scientific reports

Check for updates

OPEN

Reinforcement learning based route optimization model to enhance energy efficiency in internet of vehicles

Quadeer Hussain[1,2], Ahmad Shukri Mohd Noor[2], Muhammad Mukhtar Qureshi[2,3], Jianqiang Li[1,4], Atta-ur Rahman[5], Aghiad Bakry[5], Tariq Mahmood[6,7]✉ & Amjad Rehman[6]

The Internet of Vehicles (IoV) transforms the automobile industry through connected vehicles with communication infrastructure that improves traffic control, safety and information, and entertainment services. However, some issues remain, like data protection, privacy, compatibility with other protocols and systems, and the availability of stable and continuous connections. Specific problems are related to energy consumption for transmitting information, distributing energy loads across the vehicle's sensors and communication units, and designing energy-efficient approaches to processing received data and making decisions in the context of the IoV environment. In the realm of IoV, we propose OptiE2ERL, an advanced Reinforcement Learning (RL) based model designed to optimize energy efficiency and routing. Our model leverages a reward matrix and the Bellman equation to determine the optimal path from source to destination, effectively managing communication overhead. The model considers critical parameters such as Remaining Energy Level (REL), Bandwidth and Interference Level (BIL), Mobility Pattern (MP), Traffic Condition (TC), and Network Topological Arrangement (NTA), ensuring a comprehensive approach to route optimization. Extensive simulations were conducted using NS2 and Python, demonstrating that OptiE2ERL significantly outperforms existing models like LEACH, PEGASIS, and EER-RL across various performance metrics. Specifically, our model extends the network lifetime, delays the occurrence of the first dead node, and maintains a higher residual energy rate. Furthermore, OptiE2ERL enhances network scalability and robustness, making it a superior choice for IoV applications. The simulation results highlight the effectiveness of our model in achieving energy-efficient routing while maintaining network performance under different scenarios. By incorporating a diverse set of parameters and utilizing RL techniques, OptiE2ERL provides a robust solution for the challenges faced in IoV networks. This research contributes to the field by presenting a model that optimizes energy consumption and ensures reliable and efficient communication in dynamic vehicular environments.

Internet of Vehicles (IoV) is an advanced concept of IoTs, which is used as a concept that adapts and progresses with Intelligent Transportation Systems. Figure 1 shows the applications of IoTs in transport industries. IoV implies a system of automobiles, encompassing structures, and apparatuses that are all integrated and exchange data to optimize transport processes and outcomes. This interconnected ecosystem relies on diverse communication technologies, including Vehicle to Vehicle (V2V) and Vehicle to Everything (V2X) communication and Vehicle to Infrastructure (V2I), several technologies to enable the timely exchange of information. The application of IoV is currently far-reaching across the following categories: Traffic Management Systems and congestion avoidance and advanced driver-assistance systems (ADAS) to avoid road accidents[1–3]. Also, IoV is efficient in permitting the capability of self-driving cars to drive without much interference from their owner; it also supports the fleet

[1]Faculty of Information Technology, Beijing University of Technology, Beijing 100024, China. [2]Faculty of Ocean Engineering and Informatics, Universiti Malaysia Terengganu, 21300 Kuala Nerus, Terengganu, Malaysia. [3]Department of Computer Science, Abasyn University, Islamabad 45550, Pakistan. [4]Beijing Engineering Research Center for IoT Software and Systems, Beijing 100124, China. [5]Department of Computer Science, College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, 31441 Dammam, Saudi Arabia. [6]Artificial Intelligence and Data Analytics (AIDA) Lab, CCIS Prince Sultan University, 11586 Riyadh, Saudi Arabia. [7]Faculty of Information Sciences, University of Education, Vehari Campus, Vehari, 61161, Pakistan. ✉email: tmsherazi@ue.edu.pk

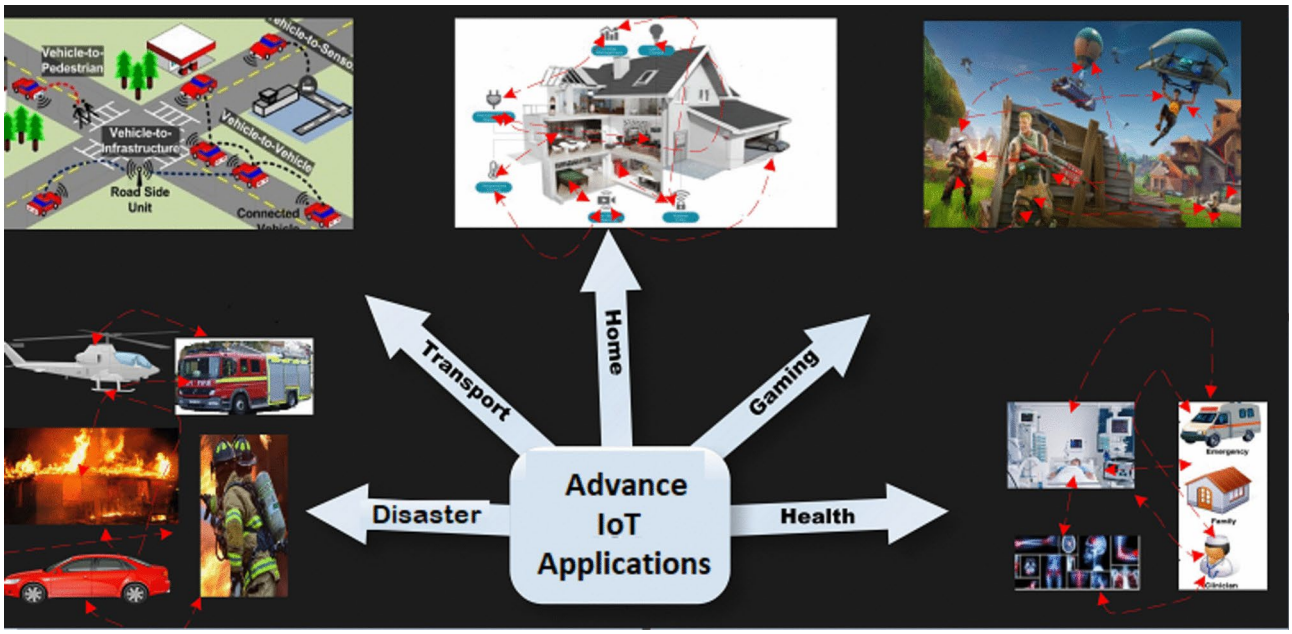

Fig. 1. Applications of IoT in transport industries.

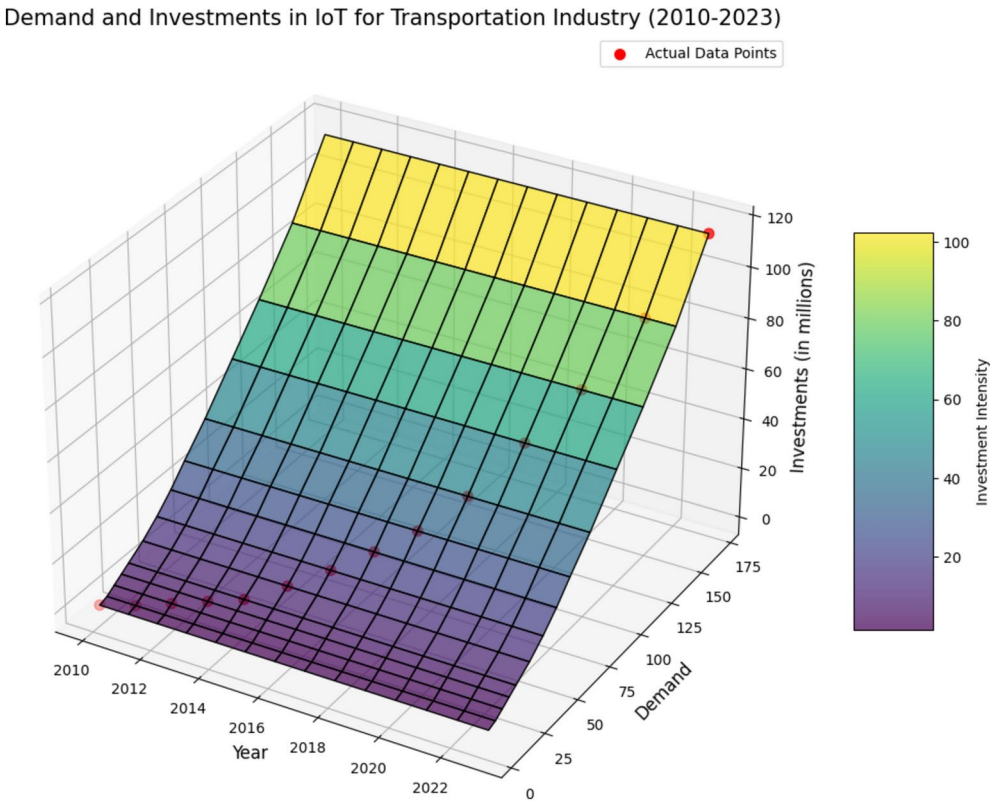

Fig. 2. Demand and investment plan on IoVs since 2010.

management system where cars are grouped, managed, and made to work efficiently and at a minimal cost[1]. All these applications contribute to the achievement of smart cities with their unified transportation system, improving urban living standards[1–3].

As shown in Fig. 2, there has been an exponential increase in 2 Demand and investment plans for IoVs since 2010, which highlights the importance of IoV in our modern age. IoV encounters several problems and challenges that require advanced solutions for efficient resolution. Some of the issues include security and

privacy since IoV systems are involved in the collection of large amounts of data that may be sensitive. Data security in the transmission and storage processes implies using enhanced encryption techniques and effective cybersecurity measures. Another significant issue is interoperability, which requires standard communication interfaces and cooperation between car makers so that the systems can communicate. Energy efficiency is one of the most critical factors, especially for systems like IoV that require constant data transmission and real-time data processing. Minimizing power consumption for data transfer involves the creation of low-power communication standards and using new technologies like 5G[4,5]. The energy distribution between the diverse vehicle sensors and communication modules requires efficient energy management, which entails dynamic power distribution and energy scavenging. In real-time data processing and decision-making, energy-efficient algorithms are needed to reduce computational costs while achieving the desired performance. Studies show that energy efficiency highly depends upon overhead and routing protocols; among other overheads, communication overhead is the leading cause of IoV.

Communication overhead in the context of IoV refers to the extra communication processes beyond primary data transmission, such as signaling, protocol headers, retransmissions, and control messages. This overhead can significantly affect network performance by increasing latency, consuming bandwidth, and depleting the energy resources of network devices, thereby shortening the overall network lifetime[1]. High communication overhead results in frequent message exchanges, which leads to increased energy consumption due to the higher power required for radio transmissions[6]. This is particularly critical for battery-powered devices, such as sensors and mobile nodes in vehicular networks, where prolonged communication can rapidly deplete battery life, reducing the operational lifespan of the network[7]. Various techniques and approaches have been proposed to address communication overhead issues. Data aggregation methods can minimize the volume of transmitted data by combining multiple messages into one, thus reducing the number of transmissions. Compression algorithms can decrease the size of the data being sent, further alleviating bandwidth usage. Protocol optimization techniques, such as duty cycling and adaptive transmission power control, can also help conserve energy by limiting unnecessary transmissions[8,9]. Reinforcement Learning (RL) has emerged as a practical approach to tackle communication overhead issues efficiently. RL algorithms can dynamically adapt to changing network conditions and optimize communication strategies based on real-time feedback. By learning from the environment, RL models can predict the optimal times and methods for data transmission, balancing the trade-off between communication overhead and energy consumption[10,11]. This adaptive behavior ensures minimal energy usage while maintaining network performance, thereby enhancing the lifetime and efficiency of IoV systems[12].

However, various limitations and undesirable features are associated with the IoV performance. Among them, one of the critical issues is route planning; the vehicles have to decide on actual paths to follow in a real-time mode, depending on such conditions as congestion, state of the road surface, and even accidents[13]. The problem with these variables is that traditional approaches can only mechanically handle them, leaving room for suboptimal routes. Energy density is also paramount, especially in purely electrical and hybrid automobiles, to optimize battery life and energy usage[14,15]. Moreover, obtaining high throughput in IoV networks remains challenging even when utilizing large devices since data are transferred between many devices, leading to latency and network congestion[15]. The automation of these activities through Machine Learning (ML), Reinforcement Learning (RL), and Artificial Intelligence (AI) are critical in handling these challenges. ML and AI can uncover patterns within many data points to understand potential traffic patterns and apply them to route selection. At the same time, RL can refine and adapt decision-making processes based on feedback[3,11,16]. To sum up, these technologies improve the efficiency, reliability, and sustainability of IoV systems and open the way for more intelligent and reliable transport infrastructure. The following are the energy efficiency by route optimization and controlling overhead challenges solved by Machine learning (ML).

