OPEN ACCESS JOURNALS

Indiana Journal of Multidisciplinary Research

Abbreviated Tittle- Ind J Multi Res

ISSN *(Online)*- 2583-3820

Journal Homepage Link- https://indianapublications.com/Journals/IJMR

DOI: https://doi.org/10.5281/zenodo.21610107

INDIANA PUBLICATIONS

PRODUCTIVE AND QUALITY RESEARCH

Research Articles

Volume-06|Issue-04|2026

Quality of Service (QoS) Optimization in 5G/6G Networks Using Neural Networks

Charis E Shiny[1], Annapurna S[2], C Lakshana[3], Anusha Fakirappa Bogur[4], Ramesh S[5], Gajanan M. Naik[6]

[1,2,3,4]Undergraduate Student, School of Computer Science and Engineering, RV University, Bengaluru, India.

[5]Faculty, School of Computer Science and Engineering, RV University, Bengaluru, Karnataka, India.

[6]Faculty, Department of Mechanical Engineering, RV Institute of Technology and Management, Bengaluru, Karnataka, India.

Article History

Received: 05.05.2026

Accepted: 22.06.2026

Published: 25.07.2026

Citation

Charis E., Annapurna S, C Lakshana, Anusha F, Ramesh S, Gajanan M Naik. (2026) Quality of Service (QoS) Optimization in 5G/6G Networks Using Neural Networks. *Indiana Journal of Multidisciplinary Research, 6*(4), 356-362.

Abstract: 5G is rolled out and next generation 6G networks are also being developed, ultra-low latency (URLL) communication as a standard is critical in supporting the plethora of applications, spanning autonomous vehicles, immersive extended reality experience, etc. However, traditional quality of service (QoS) policy support mechanisms are faced with significant limits in identifying and managing dynamic heterogeneous traffic types in the next generation wireless networks. Traffic demands will vary widely and traditional QoS will not provide the flexibility to adopt mechanisms quickly to arbitrary network conditions, along with providing a very different service requirement. This research proposes an innovative neural network based QoS optimization framework that predicts the traffic parameters from intelligence resource allocation. Even a variety of deep learning models will be used to predict the key performance indicators or latency, jitter and packet loss for different scenarios. Neural network slicing is an automation technique that manages bandwidth allocation autonomously, adapting in real time to traffic demands and service provisioning requirements. The proposed approach is expected to improve QoS by reducing latency and packet loss, and also to enhance resource utilization and service reliability in 5G and 6G networks.

Keywords: 5G, 6G, Neural Networks, QoS, Network Slicing, Traffic prediction.

INTRODUCTION

Background on 5G/6G networks and QoS requirements

Starting from the 2020s, 5G laid the foundation for network connection with ultra-low latency, IoT, smart cities, and autonomous vehicles. It has greater capacity than the previous 4G network. It is approximately 10 times faster than 4G. South Korea was the first to implement a nationwide 5G network. It uses millimetre waves. A 5G network is a cellular network since the service area is divided into smaller areas or cells. A single 5G cell can connect 1 million devices per km². The speed of 5G is 1-10 Gbps. Around 2030, 6G is expected to be implemented globally. The arrival of 6G signifies the network connection consisting of near zero latency, hyper connectivity, integration of AI (Artificial Intelligence), and providing universal coverage by connecting space, aerial and terrestrial networks. While 5G focuses on IoT (Internet of Things), 6G focuses on IoE (Internet of Everything). It is expected to deliver peak speeds of 1Tbps. It will use THz(terahertz) frequencies which are far beyond the 5G spectrum. The main challenge with higher frequencies is that they require higher power. So, active research is going on all over the world to achieve sustainability as well.

Quality of Services (QoS)

Quality of Services (QoS) is a methodology or a set priority-based rules used to regulate network traffic by allowing tasks of networks having higher importance to be executed first. It also includes a flow control mechanism to make sure that one network does not take forever to execute. No network is starved; every network is given a chance at least once. It consists of various parameters like bandwidth, latency, jitter, number of packets lost, and the throughput achieved.

