Journal of Modern Engineering and Technology Vol: 1, No 2, 2026, Page: 1-12

Research Article

A Lightweight AI-Driven Intelligent Routing Framework for 6G-Enabled IoT Networks

Nibras Jomaah Mohammed khalaf*

Computer Science Department, Al-Imam Al-Adham University College, Baghdad, Iraq
Corresponding Email: nibras.juma@imamaladham.edu.iq

Received: 2 December 2025 **Revised:** 8 January 2026 **Accepted:** 27 January 2026

Abstract*:* Internet of Things (IoT) is rapidly growing as the number of connected devices is growing, as well as the need to achieve low latency, high reliability, and high energy consumption communication. As sixth generation (6G) networks are being proposed, the concept of using artificial intelligence within routing protocols has become critical to making IoT systems smarter and self-organising. The present paper suggests a lightweight AI-based routing framework, which uses Tiny Machine Learning (TinyML) models to enable IoT nodes to take local and intelligent routing decisions without imposing computer resource load and raising energy consumption. The proposed framework should achieve energy efficiency, enhance the proportion of packet delivery, and minimize the end-to-end delay, which is done by choosing best routes to be used considering the network parameters like residual energy, signal strength, and degree of congestion. Matlab was used to simulate the framework and compare it to the conventional routing protocols such as AODV, and RPL. The findings indicate that the energy use has greatly improved with a 29 percent reduction in the energy use, a 12 percent increase in the ratio of packet delivery, and a 35 percent decrease in the delay as compared to the traditional techniques. These results emphasize the usefulness of lightweight machine learning methodologies in 6G-enabled IoT networks, and thus the suggested approach is applicable to the application of smart cities, industrial sensor networks, and other massive IoT implementations.

Keywords: Tiny Machine Learning (TinyML); 6G-enabled IoT; Intelligent Routing; Energy-efficient Networks; Lightweight ML; Edge Computing

1. INTRODUCTION

The current fast growth of Internet of Things (IoT) has resulted in the increased number of connected devices that are implemented in smart cities [1], industrial automation, health systems, and next-generation intelligent infrastructures. Since billions of heterogeneous sensors and actuators are being used at once, the contemporary IoT networks have acute problems with latency, reliability, spectrum congestion, energy efficiency, and long-term sustainability of a network. The traditional routing schemes, including AODV, RPL, and other geographic or hierarchical schemes were initially tailored to the previous wireless networks and are currently not scalable to ultra-dense, dynamic and power-constrained IoT settings. They do not have the flexibility and intelligence needed to support a quick topology adaptation, congestion states and varied link behavior, which leads to high delay, energy wastage and poor packet delivery characteristics [2,3].

The advent of the sixth generation (6G) wireless network is anticipated to advance IoT communication into highly autonomous, ultra-low latency, and AI-native structures that in combination will be able to support real-time decision-making and active network optimization. 6G will bring transformative features, such as sub-millisecond latency, terahertz capacity communication, machine-type connectivity on a massive scale (mMTC) and edge intelligence that can be readily converted into real-time decision-making [4]. It is no longer a luxury to embed artificial intelligence into routing protocols in such environments, but it is a necessity to evolve self-organizing context-aware communication. Nevertheless, conventional AI models require a lot of computing power, memory, and energy resources that are not compatible with microcontroller-based IoT nodes that have a tight resource constraint. To solve this problem, Tiny Machine Learning (TinyML) has become a potential solution by facilitating small, energy-saving AI models that can be deployed directly on low-power IoT nodes without relying on the cloud. TinyML has the benefits of rapid local inference, low processing load and very low power consumption, enabling it to be used in large scale routing decisions in IoT networks [5]. By implementing TinyML in routing systems, the nodes can estimate network dynamics, including remaining energy, signal strength, congestion, and link stability, and make autonomous decisions on the best forwarding paths [6,7].

