Multimedia Tools and Applications (2024) 83:72321–72356
https://doi.org/10.1007/s11042-024-18137-8

Vehicular communication using federated learning empowered chimp optimization (FLECO) algorithm

Ruchi Gupta[1] · **Tanweer Alam**[2]

Received: 13 April 2023 / Revised: 15 October 2023 / Accepted: 1 January 2024 /
Published online: 9 February 2024

Abstract

Recently in the field of vehicular communication, there has been a concentration of research on the integration of a vehicle-to-vehicle (V2V) network. With vehicle-to-vehicle (V2V) communication, users can directly exchange significant information with nearby vehicles. Typically, automobiles tend to travel at higher speeds on highways compared to roads at intersections. As a result, it is necessary to have a reliable system in place that can effectively and securely facilitate communication. In recent times, scientists have developed different methods for distributing information. However, these systems have various issues such as latency, reliability, mobility, and communication cost. Consequently, this results in a lack of dependability for real-time communication. Therefore, this study introduces a novel approach to Federated Learning (FL) by including the Chimp Optimization Algorithm (ChOA). Federated Learning is an approach in the field of machine learning that enables multiple devices or nodes to collaboratively train a model without the need for data exchange. In the area of vehicular communication, utilization of Federated Learning can be employed to develop a predictive model that estimates the trajectory of nearby vehicles by utilizing collected data. The Chimp Optimization Algorithm (ChOA) is designed to improve the model's efficacy. The proposed method aims to enhance the accuracy of the model's predictions regarding the conduct of nearby vehicles, while also reducing the amount of data exchanged between vehicles, by combining Federated Learning and Chimp Optimization termed FLECO. This method has the potential to enhance vehicular communication effectiveness and security, while also improving road safety and traffic management. Federated Learning facilitates the group control of a machine learning (ML) system by vehicles through the adjustment of model parameters. To enhance the energy efficiency of the system, the implementation of resource allocation and an energy-efficient algorithm is employed for Federated Learning, which integrates power and time allocation methods. This paper conducts a comprehensive analysis of the impact of enabling re-routing capabilities on (i) the mobility of vehicles and (ii) Networks for predicting traffic. To achieve this, utilize the SUMO simulator for road traffic to generate vehicle trajectories. Subsequently, we evaluate the vehicular network's connectivity employing established graph metrics. The developed system is simulated using the Python tool and experimentally validated, demonstrating its effective accuracy in vehicular communication.

✉ Ruchi Gupta
80ruchi@gmail.com

Extended author information available on the last page of the article

Keywords Federated learning · Vehicle communication · Chimp optimization algorithm · SUMO · FLECO

1 Introduction

Wireless communication techniques have undergone rapid development in smart cities, and Vehicular Ad-Hoc Networks (VANETs) are a crucial aspect of this type of communication in Intelligent Transportation Systems (ITS). Establishing effective V2V communication is a critical prerequisite for enabling autonomous and ITS. VANETs are an essential element of smart cities, and they represent a rapidly developing research field that involves collaboration between the research community and industry [1, 2]. Recent technological advancements have significantly improved the modelling of VANET communication, resulting in the ability to derive vehicular communication technology to save money, energy, and time. Due to the advancement of technology in recent times, VANETs have become essential in resolving everyday vehicular issues and establishing vehicular identities [3, 4]. Therefore, VANETs need to be upgraded to make them compatible with traditional technologies and meet the increasing demand. With the constant increase in the number of smart devices and vehicles, predicting network traffic has become an important issue that presents several challenges, such as maintaining data privacy, managing large volumes of data, and ensuring accuracy in prediction [5, 6].

In the context of a VANET, the user can utilize network resources by processing, renting, storing, and sharing them to run crucial applications through a software deployment based on a system. In addition, cloud services are utilized by vehicular nodes to acquire multimedia and traffic information. The architecture consists of two tiers, where vehicles are positioned at the network's edge, and the centralized cloud is situated at the core [7, 8]. A cloud-based VANET can control all automobile's geographic locations on a macro level. By specifying the intended recipient and targeted area of safety alerts, the system can gather real-time traffic flows [9]. The significant amount of data produced by this system is crucial data that can serve various purposes. The data provides numerous benefits concerning both commercial and eco-friendly data management system development. Therefore, a cloud-based VANET is an innovative system that can offer numerous benefits in the field of traffic management and data analysis. The data obtained from VANET can provide valuable insights to help make informed decisions about our Intelligent Transportation Systems (ITS) portfolio [10, 11]. However, with the increase in the volume of data that is expected in the ITS era, standard database systems may not be able to handle such a vast quantity of data efficiently. This creates a need to develop new and advanced data management systems that can handle large amounts of data and analyze them effectively.

Simulation is the predominant approach employed by researchers to evaluate vehicular network performance in today's world. Owing to the complexity and cost of real-world test scenarios, simulation techniques have become a common method for studying vehicular network protocols and services in ITS [12, 13]. By using simulations, researchers can assess the feasibility and effectiveness of different scenarios in a controlled environment, without incurring the costs and risks associated with real-world experiments. Simulating inter-vehicle communication in vehicular networks usually involves combining a network simulator with a realistic road traffic simulator, such as Simulation of Urban MObility (SUMO). The integration of event detection mechanisms like traffic congestion and accidents, realistic maps, mobility models, and route planning capabilities are part of this coupling [14, 15].

An open-source, space-continuous road traffic simulator known as SUMO is extensively utilized to test ITS services and VANETs. With the ability to manage vast road traffic networks, it offers features for modelling road networks and vehicular demand, including right-of-way regulations, traffic lights, and lane changing. Additionally, it includes simulations for pedestrians and public transportation. The SUMO package includes tools for defining vehicle traces manually or automatically [16].

Although machine learning (ML) algorithms are prevalent across diverse industries, standard algorithms typically need training and data storage in a cluster, single machine, or data center. FL, a new distributed ML paradigm, has been proposed to tackle this limitation, and it has gained increasing popularity and interest [17]. FL offers various benefits in terms of performance, including lower communication overhead and stronger privacy guarantees since devices don't need to transmit the entire database. Reduced communication overhead often results in lower latency for completing the training and decreased power consumption in communication when the communication bandwidth is limited. A significant role in safeguarding privacy in machine learning for self-driving cars is predicted to be filled by FL. FL enables combined training of an ML model on distributed datasets among edge devices while preserving the privacy of data on each device. Scaling large models with FL remains a challenge, despite its less communication requirements in contrast to traditional distributed learning methods. FL's ability to function with constrained communication resources, along with the mobility of edge nodes and heterogeneity of data distribution, is vital for vehicular networks to be successfully deployed. Effective use of communication resources and adoption of novel perception-based learning methods are crucial [18, 19].

Many researchers who focus on developing vehicular network models and validating results often do not consider the impact of traffic conditions when simulating the road network. This research examines the effect of vehicles' re-routing properties. To simulate traffic mobility and the effect of automatic re-routing in vehicles, we utilize the widely-used traffic simulator SUMO on a real-world map. Graph theory concepts are utilized to assess inter-vehicle connectivity. Initially, we conduct an offline analysis by converting vehicle trajectories into snapshots at various simulation times. This allows us to assess the network using established graph metrics [20]. This research, introduced FLECO (Federated Learning with Chimp Optimization Algorithm), a novel approach leveraging the Chimp Optimization Algorithm to enhance the efficiency and accuracy of Federated Learning. FLECO optimizes model parameters collaboratively across vehicular nodes, improving predictions of neighboring vehicle movements while minimizing data exchange, ultimately enhancing vehicular communication efficacy and contributing to road safety and traffic management.

1.1 Contribution of the work

- Introduced a sophisticated distributed scheme for spectrum allocation within the V2V communication system. This allocation scheme optimizes the usage of the wireless spectrum, thus significantly enhancing the overall transmission effectiveness. The approach minimizes interference and maximizes bandwidth utilization for improved communication.
- Utilized the powerful Federated Learning (FL) framework to train an advanced AI model. This approach allows for collaborative model training without the need to cen-

tralize sensitive data. By leveraging the collective intelligence of vehicular nodes, the model achieves higher accuracy and efficiency while preserving individual data privacy.