Different types of QoS includes, Best-Effort Model, Integrated Services (IntServ), Differentiated Services (DiffServ). If the network and their respective tasks are assumed to be different processes and QoS as the OS(Operating System), it could be said that the Best-Effort Model uses the FCFS algorithm (first come first serve) where as IntServ reserves the bandwidth/resources for the network similar to Priority Scheduling algorithm and DiffServ follows a pre-set order of a priority list of networks which is again similar to multi-level queue scheduling algorithms.

Limitations of existing QoS optimization techniques:

- **Scalability and Management:** The requirement of per-flow state information and resource reservation creates major scalability issues in large networks.

***Corresponding Author:** Ramesh S

DiffServ was introduced to solve this but it in turn created its own set of complexity.

- **Flexibility and Dynamic Adaptation:** Traditional QoS struggles to adapt in real-time with the changes in network conditions due to its static configuration. This makes it difficult for them to adapt to dynamic changing environments. Traditional networks are manual, static and brittle.
- **Inefficiency in heterogeneous networks:** As the networks shift from homogeneous to heterogeneous, QoS is not capable of adapting to this and performs poorly in these conditions. Applying a uniform QoS policy to a hybrid system is a suboptimal choice.
- **Encrypted Traffic:** The ability of QoS to classify traffic is reduced significantly due to the widespread increase in the use of encryption technologies. Without visibility of the packet payload, the network cannot classify traffic into appropriate QoS queues.
- **Integration with modern architectures**: Dynamic nature of traffic makes it difficult for QoS to adapt to emerging architectures. Traditional QoS is not designed to handle performance isolation or predictable latency.
- **Over-Provisioning:** The practical solution of implementing over-provision bandwidth rather than implementing a complex QoS is another limitation to the existing QoS optimization techniques.
- **End-to-end Service:** Achieving a complete end-to-end QoS is considered practically impossible in the public internet because of its multi-domain nature. Implementing an inter-domain QoS hasn't gone too much due to business and trust issues.

Motivation for using Neural Networks.

In order to provide and maintain good quality service in 5G/6G. Neural Network is used to learn, think and remember for understanding what the end user needs and shares the internet speed smartly. This concept is used to enhance the user experience on the internet faster and sustainable.

The following will give reasons for the use of Neural Networks in QoS of 5G/6G:

- Network Traffic can be predicted and classified
- Neural network will learn to share bandwidth among end – devices
- During mobility of user, the network will still perform better because of the prediction of movement of user in geo-location
- Interferences will get better management due to better and informed network lines
- Quality of Experience (QoE) is mapped to Quality of Service (QoS). Hence providing better learning for the Neural Network for enhanced 5G and 6G speeds.

Hence, we find a reason to use Neural Networks in 5G and the future of 6G lines.

MATERIALS AND METHODS

NN-based prediction of latency, jitter, and packet loss.

Latency, Jitter and Packet Loss are key metrics in QoS of optimization in 5G or 6G networks. The following are a brief definition and reasoning of the metrics:

- Packet Loss: When a data packet fails to reach its destination, it is considered as packet loss.
- Latency: It is the total delay for data to travel from source to destination.
- Jitter: It is the variation or inconsistency in delay over time.

While these are the key concepts in Computer Networks (CN), they also help in unveiling the learning potential in Neural Networks.

Neural Networks can predict these metrics to enhance network speeds that can improve QoS by replication or making a virtual copy of the neural network for testing and providing better results in the real application.

Neural-assisted network slicing for dynamic bandwidth allocation

Network Slicing is a mechanism where a single physical network is divided virtually, into slices, and each slice caters to a unique purpose/task. Each slice has its own QoS (bandwidth, latency, jitter, number of packets lost, and the final throughput achieved).

