AI-Driven QoS Optimization in 6G-Enabled IoT Networks: A Comprehensive Survey

Archana Chauhan[1]; Ranu Pandey[2]

[1]Shri Rawatpura Sarkar University, Raipur, India
[2]Shri Rawatpura Sarkar University, Raipur, India

Publication Date: 2026/05/07

Abstract: The blistering development of Internet of Things (IoT) applications, smart healthcare, industrial automation, autonomous systems, immersive communications, has put high demands on the Quality of Service (QoS), including ultra-low latency, high reliability, high connectivity, and energy efficiency. Though the fifth-generation (5G) networks have partially met these needs, they cannot be fully relied upon in terms of intelligence, scalability, and adaptability of the future IoT ecosystems. The sixth-generation (6G) wireless networks will be projected as being AI-native thus allowing autonomous, self-optimizing, and context-aware communication systems. In this survey, the author introduces a literature review of AI-driven methods of optimizing QoS of 6G-enabled IoT networks. We critically examine QoS needs of the emerging applications of IoT, important architectural characteristics of 6G networks, and application of artificial intelligence across various network layers. Research on different AI methods is discussed in terms of resource distribution, network slicing, traffic control, energy usage, and mobility. Moreover, this survey talks about widely utilized datasets, simulation, and performance assessment metrics applicable to AI-based optimization of QoS. Lastly, future directions and research challenges are brought to the fore and these are scalability, security, explainability, and real-world deployment issues. It is expected that this survey will offer researchers and practitioners a systematic review of the existing progress and will be used as a base to further research in the field of AI-enabled 6G IoT networks.

Keywords: *Artificial Intelligence (AI), 6G Wireless Networks, Internet of Things (IoT), Quality of Service (QoS), QoS Optimization, Machine Learning, Network Slicing, Edge Intelligence.*

How to Cite: Archana Chauhan; Ranu Pandey (2026) AI-Driven QoS Optimization in 6G-Enabled IoT Networks: A Comprehensive Survey. *International Journal of Innovative Science and Research Technology*, 11(4), 3550-3558.
https://doi.org/10.38124/ijisrt/26apr2075

I. INTRODUCTION

IoT has emerged to form the core of facilitating the applications of next-generation intelligent applications, such as smart healthcare, industrial automation, smart transportation systems, smart cities, and immersive multimedia services. The fast deployment of heterogeneous IoT devices and the geometric increase of data traffic, have created high Quality of Service (QoS) demands, including ultra-low latency, high reliability, massive connectivity, and energy efficiency, among others, in reference to ref5,ref22. The provision of steady QoS in the most dynamic network environments is one of the pressing problems of the modern wireless communication systems.

The new 5G wireless networks come with some of the critical technologies that are capable of meeting the requirements of various IoT services including improved mobile broadband, ultra-reliable low-latency communications, and massive machine-type communications [18, 26]. Nevertheless, the growing complexity, size, and intelligence of future IoT applications have revealed inherent shortcomings of 5G networks, especially in the areas of flexibility, real-time decision demands, and autonomous optimization remedies problems of adaptability, real-time decision-making, and autonomous optimization, especially in areas of 5G network services. The challenges have pushed the research community to visualize the sixth-generation (6G) wireless networks as a new dawn paradigm that can provide sub-milliseconds latency, very high reliability, terabit-per-second data rate, and large numbers of devices connecting to it [1, 3, 17].

In contrast to the past, the 6G networks will be AI-native, that is, AI is profoundly embedded into network design, control and management processes [4, 30]. With AI-powered networking, networking is able to self-optimize, self-heal, and respond to dynamic environments, providing data powered decisions, and those that are proactive. In the IoT, AI is crucial to the optimization of QoS due to the ability to allocate resources intelligently, dynamically slice networks, predict traffic, manage mobility, and optimize energy use through scheduling tasks intelligently [6, 7]. These functionalities are especially crucial in dense and heterogeneous IoT systems that are projected in 6G systems.

