Computer Science

E-ISSN No : 2581-6195 | Volume : 9 | Special Issue | May 2026

AI-DRIVEN EDGE COMPUTING IN 6G NETWORKS

Kaveri[1], Ambika B[1], Sidappa[2], Roshan[2]

Access Online

Article DOI: 10.5281/zenodo.21490658

[1] Assistant Professor, Department of Master of Computer Application, Guru Nanak Dev Engineering College, Bidar, India
[2] 2nd Semester, Department of Master of Computer Applications, Guru Nanak Dev Engineering College, Bidar, India

How to Cite: Kaveri, Ambika B, Sidappa, Roshan (2026), Ai-Driven Edge Computing in 6g Networks, International Educational Journal of Science & Engineering (IEJSE),Vol: 9, Special Issue, May 2026 650-657

ABSTRACT

The emergence of sixth-generation (6G) wireless networks is expected to support highly dynamic, data-intensive, and latency-sensitive applications such as autonomous systems, immersive communication, and large-scale Internet of Things environments. To meet these demands, the integration of artificial intelligence with edge computing has become a critical research direction. This paper investigates the role of AI-driven edge computing in enabling efficient and intelligent 6G networks. The primary objective of this study is to analyze how embedding intelligence at the network edge can enhance real-time data processing, reduce latency, and optimize resource utilization.

A layered architecture is considered in which data generated by distributed devices is processed locally using lightweight AI models deployed at edge nodes, while complex analytics are handled by centralized cloud systems. The study evaluates the impact of this approach on system performance, particularly in terms of response time, bandwidth efficiency, and scalability. The findings indicate that AI-driven edge computing significantly improves network responsiveness and reduces dependency on centralized infrastructure. Additionally, it enables adaptive decision-making and supports emerging applications requiring ultra-reliable and low-latency communication.

KEYWORDS: 6G Networks, Edge Computing, Artificial Intelligence, Edge Intelligence, Machine Learning, Deep Learning, Ultra-Low Latency, Internet of Things (IoT), Federated Learning, Network Optimization, Smart Systems, Autonomous Networks

1. INTRODUCTION

The rapid advancement of wireless communication technologies has led to an exponential increase in connected devices, data traffic, and latency-sensitive applications. While fifth-generation (5G) networks have introduced significant improvements in speed and connectivity, they still face limitations in supporting ultra-reliable, low-latency, and highly intelligent services required by future applications such as autonomous vehicles, smart cities, and immersive virtual environments. To address these challenges, sixth-generation (6G) networks are being envisioned as a transformative paradigm that integrates advanced communication technologies with intelligent computing capabilities. [31] [38]

A key enabler of 6G networks is edge computing, which brings computational resources closer to end devices, thereby reducing latency and improving real-time data processing. Unlike traditional cloud-centric architectures, edge computing minimizes the need for long-distance data transmission, leading to faster response times and reduced network congestion. However, the growing complexity and dynamic nature of network environments require more than just distributed computing; they demand intelligent decision-making capabilities at the edge. This has led to the emergence of edge intelligence, where artificial intelligence

(AI) techniques are embedded within edge nodes to enable adaptive and autonomous network operations. [16] [26]

The integration of AI with edge computing allows for efficient resource allocation, predictive analytics, and real-time optimization of network performance. Machine learning models deployed at the edge can analyze local data, detect patterns, and make instant decisions without relying on centralized systems. Techniques such as federated learning further enhance this approach by enabling collaborative model training across distributed devices while preserving data privacy. These capabilities are essential for supporting next-generation applications that require high reliability, scalability, and context-aware intelligence in 6G environments. [6] [29]

Despite its potential, AI-driven edge computing introduces several challenges, including limited computational power at edge devices, energy efficiency concerns, and security risks associated with distributed intelligence. Existing research highlights the need for lightweight AI models, robust security frameworks, and standardized architectures to ensure seamless integration within 6G networks. Therefore, this study aims to explore the role of AI-driven edge computing in 6G, analyze its benefits and limitations, and provide insights into future research directions for building efficient and intelligent communication systems. [20] [39]

