Global Journal of Engineering and Technology Advances

eISSN: 2582-5003

Cross Ref DOI: 10.30574/gjeta

Journal homepage: https://gjeta.com/

(REVIEW ARTICLE)

Network system optimization using machine learning analysis: A systematic review

Godwin Daniel Essien [1, *], Philip Michael Asuquo [2], Chikezie Samuel Aneke [2], Bliss Utibe-Abasi. Stephen [2], Etimbuk Emmanuel Abraham [3] and Bolanle Eunice Oduleye [1], Martha Chikodili Orazulume [3], and Somkelechi Emmanuel Ubalatu [4]

[1] *Department of Computer Engineering, Topfaith University, Mkpatak, NIGERIA.*
[2] *Department of Computer Engineering, University of Uyo, NIGERIA.*
[3] *Department of Electrical/Electronic Engineering, Topfaith University, Mkpatak, NIGERIA.*
[4] *Department of Information Technology, Federal University of Technology, Owerri, NIGERIA.*

Global Journal of Engineering and Technology Advances, 2026, 26(02), 048-061

Publication history: Received on 18 December 2025; revised on 01 February 2026; accepted on 04 February 2026

Article DOI: https://doi.org/10.30574/gjeta.2026.26.2.0022

Abstract

The contemporary digital infrastructure is based on network systems, which serve as the foundation of communication, data transfer, and service provision in various fields of application. As the network traffic, heterogeneous architectures, and quality-of-service (QoS) requirements have grown exponentially, the conventional methods of optimization have become progressively ineffective because of their lack of dynamism and computational complexity. Machine Learning (ML) has become a disruptive method of optimization of network systems as it allows making decisions that are data-driven, flexible, and intelligent. The current paper includes a systematic review of the newest developments in streamlining network systems with the help of machine learning. The review takes a systematic approach that implies identification, screening and analysis of peer-reviewed journal and conference articles published between 2021 and 2025. Supervised learning, unsupervised learning, reinforcement learning, and deep learning are considered in connection with network performance optimization aims like traffic management, routing, congestion control, resource allocation, fault prediction and energy efficiency. The findings suggest that the recent research is now dominated by reinforcement learning and deep learning models because of their ability to manage complex, dynamic, and large network environments. Nonetheless, the limitations of the current models, like data scarcity, model interpretability, scalability, and real-time deployment, are still major obstacles to their practical application. The article finishes with the future directions of research, systems such as explainable AI, federated learning, and hybrid optimization structures to increase the strength and applicability of machine learning-based network optimization systems.

Keywords: Network Optimization; Machine Learning; Deep Learning; Reinforcement Learning; Intelligent Networks; Systematic Review

1. Introduction

The accelerated speed of the development of communication technologies, cloud computing, Internet of Things (IoT), and the fifth and sixth (5G/6G) wireless networks has made modern network systems very complex. Such systems must be able to provide high performance, low latency, reliability, scalability, and energy efficiency in dynamic and unpredictable conditions. The classic methods of network optimization that were based on mathematical modeling, heuristics, and fixed configurations can hardly withstand these challenges anymore [1], [2].

Network system optimization is the process of enhancing network performance measures like throughput, latency, packet loss, energy consumption, and resource utilization. The traditional methods tend to presume deterministic traffic and an exact system model that are impractical in modern networks where a variety of systems coexist, and everything is uncertain [3]. There has thus been an increasing interest in using machine learning (ML) methods to allow intelligent, adaptive, and self-optimizing networks.

* Corresponding author: Godwin Daniel Essien

Machine learning offers the ability to acquire patterns based on historical and live data, forecast network behavior, and make the best decisions without being programmed. The recent developments in deep learning and reinforcement learning have additionally increased the applicability of ML in large, complex network settings [4]. Consequently, network optimization with the help of ML has established itself as one of the primary research topics in academic and industrial domains.

Although the literature is increasing, there is a limited systematic review that is both thorough and recent and centered on optimization of network systems, based on the use of machine learning. This gap is filled in this paper through synthesizing recent research findings, classifying research questions and methods of ML, performance trends, and the discussion of current issues.

This review has made the following principal contributions:

- The machine learning methods that have been used systematically to optimize the network.
- Discussion of the implementation issues and gaps in research.
- Determining future research directions.

2. Methodology

This part gives the systematic approach that has been followed in the review of available literature on the optimization of network systems using machine learning analysis. The methodological framework was developed to be rigorous, transparent, reproducible, and able to cover all the pertinent studies, which is consistent with the best practices of systematic review in engineering and computer science research areas. The methodology includes review design, data sources and search strategy, inclusion and exclusion criteria, article selection procedure, and data extraction and synthesis procedures.

2.1. Review Design

The present study employs a Systematic Literature Review (SLR) methodology to examine the use of machine learning methods in the optimization of the network system. In comparison to a traditional narrative review, a systematic review has a set protocol, which is clear, transparent, and reproducible, so the bias is minimized, and the methodology is rigorous [5]. This study is especially appropriate to consider the SLR methodology because the number of research outputs in the machine learning-based network optimization is rapidly developing and diverse.

The review design is based on the existing guidelines of SLR that are actively applied to the research in the engineering and computer science domains, such as the Kitchenham framework and PRISMA-based review procedures tailored to technical research [5], [7]. These guidelines note down systematic searching, critical appraisal, as well as synthesis of findings, with emphasis on structured planning.

