Optimised Q-learning for Dynamic Slot Assignment in Medium Access Control Protocol for Wireless Body Area Networks

Abdu Ibrahim Adamu, Wan Haszerila Wan Hassan, Darmawaty Mohd Ali, Wan Norsyafizan Wan Muhamad, Alwatben Batoul Rashed, and Mansir Abubakar

Original scientific article

***Abstract*—Wireless Body Area Networks (WBANs) enable continuous health monitoring through implanted and wearable sensors, but their performance hinges on an efficient Medium Access Control (MAC) scheme. Conventional protocols struggle to balance throughput, latency, and energy use, key requirements for medical data delivery. This paper introduces QLDSA-MAC, a Q-learning-driven dynamic slot-allocation MAC protocol that continuously adapts time slots to current traffic conditions. The agent maintains a Q-table of state–action values and selects slot assignments that maximize a composite reward reflecting throughput, delay, and energy consumption. Extensive simulations compare QLDSA-MAC with Time Division Multiple Access (TDMA-MAC), Concurrent MAC (C-MAC), and the IEEE 802.15.6 standard. Results show that QLDSA-MAC consistently delivers the highest throughput and the lowest packet delay across a range of traffic loads. It also reduces energy consumption, extending node lifetime in power-constrained scenarios. These gains demonstrate that reinforcement-learning (RL) methods can address WBAN challenges more effectively than fixed-rule MAC designs. Overall, QLDSA-MAC offers a practical path toward reliable, low-latency, and energy-efficient communication in healthcare WBAN deployments.**

***Index Terms*—WBANs, Medium Access Control, Dynamic Slot Allocation, Reinforcement-learning, QLDSA-MAC.**

I. Introduction

With the use of wearable and implanted sensors, Wireless Body Area Networks (WBANs) have become a game-changing technology in the healthcare industry [1], [2]. These networks improve the quality of patient care by making it easier to gather and transmit vital signs and physiological parameters, among other important health-related data. The efficient operation of WBANs is severely hindered by their unique characteristics, which include strict energy constraints, the need for reliable real-time data transmission, and dynamic traffic patterns [3]. Among these challenges, developing a successful Medium Access Control (MAC) protocol is crucial, as it has a direct impact on network performance, energy efficiency, and resource allocation [4]. WBANs have extensively utilized conventional MAC protocols, including Time Division Multiple Access (TDMA-MAC), Concurrent MAC (C-MAC), and the IEEE 802.15.6 standard. Although these protocols provide organized methods for controlling communication, they often fail to adapt to the rapidly changing circumstances that are common in WBAN situations, resulting in inefficiencies in energy consumption, throughput, and delay [5].

This study investigates the use of reinforcement learning (RL), particularly Q-learning, to develop a dynamic time slot allocation MAC protocol tailored for WBANs to overcome these constraints. Leveraging Q-learning's adaptive properties, the proposed protocol, called QLDSA-MAC (Q-learning-based Dynamic Slot Allocation MAC), optimizes time slot allocation based on current network conditions.

A preliminary version of this work was published as part of the 2025 IEEE 7th Symposium on Computers & Informatics (ISCI) and entitled "Slot Allocation in Wireless Body Area Networks (WBANs) using Q-learning Approaches." At the conference, we presented the early concept of Q-learning (QL-MAC) for adaptive slot assignments in WBANs [6]. The current paper is a significant extension of that work as it presents an improved protocol, QLDSA-MAC, that incorporates a composite reward function to jointly optimize throughput, delay, and energy. It also utilizes a realistic simulation framework on Castalia, OMNeT++, and Python (Gymnasium), and extends evaluation metrics with more theoretical support and in-depth comparison.

Four components characterize its general operation in our approach. The first is the *state*, which represents the current network situation. This state relies mainly on the traffic load of the nodes and assists in capturing the ever-changing WBAN environment so that the algorithm can react and adapt in real time. The second component, *actions*, determines which time slots to allocate or reallocate to other nodes by creating new working slots or modifying existing ones based on network traffic and priorities. The *reward function* acts as a referencing variable that encourages efficient and reliable data transmis-

Manuscript received October 14, 2025; revised November 10, 2025. Date of publication January 7, 2026. Date of current version January 7, 2026.

A. I. Adamu, D. M. Ali, and W. N. W. Muhamad are with the Wireless Communication Technology Group (WiCOT), Faculty of Electrical Engineering, Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor, Malaysia (e-mails: 2023402878@student.uitm.edu.my, darma504@uitm.edu.my, syafizan@uitm.edu.my).

W. H. W. Hassan is with the Fakulti Teknologi dan Kejuruteraan Elektronik dan Komputer, Universiti Teknikal Malaysia Melaka, Malaysia (e-mail: haszerila@utem.edu.my).

A. B. Rashed is with the Department of Information Technology, College of Computer, Qassim University, Buraydah, Saudi Arabia (e-mail: bwtban@qu.edu.sa).

M. Abubakar is with the Faculty of Computer Science and Mathematics, Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor, Malaysia (e-mail: mansir@uitm.edu.my).

Digital Object Identifier (DOI): 10.24138/jcomss-2025-0154

sion. It provides positive reinforcement when performance improves, such as higher throughput, lower latency, or reduced energy consumption, and imposes penalties when issues like packet loss or excessive delay occur. Lastly, the *agent*, which functions as the WBAN coordinator, gradually learns the most effective slot allocation strategy through continuous interaction with its environment. In certain cases, an individual sensor node may also act as a learning agent, independently optimizing its transmission timing to save energy and improve overall system efficiency [2].

The QLDSA-MAC protocol aims to enhance throughput, minimize delays, and conserve energy by dynamically adjusting resource allocation. These performance measures are vital for effectively implementing WBANs in healthcare applications. By comparing the performance of QLDSA-MAC to that of other MAC protocols, this study illustrates how effectively it addresses the limitations of traditional methods. The results highlight not only QLDSA-MAC's superiority but also the broader implications of integrating machine learning (ML) techniques into networking solutions for resource-constrained environments. Our contributions are as follows:

- Instead of using static or rule-based scheduling, we provide QLDSA-MAC, a novel MAC system that uses Q-learning in the MAC layer protocol to dynamically and intelligently allocate time slots in WBANs to improve efficiency.
- Our simulations demonstrate that QLDSA-MAC consistently outperforms IEEE 802.15.6, TDMA-MAC, and C-MAC, providing the highest throughput, the lowest average transmission latency, and the greatest energy savings. This confirms its suitability for battery-constrained and latency-sensitive healthcare applications.
- We illustrate how the discount factor controls an agent's time horizon: the closer γ is to one, the further into the future the agent looks when making decisions. In addition, we highlight the classic trade-off by comparing the α value between 0–1: a large α accelerates convergence but can be volatile in noisy settings, whereas a small α yields stability at the cost of slower adjustment.