- Developed and integrated the Chimp Optimization Algorithm (ChOA) to further optimize the efficiency of the AI model. ChOA intelligently fine-tunes model parameters, enhancing predictive accuracy and model efficiency. This integration represents a novel approach to improving machine learning models within the V2V communication context.
- Placed a significant emphasis on minimizing the latency within the V2V communication system. By optimizing algorithms and model parameters, the proposed model effectively reduces communication latency, leading to more efficient and prompt data transmission.
- Presented an extensive theoretical analysis of the proposed algorithm to validate its expected performance. The theoretical framework provides insights into how the algorithm is anticipated to perform under various conditions, laying a solid foundation for its application in practical scenarios.
- Conducted thorough experimental analysis to evaluate and validate the proposed algorithm. The experiments included a comparative study against existing algorithms, showcasing the superior performance and efficiency of the proposed model in real-world scenarios. This rigorous evaluation provides concrete evidence of the algorithm's efficacy and advantages.
- Demonstrated how the research contributes to vehicular safety by enabling intelligent communication among vehicles. Optimized communication allows for quicker response times, early collision warnings, and coordinated traffic flow. These elements collectively enhance road safety and contribute to accident prevention, ultimately making roads safer for all users.
- Focused on building a system that is highly adaptable and scalable to meet the evolving demands of V2V networks. The model can efficiently handle a growing number of vehicles and changing communication dynamics, ensuring the system's continued efficiency and relevance as vehicular networks expand and evolve.

The paper is structured as follows: Section 2 offers a summary of the reviewed literature that includes research on the design and implementation of vehicular communication systems. Section 3 outlines the stepwise methodology for the proposed approach. Section 4 presents an extensive discussion of the simulation environment, while Section 5 provides the experiments and their outcomes. The final section of the paper presents a summary of the proposed approach and the conclusion.

2 Literature survey

In this section, an analysis of different vehicular communication techniques that are currently in use is presented. Additionally, it describes the major advantages and limitations of these techniques.

A distributed Federated Learning method was developed by Samarakoon et al. [21] to achieve Ultra-Reliable Low-Latency Vehicular Communications (URLLC). By formulating the problem as a network-wide power minimization issue, while accounting for URLLC, they addressed the issue of power control and resource allocation for the V2V communication network. The researchers utilized extreme value theory to model the constraints related to URLLC. Researchers developed a distributed learning mechanism based on the principles of FL. To accurately estimate traffic flow characteristics like tail distribution, a vehicle-to-infrastructure (V2I) system relies on a complex mechanism that involves multiple components working together. By adopting the FL approach, VUEs can learn the tail distribution of the network-wide queue. Then, they presented a Lyapunov-based approach to allocate resources and manage power usage for VUEs by combining the EVT and FL approaches.

In recent years, researchers have explored a variety of optimization algorithms to improve the efficiency and reliability of vehicular communication networks. One such algorithm is the enhanced whale optimization algorithm, which was developed by Valayapalayam et al. [22]. Even though the Roadside Unit (RSU) manages the mobility factor of vehicles in the traffic system, there are still unresolved challenges related to mobility management. An algorithm was presented to improve the mobility management of the system and avoid expensive RSUs. This algorithm organizes a clustering structure and selects a suitable cluster head (CH) for VANETs. To optimize network parameters for the presented Adaptive Weighted Clustering Protocol (AWCP), researchers developed the Enhanced Whale Optimization Algorithm (EWOA) to group nodes and select the optimal Cluster Head (CH).

Sepasgozar and colleagues [23] introduced a novel approach for predicting network traffic in VANETs using the FL algorithm. Their approach, called Fed-NTP, utilizes the LSTM algorithm for local model training to achieve precise network traffic flow predictions while protecting user privacy. The authors employed a distributed approach to utilize the LSTM algorithm, employing the FL algorithm on the VANET dataset. By examining the most impactful characteristics of network congestion in both road and network environments, the model predicts network traffic.

Arya et al. [24] devised an Intruder Detection approach using Federated Learning for VANET Data Streams in Smart City Environments. Their intrusion detection technique is designed to save time and resources by utilizing the most efficient method, employing a heterogeneous neural network with a distributed FL approach. The initial phase comprises vehicles utilizing FL techniques to create DL-based Intrusion Detection System (IDS) classifiers for data streams in VANET. Upon request, they share their locally trained IDS classifiers with adjacent vehicles, which efficiently decreases communication overhead. For each vehicle, an ensemble of heterogeneous neural networks is formed through FL, which consists of classifiers trained locally and remotely. Finally, the local devices are updated by sharing the global ensemble model.

A new Clustered Vehicular FL model was introduced by Taïk et al. [25]. An architecture for FL in vehicular networks and the associated processes for learning and scheduling were presented by the researchers. To tackle the communication bottleneck, the architecture utilizes V2V resources. Simultaneously, clusters of vehicles train models, and the only information transmitted to the multi-access edge (MEC) server is the collective of each cluster. The clustering formation is tailored for both multi and single-task learning, considering both learning and communication aspects.

Yu et al. [26] proposed an FL-based scheme, Mobility-Aware Proactive Edge Caching for Connected Vehicles (MPCF), to enhance cache performance in vehicular edge networks. The scheme facilitates collaborative learning among multiple vehicles using private training data to predict content popularity with a Context-aware Adversarial Auto Encoder. Furthermore, the mobility-aware cache replacement policy is integrated into MPCF, enabling network edges to adjust content addition or eviction based on vehicle mobility patterns and preferences. This approach significantly improves cache performance, preserves users' privacy, and reduces communication costs.

In Autonomous Vehicular Networks, Ge et al. [27] proposed an approach to address URLLC. Initially, the authors introduced a function that considers both reliability and latency to analyze how they affect 5G autonomous vehicular networks when combined. The authors suggested a new network slicing approach, which spans from resource slicing to service and function slicing, to enhance the reliability and latency performance of autonomous vehicular networks in 5G. In addition, they demonstrated the interactions between latency and reliability through Monte Carlo simulations.

Yang et al. [28] formulated a Cluster-Based 3D-channel Model to analyze V2V Communications. Their model takes into consideration the spatial distribution of MPCs in both the horizontal and vertical dimensions. The extraction of MPCs is done using the Space-Alternating Generalized Expectation–maximization (SAGE) technique, and to detect and track dynamic MPC clusters, clustering, and tracking algorithms are utilized. The distribution of MPC clusters is characterized by intra and inter-cluster parameters and categorized into two types: global and scatterer-clusters. According to the model, the log-normal distribution characterizes the spread of both azimuth and elevation. Within a cluster, the power of MPCs is distributed as truncated Gaussian, while the angle of MPCs follows the Laplacian distribution.

Al-Shareeda et al. [29] presented a communication scheme for VANET that ensures privacy preservation. During initialization, the TA generates public parameters and computes private and public keys for a domain that includes multiple RSUs in a specific area. Moreover, the vehicle registration list contains the registered On-Board Units (OBUs). During the second phase, the OBU generates n lists of pseudo IDs based on its actual identity and public TA parameters. To initiate transmission and validation operations, the vehicle needs to establish mutual authentication with the nearest RSU located within the domain. The authenticity of the OBU is confirmed by the TA using the system's private key. Subsequently, a secure list of signatures for the selected timestamp, generated by the RSU, is received by the OBU. The value of n determines the level of security and anonymity for a vehicle within a region defined by the RSU, by limiting the number of unique pseudo-identities that can be utilized. Lastly, until the timestamp expires, the OBU employs the signature list.

