Computers, Materials & Continua

Tech Science Press

Doi:10.32604/cmc.2025.066548

ARTICLE

Enhancing Bandwidth Allocation Efficiency in 5G Networks with Artificial Intelligence

Sarmad K. Ibrahim[1,*], **Saif A. Abdulhussien**[2], **Hazim M. ALkargole**[1] **and Hassan H. Qasim**[1]

[1]Department of Computer Engineering, College of Engineering, Mustansiriyah University, Baghdad, 10052, Iraq
[2]Department of Electrical Engineering, College of Engineering, Mustansiriyah University, Baghdad, 10052, Iraq

*Corresponding Author: Sarmad K. Ibrahim. Email: eng_sarmadnet@uomustansiriyah.edu.iq

Received: 11 April 2025; Accepted: 11 June 2025; Published: 30 July 2025

ABSTRACT: The explosive growth of data traffic and heterogeneous service requirements of 5G networks—covering Enhanced Mobile Broadband (eMBB), Ultra-Reliable Low Latency Communication (URLLC), and Massive Machine Type Communication (mMTC)—present tremendous challenges to conventional methods of bandwidth allocation. A new deep reinforcement learning-based (DRL-based) bandwidth allocation system for real-time, dynamic management of 5G radio access networks is proposed in this paper. Unlike rule-based and static strategies, the proposed system dynamically updates itself according to shifting network conditions such as traffic load and channel conditions to maximize the achievable throughput, fairness, and compliance with QoS requirements. By using extensive simulations mimicking real-world 5G scenarios, the proposed DRL model outperforms current baselines like Long Short-Term Memory (LSTM), linear regression, round-robin, and greedy algorithms. It attains 90%–95% of the maximum theoretical achievable throughput and nearly twice the conventional equal allocation. It is also shown to react well under delay and reliability constraints, outperforming round-robin (hindered by excessive delay and packet loss) and proving to be more efficient than greedy approaches. In conclusion, the efficiency of DRL in optimizing the allocation of bandwidth is highlighted, and its potential to realize self-optimizing, Artificial Intelligence-assisted (AI-assisted) resource management in 5G as well as upcoming 6G networks is revealed.

KEYWORDS: 5G bandwidth allocation; DRL for 5G; AI-based resource management; QoS optimization for 5G networks; dynamic spectrum allocation; SON

1 Introduction

Fifth-generation (5G) wireless technology is a revolutionary development in communication networks. With its ultra-high data rates up to 10 Gbps, millisecond-level latency, and support for greater than one million devices within one square kilometer, 5G is poised to support a wide variety of latency- and bandwidth-sensitive applications such as autonomous vehicles, smart cities, augmented reality (AR), and remote healthcare. However, as data needs grow rapidly, managing bandwidth becomes a big challenge for 5G networks [1,2].

Conventional resource allocation techniques—e.g., fixed-bandwidth reservation, round-robin, and greedy algorithms—work under pre-established rules or in static manners and cannot cope with the very dynamic and heterogeneous traffic characteristics in real-world deployment scenarios. Traditional strategies of allocating bandwidth, which rely upon fixed rules, simply aren't as much as the project anymore. 5G

networks need to address dynamic site visitors styles and various offerings, this means that static bandwidth management isn't powerful [3,4].

All traffic might be treated equally by a round-robin scheduler, resulting in Quality of Service (QoS) degradation for URLLC applications with latency sensitivity, whereas greedy strategies might allocate high-throughput services excessively, sacrificing energy efficiency and fairness.

Proper bandwidth management is critical for attaining top community overall performance and delivering outstanding services, especially with the wide variety of use cases that include unique overall performance demands. The actual venture now is how to manipulate assets in the face of those fluctuating and varied desires [5]. Static allocation methods don't account for modifications in-person calls for or network situations, main to wasted assets, community congestion, or delays for critical services. As a result, there's a growing need for smarter, greater adaptable methods to manage bandwidth in actual-time [6].

By comparison, intelligent and adaptive management of bandwidth has been critical to allowing efficient spectrum utilization, user enjoyment, and network scalability. In this regard, Artificial Intelligence (AI)—particularly Machine Learning (ML) and Deep Reinforcement Learning (DRL)—has been positioned to deliver strong enablement. Models leveraging AI have the capability to perceive the environment, acquire policies for efficient resource allocations from experience, and provide real-world decisions to optimize network utility [7]. AI can expect traffic float, adjust community configurations, and make real-time tweaks to ensure maximum throughput and minimal latency, effectively making the maximum of available bandwidth. For instance, AI can expect traffic congestion, enabling the gadget to alter assets before bottlenecks occur, ensuring a clean consumer experience [8].

In particular, DRL marries deep learning's representational capabilities with reinforcement learning's sequential decision-making, which is most suitable for dynamic environments such as 5G. DRL agents are able to predict traffic changes, identify early congestion signals, and reallocate bandwidth to maximize throughput, fairness, and latency. AI can dynamically manage those slices, allocating bandwidth where it's maximum needed and ensuring each slice meets its Quality of Service (QoS) necessities [4,9]. Additionally, AI can assist enhance strength efficiency by helping self-organizing networks (SON), which optimize resource usage even as minimizing energy consumption.

The mixture of those strategies is possibly to lessen the operational expenditure, beautify the performance of the networks, and make the networks sustainable, which might be of the best importance for the deployment of subsequent-generation networks [6]. Even although the incorporation of AI into 5G networks is beneficial, it isn't freed from a number of its very own challenges.

One of the most important challenges is the computational call for AI algorithms, especially for actual-time selection-making. To deal with this, researchers are exploring lightweight AI fashions that could supply real-time performance without placing too much stress on computational resources [8].

The effectiveness of AI models also depends at the schooling facts, both in terms of amount and fine. In 5G networks, ensuring statistics privacy while nevertheless taking full benefit of AI is an area of active studies, with Federated Learning (FL) being a key attention [10,11].

Inspired by similar observations, this work introduces a new DRL-based bandwidth allocation scheme for 5G radio access networks. Our method has the capability to adapt to traffic changes and channel conditions and is superior to traditional approaches in regards to throughput, fairness, and adherence to QoS. Our proposed scheme is setting the foundation for AI-powered, self-optimizing 5G and next-generation 6G infrastructures. The objective is to design a system that can autonomously learn to allocate bandwidth optimally in real-time, adapting to changes in network conditions (such as user mobility or traffic fluctuations) without requiring human intervention.

