history of orthogonal frequency-division multiplexing [History of Communications]. *IEEE Commun. Mag.* **2009**, *47*, 26–35. [CrossRef]

65. LaSorte, N.; Barnes, W.J.; Refai, H.H. The history of orthogonal frequency division multiplexing. In Proceedings of the IEEE GLOBECOM 2008-2008 IEEE Global Telecommunications Conference, New Orleans, LA, USA, 30 November–4 December 2008; pp. 1–5.
66. Isam, S.; Darwazeh, I. Precoded spectrally efficient FDM system. In Proceedings of the 21st Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, Istanbul, Turkey, 26–30 September 2010; pp. 99–104.
67. Arrano, H.F.; Azurdia-Meza, C.A. OFDM: Today and in the future of next generation wireless communications. In Proceedings of the 2016 IEEE Central America and Panama Student Conference (CONESCAPAN), Guatemala City, Guatemala, 27–30 September 2016; pp. 1–6.
68. Hsieh, M.H.; Wei, C.H. Channel estimation for OFDM systems based on comb-type pilot arrangement in frequency selective fading channels. *IEEE Trans. Consum. Electron.* **1998**, *44*, 217–225. [CrossRef]
69. Park, B.; Cheon, H.; Kang, C.; Hong, D. A novel timing estimation method for OFDM systems. *IEEE Commun. Lett.* **2003**, *7*, 239–241. [CrossRef]
70. Katselis, D.; Kofidis, E.; Rontogiannis, A.; Theodoridis, S. Preamble-Based Channel Estimation for CP-OFDM and OFDM/OQAM Systems: A Comparative Study. *IEEE Trans. Signal Process.* **2010**, *58*, 2911–2916. [CrossRef]
71. Katselis, D. Some Preamble Design Aspects in CP-OFDM Systems. *IEEE Commun. Lett.* **2012**, *16*, 356–359. [CrossRef]
72. Chen, D.; Xia, X.G.; Jiang, T.; Gao, X. Properties and Power Spectral Densities of CP Based OQAM-OFDM Systems. *IEEE Trans. Signal Process.* **2015**, *63*, 3561–3575. [CrossRef]
73. Lee, K.F.; Williams, D.B. A space-frequency transmitter diversity technique for OFDM systems. In Proceedings of the Globecom'00-IEEE Global Telecommunications Conference. Conference Record (Cat. No. 00CH37137), San Francisco, CA, USA, 27 November–1 December 2000; Volume 3, pp. 1473–1477.
74. Batarriere, M.; Baum, K.; Krauss, T. Cyclic prefix length analysis for 4G OFDM systems. In Proceedings of the IEEE 60th Vehicular Technology Conference, Los Angeles, CA, USA, 26–29 September 2004; Volume 1, pp. 543–547. [CrossRef]
75. Cimini, L. Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing. *IEEE Trans. Commun.* **1985**, *33*, 665–675. [CrossRef]
76. Petrovic, D.; Rave, W.; Fettweis, G. Effects of Phase Noise on OFDM Systems with and without PLL: Characterization and Compensation. *IEEE Trans. Commun.* **2007**, *55*, 1607–1616. [CrossRef]
77. Ann, P.P.; Jose, R. Comparison of PAPR reduction techniques in OFDM systems. In Proceedings of the 2016 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 21–22 October 2016; pp. 1–5. [CrossRef]
78. Coleri, S.; Ergen, M.; Puri, A.; Bahai, A. Channel estimation techniques based on pilot arrangement in OFDM systems. *IEEE Trans. Broadcast.* **2002**, *48*, 223–229. [CrossRef]
79. Nagaraju, V.S.; Anusha, R.; Vallabhuni, R.R. A Hybrid PAPR Reduction Technique in OFDM Systems. In Proceedings of the 2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE), Bhubaneswar, India, 26–27 December 2020; pp. 364–367. [CrossRef]
80. Jawhar, Y.A.; Audah, L.; Taher, M.A.; Ramli, K.N.; Shah, N.S.M.; Musa, M.; Ahmed, M.S. A review of partial transmit sequence for PAPR reduction in the OFDM systems. *IEEE Access* **2019**, *7*, 18021–18041. [CrossRef]