Zhang et al. [30] proposed a self-adaptive routing service algorithm for VANET that predicts link reliability and develops a secure routing protocol to execute diverse Quality of Service (QoS) application demands. The algorithm analyzes the movement patterns of vehicles and factors contributing to link failures to propose a link duration model that evaluates link reliability. To ensure the optimal end-to-end path is maintained, the Q-Learning algorithm adaptively modifies the routing path through interactions with the environment. The self-adaptive routing algorithm called Reliable Self-Adaptive Routing (RSAR) incorporates the reliability parameter and a fine-tuned heuristic function, resulting in a significant performance in VANET. Table 1 displays the summary of the existing methods.

2.1 Problem statement and motivation of the work

In an era where technological advancements are rapidly transforming our lives, envisioning a future where roadways are not just a network for transportation, but a domain of intelligent, interconnected vehicles is within our grasp. The potential impact of such a transformation is immense: safer roads, streamlined traffic flow, reduced emissions,

Table 1 Technique, summary, pros, and cons of the reference papers

S. No	Reference Author	Technique	Summary	pros	cons
1	Samarakoon et. al., [21]	Distributed Federated Learning	The paper highlights the challenges in achieving a universally optimal design due to the complex and dynamic nature of power generation systems	Have considerable gains in reducing extreme events	Absence of non-IID training data
2	Valayapalayam Kittusamy et.al., [22]	whale optimization algorithm	The approach is based on an EWOA for organizing a cluster structure and cluster head (CH) election suitable for VANETs	High clustering efficiency	Poor mobility
3	Sepasgozar et.al., [23]	Federated Learning Algorithm and VANET	The proposed algorithm is based on the Long Short-Term Memory (LSTM) algorithm and is implemented in FL to predict network traffic accurately while preserving privacy	Lowest error in terms of network traffic	Data leaking may occur if local data is transmitted to the server
4	Arya et.al., [24]	VANET Data Streams with FL	The presented approach uses FL to estimate the tail distribution of the queue lengths for joint power and resource allocation (JPRA) in ultra-reliable low-latency vehicular communications	Can cover a large number of attacks	Not compatible

Table 1 (continued)

S. No	Reference Author	Technique	Summary	pros	cons
5	Taïk et.al., [25]	Clustered Vehicular Federated Learning	The architecture utilizes vehicular-to-vehicular (V2V) resources to bypass the communication bottleneck where clusters of vehicles train models simultaneously and only the aggregate of each cluster is sent to the multi-access edge (MEC) server	Reduced environmental impact	High network traffic
6	Yu et.al., [26]	Federated Deep Learning For Edge Caching	The proposed approach is based on federated learning and is designed to support the mobility of vehicles. The paper discusses the challenges in achieving efficient edge caching for connected vehicles and presents the proposed mobility-aware proactive edge caching scheme	Improved safety	Low clustering efficiency
7	Ge et.al., [27]	Ultra-Reliable in Autonomous Vehicular Networks	The presented approach is based on FL and is designed to support the mobility of vehicles. The paper discusses the challenges in achieving efficient edge caching for connected vehicles and presents the proposed mobility-aware proactive edge caching scheme	Better reliability	Covers only a reduced number of attacks

Table 1 (continued)

S. No	Reference Author	Technique	Summary	pros	cons
8	Yang et.al., [28]	Three-Dimensional Channel Model for v2v communication	The paper highlights the importance of V2V channel modeling for the design and performance evaluation of intelligent transportation systems. The proposed model is organized into clusters, and the channel characteristics are modeled using a combination of path loss, shadowing, and fading	Enhanced convenience	Low clustering efficiency
9	Al-Shareeda et.al., [29]	Vanet-based privacy-preserving communication scheme	The paper highlights the need for contextual privacy requirements and presents an approach based on FL to estimate the tail distribution of the queue lengths for JPRA in ultra-reliable low-latency vehicular communications	Increased efficiency	Complex model
10	Zhang et.al., [30]	self-adaptive routing service algorithm for application in VANET	The paper highlights the importance of efficient and stable routing algorithms in urban vehicular networks and provides an overview of relevant VANET routing protocols	Good mobility	Security risk

and improved overall quality of life. Vehicle-to-vehicle (V2V) communication stands as a linchpin in realizing this transformative vision. Every day, countless lives are affected by road accidents and traffic congestion. As urbanization continues to rise, the challenges of managing road safety and traffic flow become increasingly complex. These challenges necessitate innovative solutions that can seamlessly integrate with the evolving landscape of transportation.

The rapid growth of vehicular traffic has created an urgent demand for ITS. However, this surge in traffic also brings about detrimental effects such as strain on public investment, persistent traffic congestion, increased accident rates, heightened fuel consumption, and escalated environmental pollution. An effectively managed system is crucial to mitigate these challenges, and Vehicular Ad Hoc Networks (VANETs) have emerged as a promising solution proposed by researchers. VANETs utilize vehicles as network nodes, forming ad hoc networks to address traffic and transportation-related issues, and they belong to the ad hoc network family.

3 Methodology

The proposed framework proposes an approach for vehicular communication that utilizes the Federated Learning Empowered Chimp Optimization (FLECO) algorithm. The model involves the use of federated learning by vehicles to train and develop the system. The FLECO algorithm is a novel method that combines chimp optimization and federated learning to optimize communication between vehicles. The proposed model's workflow involves Data Collection, Data Preparation, a Federated Learning model, Optimization, and finally communication with vehicles. The first step is to collect data from various sources, including traffic patterns, road conditions, and vehicle speed, among other factors. The collected data is then pre-processed and prepared for training by performing data cleaning, normalization, and feature selection to ensure high-quality data that can be used to train the FLECO algorithm. The ML technique is fine-tuned using chimp optimization in the next step of the proposed workflow. The FLECO algorithm utilizes FL to train the model, which predicts the best communication path between vehicles. The training process is distributed among multiple vehicles in federated learning, and they work together to train the model without sharing their data. Once the FL process is finished, the FLECO algorithm uses chimp optimization to further enhance the machine learning model. Mimicking chimpanzees' behavior to find the best solution to a problem, chimp optimization is a nature-inspired optimization algorithm. Communication optimization is the ultimate step of the FLECO algorithm. To minimize latency, maximize throughput, and ensure the communication network's reliability, the FLECO algorithm optimizes communication between vehicles based on the machine learning model's predictions.

Overall, Fig. 1 provides an overview of the workflow of the proposed model, with each sub-section addressing a specific aspect of the methodology. Section 3.1 presents the data collection process for the proposed model. In Section 3.2, the data preparation steps are described. Section 3.3 focuses on the federated learning aspect of the model. Section 3.4 elaborates on the optimization process, and in Section 3.5, the communication optimization techniques used in the proposed model are explained.

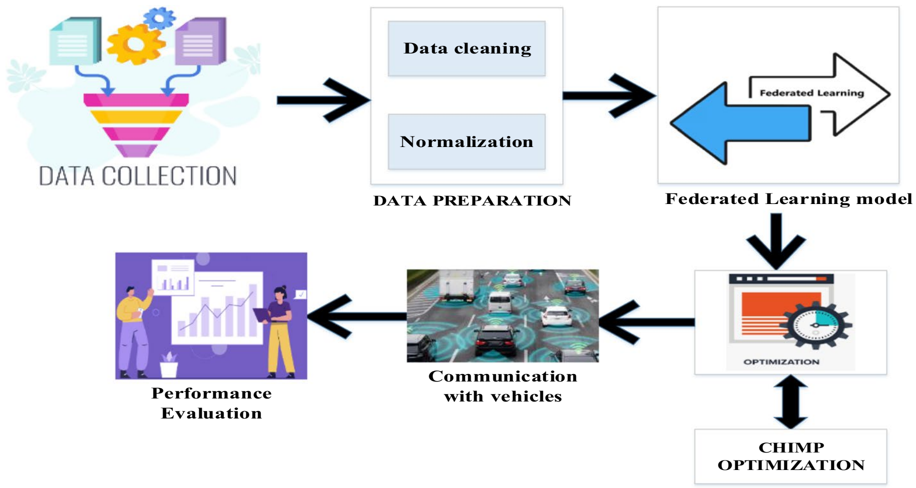


Fig. 1 Block diagram of the proposed model

3.1 Data collection

In vehicular communication, the collection of data pertains to obtaining information from vehicles within a vehicular network. Data generation from SUMO simulation software is currently underway. These data serve multiple purposes, such as the optimization of routes, enhancement of safety, and improvement of traffic flow. The "DSRC Vehicle Communications" dataset provides valuable insights into wireless communications between vehicles and roadside units, encompassing two distinct scenarios: a normal communication scenario and a scenario in the presence of a jamming attacker. In each scenario, the dataset comprises a total of 10,000 instances, capturing various communication parameters. There are five features associated with each instance, making it a relatively low-dimensional dataset. The features encompass essential communication metrics, namely:

1. Txnid (Transmitted node ID number)
2. Rxnid (Received Node ID number)
3. RSS (Received Signal Strength in dBm)
4. BER (Packet Error Rate)
5. RSSI (Received Signal Strength Indicator)
6. SNR (Signal-to-noise ratio)

These features represent real values, falling under the category of real feature types. Notably, the dataset does not contain any missing values, ensuring the completeness and reliability of the dataset for analysis and modelling purposes. The communication setups were established in accordance with IEEE 802.11p standards at a frequency of 5.9 GHz, with 10 Basic Service Messages (BSM) transmitted per second. The Control Channel (Ch172) was utilized, employing a 10 MHz channel.

3.2 Data preparation

Data is collected, pre-processed, and fully prepared for training. The pre-processing stage involves cleaning and normalizing the data, as well as selecting relevant features to ensure that the data is of high quality and can be utilized to train the FLECO algorithm. After processing the data, insights, and trends can be extracted through various techniques. Our proposed model utilizes the Federated learning model to train the data generated.

3.3 Federated learning (FL) model

The combined training of a single ML model among diverse participants on their respective local datasets is facilitated by the FL, which is a framework for distributed training that ensures privacy. Commencing the iterative training process, a centralized entity, such as a server, initializes the global model. During each iteration, i, a subset of N participants are selected to receive the current global model θ*t*. By executing Iterative stochastic gradient descent (SGD) on mini-batches extracted from their local database, each participant k trains the model. Weight-update vectors Δθ $t+1$ are produced and sent to the server after local training. Lastly, the server performs model aggregation, typically accomplished by using weighted aggregation [31]. The process is iterated until the model reaches convergence. The algorithm provides an overview of the process, with further scheme details outlined below:

Step 1- Distribute the FL model and its specifications and gather feedback

The server releases a global model, accompanied by its resource requirements for computation and data (e.g., CPU cycles, data types, and data sizes). After satisfying these requirements, each vehicle *'p'* offers affirmative feedback, including other details such as its current velocity and data diversity index I_P (as per Eq. 1) and current velocity v_k.

$$l_p = \sum_i \phi_i, p\gamma_i \tag{1}$$

with i∈{elements diversity, dataset size, age}. Other task-specific considerations can be easily incorporated into the metric.

Step 2—Select and schedule CH

Cluster heads are chosen by the server based on information received, taking into account dataset features (including dataset quality and sample quantity), as well as wireless channel status and anticipated communication duration (indicated by the stay rate). In actuality, the quality and significance of model updates are directly correlated with local dataset quality, while wireless channel state and velocity dictate the feasibility of receiving the model update during the communication round.

$$T_P = \frac{G - x_p}{v_p} \tag{2}$$

To establish communication with the gNodeB, the chosen cluster-head vehicle k must adhere to a specific standing time rate, denoted as $(t_k^{train} + t_k^{up} + T_{agg} + \delta) \leq T_k$. The estimated training time and upload time of vehicle k are represented by t_k^{train} and t_k^{up}, respectively. *Tagg* denotes the time needed for aggregation, and δ is the waiting time for collecting updates. The main factors that differ significantly between vehicles are t_k^{train} and t_k^{up}. t_k^{train} is influenced by the size of the dataset, while t_k^{up} is determined by the channel gain and resource block allocation.

Step 3- Clusters formation

Once the cluster-head selection process is complete, the remaining vehicles NH are then paired with the cluster-heads in set H. For a successful match, the total time taken for training and uploading of vehicle k should be lower than the Link Lifetime (LLT) between p and h∈H, as defined in Eq. 8. In addition, the objective of the matching process is to optimize the total weighted sum of $W_{p,h}$. The symbol $W_{p,h}$ represents the connection between p and h, and its meaning varies depending on the presence of a single overarching model or multiple variations (refer to Eq. 4). The clustering in a single joint model is determined solely by mobility. Therefore, for all pairs $k \in NH$ and $h \in H$, the value of $W_{p,h}$ is equal to 1. In contrast, every vehicle must independently train its preferred model. The preference refers to the level of accuracy exhibited by the model trained by h using the local data of *p*. The reason for this definition is that not all vehicles can take part in the clustering step of the updates (refer to Step 5).

$$LLT_{P,h} = \frac{-\Delta u_{ph} \times G_{Ph} + \left|\Delta u_{ph}\right| \times TW}{(\Delta u_{ph})^2} \tag{3}$$

In Step 3, the formation of clusters occurs by considering the inter-vehicle relationships. The determination of the scope of this relationship is based on whether a single global model is trained or if multiple iterations of the model are generated. When there are multiple models, we determine the preference of a model based on how accurately it performs on the dataset of the k^{th} vehicle. The relationship between two vehicles, denoted as $\mathrm{W}_{k,h}$, is defined as follows:

$$W_{P,h} = \begin{cases} accuracy\, of\, h \;\; if\, more\, than\, 1\, \mathrm{mod}\, el \\ \quad 1 \qquad\qquad\qquad otherwise \end{cases} \tag{4}$$

Since not all vehicles are capable of participating in the update clustering phase, the preference is established depending on the model's accuracy that is trained by *h* on *p's* local data.

Step 4—Broadcasting the model and training

Each vehicle can train on its local data for a specific number of local epochs, denoted by ϵ, and then send the update to its corresponding cluster head after the model has been aggregated to the participants. The next step involves the cluster head collecting the models it received and sending the update to the server [32]. Hierarchical FL aggregation, which promotes more participation, is commonly used to aggregate the global updates of the clusters on the server. To enhance resilience against client drop-out, it is necessary to have multiple global models that can be trained across several clusters, thus providing redundancy.

Therefore, performing update aggregation at the server level becomes a requirement. This is particularly relevant in vehicular networks.

Step 5—Updates Clustering

If the global model fails to converge or the desired accuracy is not achieved, multiple communication rounds may be executed, which will involve a significant number of vehicles in training the joint model globally. The updates are collected at the MEC server without intermediate aggregation by CHs in this step. This is because aggregated models may conceal the divergence of the various techniques. The gathered updates are subsequently employed to evaluate the similarity among participants utilizing the hierarchical clustering algorithm (defined in Eq. 5).

$$sim(p, l) = \frac{\langle \Delta\theta_p, \Delta\theta_j \rangle}{\|\theta_p\| \|\theta_j\|} \tag{5}$$

The OEM defines a maximum number of clusters, and participants are iteratively merged based on their similarity until the maximum number is reached. Limiting the number of circulating models is accomplished by establishing a predetermined upper bound for cluster formation, enabling clusters to be formed without any prior understanding of the potential distances between updates. After clusters are built, novel designs are developed via aggregation and then broadcast to available automobiles. The models are then sent back to the MEC server after being evaluated for each vehicle p using local data. Later, these values are employed to evaluate R_P and h for the particular vehicle p. Each resulting model is then individually trained using utilizing a similar process. Our work differs from the earlier works on federated learning with clustering as we evaluate preferences, which allows for partial participation instead of requiring all nodes to participate.