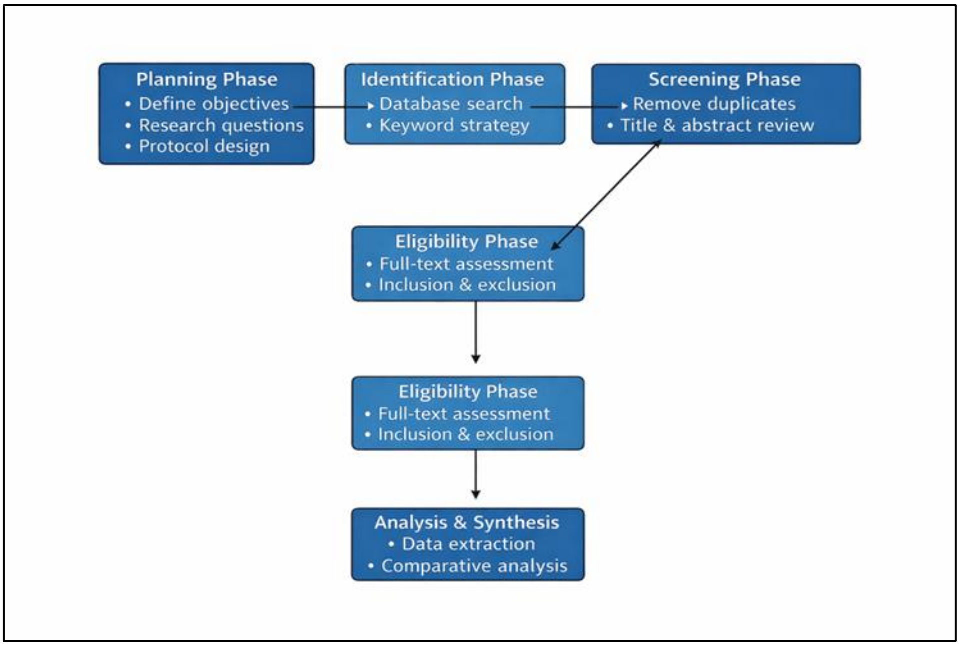
The following stages were followed in sequence in the review design:

- Planning Phase- definition of research scope, objectives, and review questions.
- Identification Phase- active search of databases containing relevant information through preset keywords.
- Screening Phase: the deletion of duplicates and irrelevant studies using titles and abstracts.
- Eligibility Phase- full-text evaluation as per inclusion and exclusion criteria.
- Analysis and Synthesis Phase - analysis and comparative analysis of selected studies, qualitative.

The fundamental research questions that will drive this review are:

- Which machine learning methods do you think are most popular in network system optimization?
- What are the network optimization goals that are solved with the use of ML?
- Which are the performance metrics applied to the ML-based optimization methods?
- What are the latest studies that report any challenges and limitations?

The use of these questions in organizing the review provides the study with the assurance that the objectives, methods, and results of the analysis are aligned [1], [4]. Figure 1 depicts the general systematic review structure used in this study, with the steps starting with planning through to synthesis. The framework emphasizes the repetitive character of the screening and eligibility review to include only high-quality and relevant studies



Source [5].

Figure 1 Systematic Review Methodology Framework

2.2. Search strategy and Data Sources.

To find the literature on network system optimization with the help of machine learning methods, a deep and systematic search strategy was utilized. Several reliable online libraries were chosen to cover as many high-quality peer-reviewed research articles as possible. It is also well known that these databases are used to publish trusted research in fields of networking, artificial intelligence, and engineering.

2.2.1. Data Sources

These were the subsequent electronic databases:

- IEEE Xplore - to access the latest research in the field of communications, networking, and machine learning.
- ScienceDirect (Elsevier) - to engineering and computer science across the board journals.
- SpringerLink - intelligent systems and optimization, theoretical and applied research.
- ACM Digital Library - of conference proceedings and journals in the field of computing systems.
- Google Scholar - to verify and include the most cited works.

All these sources are sufficient to cover both theoretical and applied research on the topic of this review [1], [2], [6].

2.2.2. Keyword Search and Formulation of Query.

The search strategy entailed the use of Boolean operators, a combination of keywords, and the expansion of synonyms to get maximum accuracy of retrieval. The main keywords were:

- Network optimization
- Machine learning
- Deep learning
- Reinforcement learning
- Intelligent networks
- Resource allocation
- Traffic management

Search query in the IEEE Xplore database:

(`network optimization` AND machine learning) OR (deep learning AND network control) OR (reinforcement learning AND routing).

In order to be relevant and up-to-date, the search was limited to the articles published in 2021-25, which portray the latest developments in the area. As they go through stringent quality control measures [3], [8], only peer-reviewed journal articles and conference papers were considered. Table 1 contains a summary of the digital libraries and the breadth of the literature search carried out in this research. It gives the databases utilized, their areas of research, and the number of articles that were found in each database, amounting to 187 publications.

Table 1 Digital Libraries and Search Scope

Database	Research Focus	Number of Articles Retrieved
IEEE Xplore	Networking, AI, ML	68
ScienceDirect	Engineering systems	41
SpringerLink	Intelligent optimization	32
ACM Digital Library	Networked systems	27
Google Scholar	Cross-disciplinary	19
Total		187

2.3. Inclusion and Exclusion Criteria

In order to keep the relevance, quality, and consistency of the reviewed studies, pre-assigned inclusion and exclusion criteria were established before the screening process. These were the same criteria used in all the retrieved articles in order to minimize selection bias and improve reproducibility.

2.3.1. Inclusion Criteria

The studies that were included in the review were studies that met the following criteria:

- Concentrated on the optimization of network systems, such as wired, wireless, cloud, Internet of Things, or software-defined network systems.
- Utilized machine learning, deep learning, or reinforcement learning methods.
- It was published as peer-reviewed journal articles or conference papers.
- Published between 2021 and 2025.
- Written in English.