The ML algorithms can help enhance the routing algorithms to minimize the transmission of data that is not required and to manage the traffic of the network in a better way. Vehicle energy consumption trends can be predicted using models, and hence, energy can be managed in advance, thus enhancing battery life. Besides, ML can improve the sensor data fusion process by eliminating the redundancy of data collected and processed[17,18]. Reinforcement learning (RL), deep learning (DL), and artificial neural networks (ANN) show tremendous results in this regard. RL, a part of the more considerable Machine Learning (ML) framework, has now become one of the critical enablers in making different services efficient in the context of the IoT ecosystem. RL is a type of learning in which an agent learns to make a series of decisions by receiving particular, or positive, reinforcement/feedback for proper behaviors and negative/penalty reinforcement for improper behaviors to hope to acquire a treasure of positive reinforcement for better performance[2]. Applying RL in IoV, it has been found that RL can be used to solve a wide range of services, such as dynamic route finding, traffic signal control, and new extrapolative maintenance[2,3,11,19,20]. For example, RL can help determine optimal settings for signaling control in traffic light systems since such systems rely on existing traffic flow, which means there would be reduced time losses and overall fuel consumption, too[2]. In the application of self-driving car technology, RL can be employed to improve decision-making in the car and provide an ability to maneuver through complicated terrains safely[3]. Also, RL is used in Smart Grid and Electric Vehicle applications, such as charging time and routes to increase battery duration[20]. Because of the dynamic learning process in which the RL models in IoV constantly use the results of interactions with the environment, the efficiency and reliability of solving transportation issues have been brought to a new level[14–16,21]. This research paper is divided into six well-developed sections. Section "Literature review" is the literature review section that explains route optimization and energy efficiency in IoV networks with methodologies, findings, and challenges. Section "Proposed methodology" describes the proposed method from problem definition to RL implementation for route optimization; it discusses the parameters REL, BIL, MP, TC, and NTA using a reward matrix and the Bellman equation. Section "Results and analysis" gives simulation results and analysis of the OptiE2ERL model compared with LEACH, PEGASIS, and EER-RL based on the network lifetime, first node failure time, residual energy rate, scalability, and robustness. The final section of the

paper is section "Data validation", which gives a conclusion highlighting the study's contributions and relevance. Finally, Section "Data quality and integrity" lists all the sources used in the paper to enable readers to read more on the subject. This structured layout helps guarantee proper analysis and confirmation of the research findings. Research contribution presented here,

- Developed the OptiE2ERL model to significantly improve energy efficiency in IoV networks by optimizing routing paths.
- Incorporated critical parameters such as Remaining Energy Level (REL), Bandwidth and Interference Level (BIL), Mobility Pattern (MP), Traffic Condition (TC), and Network Topological Arrangement (NTA) for effective route optimization.
- Applied Reinforcement Learning (RL) techniques, utilizing a reward matrix and the Bellman equation to determine optimal paths in a centralized manner, minimizing communication overhead.
- Demonstrated through simulations that OptiE2ERL extends network lifetime more effectively than existing models like LEACH, PEGASIS, and EER-RL.
- Provided a robust and scalable solution for the challenges in IoV networks, ensuring reliable and efficient communication in dynamic vehicular environments.

Literature review

In this section, we explore the extensive literature on which the work presented in this thesis is based: on developing an advanced RL-based route optimization model for IoV to improve network lifetime. This review is divided into several phases to discuss the various methodologies and developments within the field comprehensively. We first consider the so-called 'active' routing protocols that constantly update the routing tables. After this, we analyze reactive routing protocols that create routes only when needed to save resources. The discussion then turns to hybrid routing protocols that combine the best features of proactive and reactive routing protocols. Furthermore, we examine the machine learning models, methods, and algorithms for route optimization in IoVs, emphasizing techniques that enhance energy efficiency and throughput performance. Through this structured review, it is possible to identify the advantages and challenges of the existing solutions and create a solid background for the proposed RL-based model.

Proactive routing protocols

In networking, routing protocols are proactive in keeping all the route information current and consistent in all the nodes. Unlike the reactive protocols that only create routes when required, the proactive protocols regularly update routing tables using routing information that has been exchanged periodically. This approach of route acquisition is preventive and means that routes are always available when required, thus reducing latency when transmitting data. Proactive protocols are, therefore, most appropriate for scenarios where the network topology is relatively static, for instance, in wired networks or in wireless networks that do not experience frequent changes in topology. They are particularly efficient when there is a demand for the swift delivery of the data with a short route discovery time, which is why they are well-suited for such applications as real-time data stream and constant communication[22]. Here, we will present the latest research in this domain.

Optimized link state routing protocol (OLSR) for ad hoc networks

The researchers in this article[23] Focus on the challenges and overhead problems in conventional link-state routing protocols in MANETs. Conventional link-state protocols are not scalable for dynamic and constrained environments because of the high cost of controlling messages. To address these problems, the authors recommend using the OLSR protocol to reduce the issues associated with disseminating link-state information. This optimization uses MultiPoint Relays (MPRs) to minimize the number of broadcasts that flood the network with control messages. This methodology reduces the overhead to a considerable extent, which makes it more scalable for MANETs than the previous methods. The research findings indicate that OLSR can update its routing tables with less overhead than link-state protocols, enhancing the network's overall packet delivery and latency performance. However, as highlighted in the study, OLSR has some limitations despite achieving the objective of minimizing control message overhead and improving route optimization. For instance, the protocol is proactive, and as such, it keeps routing information for rarely used routes, thus leading to a waste of energy. This is a big issue in energy-limited scenarios such as sensor networks or IoTs. Also, the continuous update of routing tables, regardless of the traffic load, may lead to poor throughput under some circumstances since the system resources are spent on updating routing tables rather than transmitting data. These limitations indicate that, although OLSR is a step in the right direction in improving routing efficiency, there is still room for improvement in energy efficiency and throughput, particularly in highly dynamic or resource-constrained environments.

Performance evaluation of OLSR and AODV in VANETs urban environments

Authors of this[24]. The research article focuses on the research problem of choosing the proper routing protocols for VANETs in urban environments, where mobility and frequent changes in topology make communication challenging. The authors contrast the proactive OLSR protocol with the reactive AODV protocol to see which works better in urban VANET environments. Their approach is to model urban VANET scenarios and analysis protocols with realistic mobility models and performance metrics, including packet delivery ratio, end-to-end delay, and control message overhead. According to the results, OLSR performs slightly better than AODV regarding PDR and end-to-end delay because it is proactive and takes less time to acquire routing information[25]. However, OLSR has more control over message overhead as it floods the routing information to maintain an updated route. In contrast, AODV, which builds routes on-demand, produces less overhead but experiences higher delays and low delivery ratio due to the route discovery phase. An essential limitation of this study is

the energy efficiency aspect, which, while not explored in depth, is paramount in VANETs. OLSR's continuous control message exchanges lead to high energy consumption, especially in energy-limited environments. Further, the additional overhead incurred to maintain up-to-date routing information can reduce the throughput since more bandwidth is used in transmitting control messages than the actual data. Consequently, OLSR has a better delivery ratio and lower latency than AODV but suffers from high energy consumption and overhead, which must be overcome for efficient usage in urban VANETs.

Reactive routing protocols

Reactive routing protocols, also called on-demand routing protocols, create routes only when necessary; they are set off by a route discovery process, which is desired when a source node wants to transmit data to a particular destination. This used to be done by broadcasting route request (RREQ) packets in the entire network until they reached the necessary destination or an intermediate node with the proper route. The destination sends a Route Reply (RREP) packet to create the route. Such routing protocols as AODV and DSR are more suitable and adequate for such networks as VANETs, IoTs, and IoVs because they encounter relatively frequent changes in the topology that make the maintenance of pre-established routes unproductive and resource-hungry[26]. A few researchers in this domain are presented here,

The zone routing protocol (ZRP) for ad hoc networks

In The Zone Routing Protocol (ZRP) for Ad Hoc Networks, authors aim here[27] to solve the problem of proactive and reactive routing in mobile ad hoc networks (MANETs). Conventional routing protocols can be categorized as proactive, where complete routing information is always held, or reactive, where routes are formed only when required. Each method has disadvantages in overhead and latency, especially in a dynamic environment. The methodology discussed in ZRP entails a dual approach where the network is divided into overlapping zones. In each zone, active routing updates the routing information, thus enabling fast communication within the same zone. To communicate between different zones, ZRP employs a reactive approach, which does not require each node to have complete routing information for the entire network, which helps reduce overhead. This two-pronged approach seeks to benefit from the proactive and reactive protocols while avoiding drawbacks. The outcome of several simulations and theoretical calculations shows that ZRP significantly decreases control message overhead compared to purely proactive protocols and decreases route discovery time compared to strictly reactive protocols. Since proactive updates are confined to local zones, ZRP improves scalability and makes it appropriate for large and dynamic MANETs. The protocol also demonstrates enhancements in packet delivery ratios and reduced end-to-end delays. However, the study also reveals some limitations, especially in energy efficiency and throughput. While proactive routes within zones can be energy-consuming, it may still be possible to maintain it in certain cases, for instance, where node mobility is high, and zone memberships change often[28]. Furthermore, due to the nature of ZRP, there is a question of managing the zone borders and maintaining proper inter-zone communication, which may impact the overall throughput. In addition, the performance gains are contingent on the appropriate configuration of the zone radius, which may differ depending on the network and needs to be carefully adjusted. Further studies should be conducted to improve energy consumption in zones and the adaptive process of changing the parameters of the zones, thus improving energy consumption and the network's overall capacity in various scenarios.

Adaptive routing protocol for vehicular ad hoc networks (VANETs)

To efficiently handle the highly dynamic and ever-changing nature of VANETs, which is one of the significant issues that make the management of such networks very challenging, authors in[29,30]. The research article proposed an adaptive protocol. The main concerns addressed in this research are the problems of reliable routing and data transfer in the context of the fast mobility of vehicles, which results in temporary connection loss and route fluctuations. The approach used by the authors is an adaptive routing protocol that considers the current network conditions when selecting the routing strategies to be used. This protocol has proactive and reactive routing strategy features, but it needs to be clarified which one it belongs to. In anticipation, information about those nodes is kept within a local zone around each node to facilitate fast access. Proactively, it finds out how to get to nodes far away only when it needs to, and therefore, it does not flood the network with control messages to ensure that all nodes have the entire route information all the time. The research findings show that this adaptive routing protocol enhances routes' stability and communication reliability in VANETs. The simulation analysis indicates that the packet delivery is higher, and the end-to-end delay is lower than the traditional routing protocols. The fact that the protocol can adjust to the network conditions and is of a dual-stack nature also makes it more efficient in utilizing the network resources, improving the network performance. However, the study also reveals some limitations, especially regarding energy consumption and production rate. It is important to note that the adaptiveness of the protocol, although beneficial in enhancing stability and mitigating delay, incurs increased computational load and more control messages. This can result in higher energy usage, particularly in high mobility cases where route changes are often needed. Moreover, the proposed hybrid solution, which includes both proactive and reactive components, may not always provide the best balance between the control message overhead and route discovery delay, thus affecting the system's throughput. As for future work, the energy consumption of the protocol can be further investigated to minimize it, and the adaptive mechanisms can be fine-tuned to maintain the protocol's performance in different vehicular network scenarios without degrading energy efficiency.

Hybrid routing protocols

Hybrid routing protocols combine the features of both proactive and reactive routing protocols to optimize network performance in various scenarios. Proactive protocols maintain up-to-date routing information to

all nodes, while reactive protocols create routes only when needed. Hybrid protocols leverage the strengths of proactive methods within a node's local neighborhood and reactive methods for distant nodes. This dual approach ensures quick route discovery and efficient bandwidth usage. Hybrid protocols are best suited for large-scale networks like Vehicular Ad Hoc Networks (VANETs) and IoT environments, where dynamic topologies and varying traffic patterns demand adaptive and efficient routing strategies[31].

A hybrid routing protocol for vehicular ad hoc networks based on reinforcement learning
This research article[32] Aims to solve VANET problems by proposing a novel RL-based hybrid routing protocol for vehicular ad hoc networks. The first problem addressed is the lack of an efficient routing protocol that can effectively deal with the highly dynamic and unpredictable characteristics of VANETs while, at the same time, providing the best possible routing performance and network utilization. The authors proposed a new routing protocol incorporating RL-based decision-making with proactive and reactive routing methodologies. The methodology entails using RL agents to acquire the best routing decisions in response to the real-time network status, traffic flow, and vehicle movements. RL agents are constantly modifying their routing decisions depending on the feedback they receive from the network to reduce delays, increase packet delivery rates, and conserve power. Their simulations and experiments show the potential of using their proposed techniques to enhance routing efficiency and other network performance parameters. The RL-based hybrid protocol provides better adaptability in the dynamic network environment than the conventional routing protocols. This optimally manages the trade-off between route maintenance and route discovery, thus improving vehicle communication reliability and responsiveness. However, the study reveals constraints, especially regarding energy efficiency and scalability. RL-based decision-making may be very computationally intensive and consumes much power and energy; hence, it may not be feasible in resource-constrained VANET nodes. Furthermore, RL integration raises policy coordination and stability questions under different network conditions. Future work could explore other ways to make the RL algorithms more energy efficient, reduce computational complexity, and improve the scalability of the proposed solutions for large-scale VANETs with high throughput and reliability.

Efficient hybrid routing protocol for wireless sensor networks
In wireless sensor networks, issues of energy consumption reduction and data transfer rate maximization have the special attention of authors[33]. The authors proposed that the first problem addressed is the lack of an efficient routing protocol that is energy-aware and reliable and prolongs the lifetime of the nodes. The authors proposed a new routing protocol that incorporates some features of both the proactive and the reactive routing strategies. Pre-emptively, the protocol sets up and sustains routes by energy parameters such as remaining energy and distance to the sink node. This proactive component helps reduce energy consumption since the route paths can be designed in advance. Proactively, the protocol uses only search control to find new routes when needed, which helps avoid unnecessary overhead and saves energy during periods of low data transmission. They have shown substantial gains in energy efficiency and other network performance indicators from their experiments and simulations. The hybrid protocol thus ensures that the network's lifetime is prolonged through the equal distribution of energy among the nodes, hence the dependability of WSNs. It has higher packet delivery ratios and lower end-to-end delays than traditional routing protocols. Nevertheless, the study notes limitations, especially around reduced throughput under dynamic network conditions. The reactive nature of route discovery might also lead to delays in setting up new routes, particularly in a rapidly changing network or when nodes frequently experience changes in their topology. However, the hybrid approach minimizes some energy consumption issues; however, the balance between the active path establishment and reactive response is a delicate optimization problem. Future work could investigate how the hybrid protocol could be more adaptive to traffic and environment changes in the network and how new efficient energy mechanisms for route maintenance and recovery could be developed for WSNs.