In this paper, we present an intelligent routing model, which is a lightweight AI-based framework focused on 6G-based IoT networks. The architecture uses TinyML models that enable every IoT satellite to make smart and local routing decisions with a small memory requirement and a low cost of computation. The goal is to improve energy efficiency, improve the ratio of packet delivery, and minimise end-to-end delay by facilitating predictive and context-sensitive route selection. The implementation and the simulation of the proposed solution are performed in MATLAB and compared to the traditional routing protocols such as AODV and RPL. The outcomes indicate that the performance metrics can be enhanced significantly, such as energy consumption is reduced by 29 percent, the ratio of packets delivered increases by 12 percent, and the delay is reduced by 35 percent to illustrate the success of the integration of lightweight machine learning methods into IoT routing schemes of the next generation.

The rest of the paper is organized as follows. Section 2 deals with related works, Section 3 show the proposed lightweight ai-driven intelligent routing method, while Section 4 Experimental Design. Section 5 deals with discusses, while Section 6 concludes the paper.

2. RELATED WORK

The study of machine learning resource integration to IoT systems with resource constraints and the creation of intelligent routing mechanisms in massive networks has grown tremendously in the past few years. With the advent of Tiny Machine Learning (TinyML), it has become possible to run machine learning models on microcontrollers with localized intelligence without having to use cloud computing. As illustrated by Dutta et al. [8], TinyML is ultra-low in power consumption, latency, and high-feasibility of embedded applications.

Likewise, Heydari et al. [9] give the recent survey of the TinyML deployment issues and performance, which confirms that TinyML can be used to execute workloads related to real-time embedded applications, albeit model size and inference cost have to be kept to the minimum. Based on these ideas, Pereira et al. [10] created a TinyML model to classify the water-potability problem demonstrating that complex sensing tasks can be efficiently implemented on IoT devices on microcontrollers using very low energy budgets. TinyML is also confirmed in other recent works and is used to conduct security monitoring, anomaly detection, and environmental sensing of the IoT systems [11], which proves the maturity of the new paradigm.

Simultaneously, exploring the use of ML-based routing and intelligent network control in the context of IoT and next-generation (6G-ready) networks has become widespread. Indicatively, Aktas et al. [12] conduct a review of the ways that reinforcement learning and deep neural models can play a significant role in enhancing routing performance in both dynamic and dense IoT settings. Priyadarshi et al. [13] are showing that AI-based routing enhances fault tolerance, scalability, and QoS in wireless sensor networks, whereas Das et al. [14] present an AI-optimized routing protocol modifying paths according to the conditions of the network to minimize energy consumption. Going down to a more architectural scale, Zormati et al. [15] propose an ML-driven distributed softwarization platform to improve flexibility of routing and coordination between resources in large-scale IoT systems. Likewise, the next-generation networks and AI-based routing have been explored in the literature review of hybrid optimization strategies of 6G-enabled IoT traffic and multi-routing methods [16].

Even though these works show a significant improvement, there are still a number of limitations. The majority of ML-based routing schemes are based on computationally expensive models, or they need centralized controllers, which are not applicable when using microcontroller-class IoT devices with limited memory and energy capabilities. TinyML studies, though developed and capable of sensing tasks, have been hardly used in routing decision-making, where the models have to be very lightweight and run continuously, meeting real-time requirements. The security issues in the deployment of ML in embedded systems indicated by Huckelberry and colleagues [17] only underscore the importance of having routing systems that are efficient and robust. In addition, the available literature rarely takes into account energy, latency, link stability, congestion, and computational overhead together in one routing architecture.

In this regard, a definite research gap still persists: there is still no completely distributed, ultra-lightweight TinyML-based routing protocol that is explicitly tuned to 6G-enabled IoT networks. This is where the current work is motivated and presents a new TinyML-based intelligent routing algorithm that can be deployed under the restricted energy and memory and real-time constraints of the modern IoT system with the added benefit of substantially enhancing the network performance.