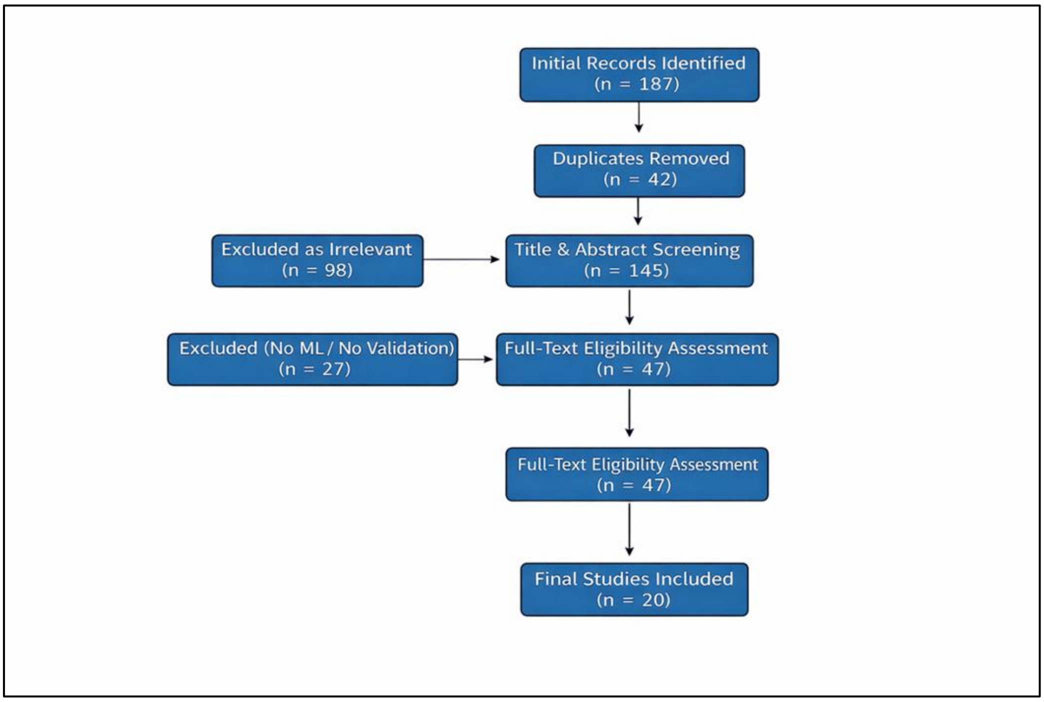
These criteria were necessary to make sure that only recent, methodologically solid, and practically important studies were taken into consideration [9], [11].

2.3.2. Exclusion Criteria

The following criteria were used to eliminate studies that had been conducted:

- Articles, technical reports, theses, or white papers that have not been peer reviewed.
- Literature on conventional optimization without the incorporation of ML.
- Papers that do not contain experimental data or performance analysis.
- Replicating the same research.
- Articles that do not focus on the optimization of networks directly.

The use of these criteria led to a great enhancement in the quality and focus of the final dataset [7], [10]. Figure 2 shows the steps of filters used on the originally searched studies, including the aspect of inclusion and exclusion criteria that narrow it down to the most useful publications



Source [5].

Figure 2 Inclusion and Exclusion Filtering Process

2.4. Article Selection Process

The process of article selection consisted of several steps to narrow down the sample of studies and only include those articles that were the most relevant and impactful in the final review.

In this phase, children's vulnerability is determined by identifying and screening them to identify their level of risk or risk factors affecting their development.

2.4.1. Identification and Screening

In this stage, the vulnerability of the children is identified by identifying them and screening them to find the level of their risk or risk factors influencing their development.

The preliminary search identified 187 articles in all the databases. Reference management tools were used to eliminate duplicate records (about 42). The rest of the articles were subjected to title and abstract screening, where irrelevant studies were sorted out in accordance with the scope and focus of research. The result of this step was 47 articles to be reviewed in full-text.

2.4.2. Eligibility Assessment

The shortlisted articles were critically checked against the grounds for inclusion and removal in their full text. Special focus was made on:

- Clarity of ML methodology
- Expertise in experimentation.
- Contribution to knowledge
- Consequently, the 20 most relevant and effective studies were chosen to be analyzed and synthesized in detail [4], [11], [15].

The following step involves extracting and synthesizing the data.

In each of the chosen studies, appropriate data were extracted systematically through the use of a pre-defined data extraction form. The attributes that were extracted were:

- Publication year and venue
- Application domain and type of network.
- Technique of machine learning adopted.
- Optimization objective
- Performance metrics
- Major results and weaknesses.

Qualitative comparative analysis was used to extract the data and then synthesize the data to identify trends, similarities, and gaps among the studies [12], [14].

2.5. Data Extraction and Synthesis

In each study that was selected, applicable data would be extracted in a predefined data extraction form. The attributes that were extracted were:

- Publication year and venue
- Type and field of application of the network.
- The technique of machine learning is applied.
- Optimization objective
- Performance metrics
- Investigative findings and limitations.

Qualitative comparative analysis helped to extract the data and synthesize the results, which made it possible to identify trends, similarities, and gaps across studies [12], [14]. Table 2 provides the data extraction attributes which will be used in the research, and these includes the type of networks (e.g., wireless, wired, SDN, IoT), machine learning methods (supervised, unsupervised, deep, and reinforcement learning), optimization goals (routing, resource allocation, energy efficiency), performance metrics (throughput, latency, energy) and validation methods (simulation, testbed, real-world experiments)