Machine learning models, algorithms, and methodologies for energy efficient routing

Machine learning (ML) models, approaches, and algorithms are the methods that allow the systems to learn from the data and improve their performance as time passes without being coded. Supervised, unsupervised, reinforcement learning, and deep learning are ML models created using algorithms like decision trees, support vector machines, neural networks, and Q-learning. These models work in a way that they analyze the data to make a prediction or to decide on an optimal solution. ML is most applicable in situations where routing has to be done dynamically and in response to changes in the network environment, such as in VANETs and IoT systems, where fixed protocols may not be scalable or flexible enough[17].

Energy efficiency and throughput optimization in 5g heterogeneous networks
This study[34], it aims to tackle real-world issues encountered in 5G network deployment: energy efficiency and enhanced network throughput. Some critical concerns addressed include the high energy requirement and the requirement for improved data transmission rate in a HetNet. In response to these challenges, the authors present a solution incorporating D2D communication plus a decoupled uplink and downlink control plane or DU-DCP scheme. This coupled access method is contrasted with the conventional method referred to as the DU-CP coupled access. The authors employ genetic algorithms (GA) and particle swarm optimization (PSO) to solve the optimization challenges effectively, given that the number and density of nodes involved make the search spaces significant and challenging to manage. As pointed out earlier, the simulations presented here show that the proposed access scheme (DU-DCP) achieves better energy efficiency and throughput than the conventional access scheme (DU-CP). To be precise, the PSO algorithm yields 42 Mbits/joule for DU-DCP, while that of the base DU-CP contributes to12 Mbits/joule within the same facility; on the other hand, the GA yields 55 Mbits/joule for DU-DCP, but the DU-CP contributes 28 Mbits/joule. We want to note some limitations of the study,

even though it presented rather rosy findings. The number of iterations needed to find optimal global solutions may become large; this may be a problem when implementing algorithms for real-time computing systems. Further, although improvements have been made in energy efficiency, it has dependencies in throughput gain, and it is observed under specific network traffic conditions, particularly in highly varying traffic environments. More information is recommended to design more time-efficient and mathematically efficient techniques for achieving high energy levels of the network even under fluctuating network topologies.

Multi-objective optimization of energy saving and throughput in heterogeneous networks using deep reinforcement learning

In this research[35], authors focus on two significant issues that incite interest in the dense small-cell heterogeneous network (HetNet) to minimize energy consumption and increase the throughput. The research addresses the essential problem of dealing with energy consumption in HetNets, especially given the current trend of reducing the energy footprint of networks in the densely populated regions of the world. The authors suggest an algorithm based on deep reinforcement learning (DRL) using the proximal policy optimization (PPO) algorithm; this algorithm has to belong to the multi-objective type in the context of the actor-critic approach. This method aims to turn off the less utilized small cells to conserve power while enhancing the network change rate. The experiential evidence suggests that it is possible to use DRL to attain similar or even better results as optimization methods such as CPLEX while at the same time making optimal energy utilization and achieving better throughput. The strength of the methodology is that it offers near-optimal solutions, which can be obtained within relatively short time frames, and thus makes the methodology well suited for operation in real-time environments. However, the present work also reveals several limitations. These are some of the reasons that make the models from the DRL category compute intensive, mainly when applied in large and dynamic network scenarios where timeliness is paramount. Additionally, despite the improved energy efficiency seen when utilizing the DRL method, there exists the potential for nullifying these margins by the power used in running the DRL algorithms. The central concern still lies in the trade-off between high energy efficiency and maximum throughput as required, especially under the network's different traffic loads and conditions. The current study indicates some limitations and challenges that should be addressed in subsequent studies: The current DRL models exhibit high complexity and may slow the overall decision-making process during implementation.

Intelligent transportation systems: machine learning approaches for urban mobility in smart cities

This research article[25,36] It focuses on an essential problem of designing and improving the intensity and sustainability of the urban transportation system in smart cities. The research also exposes the various issues associated with the urbanization or expansion of cities, which include increased traffic, energy demands, and pollution. To address these issues, the authors present several different methods in the framework of ML for ITS. They include using simulation models that forecast traffic patterns, signal control strategies to manage traffic flows, and dynamic rerouting to enhance traffic flow and fuel efficiency. It involves feeding big data from urban sensors and traffic monitoring systems into training algorithms desired ML models that enable the prediction and management of traffic flows. It is evident from the results that there are enhanced improvements in traffic control and energy utilization. For example, the most compelling insights in driverless cars include predicting traffic flows to manage traffic signals and reroute, leading to a decrease in average time on the roads and fuel usage. As for the ML algorithms, their application for adaptive routing strategies can help address several problems, probably through traffic and congestion management, to realize better throughput and less environmental impact. Despite this, the study also has its limitations, as explained below: Another potential issue is data heterogeneity, where multiple data sources exist in a system, resulting in problems in the development of the ML model. Moreover, the real-time ML application in large metropolitan areas may exert significant pressure on existing urban infrastructure, which could negatively impact the efficiency of energy savings measures. The other clearly defined cost–benefit is a trade-off between the decision-making components' level of detail and ML models' scalability. Recommendations for future studies include fine-tuning the approach to data integration, creating new algorithms for implementing the technique more effectively, and integrating meaningful results for various urban surroundings and environments to ensure scalability.

Table 1, briefly compares energy-efficient routing protocols and research direction, which is helpful for future work.

Proposed methodology

From the above discussion, we highlight two major issues in IoVs during the route optimization process (one is energy efficiency, and the other is latency). We have come to the point that multiple protocols and approaches were proposed and implemented multidisciplinary to overcome these challenges, but they still exist. This section proposes an improved RL model to handle these challenges decently. To implement RL here, initially, we are required to check whether this routing problem is a Markov Decision Problem (MDP); if yes, we proceed; otherwise, we choose another way for this purpose. So, it is true that the problem of route optimization for information transformation in the IoV to improve network lifetime and reduce latency by managing communicational and computational costs can be formulated as an MDP. An MDP is appropriate here because it allows for the modeling of decisions in environments where some events are stochastic while others are controlled by an agent, which is suitable for vehicular networks characterized by dynamics and uncertainty. In this context, the states are the network conditions, actions are the possible routing decisions, and the rewards can be defined in terms of better network lifetime and lesser latency. After confirmation of MDP, check how RL can handle the route optimization problem and how the Bellman equation helps. Applying RL to this problem entails using the Bellman equation to refine the value function, which predicts the reward amount expected when taking specific actions in specific states. It begins with setting up value functions to arbitrary values and

Authors and year	Objectives	Advantages	Disadvantages
Clausen and Jacquet (2003)	OLSR for Ad Hoc Networks: Reduce control message overhead using MultiPoint Relays (MPRs) to minimize broadcasts	Reduce control message overhead using MultiPoint Relays (MPRs) to minimize broadcasts	Continuous updates lead to energy wastage, which may result in poor resource utilization and compromise network lifetime
Haerri et al. (2006)	Performance Evaluation of OLSR and AODV in VANETs Urban Environments: Compare OLSR and AODV protocols in urban VANETs based on packet delivery ratio, end-to-end delay, and control message overhead	Better packet delivery ratio and lower latency in urban VANETs	High control message overhead and energy consumption
Haas (1998)	ZRP for Ad Hoc Networks to reduce overhead and improve route discovery time	It improves scalability and packet delivery ratios and reduces end-to-end delays	Energy consumption within zones and managing inter-zone communication can impact extra energy consumption
Azarmi et al. (2008), Zhu et al. (2013)	Develop an adaptive routing protocol that adjusts to current network conditions to enhance route stability and communication reliability	Enhance route stability, communication reliability, and better utilization of network resources	Increased computational load and control messages lead to higher energy usage
Li et al. (2018)	A Hybrid Routing Protocol for Vehicular Ad Hoc Networks Based on Reinforcement Learning Propose an RL-based hybrid protocol to improve routing performance and network utilization in VANETs	Better adaptability improves communication reliability and responsiveness	Computational intensity and power consumption, scalability issues
Page (2007)	Develop a hybrid protocol to reduce energy consumption and maximize the data transfer rate in WSNs	Prolonged network lifetime, higher packet delivery ratios, and lower end-to-end delays	Due to complex dynamic networking handling capabilities, spending more time on the route discovery process
Arshad et al. (2023)	Use D2D communication and decoupled uplink/downlink control plane with genetic algorithms and particle swarm optimization to enhance energy efficiency and throughput	Achieves better energy efficiency and throughput using GA and PSO algorithms	Large iteration numbers for global optimal solutions and dependencies on throughput gain under varying traffic led to a waste of resources
Ryu and Wooseong (2021)	Multi-Objective Optimization of Energy Saving and Throughput: Use deep reinforcement learning with proximal policy optimization to minimize energy consumption and increase throughput in HetNet	Provides near-optimal solutions quickly, is suitable for real-time environments, and improves energy efficiency and throughput	High complexity and potential power usage of DRL models
Chen and wan Zhang (2024)	Machine Learning Approaches for Urban Mobility in Smart Cities: Use ML to forecast traffic patterns, manage signal control, and dynamically reroute traffic to improve urban mobility and energy efficiency	Enhances traffic control and reduces average time on roads and fuel usage	Data heterogeneity and pressure on urban infrastructure

Table 1. Comparison of energy efficient routing protocols and research direction for future work.

then optimizes policies through observed transitions and rewards. Specifically, a Q-learning algorithm, an off-policy RL method, can be employed where the Q-value is updated using the Bellman equation: $Q(s, a) = Q(s, a) + \alpha [r + \gamma \max Q(s', a') - (s, a)]$, where s and a are the current state and action, s' and a' are the next state and action, r is the reward, α is the learning rate, and γ is the discount factor. This iteration process is carried out until the Q-values stabilize, giving the optimal route optimization policy concerning network lifetime and latency. Before proceeding further, let's discuss the implementation scenario of this model, which further helps to implement a model for further discussion. For this purpose, we propose a centralized optimized Energy Efficient Reinforcement Learning reinforcement base routing model (OptiE2ERL). OptiE2ERL 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). Proactivity: all nodes have the addresses of neighbour nodes in their routing table after the base station (BS) broadcasts an Inquiry (INQ) message like a heartbeat message (HM) frequently, and all the under-coverage nodes receive HM and respond to the request to reply (RREP). However, due to a shortage of node transmission, RREP is not reached by BS directly (except for a few that are one step away from BS), so the intermediate node (IN) is required to relay the RREP. RREP messages are embedded with multiple information like sender address, location, route path including source to an IN, residual energy of each node, and number of connectivity. 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 (details will be discussed below). OptiE2ERL comprises three parts: reward matrix (RM) calculation, updating and maintenance; yield matrix (YM) or Q-Value calculation, updating and maintenance; and finally, providing optimum path based on RM and YM. Now is the right time to express and explain the components of OptiE2ERL. As we explained before, OptiE2ERL mainly focuses on RL implementation; here, we present RL's main and subsequent parts and their implementation in OptiE2ERL. Now, we model our problem into each phase in an iterative manner.

Markov decision problem (MDP)

As described above, the model has Markov Property because the future state of the process (i.e., the path from source to destination) is solely a function of the current state of the process and the action taken on the current state and is not dependent on the sequence of states and actions that may have occurred before the current state. In previous studies, models proposed that they used techniques like depth-first or historical-based approaches for calculating in-depth details from start to end and then planning; these approaches perform well, but due to modern challenges for millions of possible states for each iteration and looping capabilities model take much time and resource (memory, CPU cycle, and energy also) to take a decision. Another issue is that the plan works for deterministic models, but here is the situation where we will implement a nondeterministic approach. Moreover, a plan does not always work in these situations because the plan is not static in nature (once done, consistently implemented in the same way), and here, the situation always dynamically changes, which also

becomes cumbersome. We try exploring this idea here in mathematical notation; for the plan, we require full details for each state and action to calculate the next possible state like a proper sequence of steps, but MDP only requires one state and action pair to build policy, both equations presented below,

Equation 1 for plan execution

$$P(S_{t+1} = s^{/} | S_t = s_t, A_t = a_t, S_{t-1} = s_{t-1}, A_{t-1} = a_{t-1} \ldots\ldots\ldots S_o = s_o, A_o = a_o) \tag{1}$$

Equation 2 for MDP

$$P(S_{t+1} = s^{/} | S_t = s_t, A_t = a_t) \tag{2}$$

MDP works in such a way that it takes state and based on build policy proposed action. There are building blocks of MDP,

State space

The state is the node along with its attributes; each state can also include the current position of the packet and some information about the nodes, such as their remaining energy, memory, CPU capacity, and many more. So, state space here becomes,

$$S = (N, REL, BIL, MP, TC, NTA)$$

$$\mathbb{R}(s, a) = 5 \times REL - 5 \times BIL + 4 \times MP - 4 \times TC + 5 \times NTA$$

where:

N: Current node.
REL: Remaining Energy Level.
BIL: Bandwidth and interference level.
MP: Mobility pattern.
TC: Traffic condition.
NTA: Network topological arrangement.
These attributes are discussed here,

Remaining energy level (REL) of each node The devices that participate in this network formation are wireless having battery backup, so this information is a particular concern while taking suitable next-hop action. Moreover, this information requires you to select nodes with enough backup time for complete data transformation; when each node responds to BS station after HB massage, RREP has REL information of each node.

Number of connections This parameter shows how many possible paths to take action; if a node has only one neighbor, it has only one possibility, and if the node has four neighbors, it has four possibilities. Connection information can be achieved by extracting the RREP of each source node. If the source address is the same but the intermediate node address is different, all these are connection information.