3.4 Chimp optimization

Our model has developed a novel algorithm, ChoA, that takes inspiration from the intelligence abilities and social dynamics exhibited by chimpanzees during group hunting. Unlike other social predators, this strategy is distinct. ChoA can be utilized to optimize vehicular communication networks, particularly in the context of VANETs. In VANETs, vehicles share data about traffic conditions, road hazards, and other relevant data by communicating with each other and with roadside infrastructure. Optimizing the communication routes between vehicles can be achieved by utilizing chimp optimization, which considers vehicle speed, network congestion, and signal strength as influencing factors. Through this technique, latency can be minimized, throughput can be maximized, and the overall reliability of the communication network can be improved. To simulate various forms of intelligence, including attacker, barrier, chaser, and driver, chimp optimization employs four distinct phases. The mathematical model of the proposed algorithm is presented in a step-by-step manner, detailing how targets or prey are driven and chased using the following set of equations (Eqns. ((6)-(7))).

$$D = \left| z.q_{prey}(n) - xa_{CHIMP}(n) \right| \tag{6}$$

$$a_{chimp}(n+1) = q_{prey} - q.d \tag{7}$$

The total number of iterations is denoted by *'n'*, and the coefficient vectors are represented by *'q', 'z',* and *'x'*. Equation (8) is used to calculate these coefficients.

$$\begin{gathered} q = 2.l.rv_1 - l \\ z = 2.rv_2 \\ x = CHOTIC_{value} \end{gathered} \tag{8}$$

During the iteration process, the chaotic vector x, which exhibits chaotic behavior, is utilized, with rv_1 and rv_2 being randomly generated values within the interval [0, 1]. Furthermore, the parameter '*l*' is gradually decreased in a non-linear manner from 2.5 to 0.

The behavior of chimps has been mathematically implemented in this step. It is assumed that the attacker, driver, barrier, and chaser possess an initial solution, as they have better information about the target's location. The locations of the remaining chimps are updated based on the best chimp locations, while the four optimal solutions that have not yet been obtained are stored for the next iteration. Mathematical Eqs. (9)-(12) illustrate this process.

$$d_{barrier} = |z_1 q_{barrier} - x_1 y| \tag{9}$$

$$d_{barrier} = |z_2 q_{barrier} - x_2 y| \tag{10}$$

$$d_{chase} = |z_3 q_{chase} - x_3 y \tag{11}$$

$$d_{drive} = |z_4 q_{drive} - x_4 y \tag{12}$$

A chimp's next location can be any point between its current location and the target or prey's location, provided that the random vectors fall within the range of [-1, 1].

$$\begin{gathered} y_1 = q_{attack} - q_1.d_{attack} \\ y_2 = q_{barrier} - q_2.d_{barrier} \\ y_3 = q_{chase} - q_3.d_{chase} \\ y_4 = q_{driver} - q_4.d_{driver} \end{gathered} \tag{13}$$

The mathematical Eq. (14) expresses how the position of the chimps is updated during the search process using the overall equations.

$$y_{n+1} = \frac{y_1 + y_2 + y_3 + y_4}{4} \tag{14}$$

The following mathematical Eq. (15) is employed to update the chimp's location during the search process within the search domain.

$$q_{CHIMP}(n+1) = \begin{cases} q_{prey}(n) - y.d, if \varphi < 0.5 \\ CHAOTIC_{Value}, if \phi > 0.5 \end{cases} \tag{15}$$

Optimizing the placement of roadside infrastructure, such as wireless access points or communication relays, using the Chimp optimization algorithm can ensure that vehicles remain within the range of a dependable communication signal at all times. Chimp optimization's capacity to adjust to dynamic conditions in real time is its key advantage. Network congestion and changes in vehicle speed, for instance, can significantly affect the communication network's performance in vehicular communication networks. The network can optimize communication routes according to changing conditions quickly by utilizing chimp optimization.

3.5 Communication optimization

In the realm of vehicular communication optimization, integrating Federated Learning (FL) stands as a fundamental strategy. Communication optimization is the ultimate step in the FLECO algorithm, which focuses on optimizing communication between vehicles and infrastructure elements, such as roadside units or access points, in vehicular communication. Communication optimization aims to enhance the efficiency, reliability, and safety of the communication network. The FLECO algorithm leverages machine learning model forecasts to minimize latency, maximize throughput, and ensure the communication network's reliability by optimizing communication between vehicles.

Federated Learning facilitates collaborative model training without the need to share raw data, preserving data privacy while enhancing the overall model. Within vehicular networks, this integration manifests through a decentralized approach: initially, a global machine learning model is established with basic parameters, followed by local model training on individual vehicles using their respective datasets. These local models' updates, in the form of model parameters, are then aggregated at a central server, allowing the construction of an improved global model through iterative optimization. This integration ensures that sensitive data remains on the respective vehicles, addressing privacy concerns, while collectively leveraging the intelligence from various vehicles to refine the global model, ultimately optimizing communication efficiency and prediction accuracy in vehicular networks.

Complementing Federated Learning, the Chimp Optimization Algorithm (ChOA) emerges as a pivotal tool for enhancing communication efficiency. The Chimp Optimization Algorithm (ChOA) draws inspiration from the foraging behavior of chimpanzees to create an efficient optimization technique. At the core of ChOA lies a population of solution agents, akin to chimpanzee communities, initially placed within the problem's search space. Each agent represents a potential solution, and their fitness, akin to food quality in chimpanzee foraging, is evaluated based on the problem being addressed. Mirroring chimpanzee behavior, the algorithm introduces collaboration among agents, enabling information sharing and a collective search for better solutions. ChOA dynamically balances exploration and exploitation by adapting its search strategy based on the quality of solutions encountered. It intelligently explores diverse areas of the solution space initially and then narrows down the search to promising regions as the algorithm progresses, imitating how chimpanzees concentrate their efforts where food is abundant. The agents iteratively update their positions based on collaboration and prior experiences, gradually converging toward optimal or near-optimal solutions. Termination occurs based on predefined criteria, signalling the end of the optimization process. This adaptive, collaborative, and nature-inspired approach in ChOA makes it a powerful tool for efficiently optimizing models and enhancing communication within vehicular networks, aligning with the algorithm's biological inspirations. The pseudocode for the proposed model is shown in algorithm 1.

Algorithm 1 Pseudocode of FLECO.

```
ChimpOptimizationAlgorithm():
    Initialize the population of solution agents randomly within the search space
    Evaluate the fitness of each agent
    repeat until the stopping criterion is met:
        Share information among agents
        Update exploration and exploitation strategy based on fitness
        for each agent in the population:
            Explore the solution space:
                Generate a random direction vector (e.g., [-1, 1, 0.5, -0.5])
                Update agent position by adding a weighted sum of the direction vector
and step size to the current position
            Evaluate fitness for the updated position
            if the fitness of the updated position is better:
                Update the agent's position to the updated position
                Store the best position found so far (global best)
    return the best solution found
```

4 Simulation environment

The SUMO (Simulation of Urban Mobility) framework is a well-known open-source tool utilized to simulate and analyze vehicular communication in urban settings. Evaluating the performance of various vehicular communication protocols and applications is made possible by simulating different traffic scenarios. Traffic flow, vehicle density, and road infrastructure can all be modeled by configuring the simulation to represent different traffic conditions. Several statistics are provided by SUMO based on simulations. We examine key traffic measures, which include: (i) the trip distance, which represents the total path travelled from starting to the ending point; (ii) the trip duration, which indicates the estimated travel time for a vehicle to complete the journey from starting point to the ending point; and (iii) the count of route modifications experienced by each vehicle.

4.1 The effect of re-routing on the mean travel time of vehicles

Redirecting vehicles to alternative routes through re-routing to avoid congestion or other road conditions can significantly affect the average travel time of vehicles. If a vehicle experiences congestion on its primary route, it might take longer to reach its destination than if it had opted for a less congested path. The total travel time can be decreased by redirecting the vehicle through re-routing. However, this reduction in travel time may require the vehicle to travel an additional distance and/or time. The influence of re-routing on the mean travel time is reliant on various factors, such as the congestion levels on the original and alternative routes, the number of vehicles using the alternative route, and the delay caused by the re-routing process.