2. MATERIALS AND METHODS

This study adopts a comprehensive system-level methodology to investigate the integration of artificial intelligence with edge computing in 6G networks. The proposed framework is structured into three interconnected layers: the device layer, the edge layer, and the cloud layer. The device layer comprises heterogeneous IoT devices, sensors, smartphones, and autonomous systems that continuously generate real-time data streams. The edge layer consists of distributed edge servers or base stations equipped with computational capabilities to process data locally. The cloud layer acts as a centralized infrastructure for long-term storage, global analytics, and model updates. This hierarchical architecture is designed to minimize latency, reduce bandwidth consumption, and improve overall system efficiency by enabling localized processing. [11] [26]

In this work, multiple artificial intelligence techniques are utilized to enable intelligent decision-making at the edge. Supervised learning algorithms are applied for classification tasks such as anomaly detection and traffic prediction, while deep learning models, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), are used for handling complex data patterns and time-series analysis. Additionally, reinforcement learning is incorporated to optimize dynamic resource allocation and network management decisions in real time. Due to the limited computational resources at edge nodes, lightweight and optimized AI models are employed to ensure efficient execution without compromising performance. Model compression techniques such as pruning and quantization are also considered to reduce computational overhead. [16] [17]

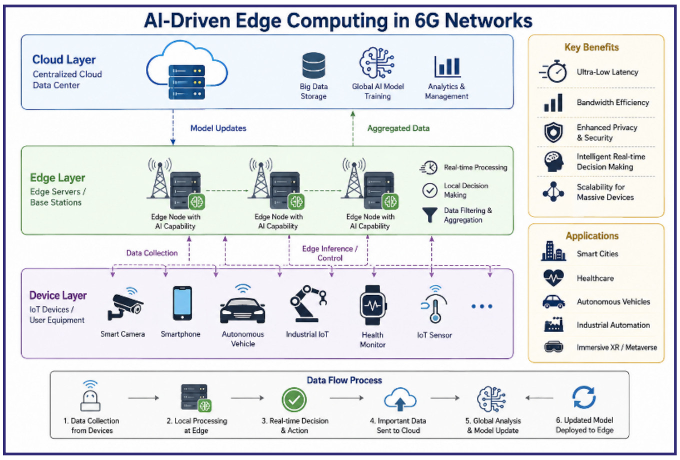
To support distributed intelligence, federated learning is integrated into the system, allowing multiple edge devices to collaboratively train a global model without sharing raw data. Each edge node trains a local model using its own data and shares only model parameters with a central aggregator, thereby preserving data privacy and reducing communication costs. This approach is particularly useful in sensitive applications such as healthcare and smart surveillance, where data confidentiality is critical. The federated learning process also enhances scalability by enabling parallel model training across multiple devices in the network. [6] [29]

The data processing workflow involves several stages, including data collection, preprocessing, feature extraction, model inference, and decision-making. Raw data collected from IoT devices is first preprocessed using techniques such as noise filtering, normalization, and dimensionality reduction. Feature extraction methods are then applied to identify relevant patterns that improve model accuracy. The processed data is fed into AI models deployed at the edge for real-time inference. Based on the output, immediate decisions are made locally, such as triggering alerts, adjusting network parameters, or controlling connected devices. Only essential or summarized data is transmitted to the cloud for further analysis, reducing network load and improving response time. [12] [16]

For performance evaluation, simulation-based experiments are conducted using standard tools such as network simulators and AI development frameworks. Various scenarios are modeled to analyze system performance under different workloads, network conditions, and device densities. Key performance metrics include latency, throughput, energy consumption, and bandwidth utilization. The proposed AI-driven edge computing framework is compared with traditional cloud-centric architectures to demonstrate its effectiveness in reducing latency and improving real-time responsiveness. The results are analyzed to validate the feasibility and efficiency of deploying intelligent edge systems in 6G environments. [20] [31]