Table 2 Data Extraction Attributes

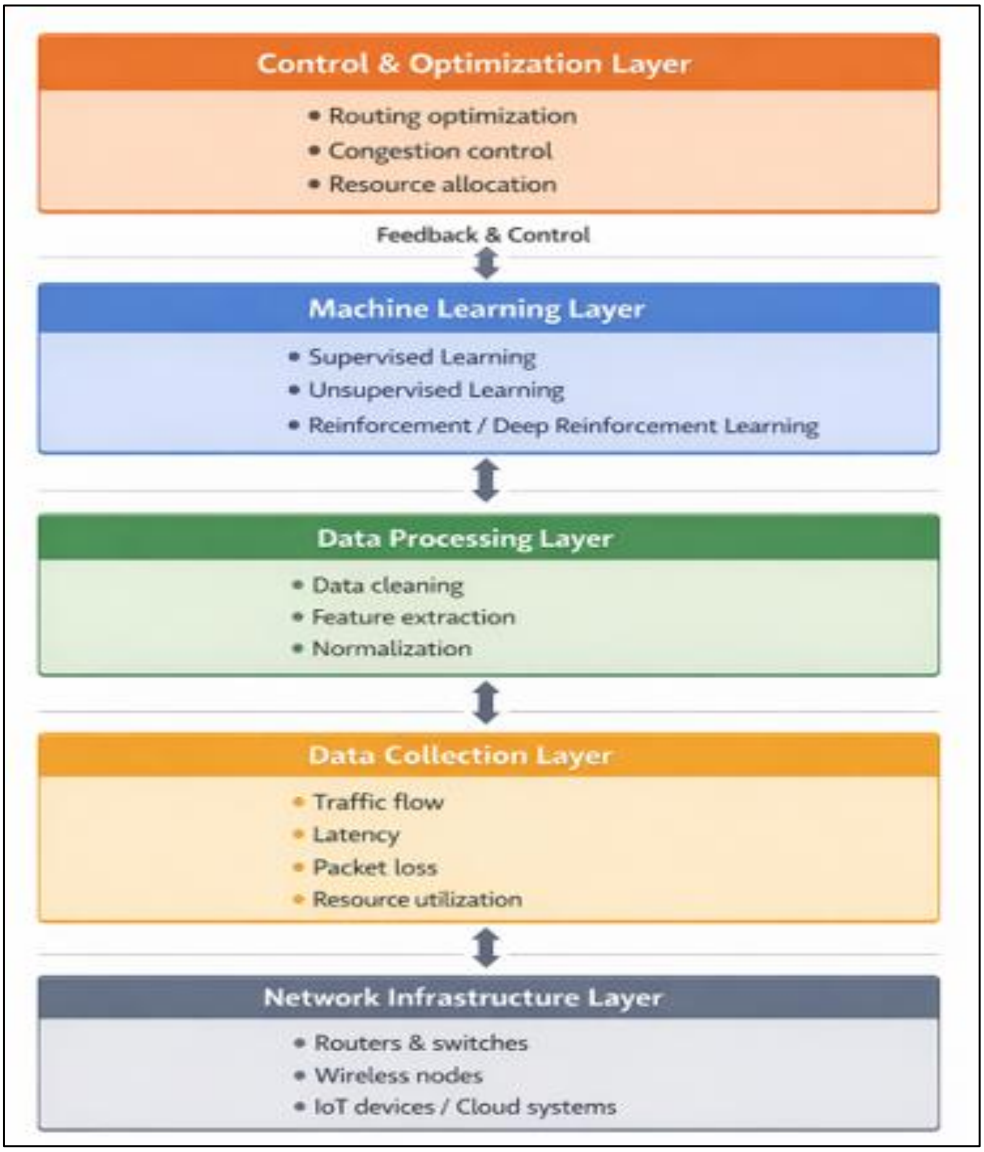
Attribute	Description
Network Type	Wireless, wired, SDN, IoT
ML Technique	SL, UL, DL, RL
Optimization Objective	Routing, resource allocation, energy efficiency
Performance Metrics	Throughput, latency, energy
Validation Method	Simulation, testbed, real world

Source [1], [2].

3. Results and discussion

3.1. Machine Learning Techniques for Network Optimization

The capability to learn by reading and adjust to changing conditions, as well as make intelligent decisions with limited human input, has made machine learning techniques central to current optimization of network systems. Various machine learning paradigms have been used in solving various optimization problems based on the availability of labeled data, learning objectives, and network dynamics. Figure 3 demonstrates a generalized system of machine learning-related network optimization, which shows how data measurement, feature derivation, training a model, and decision-making can be combined to optimize network usage. It is a framework offering a systematic way of applying the techniques of ML to increase efficiency, allocation of resources, and the overall quality of the network



Source [1].

Figure 3 Generic Machine Learning-Based Network Optimization Framework

3.1.1. Supervised Learning

One of the oldest machine learning methods used in network optimization is the supervised learning method. Support vector machines (SVM), decision trees, random forests, and artificial neural networks are the most common algorithms that have been employed to classify traffic, detect faults, predict quality of service (QoS), and estimate performance [6]. These models are based on labeled data, as the features of input, like packet headers, flow statistics, and network states, are related to pre-existing output classes or numerical values.

Supervised models of learning have a high accuracy and reliability in fairly fixed network settings, especially in a network where the traffic patterns and network behavior remain constant. Nonetheless, these models' utility is extremely contingent on the presence of large, good-quality labeled datasets. In extremely dynamic and diverse networks, e.g., 5G, IoT, and software-defined networks, it is difficult to acquire and keep labeled information up to date, thus reducing scalability and flexibility [7].

