Penerbit Akademia Baru

Journal of Advanced Research Design

Journal homepage:
https://akademiabaru.com/submit/index.php/ard
ISSN: 2289-7984

AI-Augmented Centralized Reinforcement Learning for Route Optimization and Network Longevity in IoV Environments

Quadeer Hussain[1], Ahmad Shukri Mohd Noor[1,*], Ashanira Mat Deris[1], Shakeel Ahmed[2]

[1] Department of computer science, communication and security, Faculty Ocean Engineering and Informatics, University Malaysia Terengganu, 21030, Kuala Terengganu, Malaysia
[2] Department of eLeaning & IT, Faculty, Jazan university, Gizan, 45142, Saudi Arabia

ARTICLE INFO

Article history:
Received 13 January 2025
Received in revised form 14 February 2025
Accepted 2 May 2025
Available online 23 May 2025

Keywords:
Internet of vehicles; machine learning; centralize RL base route optimization

ABSTRACT

The Internet of Vehicles (IoV) has emerged as a promising paradigm to revolutionize transportation systems by enabling seamless communication and data exchange among vehicles and infrastructure. A crucial challenge in IoV is optimizing the routes of information, which is used to enhance network efficiency, reduce congestion, and enhance network lifetime. This research presents a centralized reinforcement base learning framework that leverages the power of reinforcement learning algorithms to optimize the routes by minimizing communication overhead. To evaluate the proposed approach, extensive simulations are conducted in a realistic IoV environment, incorporating various scenarios and traffic conditions. Comparative analyses are performed against LEACH, PEGASIS and EER_RL traditional widely used route optimization algorithms and heuristic-based methods to assess the effectiveness and efficiency of the reinforcement learning-based approach. Our findings demonstrate that the reinforcement learning-based route optimization approach exhibits superior performance in terms of reducing energy consumption, minimizing communication overhead, and enhancing network lifetime compared to conventional methods. This work opens new avenues for future research in leveraging RL algorithms for optimizing various aspects of IoV systems and addressing the challenges posed by dynamic and complex vehicular networks.

1. Introduction

The Internet of Vehicles (IoV) is becoming increasingly important as research into its potential benefits continues to grow [38]. Some of the key reasons why IoV is considered important are due to these factors. Several research of IoV can help to improve road safety by enabling vehicles to communicate with each other and with roadside infrastructure. This can help to prevent accidents by providing drivers with real-time information about the traffic conditions around them. Additionally, a few proposed research models on IoV can also help to reduce traffic congestion by enabling vehicles to coordinate their movements. Moreover, IoVs enhance convenience for drivers by providing them with a variety of services, such as remote vehicle diagnostics, real-time traffic information, and even

* *Corresponding author.*
E-mail address: ashukri@umt.edu.my

https://doi.org/10.37934/ard.132.1.124139

entertainment [34]. Recently, environmentally friendly research on IoV shows that IoV help to improve environmental sustainability by reducing fuel consumption and emissions. Modern age moves from the Internet of Things (IoTs) to IoVs it provides a potential research portfolio because the proliferation of high-speed Internet and wireless communication technologies has made it easier to connect vehicles and to the Internet. This connectivity enables real-time data exchange and opens up new possibilities from IoTs to the IoVs [25]. Finally, IoV research contributes to the development of intelligent transportation systems (ITS). These systems integrate various technologies, including communication, sensing, and data analytics to optimize transportation networks. IoV plays a crucial role in ITS by enabling seamless connectivity and facilitating the exchange of information among vehicles, infrastructure, and users [27].

IoV is a network of vehicles equipped with sensors, software and technologies that mediate between these to connect & exchange data over the Internet according to agreed standards [2,39]. It can be seen that the IoV has evolved from a vehicular ad hoc network (VANET) that used a mobile ad hoc network for communication between vehicles and roadside systems to an internet of autonomous vehicles which enables autonomous, connected, electric, and shared (ACES) future mobility. It is expected that IoV will be one of the enablers for ACES future mobility [9]. Here are some advanced applications of IoVs presented. Under the domain of road safety and ITS, IoVs used for collision avoidance systems (CAS) sensors to detect other vehicles, obstacles, and objects in the closed vicinity of the car. If a collision is imminent, the system will alert the driver and take corrective action, such as braking or steering. The adaptive cruise control (ACC) system uses sensors to maintain a safe distance between the car and the vehicle in front of it. The system will automatically adjust the car's speed to maintain the desired distance. Lane departure warning systems (LWS) systems use sensors to detect when the car is about to leave its lane. If the car begins to drift, the system will alert the driver and take corrective action, such as steering the car back into its lane. Traffic signal information systems (TSIS) use sensors to detect traffic signals and provide the driver with information about the current signal status. This information can help drivers to avoid traffic tickets and improve traffic flow. Other than road safety, IoVs effectively implement traffic prediction congestion avoidance, and traveling rote management. IoVs can be used to optimize traffic flow by coordinating the movements of vehicles. For example, vehicles can be routed around traffic jams or directed to less congested roads. Moreover, IoV can be used to provide drivers with some convenience features, such as real-time traffic information, navigation assistance, and parking availability which help to make accurate decisions and save fuel consumption, and furthermore help to improve environmental emissions [15].

While IoV has the potential to revolutionize transportation systems and improve road safety, it also faces several challenges and issues. Here are some of the key challenges associated with the IoVs, and a few are presented here briefly, IoV systems are vulnerable to cyberattacks, including unauthorized access, data breaches, and malicious activities. Protecting the privacy of users' personal information and securing the communication channels within the IoV network is crucial. The interconnected nature of IoV increases the potential attack surface, making it essential to implement robust security measures [3,19]. Another challenge is data management because IoV generates vast amounts of data from various sources, including sensors, cameras, and communication devices. Managing and processing this massive volume of data in real time is a significant challenge. Effective data management solutions, such as data storage, data analytics, and data integration techniques, are required to handle the high velocity, variety, and volume of IoV data [24].

Network congestion is indeed a significant challenge in the context of the IoV to enable communication and data exchange IoV involves the interconnection of vehicles, infrastructure, and other entities. The increasing number of vehicles and the growing demand for real-time data exchange pose significant congestion issues in IoV networks. Congestion issues in IoVs are because of high device density, data volume, and bandwidth contention, latency requirements, heterogeneous network connectivity, infrastructure limitations [12,42]. Network congestion can have a few negative consequences for the IoV, including reduced performance, increased safety risks, and higher resource costs [10]. To address network congestion challenges in IoV, several strategies can be implemented traffic management and prioritization, edge computing and caching implementation, dynamic spectrum access, network optimization and resource allocation, infrastructure expansion, hybrid connectivity solutions, using congestion control and avoidance algorithms, using edge computing, using ml base efficient communication protocols [5].

Mobility issues in the context of the IoVs primarily revolve around the movement and coordination of vehicles within the network. As vehicles move around, their communication links with other vehicles and infrastructure can be disrupted. This can make it difficult to maintain a reliable network connection, which is essential for many IoV applications [40]. Vehicle heterogeneity, dynamic topological connectivity, and random mobility models are major factors that badly affect the network performance and lead to increasing packet loss, bandwidth contention, latency, security breach, flooding, and shortage of network lifetime [34]. Till now to address mobility challenges in IoV, several strategies can be implemented, like efficient routing and handover management, mobility prediction and traffic management, cooperative systems, location-based services and positioning systems, network optimization and resource management, dedicated short-range communication (DSRC).