81. Sandoval, F.; Poitau, G.; Gagnon, F. Hybrid Peak-to-Average Power Ratio Reduction Techniques: Review and Performance Comparison. *IEEE Access* **2017**, *5*, 27145–27161. [CrossRef]
82. Wang, Y.; Wang, Y.; Shi, Q. Optimized signal distortion for PAPR reduction of OFDM signals with IFFT/FFT complexity via ADMM approaches. *IEEE Trans. Signal Process.* **2018**, *67*, 399–414. [CrossRef]
83. Pachori, K.; Mishra, A. An efficient combinational approach for PAPR reduction in MIMO-OFDM system. *Wirel. Netw.* **2016**, *22*, 417–425. [CrossRef]
84. Cimini, L.J.; Sollenberger, N.R. Peak-to-average power ratio reduction of an OFDM signal using partial transmit sequences. *IEEE Commun. Lett.* **2000**, *4*, 86–88. [CrossRef]
85. Ghassemi, A.; Gulliver, T.A. PAPR reduction of OFDM using PTS and error-correcting code subblocking—Transactions Papers. *IEEE Trans. Wirel. Commun.* **2010**, *9*, 980–989. [CrossRef]
86. Xiao, Y.; Lei, X.; Wen, Q.; Li, S. A Class of Low Complexity PTS Techniques for PAPR Reduction in OFDM Systems. *IEEE Signal Process. Lett.* **2007**, *14*, 680–683. [CrossRef]
87. Zhou, Y.; Jiang, T. A Novel Multi-Points Square Mapping Combined with PTS to Reduce PAPR of OFDM Signals without Side Information. *IEEE Trans. Broadcast.* **2009**, *55*, 831–835. [CrossRef]
88. Al-Jawhar, Y.A.; Ramli, K.N.; Mustapha, A.; Mostafa, S.A.; Shah, N.S.M.; Taher, M.A. Reducing PAPR with low complexity for 4G and 5G waveform designs. *IEEE Access* **2019**, *7*, 97673–97688. [CrossRef]
89. Han, S.H.; Lee, J.H. PAPR reduction of OFDM signals using a reduced complexity PTS technique. *IEEE Signal Process. Lett.* **2004**, *11*, 887–890. [CrossRef]
90. Ho, W.S.; Madhukumar, A.; Chin, F. Peak-to-average power reduction using partial transmit sequences: A suboptimal approach based on dual layered phase sequencing. *IEEE Trans. Broadcast.* **2003**, *49*, 225–231.
91. Lim, D.W.; Heo, S.J.; No, J.S.; Chung, H. A new PTS OFDM scheme with low complexity for PAPR reduction. *IEEE Trans. Broadcast.* **2006**, *52*, 77–82. [CrossRef]
92. Bharati, S.; Podder, P. Adaptive PAPR reduction scheme for OFDM using SLM with the fusion of proposed clipping and filtering technique in order to diminish PAPR and signal distortion. *Wirel. Pers. Commun.* **2020**, *113*, 2271–2288. [CrossRef]

93. Heo, S.J.; Noh, H.S.; No, J.S.; Shin, D.J. A Modified SLM Scheme with Low Complexity for PAPR Reduction of OFDM Systems. *IEEE Trans. Broadcast.* **2007**, *53*, 804–808. [CrossRef]
94. Baig, I.; Jeoti, V. DCT precoded SLM technique for PAPR reduction in OFDM systems. In Proceedings of the 2010 International Conference on Intelligent and Advanced Systems, Kuala Lumpur, Malaysia, 15–17 June 2010; pp. 1–6. [CrossRef]
95. Breiling, H.; Muller-Weinfurter, S.H.; Huber, J.B. SLM peak-power reduction without explicit side information. *IEEE Commun. Lett.* **2001**, *5*, 239–241. [CrossRef]
96. Le Goff, S.Y.; Khoo, B.K.; Tsimenidis, C.C.; Sharif, B.S. A novel selected mapping technique for PAPR reduction in OFDM systems. *IEEE Trans. Commun.* **2008**, *56*, 1775–1779. [CrossRef]
97. Lim, D.W.; Heo, S.J.; No, J.S.; Chung, H. *On the Phase Sequences of SLM OFDM System for PAPR Reduction*; ISITA: Parma, Italy, 2004; Volume 10, pp. 230–235.
