OPTIMIZING DYNAMIC ROUTING IN 6G NETWORKS USING A MULTI-AGENT MULTI-STEP DEEP Q-LEARNING ALGORITHM

Nachimuthu senthil[1] and sumathiarumugam[2]

[1]Department of Computer Science, KPR College of Arts, Science and Research, Coimbatore- 641407, Tamil Nadu, India
[2]Department of Information Technology, KPR College of Arts, Science and Research, Coimbatore- 641407, Tamil Nadu, India

ABSTRACT

The proliferation of 6G networks poses new challenges to traditional methods of network management due to the massive volumes of data generated and the wide variety of devices they link. A paradigm change towards AI frameworks built in Machine Learning (ML) and Deep Learning (DL) is essential due to the shortcomings of these approaches. A Speed-optimized Attention-based Hybrid Graph Convolutional Network-Long Short-Term Memory (SPAH-GCN-LSTM) model and a Reinforcement Learning (RL) framework utilizing Q-Learning (QL) were developed to forecast network congestion and enhance data transmission routes, respectively. Nevertheless, in a dynamic network, uncertainty in routing decisions could be caused by the switching between policies by a single agent. The time spent training one agent is prohibitive with the increase of the size of the network. In spite of the fact that multi-agent RLhas been utilized to alleviate this problem, classical QL can potentially face the challenge of non-stationarity that arises because of the joint learning of other agents in a multi-agent stochastic-game environment. As a result, the present manuscript presents a Multi-Agent Multi-Step Deep QL(MAMS-DQL) system that is aimed at optimizing Washington routes in 6G networks. The main goal is to come up with a decentralized mechanism where every agent is able to choose its best routing strategy independently. The multi-agent dueling deep Q-network architecture is followed in this method so as to optimize routing decisions and identify the most efficient route of the network. It also uses a multi-step experience-replay strategy, which allows agents to modify their routing strategy by taking advantage of multi-step experiences across consecutive time steps of training. Lastly, the outcomes of the simulator show that the MAMS-DQL has a higher routing efficiency compared to traditional reinforcement-learning approaches.

KEYWORDS

6G networks, SPAH-GCN-LSTM, Multi-agent RL, Deep Q-network, Multi-step experience replay strategy

1. INTRODUCTION

In earlier network structures, static routing strategies were beneficial. However, such strategies were insufficient for the complex 6G networks. 6G demands innovative network management techniques to lower latency and enable cutting-edge applications like remote surgeries and driverless cars because of its large data capacity and varied device integration [1]. 6G must expand on the features set by its predecessor, while communication service providers focus on improving monetization tactics for 5G [2]. Furthermore, crowded networks cause data traffic delays, especially during rerouting. Such latency issues may hinder the operation of latency-sensitive tasks. This underscores the importance of adaptable and intelligent routing strategies for 6G networks [3, 4]. To address these issues, AI techniques, particularly ML and DL, have been

developed for 6G networks [5, 6]. In order to handle 6G networks efficiently, LSTM networks have revolutionized predictive analytics by specializing in modelling complex, non-linear interactions in time-series data. Shi et al. [7] utilized LSTM models to adjust light pathways in diverse data centre networks and implement traffic forecasting. Computing complexity and the maintenance of long-term dependencies are two of the obstacles that LSTM networks come across, despite their strengths.

To improve network performance and maximize resource use, Tshakwanda et al. [8] used dynamic routing and predictive analytics. Combining SP-LSTM and RL for 6G system forecasting and adaptive routing led to the development of the two-tier technique. RL optimizes routing patterns according to predictive outcomes, whereas SP-LSTM forecasts network congestion, facilitating proactive actions. On the other hand, SP-LSTM might not be able to handle sudden changes in network conditions. Since both spatial and temporal interconnections are crucial in 6G networks, they might impact their ability to forecast congestion in such networks. The SPAH-GCN-LSTM was proposed in [9] to anticipate congestion in 6G networks as a solution to this problem.

The model integrates global geographical correlations with local geographical factors in traffic information via the utilization of the local and global spatial-temporal modules, which improves forecast accuracy. In order to depict the spatial-temporal connection, the global module incorporates SP-LSTM and global correlation. The local module incorporates local spatial interactions by integrating a GCN, SP-LSTM and a fully linked layer. After each module's output is combined using the soft-attention method, the most important factors that lead to accurate predictions are emphasized. It is used to inform the dynamic routing of the RL framework by the use of real-time feedback and expected congestion scenarios. The QL agent keeps on learning and modifying the best routing policies.

However, in dynamic network routing, any agent can experience unpredictability in the decision made by the routing algorithm when switching between policies or when the routing algorithm is part of a real-time application. Due to the absence of a unified network view, a single agent can have difficulty of maximising several, and often competing, objectives, an issue which becomes especially acute in the 6G large scale networks. In addition, the training of a single agent may become prohibitively long as the network grows. Multi-agent RL has recently been developed to address this issue. Conventional QL approach, however, can experience challenges related to non-stationarity created by the concurrent learning actions of additional agents in a multi-agent stochastic game (SG) context.

In this regard, the article suggests a path selection and routing decision optimization method, namely, the MAMS-DQL. The problem is modelled as a multi-agent game, where the agents (reflecting different routes) are developed to come up with an effective and scalable route selection strategy. It aims to identify the optimal decentralized routing mechanism for independent agents. The multi-agent deep Q-network is presented in this structure to enhance the routing decisions by choosing the best route in the network.

The Neural Network (NN) can model the state- value functions and utilize the advantages to find the state-action value. Improved Q-function approximation through training the deep NN with system transitions to tweak the trainable parameters is possible. At every learning step, each agent feeds the DQN with the current state and calculates the Q- values after every action. In addition, the system employs a multiple-step experience replay, which enhances the regular experience replay and allows agents to train with many consecutive intervals of experience to adjust the routing strategy. This method ensures maximum use of the time association and learning effectiveness of the model thus simplifying additional optimization of the routing plans.

In turn, the suggested MAMS-DQL process can be successfully used to improve 6G routing decisions.

Here is the structure of the remaining sections: Prior research is covered in Section 2. Section 4 demonstrates the effectiveness of the MAMS-DQL method, and Section 3 describes it in detail. Future work is discussed in Section 5, which closes this study.

2. LITERATURE SURVEY

Modern network management would be incomplete without dynamic routing, which can instantly adjust to new conditions in the network. It outperforms static routing by facilitating more flexibility in intricate, dynamic networks. AI/ML technology has significantly improved the efficiency of routing operations. The most up-to-date studies on 6G network dynamic routing using AI and ML are summarized here.

In order to decrease the average time it takes for messages to go through the core network's queuing process and select an alternate route, a novel routing method that takes into account the distribution of messages throughout the network was suggested in [10]. The backbone network reduced processing overhead at intermediate routers by consolidating several messages destined for a single router into a mailbag, thus streamlining the generation of mailbags and route-finding processes inside each router. Nonetheless, the congestion problem remains unresolved, resulting in increased delay at intermediate routers. To implement multi-objective routing in 6G networks, the authors in [11] investigated the use of the quantum approximate optimization algorithm. Nonetheless, quantum computers were incapable of performing this procedure.