Dynamic Bandwidth Allocation (DBA) basically means that the bandwidth is allocated after the request is sent and based on how much data the device needs to send. The bandwidth is fairly distributed amongst different processes considering which requires more. In modern communication systems, it is especially used in PON - Passive Optical Networks and 5G networks. The same method of allocation will also be used in 6G networks. This prevents resources being wasted and improves efficiency.

A neural assisted network is backed by an Artificial Intelligence (AI) model.

A Neural-assisted network slicing for dynamic bandwidth allocation implies that an AI model is trained based on input parameters or independent variable/s such as data usage by different networks and devices in them, total time taken by the packets to complete one request-response cycle, available total bandwidth, signal quality, priority level of the slice, historical usage of the data and past traffic.

And the target variables would be the prediction of the allocation of bandwidth to each slice (done dynamically), minimizing latency, maximizing throughput, reducing the number of packets lost, making sure of fair distribution amongst the resources based on the predefined order of priority.

Comparative analysis with traditional QoS methods
A major contribution of this research is the comparison between traditional QoS methods. Traditional QoS techniques depend on static rules, fixed traffic classes, and priority-based scheduling. These methods worked well in the earlier generation networks, but they face serious challenges in 5G and 6G environments. The traffic in these networks is highly different, ranges from IoT devices to AR/VR applications and autonomous systems. Traditional methods are basically reactive, deals with poor service only after it occurs, which leads to delay, packet loss, and inefficient use of resources.

The proposed neural based framework addresses these limitations by predicting QoS parameters like latency, jitter, and packet loss before the problems occur. This approach allows the network to manage resources in advance and reduces the service interruptions. When it combines with neural network assisted network slicing, bandwidth can be allocated dynamically based on the real time traffic and service requirements. This analysis shows that the neural based methods provide better adaptability, scalability, and efficiency as compared to traditional QoS.

QoS optimization methods in 5G and 6G
Key optimization methods in 5G:

- Network Slicing: Involves creating multiple isolated networks on a shared physical infrastructure. Each slice provides different services. It ensures guaranteed resources for critical services. Slices can be configured to avoid latency and other limitations.
- Software - Defined Networking (SDN) and Network Function Virtualization (NFV): SDN enables centralized, programmable control of the entire network. NFV virtualizes network functions. SDN provides a global view of the network state. NFV allows for flexible scaling.
- Edge computing and Multi - Access Edge Computing (MEC): Processing data closer to the user instead of sending to the cloud. This serves as the main factor for reducing latency.
- Advanced Radio Resource Management and Scheduling: 5G New Radio introduces a flexible numerology. They transmit data immediately without waiting for a slot boundary.

Key optimization methods in 6G:

- AI - Native and Zero - Touch Network Management: ML algorithms will predict network congestion, user mobility, and application demand before they happen. They will reconfigure accordingly to provide seamless QoS without the need of human intervention.
- Integrated Sensing and Communication (ISAC): Wireless signals are also used for sensing the environment rather than just for communication. Sensing data can help solve the limitation of QoS not being able to adapt real-time.
- Terahertz (THz) and Novel Spectrum Bands: 6G will explore higher frequency bands to achieve extreme data rates. These also have challenges which require new QoS management techniques.
- Holographic - Type Communications and QoE Maximization: It maximizes the human-perceived experience. QoS mechanisms driven by QoE models help in understanding hologram and immersive video.

Research gap: Lack of integrated NN-based QoS prediction and slicing strategies
Several studies have worked on improving QoS in 5G networks, but they focused only on one aspect either on predicting the QoS parameters or network slicing. For example, Khudhus et al. Developed machine learning models like random forest and decision Tree to predict the QoS parameters. Their work showed that prediction can improve monitoring and service quality. But their approach did not use these predictions to guide real time bandwidth allocation or slicing decisions.