Routing is one of the most critical challenges in IoVs because efficient routing is crucial in ensuring reliable and timely communication, as well as enabling various IoV applications such as traffic management, collision avoidance, and infotainment services. Due to routing issues, the network faces overhead (spending more resources and time on the communication process instead of data transformation), delay increase, packet delivery ratio exponentially decreased, and all available resources wasted [4]. There are several research efforts underway to address the routing challenges Promising approaches and protocols include proactive, reactive, and hybrid routing protocols, ML base routing algorithms and approaches, deep learning, and Q-learning routing methods. Example of communications in IoVs is illustrated in Figure 1.

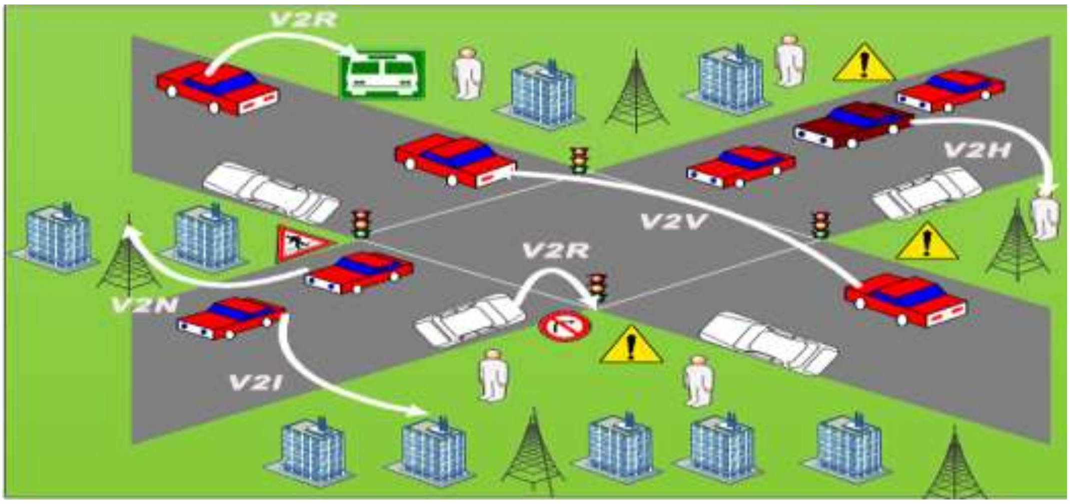

Fig. 1. Communications in IoVs

Finally, one of the key challenges of the IoVs is ensuring energy efficiency. This issue arises because IoV systems involve the integration of vehicles with various communication technologies, such as cellular networks, Wi-Fi, and DSRC. Other than technologies many challenges further lead to compromised network energy efficiency like data processing and analysis, constant heavy connectivity requirements, power-hungry sensors, and topological implementation, communication overhead (CO) network contention, random mobility, and rout disruption. Induction of these technologies and other mentioned factors consume more energy, furthermore, inefficient use of energy can lead to shorter battery life [28]. Some of the potential solutions include using more efficient sensors, power-saving modes, renewable energy sources, more efficient communication protocols, intelligent overhead avoidance models, sustainable and reliable hybrid routing protocols, and congestion avoidance ML prediction models are promising methods and protocols to achieve energy efficiency in IoVs [36,44].

As IoV networks continue to grow, CO is likely to become an increasingly important issue. CO is the state of the system when the system spends most of the time on the communication process (like sending, receiving, and processing inquiry messages or frequently initiating communication process or spending more time on communication handling all types of CO) instead of performing data transformation. CO has sworn effects on network performance like increased latency or throughput degradation, reduced operational bandwidth, flooding of unnecessary massages, and high draining of battery backup. To handle this issue there are a few ways to reduce CO in IoV networks like traffic scheduling prioritizing at an initial stage, information slicing, time sharing approach, and developing route optimization protocol for information transformation.

Route optimization plays a central role in IoVs because for timely and accurate decision-making information needs to be routed from source to destination without delay [18]. In this regard, we are facing a major challenge like communicational overhead, which further leads to delay, congests the network traffic flow, and uses much available energy which further degrades the network lifetime. Regarding communication, multiple unnecessary packets become part of the network, which reserves the resource (if throughput capacity at a given bandwidth is supposed 100 packets, then it's may possible more than 50 packets are due to CO and the system will only use half of the bandwidth). Furthermore, if communicational overhead is not tackled efficiently all resources like energy, bandwidth, computational power, and memory are futile.

So, in the research, we aim to provide optimum routing by minimizing the routing overhead. Furthermore, this minimization of routing overhead aims to minimize CO in IoV. For this purpose, we are going to propose a centralized reinforcement learning base route optimization model (CRLRO). Reinforce learning (RL) issues are formalized as Markov decision processes (MDPs) with a tuple (S,P,R), where S denotes the possible states for an agent at a given time t and A denotes the potential actions the agent might take. ρ represents the transition probability that an agent will perform action at time t and then transition from state s(t) to state s(t+1), where R represents the reward, the agent will receive for completing the action. CRLRO is centralized by nature, in which all computation (reward calculation and matrix updating, yield matrix maintenance and updating, and optimum route decision) is done by a centralized entity like a base station (entity capable of offering long-range transmission coverage and providing global internet connectivity). A base station (BS) broadcasts an Inquiry (INQ) message like a heartbeat message (HM), all the under-coverage nodes receive HM, and respond to the request to reply (RREP), but due to a shortage of node transmission RREP is not reached to BS directly (expect few which are one step away from BS) so intermediate node (IN) require to relay the RREP. RREP messages are embedded with multiple information like sender address and location, route path including source to an IN, residual energy of each node, number of connectivity, and traffic congestion level. By using the IN as a relay node, the RREP of all the nodes reach BS, after receiving RREQ, BS extracts all information and builds a reward and yield matrix (detail will be discussed in the methodology chapter). In a distributed architecture, if INQ and RREP broadcasting is the major cause of CO because at each node level broadcasting and rebroadcasting continue which becomes a bottleneck when the number of connectivity grows up, we must minimize it at an initial level as well.