A new Energy-Aware Data Collection with Routing Planning system, EADCRP-6G, was presented in [12] for 6G-enabled UAVs. To schedule routes, this method made use of an Artificial Fish Swarm-based Routing (AFSRP) technique and an Improved Red Deer Algorithm-based Clustering (IRDAC) scheme to choose the best cluster heads. In fact, there was a lot of energy lost and delay. As stated in [13], the CFTEERP was developed as a cooperative and feedback-based trustworthy energy-efficient routing protocol. The global and local trust levels of each node were determined using K-means-based feedback assessment methodologies and node attributes. Additionally, by eliminating the requirement to select the closest node for data routing, they increased the network lifespan utilizing the nearest secure node rates. Nonetheless, the computational burden remained high, necessitating the integration of ML/DL methods to enhance network performance further.

In [14], the authors addressed the issue of routing congestion among Secondary Users (SUs) in multi-hop scenarios with a known destination in the presence of Primary Users (PUs) inside 6G IoT systems. The first phase in constructing the routing traffic model was implementing the Poisson process derived from the Markov model. The routing problem was defined as the arbitrary training of non-cooperative events affecting SU routing choices. A method using distributed Non-Cooperative Learning (NCL) was subsequently used to address the problem. Conversely, packet loss and latency were high.

A Collaborative Energy-Efficient Routing Protocol (CEERP) was proposed in [15] to enhance transmission in 5G/6G wireless sensor networks. After sorting nodes according to their remaining energy, the RL chose the CH to improve data transmission. The network's performance was substantially enhanced after incorporating the CEERP with the Multi-Objective Improved Seagull Algorithm (MOISA.).However, this increased energy usage.An algorithm for load-balancing satellite routing for low Earth orbit utilizing the Markov Decision Process (MDP) was

developed in [16].The data from the current and adjacent nodes' statuses were used to iteratively identify the ideal model, considering aspects such as bandwidth consumption, latency, and the status of the current satellite node. The routing choices were unaffected by network congestion.In order to achieve end-to-end Virtual Reality (VR) communication among users, the authors of [17] presented a new concept for the effective management of VR data flow in a VR-Software Defined Network (SDN) enabled by 6G.An optimization challenge was established to enhance data rates while reducing latency and resource usage across all VR slices. However, they failed to evaluate the resource consumption and network speed. The VGradientGeocast Routing Protocol (VGGRP) was presented in [18] as a method for reducing congestion in 6G networks. They minimized delay by regularly updating the network density and buffer occupancy. Alternatively, there was a significant rate of packet loss and energy use.

Addressing long-term transmission in 5G/6G wireless sensor networks, the article [19] explained the KORP-BWKOA-SC-WSN, an improved K-means online-learning routing protocol with a black-winged kite optimization algorithm. A binarized simplicial convolutional NN was used to group the data from the sink node. To improve node coordination, the hike optimization approach chose the cluster head and then executed KORP. The optimization approach of the black-winged kite was another tool used to strengthen the system's performance. The method used a lot of energy and didn't work in large-scale heterogeneous networks.

A multi-tier multi-access edge clustering system based on fuzzy logic and QL was suggested by the authors of [20] for 5G vehicle-to-everything communication. In order to reduce vehicle contention, multi-layered multi-access edge clustering was used for clustering. A multi-hop route selection system was built using fuzzy logic and fuzzy logic techniques. Unfortunately, its performance was constrained by the use of an individual agent-based RL, which might not be suitable for networks with more nodes. Fuzzy logic was used to develop an Energy-Efficient Two-Phase Clustering and Routing Algorithm (E2PCRA) for 5G wireless networks in [21]. The fuzzy logic was used to build energy efficient clusters using the properties like residual energy, node density and signal intensity. After that, the use of fuzzy-based routing leading to the efficient paging signal transmission via the dynamic selection of the best paths took place. However, power usage and response time were still high.

For 6G networks uplink and downlink communication, the Horned Lizard Ensemble Voting Resource Allocation (HLEVRA) model was presented in [22] as an effective approach to resource allocation. To forecast user resource needs and optimise resource allocation in the NS3 environment, they used AI-based approaches. Though the number of user connections did not become minor as time went on, packet loss and communication delays remained.

ML techniques for 5G/6G network routing have been proposed in recent research, according to the examined literature. A part of these works has focused on reinforcement-learning (RL) schemes which utilize single agents to improve routing efficiency. Nevertheless, single-agent RL schemes are not suitable in the area of dynamic routing in large-scale and highly dynamic 6G networks. Overall, the 6G systems should optimize several competing goals, and the single-agent RL training does not have sufficient capabilities to resolve this issue, and the training duration grows with the size of the network. In this regard, the current research paper will be a multi-agent MAMS-DQL model combined with SPAH-GCN-LSTM to make better congestion prediction and decentralized routing decisions than single-agent RL paradigms in 6G settings.

3. PROPOSED METHODOLOGY

The MAMS-DQL method for improving 5G/6G network routing decisions is summarized in this section. Figure 1 shows an example of the proposed study. Congestion prediction in networks is initiated by means of the SPAH-GCN-LSTM model [9].

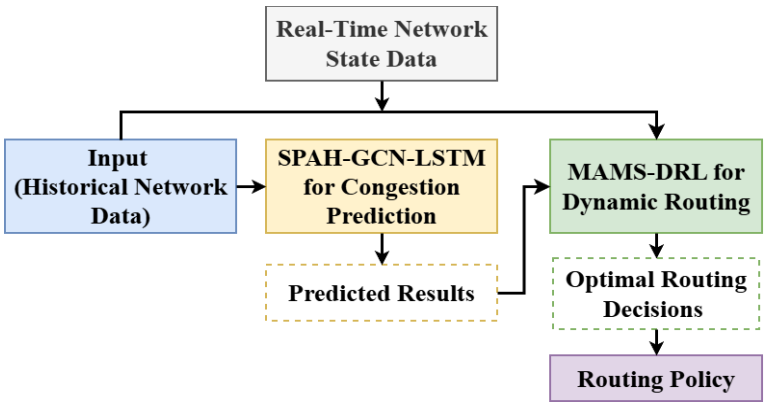


Figure 1.Building Blocks of the Suggested Research

An initial routing plan is created based on predictions from SPAH-GCN-LSTM analytics, reflecting the predicted network circumstances. Utilizing this base information, it applies RL to dynamically enhance routing choices. This research employs the MAMS-DQL algorithm, which modifies routing according to real-time network feedback.