Similarly, Kumar and Ahmad proposed a machine learning based strategy to manage resources across different network slices. Their model dynamically allocated bandwidth to meet changing traffic demands. It was effective for flexible allocation, but this method did not include predictive intelligence for QoS metrics.

From all these studies, it is clear that there is no integrated framework which combines neural network based QoS prediction with dynamic network slicing. Addressing this gap is important to meet the strict performance of 5G and 6G applications

System Model
Network Model
Multi-slice 5G/6G architecture with heterogeneous services (eMBB, URLLC, mMTC):

- eMBB: Enhanced Mobile Broadband is a service category which provides faster broadband for high data consuming processes. Thus, it provides higher throughput.
- URLLC: Ultra-Reliable Low Latency Communication is a reliable connection with minimal latency (<1 milliseconds) and is often dependent during emergency cases.
- mMTC: Massive Machine-Type Communication is the backbone of all low power consuming and low data consuming devices and connections. A huge number of devices are processed at once. So, it provides scalability and energy efficiency.

The multi-slice architecture here refers to a 5G (or even a 6G) network virtually divided into multiple slices to

*Corresponding Author: Ramesh S

cater unique, customised requirements and each slice behaves as a separate network entity.

The listed services here (eMBB, URLLC, and mMTC) are heterogeneous because they all cater to completely different requirements and utilise varying Quality of Services (QoS) metrics.

Neural Network Model

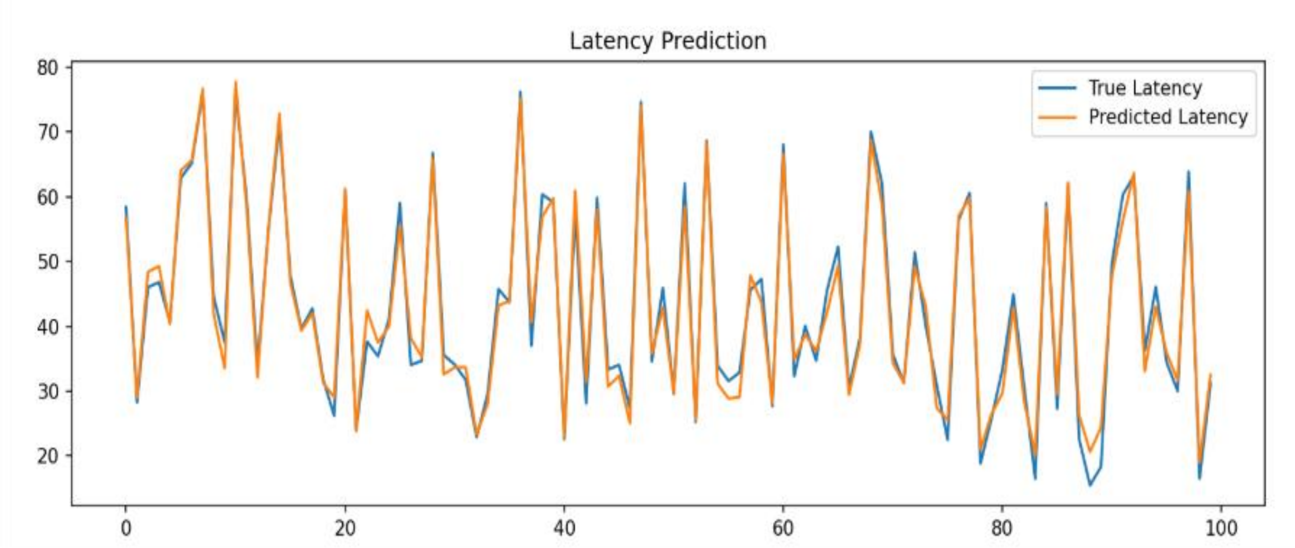
i) Input: traffic patterns, historical QoS metrics, network state.