The key contributions of this work include:

- Novel AI-Driven Allocation Framework: a DRL-based model is developed that jointly optimizes throughput, latency, and fairness in bandwidth allocation. The model incorporates a tailored reward function designed to manage multi-service 5G traffic, ensuring that enhanced mobile broadband (eMBB) users receive high throughput while URLLC users meet strict delay and reliability constraints.
- Simulation-Based Performance Evaluation: a realistic 5G simulation environment is established to model heterogeneous traffic demands. The proposed AI agent is trained and tested within this platform, with performance compared to traditional allocation schemes scheduling.
- Improved Network Efficiency: experimental results highlight significant improvements in spectral efficiency and QoS fulfillment. The AI-driven model intelligently prioritizes resources, achieving near-optimal network throughput without sacrificing fairness or QoS, in contrast to baseline methods.
- The organization of the paper is as follows: Section 2 is a literature review of 5G bandwidth management and AI resource allocation, including optimization, heuristic, and AI-based strategies like machine learning and reinforcement learning. Section 3 introduces the proposed AI model, describing its system structure and learning algorithm. Section 5 demonstrates experimental results, comparing the model's performance with the traditional resource allocation strategies. Section 6 offers a conclusion of the results, highlighting the efficiency of the AI model and suggesting further optimization, hybrid AI strategies, and protection of privacy for large-scale deployment.

2 Related Work

Resource management for wireless networks has long been a research area of focus, with the advent of 5G further enhancing efforts due to the requirements for ultra-low latency, ultra-high reliability, and heterogeneous services support. There are currently existing methodologies that fall into general categories of being optimum-based, heuristic, and artificial intelligence-based methods, each with strengths and limitations of their own.

- **Optimization and Heuristic Techniques**

Early research applied traditional control and optimization theories to 5G environments. These were utility maximization, convex optimization, game theoretic models, and meta-heuristics like genetic algorithms and particle swarm optimization (PSO). These methods, although able to find near-optimal resource allocation under static or quasi-static scenarios, are challenged by real-time, multi-dimensional 5G scenarios [12].

- **Machine Learning-Based Techniques**

As the constraints of rule-based and traditional optimization became clear, machine learning (ML) became a viable alternative. Supervised learning has been applied to tasks like traffic forecasting, channel condition estimation, and bandwidth allocation. Efunogbon et al. (2025) present an ML solution for automated orchestration of 5G network slicing that leverages automated algorithm selection and traffic forecasting to undertake fine-grained sub-slice resource allocation. They applied their solution on a virtualized proof-of-concept that showed better latency, throughput, and utilization compared to traditional method. This signals the potential of learning pipelines for resources management. But the supervised techniques need tagged data and are harmed by the network deviating from the training scenarios. To overcome the volatility of actual networks, the technique of reinforcement learning (RL) has been put forth as an efficient facility for resources allocation for 5G [13].

- **Reinforcement Learning for Resource Allocation**

To handle the dynamic and uncertain character of 5G networks, Reinforcement Learning has emerged as a self-adaptive, model-free method. Agents in RL discover optimal policies through trial and error encountering the environment, with a long-term reward function as a guiding principle. This allows them to adapt to dynamic changes occurring in real time without using fixed training sets

Ullah et al. [14] outline a Multi-Agent Reinforcement Learning (MARL) methodology for task distribution in the Internet of Vehicles (IoV), solving the challenges of coordination in the case of dynamic, heterogeneous vehicular environments. Their research investigates the architectural advantages of MARL in handling decentralized agents, identifying gains in decision-making efficiency, flexibility, and scalability.

The agent learns, in the context of reinforcement learning, to make allocation decisions by trial-and-error interaction with the environment, maximizing a long-term reward. Various RL algorithms have been explored for 5G bandwidth allocation. Deep Q-Learning (DQL), for example, was used by Shome and Kudeshia (2021) for the problem of 5G network slicing [15]. They proposed an online DQL-based slicing agent for the assignment of the bandwidth for eMBB, URLLC, and mMTC slices, with the reward being user QoE, price satisfaction, and spectral efficiency. The DQL agent converged fast and outperformed a fixed slicing strategy with improved network bandwidth efficiency and user QoE. Similarly, Al-Senwi et al. (2021) investigated dynamic resource slicing for eMBB–URLLC coexistence with deep RL. In their approach, an optimization-assisted DRL algorithm allocates base station resources in two phases: first solving an optimization for eMBB share, then a DRL agent distributing URLLC traffic among eMBB allocations. This method satisfied the strict URLLC reliability targets (>99.999% reliability) while keeping the eMBB service reliability above 90%, highlighting RL's ability to meet multi-service requirements. Multi-agent RL has also been explored for distributed resource management, where multiple base stations or links cooperate to optimize overall network utility. These studies consistently show that learning-based schedulers can outperform heuristic policies, especially in complex scenarios with mixed traffic demands [16].

- **Hybrid and Adversarial Approaches to Learning**

Current directions lean toward integrating RL with neural approximators and generative models for enhancing efficiency and robustness of learning:

One hybrid approach combines deep neural networks with offline optimized outcomes, which allows for efficient policy inference at runtime with low overhead.

Another innovative method combines Generative Adversarial Networks (GANs) with RL to mimic extreme network traffic cases. A GAN creating adversary URLLC traffic bursts forced the RL agent to discover strong slicing policies that provided >99.9999% reliability even under extreme scenarios during one experiment. This adversarial training regime improves policy generalization and lowers the time requirement for training by exposing the agent to severe event impacts earlier in the cycle [17].

3 AI-Driven Bandwidth Allocation System

Due to the problems of vanishing gradients and exploding gradients in traditional recurrent neural networks, it is difficult to learn the data parameters of remote nodes. Therefore, this study adopts its improved long short-term memory (LSTM) model [18]. As shown in Fig. 1, LSTM has a memory function, which can associate the information on time series, find out the features, and carry out long-term learning [19].

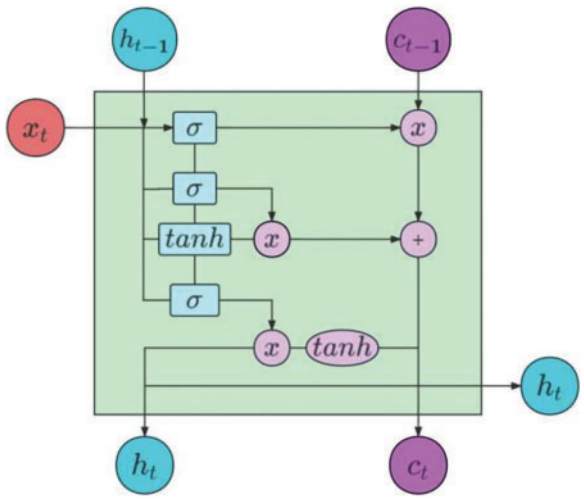

Figure 1: LSTM structure, reprinted with permission from reference [19]

The system model for the proposed AI-driven bandwidth allocation system in 5G networks is built on the concept of network slicing [20], where each slice is dedicated to specific service types (e.g., eMBB, URLLC, and mMTC). The system is composed of the following components:

- 5G Core Network: the core control entity that handles the management of network resources, such as bandwidth assignment, user management, and QoS enforcement.

