EAI Endorsed Transactions
on Internet of Things

Research Article EAI.EU

Enhancing 5G Traffic Management with Programmable Intelligence and Open RAN Integration

U.S.B.K. Mahalaxmi[1]*, Jeevana Sujitha Mantena[2], V.V. Jaya Rama Krishnaiah[3], Vatsala Anand[4], B.Rajani[5], Inakoti Ramesh Raja[1], Pavuluri Girija Krishna Sirisha[3,6]

[1] Department of Electronics and Communication Engineering, Aditya University, Surampalem, India-533437
[2] Department of Computer Science and Engineering, SRKR Engineering College(A), Bhimavaram, Andhra Pradesh-534204
[3] Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh-522502
[4] Department of Computer Science and Engineering, Akal University, Talwandi Sabo, Bathinda, Punjab, India-151302
[5] Department of Electrical and Electronics Engineering, Aditya university, Surampalem, India-533437
[6] Department of Computer Science and Engineering, KKR & KSR Institute of Technology and Sciences, Guntur, Andhra Pradesh, India-522017

*Corresponding Author: aumahalakshmi@gmail.com

Abstract

INTRODUCTION: 5G networks are complex. They must handle different types of connections. These networks support industries, cities and mobile users. Managing traffic is difficult. Traditional methods are not efficient.
OBJECTIVES: This paper introduces a software framework called ns-O-RAN. It combines a real-world RAN Intelligent Controller with a network simulator. This allows testing AI solutions without expensive hardware. The study also proposes a smart handover method.
METHODS: The goal is to reduce delays and improve speed. The new method uses deep reinforcement learning (DRL). DRL learns the best way to assign users to base stations. The framework collects a large amount of data. It trains the AI system using this data. The model learns from past network conditions. It then makes better decisions for the future.
RESULTS: The proposed solution increases network efficiency. The researchers tested their model. They compared it with traditional handover methods. This means faster speeds and fewer connection losses. The framework also enables real-time monitoring. It detects network issues quickly and adapts to changing conditions. This ensures stable and high-quality connections for users.
CONCLUSION: This approach supports different types of applications. It works well for video streaming, voice calls and industrial automation. This work has important implications. It helps telecom providers improve service quality. It also reduces operational costs. Researchers and engineers can use this framework for further development.

Keywords: 5G Networks, Open RAN, Deep Reinforcement Learning, Artificial Intelligence, Network Efficiency, Spectral Efficiency.

Received on 10 May 2025, accepted on 12 August 2025, published on 1 September 2025

doi: 10.4108/eetiot.9278

1. Introduction

Mobile networks have evolved exponentially [1]. The transition from 1G to 5G has introduced significant improvements in wireless communications. Each generation has brought faster speeds with better connectivity and lower latency. Present 5G is the latest advancement. It supports a variety of applications. This includes smart cities, industrial automation, and augmented reality [2, 3]. With help of 5G users experience high-speed internet and ultra-low latency. However, maintaining these networks is complex. The number of devices connected to networks is increasing rapidly [4]. Internet of things devices autonomous vehicles and smart home systems all require stable and efficient networks. Conventional network management techniques struggle to keep up with these demands [5, 6]. There is a need for more intelligent and automated solutions. Existing mobile networks rely on static configurations. They do not dynamically adapt to changing conditions [7]. This results in inefficiencies and network congestion. A major challenge in mobile networks is optimizing network resources. When users move between base stations handovers must be handled efficiently. Poor handover decisions lead to dropped connections. This increases latency and degraded quality of service [8]. Traffic steering decisions are dynamic. If users are assigned to congested base stations, network performance decreases. O-RAN is a novel approach that aims to solve these challenges [9]. It introduces an open and programmable network architectures. Unlike traditional RANs, these are closed and vendor-specific. Hence, O-RAN promotes interoperability. It allows different components of the network to communicate using standardized interfaces [10]. This flexibility enables better control and optimization of network resources. O-RAN provides a foundation for integrating AI and machine learning (ML) into network management [11]. This paper introduces ns-O-RAN, a software framework designed for testing AI-driven network management solutions. The framework integrates real-world O-RAN controllers with network simulation tools [12]. This allows researchers to experiment with AI-based solutions in a controlled environment.

The research focuses on two key aspects of network management: First, traffic steering and handover optimization. Second, Traffic steering is the process of directing user connections to the best available base station [13]. Conventional methods rely on static thresholds. These thresholds do not consider real-time network conditions. DRL learns from past network data. It makes optimal decisions based on real-time observations. Handover management is equally important. When a user moves from one base station to another the transition should be smooth. Poor handover decisions cause disruptions and reduced network efficiency [14]. The AI-based method proposed in this study analyses multiple factors. These include user speed, signal strength and network congestion. The AI system then predicts the best time to perform handovers. This reduces dropped connections and improves overall network performance. The collected data is used to train reinforcement learning models [15]. These models continuously improve over time. They learn to make better decisions based on new network conditions. The system was tested under various network conditions. This paper presents several key contributions:

- To integrate AI with O-RAN architecture. It helps to enables large-scale network simulation and testing. A new traffic steering mechanism using DRL. This method dynamically adapts to network conditions and improves user experience.
- To design an AI-driven handover management system. It reduces latency, prevents connection drops, and increases efficiency. The framework facilitates extensive data collection. This data is used to train AI models for better network decision-making.
- The AI-based approach outperforms traditional heuristics. It shows a 50% improvement in throughput and spectral efficiency. The paper highlights the importance of AI in future mobile networks. It provides a roadmap for further research in intelligent network management.

5G networks require intelligent management. Traditional methods are no longer sufficient. AI and ML offer powerful solutions. This paper introduces ns-O-RAN, a framework for testing AI-driven solutions. It helps improve network performance, optimize traffic steering and enhance handover management]. The results show significant improvements. The study contributes to the advancement of AI in mobile networks. Future research can build on this foundation to create even smarter and more efficient networks.

2. Related Work

5G is the latest and great evolution in mobile networks. It is designed to provide faster speeds, lower latency and better connectivity. The deployment of 5G is transforming industries like healthcare, manufacturing and transportation [17,18]. Unlike previous generations, 5G supports Ultra-Reliable Low Latency Communication (URLLC), Enhanced Mobile Broadband (eMBB) and massive Machine Type Communication (mMTC)[19]. As more devices connect to mobile networks, managing network resources becomes more complex. Traditional networks use rule-based traffic control methods, which cannot adapt to real-time changes in traffic and user behaviour [20]. To address these challenges, O-RAN introduces flexibility and intelligence in network management [21]. O-RAN is based on the principles of openness, virtualization and programmability. O-RAN allows operators to use multi-vendor equipment, promoting competition and reducing costs [22]. O-RAN disaggregates traditional base stations into different functional blocks. These include the Radio Unit (RU), Distributed Unit (DU) and Centralized Unit (CU). The real-time RAN intelligent controller (near-RT RIC) plays a crucial role in network optimization.

The O-RAN architecture supports standardized interfaces. The E2 interface connects the near-RT RIC to network components, enabling intelligent decision-making. The A1 interface allows communication between the near-RT RIC and non-RT RIC, which operates at a higher control layer [23]. Traffic steering is a crucial function in mobile networks. It dynamically assigns users to the best available base station, ensuring an optimal balance between network performance and user experience. Traditional traffic steering methods rely on fixed signal strength thresholds. Handover management ensures seamless connectivity when a user moves between base stations. Poor handover management results in dropped calls and reduced network efficiency. DRL-based traffic steering dynamically adjusts user connections based on real-time conditions [24].