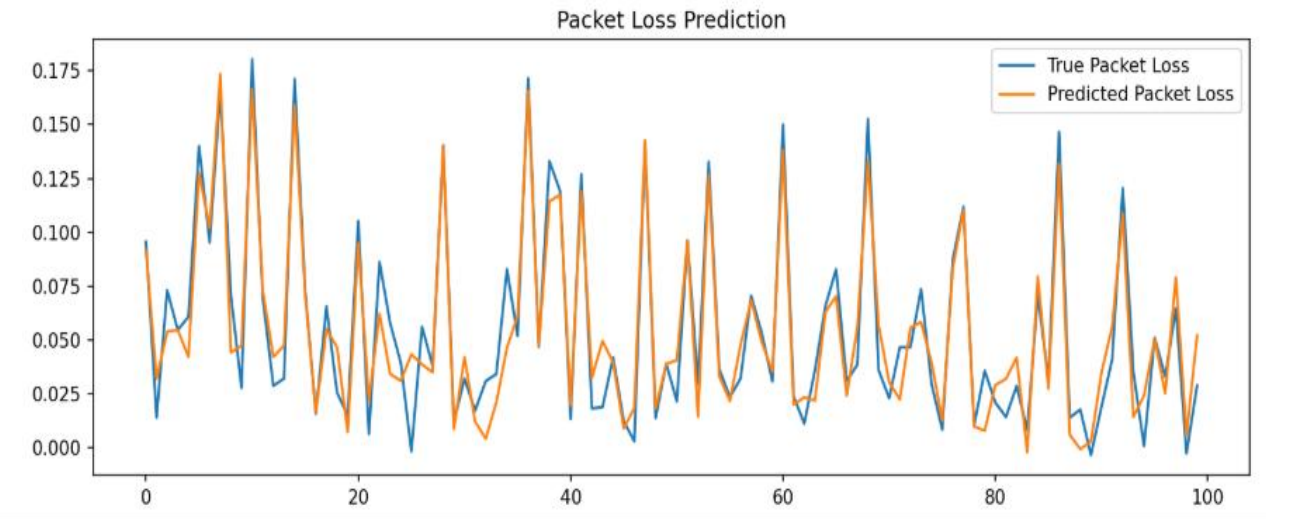
The libraries used are as follows in python:

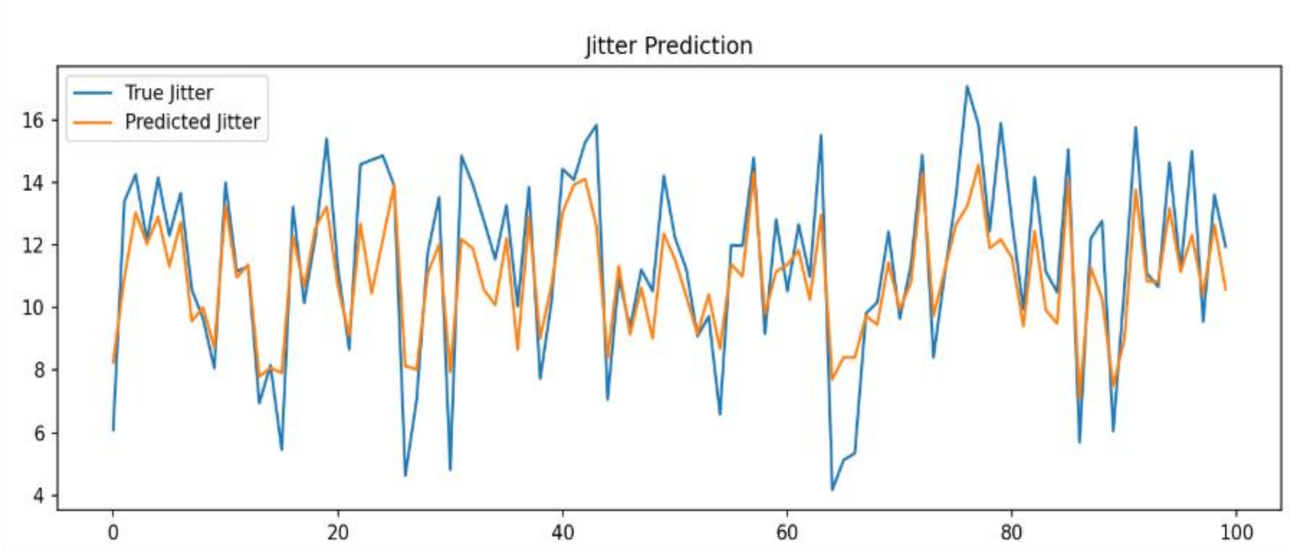
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.preprocessing import MinMaxScaler
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import LSTM, Dense, GRU
from tensorflow.keras.optimizers import Adam