Recent developments of machine learning, deep learning as well as reinforcement learning have shown considerable promise in solving the challenging QoS optimization issues in a wireless network setting [8, 9]. Moreover, new paradigms include mobile edge computing, fog computing, and edge intelligence that allow real-time AI inference nearer to the IoT devices, reducing the latency, and enhancing the quality of service delivery performance [10, 11, 19]. Another privacy-sensitive and communication-efficient technique that has been considered by federated learning is the optimization of distributed QoS in large-scale IoT networks using the federated learning approach to facilitate privacy protection in this context as well. Another application of the federated learning approach with respect to large-scale IoT networking is the optimization of distributed QoS in the context of privacy protection offered through the federated learning approach to the large-scale IoT communication and networking application, too.

In spite of these developments, the current research activities remain more or less fragmented and focusing on 5G systems or individual optimization problems. The majority of surveys that are available are focused on generic AI-enabled wireless network or QoS control in IoT and pay no attention to the specifics of 6G networks architecture and performance needs and demands [20, 27]. The new 6G technologies, including intelligent network slicing, terahertz communications, and networking based on digital twins present new challenges and opportunities

of AI-powered QoS optimization that are not clearly addressed in the literature yet [12, 24, 25].

These observations inspired this survey to identify a complete and systematic review of the use of AI in optimization of QoS in 6G-enabled IoT networks. This survey made its primary contributions as follows:

- The analysis of QoS demands and the challenges related to emerging IoT application of 6G networks.
- A broad literature review of the AI methods to perform QoS optimization, such as machine learning, deep learning, reinforcement learning, and federated learning methods.
- A summary of 6G architectural enablers, edge intelligence architectures, datasets and performance metrics of importance in AI-based QoS optimization.
- Open research challenges and future directions: A discussion of the challenges of open research and future directions, including scalability, energy efficiency, security, explainability, and real-world deployment.

The rest of this paper is structured in the following way.

In the second section of the paper, Literature Review: AI Techniques in QoS Optimization, the review of AI techniques used in QoS optimization in IoT and wireless networks is described in detail. Section III, Methodology and 6G Architectural Frameworks, details AI-based 6G network architectures and methodological frameworks to optimize QoS. The fourth section, Data sets and performance evaluation metrics, recaps the widely known datasets, simulation tools and performance evaluation measures. Section V, Open Research Challenges and Future Directions, provides an overview of issues that remain open and opportunities in research. Lastly, the paper has a conclusion, called VI, in Section VI.

II. LITERATURE REVIEW: AI TECHNIQUES FOR QOS OPTIMIZATION

The adoption of the artificial intelligence concept to the wireless communication systems has become one of the most interesting research topics due to its potential to improve the complexity of the Quality of Service (QoS) management in the realization of IoT networks. The initial research on QoS provisioning in IoT was mainly concerned with rule based mechanisms and conventional optimization techniques which were found inadequate to manage large-scale, heterogeneous and dynamic networks. With the growth of the IoT deployments, researchers started investigating the data-driven methods to enhance adaptability and scalability.

QoS optimization in wireless and IoT networks has been broadly studied with the use of machine learning (ML) methods. One of the first accounts on the relevance of ML paradigms to next-generation wireless networks was offered by Jiang et al. 6 who reported the possibilities of the technology in resource allocation, prediction of traffic, and mitigating interference. On the same note, Sun et al. (2018) reviewed the most significant ML methods and discovered the gaps in data access, the complexity of models, and real-time learning in the wireless system. Such works formed the basis of the use of AI-based solutions to the QoS management challenges.

As deep learning developed, scientists have shown better results at the modeling of Operating complex and nonlinear correlation that wireless networks exhibit. In ref. 9, Kato et al. suggested deep learning vision of traffic control in heterogeneous networks with focus on proactive congestion avoidance and QoS-aware congestion control. Deep reinforcement learning (DRL) also made it possible to optimize autonomously as proved by Ye, et al. [8] who used DRL to design dynamic resource scheduling of vehicles in vehicular communication networks, which resulted in improved latency and reliability metrics.

The concept of edge computing and fog computing has contributed greatly to the AI-based QoS optimization. Mach and Becvar [20] and Mao et al. emphasized the importance of mobile edge computing in the minimization of the latency and real-time intelligence near the Internet of Things devices [27]. Wang, et al, also conducted a survey of edge computing and deep learning convergence and proved that it was effective in applications of latency-sensitive and QoS-critical. These methods are especially applicable to 6G-enabled IoT networks, in which the minimal latency and reliability are required.