2. Proposed Methodology

In a distributed architecture, INQ and RREP broadcasting and rebroadcasting is the major cause of CO because at each node level broadcasting and rebroadcasting continue which becomes a bottleneck when the number of connectivity grows up. In the initial phase going to implement our model with centralized architecture, because BS has long-range coverage capability, heavy computation power, and no issue with backup power. While implementing a centralized approach aimed to minimize rebroadcasting during the inquiry period, here is the first layer of controlling CO. For this purpose, CRLRO, in which all computations (reward calculation, matrix updating, yield matrix maintenance and updating, and finally optimum route decision) are done by a centralized entity like base station (entity capable of offering long rang transmission coverage and providing global internet connectivity). BS broadcasts an INQ message like a HM, all the under-coverage nodes receive HM, and respond to the request to reply (RREP), but due to a shortage of node transmission, RREP is not reached to BS directly (expect few which are one step away from BS) so IN require to relay the RREP. RREP messages are embedded with multiple information like sender address and location, route path including source to an IN, its residual energy, node connectivity, and traffic congestion level of each available. By using the IN as a relay node, the RREP of all the nodes reaches BS, after receiving RREQ, BS extracts all information and builds a reward and Yield matrix. So, the working principle of CRLRO is comprised of sending HM by BS, under coverage node start RREP by putting values like position, number of connectivity, residual energy, link traffic condition, and path information (to find shortest path) from source to destination. All this information is put in the node information part (NIP) of the message header, and initiated by the edge node (which is not used as a relay node) while passing to IN, each node puts all this information into NIP and sends it to next level, till NIP reached to BS. On reception of NIP from the whole network, BS starts calculating the reward function (high reward for those who have the shortest path, high residual energy, more connectivity, and low link congestion level) and puts reward value in the reward matrix detail will be discussed in the reward matrix calculation part. Furthermore, each NIP reward matrix needs to be updated, next step is to calculate the yield function by applying Q_Learning at each NIP and update the yield matrix after each NIP. Finally, the optimum path-finding process starts, when any routing requires the source node to send the request to BS through IN, BS finds the optimum route by analyzing the yield matrix and responds as optimum RREP (ORREP), to the requesting node (detail will be discussed in optimum path finding part). So, here are calculations involving reward finding, yield calculation (Q_Matrix updating), and optimum pathfinding. Now try to elaborate on variable detail here,

To find the shortest path, instead of measuring the actual distance in meters or kilometers, the distance from a source to a destination is typically expressed in terms of "hops" (Fewer hops mean shortest path), residual energy is the remaining energy level of each node, number of connectivity is found through the neighbor table and finally link congestion level need to be found. Congestion will occur when the input exceeds the output and input data needs to wait in a memory buffer, in some cases memory buffer is also saturated, and the data packet starts dropping. So, to check congestion, check the buffer counter for each point which shows the congestion level for each link. After receiving HM, each node puts information in the index listed below and forwards it to its neighbors, till it reaches BS. When all this information reaches BS, at each HM, now BS starts applying RL to find the optimum path. As RL consists of two parts i.e., reward calculation and yield matrix maintenance and updating, Now I will explore how CRLRO calculates the reward matrix as shown in Eq. (1). Here predefine reward parameters here, high reward for those who have the shortest path (fewer hops), high residual energy, more connectivity, and low link congestion level and *vice versa*.

$$rt + 1 = (1 - P)\left[\left(\frac{1}{hp}\right) + \left(\frac{1}{Cong}\right)\right] + p[Er + com] \quad (1)$$

Here, r_{t+1} is the reward at time t+1, "ρ" is the probabilistic variable (taken value from 0 to 1) which shows the impact on other variables, like high ρ mean high chances to be selected, $h_p$is the hop count, $c_{ong}$is the congestion level, $E_r$residual energy Etotal= Etx+Erx is the number of connectivity.

Now I calculate the Q_Learning yield function by using the Q_learning matrix, here is the standard equation, the Q_Learning equation states that the Q value at a given time (t+1), while the state is (s) and action is (a), will be calculated as a reward at that state and action plus a maximum of all possible stats and corresponding actions (will select the state which has higher previous value) and multiply by Υ (which is learning rate) as shown in Eq. (2).

$$Q_{t+1}(s,a) = r_{t+1}(s,a) + \Upsilon \times max\left[Q\left(next(s), all\_possible\_actions(a^{/})\right)\right] \quad (2)$$

3. Related Work

Routing is the process of selecting paths for network traffic to travel from its source to its destination across a network, it determines the path that data packets should take to reach their intended destination efficiently and reliably [13]. Routing enables efficient and effective communication between different devices and networks by dynamically adapting to changes in network conditions, such as link failures or congestion [33]. It plays a crucial role in ensuring that data packets are delivered accurately and promptly to their intended destinations across the complex network infrastructure. Routing protocols and algorithms are used to exchange information and various metrics, such as hop count, link bandwidth, delay, or cost, to evaluate and select the most suitable path for packet delivery. Several issues and challenges can arise in routing, including, Topology changes, Link failures, Traffic congestion, Scalability, Routing loops, Routing table overload (memory leakage), and Routing protocol convergence [29]. Routing is a complex and ever-evolving field. As networks continue to grow and become complex, the challenges of routing will also increase. To handle these challenges and offer energy efficiency a few protocols are proposed under the category of proactive, reactive, and machine learning-based algorithms and models [8].

3.1 Proactive Routing Protocols

Also known as a table-driven protocol because all routing information needs to be stored and updated in short-term memory called a routing table. Table-driven routing systems rely on regular updates to ensure that all mobile nodes have accurate and up-to-date information about the network [31,32]. Table-driven routing ensures that connections are always made to every destination on the network. Every node must frequently broadcast RREQ so that all receiving nodes ensure accurate route information. Routing methods based on tables include Optimal Link State Routing (OLSR) and destination-sequenced distance vector (DSDV).

3.1.1 Sequential distance vector routing to multiple destinations (DSDV)

Each entry in the routing table is assigned a sequence number; if a link is present, the number is even, otherwise it is odd. If the destination generates the number, the sender must include it in the next update. Nodes share routing data by infrequently exchanging complete dumps and more commonly exchanging incremental changes [16]. A guarantee of loop-free DSDV pathways is provided due to odd-even number assignment. INQ and RREQ massages can be controlled because instead of keeping numerous pathways to each destination, DSDV keeps only the best one. The routing tables of this protocol need to be updated regularly, which consumes some bandwidth and battery life. A new most recent sequence number is required whenever the topology evolves. But when due to mobility, route disturbing probability increase updating can cause CO.

3.1.2 Protocol for optimal link state routing (OLSR)

OLSR uses INQ and topology control (TC) messages, it actively discovers and disseminates connection state information across the network. With this knowledge of the network's structure, each node may determine the shortest path to the next hop destination for every other node. All network destinations are known and updated proactively [11]. OLSR performs well for the small-scale network, but due to not defining any quality of a connection mechanism not fit for IoVs. Moreover, variants of OLSR which are quite suitable for IoVs consumed high Power and faced bandwidth contention.

These protocols are highly suitable for small-scale static networks; by providing fast and accurate routing information, also short-range mobility does not affect table maintenance. However, for a large stage and long-range mobility, these models do not provide suitable results. The reason behind this is that it can cause flooding into the network and wastage of available resources, Since RREQ messages are sent out even when there is no data flow, proactive routing techniques may be wasteful of bandwidth.

3.2 Reactive Routing Protocols

Due to the lack of requirement to actively seek out and maintain routes in the absence of data traffic, On-Demand (reactive) routing is greatly beneficial under this scenario. In this approach, the path between a source and a destination will be found only when it's needed, as opposed to the proactive protocols where all paths are kept indefinitely. Therefore, with reactive protocols, we can ignore paths that are currently not in use. Moreover, network traffic can be managed by not sending unnecessary control messages, which is the main difference between proactive and reactive protocols, and the cost of maintaining unused routes can be. Since reactive protocols only calculate routes when they are needed, they are significantly slower than proactive ones. Some examples of such protocols are hoc on-demand Distance Vector (AODV), Dynamic Source Routing (DSR), and so on.