The remainder of this paper is structured as follows. Section II provides an overview of existing MAC protocols for WBANs and discusses their key limitations, highlighting the need for an adaptive, learning-based approach. Section III describes the proposed Q-learning–based slot allocation method (QLDSA-MAC), along with details of the learning framework, simulation environment, and setup. Section IV presents and analyzes the results in terms of throughput, delay, and energy consumption, and further examines the impact of parameters such as the learning rate and discount factor. Finally, Section V concludes the paper by summarizing the main findings and offering directions for future research on reinforcement learning–driven MAC protocols for WBANs.

II. Related Work

Wireless Body Area Networks (WBANs) have been a game-changing technology in the medical field since they allow for remote patient care and real-time vital sign monitoring. The MAC protocol, which regulates how devices access the shared wireless medium, is a crucial part of WBANs. Conventional MAC protocols, including IEEE 802.15.6 MAC, TDMA-MAC, and C-MAC, have been extensively researched and used in WBANs [7]–[9]. Each of these protocols, however, has built-in drawbacks that impair their functionality in contexts with changing conditions and limited resources.

The available bandwidth is divided into time slots assigned to nodes using the time-based scheduling technique known as TDMA-MAC [10]. Although this organized approach prevents collisions and guarantees predictable performance, it is not flexible enough to adjust to changing traffic patterns. In situations where traffic demand varies, TDMA-MAC might not satisfy real-time needs or underutilize resources, which results in inefficiencies [11]. Likewise, devices can vie for access to the medium using contention-based protocols such as C-MAC. Despite its flexibility, this method frequently leads to retransmissions, packet collisions, and higher energy usage, especially in crowded networks [7]. Because of these limitations, C-MAC is not appropriate for applications that require low latency and high reliability.

A significant initiative to tackle the challenges of WBANs is the IEEE 802.15.6 standard, designed explicitly for these networks [8]. This standard combines both contention-based and scheduled access mechanisms to strike a balance between flexibility and performance. Nevertheless, its ability to adapt to varying traffic conditions is limited by its reliance on predetermined schedules and fixed parameters [12]. In situations with erratic data generation patterns, this rigidity can lead to inefficient resource utilization, increased delays, and elevated energy consumption [8].

While conventional MAC protocols in WBANs rely heavily on static configurations and predefined scheduling policies, such approaches often fail to adapt to dynamic network conditions and diverse traffic demands [13], [14]. To address these challenges, recent research has turned toward machine learning (ML)–based solutions, particularly reinforcement learning (RL), for adaptive decision-making. RL allows MAC protocols to continuously learn and select optimal actions in response to varying network states. For example, Kwon et al. [15] demonstrated that an RL-based contention window adjustment scheme can significantly reduce collisions and enhance throughput. Similarly, Rana et al. [16] developed a reinforcement learning–enabled multi-class MAC protocol that efficiently handles variable loads and prioritizes heterogeneous traffic under the IEEE 802.15.6 standard. In addition, a model-free RL method called Q-learning has been used to solve several networking issues, such as congestion control, spectrum allocation, and routing [17].

MAC protocols can prioritize important data streams, dynamically distribute resources, and reduce energy waste by utilizing Q-learning. Notwithstanding these developments, there is still a gap in understanding the requirements of WBAN MAC networks, as the use of RL in these protocols remains largely unexplored [18]–[20].

By introducing QLDSA-MAC, a dynamic time slot allocation algorithm grounded in Q-learning and specifically tailored for WBANs, this study aims to bridge this gap. Unlike traditional methods, QLDSA-MAC adapts to changing network conditions, ensuring efficient resource utilization while minimizing latency and energy consumption. This work pushes the boundaries of current technology in WBANs and establishes a foundation for future research in intelligent networking solutions through the integration of reinforcement learning in MAC-layer design. Table I provides a summary of various MAC protocols and their advancements in WBANs.

III. Materials and Method

A. Proposed QLDSA-MAC for Slot Allocation

The focus of this methodology is the creation and assessment of QLDSA-MAC, a dynamic time slot allocation protocol based on Q-learning that optimizes MAC in WBANs. This protocol uses Q-learning, a type of RL, to dynamically assign time slots according to current network conditions. It addresses critical WBAN challenges, including reducing latency, conserving energy, and increasing throughput, which are essential for supporting real-time health monitoring applications [26].

Q-learning, a model-free RL technique that enables the system to learn optimal strategies through trial-and-error interactions with the environment, is integrated into QLDSA-MAC. In this context, the environment is akin to the WBAN, where nodes (such as wearable sensors) compete for limited communication resources. State, action, and reward are the three fundamental building blocks upon which the Q-learning system is founded.

The *state* represents the network's current condition, including available bandwidth, traffic load, and the number of active nodes. The *action* corresponds to allocating time slots to nodes, a decision-making process that ensures efficient resource utilization. Finally, the *reward function* is designed to promote desirable outcomes, such as increased throughput, reduced energy consumption, and minimized packet collisions.

Over time, QLDSA-MAC enhances its decision-making capabilities by iteratively updating its Q-values based on observed rewards, resulting in adaptive and optimal time slot allocation. Fig. 1 shows QLDSA-MAC working in a WBAN environment.