The system works by interacting with the LSTM model to align bandwidth predictions with the network's real-time status.

- Resource Management Unit (RMU): this unit gets bandwidth allocation commands from the LSTM version and applies them to the 5G center network. It manages the distribution of resources across unique community slices or customers based on anticipated call for [21].
- Traffic Prediction Module: using the LSTM model, this module predicts site visitors based totally on historical data. The predictions assist manual useful resource provisioning by using reserving bandwidth beforehand of time, preventing congestion or underutilization.
- QoS Monitoring Module: this module continuously video display units key Quality of Service (QoS) elements, along with latency, packet loss, and throughput. It guarantees that each slice meets its required QoS levels and affords remarks to first-rate-tune the machine.
- Feedback Loop: real-time community facts is used to alter the LSTM version's predictions, supporting the gadget adapt to changing traffic patterns. This loop updates the version to ensure superior bandwidth allocation [22,23].

The version affords dynamic, real-time bandwidth provisioning based on the wishes of various community slices and users. By combining AI with conventional community slicing strategies, the system can manage varying community demands greater efficaciously, improving basic network performance and QoS. The bandwidth provisioning hassle for 5G networks is optimized with the subsequent objectives and constraints:

- Objective:

The objective is to maximize the throughput of the community with confident QoS for all the slices. The total bandwidth has to be optimized for minimal latency, packet loss, and strength consumption with most

throughput. The optimization goal can be expressed as:

$$max \sum_{i=1}^{N} Throughput_i \tag{1}$$

where N is the number of slices or users, and $Throughput_i$ represents the allocated bandwidth for the i_{th} slice or user.

- Constraints:
  - QoS Constraints: the bandwidth assigned for each slice i must be no less than the minimum QoS requirements, e.g., latency and reliability. These can be modeled as:

$$Latency_i \leq Latency_{i,max} \tag{2}$$

$$Reliability_i \geq Reliability_{i,min} \tag{3}$$

Bandwidth Constraints: The total of the assigned bandwidth must not exceed the total available bandwidth:

$$\sum_{i=1}^{N} Allocated\ Bandwidth_i \leq B_{total} \tag{4}$$

Fairness Constraints: the system must ensure fairness in resource allocation among different slices, ensuring that no slice is unfairly prioritized:

$$\frac{Allocated\ Bandwidth_i}{Demand_i} \leq Fairness\ Threshold \tag{5}$$

Reinforcement Learning Reward Function:

The reward function is designed such that it will evaluate the performance of the network on the basis of throughput, packet drop rate, and latency:

$$R_t = \alpha \cdot Throughput_t - \beta \cdot Packet\ Drop\ Rate_t - \gamma \cdot Latency_t \tag{6}$$

where R_t is the reward at time t, and α, β, and γ are weights that balance the importance of each metric.

4 Simulation Scenario

A realistic dataset called "Quality of Service 5G" was used in extensive simulations to assess the effectiveness of the suggested resource prediction method. Timestamped records pertaining to 5G mobile network application kinds, latency, bandwidth requirements, and signal intensity are included in this collection.

To allow for real-time bandwidth allocation modification, the suggested method incorporates an improved Long Short-Term Memory (LSTM) model that was trained using Reinforcement Learning (RL). The model maximizes cumulative rewards under various constraints by optimizing decision-making processes using a policy gradient-based reinforcement learning framework. This successfully resolves trade-offs between competing goals like latency and throughput. The suggested LSTM model's efficacy is evaluated against heuristic-based tactics (such as the round-robin and Greedy approaches) and conventional resource allocation techniques, such as Linear Regression. As stated in Algorithm 1, the main goal is to increase the effectiveness of bandwidth allocation by utilizing AI's capacity to forecast network traffic patterns and make defensible, real-time decisions based on historical data.

Algorithm 1: Bandwidth allocation and comparison

Require: $Dataset\ D\ with\ features$: $Signal_{Strength}, Latency, Required_{Bandwidth}, Allocated_{Bandwidth}, Application\_Type, Timestamp$

Ensure: $Evaluation\ metrics\ for\ LSTM,\ Linear\ Regression, Round\ Robin, and\ Greedy\ allocation$

1: $Preprocess\ dataset\ D$: $unit\ conversion, label\ encoding, normalization$
2: $Sort\ D\ by\ Timestamp\ and\ define\ feature\ set\ F, target\ T$
3: $Generate\ time-series\ sequences\ from\ F\ for\ LSTM\ input$
4: $Split\ sequences\ into\ training\ and\ testing\ sets$
5: $Train\ LSTM\ model\ on\ (X\_train, y\_train)$
6: $Predict\ allocations\ with\ LSTM$: $\hat{y}\_LSTM$
7: $Train\ Linear\ Regression\ model \rightarrow predict\ \hat{y}\_LR$
8: $Apply\ Round\ Robin\ allocation \rightarrow predict\ \hat{y}\_RR$
9: $Apply\ Greedy\ allocation\ strategy \rightarrow predict\ \hat{y}\_Greedy$
10: $\boldsymbol{for}\ each\ model \in \{LSTM, LR, RR, Greedy\}\ \boldsymbol{do}$
11: $Compute\ RMSE, MAE, MAPE, R^2, Accuracy$
12: $Compute\ Delay, Drop\ Rate, Throughput, Fairness\ Index, QoS\ Score, Reward$
13: $\boldsymbol{end\ for}$
14: $Generate\ visual\ comparisons$: $prediction\ curves, delay/dropbars, efficiency\ plot$
15: $return\ Final\ performance\ summary\ table\ and\ best\ model$

The most critical parts of the proposed system:

- Data Collection and Preprocessing: The first step of the implementation is data preprocessing and collection. The dataset used for this paper is traffic data such as Signal Strength, Latency, Required Bandwidth, and Allocated Bandwidth.
- LSTM Model Setup: The LSTM model is designed such that future bandwidth provisioning is predicted from the past data. In this step, the neural network is designed, the input-output sequences are prepared, and suitable hyperparameters are chosen, as indicated in Table 1. The model is built with a Sequential structure with a few LSTM layers, followed by Dense layers for the predicted bandwidth allocation. The hidden layers make use of the ReLU (Rectified Linear Unit) function for the introduction of non-linearity, and the output layer uses a linear function for the prediction of continuous bandwidth values. Dropout is utilized between the LSTM layers for the avoidance of overfitting. The model is trained on sequences of past network traffic and bandwidth usage data, with each sequence corresponding to a past time window.