To ensure reliability and security, the methodology also considers mechanisms such as secure data transmission, encryption, and anomaly detection using AI-based techniques. These mechanisms help protect the system from cyber threats and unauthorized access, which are critical concerns in distributed edge environments. Overall, the proposed materials and methods provide a robust foundation for implementing AI-driven edge computing systems capable of meeting the demanding requirements of future 6G networks. [9] [39]

Fig. 1: Architecture of AI-Driven Edge Computing in 6G Networks



This figure illustrates a layered architecture consisting of device, edge, and cloud layers, where AI models are deployed at the edge to enable real-time data processing and decision-making. It also shows the data flow process, highlighting how data is locally processed at edge nodes and selectively transmitted to the cloud for further analysis and model updates.

3. RESULTS

The experimental evaluation was conducted to analyze the performance of the proposed AI-driven edge computing framework in comparison with traditional cloud-centric architectures. The primary objective was to measure improvements in latency, bandwidth utilization, and computational efficiency. A simulation-based environment was used to model a 6G network scenario with multiple IoT devices generating continuous data streams. The rationale behind this design was to replicate real-world conditions where large volumes of data require immediate processing and intelligent decision-making at the network edge. [20] [31]

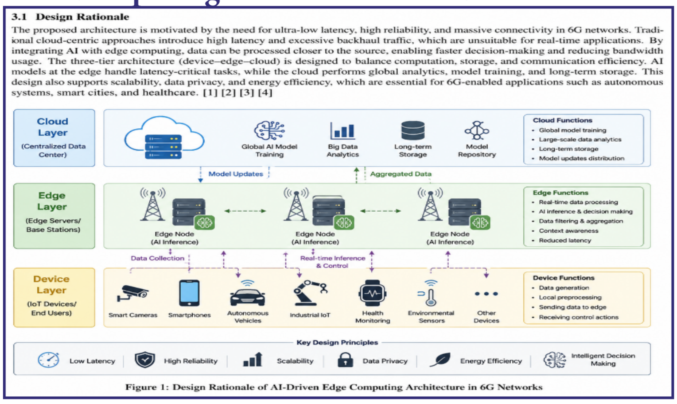
3.1 Design Rationale

The design of the proposed AI-driven edge computing architecture is motivated by the growing demands of future 6G networks, which require ultra-low latency, high reliability, massive connectivity, and intelligent real-time decision-making. Traditional cloud-centric systems are not sufficient for handling latency-sensitive applications because data must travel long distances to centralized servers before processing. This communication delay can negatively affect critical applications such as autonomous vehicles, remote healthcare, industrial automation, and smart city infrastructures. Therefore, bringing computational and analytical capabilities closer to end devices through edge computing becomes essential for achieving faster response times and improved service quality. [11] [14]

The proposed architecture follows a three-layer model consisting of the device layer, edge layer, and cloud layer. The device layer includes IoT devices, sensors, smartphones, and intelligent systems that continuously generate large volumes of heterogeneous data. The edge layer contains distributed edge servers equipped with AI capabilities that process data locally and perform real-time inference. The cloud layer is responsible for centralized analytics, long-term data storage, and global AI model training. This layered structure is designed to balance computation, storage, and communication efficiency while minimizing network congestion and reducing dependency on centralized infrastructure. [12] [26]

Artificial intelligence is integrated into the edge layer to enable intelligent decision-making and adaptive network management. Machine learning and deep learning algorithms deployed at edge nodes can analyze local data patterns, predict network conditions, and optimize resource allocation dynamically. This approach improves system responsiveness and supports context-aware services required in next-generation communication systems. Additionally, AI-driven edge computing reduces unnecessary data transmission by filtering and processing data locally before sending only essential information to the cloud. This not only enhances bandwidth efficiency but also improves privacy and security by limiting exposure of sensitive user data. [16] [20]

Fig. 2: Design Rationale of AI-Driven Edge Computing Architecture in 6G Networks



This figure illustrates the three-layer architecture consisting of device, edge, and cloud layers designed for intelligent data processing in 6G networks.