B. Frameworks for QLDSA-MAC

A clear problem statement is outlined through the Markov Decision Process (MDP). Within the MDP framework, the objective is attained through interactive learning, where S is the set of states, A is the set of actions, and R is the set of rewards that are finite in number. Both reward (R_t) and state (S_t) are discrete random variables, each following a discrete probability distribution. Notably, these variables are dependent solely on the preceding state and action. All previous states and actions are probably based on the specified values for all $s', s \in S$, $r \in R$, and $a \in A(s)$ [14], [17], [27], [28].

$$p(s', r|s, a) = \Pr(S_t = s', R_t = r \mid S_{t-1} = s, A_{t-1} = a) \tag{1}$$

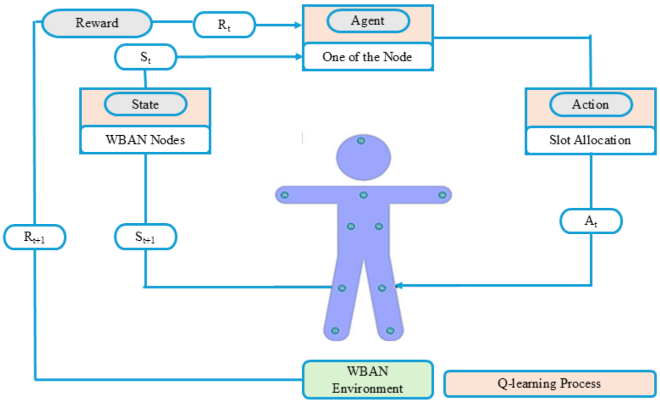


Fig. 1. A diagram of the proposed QLDSA-MAC.

$$\sum_{s\in S}\sum_{r\in R} p(s', r \mid s, a) = 1, \quad s \in S,\ a \in A(s) \tag{2}$$

One action is chosen by one environmental state and designated as a policy. If the agent follows the policy π at time t, the probability of $A_t = a$ under $S_t = s$ is denoted as $\pi(a \mid s)$. QL can solve MDPs without complete information by employing an RL technique. Hence, a QL system consists of the agent, the environment, a policy, a reward, and a Q-value function. Policy is an action that has been carried out in certain environmental conditions. Generally, a policy is a function or lookup table; it is also the foundation of an RL agent. Policy-determined behavior is adequate. In addition, the stochastic or defined probability may influence the policy for each action.

The reward signal, a single value supplied by the environment to the RL agent, determines the goal of an RL task. An agent's only goal is to maximize its reward. If a policy selects one action but the reward in the subsequent situation is low, the policy will be updated to select alternative actions in that condition. Unlike the reward function, which provides a signal for a specified time, the Q-value function produces a good signal for the state's end time. As a result, the Q-value indicates a state's cumulative reward, which an agent will use to decide which action to perform in the future [28].

In the WBAN environment, the reward in QLDSA-MAC is designed to guide the learning process toward maximizing throughput, minimizing latency, and conserving energy. Instead of a binary success/failure reward as used in the preliminary conference version [6], this work employs a multi-factor reward structure that reflects overall network performance. The reward at the time step t is formulated as:

$$R_t = w_1 \left(\frac{T_{\text{succ}}}{T_{\text{total}}}\right) - w_2 D_{\text{norm}} - w_3 E_{\text{norm}} \tag{3}$$

where T_{succ} and T_{total} denote the number of successful and total transmissions, respectively, D_{norm} represents the normalized average delay, and E_{norm} is the normalized energy consumption per node. The coefficients w_1, w_2, and w_3 are weighting factors such that $w_1 + w_2 + w_3 = 1$, controlling the trade-off among throughput, delay, and energy efficiency. A

TABLE I
SUMMARY OF MAC PROTOCOLS AND THEIR TRENDS

Ref	Protocols	IEEE Tech.	Sch.	Cont.	Poll.	Method	Traffic Type	Perf.
[7]	C-MAC	IEEE 802.15.6	-	✓	-	Simulator, numerical analysis	Concurrent Traffic	✓
[8]	IEEE 802.15.6 MAC	IEEE 802.15.6	✓	✓	✓	Network Simulator	UP (0) – UP (7)	✓
[9]	TDMA MAC	IEEE 802.15.6	✓	-	-	Network Simulator	-	✓
[21]	PA-MAC	IEEE 802.15.4	✓	✓	-	Network Sim.	Emergency, on-demand, normal, non-medical	✓
[22]	DSBS, DSBB	IEEE 802.15.6	✓	-	-	Network Simulator	Normal and emergency	✓
[23]	TA-MAC	IEEE 802.15.4	✓	✓	-	Network Simulator	Emergency, on-demand, normal, non-medical	✓
[24]	DMTM-MAC	IEEE 802.15.6	✓	✓	✓	Network Simulator	Periodic, urgent, on-demand	✓
[25]	McMAC	IEEE 802.15.4	✓	✓	✓	Network Sim.	Type 0 – Type 4	✓
[26]	ADT-MAC	IEEE 802.15.6	✓	✓	✓	Network Simulator	Emergency, periodic (High, medium, low)	✓
-	Proposed QLDSA-MAC	IEEE 802.15.6	✓	✓	✓	Network Simulator + ML	Normal and emergency	✓

higher w_1 emphasizes data reliability, while larger w_2 or w_3 penalize latency and energy wastage. This formulation allows QLDSA-MAC to dynamically adapt slot allocation decisions in response to changing network conditions and traffic loads.

The Q-value function is a key component in the agent's decision-making process. Its findings guide the agent in selecting the most appropriate action by maximizing the Q-value function. This process allows the agent to achieve the largest reward over several actions without necessarily focusing on a single reward [13]. The agent in the QL algorithm maintains a $Q(S_t, A_t)$ table. For $t = 1, 2, 3, \dots, N$, the agent observes the state S_t of the MDP in the WBANs network and selects an action A_t from the actions (A). After action A_t, the agent receives a reward $R(t)$ and then observes the next state S_{t+1}.

The agent will create an event sequence. The Q-table will be updated by the sequence of events under the $Q(S_t, A_t)$ pairs according to the QL function:

$$\begin{aligned} Q(S_t, A_t) = Q(S_t, A_t) + \alpha\big[&R_{t+1} \\ &+ \gamma \max Q(S_{t+1}, a) - Q(S_t, A_t)\big] \end{aligned} \tag{4}$$

The agent chooses an action based on the state S_t, and the maximum Q-value for the following state, S_{t+1}, may be calculated using the action A_t, which updates the current Q-value. Fig. 2 depicts the detail.

The optimal range for the discount factor γ is 0 ¡ γ 1. The learning rate, denoted by α, ranges from 0 to 1 [29]. The agent evaluates the current reward only if the discount factor γ is 0. If the discount factor γ is 1, the agent seeks a long-term reward. The learning rate determines how much the new message supersedes the old one. As the anticipated value changes, the learning rate influences the estimated speed. Fig. 6 shows that the discount factor γ scales the weight of future rewards, while Fig. 7 indicates that the learning rate α controls convergence speed. These details are depicted in the results and discussion section.