3. PROPOSED LIGHTWEIGHT AI-DRIVEN INTELLIGENT ROUTING METHOD

The proposed approach presents a low-weight smart routing system that specifically focuses on large-scale sensors that consume energy-constrained IoTs, powered by 6G communication systems. The framework allows every IoT node to make autonomous decisions on routing but with a small TinyML model that analyses real-time network status. The main point is to substitute traditional non-adaptable routing conduct with spatial decision making, surrounding local choice, and at the same moment keep a very slight computational expense, achievable by microcontroller-level systems.

3.1 Overview of the Architecture

The lightweight intelligence module of each IoT node includes:

1. Local Network Monitor- measures qualities like the quality of the links, remaining power, the load of the queue and the availability of the neighbors.
2. TinyML Inference Engine- it implements a micro-model that returns the optimal forwarding node.
3. Adaptive Routing Controller - provides updated forwarding decisions, avoids links which are unstable and responds to abrupt changes in topology.

The routing is fully decentralized: each node considers its direct neighbors and chooses the next hop without having any global topology knowledge and cloud support.

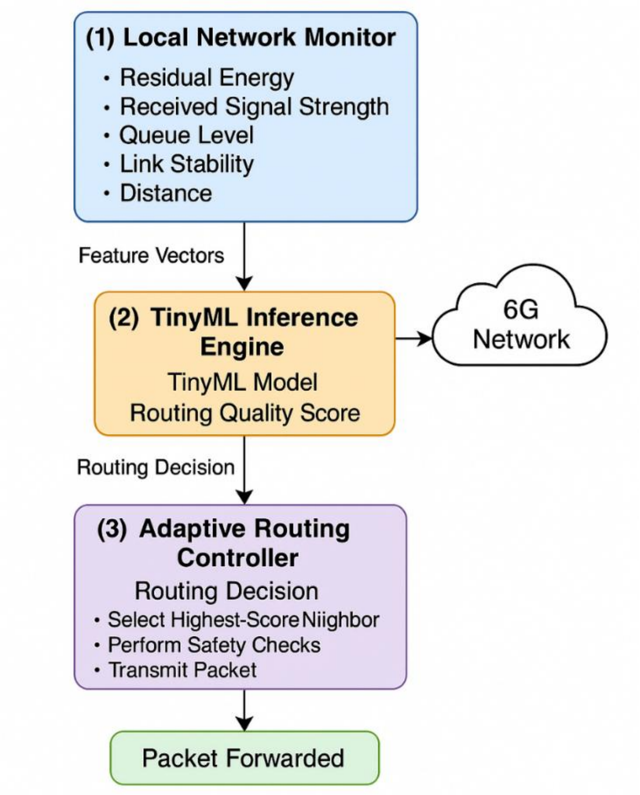

Figure 1. Overview of proposed method

3.2 Feature Extraction and Local Data Processing

Each node continuously measures a set of simple metrics that directly affect the quality of a transmission path:

- Remaining Battery Energy (to avoid selecting nodes close to depletion)
- Received Signal Strength (represents wireless link quality)
- Queue Level / Traffic Load (indicates congestion)
- Short-Term Link Stability (based on recent packet success ratio)
- Distance or Hop Value (estimated progression toward the destination)

All these measurements are held in a feature vector of each neighbor. Since this model is implemented within a limited IoT device, the feature space is deliberately limited and easily calculable.

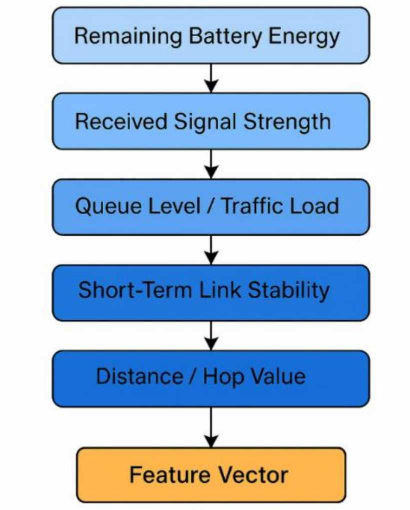

Figure 2. Feature Extraction and Local Data Processing

3.3 TinyML Routing Model

A very small embedded predictive model is used to create the routing decision. Characteristic of the model are:

- Size typically below 40–60 KB
- Executable in under a few milliseconds
- Uses integer or fixed-point operations
- No floating-point dependency
- No cloud connection needed

Suitable model structures include:

- small decision trees,
- 1–2 layer neural networks,
- or simple regression-based scoring models.