98. Lee, B.M.; Kim, Y.; De Figueiredo, R.J. Performance analysis of the clipping scheme with SLM technique for PAPR reduction of OFDM signals in fading channels. *Wirel. Pers. Commun.* **2012**, *63*, 331–344. [CrossRef]
99. Ryu, H.G.; Kim, S.K.; Ryu, S.B. Interleaving method without side information for the PAPR reduction of OFDM system. In Proceedings of the 2007 International Symposium on Communications and Information Technologies, Sydney, NSW, Australia, 17–19 October 2007; pp. 72–76. [CrossRef]
100. Jayalath, A.D.; Tellambura, C. Reducing the peak-to-average power ratio of orthogonal frequency division multiplexing signal through bit or symbol interleaving. *Electron. Lett.* **2000**, *36*, 1161–1163. [CrossRef]
101. Zhang, S.; Shahrrava, B. A selected mapping technique using interleavers for PAPR reduction in OFDM systems. *Wirel. Pers. Commun.* **2018**, *99*, 329–338. [CrossRef]
102. Sudha, V.; Syamkumar, M.; Sriram Kumar, D. A low complexity modified SLM and companding based PAPR reduction in localized OFDMA. *Wirel. Pers. Commun.* **2017**, *96*, 3207–3226. [CrossRef]
103. Wang, L.; Tellambura, C. A simplified clipping and filtering technique for PAR reduction in OFDM systems. *IEEE Signal Process. Lett.* **2005**, *12*, 453–456. [CrossRef]
104. Anoh, K.; Tanriover, C.; Adebisi, B.; Hammoudeh, M. A new approach to iterative clipping and filtering PAPR reduction scheme for OFDM systems. *IEEE Access* **2017**, *6*, 17533–17544. [CrossRef]
105. Wang, Y.C.; Luo, Z.Q. Optimized iterative clipping and filtering for PAPR reduction of OFDM signals. *IEEE Trans. Commun.* **2010**, *59*, 33–37. [CrossRef]
106. Zhu, X.; Pan, W.; Li, H.; Tang, Y. Simplified approach to optimized iterative clipping and filtering for PAPR reduction of OFDM signals. *IEEE Trans. Commun.* **2013**, *61*, 1891–1901. [CrossRef]
107. Wang, Y.; Chen, W.; Tellambura, C. Genetic algorithm based nearly optimal peak reduction tone set selection for adaptive amplitude clipping PAPR reduction. *IEEE Trans. Broadcast.* **2012**, *58*, 462–471. [CrossRef]
108. Sun, Y.; Ochiai, H. Maximum-Likelihood-Based Performance Enhancement of Clipped and Filtered OFDM Systems with Clipping Noise Cancellation. *IEEE Wirel. Commun. Lett.* **2022**, *11*, 448–452. [CrossRef]
109. Deng, S.K.; Lin, M.C. Recursive clipping and filtering with bounded distortion for PAPR reduction. *IEEE Trans. Commun.* **2007**, *55*, 227–230. [CrossRef]
110. Lee, B.M.; Kim, Y. An adaptive clipping and filtering technique for PAPR reduction of OFDM signals. *Circuits Syst. Signal Process.* **2013**, *32*, 1335–1349. [CrossRef]
111. Bakkas, B.; Benkhrouya, R.; Chana, I.; Ben-Azza, H. Palm date leaf clipping: A new method to reduce PAPR in OFDM systems. *Information* **2020**, *11*, 190. [CrossRef]
112. Jiang, T.; Wu, Y. An overview: Peak-to-average power ratio reduction techniques for OFDM signals. *IEEE Trans. Broadcast.* **2008**, *54*, 257–268. [CrossRef]
113. Tuna, C.; Jones, D.L. Tone injection with aggressive clipping projection for OFDM PAPR reduction. In Proceedings of the 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, Dallas, TX, USA 14–19 March 2010; pp. 3278–3281.