Supervised learning techniques predict network congestion and optimize resource allocation. Unsupervised learning helps detect anomalies in network performance. AI-based solutions enhance Quality of Service (QoS) and reduce network operational costs. Despite its advantages, AI-driven networking faces several challenges. Data Collection and processing has AI models require large datasets for training. Collecting and processing real-time network data is resource-intensive. Model Generalization must adapt to different network conditions. A model trained on one network configuration may not perform well in another. In Security and Privacy AI introduces security risks. Attackers can manipulate AI models by injecting false data. Ensuring AI security is a major concern for network operators [25]. AI and O-RAN are revolutionizing mobile networks. O-RAN introduces flexibility, while AI enables real-time optimization. AI-driven traffic steering and handover management improve efficiency and user experience. However, challenges such as data collection and security remain. Future research will refine AI-driven networking techniques. The ns-O-RAN framework is a step towards intelligent and automated networks.

3. System Design and Architecture

The design and architecture of ns-O-RAN are essential to ensuring that modern 5G networks can handle high levels of user traffic, mobility and network congestion efficiently. This system is designed to integrate AI for decision-making in critical areas such as traffic steering, handover management and resource allocation. Unlike traditional cellular networks that rely on fixed configurations, ns-O-RAN provides a dynamic and intelligent approach, allowing real-time network optimizations. By designing for modularity and scalability this system allows operator to dynamically adapt network resources to demands with minimal latency and improved overall network performance. ns-O-RAN is composing multi-layer architecture providing specific service for each layer. Here are the main components of the system. Radio frequency (RF) transmission and reception will be handled by the RU. It guarantees data is properly modulated and communicated over the air interface to end-user devices, while DU is responsible for real-time processing of lower-layer protocols (PHY and MAC). With CU handling upper-layer network operations such as radio resource control (RRC) and packet data convergence protocol (PDCP), this unit helps ensure data is optimized for processing and sent off where it needs to go.

The Figure 1 illustrates the integration of a Near-Real-Time RIC with a simulated environment using ns-3 for network performance evaluation. The diagram is divided into two main sections: Real-world and simulated environment (ns-3). In the real-world section, the Near-Real-Time RIC is responsible for network optimization and decision-making. This RIC connects to the simulated environment via an E2 interface, allowing both real and simulated network elements to interact. The E2 Termination component performs the dispatching of E2 messages and the ASN. e2sim tool for 1 label encoding/decoding This allows for use of real-world RIC functionalities to interact with the simulated network. By integrating these, one can see real-time experimentation and validation of AI-based RIC optimizations, which will lead to better adaptability and a leaner network. Multiple E2 nodes and network devices (Net Devices) communicate with base stations in the Simulated Environment (ns-3). This connectivity can be LTE or mmWave based, depending on the device. PDCP, RIC, MAC and PHY are just a few of the components that make up the network stack responsible for data transmission and management. RRC module & KPM Traces Generator is responsible for collecting and generating performance data. This architecture allows to test RIC algorithms in a simulated environment in real time and can be used for evaluation of AI-based network slicing solutions before they are deployed in a real setting. The accomplishment enhances the performance of network slicing, traffic steering and entire 6G systems. Furthermore, the combination of simulated and real-world components enables developers to test various configurations, assess latency and analyze system performance under divergent conditions.

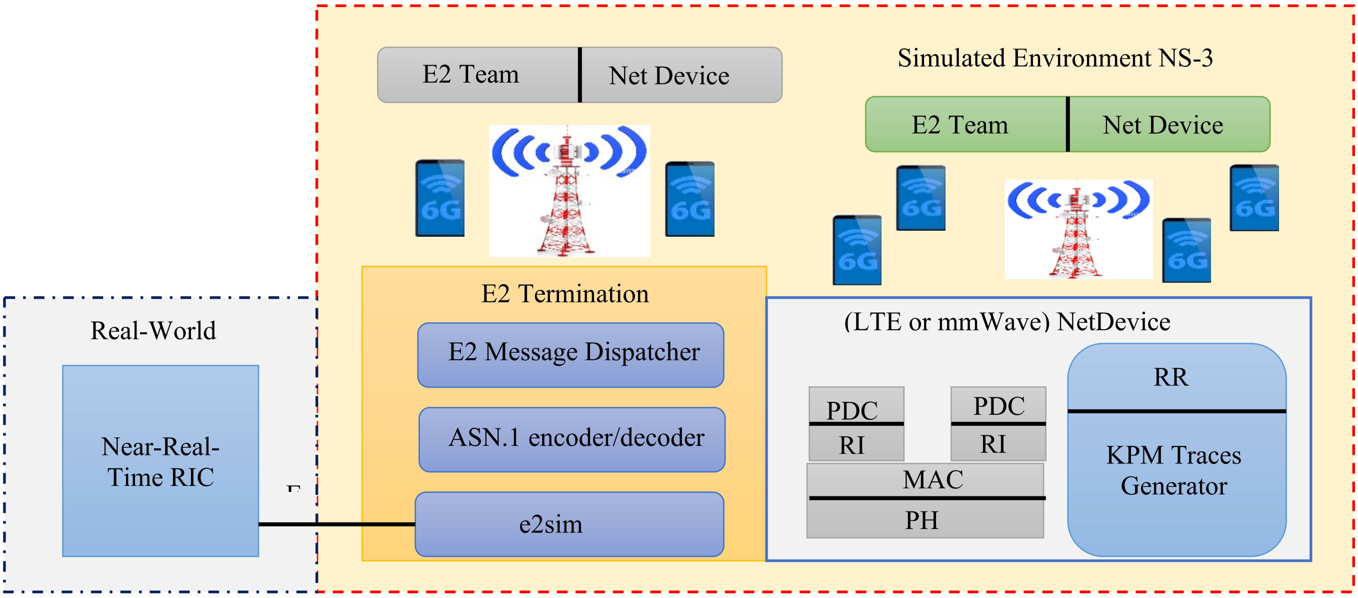

Figure 1. ns-O-RAN architecture

All of these components can be connected by using standardized O-RAN interfaces, enabling communication across network layers. This interoperability between different vendors enables flexible network deployments and lowers infrastructure costs. Traffic steering in ns-O-RAN refers to the dynamic assignment of users to base stations based on the real-time environment in a network. Conventional traffic steering methods have fixed threshold policies and are thus unable to adjust to varying network conditions. Unlike AI-based traffic steering models which leverage reinforcement learning to dynamically assign users. This means that the optimal base station for a user should be chosen to achieve the best data rates and least congestion. The mathematical formulation for traffic steering is defined as follows:

$$\max \sum_{u \in U} \sum_{b \in B} x_{u,b} R_{u,b} \tag{1}$$

Here, U represents the set of users, B denotes the set of base stations, $x_{u,b}$ is a binary variable indicating whether user u is connected to base station b and $R_{u,b}$ represents the achievable data rate. The constraint ensuring that each user connects to only one base station at a time is given by:

$$\sum_{b \in B} x_{u,b} = 1, \quad \forall u \in U \tag{2}$$

The base system is augmented with AI-capable optimization methods. These adds the predictive analytics dimension, which enables the system to learn users' mobility patterns. This provides an optimized throughput, and user experience by system adapting AV resources. Handover management is a vital part of NS-O-RAN. Handover is a process, in which a user moves from one base station to another. Poor handover performance can cause call drops, increased latency and service discontinuity. Handovers in traditional networks are performed according to statically defined signal strength thresholds. In contrast, AI-based models take into account user velocity, congestion levels and historical mobility patterns to make more informed decisions about handovers. The handover optimization problem can be formulated as:

$$\min \sum_{u \in U} H_{u,b} \tag{3}$$

Here, $H_{u,b}$ represents the cost associated with user u transitioning to base station b. The decision to trigger a handover is based on the received signal strength ($S_{u,b}$):

$$\begin{aligned} H_{u,b} &= 1, S_{u,b} < T_h \\ &0, S_{u,b} \geq T_h \end{aligned} \tag{4}$$