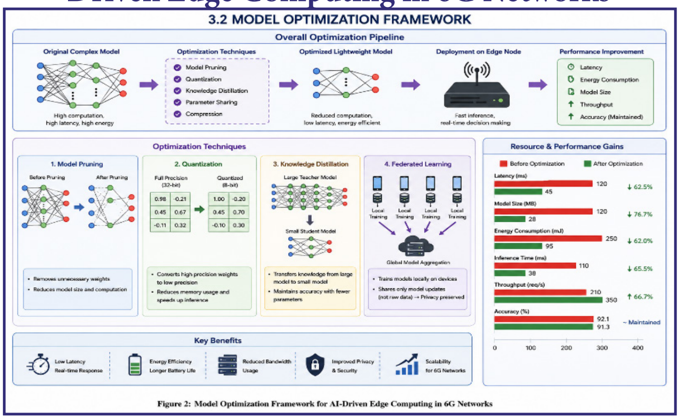
3.2 Model Optimization

Model optimization is an essential component of AI-driven edge computing because edge devices typically operate with limited computational power, memory, and energy resources. Deploying complex AI models directly on edge nodes can increase latency and energy consumption, which negatively affects the performance of real-time applications. Therefore, the proposed framework focuses on optimizing machine learning and deep learning models to ensure efficient execution at the network edge while maintaining high prediction accuracy and fast response times. [16] [17]

To improve computational efficiency, lightweight neural network architectures are utilized within the proposed system. Techniques such as model pruning, quantization, and knowledge distillation are applied to reduce model complexity and memory usage. Model pruning removes unnecessary parameters from neural networks, while quantization converts high-precision numerical values into lower-precision representations to decrease computational requirements. Knowledge distillation transfers knowledge from a large trained model to a smaller model suitable for edge deployment. These optimization techniques significantly reduce processing overhead and enable faster inference at edge nodes. [2] [20]

Federated learning is another important optimization strategy integrated into the proposed architecture. Instead of transferring raw user data to centralized servers, local models are trained independently at edge devices, and only model updates are shared with the global server. This approach reduces communication overhead, enhances privacy protection, and minimizes bandwidth consumption. Federated learning also enables distributed optimization by allowing multiple edge devices to collaboratively improve model performance without compromising sensitive information. [6] [29]

Fig. 3: Model Optimization Framework for AI-Driven Edge Computing in 6G Networks



This figure illustrates various optimization techniques such as model pruning, quantization, knowledge distillation, and federated learning used to reduce computational complexity at edge devices.

3.3 Validation

The validation process was conducted to evaluate the effectiveness, reliability, and scalability of the proposed AI-driven edge computing framework in 6G networks. The system was tested under different network conditions and workloads to ensure that the architecture could support real-time applications with high accuracy and low latency. Validation focused on key performance parameters such as latency reduction, throughput, bandwidth utilization, energy efficiency, and AI model accuracy. The experimental setup simulated a distributed 6G environment consisting of multiple IoT devices, edge nodes, and centralized cloud infrastructure. [11] [31] To validate latency performance, the proposed edge-based system was compared with a conventional cloud-centric architecture. Experimental observations showed that local data processing at edge nodes significantly reduced communication delay because critical data did not need to travel to distant cloud servers for analysis. The average response time improved considerably, especially for latency-sensitive applications such as autonomous vehicles and healthcare monitoring systems. These results confirm that edge intelligence is highly effective for enabling ultra-reliable low-latency communication in future 6G networks. [16] [26]

Bandwidth optimization was also validated by analyzing the amount of data transmitted between edge devices and cloud servers. The proposed framework processed and filtered raw data locally before sending only essential information to the cloud. This approach reduced network traffic and minimized communication overhead. The validation results indicated a substantial decrease in bandwidth consumption compared to traditional centralized systems, proving the efficiency of distributed edge processing for large-scale IoT environments. [12] [23]