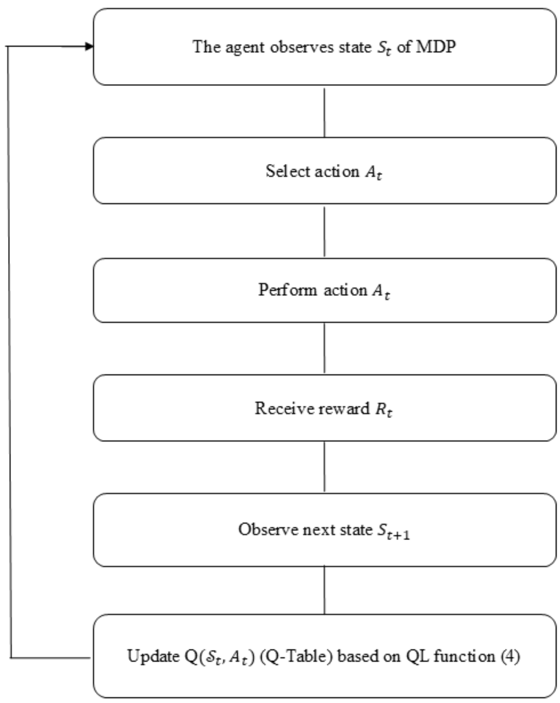


Fig. 2. A proposed QLDSA-MAC in a Markovian environment.

C. Simulation Environment for the QLDSA Mode

We run simulations of the QL algorithm in a WBAN setting, investigating different node layouts and traffic patterns. The network is initially configured with a default slot allocation, which is then adjusted by the QL agent based on feedback from a reward function. For performance testing, baseline protocols such as classic TDMA MAC, C-MAC, and IEEE 802.15.6-MAC are included, allowing for a thorough evaluation of the QL approach's effectiveness.

To assess critical performance parameters, such as packet delay, network throughput, and energy consumption, we utilized Castalia, OmNet++, and various Python libraries, including NumPy, pandas, seaborn, matplotlib, and OpenAI Gymnasium. Castalia was chosen because it is a discrete event simulator for WBANs. The simulation results were compared to benchmark protocols under identical network conditions to guarantee a rigorous and unbiased evaluation. For IEEE 802.15.6-MAC, the superframe structure used a fixed slot size, including a beacon, five exclusive access phases (EAPs) slots for emergency traffic, five random access phases (RAPs) for connection creation, and 15 TDMA-MAP slots for periodic traffic transmission.

D. Simulation Setup for the QLDSA Mode

The simulation setting was thoughtfully created to assess QLDSA-MAC's performance in authentic WBAN scenarios. The network model replicated a standard WBAN deployment by distributing several sensor nodes over a short area. These nodes mimicked the actions of medical monitoring equipment by periodically generating data packets. The simulation's duration (200 seconds), packet sizes, transmission rates, and channel characteristics were among its crucial parameters. Three popular MAC protocols, TDMA-MAC, C-MAC, and IEEE 802.15.6, were utilized to test QLDSA-MAC's performance to provide a thorough comparison. Three crucial metrics, average delay, energy usage, and throughput, were used to assess each approach. The protocol's capacity to facilitate real-time communication is demonstrated by the average latency, which calculates the time it takes for packets to move across the network. Because WBAN devices are battery-powered, energy consumption measures the entire amount of energy used by the network, which is an important factor to consider. Throughput measures how well a protocol uses available bandwidth by calculating the quantity of data that is successfully transmitted per unit of time.

The paper also offers a concise synopsis of the comparison protocols in addition to QLDSA-MAC. By using a structured time division technique, TDMA-MAC divides the available bandwidth into predetermined time slots that are allocated to nodes. Although this approach guarantees consistent performance, it is not flexible enough to adjust to shifting network conditions. Nodes vie for channel access in the contention-based mechanism of C-MAC. Although this method is easy to use, it frequently results in inefficiencies like retransmissions and packet collisions, especially in congested networks. Lastly, the standard MAC protocol for WBANs, IEEE 802.15.6, depends on preset scheduling methods. Despite being widely used, it might not handle the dynamic and diverse character of contemporary WBAN traffic patterns.

The goal of this work is to illustrate the benefits of integrating machine learning techniques into MAC layer design by contrasting QLDSA-MAC with conventional protocols. Even in challenging situations, QLDSA-MAC's adaptive nature allows it to respond dynamically to changes in network traffic, ensuring effective resource allocation. This feature sets it apart from semi-static or static protocols, such as IEEE 802.15.6 and TDMA-MAC, which may struggle to maintain good performance in dynamic environments. Moreover, QLDSA-MAC is particularly well-suited for WBANs, where reliable data transfer and energy efficiency are crucial, thanks to its ability to achieve high throughput while minimizing energy consumption.

In conclusion, the suggested approach demonstrates the creative application of Q-learning to improve WBAN MAC layer operations. The study provides valuable insights into how RL can transform resource management in wireless networks by combining a robust simulation framework with a comprehensive performance evaluation. In addition to confirming the efficacy of QLDSA-MAC, the study's findings open the door to further developments in intelligent and flexible networking systems. Table II shows the other parameters of our proposed QLDSA-based MAC protocol for WBAN networks.

TABLE II
PROPOSED QL-BASED MAC PARAMETERS [26], [30]

Slot length	10 ms
Simulation Time	200 s
Number of Nodes	10
Frequency Band	2.5 GHz
Payload	105 Bytes
Transmissions Rate	1,024 Kbps
Superframe Length	32
Transmission Power	0.01 J
Exploration Rate	0.1
Learning Rate (α)	0.1
Discount Factor (γ)	0.9
Number of Episodes	400–500

IV. RESULTS AND DISCUSSION

The QLDSA-MAC protocol's performance is contrasted with that of C-MAC [7], IEEE 802.15.6-MAC [8], and TDMA-MAC [9]. These protocols were selected for comparison due to their conceptual similarities with the main features of the suggested method. Like C-MAC, which prioritizes concurrent traffic mitigation to improve network performance and is a crucial part of QLDSA-MAC, TDMA MAC uses a dynamic superframe structure based on traffic demand and prioritizes emergency traffic above regular traffic. The superframe structure of IEEE 802.15.6-MAC serves as a benchmark for the QLDSA-MAC protocol. Moreover, the performance evaluation of the suggested protocol must consider the outstanding problems with fixed time slot allocation in the IEEE 802.15.6 standard for WBAN communication. Hence, the performance of the suggested protocol is illustrated in Figs. 3, Fig. 4, and Fig. 5.