Travel time enhancement resulting from enabling re-routing capabilities is demonstrated in Fig. 2a. The most substantial enhancement is noticeable in the case of "*R—2 min*" with 30% of the traffic demand. Figure 2b displays the mean travel time during medium traffic, indicating that the "*R—2 min*" case is the minimum scenario with an average traffic volume of 60%. Figure 2c demonstrates the mean travel time during high traffic. The reduction in mean travel time can be achieved through re-routing if the alternate path has less congestion and the delay incurred by re-routing is insignificant. The simulation framework, SUMO, is utilized to determine the impact of re-routing on mean travel time. Researchers can simulate various traffic situations and analyze the effect of re-routing on the transportation system's efficiency. By comparing the outcomes of different re-routing strategies, researchers can identify the most efficient re-routing algorithms and evaluate the potential advantages of implementing re-routing in real-world transportation systems.

4.2 Simulation scenario

For the simulations conducted in this study, a practical scenario, as presented in Fig. 3, was employed. This scenario is representative of the Pankow district and adjacent areas in Berlin, Germany, and covers an area of 9 km. The road network was created by importing a genuine map from OSM using the Net-convert SUMO tool.

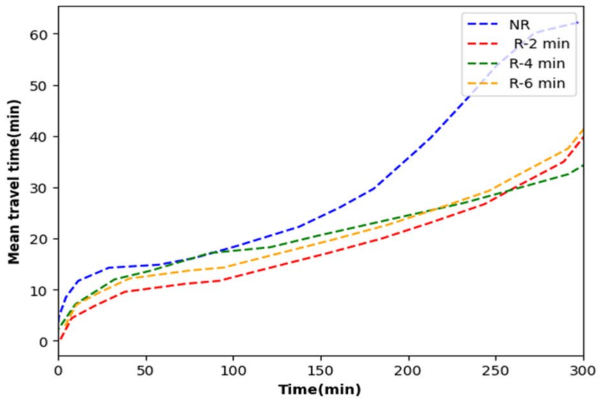
5 Performance evaluation

To evaluate the performance of vehicular communication, an analysis of the effectiveness of communication between vehicles and other vehicles or infrastructure in a specific network is required. Several factors, such as mobility, network traffic, mean score error, mean travel time, reliability, latency, accuracy, and loss rate, are utilized to assess the efficiency of vehicular communication.

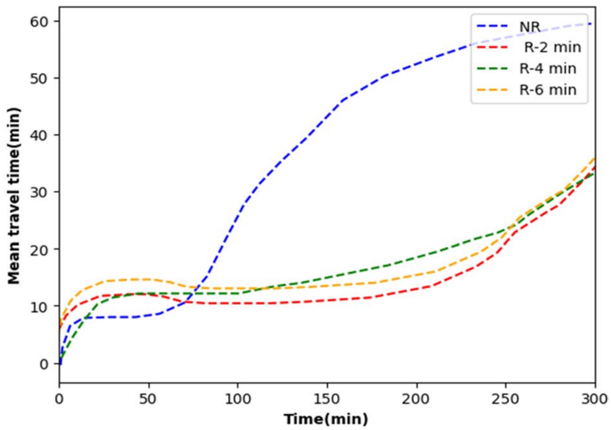
5.1 Latency model

Vehicles communicate information through vehicle-to-vehicle (V2V) links. The latency between vehicles can be classified into two categories, namely handling latency and propagation latency, without loss of generality. Therefore, the total delay in transmitting vehicular data is represented by,

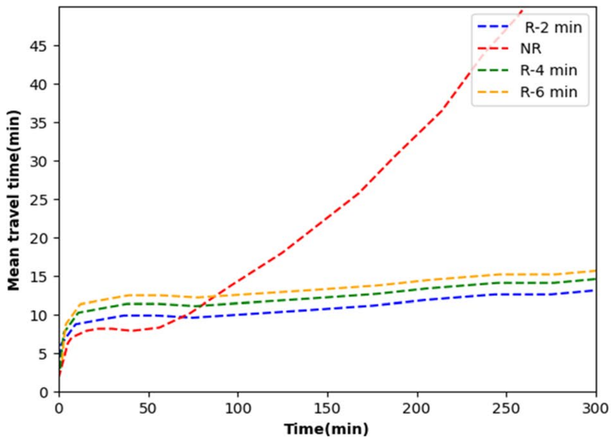
Fig. 2 **a** Mean trip time under Low traffic (30%), **b** Mean trip time under Medium traffic (60%), **c** Mean trip time under High traffic (100%)



a



b



c

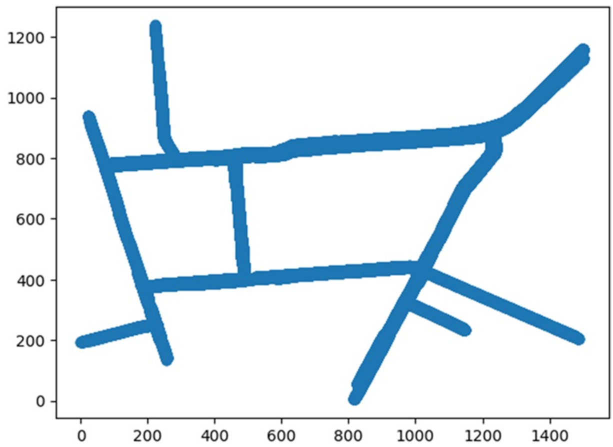


Fig. 3 Simulation scenario exported from the Pankow district of Berlin (Germany). The speed limit for each road is illustrated in meters per second (m/s)

$$T = T_{PL} + T_{HL} \tag{16}$$

where handling latency is denoted by T_{HL}, and propagation latency in the wireless connection is denoted by T_{PL}.

The density of vehicles on the road can have a major impact on latency, making it a significant challenge in vehicular communication. With an increase in the number of vehicles, the probability of congestion and network saturation rises, leading to increased latency. The time delay that arises when a signal travels from one point to another in a communication system is known as propagation latency. The distance between vehicles is one of the most crucial factors that impact propagation latency in vehicular communication, along with signal strength and the presence of obstacles or interference. As the distance between communicating vehicles increases, the signal takes longer to travel, leading to increased latency. This latency can pose a significant challenge in situations where quick response times are crucial, such as in collision avoidance systems. Several other factors can also influence propagation latency in vehicular communication, including signal strength and obstacles or interference.

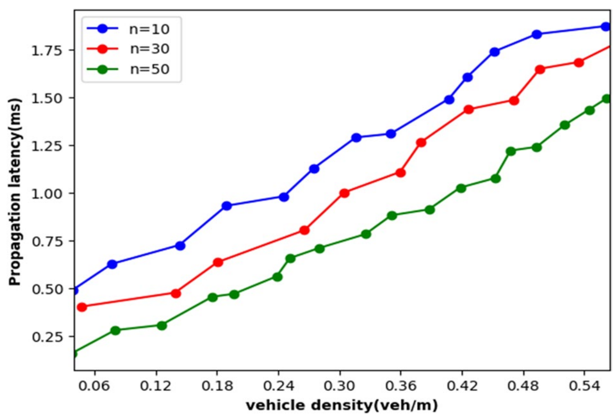


Fig. 4 Relationship between propagation delay and vehicle density

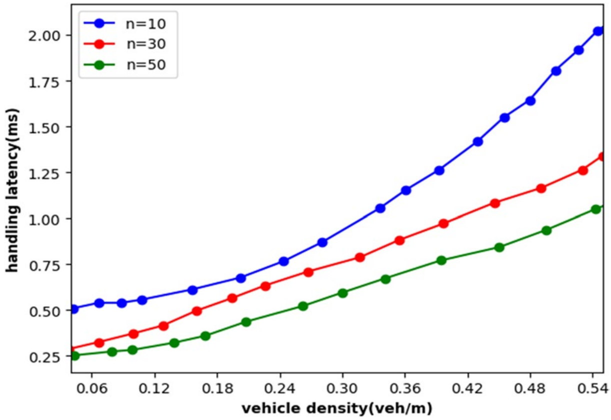


Fig. 5 Regulating latency with vehicle density

Exploring various vehicular densities, Fig. 4 displays the relationship between propagation latency and vehicle density. The chart displays the propagation latency for vehicular densities of $n=10$, 30, and 50. As the density of vehicles increases, the propagation latency also increases. Employing various techniques can help mitigate propagation latency in vehicular communication systems. Examples include the use of low-latency wireless communication protocols, optimization of communication node placement to minimize the distance between communicating vehicles, and implementation of signal amplification technologies.