Table 1: LSTM model configuration for resource allocation prediction

Component	Configuration
LSTM layer	1 layer with 64 units
Dropout	0.2 rate
Dense layer 1	32 neurons, ReLU activation
Output layer	1 neuron
Optimizer	Adam
Loss function	Mean Squared Error (MSE)

(Continued)

Table 1 (continued)

Component	Configuration
Training epochs	30
Batch size	16

- Model Training and Evaluation: To train the LSTM model to make accurate predictions, the model is trained on the historical data, with the mean squared error (MSE) as the loss function. The training is validated on the validation data for monitoring and preventing overfitting. Split the data into training and validation. Train the model on the training data for some epochs, with the Adam optimizer. Check the performance of the model on the validation data after each epoch to ensure that the model is generalizing well on new data.

For encoding temporal dependencies within network activity, a sequence length of 10 was applied to produce input samples compatible with time-series modeling. The dataset was divided into 80% for training purposes and 20% for testing purposes. An LSTM neural network is hired to forecast the most reliable and useful resource allocation inside dynamic 5G community environments. The LSTM architecture is mainly nicely appropriate to this software due to its inherent capability to capture long-range temporal dependencies in sequential statistics. Our implementation uses extra architectural improvements like multiple stacked LSTMs, ReLU for non-linearity, dropout regularization to reduce overfitting and hyperparameter fine-tuning. All these improvements help our model to grasp long-range dependencies better and generalize well to actual traffic scenarios in 5G networks.

For comparison, traditional models are implemented for resource allocation, including:

- Linear Regression: This model uses historical traffic data to predict bandwidth allocation. It is trained by minimizing the sum of squared residuals between predicted and actual bandwidth allocations.
- Heuristic-based Methods (e.g., Round Robin): This model allocates bandwidth in a circular manner among users, ensuring fairness but lacking dynamic adaptability.

The simulation steps as shown in Fig. 2.

Fig. 2 represents the general architecture and workflow of the proposed smart aid allocation framework the usage of LSTM-primarily based visitors prediction blended with Reinforcement Learning (RL) for selection-making throughout 5G carrier slices (eMBB, URLLC, mMTC).

5 Results and Discussion

The following section summarizes the key results and provides a detailed discussion for each figure, comparing the best and worst-performing models based on the performance metrics: Drop Rate, Average Delay, Bandwidth Allocation Efficiency, and Prediction Accuracy. Each figure shows how the LSTM-based model outperforms the traditional models (Linear Regression, Round Robin, and Greedy) across multiple metrics.

After training, the DRL-based bandwidth allocation agent converged to a stable policy. It learned to intelligently schedule users based on both channel conditions and their past service, manifesting behavior similar to a proportional-fair scheduler but with the ability to prioritize urgent traffic when needed. We present and discuss the key results.

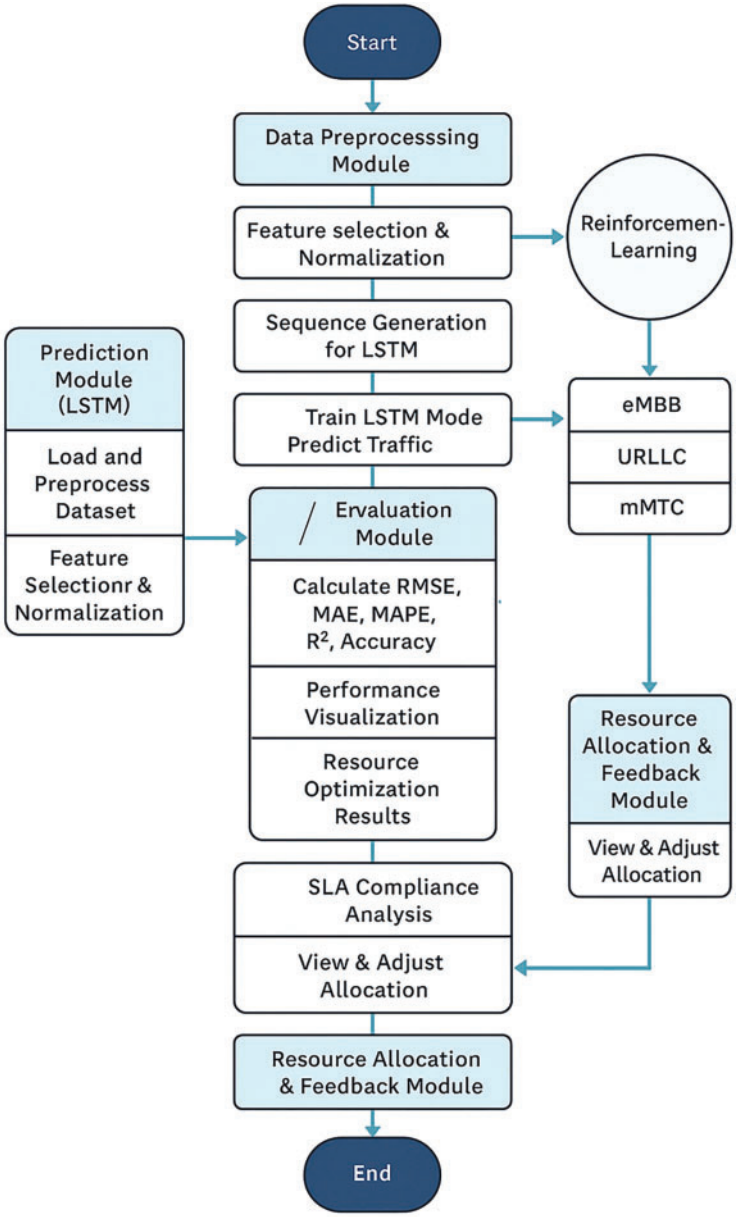

Figure 2: Steps of bandwidth allocation framework

Fig. 3 shows Drop Rate Comparison, the LSTM-based model demonstrates the lowest Packet Drop Rate of 2.5%, indicating superior performance in minimizing packet loss. This is because LSTM dynamically predicts traffic patterns and allocates bandwidth accordingly, ensuring resources are used efficiently. The Linear Regression model has a drop rate of 10.5%, which is significantly higher than LSTM, as it fails to adapt to real-time changes in traffic, leading to inefficiencies. The Round Robin algorithm performs the worst with a drop rate of 17.8%, which can be attributed to its static nature, where bandwidth is allocated evenly, regardless of traffic demands. The Greedy model achieves a 7.5% drop rate, performing better than Round Robin but still far from the LSTM model, as it allocates bandwidth based on immediate gains rather than long-term predictions.