3.1.2. Unsupervised Learning

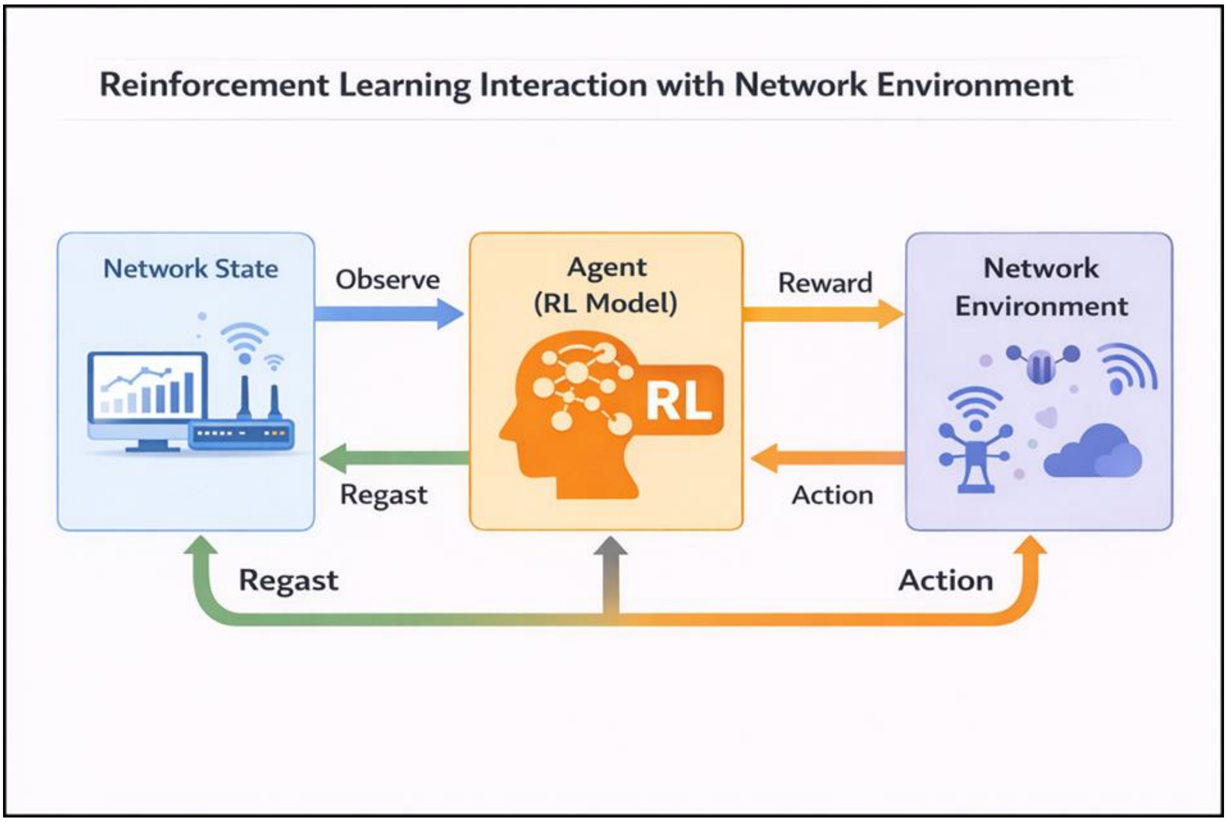
Unsupervised learning methods are used to overcome the shortage of labeled data to discover latent patterns and structures directly on unlabeled data. The algorithms are clustering (K-means, DBSCAN) and dimensionality reduction algorithms (principal component analysis (PCA)). Such methods have been effectively used in detecting anomalies, discovering traffic patterns, and network segmentation [8].

Unsupervised learning is especially useful when detecting abnormal behavior, unfamiliar attacks, or new traffic patterns without any previous knowledge. Nevertheless, the absence of clear labels complicates the process of performance evaluation, and the results may be difficult to interpret in large-scale networks.

3.1.3. Reinforcement Learning

Network optimization problems that require real-time and sequential decision-making have attracted a lot of attention to reinforcement learning (RL). Agents in RL-based systems are engaged in interaction with the network environment where they observe system states, act by bandwidth allocation or routing, and obtain rewards according to performance improvements. Routing optimization, congestion control, and spectrum management are known to fit into this learning paradigm [9], [10].

DRL, or deep reinforcement learning, is an additional version of RL that uses deep neural networks to estimate the complex relationship between states and actions, allowing the implementation of the algorithm in large-scale and high-dimensional network settings [11]. Figure 4 shows the network environment and interaction between a reinforcement learning agent and the environment. The agent monitors the states of the network, chooses actions that include routing or bandwidth allocation, and is rewarded depending on performance improvement.



Source [4].

Figure 4 Reinforcement Learning Interaction with Network Environment

3.1.4. Deep Learning

Deep learning networks, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have demonstrated excellent performance in processing high-dimensional network data as well as time-dependent network data. The models have gained popularity in traffic prediction, intrusion detection, and network fault diagnosis, due to their capability in extracting complex features on raw network data in an automated mode [12].

3.2. Optimization Objectives

The optimization of networks with the help of machine learning is designed to improve the general performance of the network by solving a variety of, and usually competing, goals: to maximize the throughput, to minimize the latency, to optimize the use of resources, energy consumption, and reliability. ML methods are adaptive and data-driven, unlike the traditional approaches that were either rule-based (or heuristic) and could not react to the condition of the network or the needs of the operation. The significant optimization goals that were discovered in the literature review are presented below.

3.2.1. Traffic Management and Routing.

Traffic control and routing continue to be the focus of optimization in today's communication networks, especially due to the growing complexity of heterogeneous and large-scale network infrastructures. ML-based routing schemes use the immediate network state information to make predictions about the congestion and update routing choices. Traffic classification and path selection have been implemented using supervised and reinforcement models with a good result in minimizing packet loss and end-to-end delay [13].

Deep reinforcement learning (DRL) has become one of the prevailing methods of adaptive routing, and agents acquire an optimal forwarding policy during recurrent interactions with the network environment. In contrast to the classical shortest-path or static routing methods, routing based on DRLs is capable of responding to traffic variations, a link failure, or different Quality of Service (QoS) policies. In a number of studies, it was found that throughput and latency reduction are significant when DRL-based routing is implemented in software-defined networks (SDNs) and wide-area networks [13], [16].

3.2.2. Resource Allocation

Wireless, cloud, and edge computing environments demand efficient allocation of resources, as they are limited in bandwidth, power, and computational resources. Resource allocation methods based on machine learning are designed to optimize network utility and provide a guarantee on QoS constraints, including delay, reliability, and fairness. Traffic demand and user behavior are predicted with the help of supervised learning models, which allow resources to be distributed proactively [14].

Deep learning and reinforcement learning methods have been demonstrated to be better in dynamically assigning resources, such as spectrum sharing, power control, and placement of virtual machines. These models are trained to learn the best allocation strategies to different network loads and user mobility patterns and are better than the traditional optimization algorithms in spectral efficiency and system throughput [14], [15]. Moreover, there is ML-based resource orchestration, which is crucial in network slicing in 5G and beyond networks to enable tailored services to various applications.

3.2.3. Energy Efficiency

The increasing size of data centers, wireless base stations, and Internet of Things (IoT) networks that consume a lot of energy has made energy efficiency a critical optimization goal. By predicting traffic demand and reconfiguring network resources (ML), intelligent energy management is achieved. Indicatively, learning-based models have the ability to dynamically toggle network elements to low-power in low-traffic periods without incurring a performance cost [16].

Deep learning algorithms have been used to optimize the use of energy consumed in cellular networks through a combination of efforts to control transmission power, user association, and base station activation. These techniques play a significant role in minimizing the cost of operation and sustainable network operation, especially in green networking projects.