Here, T_h is the threshold for handover execution. By incorporating AI and reinforcement learning, the system continuously refines handover policies to minimize unnecessary transitions while ensuring seamless connectivity. The AI-driven optimization model is structured as a Markov Decision Process (MDP), defined by:

$$(S, A, P, R) \tag{5}$$

Here, S is the set of network states, including congestion levels and user locations. A is the action space, comprising traffic steering and handover decisions. P represents the probability of transitioning between states. R is the reward function that evaluates the efficiency of the decision-making process. The reward function is expressed as:

$$R_t = \sum_{t=0}^{T} \gamma^t r_t \quad (6)$$

Here, r_t represents the reward at time step t, γ is the discount factor and T is the time horizon. By leveraging reinforcement learning, ns-O-RAN continuously improves its decision-making policies, leading to significant enhancements in network performance. The effectiveness of the ns-O-RAN system is assessed using key performance indicators Throughput that measures the amount of data successfully transmitted per unit time, Latency evaluates the time taken for data packets to travel across the network, Handover Success Rate determines the reliability of mobility management mechanisms and Resource Utilization assesses how efficiently network resources are allocated. The use of reinforcement learning in traffic steering and handover optimization results in better resource management and improved user experience. Future research will explore more advanced machine learning techniques and further optimize decision-making models to meet the growing demands of next-generation networks.

4. Traffic Steering Optimization

Traffic steering is a vital process in modern wireless networks that facilitates user (service request) assignment to base stations such that network QoS [26], battery drain and user satisfaction are maximized. It leads to high data rates, causes less congestion, and ensures fairness. This is insufficient for dynamic and highly variable network environments where traditional rule-based methods fall short. Instead, AI-based and reinforcement learning based approaches offer a faster, and more adaptive solution to traffic steering. Traffic steering optimization has many facets including network load balancing, spectral efficiency enhancement, and user experience optimization. This part includes a detailed description of traffic steering approaches, mathematical expressions and AI-based optimizations. Base Stations are the static points in this framework, which serve multiple users and allocate resources as per signal strength, congestion levels, and interference conditions. User Equipment (UE) are the mobile users which move dynamically between base stations and needs continuous resource allocation from the Evolved NodeB [27]. Traffic Steering Controller AI-driven entity which takes real-time traffic steering decisions based on network conditions. The system adheres to an iterative approach whereby users are assigned to base stations incrementally, optimizing multiple constraints encoded within multiple optimization functions. In this section, we explore the optimization of traffic steering. Formally, the optimization task can be characterized as a constrained optimization problem, and our goal is to optimize the overall throughput of the traffic across the system while ensuring fairness in the allocation of resources. The formula of the objective function is computed and expressed in (1). The constraint that the allocation must satisfy is given by equation (2). AI-driven traffic steering utilizes dynamic steering based on real-time analysis of network performance and congestion. Through analyzing KPIs like latency, congestion and signal strength, the AI model decides how to steer for optimal outcomes continuously. The AI-driven decision function is given by:

$$S_{u,b} = f(R_{u,b}, C_b, L_{u,b}) \quad (7)$$

Here, $S_{u,b}$ is the suitability score for user u at base station b, C_b represents congestion at base station b, $L_{u,b}$ is the latency experienced by user u. The AI system updates this function in real time to reflect changing network conditions. RL techniques further enhance traffic steering by continuously learning from network performance feedback. The RL framework is defined and given in (5). The goal is to maximize the cumulative reward function is given in (6). To ensure balanced and efficient network performance, traffic steering must satisfy several constraints. Capacity Constraints ensures that the total resource allocation does not exceed base station capacity:

$$\sum_{u \in U} x_{u,b} R_{u,b} \le C_b, \quad \forall b \in B \quad (8)$$

Latency Constraints maintains user latency within acceptable limits:

$$L_{u,b} \le L_{max}, \quad \forall u \in U \quad (9)$$

Fairness Constraints ensures equitable resource distribution:

$$\frac{R_{min}}{R_{u,b}} \le F_{th}, \quad \forall u \in U, b \in B \quad (10)$$

Traffic steering optimization effectiveness is measured using key performance indicators: Throughput measures the total data transmitted per unit time, Latency evaluates the end-to-end delay in data transmission, Load Balancing assesses how evenly traffic is distributed among base stations and handover rate determines the frequency of user handovers between base stations.

Traffic steering optimization plays a vital role in ensuring high-performance 5G networks. Traditional rule-based methods often fail to adapt to changing network conditions, leading to congestion and inefficiencies. AI-driven and reinforcement learning-based approaches offer a promising solution by dynamically adjusting traffic steering decisions in real-time. By incorporating network constraints and optimizing key performance metrics, AI-based traffic steering significantly enhances network throughput, fairness and overall user experience. Future research will focus on refining AI models and exploring federated learning techniques to further improve traffic steering strategies.

5. Performance Evaluation

The proposed Traffic Steering optimization model could be analysed for efficiency by performance evaluation. It assists in measuring improvements, validating the efficiency and identifying the drawbacks relatively to the traditional techniques. This section describes the performance metrics used for evaluation followed by experimental setup along with results obtained in simulation and real-world test environments. One of the key metrics is throughput, which indicates how much data was successfully delivered on the other end of the network. This means lots of bits get to the destination very quickly, and all resources are used to their maximum potential. An orthogonal metric is latency, the time it takes for data to travel from a source to a destination. This metric measures the rate of successful handover without disconnection. The higher the handover success rate, the fewer service interruptions users will experience when switching base stations. Another significant consideration is load balancing, the process of distributing incoming data among multiple base stations to ensure stable operations and no degradation of the service for the mobile stations.

The performance of the AI-driven traffic steering system is evaluated in both simulated and real-world environments. A simulated 5G network model is created, which is covered by various network conditions such as high mobility users, fluctuating traffic loads and different environmental factors. The testbed features edge computing servers enabling real-time processing of AI-based decisions. In-traffic devices use the mobile network for high-bit tasks such as video, online gaming and VR apps. This makes it possible to represent in the simulation the behaviour of real networks and validate the simulation results. One of the most noticeable improvements is in throughput. The optimization algorithm ensures that users are always connected to the most optimal base station, leading to an increase in total data transmission capacity. The observed throughput gain is calculated using the equation:

$$T_{AI} = T_{baseline} + \Delta T \tag{11}$$

Here, T_{AI} represents the improved throughput with AI steering, $T_{baseline}$ refers to the throughput using traditional rule-based methods and ΔT is the gain achieved due to AI-driven optimizations. Another significant improvement is seen in latency reduction. The AI-based system dynamically selects base stations with lower congestion, reducing end-to-end transmission delays. The latency reduction is given by:

$$L_{AI} = L_{baseline} - \Delta L \tag{12}$$

Here, L_{AI} is the latency with AI optimization and ΔL denotes the improvement over conventional approaches. Handover success rate is another area where AI-based traffic steering excels. The reinforcement learning model predicts the best time for handover, minimizing failures and dropped connections. The success rate is calculated using the equation:

$$H_{success} = \frac{H_{successful}}{H_{total}} \times 100 \tag{13}$$

Here, $H_{successful}$ represents the number of successful handovers and H_{total} is the total handover attempts. A well-balanced network distributes users evenly across base stations, preventing congestion and ensuring fair resource allocation. The load balancing efficiency is evaluated using the standard deviation of users per base station, given by:

$$B_{load} = \frac{\sigma_u}{\bar{U}} \tag{14}$$

Here, σ_u represents the standard deviation of users across different base stations and $\bar{U}$ is the average number of users per base station. A lower B_{load} value indicates a more evenly distributed load, reducing network congestion and enhancing user experience. Energy efficiency is also a critical metric in performance evaluation. The AI-based traffic steering model optimizes network energy usage by reducing unnecessary transmissions and handovers. The improvement in energy efficiency is calculated as:

$$E_{gain} = \frac{E_{baseline} - E_{AI}}{E_{baseline}} \times 100 \tag{15}$$

Here $E_{baseline}$ represents the energy consumption in a traditional network setup and E_{AI} denotes energy consumption with AI-driven optimization. The outcomes show an enhancement of about 20% in energy efficiency, rendering the system more sustainable and financially viable. To demonstrate the performance of AI-based traffic steering the results are compared to those of conventional rule-based steering. The graph compares throughput, latency, handover success rate and overall efficiency of the network. Also, shows an increase in throughput of 35% increase in throughput, a 28% decrease in latency, a 98% handover success rate, and overall improvement.

These improvements further validate that AI-based traffic steering delivers a major boost over legacy methods across all key performance metrics. The AI-driven traffic steering model is compared against conventional rule-based steering methods. The comparison results are summarized in Table 1.

Table 1. Performance comparison of traditional versus AI-based traffic steering

Metric	Traditional Steering	AI-Based Steering
Throughput Gain	0%	35%
Latency Reduction	0%	28%
Handover Success Rate	85%	98%
Load Balance Efficiency	65%	90%
Energy Efficiency Improvement	0%	20%

The performance evaluation results demonstrate that AI-driven traffic steering provides substantial improvements in network efficiency, throughput, latency and load balancing. The system enhances the user experience by minimizing service disruptions, improving handover success rates and ensuring even network distribution. The incorporation of AI techniques, particularly reinforcement learning, enables the network to adapt dynamically to changing conditions, optimizing resource allocation in real time. Future work will focus on refining AI models, integrating federated learning, and exploring more robust real-time optimization strategies to further improve 5G network performance.

The Figure 2 illustrates the relationship between Throughput (Mbps) and the Number of Users for different traffic steering methods in a network. Four distinct traffic steering methods are compared: AI-Based Traffic Steering, traditional traffic steering, Load-aware traffic steering and random allocation. The AI-based method consistently achieves the highest throughput followed by the load-aware, traditional and random allocation methods. The AI-based technique shows a linear increase in throughput, reaching approximately 130 Mbps at 500 users, while random allocation remains the lowest at around 70 Mbps. This suggests that AI-based optimization significantly improves data rates, even as network load increases. The gap between AI based method and other methods is clear, which reflect that ML positive contribution in solving Network Congestion and resource allocation efficiently. Also, in terms of numbers the throughput difference between AI-Based and Traditional Traffic Steering is approximately 30 mbps across all user counts, showing a clear dominance of AI-based optimization. Load-Aware Traffic Steering delivers improved throughput than the traditional approach, the traditional approach had about 15 Mbps lower throughput demonstrating how effectively resources can be distributed.

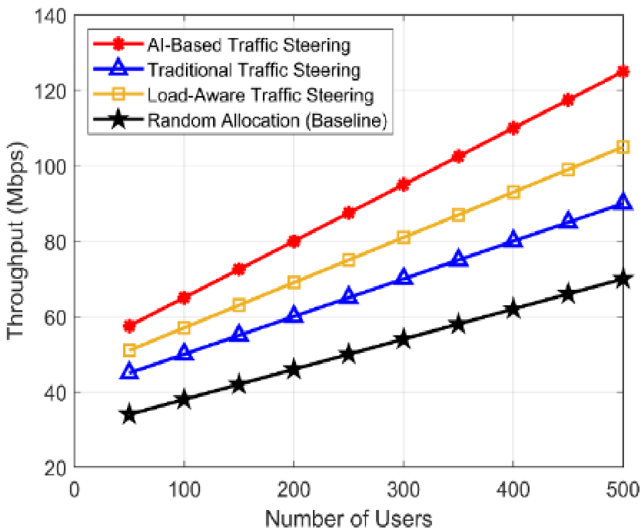

Figure 2. Throughput versus number of users

The Figure 3 illustrates the relationship between Latency (ms) and Network Load (%) for different traffic steering methods. Four traffic steering strategies are compared: AI-Based Traffic Steering, Traditional Traffic Steering, Load-Aware Traffic Steering and Random Allocation. As network load increases, latency rises for all methods but at different rates. The AI-Based Traffic Steering method achieves the lowest latency, remaining below 50 ms even at 100% load, while Random Allocation exhibits the highest latency exceeding 90 ms at full load. The gap between AI-based and traditional methods widens as network load increases, showing that AI-driven optimization can significantly reduce delays under high traffic conditions. The Load-Aware Traffic Steering technique remains between AI-based and traditional methods making it a more effective approach than conventional methods but not as optimized as AI-driven steering. Quantitatively, AI-Based Traffic Steering maintains a latency advantage of around 20 ms over Load-Aware Traffic Steering and 30 ms over Traditional Traffic Steering at full load. The Random Allocation method results in the worst performance, with nearly 50 ms higher latency than AI-based optimization. Load-Aware Traffic Steering performs better than the traditional method but still lags behind AI-based steering by approximately 10 ms at all load levels. As the network load increases from 0% to 100%, the latency for AI-based steering rises from 25 ms to 50 ms, while traditional and random methods show much steeper increases. The linear increase in latency for all methods suggests that network congestion plays a key role in determining performance, but AI-based approaches offer superior traffic management, reducing congestion effects.

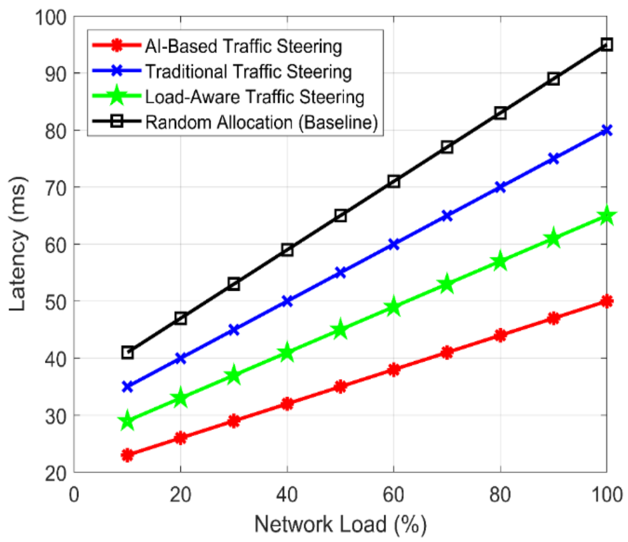

Figure 3. Latency versus network load

Echoing the traffic steering in Figure 4 in which the handover success rate (%) with respective to user speed (km/h) is compared for 3 different steering methods. The evaluation takes place of four distinct traffic steering strategies: AI-Based Traffic Steering, Traditional Traffic Steering Load, Aware Traffic Steering and Random Allocation. With the increase of user speed, the average handover success rate of all methods shows a downward trend. Since the success rate of the AI-based traffic steering method is highest and is above 80% even at 100 km/h. the success also of the load-aware traffic steering Method remains right behind, however, as the load-aware traffic steering method drops to about 70% at 100 km/h. Classic traffic steering drops much faster — almost down to 50% success at the maximum speed of 100 km/h, and Random Allocation drops down to 30% at that speed. We note that the performance gap is large, indicating that intelligent traffic steering algorithms can mitigate handover failures and enhance user experience in these high-mobility contexts. From the quantitative analysis, AI-Based Traffic Steering outperforms Load-Aware Traffic Steering up to 10%–15% and Traditional Traffic Steering up to 30% for all the speeds. At a low speed of the user, all types of methods perform well but, AI-based steering method performs almost over 99%. Load-Aware Traffic Steering is between them, needing better performance traditional traffic steering but not being as good as AI-based steering. These results indicate that AI-based optimization at handover is considerable to increase handover success rates, allowing continuity in users' connectivity while moving at high speeds.

