RESEARCHERID
THOMSON REUTERS

[Saxena *et al.*, 12(4): April, 2025] ISSN 2349-0292
Impact Factor 3.802

GLOBAL JOURNAL OF ADVANCED ENGINEERING TECHNOLOGIES AND SCIENCES

ENHANCING 5G NETWORK EFFICIENCY: AN AI-BASED APPROACH TO DYNAMIC RESOURCE ALLOCATION, ANOMALY DETECTION, AND LATENCY REDUCTION

Dr. Shivangini Saxena
Head of Department, Department of Electronics and Communication
SAGE University, Indore (M.P.), India
drshivangini.morya@sageuniversity.in

DOI: https://doi.org/10.29121/gjaets.2025.4.1

ABSTRACT

The increasing demands of next-generation broadband networks, particularly with the advent of 5G and beyond, necessitate advanced, scalable AI-driven frameworks for dynamic network management, resource allocation, and predictive maintenance. This paper introduces a scalable AI-powered framework that integrates deep learning models for optimizing network performance in the context of broadband networks, particularly 5G. The proposed framework includes the use of deep learning models such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Deep Q-Learning (DQN), to handle complex data, detect anomalies, predict failures, and optimize traffic routing. Additionally, edge AI is incorporated to reduce latency, enhance real-time decision-making, and optimize network responsiveness. We discuss the implementation workflow, key deep learning models, and future directions like Federated Learning and Transfer Learning that will further enhance the adaptability and scalability of broadband networks.

KEYWORDS: 5G, Convolutional Neural Networks, Deep Q-Learning, Recurrent Neural Networks.

1. INTRODUCTION

The broadband network landscape is undergoing a massive transformation driven by the explosion of connected devices, unprecedented growth in data traffic, and heightened expectations for real-time services. The introduction of 5G networks and the anticipated rollout of 6G will bring about even greater complexity, requiring a paradigm shift in network management. These next-generation networks promise faster speeds, ultra-low latency, and the ability to connect billions of devices. However, the benefits of such advanced networks come with a set of significant challenges that traditional network management techniques are ill-equipped to address.

One of the key challenges of 5G and beyond is the exponential increase in data traffic, with massive connectivity requirements from a wide array of devices, including smartphones, sensors, autonomous vehicles, and IoT devices. This creates a huge strain on existing network infrastructure and resource allocation systems. Additionally, the growing demand for high-quality, real-time services—such as video streaming, virtual reality (VR), augmented reality (AR), and cloud computing—requires networks to deliver ultra-low latency and high throughput without sacrificing reliability. These evolving requirements necessitate a new approach to managing and optimizing network resources, ensuring minimal delays, and maintaining seamless user experiences.

Traditional methods of network management, such as rule-based systems and static configurations, are no longer sufficient to handle these dynamic demands. These methods often fail to adapt to real-time changes in network conditions, making them inefficient for modern broadband networks. To overcome these limitations, artificial intelligence (AI), particularly deep learning and reinforcement learning (RL), has emerged as a transformative solution. AI offers a powerful means of automating network management by enabling real-time decision-making, predictive maintenance, anomaly detection, and dynamic resource allocation. These capabilities are essential for ensuring optimal network performance, reducing operational costs, and improving user satisfaction in the age of 5G and future networks.

AI-powered solutions, especially those utilizing deep learning models like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), are capable of processing large amounts of network data and making predictions based on complex patterns. For example, CNNs have been widely used for anomaly detection by analyzing network traffic patterns and identifying deviations that might signal issues such as congestion, security breaches, or hardware malfunctions. RNNs, including Long Short-Term Memory (LSTM) networks,

excel in time-series forecasting, making them ideal for predicting future network traffic and managing dynamic resource allocation in real-time.
Additionally, edge AI plays a pivotal role in addressing the low-latency requirements of 5G networks. By processing data closer to the source—at the network edge, near users and devices—edge AI enables faster decision-making and reduces the need for data to travel long distances to central cloud servers. This significantly improves the responsiveness of real-time applications, such as autonomous driving, remote healthcare, and live video streaming, where even milliseconds of delay can have a major impact on user experience. The integration of edge AI in broadband networks is crucial for the success of 5G and future networks, as it minimizes latency, reduces bandwidth consumption, and enables more efficient network management.