The current implementation includes three main parameters – Traffic patterns, Historical QoS Data and Network State parameters. The Traffic patterns include data like data rate and session duration, while latency, jitter and packet loss come under QoS data.

ii) Output: predicted QoS parameters (latency, jitter, packet loss).







iii) Model choices: LSTM/GRU (time-series), CNN (pattern recognition), or hybrid models.

QoS Optimization Algorithm

- The proposed methodology for QoS Optimization Algorithm, utilizes Neural Network (NN) to predict for improved resource allocation in 5G / 6G networks. The key parameters being – Latency, Jitter, and packet loss.
- There are two main requirements that are met through this algorithm i.e. proactive bandwidth allocation and dynamic network slicing.

Evaluation Metrics

i) Accuracy of QoS predictions.

Modern systems usually tend to predict network states before optimization. It is very important that these predictions are accurate.

Key Metrics:

- Mean Absolute error: It states the average size of the errors. Lower MAE makes the prediction more accurate.
- Root Mean Square Error: RMSE is the measure of how far off the predictions are from the actual values, on average.
- Precision and Recall: Precision states how correct are the positive predictions. Recall quantifies how many positive predictions were actually found. It is considered ideal to have a high F1 score.

High accuracy in prediction makes sure that the decisions made are based on reliable information and thereby preventing inefficient wastage of resources or failures.

ii) Improvement in resource utilization.

This measures how efficiently the network's physical resources (bandwidth, CPU, radio spectrum) are used after applying the QoS technique.

Key Metrics:

- Resource Utilization Rate - The percentage of total available capacity (e.g., bandwidth on a link, CPU cycles on a server) that is being used to carry productive traffic. The goal of optimization is to maximize this without inducing congestion.
- Blocking/Dropping Probability – The rate at which new requests are blocked, or existing flows are terminated due to lack of resources. An effective QoS technique reduces this probability significantly.
- Fairness Index – Used to quantify how resources are distributed equally among competing users or competitors. Optimal solutions aim to achieve high utilisation while maintaining fairness. Avoids starvation of any user.

Higher utilisation translates to reduced cost and improved overall network capacity.

iii) Latency reduction and throughput gain.

The following are the metrics directly impacting the application performance and quality of experience (QoE):

Key Metrics for Latency:

- End-to-End Latency – It is the total time taken for a packet to travel from the source to the destination. Reduction of this factor is directly related to saving time.
- Jitter - It refers to the variation in end-to-end delay. High jitter causes inconsistencies.
- Tail Latency – Used to measure the worst-case latency experienced by packets.

Key Metrics for Throughput:

- Average Throughput: The average rate of successful data delivery over a communication link.
- Peak Throughput: The maximum achievable data rate.

- Cell Edge Throughput: The throughput achieved by users in the worst coverage areas within a cell.

The metrics above directly define the user experience. Lower latency and higher throughput are ideal for better optimization.

RESULTS

Using simulated traffic patterns and historical QoS data, the suggested QoS neural network (NN)-based models effectively predicted latency, jitter and packet loss. Out of the models tested, the LSTM and GRU architecture pairs showed a greater capability of temporal pattern recognition relative to conventional neural, feed-forward networks. This underscores the necessity of time-series network traffic modeling.