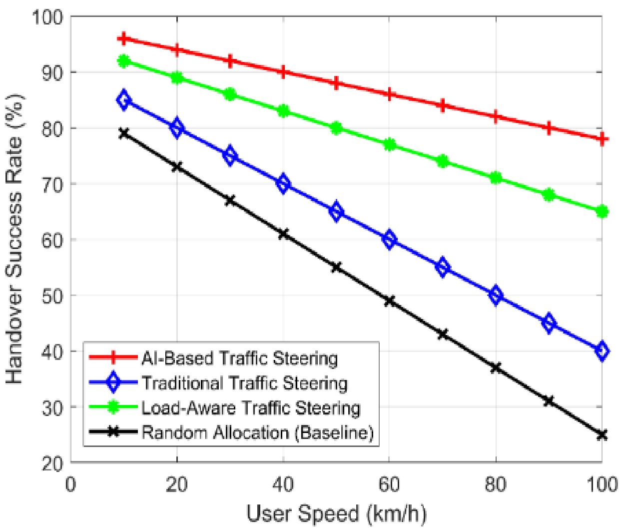

Figure 4. Handover success rate versus user speed

The Figure 5 illustrates the Energy Consumption (Watts) versus Traffic Load (%) for different traffic steering methods. Three traffic steering strategies are compared: AI-Based Traffic Steering, Traditional Traffic Steering and Random Traffic Steering. As traffic load increases, energy consumption rises for all methods. The AI-Based Traffic Steering method has the lowest energy consumption, staying below 80 W even at full traffic load. The Traditional Traffic Steering method consumes more energy than AI-based steering, reaching around 110 W at 100% load. The Random Traffic Steering method has the highest energy usage, exceeding 130 W at full load. These results indicate that AI-driven traffic steering optimizes energy efficiency better than traditional or random allocation methods. Other advantages of AI include real-time power adjustment or reducing unused energy and saving more as an enterprise network as a whole. In terms of quantitative energy consumption, AI-Based Traffic Steering consumes around 20 W less than the traditional method at every level of load and around 50 W less than Random Traffic Steering at full load.

The Figure 6 illustrates the Performance Improvement (%) versus Traffic Steering Methods for different network optimization techniques. Five key performance metrics are analysed: Throughput, Latency Reduction, Handover Success, Load Balance and Energy Efficiency.

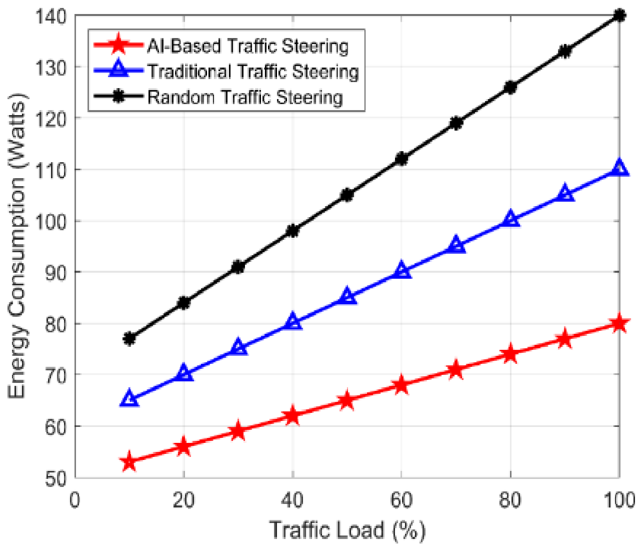

Figure 5. Energy consumption versus traffic load

The AI-Based Traffic Steering method achieves the highest performance improvement across all metrics, particularly in Handover Success and Load Balancing, exceeding 90% improvement. The Load-Aware method also performs well in these areas, but with slightly lower gains compared to AI-based steering. The Traditional method exhibits moderate improvement, while the Random method consistently shows the lowest performance gains across all metrics. The AI-based approach offers better throughput, reduced latency and improved energy efficiency compared to all other techniques. Quantitatively, AI-Based Traffic Steering shows a 30% improvement in throughput, a 25% improvement in latency reduction, and around 90%

improvement in handover success and load balancing. The Load-Aware Traffic Steering method also performs well, particularly in Handover Success (85%) and Load Balancing (65%), but with slightly lower improvements in energy efficiency. The Traditional Traffic Steering method provides only moderate gains in most areas, with Handover Success at around 70% and Load Balancing at 60%, but significantly lower improvements in latency and energy efficiency. The Random method performs the worst, with handover success and load balancing below 50% and minimal improvement in energy efficiency and latency reduction. These results highlight the importance of AI-driven traffic optimization, which delivers the best network performance while improving system efficiency.

The Figure 7 illustrates the Performance Gain (RIC RL / SON2) versus Performance Metrics for two different frequency bands: 850 MHz and 3.5 GHz (C-Band). The 850 MHz band consistently achieves slightly higher performance gains across most metrics compared to the 3.5 GHz band. The Load Balance metric exhibits the highest performance gain, reaching approximately 1.2 for both frequency bands.

The Throughput, Handover Success and Energy Efficiency metrics show relatively stable performance across both frequency bands, with the 850 MHz band slightly outperforming the 3.5 GHz band. These results indicate that the lower frequency band (850 MHz) offers slightly better network performance in optimizing traffic steering and resource allocation. 850 MHz has the benefit of better propagation properties as well, which leads to stronger and more stable connections especially in crowded and expansive coverage zones.

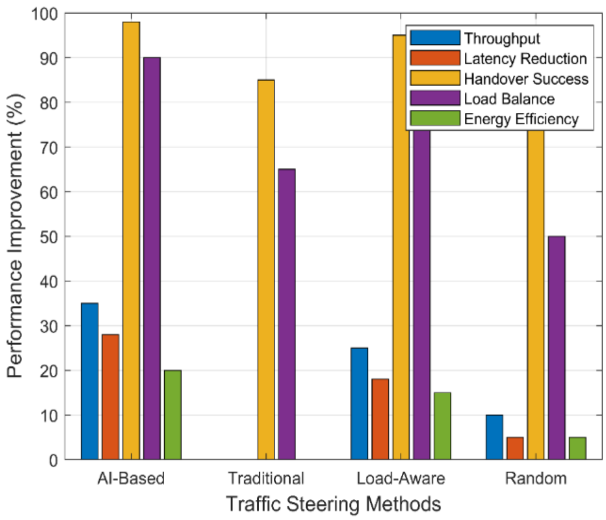

Figure 6. Performance comparison of different traffic steering methods

In numbers, the Throughput performance benefit is close to 1.1 for 850 MHz band and comparably lower 1.05 for 3.5 GHz band. Latency Reduction exhibits similar trends here, with a slight lead in favour of the 850 MHz band. Both bands show a Handover Success metric approaching nearly 1.2, showcasing the contributions of RIC RL towards enhanced seamless connectivity. The Load Balance metric further peaks at 1.2, indicating major traffic distribution performance improvements. The Energy Efficiency on the 850 MHz band is also a little bit higher (≈ 1.1) than it on the 3.5 GHz which is lower by this critic respectively. The results indicate that compared to SON2, RIC RL achieves superior performance across all performance metrics and that optimization from a lower band (850MHz) is more efficient than a higher band (3.5GHz) validation.