3.2.1 Dynamic source routing (DSR)

By eliminating the need for the periodic table-update messages mandated by the table-driven approach, the DSR protocol can reduce the amount of network bandwidth devoted to control packets, the sender provides the full route to the destination. The packets are sent along by the INs according to the paths they were given by their sources. If a route to the destination exists in the node's route [17]. A node's route to its destination can be chosen from among several possible routes, additionally, when a connection fails, an error message is sent back to its point of origin, invalidating the path.

3.2.2 Efficient power aware routing protocol (EPAR)

Protocol for Efficient and Power-Conscious Routing This is an on-demand routing protocol that reduces energy usage per packet. EPAR determines a node's capability by measuring its remaining battery life and the predicted energy required to reliably transfer data packets over a given link. EPAR employs a mini-max formulation to determine the best possible path for transmitting packets while minimizing the number of packets lost in the process. The goal of EPAR is to extend the lifetime of the network by reducing the variation in the residual energy of all the nodes. Saves power, speeds up connections, and lengthens battery life [14]. Authors show that EPAR is used to reduce residual energy, save power, speed up connections, and lengthen battery life. Despite ERAR encountering issues because while maintaining two paths active and backup, dwindling battery power, follows a circuitous routing path that lacks a backup route in case of a link breakdown.

3.3 Machine Learning-Based Energy-Efficient Routing Models

Machine learning (ML) can indeed be used for route optimization to achieve energy efficiency under diverse conditions [43]. ML algorithms can examine various factors like traffic patterns, load conditions, number of alternative suitable paths, shortest path, and most importantly energy-efficient routes by using historical data and real-time information [37]. As machine learning technology continues to develop, we can expect to see even more innovative ways to use this technology to make the model more intelligent and autonomous which adopts as per environmental impact to achieve the goal. These systems can be valuable for various IoTs applications, like health care, self-driving vehicles, multi-agent gaming apps, smart homes, etc. where energy efficiency is crucial [20,26]. Since the last decay number of research focus on this area and proposed ML bade models to achieve energy efficiency for specific objectives or handling challenges, few are presented here.

3.3.1 Routing protocol for low-power and lossy networks (RPLPL)

Ancillotti *et al.,* [6] proposed the Routing Protocol for Low-Power and Lossy Networks (RPL) link quality monitoring scheme has been proposed to keep network routing information current and respond immediately to link quality variations and topology changes brought on by nodes' mobility. To reduce the overhead brought on by active analytical operations, RL has been used. The proposed approach helps to improve packet loss rates and energy consumption, only for single-channel networks. While considering only link quality and remaining power not fit for multi hops communications, because for long ranges link quality will be compromised and energy level decreased exponentially.

3.3.2 State action reward state action (SARSA)

Yang *et al.,* [41] proposed a trusted routing system to save energy during a flooding attack period, a model built on blockchain technology and RL has been introduced to enhance secure data transactions. The proposed scheme makes use of the tamper-proof (i.e., detection of flooding of bombardment of INQ massage), decentralized, and traceability properties of the blockchain to handle the routing information, while RL assisted nodes in selecting more reliable and effective relay nodes to further develop the model. To choose dependable routing links, the RL algorithm in each routing node dynamically learns the trusted routing data on the blockchain. The study's findings demonstrate that adding RL to the blockchain system reduces energy use and improves delay performance even when there are 50% of malicious nodes in the routing environment.

3.3.3 Optimized path algorithm based on reinforcement learning (OPABRL)

Liu *et al.,* [22,23] have solved the issue of shortest paths selection in multi hops and estimated energy consumption. The authors proposed a relatively shorter path with fewer turns was analyzed and found using the OPABRL proposed in these papers. It combined prior RL technology with searching for the optimal shortest path algorithm and expected energy consumption. With a probability of 98% close to the ideal solution, OPABRL outperforms all other algorithms frequently used in intelligent robot trajectory planning in terms of the number of turns, the path lengths, and the running time. Metrics include variables like packet loss, data content, the separation between a forward node and sink node, and residual energy. OPABRL displays the best results, but they are all for a static network and understate the mobility of the nodes when we consider increased mobility packet drop rates and power consumption.

3.3.4 Prior knowledge reinforcement learning (PKRL) and RLBPLP

Liu *et al.,* [23] and Pasandideh *et al.,* [30] have created a routing algorithm PKRL and RL best path selection with length priority (RLBPLP), which deliberate metrics like packet loss, data content, the distance between a forward node and a sink node, residual energy, and these other factors are taken into consideration to ensure high throughput while using less energy. By updating the q-value in accordance with the metrics gathered, each node learns the behavior of its neighbors and avoids unfit nodes in the upcoming routing scheme. Depending on the percentage of participating device nodes, the packet delivery rate increased over time. On the other hand, the proposed approaches use less energy, because it is distributed and prefers short-range communication, this model is vulnerable to flooding in the network.

3.3.5 Clustered state action reward state action (C-SARSA)

Variants of SARSA have been proposed by Aslam *et al.,* [7] for multi-objective optimization function known as Clustering SARSA (C-SARSA) for WSN data routing, using clustering and the RL method SARSA. The main goals are to distribute energy fairly and extend its idle time, moreover data generation rate, energy consumption rate for receiving and sending the data, arrival time, charging time, travel time, and field have all been considered. Each node chooses its willingness, which indicates whether it can take part in choosing the data route, based on the remaining energy. This willingness makes this architecture for fair resource allocation which extends the lifetime but due to flat implements architecture, the suggested solution is unsuitable for large-scale WSNs and only intended for small networks with low requirements.

3.3.6 Intelligent IoT connectivity deep reinforcement learning (ICDRL) approach

Kwon *et al.,* [21] have developed a Double Deep Q-Network-based (DQN) distributed decision-making process for multi-hop wireless ad hoc networks. By choosing the best course of action based on an online network and calculating the target Q-value of that action using a target network, Double DQN addresses the issue of overestimating Q-values. To increase network throughput and reduce the amount of corresponding transmission power consumption, each relay node modifies its transmission power to widen or narrow the wave range. ICDRL handles routing optimization issues that RL techniques have addressed as path selection in terms of QoS and distance as well as routing strategies based on energy efficiency. But due to distributed nature and implementation of double calculation make it more complex and increase flooding probability.

3.3.7 Reinforcement learning-based clustering-enhanced protocol (ReLeC)

Sharma *et al.,* [35] have proposed ReLeC protocol that enables devices to make more accurate routing decisions, improve next-hop selection, and use less energy by sharing local information with the neighborhood. While the sender adds local data to the packet header, neighboring devices' routing tables are modified in accordance with the contents of the communicating device's packet header. Ids, residual energy value, geographic coordinates. ReLeC has three levels: network initialization, CH election and cluster creation, and communication phase, just like other effective, clustering-based routing protocols. Simulation results show that high throughput extends the network lifetime, but most of the resources used for CH selection, moreover if due to mobility CH disturbance probability increased, because ReLeC did not define any rule to maintain CH to avoid disturbance due to mobility.