Federated learning has proven to be a potentially viable solution to dis-tributed QoS optimization, and privacy of data. According to Samarakoon et al. [13], federated learning was tested to achieve ultra-reliable and low-latency communications demonstrating that it can decrease communication overhead and guarantee sensitive information protection. Shi et al. also discussed the efficiency of communication in the federated learning models, which is essential in the implementation of large-scale IoT solutions [21]. These papers identify federated learning as an important facilitator of scalable and privacy-conscious AI-driven optimization of QoS of future networks.

With advances in research into beyond 5G and 6G systems, the role of AI-native network architectures in a number of studies was highlighted. The vision, requirements, and challenges of 6G networks were defined by Zhang et al. [1], Akyildiz et al. [3], and Tataria et al. [17], and AI-based automation was described as one of the key design principles. In their article, Latif et al.

[30] offered a resourceful survey of AI-enabled 6G networks, emphasizing such research directions as intelligent resource management and an intelligent service provisioning.

Other recent papers have considered more advanced 6G technologies, including intelligent network slicing, digital twins and terahertz communications to optimize QoS. Zhang et al. [12] explored the application of AI-based network slicing to the 6G systems, which allow service-specific QoS provisions. Zhang et al. [24] surveyed digital twin-based networking which proved to be useful in predictive optimization of QoS and in network planning. Moreover, Elayan, etc.

Noted that terahertz communication is an essential enabler of ultra-high data rates in 6G, which brings about new challenges of propagation and reliability optimization of QoS.

Though the literature is quite extensive, the available research is still discontinuous as it tends to be concentrated on a particular AI method, or a single QoS parameter, or a single network situation. Majority of surveys refer to generic AI-supported wireless networks or QoS management in 5G systems, but they do not give a detailed analysis in accordance with the peculiarities of 6G-enabled IoT networks. The given gap encourages a systematic and holistic survey that unites AI-based methods of QoS optimization, architectural enablers, datasets, and open research challenges in the 6G IoT system context.

III. METHODOLOGY AND 6G ARCHITECTURAL FRAMEWORKS

This section reviews the methodological approaches and architectural frame-works proposed in the literature for enabling AI-driven Quality of Service (QoS) optimization in 6G-enabled IoT networks. Unlike conventional networks, 6G systems are envisioned as AI-native, where intelligence is embedded across all network layers to enable autonomous decision-making, proactive optimization, and self-adaptation [1, 3, 17].

➢ *AI-Native 6G Network Architecture*

A number of researches have emphasized that there is the necessity of paradigm shift in managing networks through centralized and static networks into de-centralized and smart networks. Zhang and colleagues [1] and Akyildiz and colleagues [3] have come up with AI-native 6G architecture where the artificial intelligence has been closely interwoven with the radio access network, core net-work, and service management layers. In this type of architecture, AI agents constantly monitor the network conditions, learn about it by past experiences, and improve the parameters of QoS, including latency, throughput, and reliability in real-time.

Tataria et al. [17] also pointed out the significance of cross-layer intelligence in 6G networks, which will allow the simultaneous optimization of the physical layer, MAC layer, network layer, and application layer. This cross-layer design enables the AI models to learn intricate connections among religious parameters of the network that is crucial to providing heterogeneous IoT services with a variety of QoS needs.

➢ *Edge Intelligence and Distributed Learning*

Edge computing is also an important attribute of 6G architecture as it facilitates low latency and context-based optimization of QoS. Mach and Becvar [20] and Mao et al. [27] showed that locating computational intelligence at the network edge could significantly lower the communication latency and backhaul congestion. The convergence of deep learning and edge computing was surveyed by Wang et al. [10], who note that the technique is particularly useful in real-time to manage the latency-sensitive IoT-based QoS.

Federated learning is a novel distributed learning approach that has been proposed to overcome the issue of scalability and privacy to optimize the QoS. Federated learning-based frameworks of ultra-reliable and low-latency communications were recently suggested by Samarakoon et al. [13]: they allow joint model training without the necessity of exchanging raw data. Shi et al. further enhanced the efficiency of communication in federated learning by Shi et al. and hence it is applicable to large scale 6G IoT application.