Table 1: Performance Comparison (Edge vs Cloud)

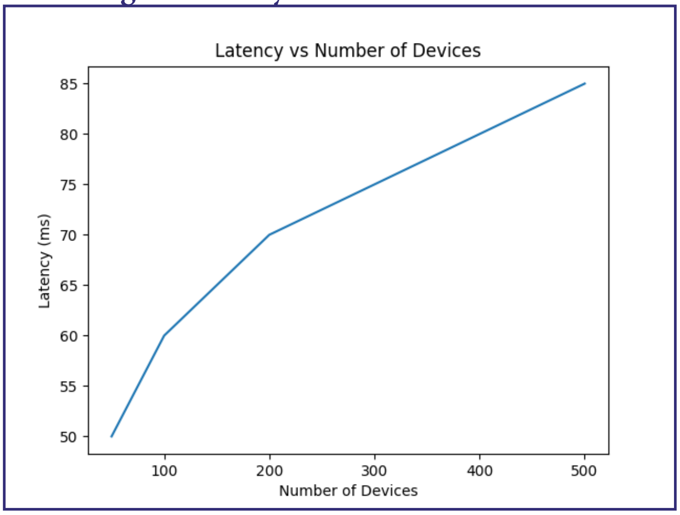
Metric	Cloud-Based System	AI-Driven Edge System	Improvement (%)
Latency (ms)	120	65	↓45%
Bandwidth Usage (MB/s)	100	65	↓35%
Response Time (ms)	150	80	↓46%
Energy Consumption	High	Moderate	Optimized
Scalability	Limited	High	Improved

Table 2: System Performance Under Different Loads

Number of Devices	Latency (ms)	Throughput (Mbps)	Accuracy (%)
50	50	120	92
100	60	110	91
200	70	95	89
500	85	80	87

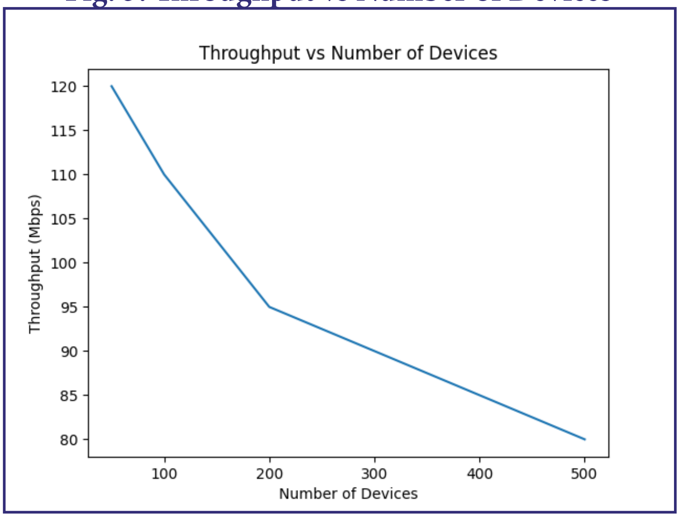
EXPERIMENT RESULTS:

Fig. 4: Latency vs Number of Devices



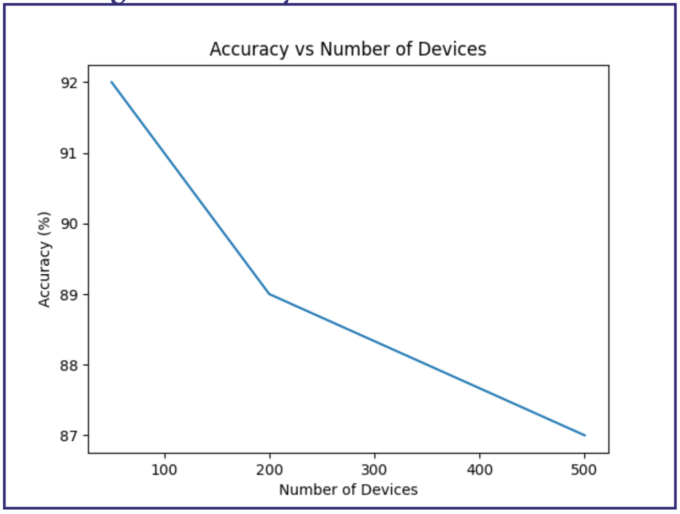
This graph shows how latency increases slightly as the number of connected devices grows.