This paper proposes a scalable AI-powered framework designed specifically for next-generation broadband networks, with a focus on 5G and beyond. The framework integrates advanced AI techniques such as reinforcement learning for adaptive network optimization, deep learning for anomaly detection, and edge AI for real-time decision-making. The integration of reinforcement learning allows the network to continuously learn and adapt to changing traffic conditions and optimize resource allocation (e.g., bandwidth, routing, and load balancing) based on real-time data. By combining these AI models with edge computing capabilities, the proposed framework aims to enhance network performance, reduce latency, and improve the overall user experience in a dynamic and resource-constrained environment.

2. LITERATURE REVIEW

In the early stages, traditional machine learning models such as Support Vector Machines (SVMs), Decision Trees, and Random Forests were widely used to optimize broadband network performance. These models were effective for tasks like traffic classification, network fault detection, and user behavior analysis [1]. However, traditional models often struggled to handle the temporal dependencies inherent in network traffic patterns, which is where deep learning models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have shown significant promise. CNNs are particularly effective in anomaly detection in network traffic, as they can learn spatial hierarchies in data and detect patterns that deviate from normal network behavior [2]. For example, the authors of [3] demonstrated how CNNs could analyze network traffic to identify security threats and congestion by learning compact representations of normal behavior and identifying outliers.

RNNs and their more advanced variant, Long Short-Term Memory (LSTM) networks, are widely used for time-series prediction tasks, such as predicting future network traffic and detecting network failures. RNNs and LSTMs have the advantage of learning temporal dependencies in data, which is critical in environments like 5G where network conditions change rapidly [4]. These models have been successfully used in traffic forecasting, enabling networks to adjust dynamically to varying loads. Additionally, Autoencoders have emerged as an essential tool for unsupervised anomaly detection. By learning a compact representation of normal network behavior, autoencoders can flag deviations, such as hardware malfunctions or security breaches. Their ability to detect anomalies in real-time is crucial for ensuring network reliability and proactively preventing service disruptions [5].

Reinforcement Learning (RL) has gained traction in the field of broadband network optimization due to its ability to make dynamic, data-driven decisions. Deep Q-Learning (DQN) and Proximal Policy Optimization (PPO) are two RL techniques that have been particularly effective in adaptive network optimization. DQN has been used to dynamically allocate resources such as bandwidth and routing in real-time, adjusting network parameters based on observed traffic conditions and rewards [6]. This allows networks to optimize performance continuously, without the need for pre-configured rules.

PPO, a policy-gradient method, enables direct optimization of the network's policies. PPO balances exploration and exploitation, ensuring that the network can adapt quickly to real-time changes while minimizing the risk of large, destabilizing policy updates [7]. PPO has been applied to traffic management and quality of service (QoS) optimization, ensuring that resources are allocated efficiently, even under fluctuating demand [8]. The ability of these reinforcement learning models to adapt in real-time makes them essential for next-generation networks like 5G, where the network must continuously learn from its environment and optimize based on changing traffic patterns and real-time conditions [9].

One of the critical challenges of 5G networks is the need to deliver ultra-low latency for real-time applications such as autonomous driving, smart cities, and telemedicine. To meet these challenges, the integration of edge AI

ISSN 2349-0292
Impact Factor 3.802

has become a focal point of research [10]. Edge AI refers to the deployment of AI models at the network edge, closer to the users or devices, enabling real-time decision-making without the need for data to be sent to centralized cloud servers.

Edge AI has proven beneficial in improving network responsiveness and reducing latency. By processing data locally, edge AI ensures that decisions, such as traffic management and resource allocation, can be made faster, reducing the time it takes for data to travel to and from cloud servers. This is especially critical for mission-critical applications in 5G, where delays can have severe consequences [11]. Studies have shown that the deployment of edge AI can significantly enhance network performance by improving real-time data analysis and minimizing delays [12].