➢ *Network Slicing and Resource Orchestration*

Network slicing is also considered an important architectural enabler of 6G networks differentiation of QoS.

Table 1: Methodological Approaches Used in Reviewed Literature

Ref.	**AI / Analytical Methodology**	**Network Architecture / Framework**	**Target QoS Objectives**
[1]	Conceptual AI-driven optimization	AI-native 6G architecture	Latency, reliability, scalability
[3]	Intelligent framework automation	Beyond-5G / 6G network model	Ultra-low latency, high data rate
[5]	Analytical QoS modeling	IoT-centric network frame-work	Latency, ciency energy effi-
[6]	Machine learning–based optimization	Next-generation wireless networks	Throughput, spectral efficiency
[7]	Supervised and unsupervised learning	Wireless traffic management system	Delay, packet loss
[8]	Deep reinforcement learning	Vehicular communication networks	Latency, reliability
[9]	Deep learning–based traffic control	Heterogeneous wireless net-works	Congestion, QoS stability
[10]	Deep learning at the edge	Edge computing framework	Latency, computation overhead
[12]	AI-enabled network slicing	Virtualized 6G architecture	Throughput, slice isolation
[13]	Federated learning	Distributed wireless net-works	Reliability, latency
[15]	Traffic ML prediction using	SDN-enabled IoT networks	Delay, channel utilization
[17]	System-level analytical modeling	6G conceptual framework	Latency, data rate, reliability
[19]	Edge intelligence frame-work	Wireless edge networks	QoS adaptability, tency la-
[20]	Computation offloading methods	Mobile edge computing architecture	Task delay, resource us-age
[21]	Communication-efficient federated learning	Edge intelligence frame-work	Convergence speed, overhead
[24]	Digital twin–assisted optimization	Digital twin-enabled net-works	QoS prediction accuracy

[25]	AI-assisted physical-layer adaptation	Terahertz communication systems	Throughput, reliability
[29]	Risk-aware optimization models	URLLC frameworks	Packet delivery, delay
[30]	AI-centric network intelligence	AI-empowered 6G networks	Latency, energy, throughput

Zhang *et al.* [12] explored AI-driven network slicing frameworks that dynamically allocate resources to heterogeneous IoT services based on their QoS requirements.

With the help of machine learning and reinforcement learning, the frame-works will allow the adaptive orchestration of resources so that service-level agreements could be ensured under different traffic conditions.

Software-defined networking (SDN) and network function virtualization (NFV) architectures have also been developed into traffic prediction and load balancing through AI. Tang et al. have shown that smart traffic load forecasting enhances channel allocation and QoS in SDN-based IoT networks, and this can be directly applied to 6G-based systems in the future, as well [15].

➢ *Digital Twin-Enabled Network Management*

Digital twin technology has become a prospective approach to predictive optimization of QoS in 6G networks. Zhang et al. surveyed digital twins networks with smart networks in which virtual representations of the physical network elements are applied to simulate, analyze, and predict network behavior [24]. Through the combination of AI models and digital twins, network operators may actively optimise the QoS parameters, anomalie and control strategy assessment prior to deployment.

➢ *Emerging Communication Technologies*

New physical-layer technologies also have a further impact on the architecture of 6G networks. Elayan et al. pointed out terahertz (THz) communication as one of the enabling factors to ultra-high data rates and presented issues to do with propagation loss, reliability, and blockage [25]. Adaptive modulation, beamforming, and power control methods implemented with the use of the AI are necessary to ensure that QoS remains high in the THz based IoT communications.

In general, such methodological approaches and architectural designs indicate that the concept of AI-based QoS optimization of 6G-enabled IoT networks is based on the close collaboration of edge intelligence, distributed learning, net-work slicing, digital twins, and novel communication technologies. All of the components comprise the basis of autonomous, scalable, and QoS-aware 6G IoT systems.