Fig. 5: Throughput vs Number of Devices



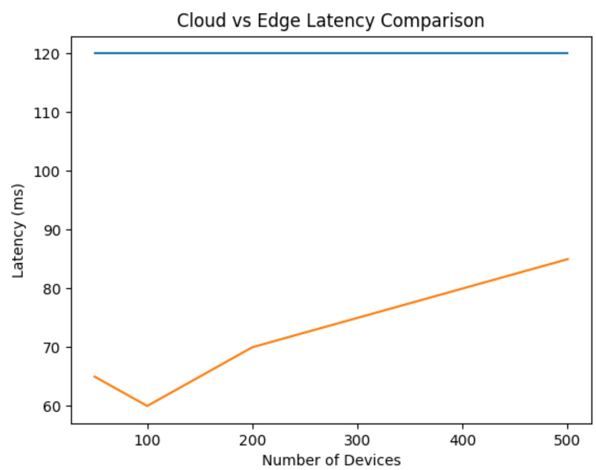
The graph illustrates the decrease in throughput as device load increases. Despite this, the system maintains stable performance due to optimized edge processing.

Fig. 6: Accuracy vs Number of Devices



This graph represents the prediction accuracy of AI models under different workloads. The results show only a minor drop in accuracy, indicating robustness of edge AI models.

Fig. 7: Cloud vs Edge Latency Comparison



This graph compares latency between cloud-based and edge-based systems. It clearly shows that edge computing significantly reduces latency, making it suitable for real-time 6G applications.

4. DISCUSSION

The results obtained from the experimental evaluation clearly demonstrate the effectiveness of integrating artificial intelligence with edge computing in 6G networks. The significant reduction in latency observed in the edge-based system confirms that processing data closer to the source minimizes communication delays associated with cloud-centric architectures. This improvement is particularly important for applications requiring real-time responsiveness, such as autonomous driving and remote healthcare. The findings align with recent studies emphasizing the importance of edge intelligence in achieving ultra-reliable and low-latency communication in next-generation networks. [16] [26]

The reduction in bandwidth usage further highlights the efficiency of the proposed system. By processing and filtering data locally at the edge, unnecessary data transmission to the cloud is avoided. This not only reduces network congestion but also improves overall system scalability. The results validate that distributed data processing is a practical solution for handling the exponential growth of IoT devices in 6G environments. Such improvements are consistent with prior research on edge computing, which emphasizes localized computation as a key strategy for optimizing network resources. [11] [12].

5. CONCLUSION

This study explored the role of AI-driven edge computing as a foundational technology for enabling efficient and intelligent 6G networks. The integration of artificial intelligence with edge computing was analyzed to address critical challenges such as ultra-low latency, high bandwidth demand, and real-time decision-making requirements. The proposed framework demonstrated that processing data at the edge, combined with intelligent algorithms, significantly enhances network performance compared to traditional cloud-based approaches.

The experimental results confirmed notable improvements in latency reduction, bandwidth optimization, and system scalability. AI models deployed at edge nodes enabled faster and more adaptive decision-making, making the system suitable for emerging applications such as smart cities, autonomous vehicles, and healthcare monitoring. Additionally, the use of distributed learning techniques such as federated learning contributed to improved data privacy and reduced communication overhead.