A. Network Throughput

The throughput (TH) for a WBAN network is defined as follows [2], [13], [28]:

$$\text{TH} = \frac{R_{\text{channel}}\, T_{\text{data}}}{T_{\text{simulation}}} \tag{1}$$

where R_{channel} is the channel data rate, T_{data} is the total transmission time of all successfully transmitted packets, and $T_{\text{simulation}}$ is the system simulation time.

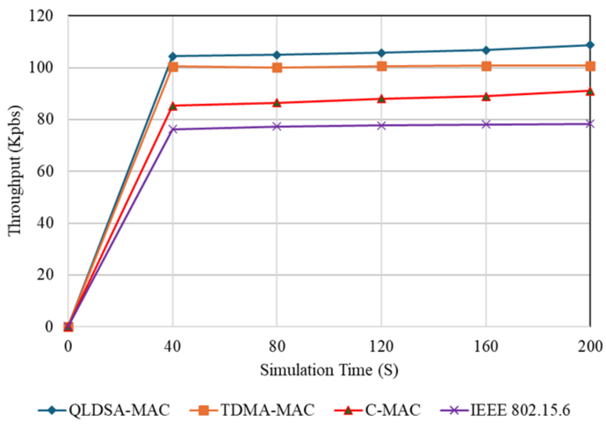


Fig. 3. Network throughput vs simulation time.

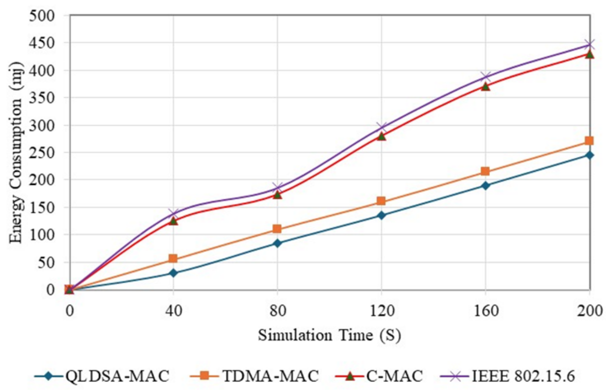


Fig. 4. Energy consumption vs simulation time.

A thorough comparison of the throughput performance of different Medium MAC protocols in Wireless Body Area Networks (WBANs) across Simulation Time is shown in Figs. 3. IEEE 802.15.6, TDMA-MAC, C-MAC, and QLDSA-MAC are among the protocols that have been assessed. A crucial parameter for evaluating the effectiveness of data transmission in WBANs, where prompt delivery of health-related data is crucial, is throughput, which is expressed in kilobits per second (kbps). The results show that these protocols function significantly differently, with QLDSA-MAC turning out to be the best option.

Fig. 3 shows the data throughput performance of each MAC protocol over 200 s. By 40 seconds, all four protocols had reached their steady-state speeds. QLDSA-MAC takes the lead, increasing from 105 kbps to 110 kbps by the end of the cycle. TDMA-MAC follows closely, averaging just around 100 kbps with little volatility. C-MAC steadily increases from 85 kbps to 90 kbps, while IEEE 802.15.6 plateaus at about 78 kbps.

In terms of performance, QLDSA-MAC beats TDMA-MAC by 9%, C-MAC by 22%, and IEEE 802.15.6 by 41%. This gain is credited to QLDSA-MAC's usage of Q-learning, which optimizes slot assignments while maintaining steady throughput. For wireless body-area networks that require quick startup and consistent performance, QLDSA-MAC is the most dependable alternative.

These results demonstrate how crucial adaptive mechanisms are to WBAN MAC schemes. By dynamically optimizing resource allocation and ensuring effective data transmission, QLDSA-MAC's incorporation of Q-learning enables it to outperform conventional protocols. In healthcare applications, where WBANs must facilitate real-time monitoring of vital signs and other crucial health-related data, this flexibility is very advantageous. QLDSA-MAC solves major issues with traditional protocols by optimizing throughput while reducing latency and energy usage.

B. Energy Efficiency

The total amount of energy used by a sensor node during its communication time across its operating states is known as its total energy consumption, denoted by E_{total}. Here, we consider only energy consumption during successful transmission and collisions [31]–[33].

$$E_{\text{total}} = E_{\text{success}} + E_{\text{collision}} \tag{2}$$

where E_{success} denotes the energy consumption of successful transmission in a WBAN network, and $E_{\text{collision}}$ denotes the energy consumption during a collision in a TDMA MAC of the WBAN network.

$$\text{Energy efficiency} = \frac{\sum D_{\text{successful}}}{T_{\text{total}}} \tag{3}$$

Here, $D_{\text{successful}}$ represents the total successfully received data in bits, and T_{total} represents the total energy consumed in Joules.

The suggested QLDSA-MAC protocol, as shown in Fig. 4, offers convincing proof of its higher efficiency when compared to more conventional MAC protocols, such as IEEE 802.15.6, TDMA-MAC, and C-MAC. Fig. 4 charts how much energy each MAC protocol burns over the 200-second run. Consumption climbs steadily for all four schemes, but the gaps open quickly. QLDSA-MAC is the most frugal throughout, creeping from zero to 270 mJ by the end. TDMA-MAC tracks slightly higher, finishing close to 290 mJ. In contrast, C-MAC and the IEEE 802.15.6 baseline draw far more power: C-MAC hits about 420 mJ at 200 s, while IEEE 802.15.6 tops the list at 450 mJ.

Translated into savings, QLDSA-MAC uses around 7% less energy than TDMA-MAC, 36% less than C-MAC, and 40% less than IEEE 802.15.6 over the full simulation. The lower draw suggests that QLDSA's Q-learning keeps retransmissions and idle listening to a minimum, making it the best choice when battery life is at a premium, especially in wearable WBAN nodes where every millijoule counts.