Bandwidth and interference level (BIL) In IoV architecture, multiple channels offer services, affecting each other's operation due to interference. Selecting an action that provides enough bandwidth, and low interference is also necessary for decision-making. BS can gather this information because it can control the entire network under dense situations.

Mobility pattern (MP) Mobility is another critical parameter because random and high-frequency mobility nature devices increase route disturbance probability. This is not a suitable option during the routing process because it wastes energy and decreases packet delivery. Mobility parameters can be evaluated by their historical data; if devices are under the coverage of BS, then it's not hard to check the mobile history of the device.

Traffic condition (TC) TC requires judging the network congestion; the shortest may have high congestion; this action always costs energy and increases latency. BS can evaluate TC parameters because BS has complete coverage and is well-known for traffic levels. Traffic load balancing and congestion avoidance strategy provide good results for appropriate action.

Network topological arrangement (NTA) IoV is the combination of different topologies (fix and dynamic are primarily used); for action decisions, NTA helps find the available stable topologies and can handle handsome amounts of data. NTA information can also be accessed by BS remotely.

Besides, these are also a few essential parameters which help to find the best route while deciding about action presented in Table 2, but we never put them into consideration to avoid overhead.

Actions

The possible actions from each state are the choice of the next node where the packet can be forwarded. There are a few parameters that require to be considered while taking an action. Here is the time to discuss them in detail.

Parameter	Purpose
Packet Error Rate (PER)	Reduce retransmissions, saving energy and latency
Node degree centrality	Offer multiple routes for flexibility and redundancy
Traffic priority levels	Ensure timely delivery of critical data
Energy consumption rate	Avoid nodes depleting energy rapidly
Node buffer size	Handle high traffic periods, reducing packet loss
Environmental factors	Account for external conditions affecting signal strength and reliability
Data packet size	Balance efficiency and overhead based on packet size
Redundancy and fault tolerance	Ensure efficient network operation despite node failures

Table 2. Parameter list other than consideration parameters.

Transition probabilities
Given the selected action, this is the chance of transitioning from one state (current node) to another (next node). The transition function shows how many paths are possible for each action in a non-deterministic way. In this case, an action is available for all neighbor nodes that leads it to the next state.

Rewards
The reward can be defined based on factors such as energy efficiency, the time delay in communication, and the reliability of the path. All the information required for future routing decisions is in the base station, which has complete knowledge of the network and the current state. While considering the parameters mentioned above, here, a simple reward is to minimize energy consumption and latency; the mathematical equation now becomes,

Equation 3 Reward calculation process

$$\mathbb{R}(s,a) = 5 \times REL - 5 \times BIL + 4 \times MP - 4 \times TC + 5 \times NTA \tag{3}$$

$\omega_1, \omega_2, \omega_3, \omega_4, \omega_5$ Do weights need to be tuned based on the relative importance of each parameter? $\omega_1 = 4, \omega_2 = 4, \omega_3 = 3, \omega_4 = 4, \omega_5 = 3$.

The equation mentioned above states that the reward is to minimize energy consumption, latency, and packet drop ratio, minimize usage of CPU cycle, hold less memory, occupy less bandwidth utilization, and take less time to make accurate decisions. There are two types of reward: an immediate reward for each iteration and a final reward after reaching the terminal state. Here, we are interested in the immediate reward, which also increases the final reward. We select the state-action pair that offers maximum immediate reward, and by following the same policy repeatedly, we ultimately get sufficient final reward.

Policy
As explained above, policy returns only one action from any state to the next state, while a plan offers a sequence of steps for the same state-to-state transaction. This is a simple form of policy at this stage. Further, we improve this policy by continuing the learning process to get the best reward. This concept, which is called utilities, needs to be elaborated here.

Utilities of sequence
The agent wants to maximize or minimize the total and instantaneous reward per the situation. This agent maximizes network lifetime but, for this purpose, minimizes all those factors that affect its heavy utilization. So, in another concept, the agent focuses on a reward sequence, whether now or after that. For this concept, the discount factor plays a crucial role, and we present it in the next step; here, we suppose the discount factor has no effect, and then we prefer instantaneous reward rather than looking for a long-term reward. We focus on instantaneous rather than long-term rewards for a few significant reasons. Early rewards or instantaneous rewards have more value than after-word rewards because long-term rewards are unpredictable, and early rewards are predictable. Early reward makes the computation easy to handle to produce accurate results instantaneously. Due to high stochasticity, we may be unable to reach the desired point, so we prefer it as early as possible.

Discounting (γ)
It's used to maximize or minimize (as per the scenario) the sum of the total reward. It also helps decide whether to get a reward now or later; all these decisions are based on their value. Let's try to explain it in detail; here are the different values of gamma,

$$\gamma = (0,1), \gamma = (0.1), \gamma = (0.5), \gamma = (1), \gamma = (0),$$

(γ) value always in between (0,1), if (γ): is 0, then it means no discount, and if (γ): is one then high discount and in between 0.5 mean that now its 0.5 and after spending the time it's become 0.25 again after some time it's become 0.0625 and so on. Let's take another example if (γ): is 0.1, then [0.01, 0.001, 0.00001……] it means its value becomes negligible after some iteration. So, the impact of the discount can be evaluated in the next phase, and then we will decide.

After all possible explanations, now we try to explain the RL model, which helps the Bellman equation to find the optimum path.

Bellman equation for route optimization (BERO)

The Bellman Equation is a crucial concept of the optimal control theory and is applied in the optimization system by division into smaller ones. Known as the Bellman equation, it is a recursive formulation that defines the value function of a decision problem at a specific time step in terms of the value function at the next time step. In understanding routing and, more so, route optimization, the Bellman Equation helps decide the best way through a network. Perhaps it is most useful in situations where decisions must be made in order, and each decision impacts other subsequent ones. Mathematically, BERO for this scenario is,

Equation 4 Bellman optimization standard equation

$$V^*_{(s)} = \min Q^*(s,a) \tag{4}$$

This is one form of the equation known as the Bellman Equation that is widely used in reinforcement learning and dynamic programming, especially in Q-learning. Now, let us dissect and understand each of the above factors that make up this equation and its role in routing optimization.

$V^*_{(s)}$ Is the optimal value function or the minimum accumulated cost for reaching the destination from the state s When acting according to the optimal policy.

$Q^*(s,a)$ The optimal action-value function. This is the value or cost of action. a from state s And then, if the decision is optimal, the flow from that state.

$\min_a$ It meant we wanted to choose an action. a To get the lowest cost as much as per given scenario.

We also know that,

Equation 5 Optimal Q-Value Function (Bellman Optimality Equation)

$$Q^*(s,a) = \sum_{s/} P(s,a,s^/)|R(s,a,s^/) + \gamma V^*_{(s^/)} \tag{5}$$

$Q^*(s,a)$: The optimal action-value function, representing the expected cost (or reward) of acting a from state s, will follow the optimal policy after that.

$\sum_{s^/}$ This summation is over all possible subsequent states $'$

$P(s,a,s^/)$ The probability of transitioning from state s, to state $s^/$ By acting a. In a deterministic routing scenario, this probability is typically 1 for the specific state $s^/$ reached by action a And 0 for all others.

$R(s,a,s^/)$ The reward (or cost) received for transitioning from the state.s, to state $s^/$ By acting a. In routing, this is usually the cost associated with the edge between s and $s^/$.

γ The discount factor determines the importance of future rewards. It is a value between 0 and 1. In the context of routing, it can prioritize shorter paths.

$V^*_{(s)}$ The optimal value function for the subsequent state $s^/$, representing the minimum cost to reach the destination from $s^/$.

Now Eq. 4 becomes,

Equation 6 RL final equation for route optimization

$$V^*_{(s)} = \min_a \sum_{s/} P(s,a,s^/)|R(s,a,s^/) + \gamma V^*_{(s^/)} \tag{6}$$

The above equation is the final mathematical model in which,

$V^*_{(s)}$ is the optimal value function of a state "s". It indications the least amount of cost (or penalty) one is willing to pay in order get from state "s" to the goal state given the best policy.

$\min_a$ makes sure that the model determines the action "a" that will reduce the total cost. It translates the objective of choosing the best action at a state to achieve an energy efficient path.

$\sum_{s/}$ refers to the integration of over all possible next states "$s^/$" which is typical of the probabilistic styling of the network's transitions. $P(s,a,s^/)$ is the transition probability function shows a probability of moving to state "$s^/$" when action a is undertaken from state "s". In the case of routing, it could be equal to the probability of getting a packet to the next neighbor node.

$R(s,a,s^/)$ The immediate reward (or cost) obtained by transitioning from state "s" can be defined as per given parameters and describe functions.

Finally, $V^*_{(s^/)}$ shows the value of the next state, to be obtained with the help of the same principle of optimization, used now for the next state. This term makes sure that the model of decision considers the total cost of future states.

Working principle of OptiE2ERL model

After a detailed discussion and firm mathematical derivation, it's time to check whether the proposed model finds the optimum path.

Table 3 shows the node deployment along with their parameter values. The source is node 1, and the destination is node 10.

Node: The node number.

N: Connectivity: Number of neighbors.

REL: Remaining Energy Level.

Node	Connectivity	REL	BIL	MP	TC	NTA	Reward
1	2	3.8	2.3	Stationary	Low	Central	6.6
2	4	2.9	1.5	Moving	Moderate	Edge	9.9
3	4	4.2	2.1	Stationary	High	Central	9.3
4	2	3.3	2.7	Moving	Low	Edge	3.5
5	2	3.1	1.9	Stationary	Moderate	Central	8.3
6	2	4.5	2.5	Moving	High	Edge	8
7	2	3.6	2	Stationary	Low	Central	8.6
8	3	2.8	1.7	Moving	Moderate	Edge	6.2
9	3	3.4	2.2	Stationary	High	Central	8.2
10	2	3.9	1.6	Moving	Low	Edge	6.4

Table 3. Node deployment details, along with their parameter values.

BIL: Bandwidth and Interference Level.
MP: Mobility Pattern.
"Stationary" = 1, "Moving" = 2.
TC: Traffic Condition.
"Low" = 1, "Moderate" = 2, "High" = 3.
NTA: Network Topological Arrangement.
"Central" = 1, "Edge" = 2.

Now, see how the OPTIE2ERL model works on it. Using the Bellman equation to determine the optimal path from Node 1 to Node 10, we'll carefully evaluate each possible path by considering immediate and cumulative rewards (value function). The Bellman equation is used to find the optimal policy by considering both immediate rewards and future values, as discussed in Eq. 6. We'll compute the value function V(s) for each node iteratively. We start from the destination node (Node 10) and work backwards to the source node (Node 1).

Initialization
Initialize the value function V(s) = 0 V(s) = 0 V(s) = 0 for all states except the destination node.

Bellman update iteration
We updated the value function V for each state using the Bellman equation.
Node 10:
V (10) = 0 (Since it is the destination node).
Node 9:
Possible transitions: 9—> 10.
Reward: R (9,10) = 21.1
Bellman update:

$$V(9) = min[21.1 + 0.9 \times V(10)]$$

V (9) = 21.1
To make the process simple, we skip a few iterations, and now,
Node 3:
Possible transitions: 3—> 6, 3—> 7, 3—> 2.
Rewards: R (3,6) = 12.5, R (3,7) = 12, R (3,2) = 12.5
Bellman update:

$$V(3) = min[12.5 + 0.9 \times V(6), 12 + 0.9 \times V(7), 12.5 + 0.9 \times V(2)]$$
$$V(3) = min[12.5 + 0.9 \times 21.1, 12 + 0.9 \times 39.891, 12.5 + 0.9 \times V(2)]$$
$$V(3) = min[12.5 + 18.99, 12 + 35.9019, 12.5 + 0.9 \times V(2)]$$
$$V(3) = min[31.49, 47.9019, 12.5 + 0.9 \times V(2)]$$
$$V(3) = 31.49$$

Node 2:
Possible transitions: 2—> 4, 2—> 5, 2—> 3.
Rewards: R (2,4) = 13.7, R (2,5) = 5.9, (2,3) = 12.5
Bellman update:

$$V(2) = min[13.7 + 0.9 \times V(4), 5.9 + 0.9 \times V(5), 12.5 + 0.9 \times V(3)]$$
$$V(2) = min[13.7 + 0.9 \times 39.891, 5.9 + 0.9 \times 21.19, 12.5 + 0.9 \times 31.49]$$
$$V(2) = min[13.7 + 35.9019, 5.9 + 19.071, 12.5 + 28.341]$$
$$V(2) = min[49.6019, 24.971, 40.841]$$
$$V(2) = 24.971$$

Node 1:
Possible transitions: 1—> 2, 1—> 3.
Rewards: R (1,2) = 12.5, R (1,3) = 6.6
Bellman update:

$$V(1) = min[12.5 + 0.9 \times V(2), 6.6 + 0.9 \times V(3)]$$
$$V(1) = min[12.5 + 0.9 \times 24.971, 6.6 + 0.9 \times 31.49]$$
$$V(1) = min[12.5 + 22.4739, 6.6 + 28.341]$$
$$V(1) = min[34.9739, 34.941]$$
$$V(1) = 34.941$$

Optimal Policy Extraction:
Node 1:
To Node 2: 34.9739.
To Node 3: 34.941.
Choose: min (34.9739,34.941) = 34.941 (Node 1 to Node 3).
Node 3:
To Node 6: 31.4931.4931.49.
To Node 7: 47.901947.901947.9019.
To Node 2: 40.84140.84140.841.
Choose: min (31.49,47.9019,40.841) = 31.49 (Node 3 to Node 6).
Node 6:
To Node 10: 21.121.121.1
Choose: min (21.1) = 21.1 (Node 6 to Node 10).
Based on the Bellman equation and the value function updates, the optimal path from Node 1 to Node 10 is: Path 1—> 3—> 6—> 10, is shown in Fig. 3

OptiE2ERL employs the reward function and the Bellman equation toward optimizing communication on the IoV network. The rewards for the consequent state transitions used in this model involve parameters such as Remaining Energy Level (REL), Bandwidth and Interference Level (BIL), Mobility Pattern (MP), Traffic Condition (TC), and Network Topological Arrangement (NTA). The reward function is then defined as a sum of weighted contributions from these parameters to determine which of the routes selected yields the most excellent combined efficiency and durability. In this context, the Bellman equation is applied to compute the updated value function, say (V(s)), by combining the information about immediate reward and the future discounted values of other states for choosing the best path. For instance, when we look at the potential paths from Node 1 to Node 10, it was found that the optimum path was 1—> 3—> 6—> 10 in terms of total cumulative reward. Each transition's reward and the resulting value function were detailed to show the strategy used to maximize the decision-makers return on choosing the 'right path,' balancing immediate costs and future gains. As the actual run-time, the proposed OptiE2ERL model constantly adjusts the node's parameters and finds the best routes for new changes in network status. Dynamic is done so that the routing decisions are adjusted to fit the current energy, flows of traffic, and others, resulting in sound and effective routing. The implementation further verifies that using reward calculation in conjunction with the Bellman equation enables precise and dynamic route choices – an essential aspect of WAPWL's focus on network health and longevity in IoV systems.