114. Alsusa, E.; Yang, L. A Low-Complexity Time-Domain Linear Symbol Combining Technique for PAPR Reduction in OFDM Systems. *IEEE Trans. Signal Process.* **2008**, *56*, 4844–4855. [CrossRef]
115. Guel, D.; Palicot, J. FFT/IFFT Pair Based Digital Filtering for the Transformation of Adding Signal PAPR Reduction Techniques in Tone Reservation Techniques. In Proceedings of the 2009 Fifth International Conference on Wireless and Mobile Communications, Cannes/La Bocca, France, 23–29 August 2009; pp. 200–204. [CrossRef]
116. Singh, V.; Sharma, R. PAPR reduction techniques in LTE/LTE-A systems. In Proceedings of the 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies, Ramanathapuram, India, 8–10 May 2014; pp. 688–692. [CrossRef]
117. Wattanasuwakul, T.; Benjapolakul, W. PAPR reduction for OFDM transmission by using a method of tone reservation and tone injection. In Proceedings of the 2005 5th International Conference on Information Communications and Signal Processing, Bangkok, Thailand, 6–9 December 2005; pp. 273–277.
118. Wang, Y.; Zhang, R.; Li, J.; Shu, F. PAPR Reduction Based on Parallel Tabu Search for Tone Reservation in OFDM Systems. *IEEE Wirel. Commun. Lett.* **2019**, *8*, 576–579. [CrossRef]
119. Jiao, Y.; Liu, X.; Wang, X. A novel tone reservation scheme with fast convergence for PAPR reduction in OFDM systems. In Proceedings of the 2008 5th IEEE Consumer Communications and Networking Conference, Las Vegas, NV, USA, 10–12 January 2008; pp. 398–402.

120. Moon, J.H.; Nam, Y.R.; Kim, J.H. PAPR Reduction in the FBMC-OQAM System via Segment-Based Optimization. *IEEE Access* **2018**, *6*, 4994–5002. [CrossRef]
121. Elavarasan, P.; Nagarajan, G. A summarization on PAPR techniques for OFDM systems. *J. Inst. Eng. (India) Ser. B* **2015**, *96*, 381–389. [CrossRef]
122. Chen, J.C.; Li, C.P. Tone reservation using near-optimal peak reduction tone set selection algorithm for PAPR reduction in OFDM systems. *IEEE Signal Process. Lett.* **2010**, *17*, 933–936. [CrossRef]
123. Abdalmunam, A.; Anuar, M.S.; Junita, M.N. Use of Peak-to-Average Power Ratio in Orthogonal Frequency Division Multiplexing Techniques: A Review of the Drawbacks. *J. Theor. Appl. Inf. Technol.* **2019**, *97*, 3170–3180.