These results demonstrate the shortcomings of static and semi-static MAC protocols in WBANs, where sensor nodes run on batteries, and energy economy is crucial. Because QLDSA-MAC is adaptive, it can successfully handle these issues by dynamically modifying time slot allocations in response to the state of the network. The protocol's use of Q-learning reduces delays and energy waste, guaranteeing

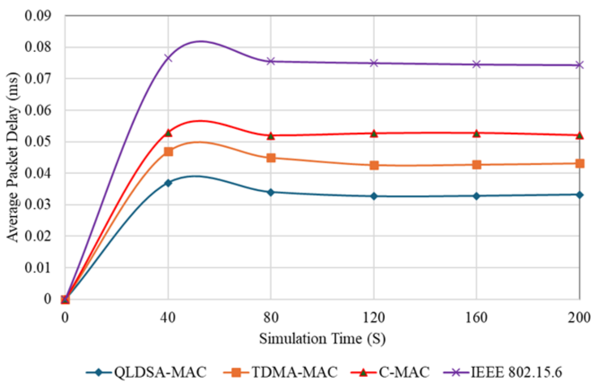


Fig. 5. Packet delay vs simulation time.

the prompt and dependable supply of vital health-related data while prolonging the life of WBAN devices. Energy efficiency has a direct impact on the durability and usability of wearable and implanted sensors, making this skill very useful in healthcare applications.

C. Packet Delay

The average amount of time that passes between a packet being generated at the sensor node and being received at the hub is known as the packet delay of that node. A lower packet delay shows better protocol performance. It is computed as follows [10], [34]:

$$\text{Delay} = \sum_{i=i}^{C_{\text{Success}}} \sum_{j=1,\ \text{connection } i}^{n_{\text{trial}}} t^{j}_{\text{delay}} \tag{4}$$

where C_{Success} denotes the total number of successful connections, n_{trial} is the number of attempts required to achieve successful transmission, and t^{j}_{delay} denotes the contention delay for the j^{th} trial for a connection involved in a TDMA MAC of a WBAN network.

The performance evaluation of the suggested QLDSA-MAC protocol, as shown in Fig. 5, provides convincing evidence of its superiority over conventional MAC protocols, such as IEEE 802.15.6, TDMA-MAC, and C-MAC. Fig. 5 traces the average packet delay for the same four protocols over the 200-second run. All curves climb quickly during start-up, peaking near the 40-second mark, and then settle into a gentle plateau.

By the 200-second point, QLDSA-MAC's delay is about 23% lower than TDMA-MAC, 38% lower than C-MAC, and 56% lower than IEEE 802.15.6. The consistently shorter queues suggest that QLDSA's Q-learning quickly discovers efficient slot assignments, trimming contention and retransmissions. For latency-sensitive WBAN traffic, this makes QLDSA-MAC the clear winner.

These results demonstrate the shortcomings of static and semi-static MAC protocols in WBANs, where effective resource management and real-time data transfer are critical. Because QLDSA-MAC is adaptive, it can successfully handle these issues by dynamically modifying time slot allocations in response to the state of the network. The protocol ensures dependable and timely delivery of vital health-related data by utilizing Q-learning to reduce packet collisions and delays. In healthcare applications, where delays can have serious repercussions for patient monitoring and treatment, this capacity is beneficial.

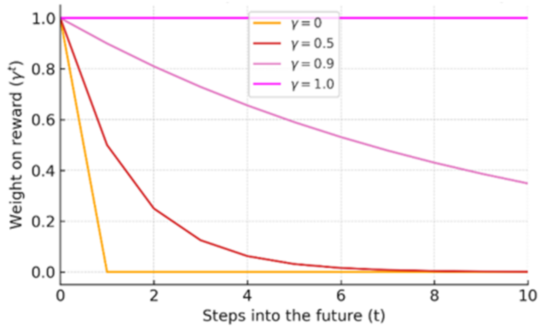


Fig. 6. Discount factor γ scales future reward weight.

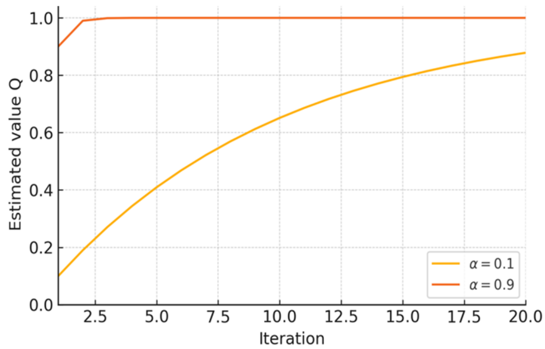


Fig. 7. Learning rate α controls convergence speed.

D. Choosing Value for Learning Rate and Discount Factor

Fig. 6 plots the weight assigned to future rewards (vertical axis) against the number of steps into the future (horizontal axis) for four discount factors: γ = 0, 0.5, 0.9, and 1.0. When γ = 0, the curve drops immediately to zero, showing that only the immediate reward is valued. With γ = 0.5, the weight falls off quickly but not instantaneously, reflecting modest concern for near-term rewards. For γ = 0.9, the decline is gradual, indicating that the agent still values reward many steps ahead, though each is slightly less important than the last. At γ = 1.0, the line remains flat at one, meaning every future reward is treated the same as the present reward. Together, these curves illustrate how the discount factor controls an agent's time horizon: the closer γ is to one, the further into the future the agent looks when making decisions.

Fig. 7 shows how the learning rate α influences the speed at which an estimated value (vertical axis) converges to the true value of 1.0 (dashed reference line) over successive updates (horizontal axis). With a high learning rate (α = 0.9), the estimate leaps toward the target within just a few iterations, demonstrating rapid adaptation. By contrast, a low learning rate (α = 0.1) causes the estimate to rise slowly and smoothly, illustrating cautious, incremental learning. The comparison highlights the classic trade-off: a large α accelerates convergence but can be volatile in noisy settings, whereas a small α yields stability at the cost of slower adjustment.

V. Conclusion

This study shows that the suggested QLDSA-MAC protocol outperforms traditional MAC techniques in WBANs. QLDSA-MAC surpasses TDMA-MAC, C-MAC, and IEEE 802.15.6 in throughput, achieving the highest data rates and continuing to improve steadily as simulation time progresses. QLDSA-MAC also exhibits the lowest average transmission delays, suggesting more efficient channel access and prompt data delivery, which is crucial for real-time healthcare applications. The advantage of QLDSA-MAC is further illustrated by energy consumption analysis, which indicates that the protocol utilizes the least amount of energy over the simulation period, thereby extending network lifetime and conserving node battery resources. These results showcase the effectiveness of integrating Q-learning into MAC protocols to facilitate dynamic slot allocation that intelligently adapts to changing network conditions. Consequently, QLDSA-MAC presents a viable approach to enhance the sustainability, efficiency, and reliability of future WBAN deployments. Although there have been significant improvements in throughput, latency, and energy efficiency with the QLDSA-MAC protocol, several avenues for future research remain to further enhance its capabilities. For instance, applying the Q-learning paradigm to a Deep Q-Network (DQN) in larger and more dynamic WBAN contexts may improve scalability and learning efficiency.