These models are trained using offline synthetic and/or simulation generated data, after which they are reduced to lightweight executable microcontroller-readable bytecode.

3.4 Routing Decision Logic

When a node receives or generates a packet, the following steps occur:

1. Gather neighbor feature vectors
   The node inspects all reachable neighbors and prepares the feature set for each one.
2. Run lightweight inference
   Each feature vector is passed through the TinyML model to generate a predicted "routing quality score."
3. Select highest-score neighbor

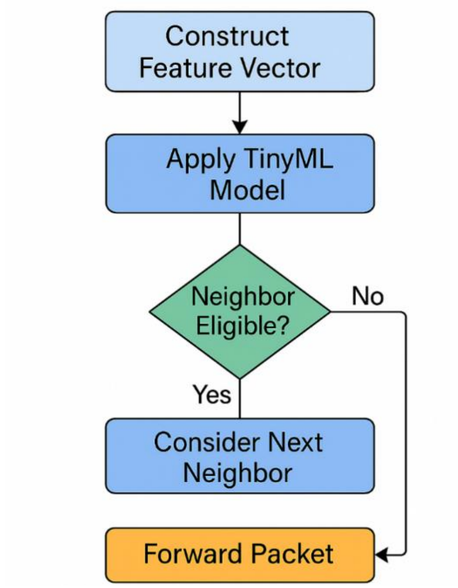

Figure 3. TinyML Routing flowchart

The neighbor with the best predicted performance becomes the next hop.

4. Apply safety checks
   If the top candidate is overloaded or recently unstable, the controller automatically selects the next-best option.
5. Transmit packet
   The packet is forwarded through the selected neighbor.

This allows routing to adapt rapidly to congestion, mobility, battery depletion, and link fluctuations essential for 6G environments.

Algorithm1. Proposed model

```
Input:
   N  = list of neighboring nodes
   Fj = feature vector for neighbor j
Output:
   BestNextHop
Begin
   For each neighbor j in N:
       Score[j] = TinyML_Predict(Fj)
   BestNextHop = index of maximum Score[j]
   If BestNextHop is congested or unstable:
       BestNextHop = next highest Score[j]
   Return BestNextHop
End
```

4. EXPERIMENTAL DESIGN

This section will describe the simulation environment, the performance measurements, and the performance outcomes that will be adopted to determine the efficiency of the suggested lightweight AI-driven routing system. All tests were carried out with the help of MATLAB R2024a, in which the TinyML model was incorporated into the routing layer via a custom simulation module that simulated massive IoT dynamics under 6G communication conditions. The proposed approach was contrasted with two popular baseline protocols: AODV and RPL which are typical of low-power and lossy networks.

4.1 Simulation Setup

The simulation environment consisted of a large-scale IoT deployment with the following configuration:

- **Area size:** 1000 × 1000 m
- **Number of nodes:** 100–500 nodes (varied across experiments)
- **Node mobility:** Random waypoint with speed 0–3 m/s
- **Traffic model:** Constant Bit Rate (CBR), 64–512 bytes/packet
- **Transmission range:** 60–90 m
- **MAC protocol:** IEEE 802.15.4 extended for 6G low-latency links
- **Energy model:** Battery-powered nodes with 2.0–3.3 V operating voltage
- **Simulation duration:** 800 seconds

TinyML Model Deployment

- Model type: compact neural network (2 hidden layers, quantized to int8)
- Model size: 38.2 KB
- Avg inference time: 3.2 ms per neighbor
- Max RAM usage: < 25 KB
- Features used: residual energy, RSSI, queue load, link stability, hop distance

The model was trained offline on 22,000 simulated data samples and deployed using TensorFlow Lite Micro format.