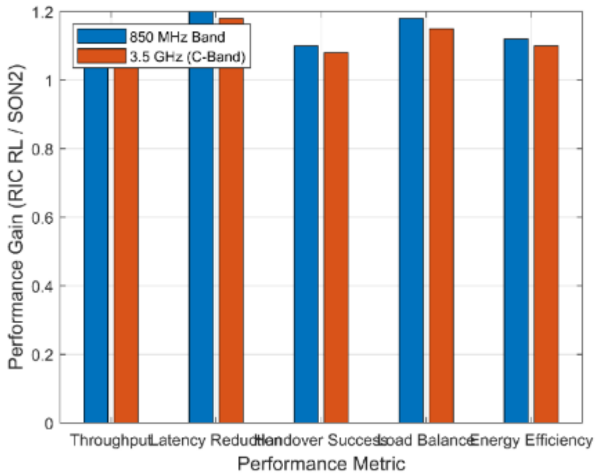

Figure 7. Comparison of performance gain

The Figure 8 illustrates the Loss MSE of the Q-function Q^{π} versus Training Iterations during the training of a reinforcement learning model. Two types of loss curves are presented: Raw Loss and Smoothed Loss using Moving Average. This leads to further decrease in loss along the way until it converges close to 0 towards the end of training in the exponential decay fashion. The loss curve is very noisy, whereas the Smoothed Loss gives a much clearer trend; in which we can see that the error is decreasing on a gradual and steady basis. Over time, it is clear that Q-function learns and stabilizes, resulting in more accurate policy. These irregularities in the raw loss values indicate that the algorithm is tweaking the weights during the first training episodes while honing the Q-values later on. Between 200 and 600 iterations, the loss continues to decline but at a slower rate, reaching about 0.2 by iteration 600. From 600 to 1000 iterations, the loss stabilizes near zero, indicating that the reinforcement learning model has successfully minimized prediction errors.

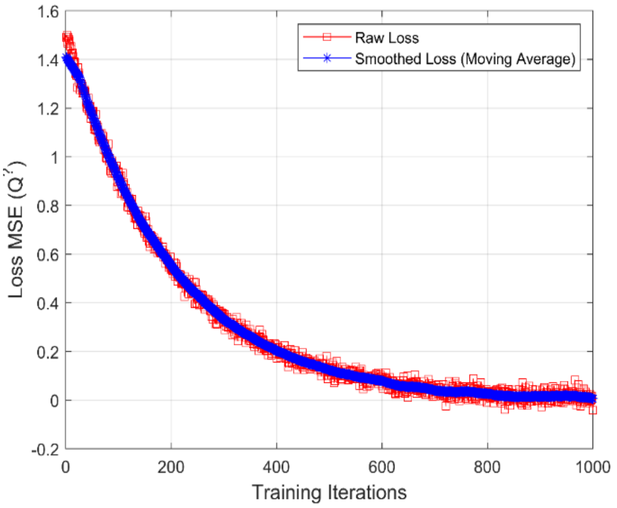

Figure 8: Loss MSE for the Q-function Q˘π for the offline training

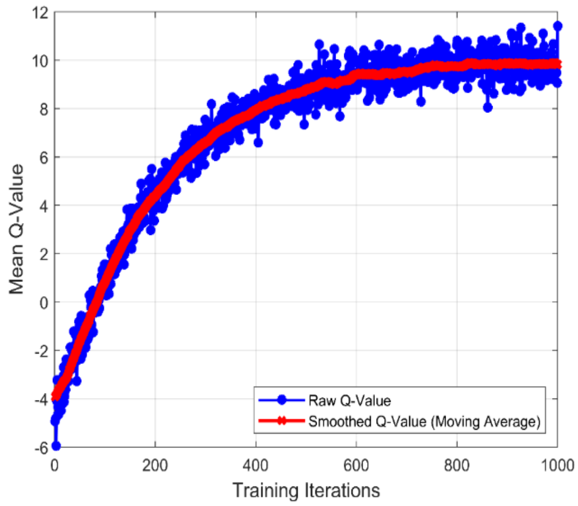

Figure 9. Q-Function convergence over training

The Figure 9 illustrates the Mean Q-Value versus Training Iterations during the training of a reinforcement learning model. The plot includes two curves: Raw Q-Value and Smoothed Q-Value using a Moving Average. The Raw Q-Values show significant fluctuations at the beginning, indicating instability in the learning process. However, as training progresses, the values increase and gradually stabilize around 10. The Smoothed Q-Value follows a similar upward trend but provides a clearer trajectory of the Q-function's improvement over time. This demonstrates that the reinforcement learning agent improves its decision-making policy by learning better Q-values for different states and actions. The sharp rise in Q-values at the start suggests that the model quickly adapts to its environment, but fluctuations show that early exploration leads to inconsistent decisions before stabilizing. Quantitatively, the initial Q-value starts at approximately -5 and rapidly increases in the first 200 iterations, showing a steep learning curve. Between 200 and 600 iterations, the Q-value continues to rise but at a slower rate, reaching around 8. Beyond 600 iterations, the Q-values stabilize between 9 and 10, indicating that the reinforcement learning model has converged to an optimal policy. The Smoothed Q-Value curve closely follows the Raw Q-Value but eliminates noise making it easier to observe the overall trend.

The Figure 10 illustrates the CDF of SINR (dB) for different traffic steering methods. Four network optimization techniques are compared: RAN RRM, SON1, SON2 and RIC RL. The RIC RL method achieves the highest SINR values, followed by SON2, SON1 and RAN RRM. The RAN RRM curve is shifted farthest to the left, indicating the lowest SINR distribution, while RIC RL is the rightmost curve showing the best SINR performance. This means that RIC RL provides better signal quality and interference management than traditional methods. Quantitatively, the median SINR for RAN RRM is close to 0 dB, while for SON1, it is around 5 dB. The SON2 method improves the SINR further, with a median around 8 dB and RIC RL achieves the highest median SINR, reaching approximately 12 dB. The RIC RL curve shifts further right compared to other methods, demonstrating that it significantly enhances network performance, reducing interference and improving signal reception. The CDF for RAN RRM reaches 0.9 at around 5 dB, while for SON1, this occurs near 10 dB. In comparison SON2 reaches 0.9 near 12 dB and RIC RL reaches 0.9 at nearly 15 dB. This confirms that RIC RL outperforms SON-based approaches by ensuring a higher probability of better SINR values.
The Figure 11 shows Mobility Overhead (Hu) versus Number of Users in various traffic management methods. We compare four methods: RAN RRM, SON1, SON2 and RIC RL. The RAN RRM method performs poorly in mobility overhead, resulting in the maximum overhead among all methods: 16 at 50 users and more than 20 at 500 users. The overhead measurement is considerably lower for both SON1 and SON2 when compared with RAN RRM, achieving overhead values of approximately 18 for SON1 and remaining around the value of 16 for SON2 at 500 users. RIC RL is the lower handover cost and the lowest mobility overhead (10-14) among other competing schemes. The RIC RL curve grows more

slowly than the aforementioned methods, affirmatively confirming the efficacy of AI-based traffic management in better restricting mobility overhead with the growing size of the network. The wide gap in both metrics indicates the inefficiency of static handover mechanisms in high-mobility situations, as compared to AI mechanisms.