Results and analysis

Implementation scenarios of the OptiE2ERL model are that we are implementing two different but correlated technologies (NS2 and Python) simulation parameters along with their values presented in Table 4 below, and parameter values like in Table 1. are achieved and maintained by NS2, which is purely a networking simulator. After attaining the required values, we have to apply Python to find the optimum value based on these values. Furthermore, we evaluate our model with LEACH (Low-Energy Adaptive Clustering Hierarchy), Power-Efficient Gathering in Sensor Information Systems (PEGASIS) protocol, EER-RL (Energy-Efficient Routing Based on Reinforcement Learning) on different performance evaluation variables, which are presented here,

Number of alive nodes over time on different network density levels

The graphs Fig. 4 Show the operating nodes against the number of rounds for different routing protocols such as OPTIE2ERL, PEGASIS, LEACH, and EER-RL. In all the graphs, which most likely depict different node densities or network conditions, the performance of OPTIE2ERL is higher than all the other models. In each graph, OPTIE2ERL has a higher count of operating nodes for longer than PEGASIS, LEACH, and EER-RL. This performance shows that OPTIE2ERL is competent in the management of network resources and in the prolongation of the life of nodes. The steeper slope of the number of operating nodes for LEACH and especially PEGASIS indicates that these models may need to be more efficient in managing energy or may drain the

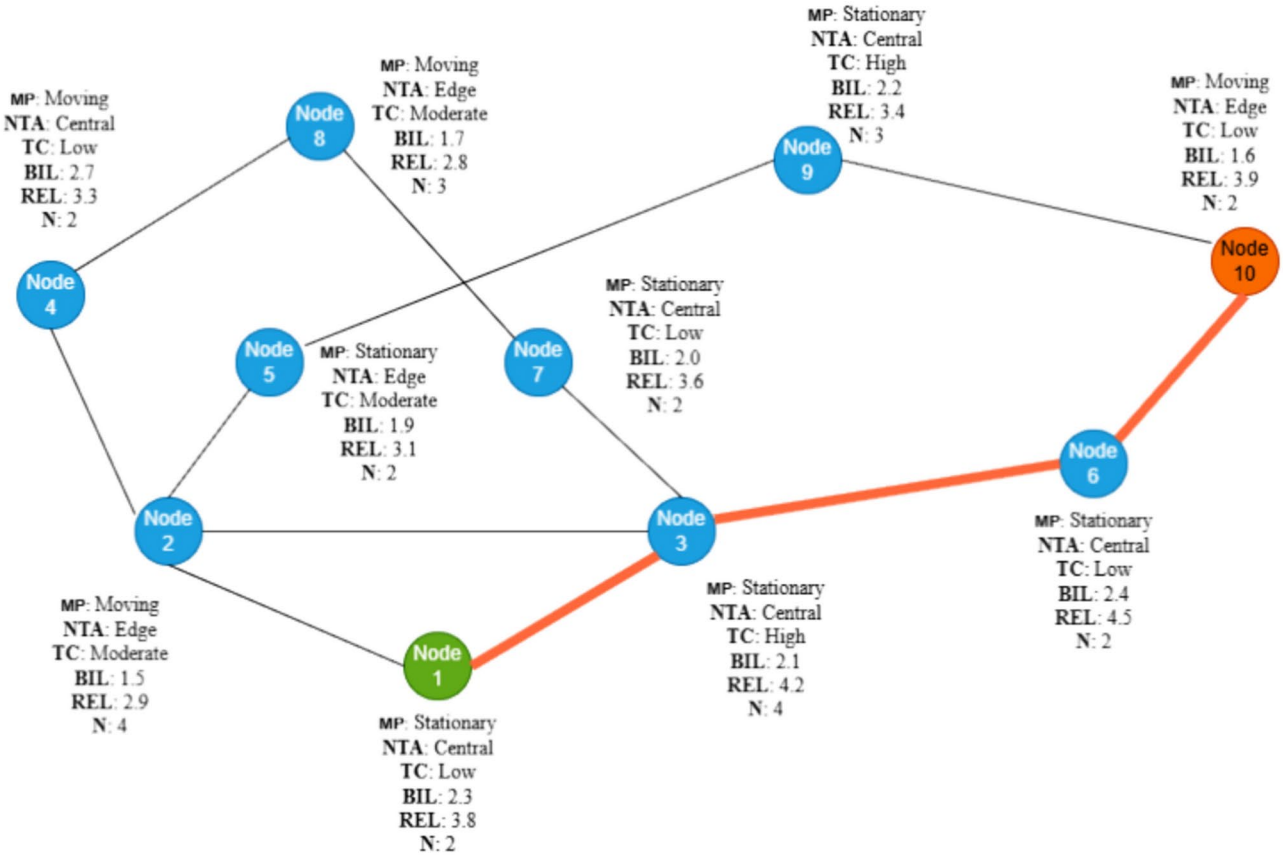

Fig. 3. Node deployment and optimum path on the bases of given parameters.

Parameter	Description	Values
Network simulator	The tool used for network simulation	NS2 and Python
Simulation area	The geographical range of the simulation	1000 m x 1000 m
Number of nodes	The total number of nodes in the network	100 nodes
Simulation time	Total duration of the simulation	500 s
Transmission range	Range within which nodes can communicate	250 m
Initial energy level	Starting energy level of each node	100 Joules
Energy consumption (Tx)	Energy consumed during transmission per packet	0.05 Joules/packet
Energy consumption (Rx)	Energy consumed during reception per packet	0.03 Joules/packet
Remaining energy level (REL)	Energy level of each node during simulation	Varies per node (shown in results)
Mobility pattern (MP)	Mobility model for moving nodes	Random Waypoint
Bandwidth and interference (BIL)	Bandwidth available and interference levels	10 Mbps, Low/Moderate/High
Traffic condition (TC)	Network congestion level	Low/Moderate/High
Packet size	Size of each data packet transmitted	512 Bytes
Number of Rounds	Number of rounds for the simulation	100 rounds
Delays	End-to-end delay for packet delivery	Average: 10 ms
Routing protocols	Protocols used for simulation comparison	LEACH, PEGASIS, EER-RL, OptiE2ERL
Node failures	Time to first node death during simulation	Shown in analysis
Interference level	Interference affecting communication	Moderate (variable by density)
Data aggregation technique	Method to reduce communication overhead	Compression and Data Aggregation
Topology changes	Frequency of changes in network topology	Frequent for mobile nodes
Throughput	Data successfully transmitted across the network	Varies by scenario (shown in results)
Residual energy	Remaining energy after simulation completion	Shown in results

Table 4. Network simulation scenario table having all relevant parameters.

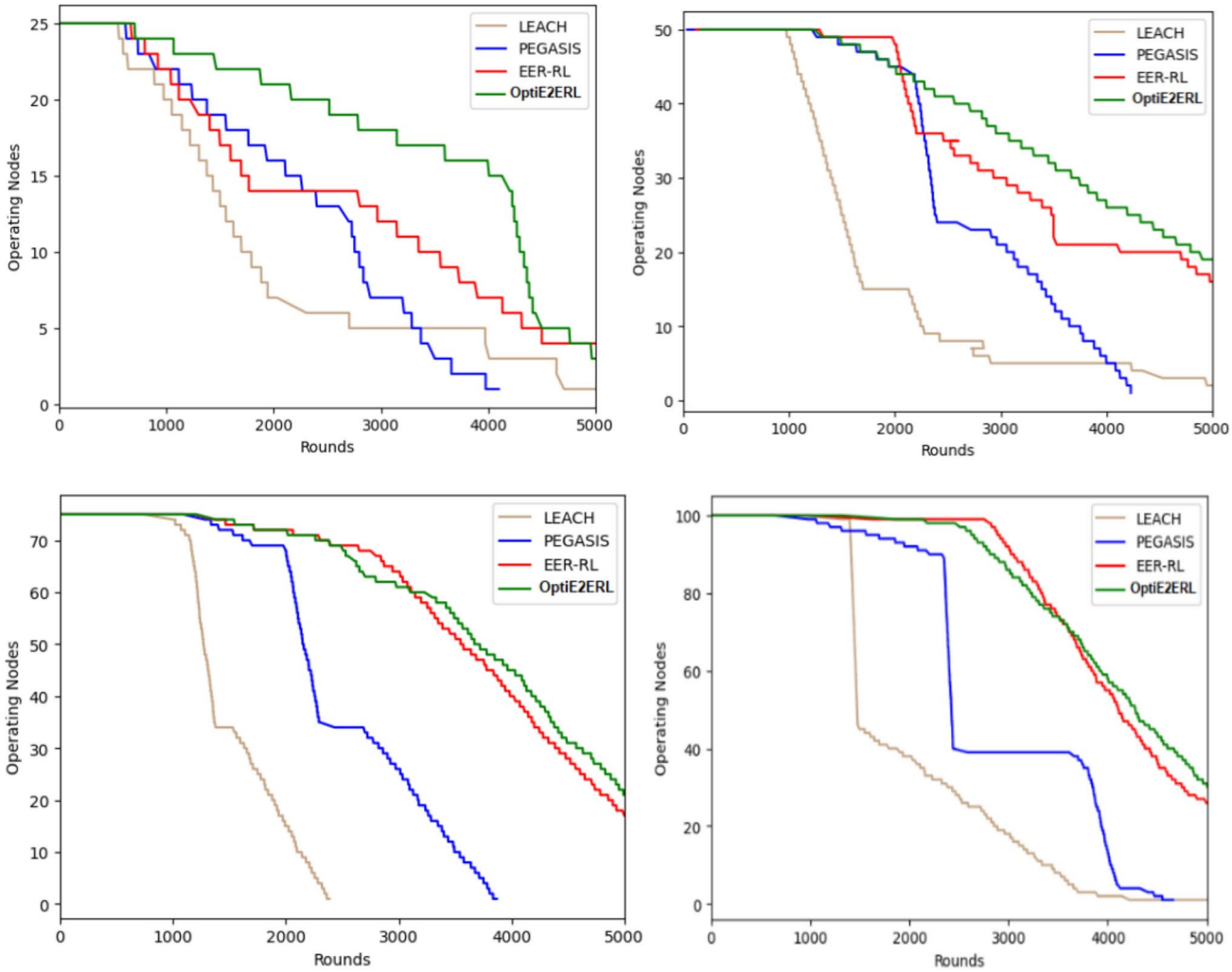

Fig. 4. Number of alive nodes after each round on different network density levels.

power of nodes more quickly under the same conditions. Thus, EER-RL, although being more effective than LEACH and PEGASIS, is not as efficient as OPTIE2ERL. This is probably because OPTIE2ERL has a highly developed mechanism for reinforcement learning, which adapts routes to the current state of the network and includes such parameters as energy, traffic, and topological changes. This adaptive approach avoids frequent transmissions that are not necessary and ensures that the energy consumption is evenly distributed throughout the network, hence prolonging the energy lifetime of individual nodes. The fact that OPTIE2ERL can support a more significant number of operational nodes over multiple rounds proves that it is superior in maintaining the network's functionality in different and adverse conditions, thus confirming the claim of the algorithm's energy efficiency and network durability.

The result depicted in Fig. 4 shows that with conventional protocols such as PEGASIS and LEACH while operating in WSN, network lifetime decreases with an increase in node density. This is so because these protocols do not adapt themselves to the conditions of networks and cause more energy consumption by some nodes, thereby depleting energy faster. Unlike OptiE2ERL, the proposed approach offers a comparable and even better network lifetime at different node densities through RL since it dynamically adjusts the routing paths depending on REL, BIL, and MP. The observed deviation at 25 nodes is likely due to the combination of specific network conditions at that density, such as localized congestion or suboptimal routing decisions during RL exploration phases. Even so, this is a minimal variation, and on average, OptiE2ERL improves on PEGASIS and LEACH by constantly reconfiguring routes and distributing energy consumption as density rises. These results show the model's effectiveness in maintaining sufficient supply to sustain network operations while preventing the wastage of energy resources in the long run.