The results indicate that the proactive approach to bandwidth allocation, using NN predictions, reduced performance loss during peak usage times. Rather than being reactive to congestion once it was present, the system predicted possible quality of service (QoS) lapses and allocated bandwidth. This reduced the spurts of latency and made the traffic flow more even.

DISCUSSION

With the rapid evolution of modern communication networks such as 5G and emerging 6G, the integration of deep learning techniques has become essential for efficient network management and optimization. However, despite their potential, deploying neural models in such dynamic and large-scale environments introduces several practical challenges. These challenges arise from the complexity of network data, the resource-intensive nature of model training, and the need to maintain performance in constantly changing conditions. The key issues are outlined below:

- Massive Data Requirements**:** 5G and 6G networks generate a large amount of heterogeneous traffic data from IoT devices, autonomous systems and immersive applications. To guide neural models to get accurate output large datasets must be collected and need to be pre- processed. This process requires more time, effort and storage resources.
- High Computational Cost**:** training the deep learning models on such large datasets requires specific hardware like GPUs or TPUs. It is costly to acquire and maintain these infrastructures across distributed network domains. Large scale training consumes more energy, which creates the issues in cost and environmental sustainability.
- Need for Continuous Retraining: Network traffic patterns change quickly due to user movement, different service, and with changing demand. Older datasets may cause models to become outdated quickly. Neural models need to be retrained frequently to maintain accuracy, which increases computational and time overhead.
- Scalability challenges: It's challenging to integrate retrained models into existing network environments without service interruptions. As network scale increases, it is difficult to deploy the updated models across multiple base stations by limiting real time adaptability and scalability.

CONCLUSION

Prediction of the latency, jitter and packet loss was successful by the neural network in simulated data. Hence, we can get QoS – neural network based smarter and proactive resource allocation for 5G / 6G networks.

The journey towards 6G is helpful in crossing boundaries of what is possible in telecommunication.

i) Federated neural networks for distributed QoS optimization.

One of the core limitations in the optimization of QoS is the use of centralized data collection. This leads to compromising the user's privacy. Federated learning offers a revolutionary distributed machine learning framework that directly addresses these challenges. In this model, the training process is inverted: instead of pooling user data into a central cloud server for model training, the AI model itself is dispatched to the data sources located at the network edge. This means the network can develop highly accurate, real-time predictive models for traffic, congestion, and resource needs that are informed by a vast array of local conditions without ever compromising user confidentiality. The primary research challenges moving forward will be to enhance the efficiency of the communication process between the edge and the aggregator.

ii) Integration with reinforcement learning for self-optimizing 6G networks.

As complexity of networks increases, traditional optimization techniques don't meet the requirements. To solve this, RL (Reinforcement learning) is introduced. Reinforcement learning provides a pathway to independent and self-optimized 6G networks. When a RL agent is deployed within the network, it continuously monitors its state which includes metrics like traffic load, latency, and channel conditions, and takes actions to optimize them. The approach of implementing or deploying a RL agent is ideal for managing the rigid, clashing demands of 6G use cases. Future research is expected to explore techniques that puts into methods like multi-agent reinforcement learning (MARL) into a practical use. Here, multiple agents across the network must collaborate and compete to achieve an optimal outcome. Major setbacks these methods might face are to overcome designing reward functions that accurately represent complex network goals and scaling these solutions to the massive size and speed of 6G networks. In conclusion, combining federated learning and reinforcement learning is expected to unlock a new era of intelligent QoS management. Together, they will enable the creation of 6G networks that are more

***Corresponding Author:** Ramesh S

efficient and responsive. They would also be privacy-conscious and become capable of self-driven evolution to meet the demands of the future.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

The authors have no acknowledgments to declare.