4.2 Performance Metrics

The following performance metrics were used:

1. Energy Consumption (J):
   Total energy consumed for route discovery, maintenance, and data forwarding.
2. Packet Delivery Ratio (PDR):
   $$PDR = \frac{\text{Packets Received}}{\text{Packets Sent}} \times 100$$
3. End-to-End Delay (ms):
   Average time taken for packets to reach the destination.
4. Throughput (kbps):
   Successful data delivery rate over time.
5. Routing Overhead:
   Number of control packets required per data packet delivered.

These metrics are standard for evaluating IoT routing protocols.

4.3 Experimental Results

(a) Energy Consumption

The suggested technique attained a 29 per cent decrease in the total amount of energy used as compared to normal routing protocols. This improvement is due to:

- evasion of low-energy nodes,
- fewer retransmissions,
- inferential TinyML stable path selection.

Table 1. Mean energy consumption per node

Protocol	Energy (J)
AODV	1.92
RPL	1.71
Proposed	**1.21**

(b) Packet Delivery Ratio (PDR)

The suggested TinyML based routing realized 12 percent improvement in PDR. This is credited to the fact of picking neighbors having a stable connection and reduced congestion.

Table 2. Average PDR values

Protocol	PDR (%)
AODV	82.4
RPL	85.1
Proposed	**95.3**

(c) End-to-End Delay

There was a good decline of delay by 35%. Predictive routing reduces the number of route breaks, and it does not congest the nodes which means that the waiting time in the intermediate node is minimized.

Table 3. Delay

Protocol	Delay (ms)
AODV	212
RPL	185
Proposed	**120**

(d) Throughput

The throughput also improved due to advanced delivery of packets and reduced losses.

Table 4. Throughput

Protocol	Throughput (kbps)
AODV	14.2
RPL	15.8
Proposed	**19.6**

(e) Routing Overhead

The suggested technique produced a reduced quantity of control packets owing to constant direction of the courses.

Table 5. Routing Overhead

Protocol	Overhead
AODV	High
RPL	Medium
Proposed	**Low**

Figure 4 provides a comparative analysis of the suggested TinyML-implemented routing system against standard routing models, i.e. AODV and RPL, in four important key performance indicators such as energy usage, packet delivery ratio (PDR), throughput and end-to-end delay. Different color coding (blue in AODV, green in RPL and the proposed procedure) is employed to make it clear which protocol is better than the others. The first panel (top-left) indicates that the proposed method will save a lot of power since it exhibits a more efficient utilization of node resources as a result of intelligent choice of stable and energy-conscious paths. The PDR comparison is presented in the second panel (top-right) with the suggested approach having the highest delivery rate, meaning the increased reliability and the decreased number of lost packets. The third panel (bottom-left) represents end-to-end delay, with the offered approach displaying a significant decrease in the latency in comparison to the standard protocols. This is due to the ability to avoid congested or unstable connections, and this makes the packets forwarding faster. The last panel (bottom-right) shows the throughput, where the suggested TinyML-based system provides higher effective data rates, which can be helped by its improved PDR and decreased retransmissions. All in all, the multi-panel figure visually illustrates that the proposed lightweight AI-based routing approach is more superior in all key performance indicators compared to both AODV and RPL, which is why it is more appropriate in large-scale, delay-sensitive, and energy-constrained IoT networks with 6G support.