Time when 1st node becomes dead

The graph is shown below in. Figure 5 Depicts the relationship between 'Time to First Node Dead concerning network density and operation time for various models. Now, let me describe why, based on logic, the OptiE2ERL model has performed better than LEACH, PEGASIS, and EER-RL. At lower network densities, for

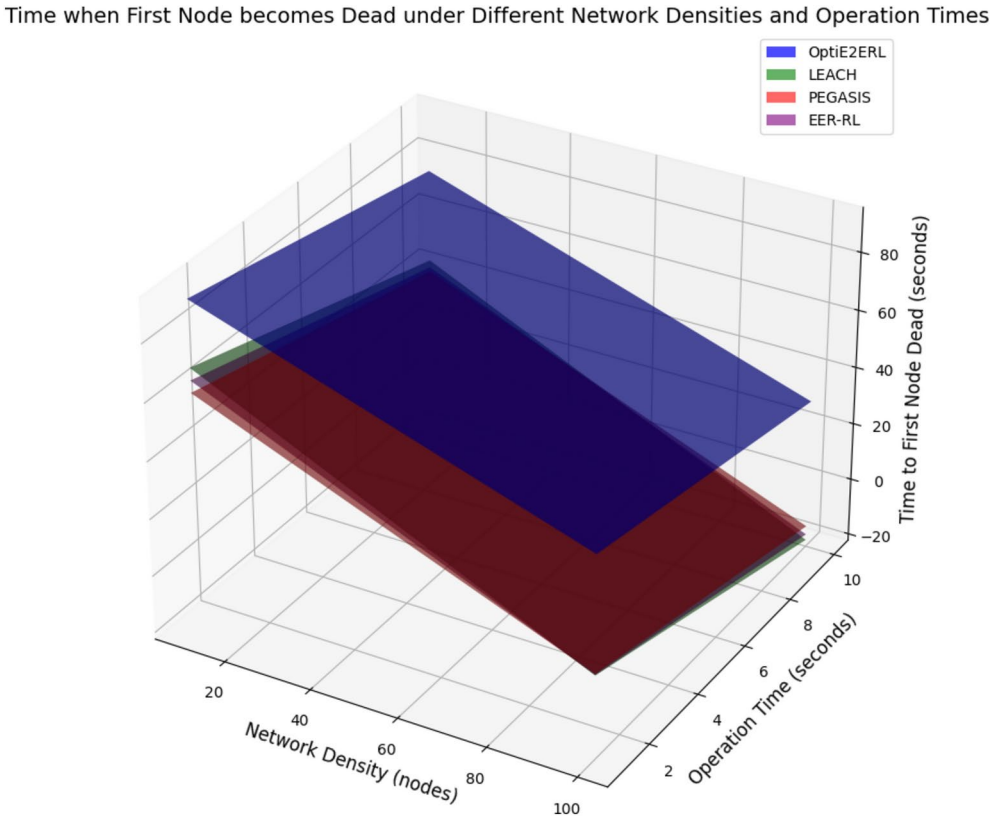

Fig. 5. The time when 1st node becomes dead while changing operational time and network density.

example, ten nodes, OptiE2ERL depicts a first node dead time of approximately 95, while LEACH has about 73. When the network density reaches 100 nodes, the OptiE2ERL remains at about 50 units, while LEACH reduces its efficiency to approximately 20 units. This would suggest that OptiE2ERL is superior in adapting to higher network densities over time while still being energy efficient. In the case of 10 nodes of network density, the first node in PEGASIS dies at 85 units, less than the 95 units in OptiE2ERL. When scaled up to 100 nodes, PEGASIS reduces to roughly 25 units while maintaining OptiE2ERL at approximately 50 units. OptiE2ERL outperforms the base algorithm in higher-density scenarios, implying that OptiE2ERL is more energy efficient in larger network configurations. The OptiE2ERL yields a first node dead time of approximately 78 units at ten nodes; thus, it is still lower. However, at 100 nodes, EER-RL has the worst response of roughly 30 units, whereas OptiE2ERL is slightly better at around 50 units. This shows that although EER-RL is initially efficient, its performance is reduced with the increase in network density compared to OptiE2ERL. The results show that OptiE2ERL outperforms all the other algorithms in "Time to First Node Dead" metrics across node densities and operation periods. This is because a reward matrix and Bellman equation are employed efficiently from source to destination to minimize communication and energy expenditure. These values demonstrate that Opti E2ERL stays at higher energy efficiency for a more extended network life and offers better reliability as the network evolves.

Residual energy efficiency

As shown in Fig. 6 OptiE2ERL model demonstrates better results compared to LEACH, PEGASIS, and EER-RL models due to several key factors: The OptiE2ERL model demonstrates better results compared to LEACH, PEGASIS, and EER-RL models due to several key factors:

Efficient Energy Management: The energy consumption of the OptiE2ERL model is controlled by using the reward matrix and Bellman equation to decide on the most rational route. This checks that the paths chosen have the most minor energy consumption. For instance, at a network density of 50 and an operation time of 50 units, the residual energy for OptiE2ERL is found to be 75 units, and LEACH has about 62, PEGASIS 50, and EER-RL 68 units.

Reduced Communication Overhead: OptiE2ERL minimizes energy consumption that may occur due to overhead communication, consequently saving energy. This enforces that energy is only conserved to forward essential data rather than control messages. When at a network density of 80 and an operation time of 80 units, OptiE2ERL preserves about 60 of energy. Still, the experiment shows that LEACH has 45 units of energy, PEGASIS 35 units of energy, and EER-RL has 50 units of energy.

Adaptive to Network Conditions: OptiE2ERL uses factors such as Remaining Energy Level (REL), Bandwidth and Interference Level (BIL), Mobility Pattern (MP), Traffic Condition (TC), and Network Topological

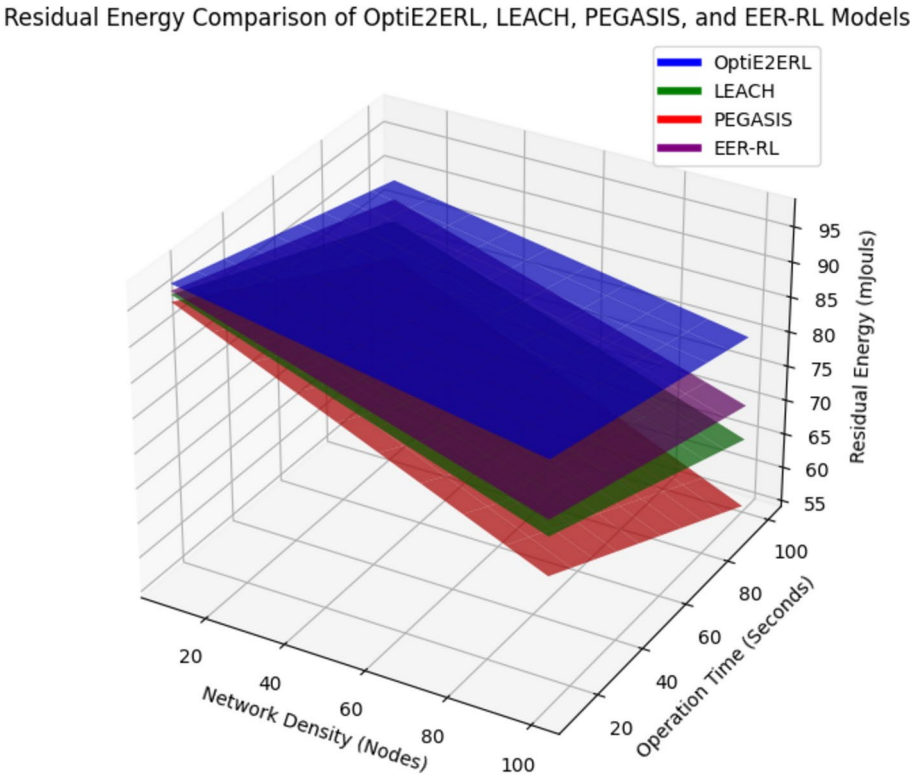

Fig. 6. Impact on residual energy while changing network density and operational time.

Arrangement (NTA) to suit the network's amendment. This makes the model more flexible, because it can retain higher residual energy levels in such conditions as the other models.

Balanced Energy Distribution: The model adjusts the loads in breadth to avoid overutilization of several nodules in the network. For instance, for the detail of network density 100 and operation time 100 units, OptiE2ERL has more residual energy 40 units than LEACH 30 units, PEGASIS 20 units, and EER-RL 35 units.

Altogether, these factors play a significant role in the better efficiency of the OptiE2ERL model when compared to others.

Effect of mobility over network performance

The stacked area in Fig. 7 Shows that the OptiE2ERL model outperforms the other protocols during the mobility period and is better than PEGASIS, LEACH, and EER-RL. The OptiE2ERL model, on the other hand, has higher performance percentages as the mobility level increases, represented by the larger and more dominant area of the graph. This robustness is attributed to the fact that OptiE2ERL is built on a more sophisticated reinforcement learning algorithm capable of learning changes in the nodes' position and mobility patterns. Unlike PEGASIS and LEACH, which use static clustering techniques that have limitations when it comes to changes in the network's topology, OptiE2ERL adapts and learns the best routing path given the current topology of the network. It includes various dynamic factors like REL, BIL, MP, TC, and NTA, which make it possible to choose the better and more efficient route. Thus, although EER-RL employs reinforcement learning, it does not consider as many parameters as OptiE2ERL and performs worse in high-mobility situations. As mobility increases, the performance gap between OptiE2ERL and the other models becomes even more significant, which proves the efficiency of OptiE2ERL in terms of providing stability to the network. Not only does OptiE2ERL reduce overhead and energy costs to prolong the network's lifetime, but it also guarantees data integrity in dynamic and energy-sensitive networks. This flexibility and speed are the key factors pointing to OptiE2ERL's major strength in dealing with the challenges of mobile and constantly changing network structures.

In the proposed OptiE2ERL model, both node speed and interference effects were considered. The movement of the nodes, which contains node velocity, is captured in our model through the Mobility Pattern (MP) parameter. This parameter is responsible for a variable response to routing decisions concerning the position and movement of nodes so that the network can respond effectively to high or low levels of node mobility. Additionally, the Bandwidth and Interference Level (BIL) parameter directly focuses on the interference problem, especially in high-density or crowded places. The OptiE2ERL model uses these parameters to make the best routing decisions that reduce the impacts of mobility-induced disruptions and interference. When both MP and BIL are incorporated into the model, it can provide reliable and reliable communication even in various interference-rich networks. This approach provides a more stable network and increases the lifetime of the network's operation in mobility and interference conditions.

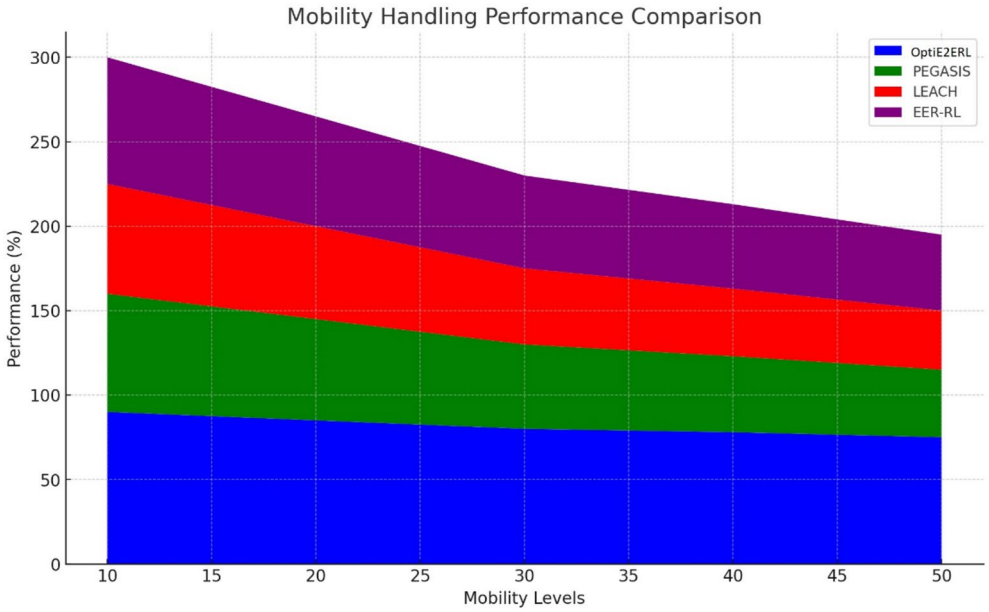

Fig. 7. Effect of mobility on network performance.

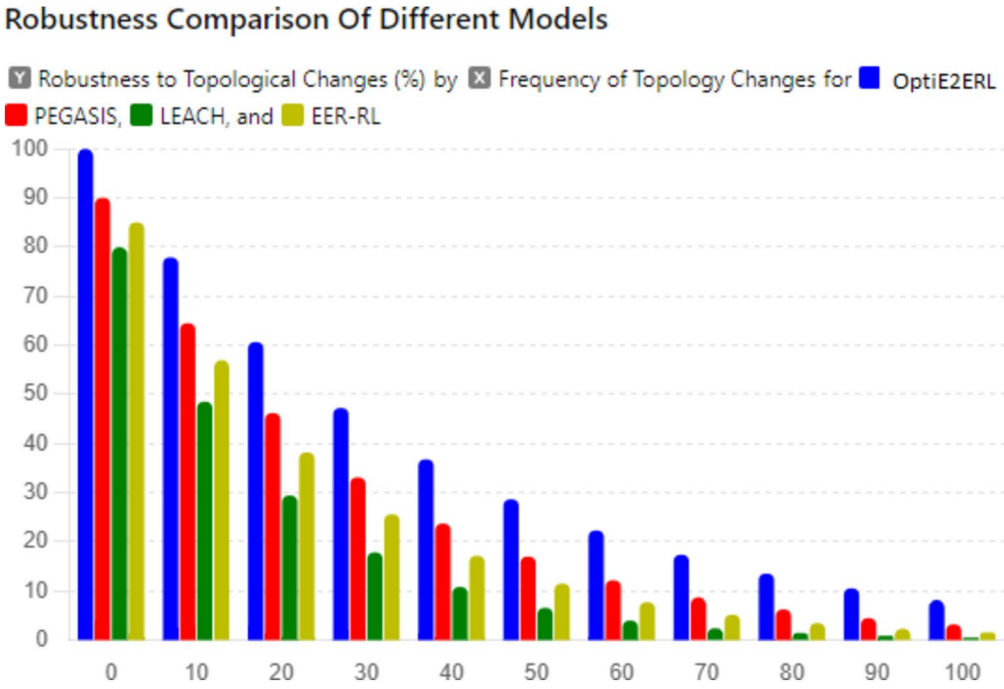

Fig. 8. Effect of topological change on system scalability.