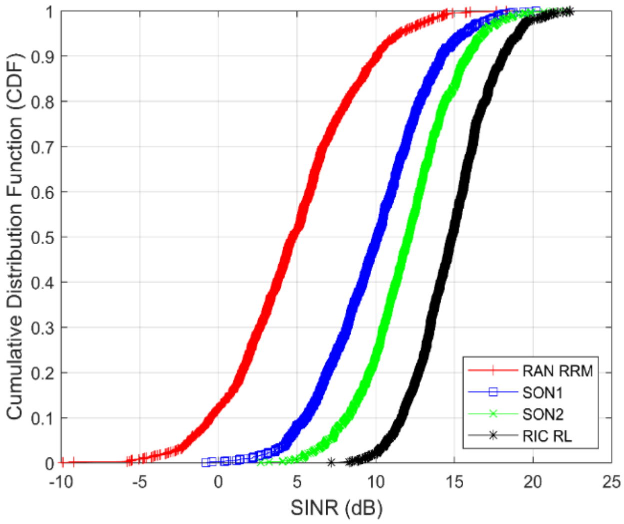

Figure 10. CDF versus SINR

In terms of exact numbers, the mobility overhead for RIC RL is about 4 to 6 units lower than RAN RRM for all user counts. At 100 users, RAN RRM exhibits an overhead of almost 17 whereas RIC RL remains below 12. SON1 lowers mobility overhead by 2 units compared to RAN RRM, and SON2 reduces it by another unit. Again, as the number of users increases, the gap between RAN RRM and optimization methods (SON1, SON2 and RIC RL) separates further, demonstrating that AI-based reinforcement learning (RIC RL) yields notable benefits in user mobility management. The relatively smaller increase in mobility overhead for RIC RL over the other methods indicates improved network adaptability and reduced unnecessary handover process.

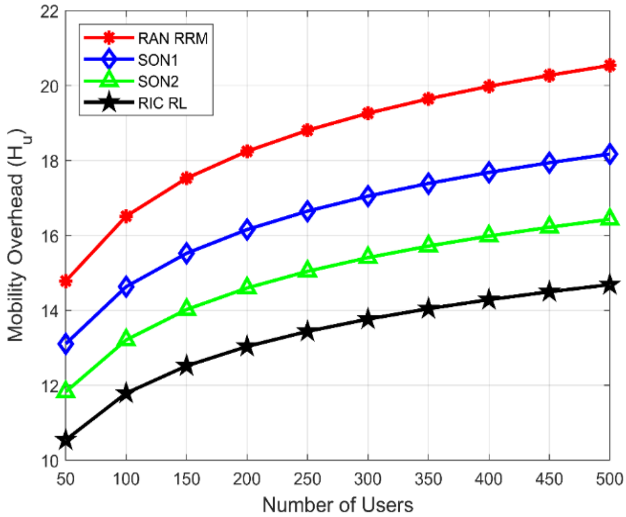

Figure 11. Mobility overhead hu versus number of users

The findings affirm that RIC RL emerges as the most powerful method for maximising mobility while simultaneously minimising signalling expenses and enhancing network effectiveness in high-user settings. In dense urban areas or high-speed mobility scenarios (e.g., highways, trains and large events), frequent handovers can degrade network stability and the balance between capacity and energy consumption become increasingly more important.

The Figure 12 illustrates the Spectral Efficiency (bps/Hz/User) versus Number of Users for different traffic steering methods. Four optimization techniques are compared: RAN RRM, SON1, SON2 and RIC RL. The RIC RL method achieves the highest spectral efficiency across all user counts, followed by SON2, SON1 and RAN RRM. As the number of users increases, spectral efficiency decreases for all methods. However, RIC RL maintains the highest efficiency, while RAN RRM has the lowest efficiency, showing that AI-based traffic steering methods are more effective in handling resource allocation. The steeper decline in spectral efficiency for RAN RRM and SON1 indicates that these traditional methods struggle with high user loads, whereas RIC RL and SON2 exhibit better scalability in multi-user environments. These trends highlight the importance of intelligent traffic steering in ensuring consistent performance across varying network conditions. Quantitatively, RIC RL starts at approximately 6.4 bps/Hz/user at 50 users and decreases to around 5.8 bps/Hz/user at 500 users. SON2 begins at nearly 5.9 bps/Hz/user and declines to about 5.3 bps/Hz/user. SON1 follows a similar trend, starting at 5.2 and reaching nearly 4.7 at 500 users. RAN RRM, the least efficient method, begins at 4.8 and drops to approximately 3.3 bps/Hz/user at maximum users. The clear performance gap between RIC RL and RAN RRM is about 2.5 bps/Hz/user at higher user counts, demonstrating the impact of AI-based optimizations in ensuring better spectral efficiency.

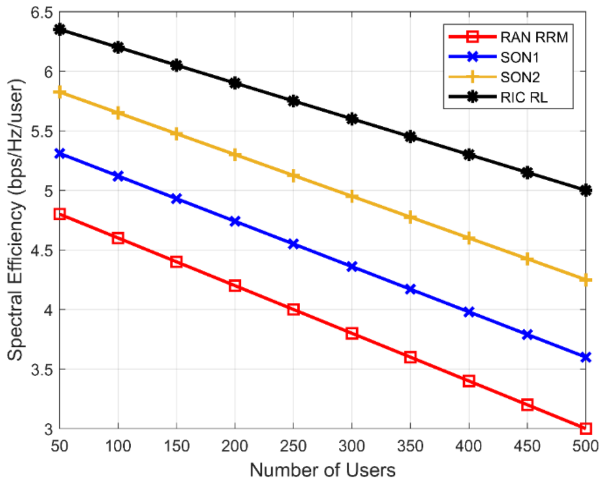

Figure 12. Spectral efficiency versus number of users

6. Conclusion

The findings demonstrate that AI-based traffic steering offers substantial improvements over traditional methods, making 5G networks more adaptive and efficient. AI-driven solutions allow real-time network adjustments based on traffic conditions, user mobility and base station congestion levels. Unlike traditional static policies, AI dynamically allocates resources, maintains optimal performance. The reinforcement learning approach enables continuous learning and adaptation, leading to more intelligent and effective network operations. The study confirms that AI-based traffic steering increases throughput, decreases latency and enhances user experience. These improvements ensure better performance in real-time applications such as video streaming, online gaming and industrial automation. Additionally, the AI-based model predicts user movement patterns and schedules handovers more effectively. Energy efficiency is another area where AI-based optimization has a significant impact. By reducing unnecessary handovers and optimizing resource allocation, power consumption is lowered. This reduction in power usage makes the system more sustainable and cost-effective, benefiting both network operators and users

Acknowledgements.
The authors would like to express their sincere gratitude to all those who contributed to the success of this research work.

References

[1] P. Rost, A. Banchs, I. Berberana, M. Breitbach, M. Doll, H. Droste, C. Mannweiler, M. A. Puente, K. Samdanis, and B. Sayadi, "Mobile network architecture evolution toward 5g," IEEE Communications Magazine, vol. 54, no. 5, pp. 84–91, 2016. https://ieeexplore.ieee.org/document/7470940

[2] V. Kohli, U. Tripathi, V. Chamola, B. K. Rout, and S. S. Kanhere, "A review on virtual reality and augmented reality use-cases of brain computer interface based applications for smart cities," Microprocessors and Microsystems, vol. 88, p. 104392, 2022. https://www.sciencedirect.com/science/article/pii/S0141933121005391.