3.2.4. Fault Detection and Network Reliability

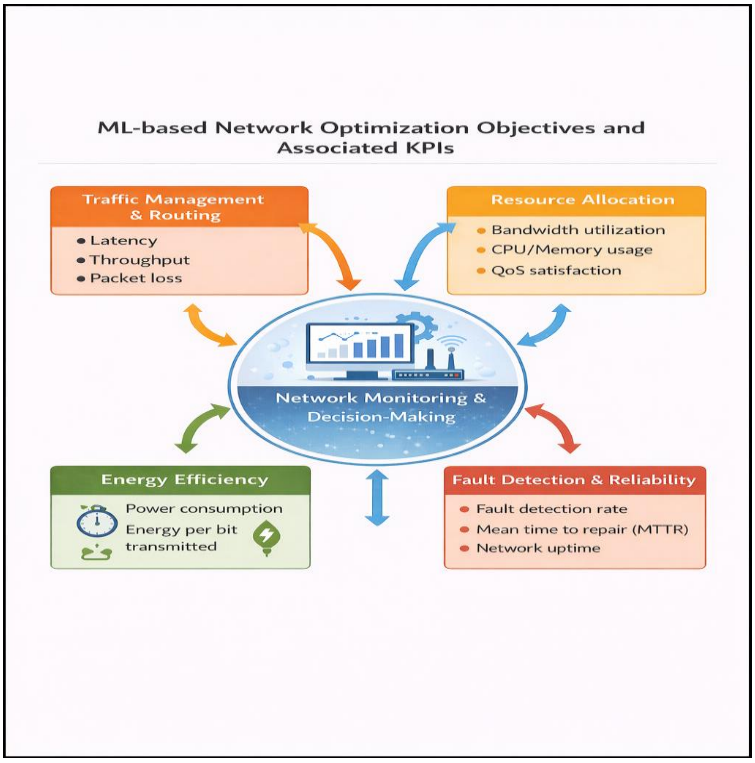
The mission-critical applications require high reliability of the network. Predictive maintenance and fault detection systems are based on ML to process both historical and real-time network data to detect an anomaly and possible failure. Detecting abnormal traffic patterns, equipment degradation, and performance bottlenecks are typically performed with unsupervised learning and deep learning techniques [17].

ML-based reliability optimization saves downtime and maintenance expenses by allowing the detection of faults early and automatic corrective measures. Such an active model can be used to increase the robustness of the network in general and to ensure sustaining network service when dealing with complex and highly dynamic network environments. Table 3 projects the main goals of network optimization to the applied machine learning methods. Each objective (traffic management, resource allocation, energy efficiency, and fault detection) is presented in the table with the focus on the ML methods applied to it and its main advantages, i.e., dynamic congestion adaptation, optimized resource use, reduced energy consumption, and predictive maintenance.

Table 3 Mapping Optimization Objectives to Machine Learning Techniques

Optimization Objective	ML Techniques Used	Key Benefits / Notes	References
Traffic Management & Routing	Reinforcement Learning (Q-learning, DQN), Deep RL	Dynamic adaptation to congestion, reduced latency	[13], [14]
Resource Allocation	Supervised Learning, Reinforcement Learning, Deep Learning	Optimized bandwidth, power, and CPU allocation	[14], [15]
Energy Efficiency	Supervised Learning, Deep RL	Minimized energy consumption in wireless/data centers	[16]
Fault Detection & Reliability	Supervised Learning, Unsupervised Learning	Predictive maintenance, early fault detection	[17], [12]

Figure 5 exemplifies the quality of network systems and the primary optimization goals, along with the KPIs that are common to determine the efficiency of such systems. Arrows are used to show the impact of machine learning methods on these goals in network monitoring and decision-making



Source [1], [2].

Figure 5 Optimization Objectives and Key Performance Indicators (KPIs)

3.3. Performance Evaluation Metrics

Performance evaluation metrics are highly important in the process of measuring the performance of some methods for machine learning in network system optimization. These metrics give objective metrics to measure the improvement in network performance, resource efficiency, reliability, and adaptability. Throughput, latency, rate of packet loss, energy consumption, resource utilization, and learning convergence time are some of the key measures that are shared across the reviewed literature. The interpretation of the result of an experiment and the comparison of the performance of the ML-based network optimization methods require knowing these metrics [18], [19].

3.3.1. Throughput

Throughput is the data rate of how many data packets are successfully sent across the network, expressed in bits per second (bps). High throughput implies effective use of the network resources and low congestion. Throughput is commonly employed as a measure of the efficiency of routing and resource allocation algorithms in ML-based optimization. As an illustration, routing schemes based on reinforcement learning are developed to choose the route that will maximize the throughput and keep the congestion low [13], [20]. Cloud and wireless networks also apply throughput to estimate the effects of dynamic bandwidth allocation strategies applied with the help of deep learning models.

3.3.2. *Latency*

The time taken by a packet between the source and destination is referred to as latency and is measured in milliseconds. Real time applications like video streaming, online gaming, and industrial IoT highly require low latency. In order to minimize the latency during the evolving traffic conditions, ML methods (and in particular the reinforcement learning and deep learning models) are used to optimize routing and traffic management [14]. Measures of latency, like average latency, maximum latency, and 95th percentile latency, are frequently reported to provide both normal and worst-case network delays.

3.3.3. Packet Loss Rate

Packet loss rate is used to determine the percentage of packets sent that fail to arrive at the final destination. A large amount of packet loss lowers the performance of the network, lowers throughput, and impacts application QoS. ML models are often used to control congestion, predict traffic, and identify errors to reduce the loss of packets [15]. The assessment of packet loss enables the researcher to evaluate the stability and consistency of the optimization algorithms in different network conditions.

3.3.4. Energy Consumption

The wireless networks, data centers, and IoT systems require a critical consideration of energy efficiency. The optimization is based on ML to minimize power consumption at the expense of performance. Measures of energy consumption are commonly denoted in watts, or joules per bit transmitted. Deep reinforcement learning and other algorithms will be able to adaptively regulate transmission power, idle mode, and resource allocation to save on energy without affecting the QoS [16]. The measure is especially applicable to the field of green networking and sustainable system design.