IV. DATASETS AND PERFORMANCE EVALUATION METRICS

This part summarises popular datasets, simulation environments, and performance measurement indicators that had been utilised in the literature on evaluating AI-based Quality of Service (QoS) optimization in the IoT and next-generation wireless networks. Most current studies are based on a mix of a synthetic data generation and evaluation as well as on public datasets because there were not yet large-scale real-world 6G deployments [1, 3, 17].

➢ *Datasets for AI-Driven QoS Optimization*

There are still limited publicly available datasets of 6G-enabled IoT networks. As a result, network trace, IoT sensor and synthetic datasets are frequently used to train and evaluate AI models. Traffic recordings of packet arrival rates, throughput, latency and congestion pattern are typically used in QoS prediction and traffic classification exercises, both of which are employed in traffic congestion management tasks, as well as in traffic congestion forecasting tasks [7, 9]. The datasets of IoT sensors are often employed to model heterogeneous behavior of devices, patterns of data generation, and energy consumption properties of devices [5].

Besides real-life data, the use of synthetic data generated by network simulators is a universal practice to simulate 6G conditions. Paper like on-[12, 15] use simulation based datasets to simulate dynamic network slicing, resource al-location as well as changes in traffic loads and their changes under various QoS constraints. Digital twin methods also facilitate larger-scale and realistic generation of data, generating virtual copies of real-life networks to simulate data to train and validate AI models and systems more realistically and at scale, allowing the generation of large-scale datasets that are both realistic and ac-curate to train and evaluate AI models and systems with neural networks and reinforcement learning algorithms, respectively.

➢ *Simulation Platforms and Testbeds*

Simulation platforms are important in testing the AI-based QoS optimization methods. The model is typically simulated using network simulators like NS-3 and OMNeT++, which can analyze latency, throughput, and packet delivery performances in detail due to the possibility to model wireless networks and IoT networks. Doing these platforms are commonly combined with machine learning systems to enable closed-loop learning and control.

Simulation environments based on MATLAB are also extensively applicable in modelling physical-layer and system-level behaviour of 5G and beyond-5G networks, such as spectrum allocation, beamforming and channel modelling [26]. Recent research uses edge computing testbeds and virtualized settings to test AI models in realistic latency and resource constraints settings with proper latency and resources available to them through virtualized environments and testbeds. Nevertheless, the absence of standardized 6G testbeds is still a major problem with regard to reproducible assessment.

- ***Performance Evaluation Metrics***

The effectiveness of AI-based methods of QoS optimization is evaluated over a set of diverse metrics that indicate both service-level and network-level goals.

Delay-sensitive IoT applications mostly use latency and jitter as key metrics, especially when it comes to ultra-reliable and low-latency communication applications .

Such measures as the ratio of packet delivery and the probability of outage are metric reliability measures that are commonly used to evaluate the strength of AI-based optimization solutions [8]. IoT devices that have limited power availability such as energy efficiency measurements such as the consumption per transmitted bit and network lifetime are important in the scale of battery-constrained devices [14]. Also, Quality of Experience (QoE) measures are be-ginning to be regarded to characterise user-centric performance in addition to the conventional QoS measures. [23].

Table 2: Datasets and Data Sources Used in Reviewed Literature

Ref.	Dataset / Data Source	Application Context	Key QoS Metrics
[5]	IoT sensor data (pub-lic/synthetic)	QoS modeling in IoT networks	Latency, reliability, energy
[6]	Simulated wireless network data	ML-based resource allocation	Throughput, spectral efficiency
[7]	Network traffic traces	QoS prediction and traffic classification	Delay, packet loss
[8]	Simulated V2V datasets	DRL-based resource allocation	Latency, reliability
[9]	Heterogeneous traffic data	Deep learning traffic control	Congestion, delay
[10]	Edge computing simulation data	Edge intelligence evaluation	Latency, computation cost
[12]	Synthetic network slicing data	AI-driven network slicing	Throughput, slice isolation
[13]	Distributed wireless datasets	Federated learning for URLLC	Latency, communication cost
[14]	Energy consumption datasets	Energy-efficient communications	Energy efficiency
[15]	SDN-based IoT traffic data	Intelligent traffic prediction	Delay, channel utilization
[17]	Simulated 6G scenarios	6G requirement analysis	Latency, data rate
[24]	Digital twin–generated data	Predictive network optimization	QoS prediction accuracy
[25]	THz channel simulation data	Terahertz communication analysis	Throughput, reliability
[27]	MEC simulation datasets	MEC-based QoS optimization	Latency, resource usage
[29]	URLLC simulation data	Ultra-reliable communications	Packet delivery ratio
[30]	Multi-source synthetic datasets	AI-empowered 6G evaluation	Latency, energy, throughput