3.3.8 Reinforcement-learning-based energy efficient control and routing protocol (RLBEEP)

Abadi *et al.,* [1] have utilized reinforcement learning (RLBEEP) to maximize the long-term reward received by each node when optimizing routing policies. Three energy management strategies have been suggested to extend the lifespan of wireless sensor networks. The first strategy is to use reinforcement learning to correctly navigate to shorten routes and reduce energy consumption. The second strategy is to increase node energy consumption by utilizing a sleep scheduling technique. The final method is used to limit each node's data transmission based on the received data change rate. Simulation results demonstrate that the proposed method significantly outperforms previously reported methods in terms of network lifespan. On the other hand, due distribution nature and implementation of three strategies make the implementation more complex and time-consuming. CO was also introduced because for all three techniques, more and more INQs of different natures need to be broadcast which is the main source of CO, and this overhead reduces.

4. Results

We used the NS3 simulation tool and will deploy up to 100 nodes at random over a sensing field in a random fashion under the coverage of BS, in addition, we considered the network to be heterogeneous, with devices ranging in energy up to 5 joules maximum as shown in Table 1. All nodes can communicate with their neighbor nodes (except nodes that are 1 hop away from BS). Because it's an iterative process, model will run thousands of times to find the optimum path, but when the model reaches BS through any route it is called an episode or round. We compare proposed model CRLRO with Low-energy adaptive clustering hierarchy (LEACH), Power-Efficient Gathering in Sensor Information Systems (PEGASIS) and Energy-Efficient Routing Based on Reinforcement Learning (EER_RL), which is currently inducted and implemented also having the same circumstances.

Table 1
Parameters and its value

Parameter	Value
Network operation area	1 km square
Number of nodes	25-100
Initial energy level	Up to 5 joules maximum
Maximum data size	4000 bits
E_{Tx}	100 Neno Joules per bit
E_{Rx}	50 Neno Joules per bit
ρ	100*10-9

Network lifetime: The time till the network can transmit and receive data, so to check network lifetime we carry out multiple performance checks as compared to LEACH and flat EER-RL. We use four matrices in this regard, number of alive nodes vs. time (number of episodes), energy consumption at each round, and the time when the first node becomes dead while changing the node density (from 25 to 100) This matrix shows the lifetime over time. To check which phase of the model consumes how much energy, a phase-wise energy consumption graph will be drawn and evaluated.

4.1 Number of Alive Nodes Over Time

We are going to compare lifetime by evaluating some alive nodes after each episode or round. We are putting some episodes at the x-axis and node density at the y-axis, in the first two graphs node density is low (25 and 50) then LEACH and flat EER-RL perform well because LEACH use the clustering technique, and small networks with few clusters are formed, and the network will manage energy efficiently. Moreover, flat EER-RL uses distributed RL with multiple agents' policy, so after a few iterations model will find the optimum path with no more energy consumed. On the other hand, when density increases CRLRO shows tremendous results because most of the energy is implemented during the reward and Q_matrix updating process which is done by BS, so most of the energy is conserved. But in the case of LEACH, the number of cluster increases, and the formation of cluster and the selection of cluster head consume much energy, also in the case of Flat EER-RL, due to its distributed nature CO occur which consume much energy as shown in Figure 2.

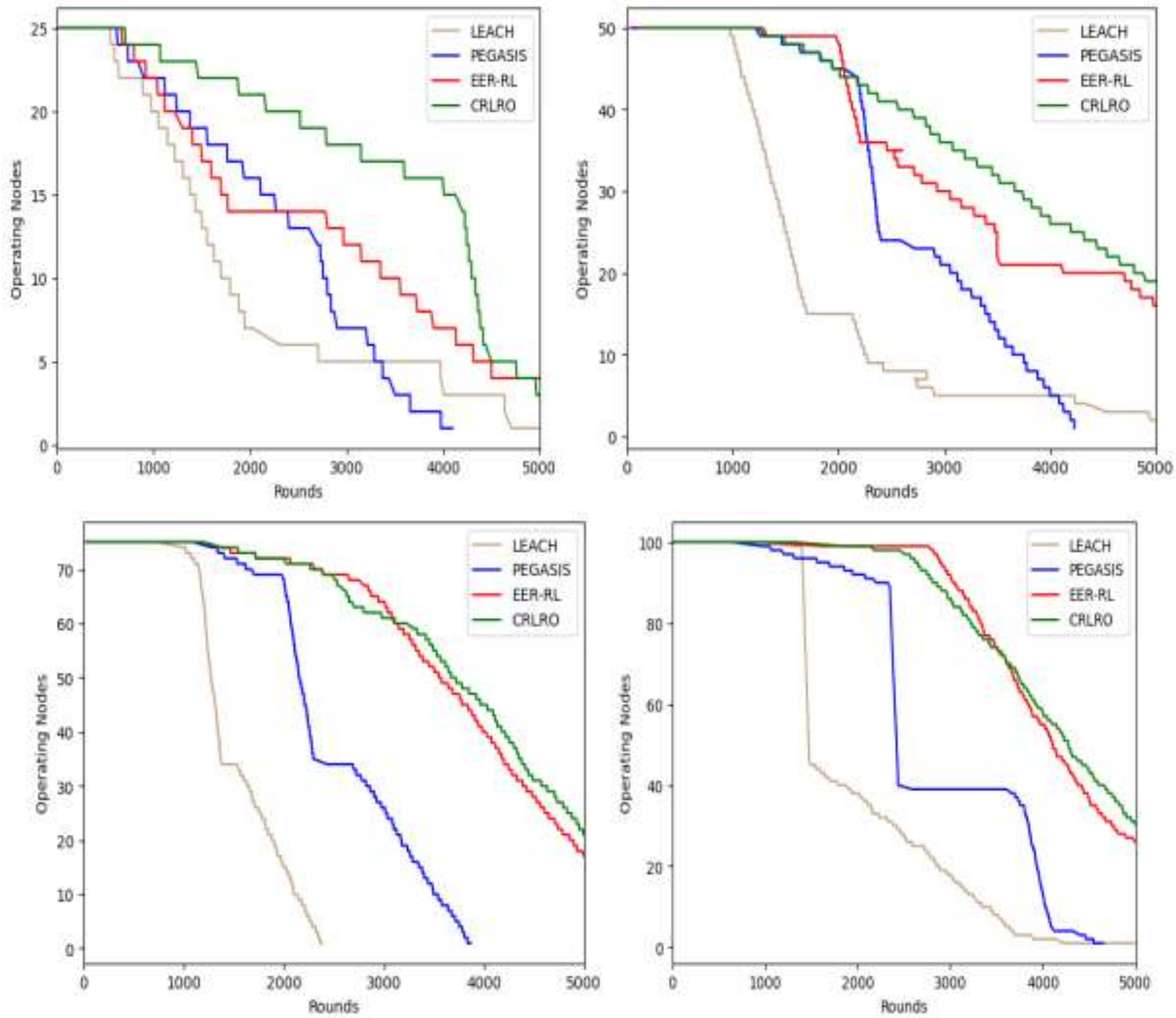

Fig. 2. Network lifetime w.r.t network density

4.2 The Time When the First Node Becomes Dead

To evaluate the lifetime, we draw a bar graph, divide this graph into four patches as per node density, and calculate the time for each patch as shown in Figure 3. We notice that LEACH performs consistently around (550-600 seconds) at each density level, but this time is quite shorter. Flat EER_RL long time at basic three density levels (25,50 and 75) but while reaching 100 exponentially decreased, because at 100 node level, CO occurs (in distributed model while increasing node density flooding of INQ will start which will cause energy consumption). While CRLCO performs well at high-density level, due to reaching request message to BS quickly and sharply, high, and accurate response of BS on each request only one-way INQ message is required (instead of two-way INQ for LEACH, Flat EER_RL), which conserves the weighty energy.