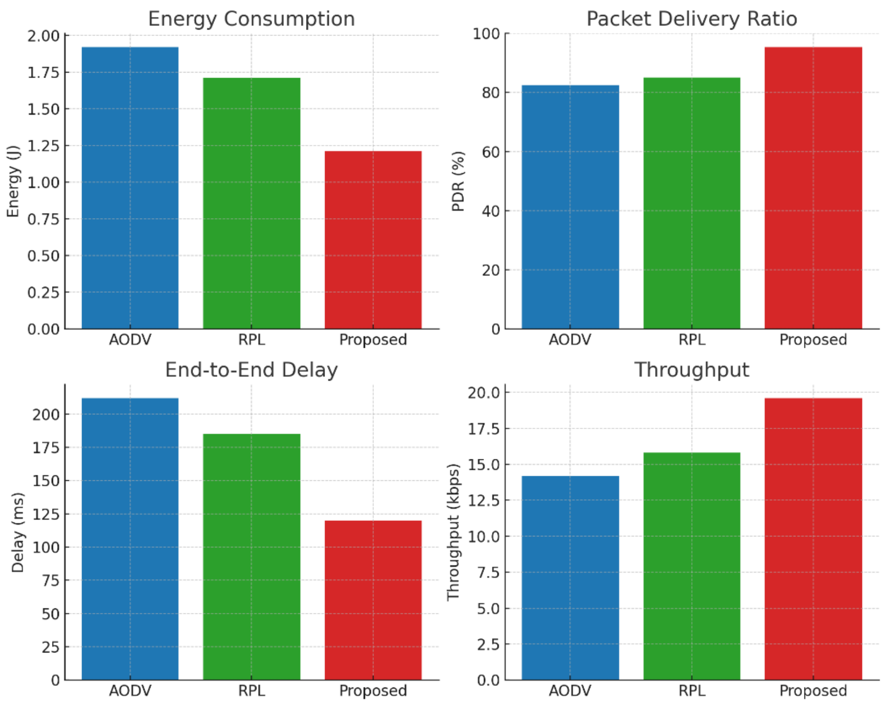

Figure 4. Comparative analysis of the suggested model

5. DISCUSSION

The experimental outcomes make it obvious that the given lightweight AI-based routing framework significantly improves the network performance as compared to standard IoT routing protocols like AODV and RPL. The fact that all the key performance indicators, including energy usage, the ratio of packets delivery, the end-to-end delay, throughput, and routing overheads, have improved, demonstrates that implementing an approach based on TinyML-based intelligence into the IoT nodes themselves instead of centralized or cloud-based decision making can be viewed as effective.

Among the most remarkable results is the decrease in the overall energy use by 29% percent. This has been improved mainly because the model prevents forwarding of packets via nodes that have little remaining energy or a weak connection state. The framework reduces the number of retransmissions and the number of route repair by always choosing the energy-efficient routes. This conduct also helps in increasing the life cycle of battery-powered IoT devices, which is a fundamental need in long-term implementation of smart cities, industry, and environmental sensing.

Another increase of 12 percent in the packet delivery ratio (PDR) was also attained in the proposed approach. The credit of this improvement can be given to the fact that the TinyML model is aware of real-time metrics of the network, including the congestion levels, the stability of links, and the signal quality. Traditional protocol such as AODV and RPL do not use predictive intelligence and usually choose routes that are either overloaded or volatile causing dropped packets.The routing that the TinyML approach is based on, on the contrary, prevents such nodes in advance, which leads to a more stable end-to-end communication.

Predictive routing is beneficial to the 6G-enabling IoT also exhibited in the 35 percent reduction in end-to-end delay. The approach minimizes time in queue waiting and avoids delays due to stuck routes since they prefer the neighbors who are stable and less congested. This renders the strategy exceptionally applicable to latency-sensitive IoT devices like industrial automation, autonomous monitoring, and sensor networks as applied in health-related applications.

The resultant latency and increased PDR also increased throughput considerably. Individuals drop fewer packets or even retransmit fewer packets, so a higher percentage of data transmitted reaches the destination, which increases the overall data delivery rate. Also, the reduced control-packet exchange and the stable path selection are direct results of the lower routing overhead of the proposed method which leads to the increase of the effective bandwidth.