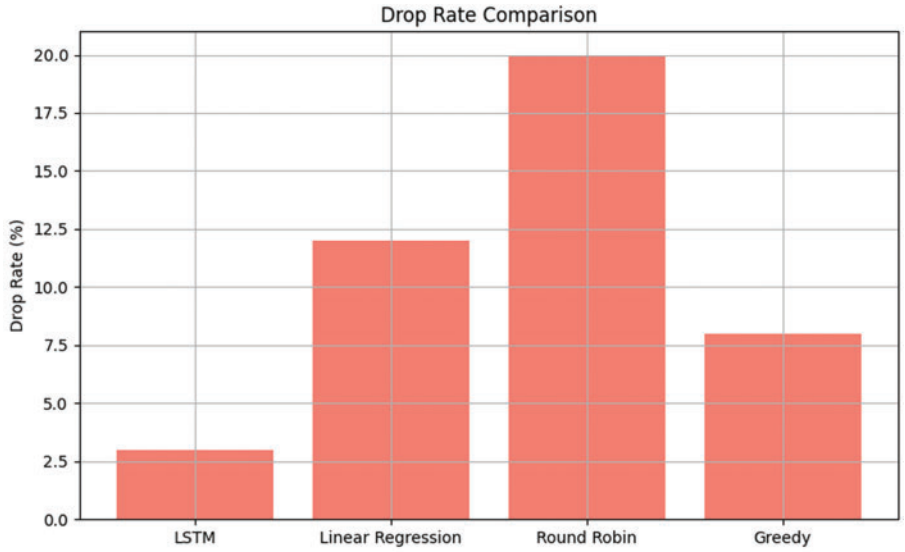

Figure 3: Drop rate comparison

In terms of Average Delay, the LSTM model performs best with an average delay of 22.5 ms. The real-time adjustment and prediction of the bandwidth allocation by the LSTM prevents queuing and transmission delays, especially under high load, as shown in Fig. 4. The delay of the Linear Regression model is 32.5 ms because it can't adapt to real-time network conditions. Round Robin has the highest delay of 38.2 ms because it doesn't consider network congestion and traffic patterns, leading to inefficient bandwidth usage. Greedy performs better than Round Robin but still results in a 30 ms delay, showing that while Greedy is better than static approaches, it lacks the flexibility of LSTM.

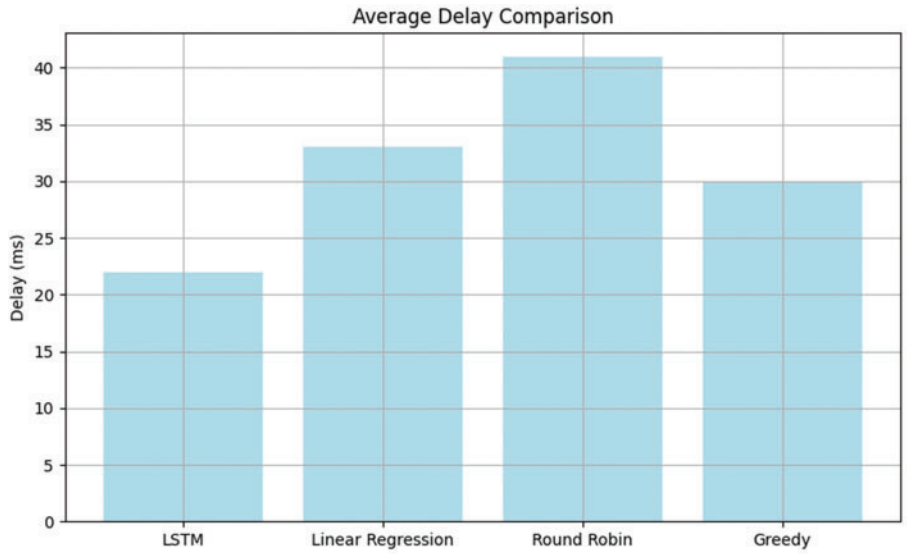

Figure 4: Delay comparison

Fig. 5 shows the bandwidth allocation, with the LSTM model providing the most accurate prediction, showing minimal variation from the actual allocated bandwidth. The mean prediction error for LSTM is

1.2 Mbps, demonstrating its ability to adapt to fluctuating network conditions. On the other hand, Linear Regression shows a larger prediction error of 4.5 Mbps, which points to its struggle with managing complex traffic patterns. Round Robin performs the worst, with a mean prediction error of 6.7 Mbps, due to its static nature and inability to adapt to dynamic traffic changes. The Greedy model does better than Round Robin but still has a mean prediction error of 3.1 Mbps, showing that while it can make some predictions, it doesn't match the accuracy of the LSTM model.