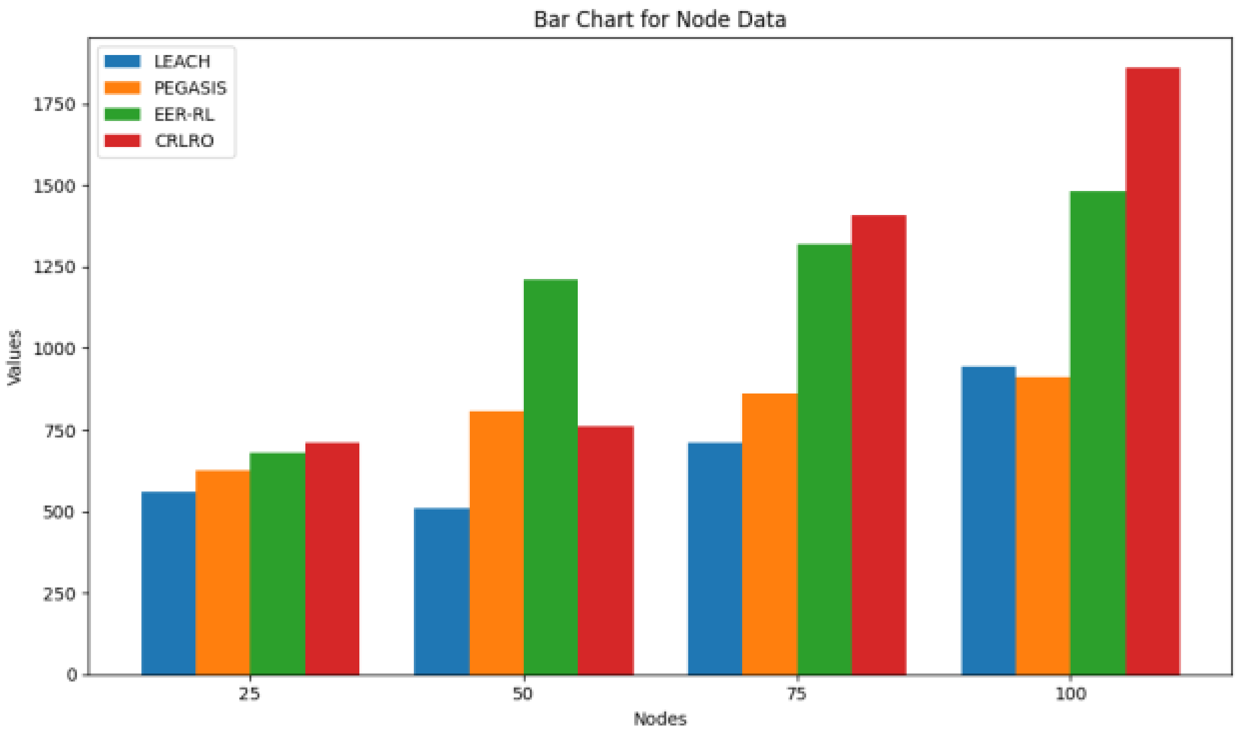

Fig. 3. Time when first node become dead

4.3 Long-Term Energy Consumption Evaluation

Now it's time to evaluate energy consumption to complete the network for the long term. We observed heavy fluctuations in LEACH because its focus on short-term cluster forming and long-term cluster maintenance process consumes much energy as depicted in Figure 4. Also, the Flat EER-RL model focuses on short-term rewards instead of long-term rewards, due to short-term reward and distributed architecture Flat EER-RL consistently consumes more energy. But we see a steady state line in the CRLCO model because at each level energy computation is the same because all maintenance and updating are done in BS, the source node just sends INQ when required and BS will provide a complete path efficiently, no use of much energy. We see in LEACH, initially steady, after those heavy fluctuations for much time, then calms down and again at the end energy fluctuations. Because of that LEACH applies a cold start strategy (state line at the initial point), then clustering and cluster head process start which requires heavy computation and communication for a long time (heavy fluctuation in the middle), and finally due to no specific cluster head selection process, may require cluster head selection at the end also (again fluctuation at the end also). In the case of the Flat EER_RL model, we see a little bit of high consumption at the starting state, then consumption variation at the middle and end points. It's because, in the start stage, each node starts broadcasting and rebroadcasting to build the RL model (high energy consumption at the initial point), after the RL model starts the optimization process which continues till the end and consumes energy. In our model cold steady start, and no need for any broadcasting and rebroadcasting message why steady state line (starting from the initial point to the middle). After that HM is initiated by BS and all nodes start responding that why energy fluctuation is observed in the middle, after BS build a model and optimize it by using its energy (BS energy is not included in network energy).

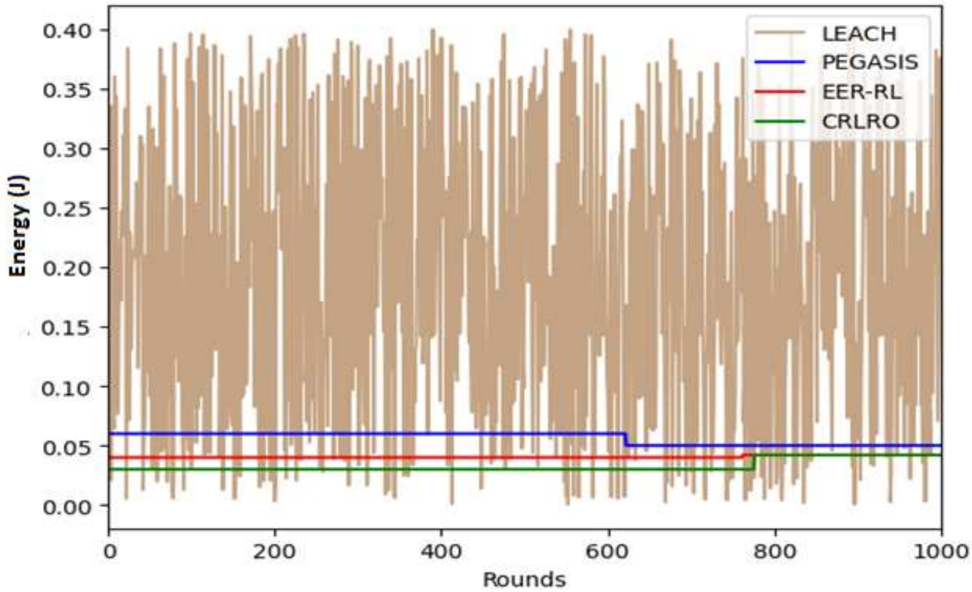

Fig. 4. Phase wise energy consumption

5. Conclusions

Our research presents a route optimization model under diverse conditions, that successfully improves the network lifetime on the Internet of Vehicles by minimizing communication overhead. By implementing RL approach in a centralized form, we successfully control the communication overhead which further leads to minimize energy consumption. We evaluated our model in multiple directions and found much better results under different possible situations with respect to LEACH, PEGASIS and EER_RL. The implications of this research use its knowledge of the network to make optimal extend beyond the IoV, as the concepts and methodologies can be applied to other dynamic and resource-constrained networks as well. As the IoV continues to evolve and expand, our model can serve as a valuable tool to optimize communication, promote sustainability, and ensure a seamless and reliable vehicular networking experience.