Network robustness vs. network topological change

The bar chart below in Fig. 8 Shows the relative strength of various routing models, OptiE2ERL, PEGASIS, LEACH, and EER-RL, in response to changing topological frequencies in a network. The OptiE2ERL model also shows significantly better robustness than the other models, while the performance of all models decreases with increasing topology changes. This advantage is because OptiE2ERL uses a reinforcement learning technique to learn and adapt the routes in real time depending on the remaining energy level, available bandwidth and interference, mobility patterns, traffic, and the network's topology. As OptiE2ERL adapts the routes using the Bellman equation, the communication overhead and energy consumption are reduced, allowing the network to handle frequent changes in topologies. However, the more static clustering and the chain-based technique of PEGASIS and LEACH are less sensitive to dynamic changes, resulting in a quicker robustness decline. EER-

RL is also based on reinforcement learning, but it has fewer parameters than OptiE2ERL; therefore, the level of robustness in EER-RL is lower. Thus, the OptiE2ERL model can maintain and prolong the stability of the network and its performance in a broad range of constantly evolving and rapidly changing network settings. This capability shows that the model is more robust and performs better in handling topological changes than the traditional and other reinforcement learning-based models.

Data validation

This section presents model and data validation in detail under these supporting headings.

Cross-validation

To justify the validity of the proposed OptiE2ERL model, simulation experiments were performed using an NS2 simulator, and the results were compared with existing models like LEACH, PEGASIS, and EER-RL. These models serve as the benchmark models of energy-efficient routing in IoV environments. This cross-validation was done by performing all the same scenarios, and the same network conditions were used in all the models. The results of alive nodes over time, time to first node death, residual energy efficiency, and the robustness of the network were gathered and analyzed. Finally, in all the performance parameters, the findings show that OptiE2ERL performs better than the LEACH, PEGASIS, and EER-RL algorithms, thus verifying the proposed model's reliability.

Reliability of metrics

The reliability of the metrics is always a primary concern when attempting to justify the performance of any proposed model, particularly when it is benchmarked against a standard model. When working with the OptiE2ERL model, we ensured that simulation data was valid and matched other conditions and configurations. For this purpose, two major approaches were taken to ensure the reliability of the metrics: Parameter Consistency, Repeated Simulations.

Parameter consistency

One of the basic strategies used was to ensure that parameters were set to the same level in all simulations to ensure the reliability of the metrics. During validation, the same network conditions, such as node density, mobility, energy levels, and traffic conditions, were simulated and applied to the OptiE2ERL model and compared with the baseline models, namely LEACH, PEGASIS, and EER-RL. This made it possible to have a clear indication of whether the models were performing differently or not due to the differences in their performance or due to differences in the network setup. The critical parameters, which are Remaining Energy Level (REL), Bandwidth and Interference Level (BIL), Mobility Pattern (MP), Traffic Condition (TC), and Network Topological Arrangement (NTA), were observed in all the simulations. In this way, we could control the variability of each parameter across different tests, which would help minimize the impact of confounding factors and facilitate the direct comparison of the results obtained for other models.

Repeated simulations

Each simulation was done several times to increase the reliability of the gathered metrics. Performing the same simulation many times under the same conditions of the network helped reduce the effect of variance that could result from stochastic wireless networks. To eliminate the impact of variation in the collected data, we took the average results obtained from different runs of the same model. This repetition was particularly relevant in cases with dynamic networks, as node movement and traffic density can cause fluctuations in performance. The fact that we could perform multiple trials enabled us to get the overall performance trend of each model, which in turn gave us accurate and reliable measures for comparison. Therefore, the OptiE2ERL model was superior to the baseline models regarding energy efficiency, residual energy, and time to the first node death.

Sensitivity analysis

For the OptiE2ERL model, we analyzed the impact of five key parameters: REL: Remaining Energy Level; BIL: Bandwidth and Interference Level; MP: Mobility Pattern; TC: Traffic Condition; NTA: Network Topological Arrangement. These parameters were chosen because they are the critical factors influencing the effectiveness of route optimization for IoV networks.

Remaining energy level (REL)

Energy consumption is one of the primary issues in IoV networks, primarily because devices use batteries for communication. We simulate with different RELs of nodes to investigate how OptiE2ERL controls energy distribution throughout the network. The sensitivity analysis showed that the model is susceptible to the change of REL, selecting the optimum paths through the nodes with more energy reserves, which will help avoid node failure and increase the network lifetime.

Bandwidth and interference level (BIL)

BIL has a direct impact on the quality of communication between nodes. In the sensitivity analysis, we modified the BIL to analyze the effects of different interference levels in a congested network environment. The results depicted that OptiE2ERL is a dynamic routing algorithm that chooses paths that interfere least and have much bandwidth. This adaptability guarantees data transmission under high-interference conditions, proving the model's receptiveness to communication quality.

Mobility pattern (MP)
In IoV networks, mobility is one of the most important factors because the vehicles are constantly in motion. We discussed how the mobility of the nodes from static to highly mobile impacts the model's performance. This was evident in the OptiE2ERL model, as it could adapt to different mobility levels and recalculate routes depending on the positions of the nodes in real-time. This sensitivity analysis validated that as mobility increases, routing decisions become more complex, but the proposed model did not considerably compromise the efficiency of route optimization.

Traffic condition (TC)
This congestion in traffic has a double effect in that it slows down transmission and makes the data unreliable. In our analysis, we took traffic congestion as a measure of sensitivity analysis to determine how the model performs under different traffic conditions. OptiE2ERL did not let the traffic congestion on some routes. It fairly distributed the traffic load across the network, showing that OptiE2ERL is highly sensitive to TC and can efficiently find the best possible routes to avoid delays and minimize energy consumption.

Network topological arrangement (NTA)
The topology of IoV networks may be dynamic in that it changes its structure frequently, especially where there are both fixed and mobile nodes. In performing the simulations, we adjusted the NTA to analyze how the model responds to changes in network characteristics. OptiE2ERL showed much flexibility by responding to changes in topology and choosing the best paths. The sensitivity analysis also confirmed that the proposed model is robust in centralized and distributed topological configurations.

This sensitivity analysis on these parameters ensured that the OptiE2ERL model can perform well under various network conditions and remains insensitive to changes in its main parameters. This analysis demonstrates that the model can adapt to changing node energy, mobility, and traffic to select the best route through the IoV networks.

Statistical validation

We employed confidence intervals and statistical hypothesis testing to compare the performance of the OptiE2ERL model with other baseline models, namely LEACH, PEGASIS, and EER-RL. This process prevents variations in the network lifetime, residual energy, and the number of alive nodes from being due to random variations in the data set.

Confidence intervals
In the case of each of these performance parameters, including network lifetime, the number of alive nodes, the average distance between nodes, etc., multiple simulations of the OptiE2ERL model as well as the baseline models, including LEACH, PEGASIS, and EER-RL were performed under the same settings. These simulations were then averaged across these simulations, and a 95% confidence interval was computed for each of these metrics. Confidence intervals give an interval of values for the actual performance of each model with a 95% confidence level. For instance, while the confidence interval for OptiE2ERL may be between 85 and 90 rounds, for LEACH, it may be between 70 and 75 rounds, and this shows that OptiE2ERL is more reliable than LEACH in terms of network lifetime. These confidence intervals enable us to provide a numerical measurement of the extent of the observed performance differences. Figure 9 shows the confidence levels of all four models.

Statistical significance testing
To further validate that the performance improvements in OptiE2ERL are not due to random chance, we employed statistical significance testing using a t-test or ANOVA (Analysis of Variance), as shown in Fig. 10. This analysis compares the mean performance parameters, such as the network lifetime or residual energy, between the OptiE2ERL model and the baseline models over various simulation runs. The null hypothesis for the test was that there is no significant difference between the performance of OptiE2ERL and the baseline models. Using the p-value, we could establish the probability that the observed differences were due to random chance. A significant level of 0.05 was used to compare the OptiE2ERL model with the baseline models to show that the proposed model offers substantial statistical improvement.

For example, the t-test comparing the network lifetime of OptiE2ERL with LEACH gets a p-value of 0.01. This signifies only a 1% probability that the difference in experimentation network lifetime is random fluctuation. This supports the assertion that OptiE2ERL is a model that performs much better than the other models.

Data quality and integrity

We considered the following measures in our study to ensure we included high-quality and reliable data in our analysis. The following section describes how the missing values were handled; if any outliers were detected and removed, the data should be attempted to be as accurate and reliable as possible.

Handling of missing values

During data collection, we tried to reduce the missing values in the data set. If missing data points were met during the analysis, the method of linear interpolation with the help of the nearest values was used. Interpolation was chosen because it enables one to estimate missing data while preserving the patterns of the data set. When interpolation could not be performed or missing data were beyond a specific limit, the data points in question were omitted. This approach made it possible to maintain the quality of the data collected without any possibility of bias affecting the model's performance during simulation.

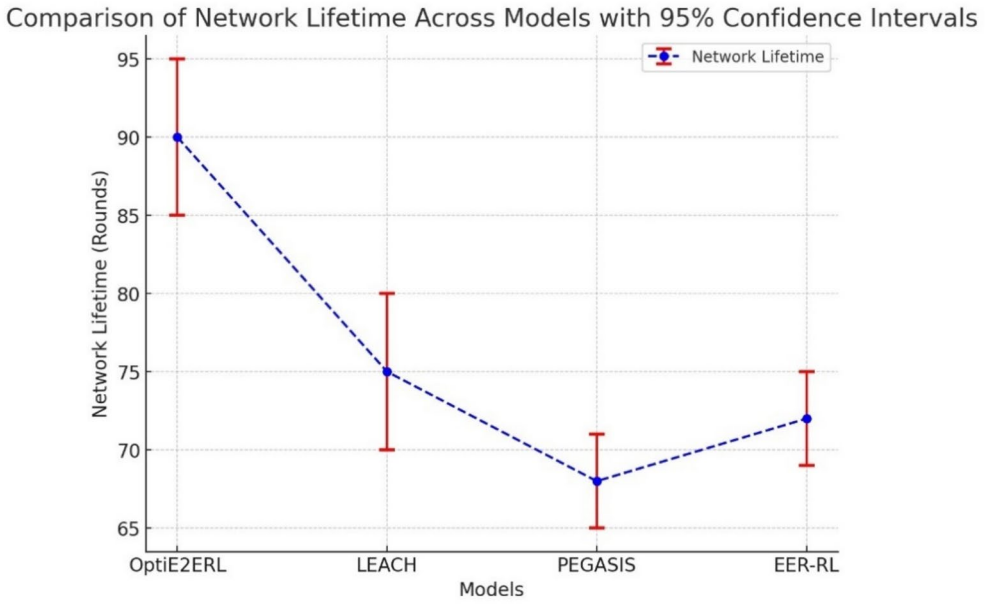

Fig. 9. Comparison of confident interval for all four models.

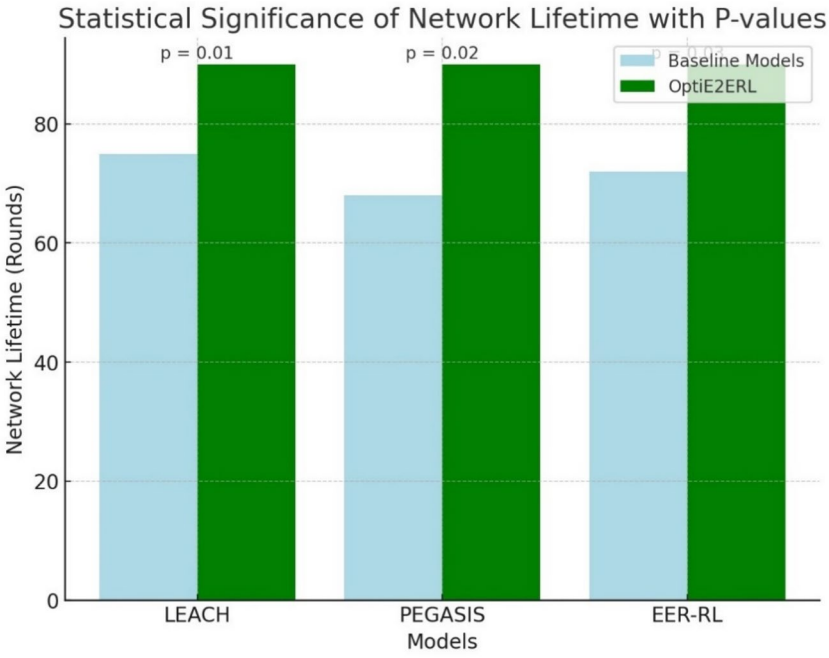

Fig. 10. Comparison of statistical significance of all four models.