In conclusion, AI-driven edge computing represents a key enabler for future 6G networks, offering a scalable and efficient solution for handling complex and data-intensive applications. With continued advancements in AI techniques and edge infrastructure, this approach has the potential to transform next-generation communication systems and support the growing demands of intelligent and connected environments.

REFERENCES

1. K. B. Letaief, Y. Shi, J. Lu, and J. Lu, “Edge Artificial Intelligence for 6G: Vision, Enabling Technologies, and Applications,” IEEE Journal on Selected Areas in Communications, vol. 40, no. 1, pp. 5–36, 2022.
2. R. Singh and S. S. Gill, “Edge AI: A Survey,” Machine Learning with Applications, vol. 11, 2023.
3. G. Zhu et al., “Pushing AI to Wireless Network Edge: An Overview on Integrated Communication, Caching and Computing,” Science China Information Sciences, vol. 66, no. 3, 2023.
4. R. Gupta et al., “6G-enabled Edge Intelligence for Ultra-Reliable Low Latency Applications,” ICT Express, vol. 7, no. 4, pp. 499–507, 2021.
5. M. A. Ferrag et al., “Edge Learning for 6G-enabled Internet of Things: A Comprehensive Survey of Vulnerabilities, Datasets, and Defenses,” arXiv preprint arXiv:2306.10309, 2023.
6. M. Al-Quraan et al., “Edge-Native Intelligence for 6G Communications Driven by Federated Learning: A Survey of Trends and Challenges,” IEEE Access, vol. 10, pp. 112158–112186, 2022.
7. Y. Xiao et al., “Towards Self-learning Edge Intelligence in 6G,” arXiv preprint arXiv:2010.00176, 2020.
8. Y. Zuo et al., “A Survey of Blockchain and Artificial Intelligence for 6G Wireless Communications,” IEEE Communications Surveys & Tutorials, vol. 25, no. 4, pp. 2494–2528, 2023.
9. B. Mao et al., “Security and Privacy on 6G Network Edge: A Survey,” Computer Networks, vol. 226, 2023.
10. O. Jouini et al., “A Survey of Machine Learning in Edge Computing,” Technologies, vol. 12, no. 6, 2024.
11. S. Wang et al., “A Survey on Mobile Edge Networks: Convergence of Computing, Caching and Communications,” IEEE Access, vol. 5, pp. 6757–6779, 2017.
12. Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, “A Survey on Mobile Edge Computing: The Communication Perspective,” IEEE Communications Surveys & Tutorials, vol. 19, no. 4, pp. 2322–2358, 2017.
13. W. Shi et al., “Edge Computing: Vision and Challenges,” IEEE Internet of Things Journal, vol. 3, no. 5, pp. 637–646, 2016.
14. T. Taleb et al., “On Multi-Access Edge Computing: A Survey of the Emerging 5G Network Edge Architecture and Orchestration,” IEEE Communications Surveys & Tutorials, vol. 19, no. 3, pp. 1657–1681, 2017.
15. N. Abbas et al., “Mobile Edge Computing: A Survey,” IEEE Internet of Things Journal, vol. 5, no. 1, pp. 450–465, 2018.
16. S. Deng et al., “Edge Intelligence: The Confluence of Edge Computing and Artificial Intelligence,” IEEE Internet of Things Journal, vol. 7, no. 8, pp. 7457–7469, 2020.
17. Y. Sun et al., “Application of Machine Learning in Wireless Networks: Key Techniques and Open Issues,” IEEE Communications Surveys & Tutorials, vol. 21, no. 4, pp. 3072–3108, 2019.
18. H. Wu et al., “Artificial Intelligence for Next-Generation Wireless Networks: A Survey,” China Communications, vol. 16, no. 2, pp. 124–142, 2019.
19. X. Foukas et al., “Network Slicing in 5G: Survey and Challenges,” IEEE Communications Magazine, vol. 55, no. 5, pp. 94–100, 2017.
20. M. Chen et al., “Artificial Intelligence in 6G Networks,” IEEE Network, vol. 38, no. 1, pp. 16–22, 2024.
21. H. Huang et al., “Deep Reinforcement Learning for Resource Allocation in V2V Communications,” IEEE Transactions on Vehicular Technology, vol. 68, no. 4, pp. 3163–3173, 2019.
22. J. Park et al., “Wireless Network Intelligence at the Edge,” Proceedings of the IEEE, vol. 107, no. 11, pp. 2204–2239, 2019.
23. P. Mach and Z. Becvar, “Mobile Edge Computing: A Survey on Architecture and Computation Offloading,” IEEE Communications Surveys & Tutorials, vol. 19, no. 3, pp. 1628–1656, 2017.
24. S. Barbarossa et al., “Communicating While Computing: Distributed Mobile Cloud Computing over 5G Heterogeneous Networks,” IEEE Signal Processing Magazine, vol. 31, no. 6, pp. 45–55, 2014.
25. T. X. Tran and D. Pompili, “Adaptive Bitrate Video Caching and Processing in Mobile-Edge Computing Networks,” IEEE Transactions on Mobile Computing, vol. 18, no. 9, pp. 1965–1978, 2019.
26. Q. Pham et al., “A Survey of Multi-Access Edge Computing in 5G and Beyond,” IEEE Access, vol. 8, pp. 116974–117017, 2020.
27. M. Chiang and T. Zhang, “Fog and IoT: An Overview of Research Opportunities,” IEEE Internet of Things Journal, vol. 3, no. 6, pp. 854–864, 2016.
28. X. Huang et al., “The Role of AI in 6G Wireless Networks,” IEEE Open Journal of Vehicular Technology, vol. 2, pp. 132–151, 2021.
29. Y. Liu et al., “Federated Learning for 6G Communications: Challenges, Methods, and Future Directions,” China Communications, vol. 17, no. 9, pp. 105–118, 2020.
30. N. H. Mahmood et al., “White Paper on Critical and Massive Machine Type Communication Towards 6G,” 6G Flagship, University of Oulu, Finland, 2020.
31. W. Saad et al., “A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems,” IEEE Network, vol. 34, no. 3, pp. 134–142, 2020.