Reference

[1] Abadi, Ali Forghani Elah, Seyyed Amir Asghari, Mohammadreza Binesh Marvasti, Golnoush Abaei, Morteza Nabavi, and Yvon Savaria. "RLBEEP: Reinforcement-learning-based energy efficient control and routing protocol for wireless sensor networks." *IEEE Access* 10 (2022): 44123-44135. https://doi.org/10.1109/ACCESS.2022.3167058

[2] Abbasian Dehkordi, Soroush, Kamran Farajzadeh, Javad Rezazadeh, Reza Farahbakhsh, Kumbesan Sandrasegaran, and Masih Abbasian Dehkordi. "A survey on data aggregation techniques in IoT sensor networks." *Wireless Networks* 26, no. 2 (2020): 1243-1263. https://doi.org/10.1007/s11276-019-02142-z

[3] Aburashed, Laila, Marah AL Amoush, and Wardeh Alrefai. "SQL Injection Attack Detection using Machine Learning Algorithms." *Semarak International Journal of Machine Learning* 2, no. 1 (2024): 1-12. https://doi.org/10.37934/sijml.2.1.112

[4] Agbaje, Paul, Afia Anjum, Arkajyoti Mitra, Emmanuel Oseghale, Gedare Bloom, and Habeeb Olufowobi. "Survey of interoperability challenges in the internet of vehicles." *IEEE Transactions on Intelligent Transportation Systems* 23, no. 12 (2022): 22838-22861.https://doi.org/10.1109/TITS.2022.3194413

[5] Ali, Elmustafa Sayed, Mohammad Kamrul Hasan, Rosilah Hassan, Rashid A. Saeed, Mona Bakri Hassan, Shayla Islam, Nazmus Shaker Nafi, and Savitri Bevinakoppa. "Machine learning technologies for secure vehicular communication in internet of vehicles: recent advances and applications." *Security and Communication Networks* 2021, no. 1 (2021): 8868355. https://doi.org/10.1155/2021/8868355

[6] Ancillotti, Emilio, Carlo Vallati, Raffaele Bruno, and Enzo Mingozzi. "A reinforcement learning-based link quality estimation strategy for RPL and its impact on topology management." *Computer Communications* 112 (2017): 1-13. https://doi.org/10.1016/j.comcom.2017.08.005

[7] Aslam, Nelofar, Kewen Xia, and Muhammad Usman Hadi. "Optimal wireless charging inclusive of intellectual routing based on SARSA learning in renewable wireless sensor networks." *IEEE Sensors Journal* 19, no. 18 (2019): 8340-8351. https://doi.org/10.1109/JSEN.2019.2918865

[8] Biersack, Ernst, Christian Callegari, and Maja Matijasevic. *Data traffic monitoring and analysis*. Heidelberg, Germany: Springer Berlin Heidelberg, 2013. https://doi.org/10.1007/978-3-642-36784-7

[9] Campolo, Claudia, Antonella Molinaro, and Riccardo Scopigno. "Vehicular ad hoc Networks." *Standards, Solutions, and Research* (2015). https://doi.org/10.1007/978-3-319-15497-8

[10] Chen, Min, Yuanwen Tian, Giancarlo Fortino, Jing Zhang, and Iztok Humar. "Cognitive internet of vehicles." *Computer Communications* 120 (2018): 58-70. https://doi.org/10.1016/j.comcom.2018.02.006

[11] Clausen, Thomas, and Philippe Jacquet. *Optimized link state routing protocol (OLSR)*. No. rfc3626. 2003. https://doi.org/10.17487/rfc3626

[12] El Madani, Samira, Saad Motahhir, and Abdelaziz El Ghzizal. "Internet of vehicles: concept, process, security aspects and solutions." *Multimedia Tools and Applications* 81, no. 12 (2022): 16563-16587. https://doi.org/10.1007/s11042-022-12386-1

[13] Elappila, Manu, Suchismita Chinara, and Dayal Ramakrushna Parhi. "Survivable path routing in WSN for IoT applications." *Pervasive and Mobile Computing* 43 (2018): 49-63. https://doi.org/10.1016/j.pmcj.2017.11.004

[14] Golla, Varaprasad, G. Jayanthi, and H. N. Shivashankar. "Designing energy routing protocol with power consumption optimization in MANET'." *IEEE Transactions on Emerging topics in Computing* 2, no. 2 (2014): 192-197. https://doi.org/10.1109/TETC.2013.2287177

[15] Gupta, Nishu, Arun Prakash, and Rajeev Tripathi, eds. *Internet of vehicles and its applications in autonomous driving*. Cham, Switzerland: Springer, 2021. https://doi.org/10.1007/978-3-030-46335-9

[16] Hu, Yih-Chun, David B. Johnson, and Adrian Perrig. "SEAD: Secure efficient distance vector routing for mobile wireless ad hoc networks." *Ad hoc networks* 1, no. 1 (2003): 175-192. https://doi.org/10.1016/S1570-8705(03)00019-2

[17] Johnson, D. "Dynamic source routing in ad hoc wireless networks." *Mobile Computing/Kluwer Academic Publishers* (1996).

[18] Kayarga, Tanuja, and S. Ananda Kumar. "A study on various technologies to solve the routing problem in Internet of Vehicles (IoV)." *Wireless Personal Communications* 119 (2021): 459-487. https://doi.org/10.1007/s11277-021-08220-w

[19] Kimani, Kenneth, Vitalice Oduol, and Kibet Langat. "Cyber security challenges for IoT-based smart grid networks." *International journal of critical infrastructure protection* 25 (2019): 36-49. https://doi.org/10.1016/j.ijcip.2019.01.001

[20] Kopytko, Vasyl, Lyubov Shevchuk, Larysa Yankovska, Zhanna Semchuk, and Rostyslav Strilchuk. "Smart home and artificial intelligence as environment for the implementation of new technologies." *Traektoriâ Nauki= Path of Science* 4, no. 9 (2018): 2007-2012. https://doi.org/10.22178/pos.38-2

[21] Kwon, Minhae, Juhyeon Lee, and Hyunggon Park. "Intelligent IoT connectivity: Deep reinforcement learning approach." *IEEE Sensors Journal* 20, no. 5 (2019): 2782-2791. https://doi.org/10.1109/JSEN.2019.2949997

[22] Liu, Xiao-huan, De-gan Zhang, Ting Zhang, and Yu-ya Cui. "New method of the best path selection with length priority based on reinforcement learning strategy." In *2019 28th International Conference on Computer Communication and Networks (ICCCN)*, pp. 1-6. IEEE, 2019. https://doi.org/10.1109/ICCCN.2019.8847049