3.3.5. Resource Utilization

Resource utilization measurements measure the efficiency of the resources being assigned to the network, i.e., bandwidth, CPU cycles, and memory. The utilization of the resources is optimized, which helps to provide the needed performance level without the over-provisioning of the network and allows for a decrease in the cost of its functioning. Supervised and reinforcement learning methods are examples of machine learning methods that have been applied to dynamically assign resources according to anticipated demand patterns [14], [15]. Measures usually consist of the bandwidth consumption rates and CPU load, as well as memory consumption rates.

3.3.6. Learning Convergence Time

Convergence time is a measure of the time taken by an ML model before it becomes stable and starts giving valid decisions. The shorter convergence times are desired, particularly for real-time or highly dynamic networks. Convergence time is used to measure the efficiency of an algorithm in reinforcement learning and a deep learning model to adapt to varying network conditions [9], [11]. The time of convergence is an important parameter for measuring the feasibility of the implementation of the ML algorithms in the functional networks.

The combination of these metrics gives a complete frame of analysis of network optimization of the ML-based. Packet loss, delay, and throughput concentrate on the quality of performance, efficiency is considered by energy consumption and resource utilization, and learning convergence time is an indicator of adaptability by the algorithm. The choice of relevant metrics is essential to be able to compare techniques properly and make the decisions to deploy them in various network settings [18].

3.4. Challenges and Limitations

Although machine learning (ML)-based network optimization methods have yielded promising performance and made great progress, there exist a number of critical challenges and limitations. It is necessary to address these issues to make it possible to involve the methods of ML in the real network setting and guarantee efficient, reliable, and safe functioning. The major issues are quality and availability of data, interpretability of the model, scalability, real-time implementation, security, and privacy.

The data available will be of high quality, and it will be reliable. The use of the machine learning models, especially the supervised and deep learning methods are highly reliant on high-quality labelled data to train. Nevertheless, it is not always easy to gather large-scale, precise, and current network data because of network heterogeneity, dynamism in traffic patterns, and privacy laws [18], [19]. The historical traffic logs can be incomplete or erroneous in most instances, resulting in biased or sub-optimal models. Moreover, network anomalies or faults may be labelled manually, which can be time-consuming and requires expertise. In wireless, IoT, and cloud networks, the problem is enhanced by the steady stream of high-dimensional data, hence posing substantial challenges to model training and validation.

3.4.1. Model Interpretability

The other important constraint is the interpretability of machine learning models, especially the deep neural networks and reinforcement learning agents. Although these models may work well, their internal decision-making methods are mostly black box and thus, the network operators cannot comprehend, trust, and validate their results [18]. The inability to interpret the result is a drawback of both troubleshooting and meeting regulatory requirements as well as assuring safety in the critical domains of industrial internet of things, medical, and autonomous systems. Explainable AI (XAI) methods are under development to overcome this weakness, but these methods are usually associated with trade-offs of model complexity or performance.

3.4.2. Scalability and Real-Time Deployment.

One important issue related to applying ML models to large networks is scalability. Most ML algorithms are expensive to scale to large-scale computing and inference, making them infeasible to apply to dynamic network problems in real-time. As an example, deep reinforcement learning agents can be quite slow to converge, hence not suited to latency-sensitive applications [11], [20]. Also, the dimensionality of network states in large-scale distributed systems is high, making it difficult to meet the memory and processing overhead and communication latency between network nodes. The question of how to ensure that ML models are efficient in scaling but preserving performance is an open research problem.

3.4.3. Data protection and privacy issues.

Optimization methods of networks using them as ML create new attack points and privacy threats. Adversarial attacks may be used to influence the performance of models by manipulating their inputs to either low performance or security breach [18]. Equally, the models can be trained on sensitive traffic data, causing possible privacy breach, and this is particularly so when shared or in multi-tenant setups. Such techniques as federated learning, differential privacy, and secure multi-party computation are under consideration to address these threats, but are frequently associated with additional computational and communication costs [19].

3.4.4. Summary:

Although machine learning has a transformational potential of being the key to optimizing a network, its practical implementation is limited by issues related to the data, interpretability, scalability, and security. These limitations are important to overcome in order to make ML-based network optimization frameworks reliable, transparent, and capable of functioning in the real world. Future studies should be interested in data-efficient models, explainable algorithms, scalable architectures, and privacy-preserving methods to realize all the advantages of intelligent network systems in detail [18] [20]. Table 4 highlights the major issues in machine learning-based network optimization, namely, the small amount of data, the model interpretability, scalability, and security/privacy concerns. Potential solutions to every challenge, including federated learning, explainable AI, hierarchical ML, and privacy-preserving solutions, are proposed .

Table 4 Challenges in ML-Based Network Optimization

Challenge	Description	Potential Solutions
Data Scarcity	Lack of labeled data	Federated learning
Model Interpretability	Black-box nature	Explainable AI
Scalability	Large-scale networks	Hierarchical ML
Security & Privacy	Data exposure risks	Privacy-preserving ML

Source [1],[4].

4. Conclusion

The systematic review has examined the latest developments in the optimization of a network system using machine learning (ML) methods and the revolutionary nature of machine learning in current communication networks. All the reviewed papers provide sufficient evidence that ML-based approaches, in particular, deep learning and reinforcement learning, are much more efficient than the conventional rule-based and heuristic optimization methods. Their capability to learn massive network data, adaptability to changing traffic patterns, and proactive optimization is possible to enhance their performance with respect to all major goals, including the minimization of latency, effective utilization of resources, energy savings, and increased reliability.

Deep learning models have shown to be particularly useful in processing high-dimensional and complex network data, making it possible to predict traffic correctly, detect anomalies, and model performance. Equally, reinforcement learning has demonstrated excellent ability in dynamic decision-making processes, including routing and congestion control, and spectrum allocation, among others, whereby agents are able to interact with the network environment continuously and optimize long-term performance. Such features are very appropriate to new network paradigms, such as, but not limited to, software-defined networking (SDN), network function virtualization (NFV), 5G/6G systems, Internet of Things (IoT), and cloud-based infrastructures.