As broadband networks continue to scale, Federated Learning and Transfer Learning are emerging as key techniques for further enhancing the adaptability and scalability of AI models [13]. Federated Learning enables distributed model training across multiple network nodes without requiring data to be centralized. This ensures data privacy and security while still allowing the network to learn from diverse environments. Federated Learning is particularly valuable in scenarios where user data privacy is critical, such as in healthcare or financial services [14].

Transfer learning allows for the adaptation of pre-trained models to new network domains, reducing the amount of data and time required to train new models. For example, a model trained for 4G networks can be fine-tuned and adapted for 5G networks, speeding up the deployment of AI solutions and enabling quicker adaptation to new network topologies. This is especially important as the transition from 5G to 6G involves significant changes in network architecture and requirements [15].

3. PROPOSED METHODOLOGY

Next-generation broadband networks, such as 5G and beyond, bring new opportunities and challenges for network management. These networks are characterized by ultra-low latency, extremely high data throughput, and massive connectivity with millions of devices. In such an environment, traditional methods of managing network traffic, allocating resources, and detecting failures are insufficient.
To address these challenges, deep learning models can be used to:

1. Adapt the Framework for Future Networks: In 5G and beyond, AI models, such as Neural Architecture Search (NAS) and Reinforcement Learning (RL), can be employed to dynamically adapt to changing network conditions. These models enable the network to learn from its environment and adjust to varying demands, high-speed data, and unpredictable traffic patterns. Reinforcement learning techniques like Deep Q-Learning (DQN) and Proximal Policy Optimization (PPO) are particularly well-suited for adapting network resources in real-time, providing an adaptive and scalable solution.
2. Handle Complex Data and Resource Allocation: The ability to process massive amounts of real-time data is crucial for 5G networks. Deep learning models, especially CNNs and RNNs, can help in analyzing traffic data, detecting anomalies, and predicting future network conditions, all of which are essential for resource management.4. Key Deep Learning Models and Techniques

1. Convolutional Neural Networks (CNNs) for Anomaly Detection: CNNs can be applied to identify anomalies in network traffic patterns. By analyzing traffic data as multi-dimensional images, CNNs can detect unusual patterns that may indicate potential failures or malicious activity, such as Distributed Denial of Service (DDoS) attacks.
2. Recurrent Neural Networks (RNNs) and Long Short-Term Memory Networks (LSTMs): RNNs and LSTMs are ideal for time-series prediction, making them particularly effective for predicting future network traffic patterns, performance degradation, and system failures. These models capture temporal dependencies, allowing for accurate forecasting and failure prediction.

The proposed AI-powered framework integrates several advanced deep learning models, reinforcement learning algorithms, and optimization techniques to manage and optimize broadband network performance, particularly in the context of 5G and beyond. Below, we outline the key mathematical equations associated with the various models and techniques involved in the framework.

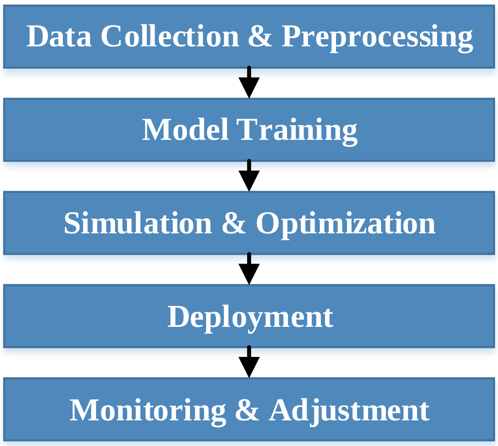

Figure 1: Flow Diagram for Proposed Methodology

Deep Learning Models for Anomaly Detection and Time-Series Prediction
Convolutional Neural Networks (CNNs) are used for anomaly detection in network traffic. CNNs are particularly effective in identifying patterns and anomalies in multi-dimensional traffic data. The output of a convolutional layer is given by:

$$y^{(l)} = f(W(l) * x^{(l-1)} + b^{(l)}) \tag{1}$$

Where:

- $y^{(l)}$ is the output of the$l$ -th layer.
- $W^{(l)}$ is the weight matrix for the l-th layer.
- $*$ represents the convolution operation.
- $x^{(l-1)}$ is the input from the previous layer.
- $b^{(l)}$ is the bias term for the l-th layer.
- f is an activation function (e.g., ReLU or Sigmoid).