Convergence speed, training overhead, model accuracy, and communication cost are learning-related metrics that are typically mentioned to indicate the feasibility of AI solutions in large-scale IoT deployments (AI) (ref13,ref21). These measures are of special concern to federated and edge learning systems, in which resource limitations and communication costs have a direct effect on the quality of service.

Altogether, the datasets and metrics of performance analysis reviewed indicate that the standardization of benchmarks, realistic datasets, and single evaluation frameworks preconditioned by 6G-enabled IoT networks are necessary. These are the issues that need to be tackled to allow making fair comparisons, ensure reproducibility and practical applications of AI-based QoS optimization methods.

V. OPEN RESEARCH CHALLENGES AND FUTURE DIRECTIONS

Although there is a huge advancement in artificial intelligence (AI)-based 6G and IoT-based communication systems, various unanswered research questions still exist, and they need to be answered to achieve the ultimate goal of fully autonomous, robust, and scalable 6G next-generation networks.

➢ *Scalability and Heterogeneity Management*

The next generation of the IoT networks, which will be 6G, will accommodate a very huge number of devices, their traffic patterns, requirements, and hardware performance will be very heterogeneous. Current AI models are not always able to perform well at scale when deployed to ultra-dense and heterogeneous environments. The ability to design lightweight, adaptive and scalable learning models that are capable of generalizing to a variety of network conditions is also a challenge [1, 3, 30].

➢ *Real-Time Intelligence and Latency Constraints*

Several AI-based solutions use centralized or cloud-based processing and this can add unacceptable latencies to ultra-reliable low-latency communication (URLLC) services. To attain real-time intelligence and also satisfy strict latency requirements, it is necessary to conduct more research on edge-native learning, low-complexity inference models, and fast convergence algorithms that can be used in time-critical applications [8, 10, 19].

➢ *Data Availability, Quality, and Privacy*

AI-based 6G systems depend heavily on large volumes of high-quality data for training and optimization. However, data scarcity, non-IID data distribution, and privacy concerns limit the effectiveness of centralized learning approaches. Federated and privacy-preserving learning techniques show promise, but challenges related to communication overhead, convergence stability, and security vulnerabilities remain open [13, 21].

➢ *Energy Efficiency and Sustainable Intelligence*

The integration of AI into 6G networks significantly increases computational and energy demands, particularly at edge devices with limited resources. Balancing intelligence with energy efficiency is a critical research issue. Future work must explore green AI models, energy-aware learning strategies, and hardware-efficient architectures to support sustainable 6G deployments [5, 20, 30].

➢ *Robustness, Reliability, and Trustworthiness of AI Models*

AI-driven network management systems must operate reliably under dynamic and unpredictable conditions, including channel variations, traffic surges, and adversarial attacks. Ensuring robustness, interpretability, and trustworthiness of AI models is essential for mission-critical applications. Research on explain-able AI, fault-tolerant learning, and adversarial resilience is still in its early stages [9, 29].

➢ *Integration of Emerging 6G Technologies*

Future 6G networks will incorporate advanced technologies such as terahertz communications, intelligent reflecting surfaces, digital twins, and integrated sensing and communication. Seamlessly integrating AI-driven optimization with these emerging technologies poses architectural and algorithmic challenges, particularly in terms of coordination, standardization, and real-time control [24, 25].

➢ *Future Research Directions*

To address the aforementioned challenges, future research should focus on developing distributed and collaborative intelligence frameworks, cross-layer optimization techniques, and unified AI-native network architectures. Additionally, standardized datasets, realistic simulation platforms, and real-world testbeds are required to enable fair performance evaluation and accelerate the transition from theoretical models to practical 6G deployments.