Although these are the benefits, the review also pinpoints a number of severe challenges that surround extensive real-life implementation. The large amounts of high-quality and labeled data needed by ML models are not always readily available because of privacy and operational issues, and network heterogeneity. Also, training and deployment of some complex ML models can have a large computational cost, which can be too expensive for network components with limited resources or that need low latency. Lack of interpretability of most ML models is another critical issue because it lowers the level of trust and makes automated decisions not easy to comprehend, validate, and troubleshoot by network operators.

In order to overcome these limitations, it is suggested that future studies focus more on the creation of explainable AI methods that will add clarity to the process of optimization decision-making in ML. Federated learning represents a good perspective that allows joint training of models among distributed network nodes without compromising the privacy of data. Also, there is a possibility of balanced hybrid optimization structures that combine both ML and standard analytical and heuristic procedures, as these can provide both learning-based flexibility and the knowledge of the domain and compute efficiency. By overcoming these challenges, network optimization systems that are based on ML will be able to be more robust, trustworthy, and practically applicable in the future, which will provide the future needs of the communication networks of the next generation.

Compliance with ethical standards

Disclosure of conflict of interest

The Authors proclaim no conflict of interest.

References

[1] Zhao T, Xie Y, Wang Y, Cheng J, Guo X, Hu B. A Survey of Deep Learning on Mobile Devices - Applications, Optimizations, Challenges, and Research Opportunities. Proceedings of the IEEE. 2022, 110(3): 334-354.

[2] Hu S, Chen X, Ni W, Hossain E, Wang X. Distributed machine learning for wireless communication networks - Techniques, architectures, and applications. IEEE Communications Surveys & Tutorials. 2021, 23(3): 1458-1493.

[3] Otieno H, Malila B, Mwangama J. Deployment and Management of Intelligent End-to-End Network Slicing in 5G and Beyond 5G Networks - A Systematic Review. IEEE Access. 2024, 12: 190411-190433.

[4] Li J, Wu H, Huang X, Huang Q, Huang J, Shen X. Toward reinforcement-learning-based intelligent network control in 6G networks. IEEE Network. 2023, 37(4): 104-111.

[5] Pena C, Cartaxo B, Steinmacher I, Badampudi D, Silva D, Ferreira W, Soares S. Comparing the Efficacy of Rapid Review with a Systematic Review in the Software Engineering Field. Journal of Software Evolution and Process. 2025, 37(1): e2748.

[6] Sheikh M, Peng, Y. Procedures, criteria, and machine learning techniques for network traffic classification - a survey. IEEE Access. 2022, 10: 61135-61158.

[7] Ajibuwa O, Hamdaoui B, Yavuz A. A survey on AI/ML-driven intrusion and misbehavior detection in networked autonomous systems - techniques, challenges, and opportunities. arXiv preprint arXiv:2305.05040. 2023

[8] Cui Y, Liu Z, Lian S, A Survey on Unsupervised Anomaly Detection Algorithms for Industrial Images. IEEE Access. 2023, 11: 55297-55315.

[9] Zhang J, Zhao J, Cai K. Deep Reinforcement Learning for Routing and Scheduling Optimization in Fanet-Assisted IoT Networks. The Integrated Communications, Navigation and Surveillance Conference (ICNS). 2025, 1: 1-7.

[10] Tian H, Liao X, Zeng C, Sun D, Zhang J, Chen K. Efficient DRL-Based Congestion Control with Ultra-Low Overhead. IEEE/ACM Transactions on Networking. 2024, 32(3): 1888-1903.

[11] Alwarafy A, Abdallah M, Çiftler B, Al-Fuqaha A, Hamdi M. The Frontiers of Deep Reinforcement Learning for Resource Management in Future Wireless HetNets - Techniques, Challenges, and Research Directions. IEEE Open Journal of the Communications Society. 2022, 3: 322-365.

[12] Elsayed S, Mohamed K, Madkour M. A Comparative Study of Using Deep Learning Algorithms in Network Intrusion Detection. IEEE Access. 2024, 12: 58851-58870.

[13] Santos CL, Mezher AM, León JP, Barrera JC, Guerra EC, Meng J. ML-RPL: Machine Learning-Based Routing Protocol for Wireless Smart Grid Networks. IEEE Access. 2023, 11: 57401-57414.

[14] Hamedoon S, Chattha J, Rashid U, Kazmi S, Mazzara M. AI-Driven Resource Allocation for RIS-Assisted NOMA in IoT Networks. IEEE Access. 2025, 13: 68152-68171.

[15] Krishna K, Siddhey M, Md-Abul K, Goel O, Jain A. Optimizing Machine Learning Models for Predictive Analytics in Cloud Environments. JRPS. 2022, 13(5): 611–642.

[16] Hossain M, Hossain A, Ansari N. AI in 6G: Energy-Efficient Distributed Machine Learning for Multilayer Heterogeneous Networks. IEEE Network. 2022, 36(6): 84-91.

[17] Rahimpour S, Husev O, Vinnikov D, Kurdkandi N, Tarzamni H. Fault Management Techniques to Enhance the Reliability of Power Electronic Converters - An Overview. IEEE Access. 2023 11: 13432-13446.

[18] Ghadi Y, Mazhar T, Shloul T, Shahzad T, Ahmed U. Machine Learning Solutions for the Security of Wireless Sensor Networks - A Review. IEEE Access. 2024, 12: 12699-12719.

[19] Sun H, Liu Y, Nag A. Advancing 6G - Survey for Explainable AI on Communications and Network Slicing. IEEE Open Journal of the Communications Society. 2025, 6: 1372-1412.

[20] Qi Q, Chen X. Robust Design of Federated Learning for Edge-Intelligent Networks. IEEE Transactions on Communications. 2022, 70(7): 4469-4481.