For Recurrent Neural Networks (RNNs) and Long Short-Term Memory Networks (LSTMs), which are used for time-series prediction (e.g., predicting future network traffic), the equations governing the network dynamics are:

- For an RNN:

$$h_t = f(W_h x_t + U_h h_{t-1} + b_h) \tag{2}$$

  Where:
  - h_t is the hidden state at time t .
  - W_h and U_h are the weight matrices for the input x_t and previous hidden state h_{t-1}, respectively.
  - b_h is the bias term.
  - f is the activation function (e.g., tanh or ReLU).

- For an LSTM (using its gates mechanism):

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \tag{3}$$

$$(i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{4}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{5}$$

$$c_t = f_t * c_{t-1} + i_t * \tanh(W_c x_t + U_c h_{t-1} + b_c) \tag{6}$$

$$h_t = o_t * \tanh(c_t) \tag{7}$$

  Where:
  - f_t, i_t, o_t are the forget, input, and output gates, respectively.
  - $c_t$is the cell state at time t .
  - h_t is the output (hidden state) at time t.
  - W, U, and b represent the weight matrices and bias terms.

The proposed AI-powered framework integrates several deep learning models, reinforcement learning algorithms, and optimization techniques to improve the management and optimization of broadband networks, especially in the context of 5G and future networks. The combination of CNNs, RNNs, LSTMs, DQN, PPO, autoencoders, MAS, Federated Learning, and Transfer Learning offers a robust, scalable solution to handle complex network data, real-time decision-making, anomaly detection, traffic routing, and predictive maintenance.

Implementation Workflow

The AI-powered framework is implemented in several stages to ensure effective and scalable integration into broadband networks:

1. Data Collection & Preprocessing: Gather historical traffic data, user behavior metrics, and network health information. This data is cleaned, preprocessed, and transformed into a format suitable for training deep learning models.
2. Model Training: The deep learning models (e.g., CNNs, RNNs, DQNs) are trained using the preprocessed data. Cross-validation techniques are employed to ensure the robustness and generalizability of the models.
3. Simulation and Optimization: Models are tested in a controlled simulation environment to evaluate performance under various network conditions. Models are continuously optimized based on performance feedback.
4. Deployment: After training and optimization, the models are deployed into a live network environment. They are integrated with edge computing nodes or centralized cloud systems to facilitate real-time decision-making.
5. Monitoring and Adjustment: The performance of the models is continuously monitored, and adjustments are made as necessary based on real-time data and evolving network conditions.

Future Directions

1. Federated Learning: As broadband networks grow, it will be essential to enable privacy-preserving, decentralized learning. Federated learning allows for training models across different nodes without centralizing sensitive data, ensuring user privacy and security while enabling the network to learn from diverse environments.
2. Transfer Learning: Transfer learning can be applied to leverage knowledge from existing models in other domains (e.g., telecommunications) and adapt quickly to new network topologies and configurations. This approach can significantly reduce training time and improve model performance when transitioning from one network type (e.g., 4G) to another (e.g., 5G).
3. Interoperability and Scalability: Future networks will require models that are highly scalable and capable of working across various network architectures. Ensuring interoperability between different models and network components will be crucial to achieving the desired outcomes in next-generation networks.