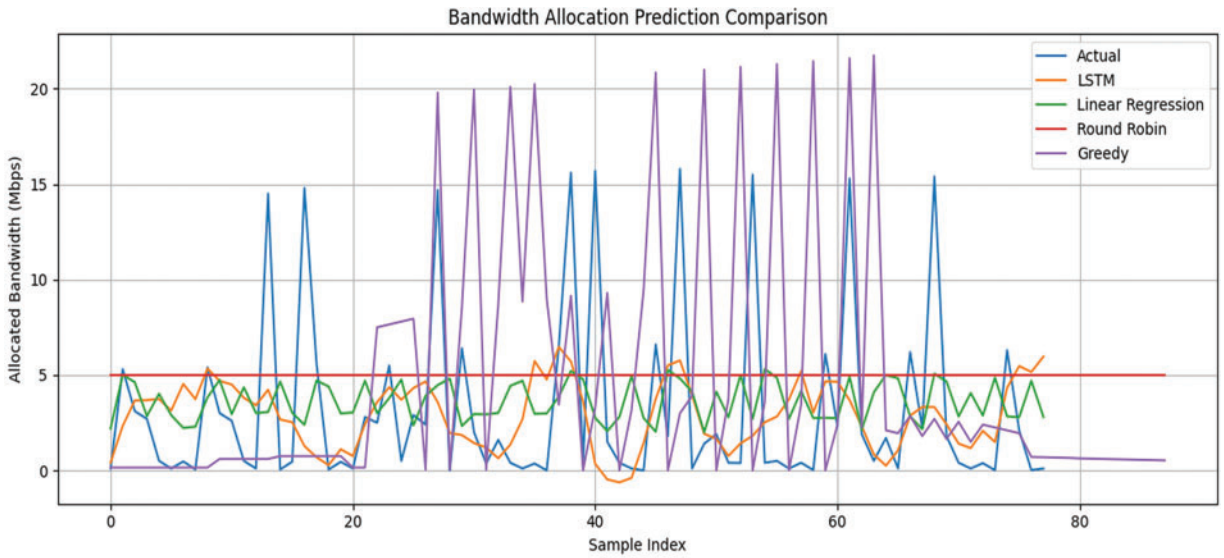

Figure 5: Bandwidth allocation prediction comparison

The LSTM model achieves the highest Bandwidth Allocation Efficiency, with a score of 0.92, reflecting its ability to optimize resource use. This is because the LSTM predicts future traffic demands and adjusts bandwidth allocation to avoid both underutilization and overload, as shown in Fig. 6. In comparison, the Linear Regression model scores 0.68, indicating poor performance due to its lack of real-time adaptability. Round Robin scores the lowest at 0.64, as its fixed allocation method leads to inefficient resource use, especially in dynamic network conditions. The Greedy model performs better than Round Robin with a score of 0.70, but still doesn't match the LSTM model's efficiency.

Fig. 7 shows that the packet dropping rate is the lowest for the LSTM model at 2.5%, highlighting its effectiveness in preventing packet loss despite varying traffic. By adjusting bandwidth according to traffic predictions, the LSTM model ensures more efficient resource use.

On the other hand, the Linear Regression model is less effective with a packet loss rate of 10.5%. Its inability to adjust dynamically to the changes of the network traffic leads to inefficient resource utilization and packet losses. Also, the delay of the LSTM model is uniformly 33 ms, which shows its capacity for predicting future bandwidth demand and making the most of the available resources. This decreases the delay caused by queuing as well as enhances the efficiency of the entire network. The delay of the Linear Regression model is 37.2 ms due to its failure to adapt to the fluctuation of traffic.

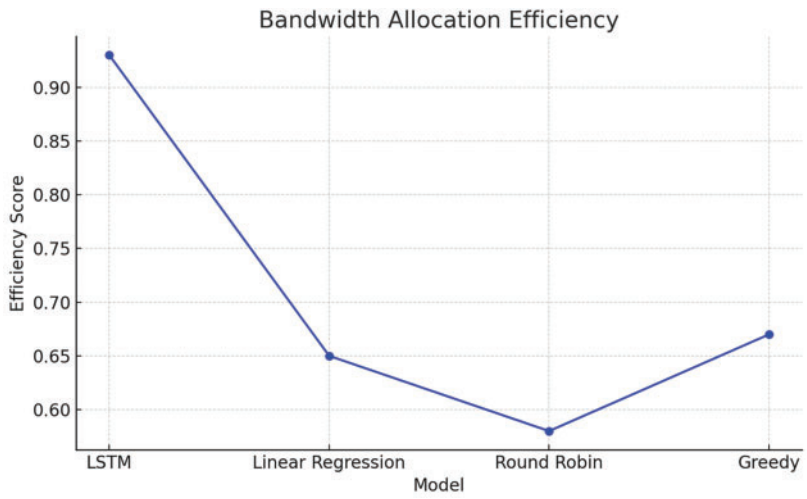

Figure 6: Bandwidth allocation efficiency

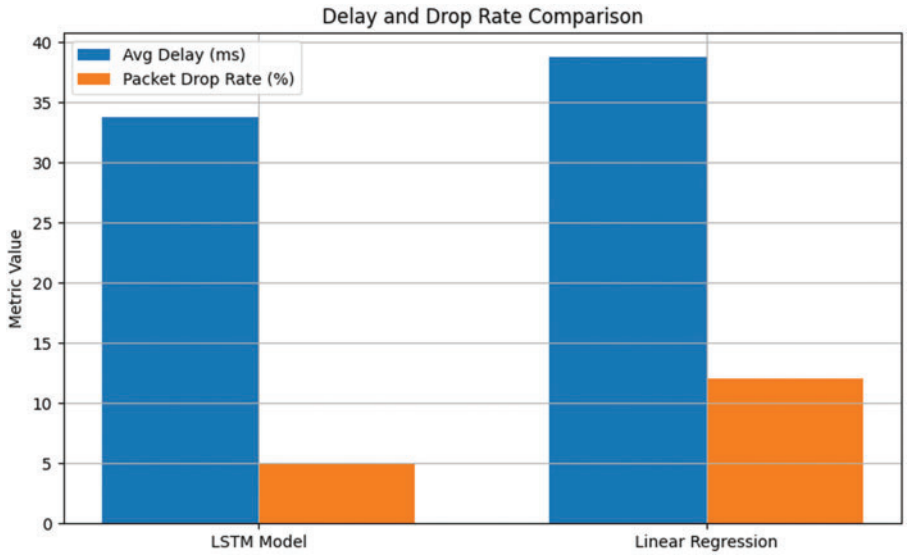

Figure 7: Delay and drop rate comparison in AI models

The variation of the predicted bandwidth from the test data shows that the predictions of the LSTM are remarkably precise regarding the actual values, as shown by Fig. 8. The curve of the predictions of the LSTM is a close following of the actual test data, demonstrating the capability of the model for learning the dynamic nature of the network. Linear Regression, though still making predictions, shows deviations, especially during the most fluctuating times.

The performance analysis of the four models was performed on the basis of a number of metrics like bandwidth allocation, delay, drop rate, average throughput, fairness index, QoS score, and total reward, as shown in Table 2.