Figure 5 shows the handling delay with the vehicle density at various densities. As the density of vehicles increases, the handling latency also increases. The total latency of a vehicular communication system is the combination of both the propagation latency and the handling latency.

Figure 6 shows the overall delay for various RSU densities with the density of vehicles. Keeping the number of vehicles stable, an increase in vehicular density results in an increase in the total latency. However, as the number of RSUs increases, the total latency decreases for a given vehicle density. Therefore, reducing total latency in a communication system can enhance the system's responsiveness and reliability.

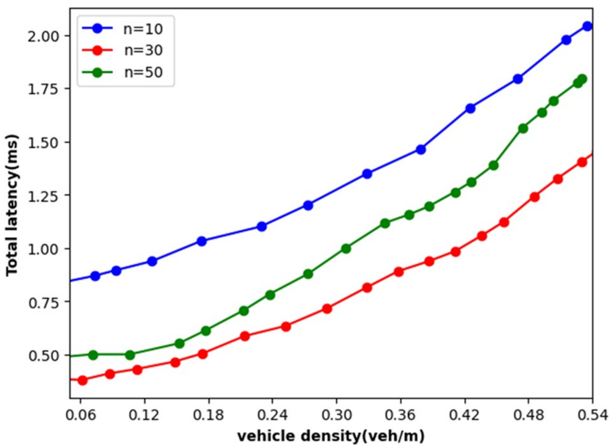


Fig. 6 Relationship between total latency and vehicle density

5.2 Reliability

Ensuring reliability in vehicular communication systems is crucial as it directly impacts the effectiveness and safety of the system. Several factors can affect reliability in vehicular communication systems, including interference, mobility, and communication range. The reliability of V2V communication is a crucial factor in achieving widespread deployment of cooperative vehicular systems. Ensuring the reliability of vehicular communication systems requires extensive testing and validation. This process can detect possible issues and guarantee that the system complies with the necessary safety and reliability standards.

With regards to the density of vehicles, Fig. 7 shows the reliability in relation to the number of vehicles *'n'*. When the value of *'n'* is kept constant, the reliability increases as the density of vehicles on the road network increases. On the other hand, reliability rises as 'n' grows while the vehicle density is constant. To enhance the reliability of communication systems, a combination of design, testing, and operational strategies is necessary. Achieving high levels of reliability for safety–critical applications in vehicular communication systems can be accomplished by implementing redundancy, monitoring and diagnostics, and fault tolerance. One common metric for reliability is the Packet Reception Ratio (PRR), which represents the ratio of successfully received packets to the total transmitted packets. The formula for PRR is:

$$PRR = \frac{Number\ of\ succesfully\ received\ packets}{Total\ number\ of\ transmitted\ packets} \times 100\% \tag{17}$$

5.3 Mobility analysis

The effectiveness and dependability of communication systems in vehicles, which involve transmitting and receiving data, can be optimized by analyzing *g* the mobility patterns and the potential impact of these patterns on the system. Mobility analysis in vehicular communication involves assessing how the movement of vehicles influences communication performance. Mobility in V2V communication is often quantified using metrics related

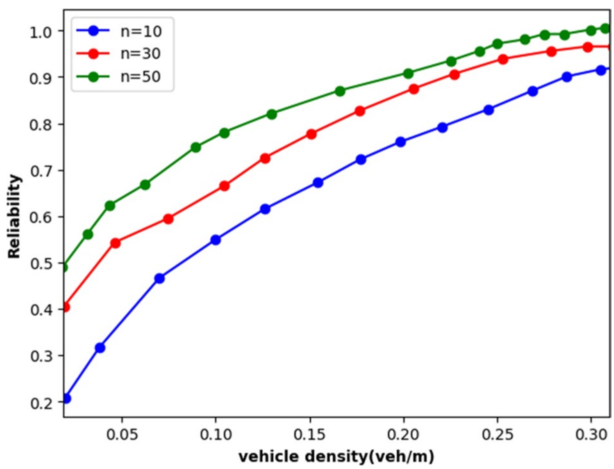


Fig. 7 Reliability in terms of vehicle density

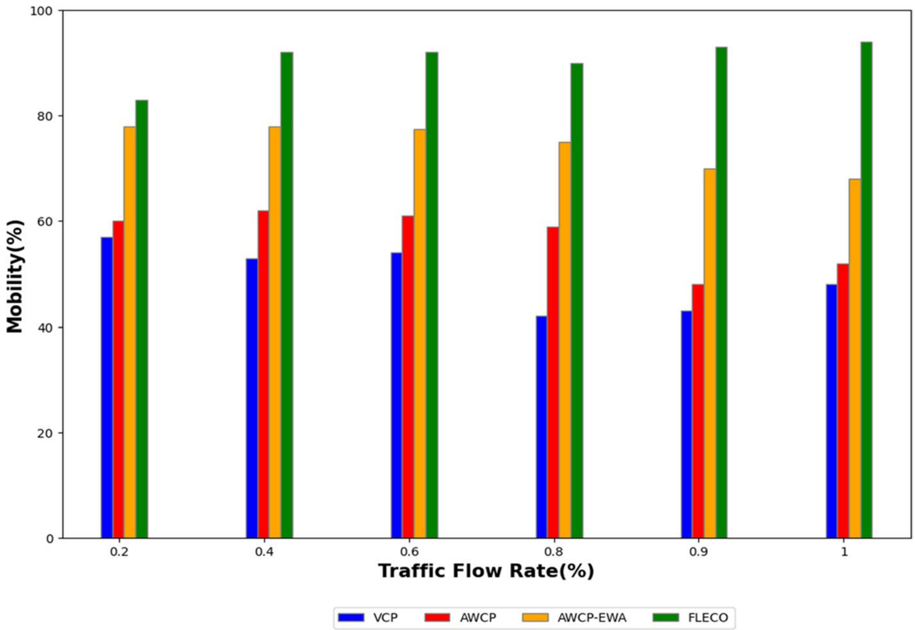


Fig. 8 Mobility analysis with traffic flow rate

to vehicle movement and its impact on communication. One common metric for mobility is Relative Velocity, which measures the velocity difference between the sender and the receiver. The formula for Relative Velocity *(V_{rel})* is:

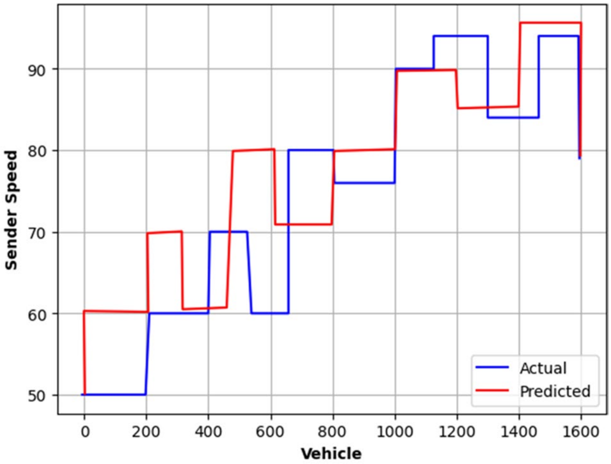
$$V_{rel} = \left|V_{sender} - V_{receiver}\right| \tag{18}$$

A graphical representation of mobility analysis with respect to traffic flow rate can be seen in the figure below.