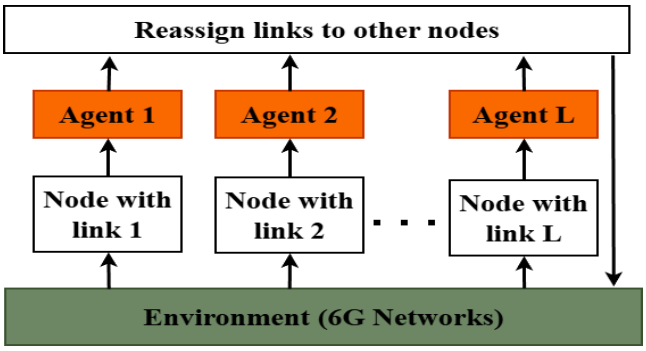


Figure 2.Diagrammatic Representation of MADRL for Route Optimization

3.1. Preliminaries on Multi-Agent Deep Reinforcement Learning

Distributed solutions, made possible by multi-agent systems, have the potential to outperform centralized systems in cost and efficiency. MADRL emphasizes sequential decision-making among multiple agents in a collaborative setting, where the actions of others impact each agent's reward. To maximize long-term benefits, agents must evaluate their counterparts' strategies. For the multi-agent circumstances, SG adjusts the MDP as a tuple $(S, U^1, \dots, U^n, r^1, \dots, r^n, \mathcal{P})$where n is the agent count and U^i, $i = 1, \dots, n$is the set of limited actions that agents can perform, with U =1indicating the mutual action set.

The agent's reward functions are denoted as $U^1 \times U^2 \times U^n$, where $i = 1, \dots, n$, and the state transition probability function is denoted as $\mathcal{P}_a: S \times U \times S \rightarrow [0,1]$.Every agent's Q-function is defined by their combined actions and strategic decisions. Additionally, in an entire cooperative

SG, the reward function is the same for all agents.Figure 2 shows the 5G/6G network block diagram for MADRL-based path optimization.

3.2. Proposed MAMS-DRL for Route Optimization

3.2.1. Problem Formation

SG models the agents' interactions and the environment changes in reaction to player actions. This study demonstrates that routing choices in 5G/6G networks may be characterized as SGs. Comprehensive definitions are presented below.

- Set of agents: In the routing optimization issue, each node is modelled as an agent, which attempts to choose an appropriate sequence of links (i.e., route) from N available links. In this novel approach, every node communicates with its wireless surroundings and selects its own path depending on the actual and expected state of the network.
- The State of space (S): Every states in S is represented as a vector and includes attributes such as link ID, time, and performance measures. The state at any particular time t in a network with N links and T time steps for throughput metrics is denoted as $s_t = [link_{1-N}, time_{1-T}, throughput_{1-T}]$.
- Action of space (A): The current configuration of the network dictates the action $a \in A$, which is associated with decisions about routing.Any of the directly connected connections can be used to route to any other node in the network, given that node i. Therefore, there will be Lpotential routing decisions in the action space for a node with L linked connections.
- Reward function(R): One possible benefit of each routing decision is related to the reward function, which is based on the expected moving average throughput of the selected links. In order to determine the optimal routing path due to the current state of the network, one uses mathematics to determine the $R(s\_t, a\_t)$ status-action pair.
- Target network: In DQN, two independent NNs, such as main and target, are considered.
- In case of specific state-action pairs, the main network generates Q-values, while the objective network generates Q-targets. In order to generate the Q-targets, the core network needs to be updated continually. At each iteration, the agent may modify the weights (θ') of the target network by using the main network's weights.
- Experience relay: These interactions can be temporarily stored in the agent's experience memory. Selecting a small subset of experiences at random from the experience memory is the next stage before updating the main network.

3.2.2. Multi-Agent Duelling Deep Q-Network Model

The conventional DQN technique calculates the value of each action inside a given state. But in certain places, a similar value function could be the outcome of different regulations. This pattern of behaviour can make it harder to figure out how to respond optimally in specific situations. To address this issue, fighting with a DQN has been suggested. Two streams of Q-functions are present in the Q-network of the dueling DQN, an enhanced variant of the DQN:

1. For each state, the state value function measures the anticipated benefit from policy π.
2. The action advantage function A assigns a relative value to each action (s, a).

In order to maximize efficiency and speed up convergence, these two streams are combined using an aggregation module and a single-output Q-function. The result is the duelling network's output. The state-action value function shows how much state*s* stands to gain from action an under policy π, while the state value function shows how effective the policy is overall in

states.Their disparity illustrates the additional benefit of selecting an in s, as delineated in Eq. (1).

$$A\hat{}\pi\ (s,a) = Q\hat{}\pi\ (s,a) - V\hat{}\pi\ (s) \quad (1)$$

Two independent data streams are utilized in this duelling network concept: The value of the action advantage $A(s,a;\theta,\alpha)$is produced by one stream, whereas the value of the state $V(s;\theta,\beta)$, is produced by the other stream.Here, θ controls how the input layer processes the given data, α refers to the state value stream, and β pertains to the advantage stream.This is the dueling network architecture representation of DQN's output:

$$Q(s,a;\theta,\alpha,\beta) = V(s;\theta,\beta) + A(s,a;\theta,\alpha) \quad (2)$$

Given that the Q-value is generated by the network, the values of V and A are left undefined.Through the utilization of central processing of actions, the identifieability is enhanced, leading to better optimization stability and maintained performance. In this context, a' denotes all potential actions, whereas $avgA(s,a';\theta,\alpha)$ signifies the average value of the advantage function every actions. A modified definition of the Q-value is outlined below:

$$\mathrm{Q(s,a;\theta,\alpha,\beta) = V(s;\theta,\beta) + A(s,a;\theta,\alpha) - avgA(s,a\hat{}';\theta,\alpha)} \quad (3)$$

The objective function is modelled as follows:

$$y = E\_((s,a,r,s\hat{}')\ )\ [r + \gamma \times Q\_T\ (s\hat{}', argmax\ \ [\![Q\_M\ (s\hat{}', a\hat{}';\theta,\alpha,\beta)]\!]\ ;\theta\hat{}',\alpha,\beta)] \quad (4)$$

This is also how the loss function is defined for adjusting the network's parameters:

$$\mathrm{L(\theta,\alpha,\beta) = E\_((s,a,r,s\hat{}')\ )\ [y - \ [\![Q\_M\ (s\hat{}', a\hat{}';\theta,\alpha,\beta)]\!]\ \hat{}2\ ]} \quad (5)$$

In Eqns. (4) and (5), In a network, θ represents the primary parameter while θ' stands for the target parameter.In the model of the duelling network, α and β are the parameters of state value function V and action advantage function A respectively.The target network's action value function is Q_T, while the main network's action value function is Q_M.Furthermore, the maximization of the Q-value of the main network in states' is represented as argmax $Q_M(s',a';\theta,\alpha,\beta)$.

3.2.3. Multi-Step Experience Replay Mechanism

In NN training, agents can make better use of their past experiences due to the experience replay buffer, which stores and retrieves these memories. In contexts with several agents, it can ensure stable training. Using a random subset of past encounters from the buffer, agents can execute batch learning and modify DQN parameters with actual interaction events using experience replay. By limiting the number of interactions and using rational strategies for environmental engagement, this improves sample efficiency and stops the model from learning just about specific contexts. This allows for a better exploration of the environment and better strategy optimization.