Acknowledgement

This research was supported by the Ministry of Higher Education Malaysia (MOHE) and the Faculty of Electrical Engineering, Universiti Teknologi MARA (UiTM), Shah Alam, Selangor, Malaysia, under the Fundamental Research Grant Scheme (FRGS), Grant No. FRGS/1/2023/TK07/UITM/02/29. The authors acknowledge the institutional support provided through the Research Fellow Programme from Universiti Teknikal Malaysia Melaka (UTeM).

References

[1] B. Alte and A. Vahdat, "Secure emergency mac protocol for wireless body area networks," *International Journal of Communication Networks and Distributed Systems*, vol. 31, no. 3, pp. 251–276, 2025.

[2] A. I. Adamu, P. K. Donta, D. M. Ali, S. Sarang, G. M. Stojanović, and S. S. Sarnin, "A systematic literature review of advanced machine learning techniques in wireless body area networks: Application, challenges, and future directions," *IEEE Access*, pp. 1–1, 2025.

[3] S. Karthika and K. J. Gnanaselvi, "A review of forensics security hazards and challenges in wban and healthcare systems," in *Security, Privacy, and Trust in WBANs and E-Healthcare*, 2024, pp. 63–82.

[4] T. Vikash, G. Suman, M. Tolani, A. T. Tumunuri, and P. Kumar, "Bitmapping-based security-aware energy-efficient mac protocol for lorawan," *IEEE Access*, 2025.

[5] W. H. Wan Hassan, D. Mohd Ali, J. Mohd Sultan, and M. Kassim, "Medium access control protocol based on time division multiple access scheme for wireless body area network," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 14, no. 3, pp. 2762–2770, 2024.

[6] A. I. Adamu, D. M. Ali, S. I. B. Kamarudin, S. S. Sarnin, and W. H. Wan Hassan, "Slot allocation in wireless body area networks (wbans) using q-learning approaches," in *2025 IEEE 7th Symposium on Computers & Informatics (ISCI)*, 2025, pp. 41–46.

[7] R. Zhang, H. Moungla, J. Yu, and A. Mehaoua, "Medium access for concurrent traffic in wireless body area networks: Protocol design and analysis," *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2586–2599, 2017.

[8] D. M. G. Preethichandra, L. Piyathilaka, U. Izhar, R. Samarasinghe, and L. C. De Silva, "Wireless body area networks and their applications - a review," *IEEE Access*, vol. 11, pp. 9202–9220, 2023.

[9] H. Fourati, H. Idoudi, and L. A. Saidane, "Intelligent slots allocation for dynamic differentiation in ieee 802.15.6 csma/ca," *Ad Hoc Networks*, vol. 72, pp. 27–43, 2018.

[10] K. J. Mystica and J. M. L. Manickam, "Learning to allocate: a delay and temperature-aware slot allocation framework for wban with tdma-mac," *Wireless Networks*, vol. 31, no. 1, pp. 165–183, 2025.

[11] R. Negra, I. Jemili, and A. Belghith, "Novel mac protocol for handling correlation and dynamics of medical traffic in wbans," *Ad Hoc Networks*, vol. 151, p. 103236, 2023.

[12] M. Hernandez, R. Kohno, T. Kobayashi, and M. Kim, "New revision of ieee 802.15.6 wireless body area networks," in *International Symposium on Medical Information and Communication Technology (ISMICT)*, 2022.

[13] M. Roy, D. Biswas, N. Aslam, and C. Chowdhury, "Reinforcement learning based effective communication strategies for energy-harvested wban," *Ad Hoc Networks*, vol. 132, p. 102880, 2022.

[14] D. D. Olatinwo, A. M. Abu-Mahfouz, G. P. Hancke, and H. C. Myburgh, "Markov-decision process-based energy-aware mac protocol for iot wban systems," *IEEE Sensors Journal*, vol. 24, no. 17, pp. 27 981–27 997, 2024.

[15] J.-H. Kwon, D. Kim, and E.-J. Kim, "Reinforcement learning-based contention window adjustment for wireless body area networks," in *2023 International Conference on Big Data Analytics and Practices (IBDAP)*, 2023, pp. 1–4.

[16] S. U. Rana, M. M. Hossain, S. Kazary, and M. O. Rahman, "Multi-class multi-load handling mac protocol for wban based on ieee 802.15.6 standard using reinforcement learning," in *2024 IEEE International Conference on Computing, Applications and Systems (COMPAS)*, 2024, pp. 1–6.

[17] C.-M. Wu, Y.-C. Kao, K.-F. Chang, C.-T. Tsai, and C.-C. Hou, "A q-learning-based adaptive mac protocol for internet of things networks," *IEEE Access*, vol. 9, pp. 128 905–128 918, 2021.

[18] V. Aruna, L. Anjaneyulu, and C. Bhar, "Deep-q reinforcement learning based resource allocation in wireless communication networks," in *2022 IEEE International Symposium on Smart Electronic Systems (iSES)*, 2022, pp. 66–72.

[19] P. Consul, I. Budhiraja, R. Arora, S. Garg, B. J. Choi, and M. S. Hossain, "Federated reinforcement learning based task offloading approach for mec-assisted wban-enabled iomt," *Alexandria Engineering Journal*, vol. 86, pp. 56–66, 2024.

[20] Y.-H. Xu, J.-W. Xie, Y.-G. Zhang, M. Hua, and W. Zhou, "Reinforcement learning (rl)-based energy efficient resource allocation for energy harvesting-powered wireless body area network," *Sensors*, vol. 20, no. 1, p. 44, 2019.

[21] S. Bhandari and S. Moh, "A mac protocol with dynamic allocation of time slots based on traffic priority in wireless body area networks," *SSRN Electronic Journal*, 2019.

[22] M. Salayma, A. Al-Dubai, I. Romdhani, and Y. Nasser, "Reliability and energy efficiency enhancement for emergency-aware wireless body area networks (wbans)," *IEEE Transactions on Green Communications and Networking*, vol. 2, no. 3, pp. 804–816, 2018.