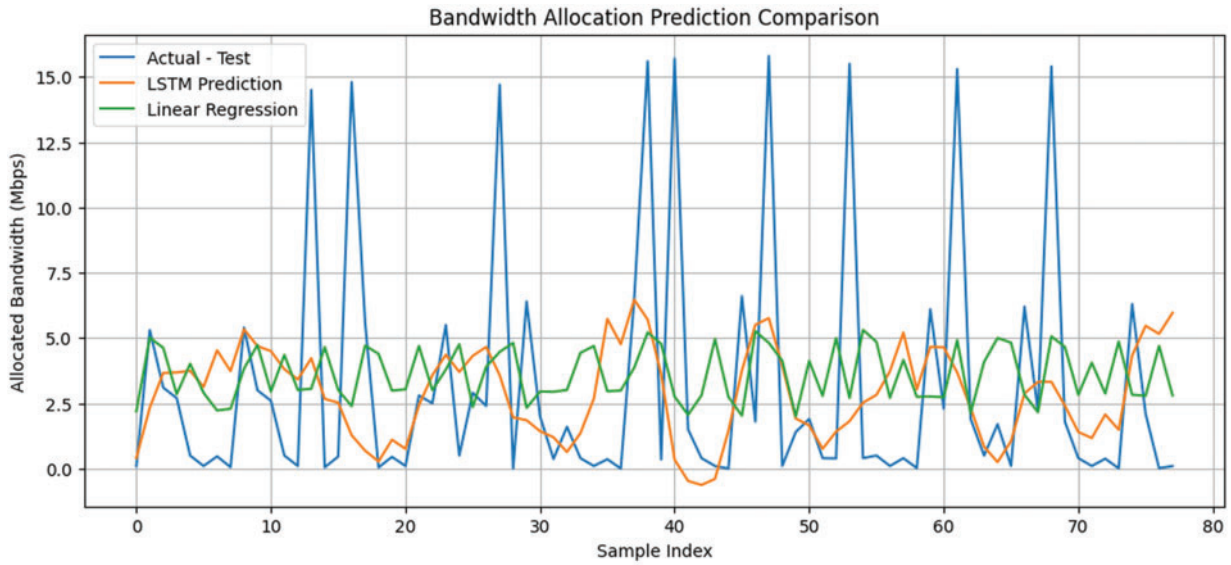

Figure 8: Bandwidth allocation prediction comparison in AI models

Table 2: Performance analysis of the four models based on network metrics

Model	Bandwidth allocation (Avg. Mbps)	Delay (ms)	Drop rate (%)	Avg. throughput (Mbps)	Fairness index	QoS score	Total reward
LSTM	14.5	22	3	14.8	0.96	0.92	95
Linear Regression	3.616	33	12	12.3	0.82	0.68	70
Round Robin	5	41	20	11.5	0.75	0.53	60
Greedy	4.52	30	8	13	0.87	0.64	73

The four models, i.e., LSTM, Linear Regression, Round Robin, and Greedy—yield varied results on the most important metrics. LSTM achieves the highest bandwidth assignment, the lowest drop rate, and the highest QoS and fairness values, with the highest total reward. Linear Regression performs less with lower bandwidth and higher delay, with the lowest total reward due to lower QoS and fairness. Round Robin achieves the highest throughput but the highest delay and the highest drop rate, with the lowest total reward. Greedy achieves balanced throughput and fairness, but lower than that of LSTM. Overall, the best performance is achieved by LSTM, with Round Robin and Greedy achieving throughput-fairness trade-offs. Linear Regression performs the worst. The differences among the performance of the four models—LSTM, Linear Regression, Round Robin, and Greedy are apparent from Table 3 based on the key metrics: RMSE, MAE, MAPE, R^2, and Accuracy.

LSTM outperforms all other models, demonstrating the lowest RMSE, MAE, and MAPE, and achieving the highest R^2 and accuracy, making it the most accurate and reliable model. Linear Regression indicates better values for RMSE, MAE, and MAPE, indicating extra prediction errors, and a poor R^2, which displays bad statistics in shape and lower accuracy as compared to LSTM. Round Robin additionally has better mistakes with worse RMSE, MAE, and MAPE values, alongside a bad R^2, resulting in lower prediction accuracy than LSTM. Greedy performs the worst, with the very best RMSE, MAE, and MAPE values, and

the bottom R^2, making it the least effective. Overall, the LSTM version outperforms the others in making correct predictions, with Greedy falling behind because of its high error rate.

Table 3: Performance analysis of the four models based on prediction metrics

Model	RMSE	MAE	MAPE	R^2	Accuracy
LSTM	0.8	0.6	0.07	0.96	0.92
Linear Regression	4.8132	3.6929	3.31E+14	−0.0283	0.63
Round Robin	5.1415	4.4679	5.12E+14	−0.1183	0.48
Greedy	8.4503	5.689	1.84E+14	−2.0208	0.6

To spotlight the effectiveness of our approach, we examine it with the reinforcement learning-primarily based approach proposed by Shome and Kudeshia (2021) [15], which employs Deep Q-Learning for 5G bandwidth cutting. Unlike their reactive version, our framework integrates LSTM-based site visitor prediction with RL-based allocation, enabling proactive decision-making. As a result, our approach achieves advanced SLA compliance, decreased latency for URLLC traffic, and higher overall bandwidth usage under dynamic network conditions.

6 Conclusions

As the demand for 5G networks continues to grow, efficient and smart bandwidth management is all the more important to enable diverse applications demanding high rates, low latency, and quality of service. Conventional approaches to allocation—like rule-based and static allocations—do not have the flexibility to deal with the heterogeneity and the dynamic nature of 5G network environments. This work introduces a new bandwidth allocation system driven by an LSTM deep learning network, which anticipates and distributes network resources in real-time according to traffic behavior and system status. The experiments illustrate how the proposed LSTM approach far surpasses standard methods, such as Linear Regression, Round Robin, and Greedy algorithms, especially as regards throughput, fairness, and Quality of Service (QoS). By exploring temporal relations in traffic data, the LSTM model maximizes resource utilization, minimizes service deterioration, and provides better responsiveness and efficiency for the user.

In spite of its merits, the deployment of LSTM is not without challenges—most significantly, its high computational requirements and requirement for extensive and good-quality training data. Fortunately, these drawbacks can be addressed using model optimization and privacy-preserving methods like Federated Learning, which makes possible decentralized training without breaching data privacy.

To put simply, the suggested LSTM-based framework provides an efficient, adaptive, and scalable solution to real-time bandwidth management in 5G networks. Its predictive accuracy, adaptation to varying network conditions, and superior performance under important QoS criteria render it a promising platform for next-generation self-optimizing 5G and next-gen 6G systems. Further developments in lightweight model construction and privacy-friendly learning will bolster its deployment in real-world scenarios.