Information gathered at random during experience replay can be represented as a tuple (s_t, a_t, r_t, s_{t+1}) that includes the action, current state, and reward following state.But selecting a single element of network data at random can ruin the state sequence's temporal correlation. At any one moment, the network's current status could be intricately related to its previous or subsequent states, depending on the route (link) selection process. The loss of temporal structure in the current state information due to random sampling could reduce learning efficacy.

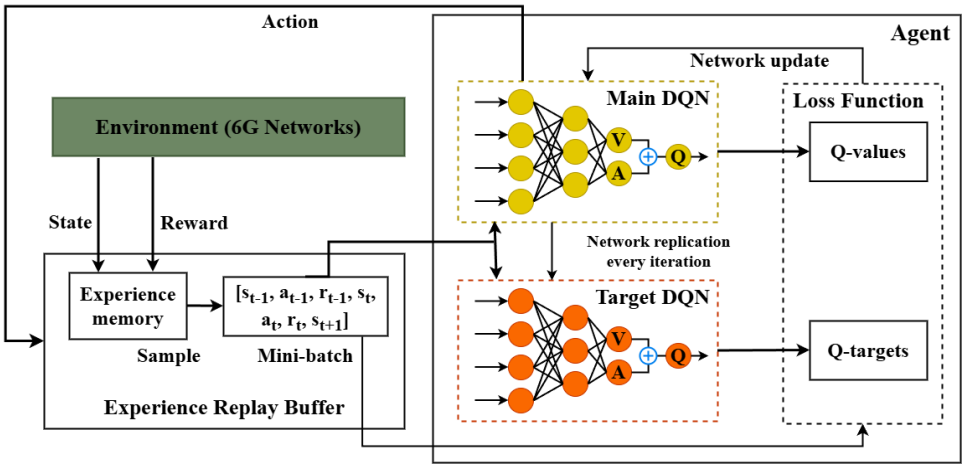


Figure 3.Architecture of MAMS-DRL Model for Path Optimization

Table 1. Parameter Settings for MAMS-DRL

Parameters	Values
Rate of learning (α)	0.01
Reduction factor (γ)	0.9
Maximum greedy factor (ϵ)	0.9
Number of episodes (N_{epi})	100
Size of memory buffer (N_D)	20000
Size of mini-batch	128
Memory replay period (v)	4 iteration steps
Length of memory replay period	5000
Number of hidden layers	2

In certain cases, a continuous chain of states may give more environmental context than a randomly selected set of states. In this study, the dueling DQN is employed in conjunction with a multi-step experience replay approach that places an emphasis on tracking a series of continuous network states instead of discrete nodes.

Updating the utilization of temporal data and making it easier to acquire more contextual insight are both made possible by maintaining a certain amount of continuous state experiences.Figure 3 shows the complete MAMS-DRL model architecture. Table 1 also contains the parameters and settings for this MAMS-DRL.

The pseudocode for this MAMS-DRL-based path optimization is given in Algorithm 1.

Algorithm 1: MAMS-DRL-Based Path Optimization
Input: Real-time network state data and predicted network congestion status
Output: Optimal path with stable links

1. Begin
2. For every agent, randomly assign weights θ to the main network $Q(s,a|\theta)$and weights $\hat{\theta} = \theta$ to the target network $Q(s,a|\hat{\theta})$.
3. Set the size of the multi-step experience memory D to N_D;
4. Set the state and action;
5. for every episode e in the interval $[1, N_{epi}]$.

6. Remove all previous states and restore the beginning state s_t;
7. for every value of t in the interval $[1, T]$
8. $for(each\ agent\ i = 1{:}\ N)$
9. action←MARDL.action(s);
10. Choose path optimization as a_i and execute a_i;
11. $end\ for$
12. Track the outcome $r_t$and the state that follows s_{t+1};
13. s_t is equal to s_{t+1};
14. Keep the following information in the experience replay buffer: $[s_{t-1}, a_{t-1}, r_{t-1}, s_t, a_t, r_t, s_{t+1}]D$;
15. for every agent □ from 1 to □
16. Randomly select □ samples from □ to form a mini-batch;
17. Set □ as Eq. (4);
18. Update the main network by minimizing loss $\mathcal{L}(\square_{\square}, \square, \square)$ based on gradient descent scheme;
19. terminate after
20. It is necessary to update the target network parameters $\widehat{\square}$ for every agent and every □ steps.
21. terminate after
22. terminate after
23. Finish

4. RESULTS AND DISCUSSION

In this part, compared the MAMS-DRL method to other 6G routing algorithms, including EADCRP, CFTEERP, CEERP, DQNLLRA, VGGRP, and E2PCRA. Running Windows 10 64-bit, the virtual desktop included a 1 TB hard drive, 8 GB of RAM, and an Intel® CoreTM i5-4210 CPU with a clock speed of 3 GHz.

4.1. 5G/6G Simulation Environment

The Python modifies the 6G network design by incorporating nodes, connections, and critical components. At the user level, the UPF handles traffic optimization, the SMF handles accessibility and mobility, and the gNB provides base station operations, transmits signals, and manages radio resources. User components are able to resemble real-life mobility situations by making use of several mobility models. The integration of technologies such as web slicing, edge systems, SDNs, VNFs, massive MIMO, and mmWave communications allows for the evaluation of methods for system control and adjustment. The experimental setup mimics the characteristics of future 6G applications and infrastructures by including real-world traffic patterns and other kinds of traffic, such as interactions between the Internet of Things (IoT), mission-critical systems, and ultra-high-definition audio-visual streaming.

4.1.1. Network Topology

The replication of the heterogeneity and size interconnectedness of real 6G networks is the main objective of the suggested network design. Due to the lack of a definitive 6G Network design, this topology is based on existing and future networking technologies. Instead than accurately depicting an L-3 topology, the design places an emphasis on connectedness. To employ the full-mesh setup as a theoretical paradigm to delineate ideal connection and related intricacies. Common measures used to evaluate the MAMS-DRL algorithm's performance include computational overhead, resource consumption, average throughput, and percentile rate among nodes.

4.1.2. Robustness Evaluation for Path Optimization Using MAMS-DRL Algorithm

Improving network management is possible by comparing the link order preceding and following predictions in relation to moving_mean_throughput is listed in Table 2.The rearrangement of connections indicates possible network congestion problems, enabling preemptive resource modifications. By comparing real-time network data with predictions of future congestion, the MAMS-DRL algorithm finds the optimal paths for data transmission across networks.

To better manage networks, the moving_mean_throughput evaluates link rankings both preceding and following congestion forecasts using SPAH-GCN-LSTM.

The best paths for data transfer between gNB nodes are found using the MAMS-DRL. The efficiency of the route is assessed by total_rewards and route_length. The total rewards determine the efficiency of route selection by considering both immediate and prospective benefits. Meanwhile, route_length dictates transmission velocity and resource consumption, both of which are crucial in time-sensitive scenarios.