REFERENCES

1. Gu, Z., She, C., Hardjawana, W., Lumb, S., McKechnie, D., Essery, T., & Vucetic, B. (2020). Knowledge-assisted deep reinforcement learning in 5G scheduler design: From theoretical framework to implementation. *arXiv preprint*, arXiv:2009.08346.
2. Dong, R., She, C., Hardjawana, W., Li, Y., & Vucetic, B. (2020). Deep learning for radio resource allocation with diverse quality-of-service requirements in 5G. *arXiv preprint*, arXiv:2004.00507.
3. Khan, S., Khan, S., Ali, Y., Khalid, M., Ullah, Z., & Mumtaz, S. (2021). Highly accurate and reliable wireless network slicing in 5th generation networks: A hybrid deep learning approach. *arXiv preprint*, arXiv:2111.09416.
4. Ghuge, M., Ranjan, N., Mahajan, R. A., Upadhye, P. A., Shirkande, S. T., & Bhamare, D. (2023). Deep learning driven QoS anomaly detection for network performance optimization. *Journal of Electrical Systems, 19*(2), 1–10.
5. Srinivasan, R., & Garcia-Luna-Aceves, J. J. (2023). Optimized network coding with real-time loss prediction for hybrid 5G networks. *Lecture Notes in Computer Science, 13853*, 82–97.
6. Parada, R., Abu-Helalah, E., Serra, J., Aguilar, A., & Dini, P. (2025). Enhancing 5G O-RAN communication efficiency through AI-based latency forecasting. *arXiv preprint*, arXiv:2502.18046.
7. Biernacki, A. (2024). Throughput prediction of 5G network based on trace similarity for adaptive video. *Applied Sciences, 14*(5), 1962.
8. Anonymous. (2025). Probabilistic delay forecasting in 5G using recurrent and attention-based architectures. *arXiv preprint*, arXiv:2503.15297.
9. Alsulami, H., Serbaya, S. H., Abualsauod, E. H., Othman, A. M., Rizwan, A., & Jalali, A. (2022). A federated deep learning empowered resource management method to optimize 5G and 6G quality of services (QoS). *Wireless Communications and Mobile Computing, 2022*(1), 1352985.
10. Qian, Y., Liu, G., & Shi, H. (2024). Deep learning method to optimize the quality of services in UAV-aided 5G/6G networks. *Wireless Networks, 30*(8), 7003–7013.
11. Fadlullah, Z. M., Mao, B., & Kato, N. (2022). Balancing QoS and security in the edge: Existing practices, challenges, and 6G opportunities with machine learning. *IEEE Communications Surveys & Tutorials, 24*(4), 2419–2448.
12. Chahbi, K., Mahfoudi, M., El Mendili, F., Fattah, M., Mazer, S., & El Bakkali, M. (2024). Enhancing quality of service in 5G/6G networks: Innovative approaches. *Lecture Notes in Computer Science*, 129–136.
13. Mahmood, A. N., Mahmud, M. N., & Salleh, M. F. M. (n.d.). Challenges and solutions for video streaming services quality of experience (QoE) prediction toward 6G wireless networks.
14. Ahmed, R., Mahmood, M. R., & Matin, M. A. (2023). Challenges in meeting QoS requirements toward 6G wireless networks: A state of the art survey. *Transactions on Emerging Telecommunications Technologies, 34*(2), e4693.
15. Alhammadi, A., Shayea, I., El-Saleh, A. A., Azmi, M. H., Ismail, Z. H., Kouhalvandi, L., & Saad, S. A. (2024). Artificial intelligence in 6G wireless networks: Opportunities, applications, and challenges. *International Journal of Intelligent Systems, 2024*(1), 8845070.
16. Maragatharajan, M., Sureshkumar, A., Dhanaraj, R. K., Nirmala, E., Sayeed, M. S., Quasim, M. T., & Basheer, S. (2025). Hybrid feed forward neural networks and particle swarm optimization for intelligent self-organization in the industrial communication networks. *IEEE Open Journal of the Communications Society*.

***Corresponding Author:** Ramesh S