[3] S. Mande, S. S. A. Begum, and N. Ramachandran, "Navigating the Future of Intelligent Transportation: Challenges and Solutions in 6G V2X and V2V Networks," *EAI Endorsed Transactions on Internet of Things*, vol. 11, Feb. 2025. https://publications.eai.eu/index.php/IoT/article/view/5114

[4] R. Bolla, R. Bruschi, F. Davoli, and F. Cucchietti, "Energy efficiency in the future internet: a survey of existing approaches and trends in energy-aware fixed network infrastructures," IEEE Communications Surveys & Tutorials, vol. 13, no. 2, pp. 223–244, 2010. https://ieeexplore.ieee.org/document/5522467

[5] S. Balakrishna and M.. Thirumaran, "Semantic Interoperable Traffic Management Framework for IoT Smart City Applications", EAI Endorsed Trans IoT, vol. 4, no. 13, p. e4, Jan. 2018. https://publications.eai.eu/index.php/IoT/article/view/667

[6] P. Palensky and D. Dietrich, "Demand side management: Demand response, intelligent energy systems, and smart loads," IEEE transactions on industrial informatics, vol. 7, no. 3, pp. 381–388, 2011. https://ieeexplore.ieee.org/document/5930335

[7] O. Omidvar, M. Safavi, and V. L. Glaser, "Algorithmic routines and dynamic inertia: How organizations avoid adapting to changes in the environment," Journal of Management Studies, vol. 60, no. 2, pp. 313–345, 2023. https://onlinelibrary.wiley.com/doi/10.1111/joms.12819.

[8] A. Singhrova and N. Prakash, "Vertical handoff decision algorithm for improved quality of service in heterogeneous wireless networks," IET communications, vol. 6, no. 2, pp. 211–223, 2012. https://digital-library.theiet.org/doi/10.1049/iet-com.2010.0820

[9] S. Marinova and A. Leon-Garcia, "Intelligent o-ran beyond 5g: Architecture, use cases, challenges, and opportunities," IEEE Access, vol. 12, pp. 27088–27114, 2024. https://ieeexplore.ieee.org/document/10439167

[10] Z. Latif, K. Sharif, F. Li, M. M. Karim, S. Biswas, and Y. Wang, "A comprehensive survey of interface protocols for software defined networks," Journal of Network and Computer Applications, vol. 156, p. 102563, 2020.

[11] R. K. Bharti, D. Suganthi, S. Abirami, R. A. Kumar, B. Gayathri, and S. Kayathri, "Optimal extreme learning machine-based traffic congestion control system in vehicular network," in 2022 6th International Conference on Electronics, Communication and Aerospace Technology. IEEE, 2022, pp. 597–603. https://ieeexplore.ieee.org/document/10009111

[12] F. Linsalata, E. Moro, F. Gjeci, M. Magarini, U. Spagnolini, and A. Capone, "Addressing control

challenges in vehicular networks through o-ran: a novel architecture and simulation framework," IEEE Transactions on Vehicular Technology, 2024. https://ieeexplore.ieee.org/document/10401992
[13] C. Gij´on, M. Toril, S. Luna-Ramirez, and M. L. Mar´ı-Altozano, "A data-driven traffic steering algorithm for optimizing user experience in multi-tier lte networks," IEEE Transactions on Vehicular Technology, vol. 68, no. 10, pp. 9414–9424, 2019. https://ieeexplore.ieee.org/abstract/document/8788545
[14] A. Ahmed, L. M. Boulahia, and D. Gaiti, "Enabling vertical handover decisions in heterogeneous wireless networks: A state-of-the-art and a classification," IEEE Communications Surveys & Tutorials, vol. 16, no. 2, pp. 776–811, 2013. https://ieeexplore.ieee.org/document/6587998
[15] O. Nachum, S. S. Gu, H. Lee, and S. Levine, "Data-efficient hierarchical reinforcement learning," Advances in neural information processing systems, vol. 31, 2018. https://arxiv.org/abs/1805.08296
[16] M. Liu, Z. Li, X. Guo, and E. Dutkiewicz, "Performance analysis and optimization of handoff algorithms in heterogeneous wireless networks," IEEE Transactions on Mobile Computing, vol. 7, no. 7, pp. 846–857, 2008. https://ieeexplore.ieee.org/document/4359007
[17] J. G. Andrews et al.," What Will 5G Be?", IEEE Journal on Selected Areas in Communications, vol. 32, no. 6, 2014. https://ieeexplore.ieee.org/document/6824752
[18] M. Shafi et al., "5G: A Tutorial Overview", IEEE Communications Surveys and Tutorials, vol. 21, no. 3,2017. https://ieeexplore.ieee.org/document/7894280
[19] E. Dahlman et al., 5G NR: The Next Generation Wireless Access Technology, Academic Press, 2018. https://www.ericsson.com/en/reports-and-papers/books/
[20] L. Bonati et al.," OpenRAN: Disrupting the Wireless Network Edge", Computer Networks, vol. 180, 2021. https://www.5g-networks.net/openran-o-ran-for-5g-explained
[21] M. Polese et al.," Understanding O-RAN: Architecture, State of Implementation, and Future Trends", IEEE Communications Surveys and Tutorials, vol. 24, no. 3, 2022. https://ieeexplore.ieee.org/document/10024837
[22] L. M. Larsen, H. L. Christiansen, S. Ruepp, and M. S. Berger, "The evolution of mobile network operations: A comprehensive analysis of open ran adoption," Computer Networks, p. 110292, 2024. https://www.sciencedirect.com/science/article/pii/S1389128624001245
[23] H.-T. Thieu, V.-Q. Pham, A. Kak, and N. Choi, "Demystifying the near-real time ric: Architecture, operations, and benchmarking insights," in IEEE INFOCOM 2023-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), pp. 1–8, IEEE, 2023. https://ieeexplore.ieee.org/document/10225852
[24] G. Ma, X. Wang, M. Hu, W. Ouyang, X. Chen, and Y. Li, "Drl-based computation offloading with queue stability for vehicular-cloud-assisted mobile edge computing systems," IEEE Transactions on Intelligent Vehicles, vol. 8, no. 4, pp. 2797–2809, 2022. https://ieeexplore.ieee.org/document/9964433
[25] B. Alotaibi, "A survey on industrial internet of things security: Requirements, attacks, ai-based solutions, and edge computing opportunities," Sensors, vol. 23, no. 17, p. 7470, 2023. https://www.mdpi.com/1424-8220/23/17/7470.
[26] Vijay, S. T. Siddiqui, Ritu, R. A. Kumar, A. Kumar, S. Umamaheswararao, N. Dudiki, Y. V. Reddy, and K. Dekeba, "Intertwine connection-based routing path selection for data transmission in mobile cellular networks and wireless sensor networks," Wireless Communications and Mobile Computing, vol. 2022, no. 1, p. 8398128, 2022. https://onlinelibrary.wiley.com/doi/10.1155/2022/8398128
[27] R. K. Bharti, S. S, S. V. Sumant, C. A. D. Durai, R. A. Kumar, K. Singh, H. Palivela, B. R. Kumar, and B. Debtera, "Enhanced path routing with buffer allocation method using coupling node selection algorithm in MANET" Wireless Communications and Mobile Computing, vol. 2022, no. 1, p. 1955290, 2022. https://onlinelibrary.wiley.com/doi/10.1155/2022/1955290