Table 2. Pre- and Post-Congestion-Prediction Organization of Network Links

Order	Referral ID (Before Prediction)	Referral ID(After Prediction)
1	5	14
2	10	8
3	2	9
4	1	6
5	14	-
6	3	2
7	6	-
8	8	4
9	4	-
10	9	-

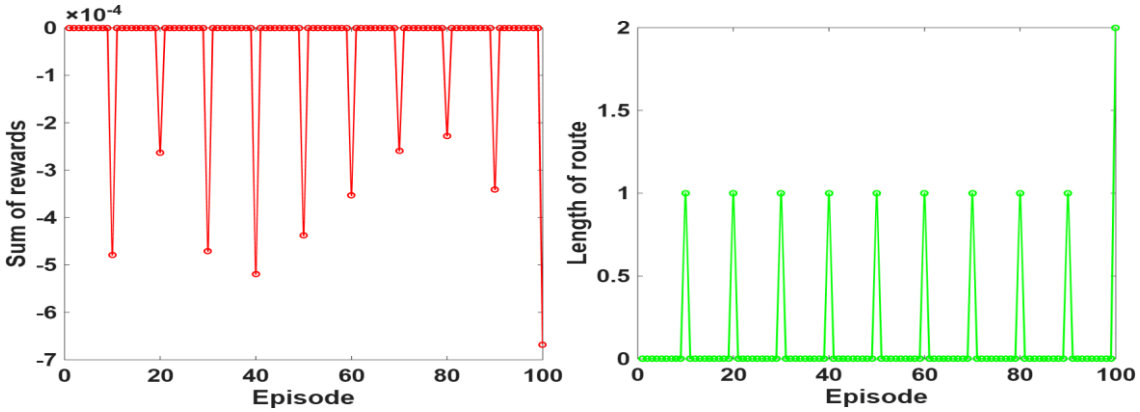
The findings shown in Figure 4 and Table 3 indicate that a greater number of episodes exhibited an absence of a stable route discovery, as evidenced by zero values for both incentives and route length. The agents found the best possible routes with the highest possible rewards in episodes 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100.

All episodes, with the exception of episode 100, had a routelength of 1, illustrating the algorithms' efficacy inidentifying optimal routes. Episode 100 serves as a learning signal for the agents to choose shorter routes, with a route length of two and a more substantial negative payout. Even when several instances fail to provide favorable paths, the MAMS-DRL algorithm consistently identifies optimal paths, demonstrating its robustness.

Table 3. Episodes with Non-Zero Total Rewards and Route Length

Episodes	Total rewards	Route length
100	-0.000668107	2
90	-0.000341379	1
80	-0.000228051	1
70	-0.000259624	1
60	-0.000353107	1
50	-0.000437912	1

40	-0.000519438	1
30	-0.000471090	1
20	-0.000263251	1
10	-0.000478953	1



(b)

Figure 4. Dissimilarities of Total Rewards and Path Length per Episodes in MAMS-DRL for Path Optimization

The MAMS-DRL algorithm's success in optimizing resource use is seen in Table 4, which indicates a total weight of around 0.2056 for the optimal route. This weight, computed from individual link weights, illustrates the significant contributions from the gNB_7 to the L3_Switch_5 connections and the L3_Switch_5 to the gNB_43 link. Additionally, a mean reward of -4.02e-05 suggests that the system may optimize pay-outs along certain trajectories.

The method finds quicker and more effective paths with lower link traversals than other methods, with a mean route length of 0.1.The anticipated congestion levels prompt the MAMS-DRL to adjust weights dynamically. Consequently, network management becomes more flexible, enabling dynamic route modifications in response to changing network conditions.

Table 4. MAMS-DRL in Action: Validating Resource Use via Choosing Optimal Path

Matrices	Values
Finest path	[gNB_7, L3_Switch_5, gNB_43]
Weight of(gNB_7, L3_Switch_5)	0.07793
Weight of (L3_Switch_5, gNB_43)	0.128
Total weight	0.20251
Average reward	-4.02e-05
Mean route length	0.1

4.1.3. Evaluation Measures

The efficacy of dynamic optimal routing is evaluated using the following metrics:

- Mean throughput: It is the amount of data sent to the destination node in a certain amount of time that has been successfully transmitted.

$$\text{Mean throughput} = \sum_{i=1}^{n} d_i / t \quad (6)$$

Regarding Eq. (6), Amount of accurate data transfers is denoted by n, data transmitted in the i^{th} transfer is denoted by d_i, and time is represented by t.

- Usage of resources: In order to optimize the path, we measure the computation, memory, and bandwidth consumption of the MAMS-DRL. These metrics are determined using Eq. (7).

$$\square_{\square\square\square\square\square} = \square_{\square\square\square} + \square_{\square\square\square\square\square\square} + \square_{\square\square\square\square\square\square\square\square\square_h} \quad (7)$$

In Eq. (7), $\square_{\square\square\square}$, $\square_{\square\square\square\square\square\square}$, and $\square_{\square\square\square\square\square\square\square\square\square_h}$ are the resources utilized by the CPU, memory, and bandwidth, respectively, for the MAMS-DRL's communication with the nodes.

- The average throughput rate as a percentage of all nodes: In a system, it specifies the amount of nodes achieve a certain performance measure, such throughput.
- Computation overhead: It is the additional time needed by the MAMS-DRL algorithm to make appropriate routing decisions.

4.1.4. Simulation Results

Results from the MAMS-DRL's comparison with the EADCRP, CFTEERP, CEERP, DQNLLRA, VGGRP, and E2PCRA simulations are shown in this section. Figure 5 shows the mean throughput of proposed and current approaches for path optimization in 5G/6G networks, using actual-time network data and the congestion status predictions from the SPAH-GCN-LSTM. The MAMS-DRL enhances the mean throughput by 78.13%, 53.11%, 43.31%, 34.55%, 23.18%, and 11.67% in comparison to the EADCRP, CFTEERP, CEERP, VGGRP, E2PCRA, and DQNLLRA, respectively. The MAMS-DRL algorithm delivers superior performance by dynamically and decentralizing changing routing rules. This is achieved by efficiently organizing destinations according to relational BS linkages, hence facilitating agent collaboration and enhancing throughput. Figure 6 illustrates a comparison of percentile rates across nodes for the proposed and current approaches for route optimization in 6G networks. The proposed MAMS-DRL enhances the percentile throughput across nodes by 46.15%, 33.8%, 23.38%, 13.1%, 6.74%, and 3.26% relative to the EADCRP, CFTEERP, CEERP, VGGRP, E2PCRA, and DQNLLRA, respectively.