VI. CONCLUSION

This survey presented a comprehensive review of artificial intelligence (AI)-driven quality of service (QoS) optimization techniques for 6G-enabled Internet of Things (IoT) networks. By systematically analyzing the evolving QoS requirements, AI-based optimization strategies, and emerging 6G architectural frameworks, this paper highlighted how intelligent networking paradigms are reshaping next-generation wireless communication systems.

The review demonstrated that machine learning, deep learning, reinforcement learning, and federated learning have become key enablers for addressing complex QoS challenges such as latency reduction, reliability enhancement, energy efficiency, and resource management in ultra-dense and heterogeneous IoT environments. Furthermore, the analysis of AI-enabled 6G architectures emphasized the growing role of edge intelligence, network slicing, and cross-layer optimization in supporting diverse service requirements.

In addition, this survey summarized commonly used datasets and performance evaluation metrics, revealing a lack of standardized benchmarks and real-world datasets tailored to 6G IoT scenarios. This gap underscores the need for unified evaluation frameworks to ensure fair comparison and reproducibility of future research outcomes.

Despite notable advancements, several open challenges persist, including scalability, data privacy, energy-efficient intelligence, robustness of AI models, and the integration of emerging 6G technologies such as terahertz communication and intelligent reflecting surfaces. Addressing these challenges will require collaborative intelligence frameworks, AI-native network designs, and closer alignment between theoretical research and practical deployments.

Overall, this survey provides a structured reference for researchers and practitioners by consolidating recent advances, identifying research gaps, and out-lining promising future directions. It is anticipated that the insights presented in this work will facilitate the development of intelligent, reliable, and scalable QoS optimization solutions for next-generation 6G-enabled IoT networks.

REFERENCES

[1]. Z. Zhang, Y. Xiao, Z. Ma, et al., “6G Wireless Networks: Vision, Requirements, Architecture, and Key Technologies,” *IEEE Vehicular Technology Magazine*, vol. 14, no. 3, pp. 28–41, Sep. 2019.
[2]. M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, “Disease Prediction by Machine Learning over Big Data from Healthcare Communities,” *IEEE Access*, vol. 5, pp. 8869–8879, 2017.
[3]. I. F. Akyildiz, A. Kak, and S. Nie, “6G and Beyond: The Future of Wireless Communications Systems,” *IEEE Access*, vol. 8, pp. 133995–134030, 2020.
[4]. S. Dang, O. Amin, B. Shihada, and M.-S. Alouini, “What Should 6G Be?,” *Nature Electronics*, vol. 3, pp. 20–29, 2020.
[5]. M. Tatipamula, D. A. Medhi, and B. D. Davie, “Quality of Service (QoS) in Internet of Things,” *Computer Networks*, vol. 128, pp. 172–185, 2017.
[6]. C. Jiang, H. Zhang, Y. Ren, Z. Han, K.-C. Chen, and L. Hanzo, “Ma-chine Learning Paradigms for Next-Generation Wireless Networks,” *IEEE Wireless Communications*, vol. 24, no. 2, pp. 98–105, Apr. 2017.
[7]. Y. Sun, M. Peng, Y. Zhou, Y. Huang, and S. Mao, “Application of Machine Learning in Wireless Networks: Key Techniques and Open Issues,” *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3072–3108, 2019.
[8]. H. Ye, G. Y. Li, and B.-H. Juang, “Deep Reinforcement Learning Based Resource Allocation for V2V Communications,” *IEEE Transactions on Vehicular Technology*, vol. 68, no. 4, pp. 3163–3173, 2019.
[9]. N. Kato, Z. M. Fadlullah, B. Mao, F. Tang, and O. Akashi, “The Deep Learning Vision for Heterogeneous Network Traffic Control: Proposal, Challenges, and Future Perspective,” *IEEE Wireless Communications*, vol. 24, no. 3, pp. 146–153, Jun. 2017.
[10]. X. Wang, Y. Han, V. Leung, et al., “Convergence of Edge Computing and Deep Learning: A Comprehensive Survey,” *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 869–904, 2020.
[11]. M. Chiang, S. Ha, C.-L. I, and F. Risso, “Clarifying Fog Computing and Networking: 10 Questions and Answers,” *IEEE Communications Maga-zine*, vol. 55, no. 4, pp. 18–20, Apr. 2017.
[12]. Q. Zhang, Q. Chen, and Y. Li, “Intelligent Network Slicing for 6G Net-works: Architecture and Challenges,” *IEEE Network*, vol. 35, no. 4, pp. 203–209, 2021.
[13]. S. Samarakoon, M. Bennis, W. Saad, and M. Debbah, “Federated Learning for Ultra-Reliable Low-Latency V2V Communications,” *IEEE Transactions on Communications*, vol. 68, no. 2, pp. 1146–1159, 2020.
[14]. K. Yang, S. Martin, and G. Zheng, “Energy-Efficient Wireless Communications: Tutorial, Survey, and Open Issues,” *IEEE Communications Surveys & Tutorials*, vol. 17, no. 3, pp. 1508–1529, 2015.
[15]. F. Tang, Z. M. Fadlullah, B. Mao, and N. Kato, “An Intelligent Traffic Load Prediction-Based Adaptive Channel Assignment Algorithm in SDN-IoT,” *IEEE Internet of Things Journal*, vol. 5, no. 6, pp. 5141–5154, 2018.
[16]. Y. Liu, K. Wang, Y. Lin, and W. Xu, “Lightweight Deep Learning Models for Resource Allocation in 6G IoT Networks,” *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 123–135, 2022.
[17]. H. Tataria et al., “6G Wireless Systems: Vision, Requirements, Challenges, Insights, and Opportunities,” *Proceedings of the IEEE*, vol. 109, no. 7, pp. 1166–1199, Jul. 2021.