124. Youssef, M.I.; Tarrad, I.F.; Mounir, M. Performance evaluation of hybrid ACE-PTS PAPR reduction techniques. In Proceedings of the 2016 11th International Conference on Computer Engineering and Systems (ICCES), Cairo, Egypt, 20–21 December 2016; pp. 407–413. [CrossRef]
125. Kang, B.M.; Ryu, H.G.; Ryu, S.B. A PAPR Reduction Method using New ACE (Active Constellation Extension) with Higher Level Constellation. In Proceedings of the 2007 IEEE International Conference on Signal Processing and Communications, Dubai, United Arab Emirates, 24–27 November 2007; pp. 724–727. [CrossRef]
126. Krongold, B.S.; Jones, D.L. PAR reduction in OFDM via active constellation extension. *IEEE Trans. Broadcast.* **2003**, *49*, 258–268. [CrossRef]
127. Van der Neut, N.; Maharaj, B.T.; De Lange, F.; González, G.J.; Gregorio, F.; Cousseau, J. PAPR reduction in FBMC using an ACE-based linear programming optimization. *EURASIP J. Adv. Signal Process.* **2014**, *2014*, 172. [CrossRef]
128. Zhou, Y.; Jiang, T. A novel clipping integrated into ACE for PAPR reduction in OFDM systems. In Proceedings of the 2009 International Conference on Wireless Communications and Signal Processing, Nanjing, China, 13–15 November 2009; pp. 1–4. [CrossRef]
129. Liu, Y.; Wang, Y. Novel ACE Scheme for PAPR Reduction of High Broadband OFDM Systems. In Proceedings of the 2019 IEEE 19th International Conference on Communication Technology (ICCT), Xi'an, China, 16–19 October 2019; pp. 1–5. [CrossRef]
130. Deumal, M.; Behravan, A.; Eriksson, T.; Pijoan, J.L. Evaluation of performance improvement capabilities of PAPR-reducing methods. *Wirel. Pers. Commun.* **2008**, *47*, 137–147. [CrossRef]
131. Kashyap, P.; Thakur, K. Improved ACE method for reducing PAPR in OFDM system. In Proceedings of the 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, India, 20–21 April 2018; pp. 402–406. [CrossRef]
132. Wang, S.H.; Lin, W.L.; Huang, B.R.; Li, C.P. PAPR Reduction in OFDM Systems Using Active Constellation Extension and Subcarrier Grouping Techniques. *IEEE Commun. Lett.* **2016**, *20*, 2378–2381. [CrossRef]
133. Jones, D.L. Peak power reduction in OFDM and DMT via active channel modification. In Proceedings of the Conference Record of the Thirty-Third Asilomar Conference on Signals, Systems, and Computers (Cat. No. CH37020), Pacific Grove, CA, USA, 24–27 October 1999; Volume 2, pp. 1076–1079.
134. Kalinov, A.; Bychkov, R.; Ivanov, A.; Osinsky, A.; Yarotsky, D. Machine Learning-Assisted PAPR Reduction in Massive MIMO. *IEEE Wirel. Commun. Lett.* **2021**, *10*, 537–541. [CrossRef]
135. Gökceli, S.; Riihonen, T.; Levanen, T.; Valkama, M. Machine Learning Based Tuner for Frequency-Selective PAPR Reduction. *IEEE Trans. Veh. Technol.* **2023**, *72*, 5378–5383. [CrossRef]
136. Chen, M.; Challita, U.; Saad, W.; Yin, C.; Debbah, M. Artificial neural networks-based machine learning for wireless networks: A tutorial. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 3039–3071. [CrossRef]
137. Hu, W.W.; Lee, D.H. PAPR reduction for visible light communication systems without side information. *IEEE Photonics J.* **2017**, *9*, 1–11. [CrossRef]
138. Chacko, A.; Jeevitha, J. PAPR reduction techniques for performance improvement in OFDM systems. In Proceedings of the 2014 International Conference on Electronics and Communication Systems (ICECS), Coimbatore, India, 13–14 February 2014; pp. 1–4.
139. Wang, P.; Lin, S.C.; Luo, M. A framework for QoS-aware traffic classification using semi-supervised machine learning in SDNs. In Proceedings of the 2016 IEEE International Conference on Services Computing (SCC), San Francisco, CA, USA, 27 June–2 July 2016; pp. 760–765.