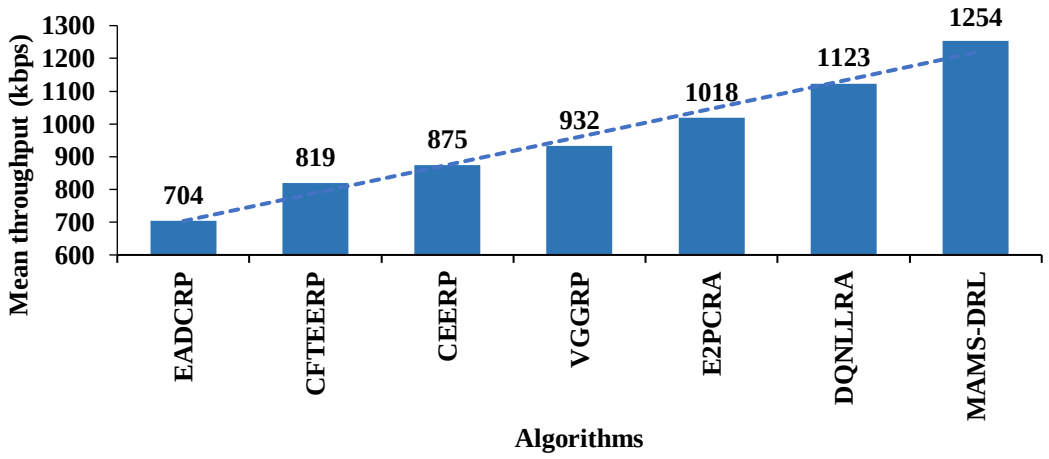


Figure 5. Analysis of Mean Throughput

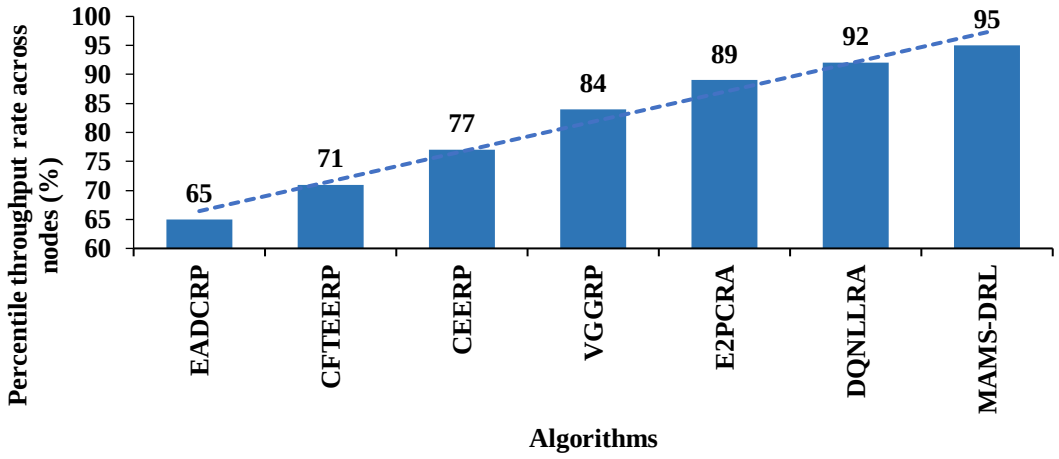


Figure 6. Evaluation of Percentile Throughput Rate Across Nodes

Figure 7 illustrates the comparative analysis of several 6G route optimization techniques regarding resource use. The proposed MAMS-DRL algorithm reduces resource use by 66.67%, 63.79%, 59.62%, 53.33%, 43.24%, and 25% when compared to EADCRP, CFTEERP, CEERP, VGGRP, E2PCRA, and DQNLLRA, respectively. The MAMS-DRL algorithm does this by reducing unnecessary resource use during routing via localized decision-making.

Figure 8 compares the computational overhead of several 6G route optimization algorithms. The findings indicate that, in comparison to EADCRP, CFTEERP, CEERP, VGGRP, E2PCRA, and DQNLLRA, the proposed MAMS-DRL substantially reduces computational overhead by 66.1%, 60.78%, 56.52%, 48.72%, 37.5%, and 23.08%, respectively. By combining MADRL with amulti-step experience replay mechanism in 6G routing, the MAMS-DRL algorithm achieves outstanding results by balancing networking speed and resource efficiency.

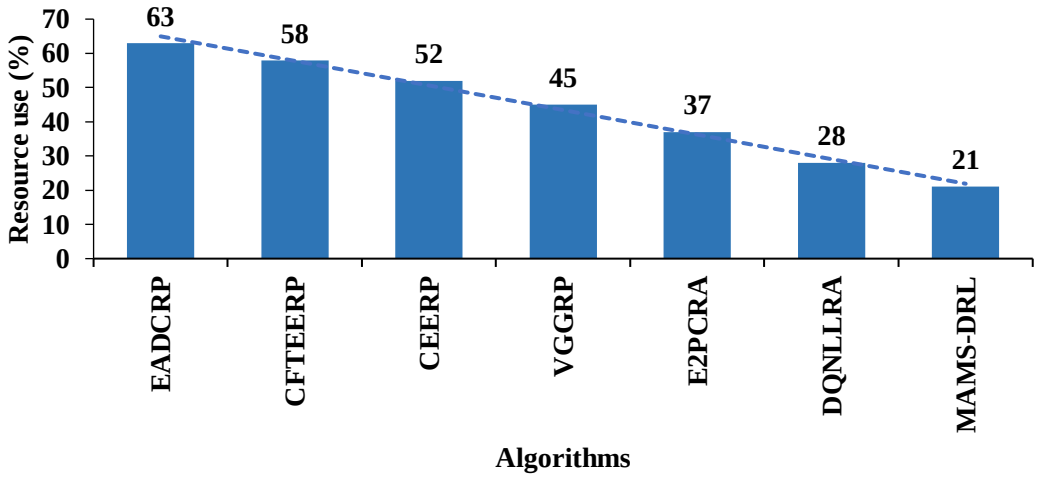


Figure 7. Evaluation of Resource Use

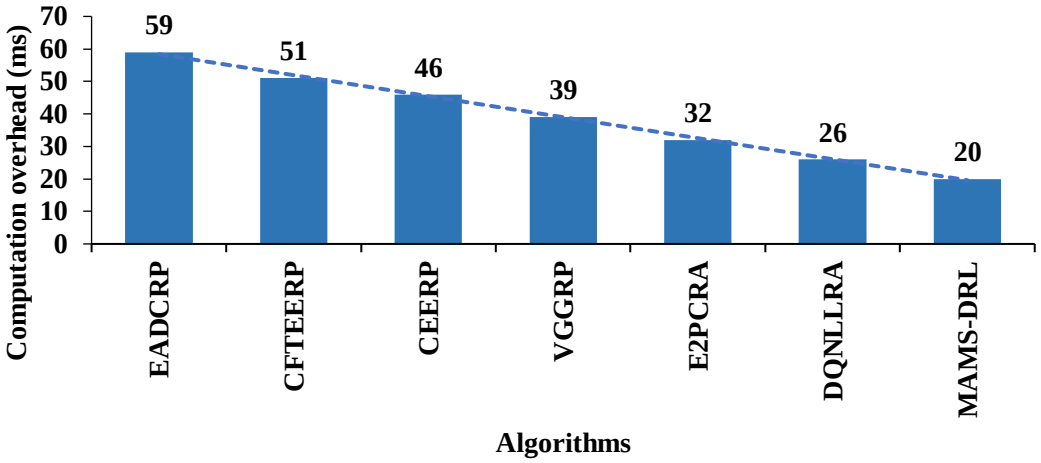


Figure 8. Evaluation of Computation Overhead

4.2. Ablation Study

4.2.1. Results from MAMS-DRL Simulations in 6G Wireless Sensor Networks

The settings for the WSN simulation are displayed in Table 5. With an initial energy of 100 joules each, the 500 node of sensors were dispersed throughout a 500 x 500 m2 region. There were 1200 simulation iterations, and each data packet were 500 bytes in size.