[18]. A. Gupta and R. K. Jha, “A Survey of 5G Network: Architecture and Emerging Technologies,” *IEEE Access*, vol. 3, pp. 1206–1232, 2015.

[19]. J. Park, S. Samarakoon, M. Bennis, and M. Debbah, “Wireless Network Intelligence at the Edge,” *Proceedings of the IEEE*, vol. 107, no. 11, pp. 2204–2239, Nov. 2019.

[20]. 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.

[21]. Y. Shi, K. Yang, T. Jiang, et al., “Communication-Efficient Federated Learning for Wireless Edge Intelligence,” *IEEE Communications Maga-zine*, vol. 58, no. 12, pp. 70–76, Dec. 2020.

[22]. A. Yadav and O. Kaiwartya, “AI-Enabled QoS-Aware Routing in IoT Net-works: A Survey,” *Computer Communications*, vol. 176, pp. 154–173, 2021.

[23]. R. Li, Z. Zhao, X. Zhou, et al., “Intelligent 5G: When Cellular Networks Meet Artificial Intelligence,” *IEEE Wireless Communications*, vol. 24, no. 5, pp. 175–183, Oct. 2017.

[24]. S. Zhang, P. Yang, and J. Zhang, “Digital Twin-Enabled Smart Networks: A Survey,” *IEEE Communications Surveys & Tutorials*, vol. 24, no. 2, pp. 809–848, 2022.

[25]. H. Elayan, O. Amin, B. Shihada, and M.-S. Alouini, “Terahertz Band: The Last Piece of RF Spectrum Puzzle for Communication Systems,” *IEEE Open Journal of the Communications Society*, vol. 1, pp. 1–32, 2020.

[26]. J. G. Andrews et al., “What Will 5G Be?,” *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, 2014.

[27]. Y. Mao, C. You, J. Zhang, et al., “A Survey on Mobile Edge Computing: The Communication Perspective,” *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2322–2358, 2017.

[28]. X. Xu, H. Sun, and L. Qian, “AI-Driven Network Slicing for 6G IoT Net-works,” *IEEE Network*, vol. 36, no. 1, pp. 110–117, 2022.

[29]. M. Bennis, M. Debbah, and H. V. Poor, “Ultrareliable and Low-Latency Wireless Communication: Tail, Risk, and Scale,” *Proceedings of the IEEE*, vol. 106, no. 10, pp. 1834–1853, Oct. 2018.

[30]. S. Latif, Z. Idrees, J. Ahmad, et al., “AI-Empowered 6G Networks: A Survey,” *IEEE Access*, vol. 9, pp. 155002–155022, 2021.