140. Le Goff, S.Y.; Al-Samahi, S.S.; Khoo, B.K.; Tsimenidis, C.C.; Sharif, B.S. Selected mapping without side information for PAPR reduction in OFDM. *IEEE Trans. Wirel. Commun.* **2009**, *8*, 3320–3325. [CrossRef]
141. Shi, L.; Zhang, X.; Wang, W.; Wang, Z.; Vladimirescu, A.; Zhang, Y.; Wang, J. PAPR reduction based on deep autoencoder for VLC DCO-OFDM system. In Proceedings of the 2019 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Jeju, Republic of Korea, 5–7 June 2019; pp. 1–4.
142. Zhang, M.; Liu, M.; Zhong, Z. Neural network assisted active constellation extension for PAPR reduction of OFDM system. In Proceedings of the 2019 11th International Conference on Wireless Communications and Signal Processing (WCSP), Xi'an, China, 23–25 October 2019; pp. 1–6.
143. Hao, L.; Wang, D.; Cheng, W.; Li, J.; Ma, A. Performance enhancement of ACO-OFDM-based VLC systems using a hybrid autoencoder scheme. *Opt. Commun.* **2019**, *442*, 110–116. [CrossRef]
144. Chakravarty, S.; Kumar, A. PAPR Reduction of GFDM Signals Using Encoder-Decoder Neural Network (Autoencoder). *Natl. Acad. Sci. Lett.* **2023**, *46*, 213–217. [CrossRef]

145. Liu, Z.; Hu, X.; Han, K.; Zhang, S.; Sun, L.; Xu, L.; Wang, W.; Ghannouchi, F.M. Low-complexity PAPR reduction method for OFDM systems based on real-valued neural networks. *IEEE Wirel. Commun. Lett.* **2020**, *9*, 1840–1844. [CrossRef]
146. Mohamed, A.; Eldien, A.S.T.; Fouda, M.M.; Saad, R.S. LSTM-Autoencoder Deep Learning Technique for PAPR Reduction in Visible Light Communication. *IEEE Access* **2022**, *10*, 113028–113034. [CrossRef]
147. Miao, P.; Zhu, B.; Qi, C.; Jin, Y.; Lin, C. A model-driven deep learning method for LED nonlinearity mitigation in OFDM-based optical communications. *IEEE Access* **2019**, *7*, 71436–71446. [CrossRef]
148. Wang, B.; Si, Q.; Jin, M. A novel tone reservation scheme based on deep learning for PAPR reduction in OFDM systems. *IEEE Commun. Lett.* **2020**, *24*, 1271–1274. [CrossRef]
149. Wang, X.; Jin, N.; Wei, J. A model-driven DL algorithm for PAPR reduction in OFDM system. *IEEE Commun. Lett.* **2021**, *25*, 2270–2274. [CrossRef]
150. Mitsiou, N.A.; Diamantoulakis, P.D.; Sarigiannidis, P.G.; Karagiannidis, G.K. Energy Efficient OFDM with Intelligent PAPR-aware Adaptive Modulation. *IEEE Commun. Lett.* **2023**, *27*, 3290–3294. [CrossRef]
151. Van Luong, T.; Zhang, X.; Xiang, L.; Hoang, T.M.; Xu, C.; Petropoulos, P.; Hanzo, L. Deep learning-aided optical IM/DD OFDM approaches the throughput of RF-OFDM. *IEEE J. Sel. Areas Commun.* **2021**, *40*, 212–226. [CrossRef]
152. Nguyen, M.T.; Kaddoum, G.; Selim, B.; Srinivas, K.; de Araujo-Filho, P.F. Deep Unfolding Network for PAPR Reduction in Multicarrier OFDM Systems. *IEEE Commun. Lett.* **2022**, *26*, 2616–2620. [CrossRef]
153. Omar, M.S.; Qi, J.; Ma, X. Mitigating Clipping Distortion in Multicarrier Transmissions Using Tensor-Train Deep Neural Networks. *IEEE Trans. Wirel. Commun.* **2022**, *22*, 2127–2138. [CrossRef]
154. Aoudia, F.A.; Hoydis, J. Waveform learning for next-generation wireless communication systems. *IEEE Trans. Commun.* **2022**, *70*, 3804–3817. [CrossRef]
155. Ye, S.; Wang, T.; Wang, S. PAPR Reduction for OFDM Signals: A Monte Carlo Tree Search Method. In Proceedings of the 2021 IEEE/CIC International Conference on Communications in China (ICCC), Xiamen, China, 28–30 July 2021; pp. 770–775.