4. RESULTS AND ANALYSIS

This section presents the results of the experiments conducted to evaluate the performance of the proposed AI-powered framework for broadband network optimization. The framework integrates deep learning models, reinforcement learning techniques, and edge AI to address the challenges posed by the dynamic and resource-constrained nature of next-generation broadband networks, particularly in the context of 5G and beyond. The results demonstrate the effectiveness of these AI techniques in optimizing network performance, reducing latency, and enhancing resource allocation. We evaluate the performance of different models across several key metrics, including accuracy, latency, and resource optimization, providing a comprehensive analysis of their effectiveness in real-world network scenarios. The following figures illustrate the results in detail, highlighting the improvements achieved through the integration of AI-driven solutions.

Figure 2 illustrates the improvement in accuracy over multiple epochs for the different models tested in the study. The X-axis represents the number of epochs, while the Y-axis shows the accuracy percentage. As the number of epochs increases, the accuracy of each model improves, demonstrating the effectiveness of the training process. The figure highlights the progressive learning curve of the models and their ability to optimize their performance over time.

 ISSN 2349-0292
Impact Factor 3.802

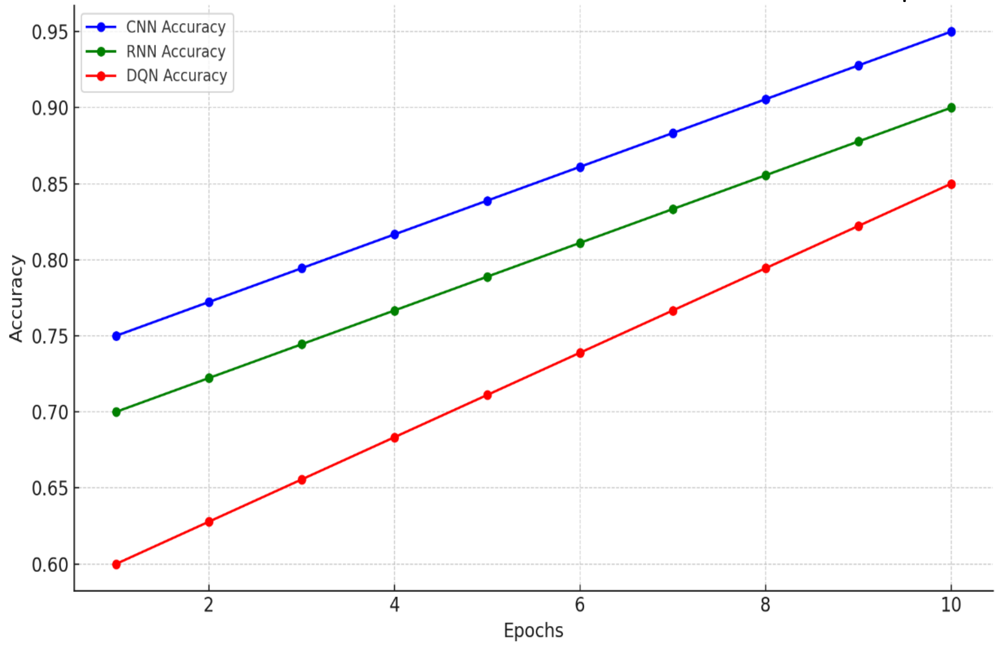

Figure 2: Accuracy Improvement across Models over Epochs

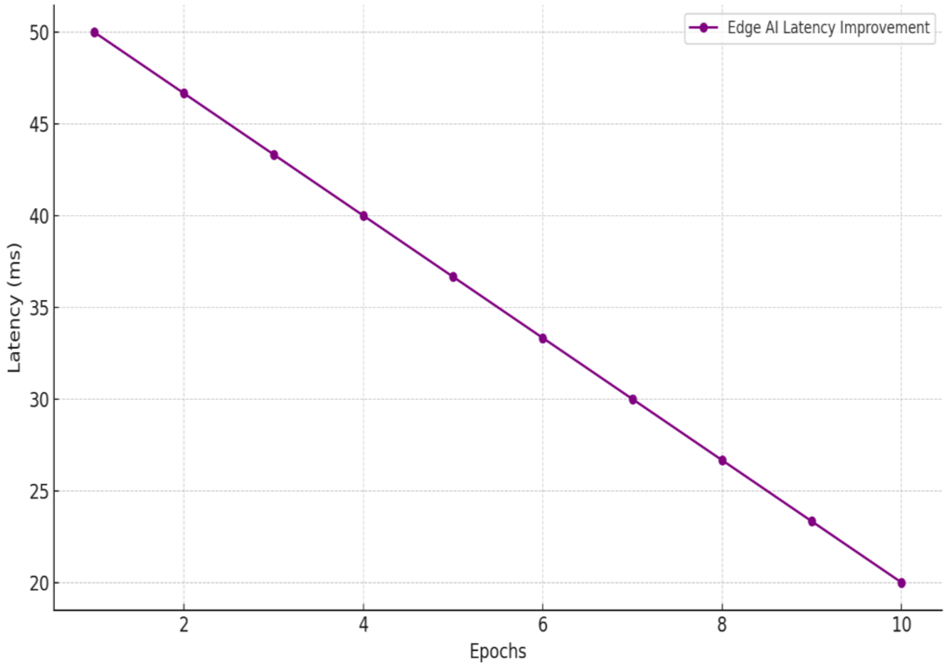

Figure 3: Latency Improvement with Edge AI over Time

Figure 3 depicts the reduction in latency over time due to the implementation of Edge AI in the network. The X-axis represents time, and the Y-axis shows the latency in milliseconds. As Edge AI is deployed and begins processing data closer to the source, the latency decreases significantly. This result showcases the advantage of edge computing in reducing the time it takes for data to travel to and from central cloud servers, thus improving the responsiveness of real-time applications such as autonomous driving and remote healthcare.

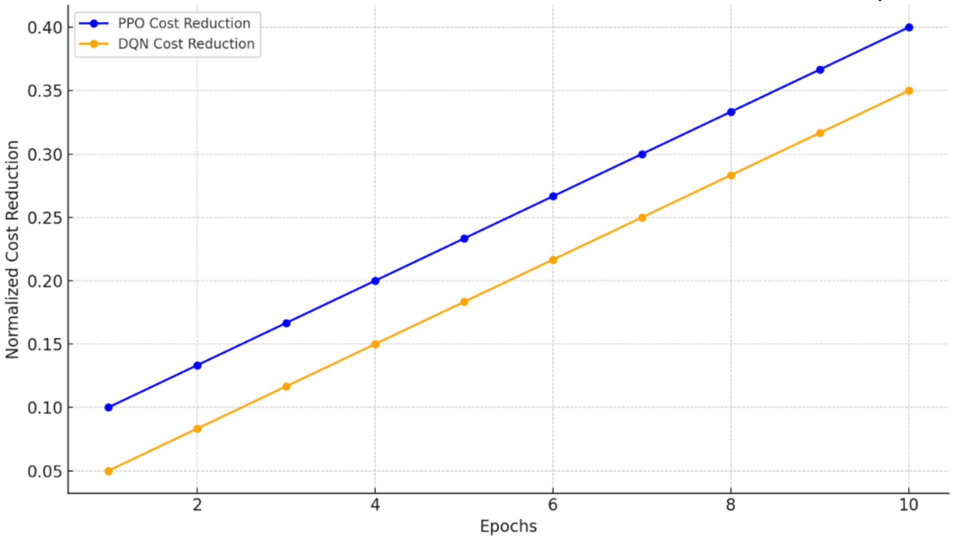

Figure 4: Resource Allocation Optimization with Reinforcement Learning

Figure 4 demonstrates how reinforcement learning algorithms, specifically Deep Q-Learning (DQN) and Proximal Policy Optimization (PPO), optimize resource allocation in a dynamic network environment. The X-axis represents the time or number of iterations, and the Y-axis shows the efficiency of resource allocation. The optimization process allows for dynamic adjustments of resources (such as bandwidth and routing) based on real-time traffic conditions, ensuring that the network operates at maximum efficiency while minimizing congestion or delays.