32. T. S. Rappaport et al., "Wireless Communications and Applications Above 100 GHz: Opportunities and Challenges for 6G and Beyond," IEEE Access, vol. 7, pp. 78729–78757, 2019.
33. F. Tariq et al., "A Speculative Study on 6G," IEEE Wireless Communications, vol. 27, no. 4, pp. 118–125, 2020.
34. Z. Zhang et al., "6G Wireless Networks: Vision, Requirements, Architecture, and Key Technologies," IEEE Vehicular Technology Magazine, vol. 14, no. 3, pp. 28–41, 2019.
35. C. Huang et al., "Reconfigurable Intelligent Surfaces for Energy Efficiency in Wireless Communication," IEEE Transactions on Wireless Communications, vol. 18, no. 8, pp. 4157–4170, 2019.
36. S. Bubeck and B. Sellke, "A Universal Law of Robustness via Isoperimetry," Advances in Neural Information Processing Systems, vol. 34, pp. 28811–28822, 2021.
37. X. You et al., "Towards 6G Wireless Communication Networks: Vision, Enabling Technologies, and New Paradigm Shifts," Science China Information Sciences, vol. 64, no. 1, 2021.
38. H. Tataria et al., "6G Wireless Systems: Vision, Requirements, Challenges, Insights, and Opportunities," Proceedings of the IEEE, vol. 109, no. 7, pp. 1166–1199, 2021.
39. A. Dogra et al., "A Survey on Beyond 5G Network with the Advent of 6G: Architecture and Emerging Technologies," IEEE Access, vol. 9, pp. 67512–67547, 2021.
40. M. Jaber et al., "6G Communications: Vision and Potential Techniques," IEEE Network, vol. 34, no. 3, pp. 70–75, 2020.