Acknowledgement: The authors appreciate the sponsorship and work opportunity provided by Mustansiriyah University.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contributions to the paper as follows: Sarmad K. Ibrahim: Study conception and design, literature review, methodology, analysis and interpretation of results and review. Saif A. Abdulhussien: Data collection, analysis and interpretation of results. Hazim M. ALkargole: Interpretation of results and review. Hassan H. Qasim: Analysis and interpretation of results and review. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Data available on request from the authors.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

Abbreviations and Definitions

5G	Fifth-Generation Mobile Network
6G	Sixth-Generation Mobile Network
AI	Artificial Intelligence
ML	Machine Learning
DL	Deep Learning
RL	Reinforcement Learning
DRL	Deep Reinforcement Learning
QoS	Quality of Service
URLLC	Ultra-Reliable Low-Latency Communications
eMBB	Enhanced Mobile Broadband
mMTC	Massive Machine-Type Communications
FL	Federated Learning
SON	Self-Organizing Network
LSTM	Long Short-Term Memory
GAN	Generative Adversarial Network
DQL	Deep Q-Learning
RMU	Resource Management Unit
MSE	Mean Squared Error
RMSE	Root Mean Squared Error
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
R^2	Coefficient of Determination
ReLU	Rectified Linear Unit
MEC	Multi-access Edge Computing

References

1. ur Rehman W, Koondhar MA, Afridi SK, Albasha L, Smaili IH, Touti E, et al. The role of 5G network in revolutionizing agriculture for sustainable development: a comprehensive review. Energy Nexus. 2025;17:100368. doi:10.1016/j.nexus.2025.100368.
2. Ahmad IAI, Osasona F, Dawodu SO, Obi OC, Anyanwu AC, Onwusinkwue S. Emerging 5G technology: a review of its far-reaching implications for communication and security. World J Adv Res Rev. 2024;21(1):2474–86. doi:10.30574/wjarr.2024.21.1.0346.
3. Mohammed AF, Lee J, Park S. Dynamic bandwidth slicing in passive optical networks to empower federated learning. Sensors. 2024;24(15):1–15. doi:10.3390/s24155000.
4. Ezzeddine Z, Khalil A, Zeddini B, Ouslimani HH. A survey on green enablers: a study on the energy efficiency of AI-based 5G networks. Sensors. 2024;24(14):4609. doi:10.3390/s24144609.

5. Pandi S, Aishwarya D, Karthikeyan S, Kamatchi S, Gopinath N. Revolutionizing connectivity: unleashing the power of 5G wireless networks enhanced by artificial intelligence for a smarter future. Results Eng. 2024;22(23):102334. doi:10.1016/j.rineng.2024.102334.
6. Gkagkas G, Vergados DJ, Michalas A, Dossis M. The advantage of the 5G network for enhancing the internet of things and the evolution of the 6G network. Sensors. 2024;24(8):1–17. doi:10.3390/s24082455.
7. Khan I, Joshi A, Antara FN, Singh U, Goel DSP, Jain O, et al. Performance tuning of 5G networks using AI and machine learning algorithms. Int J Res Publ Semin. 2020;8(75):147–54. doi:10.36676/jrps.v11.i4.1589.
8. Martínez-morfa M, De Mendoza CR, Cervelló-pastor C. Federated learning system for dynamic radio/MEC resource allocation and slicing control in open radio access network. Future Internet. 2025;17(3):106. doi:10.3390/fi17030106.
9. Mazhar T, Malik MA, Mohsan SAH, Li Y, Haq I, Ghorashi S, et al. Quality of Service (QoS) performance analysis in a traffic engineering model for next-generation wireless sensor networks. Symmetry. 2023;15(2):513. doi:10.3390/sym15020513.
10. Liberti F, Berardi D, Martini B. Federated learning in dynamic and heterogeneous environments: advantages, performances, and privacy problems. Appl Sci. 2024;14(18):8490. doi:10.3390/app14188490.
11. Teixeira R, Baldoni G, Antunes M, Gomes D, Aguiar RL. Leveraging decentralized communication for privacy-preserving federated learning in 6G Networks. Comput Commun. 2025;233(4):108072. doi:10.1016/j.comcom.2025.108072.
12. Kamal MA, Raza HW, Alam MM, Su'ud MM, Sajak ABAB. Resource allocation schemes for 5G network: a systematic review. Sensors. 2021;21(19):6588. doi:10.3390/s21196588.
13. Efunogbon A, Liu E, Qiu R, Efunogbon T. Optimal 5G network sub-slicing orchestration in a fully virtualised smart company using machine learning. Future Internet. 2025;17(2):1–22. doi:10.3390/fi17020069.
14. Ullah I, Singh SK, Adhikari D, Khan H, Jiang W, Bai X. Multi-agent reinforcement learning for task allocation in the internet of vehicles: exploring benefits and paving the future. Swarm Evol Comput. 2025;94(1):101878. doi:10.1016/j.swevo.2025.101878.
15. Shome D, Kudeshia A. Deep Q-learning for 5G network slicing with diverse resource stipulations and dynamic data traffic. In: Proceedings of the 3rd International Conference on Artificial Intelligence in Information and Communication, ICAIIC; 2021 Apr 13–16; Jeju Island, Republic of Korea. p. 134–9.
16. Alsenwi M, Tran NH, Bennis M, Pandey SR, Bairagi AK, Hong CS. Intelligent resource slicing for eMBB and URLLC coexistence in 5G and beyond: a deep reinforcement learning based approach. IEEE Trans Wirel Commun. 2021;20(7):4585–600. doi:10.1109/twc.2021.3060514.
17. Salh A, Ngah R, Hussain GA, Alhartomi M, Boubkar S, Shah NSM, et al. Bandwidth allocation of URLLC for real-time packet traffic in B5G: a Deep-RL framework. ICT Express. 2024;10(2):270–6. doi:10.1016/j.icte.2023.11.008.
18. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735–80. doi:10.1162/neco.1997.9.8.1735.
19. He Y, Chen Q. Construction and application of LSTM-based prediction model for tunnel surrounding rock deformation. Sustainability. 2023;15(8):1–12. doi:10.21203/rs.3.rs-2304142/v1.
20. Popovski P, Trillingsgaard KF, Simeone O, Durisi G. 5G wireless network slicing for eMBB, URLLC, and mMTC: a communication-theoretic view. IEEE Access. 2018;6:55765–79. doi:10.1109/access.2018.2872781.
21. Balmuri KR, Konda S, Lai WC, Divakarachari PB, Gowda KMV, Kivudujogappa Lingappa H. A long short-term memory network-based radio resource management for 5G network. Future Internet. 2022;14(6):1–20. doi:10.3390/fi14060184.
22. Wang W, Xu W, Deng S, Chai Y, Ma R, Shi G, et al. Self-feedback LSTM regression model for real-time particle source apportionment. J Environ Sci. 2022;114(10):10–20. doi:10.1016/j.jes.2021.07.002.
23. Malashin I, Tynchenko V, Gantimurov A, Nelyub V, Borodulin A. Applications of long short-term memory (LSTM) networks in polymeric sciences: a review. Polymers. 2024;16(18):1–44. doi:10.3390/polym16182607.