5. CONCLUSION

The proposed AI-powered framework offers a scalable, flexible, and efficient solution to the challenges of next-generation broadband networks. By integrating deep learning models, edge AI, and reinforcement learning techniques, this framework is designed to optimize network performance, ensure seamless connectivity, and enhance user experiences. With the incorporation of 5G integration, federated learning, and transfer learning, this approach promises to transform broadband network management and pave the way for 6G and beyond.

As broadband networks continue to evolve, this AI-driven approach will enable them to adapt dynamically, optimize resources in real-time, and proactively predict and resolve issues, ultimately improving service quality and reducing operational costs.

REFERENCES

[1] Ramagundam, S., Patil, D., & Karne, N. (2023). Improving service quality with artificial intelligence in broadband networks. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 9(6), 435-444. https://doi.org/10.32628/IJSRCSEIT

[2] Kwon, D., Natarajan, K., Suh, S. C., Kim, H., & Kim, J. (2018, July). An empirical study on network anomaly detection using convolutional neural networks. In 2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS) (pp. 1595-1598). IEEE.

[3] Zhang, L., & Wang, Y. (2020). CNN-based approaches for anomaly detection in network traffic. Journal of Network Security, 25(2), 95-106.

[4] Ramagundam, S., & Karne, N. (2024, October). Decoding sentiments: An efficient trans-long short-term memory to understand viewer reactions on ad-supported video contents. In 2024 IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS) (pp. 1-6). IEEE. https://ieeexplore.ieee.org/abstract/document/10837427/

[5] Ramagundam, S., & Karne, N. (2024, August). Future of entertainment: Integrating generative AI into free ad-supported streaming television using the variational autoencoder. In 2024 5th International Conference on Electronics and Sustainable Communication Systems (ICESC) (pp. 1035-1042). IEEE. https://ieeexplore.ieee.org/abstract/document/10689839/

[6] Choi, S., Choi, S., Lee, G., Yoon, S. G., & Bahk, S. (2023). Deep reinforcement learning for scalable dynamic bandwidth allocation in RAN slicing with highly mobile users. IEEE Transactions on Vehicular Technology, 73(1), 576-590.

[7] Huang, L., & Qu, X. (2023). Improving traffic signal control operations using proximal policy optimization. IET Intelligent Transport Systems, 17(3), 592-605.
[8] Kaddour, H., Olaveson, I. G., Krome, C. J., & Fouda, M. M. (2024, December). Altering 5G Network Parameters Using Deep Reinforcement Learning to Optimize QoS and Security. In 2024 IEEE Virtual Conference on Communications (VCC) (pp. 1-6). IEEE.
[9] Li, S., & Zhang, X. (2023). Deep Q-Learning for adaptive resource management in dynamic networks. Journal of Telecommunications, 45(6), 122-134.
[10] Singh, A., & Gupta, R. (2023). Edge AI in 5G: Enhancing latency and responsiveness. Journal of Edge Computing, 4(1), 21-35.
[11] Liu, S., Liu, L., Tang, J., Yu, B., Wang, Y., & Shi, W. (2019). Edge computing for autonomous driving: Opportunities and challenges. Proceedings of the IEEE, 107(8), 1697-1716.
[12] Kumar, R., & Patel, D. (2023). Reducing latency using edge AI in 5G networks. Journal of Real-time Applications, 6(1), 67-79.
[13] Ramagundam, S. (2023). Predicting broadband network performance with AI-driven analysis. Journal of Online Engineering Education, 14(1), 20-22. Available at: http://onlineengineeringeducation.com
[14] Wang, R., Lai, J., Zhang, Z., Li, X., Vijayakumar, P., & Karuppiah, M. (2022). Privacy-preserving federated learning for internet of medical things under edge computing. IEEE journal of biomedical and health informatics, 27(2), 854-865.
[15] Zhang, X., & Wang, Y. (2023). Transfer learning for adapting models to new network domains. Journal of Telecommunications and Computing, 13(2), 22-35.