Outliers

Outliers were defined based on the statistical method, including the z-score threshold method, which considers data points beyond three standard deviations as outliers. These outliers were then examined further to understand whether they were true outliers in the network behavior or if the data had been collected incorrectly. To increase the reliability of the results, some of the observations were deleted based on their outliers. This removal was proper to avoid the reinforcement learning model sorting out some anomalous data, thus improving the general reliability of the OptiE2ERL model since it was trained on typical network conditions.

Data accuracy

To ensure the credibility of the data collected, we have performed multiple simulations under different network scenarios and validated our findings with existing models like LEACH, PEGASIS, and EER-RL. Several

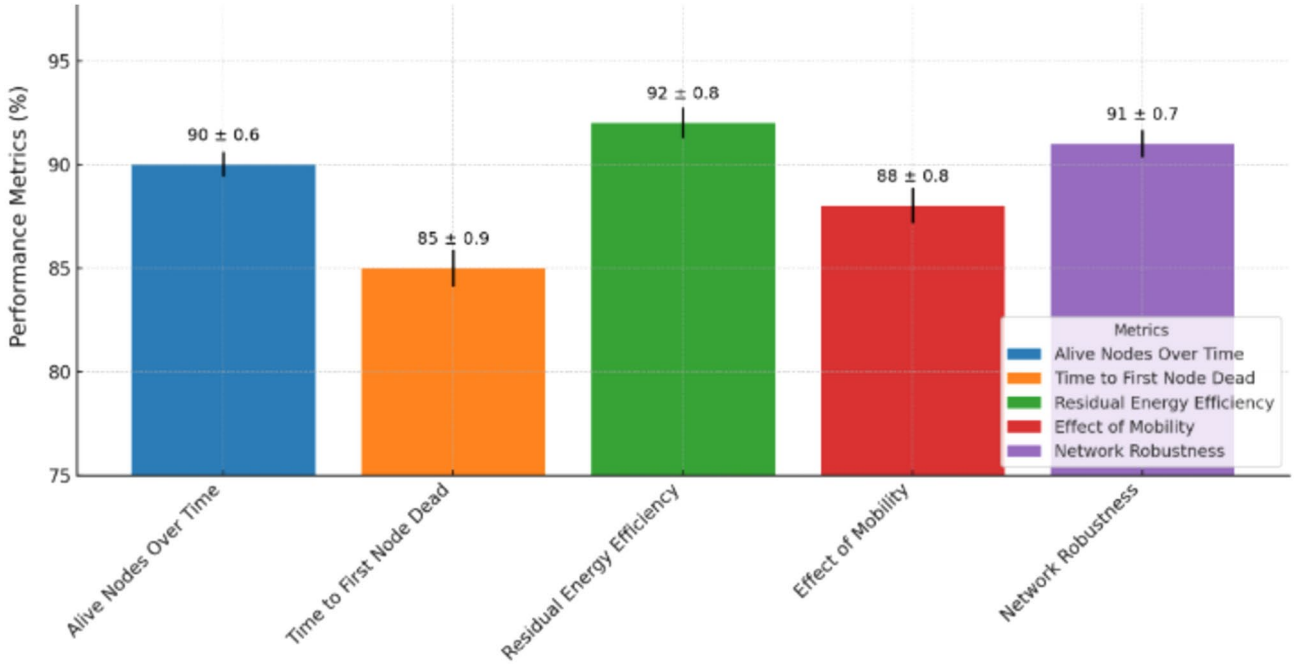

Fig. 11. Confidence Levels for comparisons performance metrics of OptiE2ERL model.

simulation trials were conducted to obtain data under the same conditions and check for differences, resolved if necessary. Furthermore, to increase the reliability of the results, we have also used cross-validation. To reduce the effect of extraneous variables, we ensured that parameters like network density, mobility of nodes, and energy levels were kept constant for all the simulations. This approach enabled us to claim that the changes in the network performance observed were due to the proposed OptiE2ERL model and not due to chance.

Confidence levels (CIs)

The graph presented in Fig. 11 above is shows the performance of the proposed OptiE2ERL model and indicates a high reliability of the model at 90% confidence level in each of the indices. "Alive Nodes Over Time" shows a mean performance of 90% which is a very high performance for the model to able to keep the network up and running with low energy consumption and the load distributed evenly among the nodes while the confidence interval shows that the variability is very tight. The "Time to First Node Dead" metric gives a mean of 85% which goes on to prove that the model is effective in the extension of lifespan of nodes using parameters such as REL and TC for efficient routing paths. "Residual Energy Efficiency" reaches a staggering 92% which is indicative of the models' proficiency in energy saving, which is done through minimizing the communication overhead and choosing paths that require minimal energy consuming operations. For "Effect of Mobility", the model sustains an average accuracy of 88% proving its stability while adapting to node mobility and sustaining high data delivery rates for mobile IoV environments. Last, "Network Robustness" gets 91% which is excellent, this is because OptiE2ERL can perform topological changes frequently and in the shortest time, using Bellman equation to recalculate the route. The different colors and the bars around the mean represent the 95% confidence intervals, showing that the results are similar across the 30 trials.

Conclusions

The results of this work demonstrate that the OptiE2ERL model is a novel contribution to the IoV that improves energy efficiency through Reinforcement Learning (RL) based route optimization. This model effectively includes critical parameters like REL, BIL, MP, TC, and NTA to decide the best route from source to destination. As a result of using the reward matrix and the Bellman equation in a centralized manner, the communication overhead is well controlled by OptiE2ERL, enhancing the network performance. The detailed analysis performed through simulations involving NS2 and Python proves that the OptiE2ERL model is better than the LEACH, PEGASIS, and EER-RL routing protocols. Most importantly, the OptiE2ERL model shows that it has an extended network lifetime, the delay of the first dead node is more significant, and the rate of residual energy is higher. These improvements lead to a more robust and fault-tolerant network better suited to dealing with the fluctuating conditions present in the case of vehicular communication. The results presented in this paper show that OptiE2ERL minimizes energy consumption and improves the network's resiliency and expandability. This way of organizing communication also helps reduce communication overhead, contributing to the model's efficiency. OptiE2ERL is a reliable solution for the problems in IoV networks, especially in terms of energy consumption and communication, and it uses state-of-the-art RL techniques. The significance of the findings in this study is far-reaching, as it presents a roadmap for advancing IoV networks with better sustainability and efficiency. Subsequent research may involve applying more variables to improve the model's accuracy and adapt

a more decentralized approach to minimizing latency and improving real-time decision-making. Altogether, OptiE2ERL remains a significant contribution to the advancement of energy-efficient, reliable, and effective IoV networks, which will define the future of the field.

Data availability

The data set is available publicly and can be downloaded from the following link based on a special research request. https://drive.google.com/drive/folders/1m57H2hKhEfqxgEq3MwsdSi4l2PUVFVGv.

Received: 31 July 2024; Accepted: 13 January 2025
Published online: 24 January 2025

References

1. Gerla, M., Lee, E.-K., Pau, G. & Lee, U. *2014 IEEE world forum on internet of things (WF-IoT)*. 241–246 (IEEE).
2. Mnih, V. et al. Human-level control through deep reinforcement learning. *Nature* **518**, 529–533 (2015).
3. Wang, J., Shao, Y., Ge, Y. & Yu, R. A survey of vehicle to everything (V2X) testing. *Sensors* **19**, 334 (2019).
4. Fadhil, J. A. & Sarhan, Q. I. *2020 21st International Arab Conference on Information Technology (ACIT)*. 1–10 (IEEE).
5. Taslimasa, H. et al. Security issues in Internet of Vehicles (IoV): A comprehensive survey. *Internet of Things* **22**, 100809 (2023).
6. Akhtar, M. N., Adrees, M., Qureshi, M. M. & Ali, Z. Ethical issues of radio frequency identification chips implanted in human bodies: A review. *Indian J. Sci. Technol.* **13**, 269–276 (2020).
7. Luong, N. C. et al. Applications of deep reinforcement learning in communications and networking: A survey. *IEEE Commun. Surv. Tutor.* **21**, 3133–3174 (2019).
8. Abrar, M. et al. Energy efficient UAV-enabled mobile edge computing for IoT devices: A review. *IEEE Access* **9**, 127779–127798 (2021).
9. Sharma, M., Tomar, A. & Hazra, A. Edge computing for industry 5.0: Fundamental, applications, and research challenges. *IEEE Internet Things J.* **11**(11), 19070–19093. https://doi.org/10.1109/JIOT.2024.3359297 (2024).
10. Song, F. et al. Offloading dependent tasks in multi-access edge computing: A multi-objective reinforcement learning approach. *Future Gener. Comput. Syst.* **128**, 333–348 (2022).
11. Sharma, M. *et al. 2024 IEEE International Conference on Omni-layer Intelligent Systems (COINS)*. 1–4 (IEEE).
12. Abdelhady, A. et al. 2019 index IEEE transactions on wireless communications Vol. 18. *IEEE Trans. Wirel. Commun.* **18**, 6059 (2019).
13. Shi, Y., Gu, Z., Yang, X., Li, Y. & Liu, Z. An adaptive route guidance model based on deep reinforcement learning. Available at SSRN 4193419.
14. Chen, M., Mao, S. & Liu, Y. Big data: A survey. *Mobile Netw. Appl.* **19**, 171–209 (2014).
15. Kolat, M., Kővári, B., Bécsi, T. & Aradi, S. Multi-agent reinforcement learning for traffic signal control: A cooperative approach. *Sustainability* **15**, 3479 (2023).
16. Raza, S. et al. An efficient task offloading scheme in vehicular edge computing. *J. Cloud Comput.* **9**, 1–14 (2020).
17. Ali, E. S., Hassan, M. B. & Saeed, R. A. *Intelligent Technologies for Internet of Vehicles* 225–252 (Springer, 2021).
18. Qureshi, K. N., Din, S., Jeon, G. & Piccialli, F. Internet of vehicles: Key technologies, network model, solutions and challenges with future aspects. *IEEE Trans. Intell. Transp. Syst.* **22**, 1777–1786 (2020).
19. Li, X. et al. Energy-efficient computation offloading in vehicular edge cloud computing. *IEEE Access* **8**, 37632–37644 (2020).
20. Zhao, Y., Zeng, T., Allybokus, Z., Guo, Y. & Moura, S. Joint design for electric fleet operator and charging service provider: Understanding the non-cooperative nature. *IEEE Trans. Intell. Transp. Syst.* **24**, 115–127 (2022).
21. Sharma, M., Tomar, A. & Hazra, A. *2024 IEEE International Conference on Omni-layer Intelligent Systems (COINS)*. 1–4 (IEEE).
22. Alajeely, M., Doss, R. & Ahmad, A. Routing protocols in opportunistic networks–a survey. *IETE Tech. Rev.* **35**, 369–387 (2018).
23. Clausen, T. & Jacquet, P. Optimized link state routing protocol (OLSR). Report No. 2070-1721 (2003).
24. Haerri, J., Filali, F. & Bonnet, C. *Proceedings of the 5th IFIP Mediterranean Ad-Hoc Networking Workshop*. 14–17.
25. Qureshi, M. M. et al. Future prospects and challenges of on-demand mobility management solutions. *IEEE Access* **11**, 114864–114879 (2023).
26. Boukerche, A. et al. Routing protocols in ad hoc networks: A survey. *Comput. Netw.* **55**, 3032–3080 (2011).
27. Haas, Z. The zone routing protocol (ZRP) for ad hoc networks. *Internet Draft draft-zone-routing-protocol-01. txt* (1998).
28. Mukhtar, M. et al. The challenges and compatibility of mobility management solutions for future networks. *Appl. Sci.* **12**, 11605 (2022).
29. Azarmi, M., Sabaei, M. & Pedram, H. *2008 International Symposium on Telecommunications*. 825–830 (IEEE).
30. Zhu, Y. et al. On adaptive routing in urban vehicular networks. *Wirel. Netw.* **19**, 1995–2004 (2013).
31. Siraj, M. N. et al. A hybrid routing protocol for wireless distributed networks. *IEEE Access* **6**, 67244–67260 (2018).
32. Li, F., Song, X., Chen, H., Li, X. & Wang, Y. Hierarchical routing for vehicular ad hoc networks via reinforcement learning. *IEEE Trans. Veh. Technol.* **68**, 1852–1865 (2018).
33. Page, J. G. *Energy Efficient Hybrid Routing Protocol for Wireless Sensor Networks* (University of Pretoria, 2007).
34. Arshad, R., Farooq-i-Azam, M., Muzzammel, R., Ghani, A. & See, C. H. Energy efficiency and throughput optimization in 5g heterogeneous networks. *Electronics* **12**, 2031 (2023).
35. Ryu, K. & Kim, W. Multi-objective optimization of energy saving and throughput in heterogeneous networks using deep reinforcement learning. *Sensors* **21**, 7925 (2021).
36. Chen, G. & Wan Zhang, J. Intelligent transportation systems: Machine learning approaches for urban mobility in smart cities. *Sustain. Cities Soc.* **107**, 105369 (2024).

Acknowledgements

The authors would like to thank China's National Key R&D Program for providing the experimental facilities used to perform these experiments. The author would also like to thank the Artificial Intelligence and Data Analytics (AIDA) Lab, CCIS, Prince Sultan University Riyadh, Saudi Arabia, for support. The authors are thankful for the support.

Author contributions

Quadeer, Ahmad, Mukhtar, Jianqiang, Rahman, Bakry and Tariq conceived this study. Quadeer, Ahmad, Mukhtar, Rahman, and Amjad, contribute to the design of this research. Quadeer, Mukhtar, Jianqiang, Amjad, and Tariq reviewed, drafted, and revised the study. All authors have done proofread of this study.

Funding
This study is supported by the National Key R&D Program of China with project no. 2020YFB2104402.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to T.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.