[23] S. Bhandari and S. Moh, "A priority-based adaptive mac protocol for wireless body area networks," *Sensors*, vol. 16, no. 3, p. 401, 2016.

[24] R. Negra, I. Jemili, and A. Belghith, "Novel mac protocol for handling correlation and dynamics of medical traffic in wbans," *Ad Hoc Networks*, vol. 151, p. 103236, 2023.

[25] M. Monowar, M. Hassan, F. Bajaber, M. Al-Hussein, and A. Alamri, "Mcmac: Towards a mac protocol with multi-constrained qos provisioning for diverse traffic in wireless body area networks," *Sensors*, vol. 12, no. 11, pp. 15 599–15 627, 2012.

[26] W. H. Wan Hassan, S. Sarang, D. Mohd Ali, G. M. Stojanovic, W. N. S. Wan Muhammad, and N. Ya, "Adaptive medium access control protocol for dynamic medical traffic with quality-of-service provisioning in wireless body area network," *IEEE Access*, vol. 12, pp. 191 461–191 479, 2024.

[27] J. Clifton and E. Laber, "Q-learning: Theory and applications," *Annual Review of Statistics and Its Application*, vol. 7, no. 1, pp. 279–301, 2020.

[28] E. D. and A. Louette, "Introduction to reinforcement learning," 2024.

[29] M. A. Jumaah, Y. H. Ali, and T. A. Rashid, "Efficient q-learning hyperparameter tuning using fox optimization algorithm," *Results in Engineering*, vol. 25, p. 104341, 2025.

[30] A. I. Adamu *et al.*, "Dynamic slot allocation in wireless body area networks: Exploring q-learning approaches," *Journal of Communications*, vol. 20, Aug 2025.

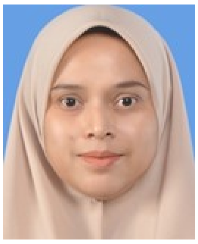
[31] L. Wang, G. Zhang, J. Li, and G. Lin, "Joint optimization of power control and time slot allocation for wireless body area networks via deep reinforcement learning," *Wireless Networks*, vol. 26, no. 6, pp. 4507–4516, 2020.

[32] M. Zheng, L. Chen, W. Liang, H. Yu, and J. Wu, "Energy-efficiency maximization for cooperative spectrum sensing in cognitive sensor networks," *IEEE Transactions on Green Communications and Networking*, vol. 1, no. 1, pp. 29–39, 2017.

[33] Z. Sadreddini, Makul, T. Çavdar, and F. B. Günay, "Performance analysis of licensed shared access based secondary users' activity on cognitive radio networks," in *2018 Electric Electronics, Computer Science, Biomedical Engineering's Meeting (EBBT)*, 2018, pp. 1–4.

[34] S. Goel, K. Guleria, S. N. Panda, F. S. Alharithi, A. Singh, and A. Ali, "An improved routing technique for energy optimization and delay reduction for wireless body area networks," *Egyptian Informatics Journal*, vol. 29, p. 100630, 2025.

Abdu Ibrahim Adamu obtained his Bachelor of Science in Computer Science (Hons) in 2014 from Kano University of Science and Technology, Nigeria, which has now changed to Dangote University of Science and Technology, Kano State, Nigeria. In 2023, he obtained his MSc in Computer Science at Universiti Teknologi Mara, Shah Alam, Selangor, Malaysia. He is pursuing his Ph.D. in Electrical Engineering at Universiti Teknologi Mara, Shah Alam, Selangor, Malaysia. His research interests include Wireless Communication, Artificial Intelligence, Machine Learning, Big Data Analysis, and Cloud Computing.



Wan Haszerila Wan Hassan received the first degree in electrical and electronic engineering and the master's degree in electrical, electronic, and telecommunication engineering from Universiti Teknologi Malaysia (UTM), Malaysia, in 2011 and 2013, respectively. She completed her Ph.D. degree with the Faculty of Electrical Engineering, Universiti Teknologi MARA (UiTM), Malaysia, in 2025. She is also a Lecturer at the Fakulti Teknologi dan Kejuruteraan Elektronik dan Komputer, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia. Her research interests include wireless communication and QoS provisioning in wireless networks.

Darmawaty Mohd Ali is an Associate Professor at the Faculty of Engineering, Universiti Teknologi MARA (UiTM). In 2012, she obtained her Ph.D. from Universiti Malaya (UM). She received her first degree from Universiti Kebangsaan Malaysia (UKM), with honors in Electrical, Electronic, and Systems Engineering in 1999. She began her career as a product engineer before pursuing her master's degree at Universiti Teknologi Malaysia (UTM). Her research interests include wireless access technology and quality of service provisioning in wireless networks.

Wan Norsyafizan Wan Muhamad received her bachelor's degree in electrical engineering from Universiti Malaya (UM), Malaysia, in 2002. She completed her Master of Electrical Engineering at Universiti Malaya (UM) in 2009. She obtained her Ph.D. in 2017 from the University of Newcastle, Australia. She is currently a Senior Lecturer in the Faculty of Electrical Engineering at Universiti Teknologi MARA (UiTM), Malaysia. Her research interest is in wireless communication (Physical and MAC cross-layer optimization). She is a member of IEEE.

Alwatben Batoul Rashed holds a Ph.D. in Computer Science from the Faculty of Computer Science and Information Technology, Universiti Putra Malaysia (UPM), specializing in Intelligent Computing. She is currently an Assistant Professor at the College of Computer, Qassim University, Saudi Arabia. Her work focuses on intelligent computing, data mining, and algorithm optimization. She has published numerous articles advancing machine learning, predictive analytics, and data-driven decision-making.

Mansir Abubakar is a Senior Lecturer in the School of Computing, College of Computing, Informatics, and Mathematics at Universiti Teknologi Mara (UiTM), Shah Alam. He earned his Ph.D. in Intelligent Computing from Universiti Putra Malaysia (UPM), an M.Sc. in Computer Science from Bayero University, Kano, and a B.Sc. in Computer Science from Al-Qalam University, Katsina. He is a certified LMS facilitator and has been an IEEE member since 2016. His research interests include Semantic Web Technology, Ontology Integration, NLP, Machine Learning, and AI applications in agriculture.