Table 5. Simulation parameters for WSN

Parameter	Value
Rounds of Simulation	1200
Deployment Zone	500 × 500 m2
Size of data packet	500 bytes
Staring energy for each node	100 Joules
Quantity of sensor nodes	500
Simulation time	200 sec
Node transmission range	10m
Mobility Model	Random waypoint

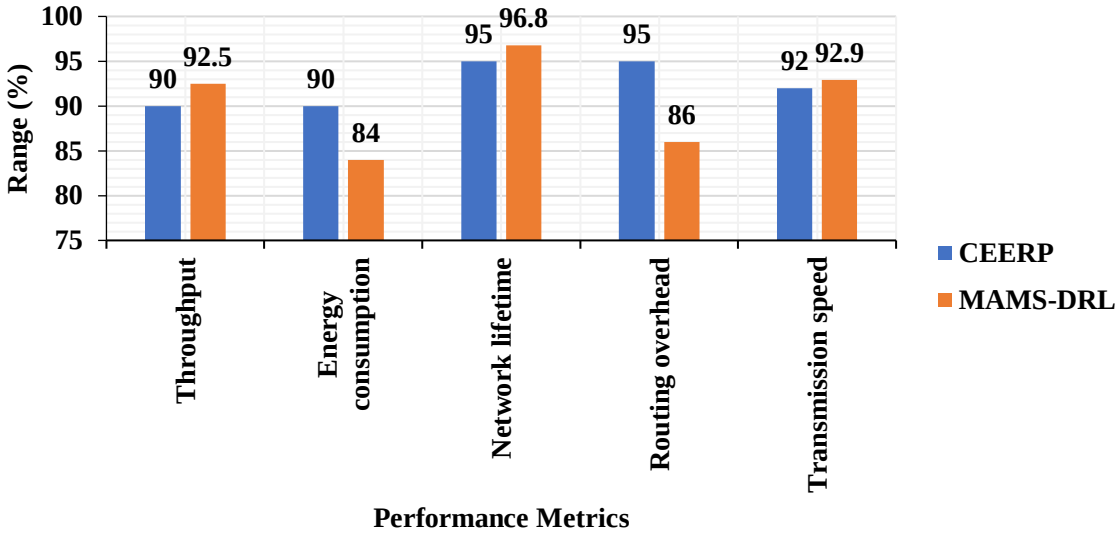


Figure 9. Evaluation of MAMS-DRL and CEERP in 6G WSN Scenario

Throughput, energy consumption, transmission velocity, network lifespan, routing overhead, and MAMS-DRL together with the CEERP [15] are evaluated in Figure 9 for 5G/6G WSNs. Energy consumption refers to the amount of power needed to send and receive data packets. The lifetime of a network is the amount of time it takes to finish a task successfully. Additionally, it specifies when the node stops responding during the data transmission. The term "routing overhead" describes the amount of data packets transferred for the purposes of route discovery and maintenance. The rate at which information packets are transmitted between computers is called the rate of transmission of a network connection.

The results show that the suggested MAMS-DRL is more efficient than the CEERP in every metric. This is because the pathways between both destination and source nodes for data transmission have been optimized. To significantly improve the network's persistence and energy efficiency, it is essential in 5G/6G networks to determine the best shortest route for data transfer between nodes. For secure data transfer in 5G/6G WSNs, this MAMS-DRL method is ideal.

4.2.2. Simulation Results of MAMS-DRL in 6G Vehicle Ad hoc Networks

The suggested MAMS-DRL model is tested with the 6G oriented Vehicle Ad hoc protocols VGGRP in a 6G network scenario for Vehicle Ad hoc Networks. The parameters that were initialized are displayed in Table 6.

Table 6. Parameters for the simulation and their values

Requirements for simulation	values
The quantity of vehicle nodes	200
Starting Energy for each node	300J
Range of Bandwidth	2–20 MHz
Size of the Block	30KB
Epochs	10
Learning Rate	0.01

Deployment Zone	500 × 500 m2
Count of Road side unit	10

Table 7. Evaluation of MAMS-DRL and VGGRP in 6G VANET Scenario

Metrics	VGGRP	Proposed MAMS -DRL
Throughput (kbps)	210	235
Energy consumption (J)	33	30.1
PDR (%)	91	92.3
Packet drop ratio (%)	15	13.6
End-to-end delay (Ms)	5950	5572
Network lifetime (sec)	230	264

Energy consumption, Throughput, packet loss rate, end-to-end latency and network lifetime are the primary metrics examined in Table 7, which outlines the effectiveness of MAMS-DRL utilizing the VGGRP inside a 6G VANET scenario including 200 vehicle nodes.The findings demonstrate that MAMS-DRL improves routing efficiency in VANETs relative to VGGRP. While the VGGRP mitigates network congestion, it is unable to assimilate both geographical and temporal attributes from the traffic data.

The SPAH-GCN-LSTM successfully learns spatiotemporal correlations in traffic data to anticipate and mitigate congestion, while the MAMS-DRL identifies the ideal route based on the expected congestion results. This leads to increased throughput, packet delivery ratio, and network longevity while significantly diminishing packet loss and end-to-end latency. Consequently, the MAMS-DRL is suitable for 6G VANETs to enhance routing efficiency.

5. CONCLUSION

The MAMS-DQL algorithm was created in this research to improve routing choices in 6G networks. In the beginning, the SPAH-GCN-LSTM network was able to predict if there would be network congestion. The multi-agent battling DQN model was then used to improve routing decisions and determine the optimal path for the network. Agents can adjust the training routing strategy by referencing past experiences from successive time steps, due to a multiple-step experience replay technique. Lastly, simulation results demonstrated that the MAMS-DQL outperformed the traditional RL techniques. In comparison to the traditional techniques, this MAMS-DRL increased throughput by 1254 kbps on average, reduced resource consumption by 21%, and reduced calculation overhead by 26 milliseconds. The percentage of throughput across nodes is 95%.In contrast, determining the appropriate discount factor to balance current and future rewards may be difficult in dynamic network situations. Future study will concentrate on using a metaheuristic approach to choose the right parameters, including the discount factor, for MAMS-DRL, resulting in lower computing overhead and improved route optimization efficiency.