156. Marugán, A.P.; Márquez, F.P.G.; Perez, J.M.P.; Ruiz-Hernández, D. A survey of artificial neural network in wind energy systems. *Appl. Energy* **2018**, *228*, 1822–1836. [CrossRef]
157. Mas'ud, A.A.; Albarracín, R.; Ardila-Rey, J.A.; Muhammad-Sukki, F.; Illias, H.A.; Bani, N.A.; Munir, A.B. Artificial neural network application for partial discharge recognition: Survey and future directions. *Energies* **2016**, *9*, 574. [CrossRef]
158. Mittal, S. A survey on modeling and improving reliability of DNN algorithms and accelerators. *J. Syst. Archit.* **2020**, *104*, 101689. [CrossRef]
159. Hussain, H.; Tamizharasan, P.; Rahul, C. Design possibilities and challenges of DNN models: A review on the perspective of end devices. *Artif. Intell. Rev.* **2022**, *55*, 5109–5167. [CrossRef]
160. Świechowski, M.; Godlewski, K.; Sawicki, B.; Mańdziuk, J. Monte Carlo tree search: A review of recent modifications and applications. *Artif. Intell. Rev.* **2023**, *56*, 2497–2562. [CrossRef]
161. Balatsoukas-Stimming, A.; Studer, C. Deep unfolding for communications systems: A survey and some new directions. In Proceedings of the 2019 IEEE International Workshop on Signal Processing Systems (SiPS), Nanjing, China, 20–23 October 2019; pp. 266–271.
162. Manisha; Dhull, S.K.; Singh, K.K. ECG beat classifiers: A journey from ANN to DNN. *Procedia Comput. Sci.* **2020**, *167*, 747–759. [CrossRef]
163. Lorentz, R.J. Improving monte-carlo tree search in havannah. In Proceedings of the International Conference on Computers and Games, Kanazawa, Japan, 24–26 September 2010; Springer: Berlin/Heidelberg, Germany, 2010; pp. 105–115.
164. Jagannath, A.; Jagannath, J.; Melodia, T. Redefining wireless communication for 6G: Signal processing meets deep learning with deep unfolding. *IEEE Trans. Artif. Intell.* **2021**, *2*, 528–536. [CrossRef]
165. Nguyen, N.T.; Ma, M.; Lavi, O.; Shlezinger, N.; Eldar, Y.C.; Swindlehurst, A.L.; Juntti, M. Deep unfolding hybrid beamforming designs for THz massive MIMO systems. *IEEE Trans. Signal Process.* **2023**, *71*, 3788–3804. [CrossRef]
166. Varshney, P.; Babu, P.; Stoica, P. Low-PAPR OFDM Waveform Design for Radar and Communication Systems. *IEEE Trans. Radar Syst.* **2023**, *1*, 69–74. [CrossRef]
167. Shang, H.; Chen, R.; Zhang, H.; Ma, G.; He, R.; Ai, B.; Zhong, Z. OTFS modulation and PAPR reduction for IoT-railways. *China Commun.* **2023**, *20*, 102–113. [CrossRef]
168. Qiu, C.; Chen, Z.; Ren, X.; Dai, Z.; Zhang, C.; Wang, X. AI-mers-6G: AI-Driven Region-temporal Resource Provisioning for 6G Immersive Services. *IEEE Wirel. Commun.* **2023**, *30*, 196–203. [CrossRef]
169. Tao, Y.; Wu, J.; Lin, X.; Yang, W. DRL-Driven Digital Twin Function Virtualization for Adaptive Service Response in 6G Networks. *IEEE Netw. Lett.* **2023**, *5*, 125–129. [CrossRef]
170. Brito, J.M.C.; Mendes, L.L.; Gontijo, J.G.S. Brazil 6G Project—An Approach to Build a National-wise Framework for 6G Networks. In Proceedings of the 2020 2nd 6G Wireless Summit (6G SUMMIT), Levi, Finland, 17–20 March 2020; pp. 1–5. [CrossRef]
171. Wang, C.X.; You, X.; Gao, X.; Zhu, X.; Li, Z.; Zhang, C.; Wang, H.; Huang, Y.; Chen, Y.; Haas, H.; et al. On the Road to 6G: Visions, Requirements, Key Technologies, and Testbeds. *IEEE Commun. Surv. Tutor.* **2023**, *25*, 905–974. [CrossRef]
172. Ericsson. *Defining AI Native: A Key Enabler for Advanced Intelligent Telecom Networks*; White Paper; Ericsson: Stockholm, Sweden, 2023.