[23] Liu, Xiao-huan, De-gan Zhang, Ting Zhang, and Yu-ya Cui. "Novel approach of the best path selection based on prior knowledge reinforcement learning." In *2019 IEEE International Conference on Smart Internet of Things (SmartIoT)*, pp. 148-154. IEEE, 2019. https://doi.org/10.1109/SmartIoT.2019.00031

[24] Ma, Meng, guptaPing Wang, and Chao-Hsien Chu. "Data management for internet of things: Challenges, approaches and opportunities." In *2013 IEEE International conference on green computing and communications and IEEE Internet of Things and IEEE cyber, physical and social computing*, pp. 1144-1151. IEEE, 2013. https://doi.org/10.1109/GreenCom-iThings-CPSCom.2013.199

[25] Mahmood, Zaigham. "Connected vehicles in the internet of things." *Cham, Switzerland: Springer* (2020). https://doi.org/10.1007/978-3-030-36167-9

[26] Mohamad, Muhammad Arif, and Muhammad Aliif Ahmad. "Handwritten Character Recognition using Enhanced Artificial Neural Network." *Journal of Advanced Research in Computing and Applications* 36, no. 1 (2024): 1-9. https://doi.org/10.37934/arca.36.1.19

[27] Muthuramalingam, S., A. Bharathi, S. Rakesh Kumar, N. Gayathri, R. Sathiyaraj, and B. Balamurugan. "IoT based intelligent transportation system (IoT-ITS) for global perspective: A case study." *Internet of things and big data analytics for smart generation* (2019): 279-300. https://doi.org/10.1007/978-3-030-04203-5_13

[28] Mutombo, Vially Kazadi, Seungyeon Lee, Jusuk Lee, and Jiman Hong. "EER-RL: Energy-Efficient Routing Based on Reinforcement Learning." *Mobile Information Systems* 2021, no. 1 (2021): 5589145. https://doi.org/10.1155/2021/9817562
[29] Parissidis, Georgios, Merkourios Karaliopoulos, Rainer Baumann, Thrasyvoulos Spyropoulos, and Bernhard Plattner. "Routing metrics for wireless mesh networks." *Guide to wireless mesh networks* (2009): 199-230. https://doi.org/10.1007/978-1-84800-909-7_8
[30] Pasandideh, Faezeh, João Paulo J. da Costa, Rafael Kunst, Nahina Islam, Wibowo Hardjawana, and Edison Pignaton de Freitas. "A review of flying ad hoc networks: Key characteristics, applications, and wireless technologies." *Remote Sensing* 14, no. 18 (2022): 4459. https://doi.org/10.3390/rs14184459
[31] Perkins, Charles E. *Ad hoc networking*. Pearson Education India, 2008.
[32] Perkins, Charles E., and Pravin Bhagwat. "Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers." *ACM SIGCOMM computer communication review* 24, no. 4 (1994): 234-244. https://doi.org/10.1145/190809.190336
[33] Raj, Jennifer S., and Abul Basar. "QoS optimization of energy efficient routing in IoT wireless sensor networks." *Journal of ISMAC* 1, no. 01 (2019): 12-23. https://doi.org/10.36548/jismac.2019.1.002
[34] Sharma, Surbhi, and Baijnath Kaushik. "A survey on internet of vehicles: Applications, security issues & solutions." *Vehicular Communications* 20 (2019): 100182. https://doi.org/10.1016/j.vehcom.2019.100182
[35] Sharma, Tripti, Archana Balyan, Rajit Nair, Paras Jain, Shivam Arora, and Fardin Ahmadi. "ReLeC: A Reinforcement Learning-Based Clustering-Enhanced Protocol for Efficient Energy Optimization in Wireless Sensor Networks." *Wireless Communications and Mobile Computing* 2022, no. 1 (2022): 3337831. https://doi.org/10.1155/2022/3337831
[36] Shayea, Ibraheem, Mustafa Ergen, Marwan Hadri Azmi, Sultan Aldirmaz Çolak, Rosdiadee Nordin, and Yousef Ibrahim Daradkeh. "Key challenges, drivers and solutions for mobility management in 5G networks: A survey." *IEEE access* 8 (2020): 172534-172552. https://doi.org/10.1109/ACCESS.2020.3023802
[37] Soundari, A. Gnana, and V. L. Jyothi. "Energy efficient machine learning technique for smart data collection in wireless sensor networks." *Circuits, Systems, and Signal Processing* 39, no. 2 (2020): 1089-1122. https://doi.org/10.1007/s00034-019-01181-3
[38] Storck, Carlos Renato, and Fátima Duarte-Figueiredo. "A survey of 5G technology evolution, standards, and infrastructure associated with vehicle-to-everything communications by internet of vehicles." *IEEE access* 8 (2020): 117593-117614. https://doi.org/10.1109/ACCESS.2020.3004779
[39] Utama, Sunariya, Israa Shakir Seger, Sajad Muhil Abd, Roshidi Din, Alaa Jabbar Qasim Almaliki, and Jabbar Qasim Almaliki. "Analytical and Empirical Insights into Wireless Sensor Network Longevity: the Role MAC Protocols and Adaptive Strategies." *Journal of Advanced Research in Computing and Applications* 36, no. 1 (2024): 52-60. https://doi.org/10.37934/arca.36.1.5260
[40] Xu, Wenchao, Haibo Zhou, Nan Cheng, Feng Lyu, Weisen Shi, Jiayin Chen, and Xuemin Shen. "Internet of vehicles in big data era." *IEEE/CAA Journal of Automatica Sinica* 5, no. 1 (2017): 19-35. https://doi.org/10.1109/JAS.2017.7510736
[41] Yang, Jidian, Shiwen He, Yang Xu, Linweiya Chen, and Ju Ren. "A trusted routing scheme using blockchain and reinforcement learning for wireless sensor networks." *Sensors* 19, no. 4 (2019): 970. https://doi.org/10.3390/s19040970
[42] Yin, Xiuwen, Jianqi Liu, Xiaochun Cheng, and Xiaoming Xiong. "Large-size data distribution in IoV based on 5G/6G compatible heterogeneous network." *IEEE Transactions on Intelligent Transportation Systems* 23, no. 7 (2021): 9840-9852. https://doi.org/10.1109/TITS.2021.3118701
[43] Yuan, Tingting, Wilson da Rocha Neto, Christian Esteve Rothenberg, Katia Obraczka, Chadi Barakat, and Thierry Turletti. "Machine learning for next-generation intelligent transportation systems: A survey." *Transactions on emerging telecommunications technologies* 33, no. 4 (2022): e4427. https://doi.org/10.1002/ett.4427
[44] Yunus, Mawarni Mohamed, Khairina Nazihah Azahar, Mas Haslinda Mohamad, and Yulindon Khaidir. "Evaluating 4G Network Performance at UTeM Campus to Facilitate 5G Implementation in Malaysia." *Semarak International Journal of Electronic System Engineering* 3, no. 1 (2024): 15-27. https://doi.org/10.37934/sijese.3.1.1527