CONFLICTION OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] Ishteyaq, I., Muzaffar, K., Shafi, N., &Alathbah, M. A. (2024) "Unleashing the power of tomorrow: exploration of next frontier with 6G networks and cutting-edge technologies", IEEE Access, Vol. 12, pp. 29445–29463.

[2] Chataut, R., Nankya, M., & Akl, R. (2024) "6G networks and the AI revolution—exploring technologies, applications, and emerging challenges", *Sensors*, Vol. 24, No. 6, p. 1888.

[3] Bhide, P., Shetty, D., & Mikkili, S. (2025) "Review on 6G communication and its architecture, technologies included, challenges, security challenges and requirements, applications, with respect to AI domain", *IET Quantum Communication*, Vol. 6, No. 1, p. e12114.

[4] Das, S. R., Sarma, S. S., Khuntia, M., Roy, I., Sinha, K., & Sinha, B. P. (2022) "A novel routing strategy towards achieving ultra-low end-to-end latency in 6G networks", *International Journal of Computer Networks & Communications (IJCNC)*, Vol. 14, No. 1, pp. 1-24.

[5] Oukebdane, M. A., Shah, A. S., Azad, A. K., Ekoru, J., &Madahana, M., (2025) "Unraveling the nexus of ML and 6G: challenges, opportunities, and future directions", *IEEE Access*, Vol. 13, pp. 114934–114958.

[6] Alhammadi, A., Shayea, I., El-Saleh, A. A., Azmi, M. H., Ismail, Z. H., Kouhalvandi, L., & Saad, S. A., (2024) "Artificial intelligence in 6G wireless networks: opportunities, applications, and challenges", *International Journal of Intelligent Systems*, Vol. 2024, No. 1, p. 8845070.

[7] Shi, H., & Wang, C., (2018) "LSTM-based traffic prediction in support of periodically light path reconfiguration in hybrid data center network", *In IEEE 4th International Conference on Computer and Communications*, pp. 1124–1128.

[8] Tshakwanda, P. M., Arzo, S. T., &Devetsikiotis, M., (2024) "Advancing 6G network performance: AI/ML framework for proactive management and dynamic optimal routing", *IEEE Open Journal of the Computer Society*, Vol. 5, pp. 303–314.

[9] Senthil, N., & Arumugam, S., (2025) "Leveraging global and local spatial-temporal correlations of traffic to improve congestion prediction and routing in 6G networks", *International Journal of Computer Networks and Applications*, Vol. 12, No. 1, pp. 93–105.

[10] Das, S. R., Sarma, S. S., Khuntia, M., Roy, I., Sinha, K., & Sinha, B. P., (2022) "A novel routing strategy towards achieving ultra-low end-to-end latency in 6G networks", *International Journal of Computer Networks & Communications*, Vol. 14, No. 1, pp. 1–24.

[11] Urgelles, H., Picazo-Martinez, P., Garcia-Roger, D., & Monserrat, J. F., (2022) "Multi-objective routing optimization for 6G communication networks using a quantum approximate optimization algorithm", *Sensors*, Vol. 22, No. 19, p. 7570.

[12] Duhayyim, M. A., Obayya, M., Al-Wesabi, F. N., Hilal, A. M., Rizwanullah, M., & Eltahir, M. M., (2022) "Energy aware data collection with route planning for 6G enabled UAV communication", *Computers, Materials & Continua*, Vol. 71, No. 1, pp. 825–842.

[13] Gayathri, A., Prabu, A. V., Rajasoundaran, S., Routray, S., Narayanasamy, P., Kumar, N., & Qi, Y., (2022) "Cooperative and feedback based authentic routing protocol for energy efficient IoT systems", *Concurrency and Computation: Practice and Experience*, Vol. 34, No. 11, p. e6886.

[14] Malik, T. S., Malik, K. R., Sanaullah, M., Hasan, M. H., & Aziz, N., (2022) "Non-cooperative learning based routing for 6G-IoT cognitive radio network", *Intelligent Automation and Soft Computing*, Vol. 33, No. 2, pp. 809–824.

[15] Gururaj, H. L., Natarajan, R., Almujally, N. A., Flammini, F., Krishna, S., & Gupta, S. K., (2023) "Collaborative energy-efficient routing protocol for sustainable communication in 5G/6G wireless sensor networks", *IEEE Open Journal of the Communications Society*, Vol. 4, pp. 2050–2061.

[16] Dong, F., Song, J., Zhang, Y., Wang, Y., & Huang, T., (2023) "DRL-based load-balancing routing scheme for 6G space–air–ground integrated networks", *Remote Sensing*, Vol. 15, No. 11, p. 2801.

[17] Naguib, K. M., Ibrahim, I. I., Elmessalawy, M. M., & Abdelhaleem, A. M., (2024) "Optimizing data transmission in 6G software defined networks using deep reinforcement learning for next generation of virtual environments", *Scientific Reports*, Vol. 14, No. 1, p. 25695.

[18] Sahoo, A., & Tripathy, A. K., (2025) "Congestion avoidance in 6G networks with V gradient geocast routing protocol", *Scientific Reports*, Vol. 15, No. 1, p. 595.

[19] Gupta, N., Mazzei, M., Mäkelä, J., &Uitto, M., (2025) "Optimized K-means routing protocol with black-winged kite algorithm for sustainable 5G/6G sensor networks", Internet of Things, p. 101792.

[20] Alagumani, S., & Natarajan, U. M., (2025) "Q-learning and fuzzy logic multi-tier multi-access edge clustering for 5G V2X communication", *Network: Computation in Neural Systems*, Vol. 36, No. 1, pp. 174–197.

[21] Meher, S., Sahu, B. J. R., Dash, S., & Mohanty, B., (2025) "E2PCRA: a fuzzy-based energy-efficient paging using clustering and routing algorithm for 5G wireless networks", *IEEE Access*, pp. 1–26.

[22] Kavyashree, M. K., &Shankaraiah, N. N. (2026) "HLEVRA: An efficient resource allocation strategy for uplink and downlink in 6G networks", *International Journal of Computer Networks & Communications (IJCNC)*, Vol. 18, No. 1, pp. 89–108.

AUTHORS

Mr. N.Senthil is working as an Assistant Professor in Kangeyam Institute of commerce. he has completed his M.Sc. (Computer Science) in the year April 2005 and M.Phil. in networks in the year 2008.he has 17 years of academic experience. Currently he is her Ph.D. in networks from KPR College of Arts Science and Research, Coimbatore.

Dr. A. Sumathi is an Associate Professor and Head of the IT Department at KPR College of Arts, Science and Research in Coimbatore. She completed her PG degree in 2003, her M.Phil. in 2005, and her Ph.D. in 2019. She has 21 years of experience in the teaching field, with a specialization in Data Mining and Machine Learning.