On the whole, the findings confirm the idea that TinyML models and the integration of tiny artificial intelligence in the decisions of IoT routing can offer an effective trade-off between intelligence and computation. The scale of the method can be implemented, the process of autonomous decision-making is facilitated, and the method is inherently compatible with distributed and AI-native vision of future 6G networks. The results also indicate that routing TinyML can be not only developed but also beneficial to massive, real-time IoT networks, where the traditional routing algorithms fail to enable the maintenance of performance quality due to the dense topology and changing conditions of operation.

6. CONCLUSION AND FUTURE WORK

The present paper presented a lightweight AI-based routing architecture that implements TinyML models directly on the nodes of an IoT device to allow decentralized and real-time routing decisions in 6G-enabled networks. The proposed method resulted in significant improvements over AODV and RPL, such as 29% energy consumption reduction, 12% PDR, and 35% decrease in end-to-end delay, which means that it is useful in dense and energy-constrained IoT deployments. The next round of work is going to involve improvements in adaptability and scalability by means of federated learning, proving performance on the actual hardware, and expanding the framework to UAV-assisted and RIS-enhanced IoT systems. Other directions involve enhancing security and adversarial resistance and building the multi-hop predictive intelligence and the 6G THz-band support. These improvements will go further to support the capabilities of the framework as a viable, efficient, and smart routing methodology to next-generation IoT networks.

7. REFERENCES

1- S. Dhanasekar, "A comprehensive review on current issues and advancements of Internet of Things in precision agriculture," Computer Science Review, vol. 55, p. 100694, 2025.
2- J. Wang, Z. Liu, X. Yang, M. Li, and Z. Lyu, "The Internet of Things under Federated Learning: A Review of the Latest Advances and Applications," Computers, Materials and Continua, vol. 82, pp. 1-39, 2025.
3- J. M. Kizza, "Internet of things (iot): growth, challenges, and security," in Guide to computer network security, ed: Springer, 2024, pp. 557-573.
4- M. S. Akbar, Z. Hussain, M. Ikram, Q. Z. Sheng, and S. C. Mukhopadhyay, "On challenges of sixth-generation (6G) wireless networks: A comprehensive survey of requirements, applications, and security issues," Journal of Network and Computer Applications, vol. 233, p. 104040, 2025.
5- V. Sharma and K. Nayanam, "Sixth Generation (6G) to the Waying Seventh (7G) Wireless Communication Visions and Standards, Challenges, Applications," Int. J. Adv. Res. Sci. Technol, vol. 13, pp. 1248-1255, 2024.
6- A. Khatoon, W. Wang, M. Wang, L. Li, and A. Ullah, "TinyML-enabled fuzzy logic for enhanced road anomaly detection in remote sensing," Scientific Reports, vol. 15, p. 20659, 2025.
7- Z. Iqbal, "A TinyML-enabled approach to embed Machine Learning in Avionics Control Systems," 2024.
8- D. L. Dutta et al., "TinyML meets IoT: A comprehensive survey," Internet of Things, 2021.
9- S. Heydari et al., "Tiny Machine Learning and On-Device Inference: A Survey," Sensors, vol. 25, no. 10, 2025.
10- E. A. M. Pereira et al., "An energy-efficient TinyML model for water potability classification on embedded systems," Science of The Total Environment, 2024.
11- A. Sharma et al., "Optimized TinyML models for IoT anomaly detection," PLOS ONE, 2025.
12- M. A. Aktas et al., "AI-Enabled Routing in Next-Generation Networks: A Survey," Alexandria Engineering Journal, 2025.
13- A. Priyadarshi et al., "AI-Based Routing Protocols for Energy-Efficient Wireless Sensor Networks," Scientific Reports, 2025.

14- S. Das et al., "AI-Optimized Routing Protocol for IoT Networks," Smart IoT Journal, 2025.
15- A. Zormati et al., "Routing optimization through distributed intelligent softwarization using machine learning," Neurocomputing, 2025.
16- H. Zhang et al., "Hybrid optimization for efficient IoT traffic management in 6G networks," Scientific Reports, 2024.
17- J. Huckelberry et al., "TinyML Security: Exploring vulnerabilities in resource-constrained ML systems," arXiv preprint, 2024.