173. Mueck, M.D.; On, A.E.B.; Du Boispean, S. Upcoming European Regulations on Artificial Intelligence and Cybersecurity. *IEEE Commun. Mag.* **2023**, *61*, 98–102. [CrossRef]

174. Yrjölä, S.; Ahokangas, P.; Matinmikko-Blue, M. (Eds.) White Paper on Business of 6G; Number 3 in 6G Research Visions, University of Oulu. 2020. Available online: <http://urn.fi/urn:isbn:9789526226767> (accessed on 1 February 2024).
175. Wu, W.; Yang, F.; Zhou, F.; Wu, Q.; Hu, R.Q. Intelligent Resource Allocation for IRS-Enhanced OFDM Communication Systems: A Hybrid Deep Reinforcement Learning Approach. *IEEE Trans. Wirel. Commun.* **2023**, *22*, 4028–4042. [CrossRef]
176. Siddiki Abir, M.A.B.; Chowdhury, M.Z.; Jang, Y.M. Software-Defined UAV Networks for 6G Systems: Requirements, Opportunities, Emerging Techniques, Challenges, and Research Directions. *IEEE Open J. Commun. Soc.* **2023**, *4*, 2487–2547. [CrossRef]
177. Shen, S.; Yu, C.; Zhang, K.; Ni, J.; Ci, S. Adaptive and Dynamic Security in AI-Empowered 6G: From an Energy Efficiency Perspective. *IEEE Commun. Stand. Mag.* **2021**, *5*, 80–88. [CrossRef]
178. Zhou, D.; Sheng, M.; Li, J.; Han, Z. Aerospace Integrated Networks Innovation for Empowering 6G: A Survey and Future Challenges. *IEEE Commun. Surv. Tutor.* **2023**, *25*, 975–1019. [CrossRef]
179. Shi, Y.; Lian, L.; Shi, Y.; Wang, Z.; Zhou, Y.; Fu, L.; Bai, L.; Zhang, J.; Zhang, W. Machine Learning for Large-Scale Optimization in 6G Wireless Networks. *IEEE Commun. Surv. Tutor.* **2023**, *25*, 2088–2132. [CrossRef]
180. Xu, E. *Envisioning and Defining 6G Together*; Huawei: Shenzhen, China, 2021.
181. Li, X.; Zhang, H.; Peng, C.; Liu, Z.; Wang, F. *NET4AI: Supporting AI as a Service in 6G*; HuaweiTech Future Technologies: Shenzhen, China, 2022.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

MDPI AG
Grosspeteranlage 5
4052 Basel
Switzerland
Tel.: +41 61 683 77 34

MDPI Books Editorial Office
E-mail: books@mdpi.com
www.mdpi.com/books



Disclaimer/Publisher's Note: The title and front matter of this reprint are at the discretion of the Topic Editors. The publisher is not responsible for their content or any associated concerns. The statements, opinions and data contained in all individual articles are solely those of the individual Editors and contributors and not of MDPI. MDPI disclaims responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Academic Open
Access Publishing

mdpi.com

ISBN 978-3-7258-8820-7