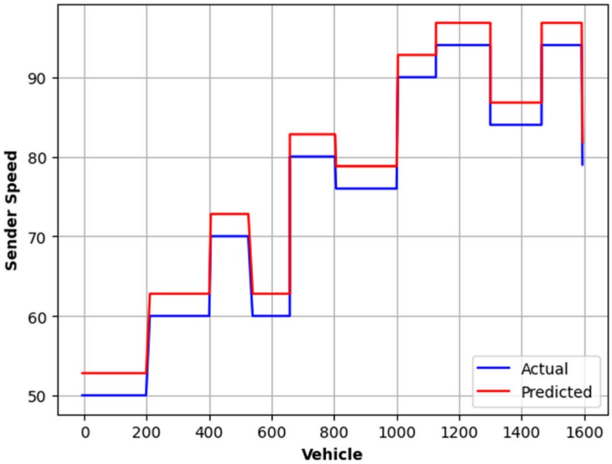
Different techniques, including AWCP, AWCP-EWA, VCP, and the novel FLECO Algorithm, were used to analyze the mobility of VANET systems, as illustrated in Fig. 8. The mobility of nodes in VANET is heavily influenced by the road structure and vehicle density in urban environments. The FLECO Algorithm, which employs an enhanced chimp optimization process, achieved the highest mobility compared to existing algorithms. The figures above demonstrate that the proposed FLECO algorithm's mobility analysis improves the system's efficiency, surpassing the compared methods in various aspects, such as energy efficiency.

5.4 prediction of network traffic

Monitoring network traffic in vehicular communication is crucial for optimizing communication system performance and improving transportation network efficiency. Accurately predicting network traffic has become increasingly important in modern times for effective traffic analysis. The need for monitoring network traffic has grown significantly over the years. In the past, administrators only monitored a small number of network devices or



a



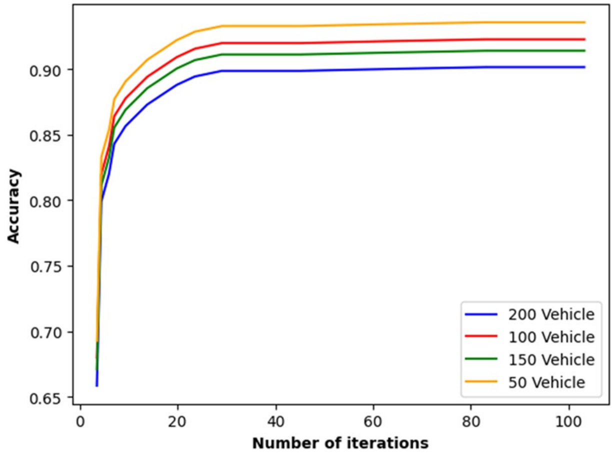
b

Fig. 9 **a** Results of network traffic prediction for the existing Fed-NTP, **b**. Results of network traffic prediction for the novel FLECO model

less than a thousand computers. To predict network traffic, Fig. 9 depicts the correlation between the number of vehicles and the sender's speed. The figure also shows the actual and predicted range of network traffic.

In Fig. 9a, the existing Fed-NTP model is used for network prediction, while in Fig. 9b, the proposed FLECO algorithm is employed. The graph in Fig. 9a displays the predicted outcomes for the "sender speed" parameter in network traffic flow. The predicted model is represented by the red line and the real data is represented by the blue line. As demonstrated in Fig. 9b, the proposed FLECO algorithm outperforms other algorithms with the smallest disparity between the predicted and actual data.

Fig. 10 Accuracy curve



5.5 Accuracy and loss curve

Evaluating the performance of learning models in vehicular communication involves using accuracy and loss curves. The loss curve measures how well the model fits the data, while the accuracy curve evaluates the model's prediction ability. The accuracy curve typically illustrates how the model's accuracy changes over time on both the training and validation data. As the model is trained on more data, its accuracy on the training data increases, while its accuracy on the validation data may eventually level off or decrease if the model starts to over-fit the data. Overfitting occurs when the model becomes too complex, fitting the noise in the data rather than the underlying patterns.

The accuracy curve of the federated learning model is depicted in Fig. 10, which illustrates that the accuracy value varies as the number of iterations (i.e., vehicles) increases. The accuracy of prediction is low for high vehicle density, while it is high for low vehicle density. Consequently, there is not much variation in accuracy between them.

Fig. 11 Loss Curve

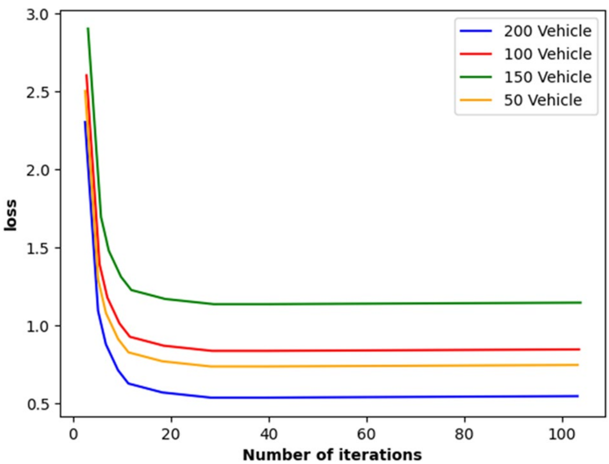


Fig. 12 Comparison of Accuracy

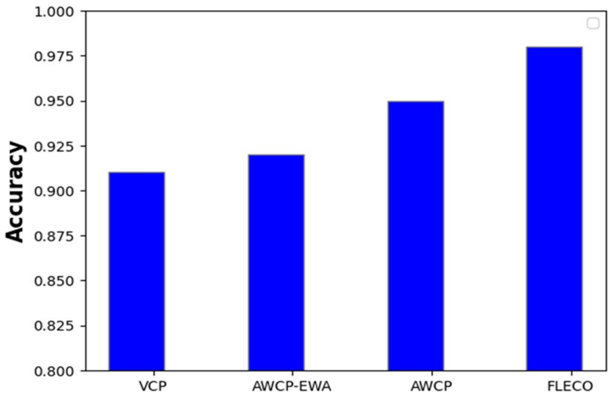
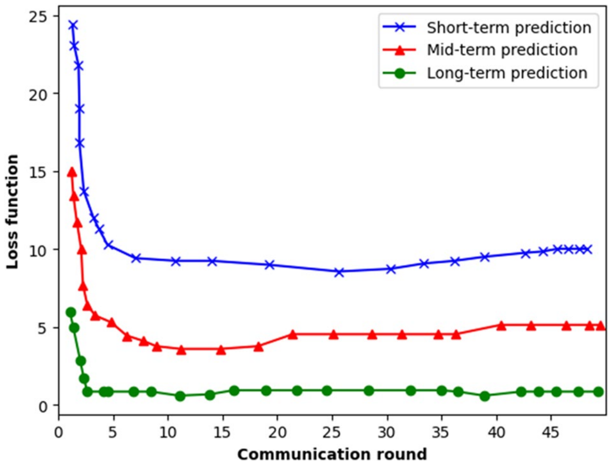


Fig. 13 Variation of Loss function as the FL proceeds



A federated learning model's error on the training and validation data is typically displayed in the loss curve over time. The error is determined by comparing the model's predictions to the actual values in the data. As more data is used to train the model, the error on the training data typically decreases. However, if the model begins to over-fit the data, the error on the validation data may eventually increase. A good federated learning model for vehicular communication should ideally have high accuracy on both the training and validation data, with a low loss on both sets. This implies that the model can generalize well to new data without overfitting the training data.

The loss curve of the federated learning model is shown in Fig. 11, where it can be observed that the loss value varies as the number of iterations (i.e., amount of vehicles) increases. For high vehicle density, the loss of prediction is high, whereas, for low vehicle density, the obtained loss is small. Hence, accuracy and loss curves are valuable tools for evaluating the performance of federated learning models in vehicular communication and can assist in identifying potential problems. The proposed approach utilizes ensemble learning to enhance the accuracy by integrating predictions from multiple FL models.

The proposed system has exhibited significant enhancement in performance in comparison to benchmark algorithms and existing techniques. This improvement is achieved by

Fig. 14 **a.** MSE of trained and tested data for the existing Fed-GRU Algorithm, **b**. MSE of trained and tested data for the existing Fed-NTP, **c**. MSE of trained and tested data for FLECO

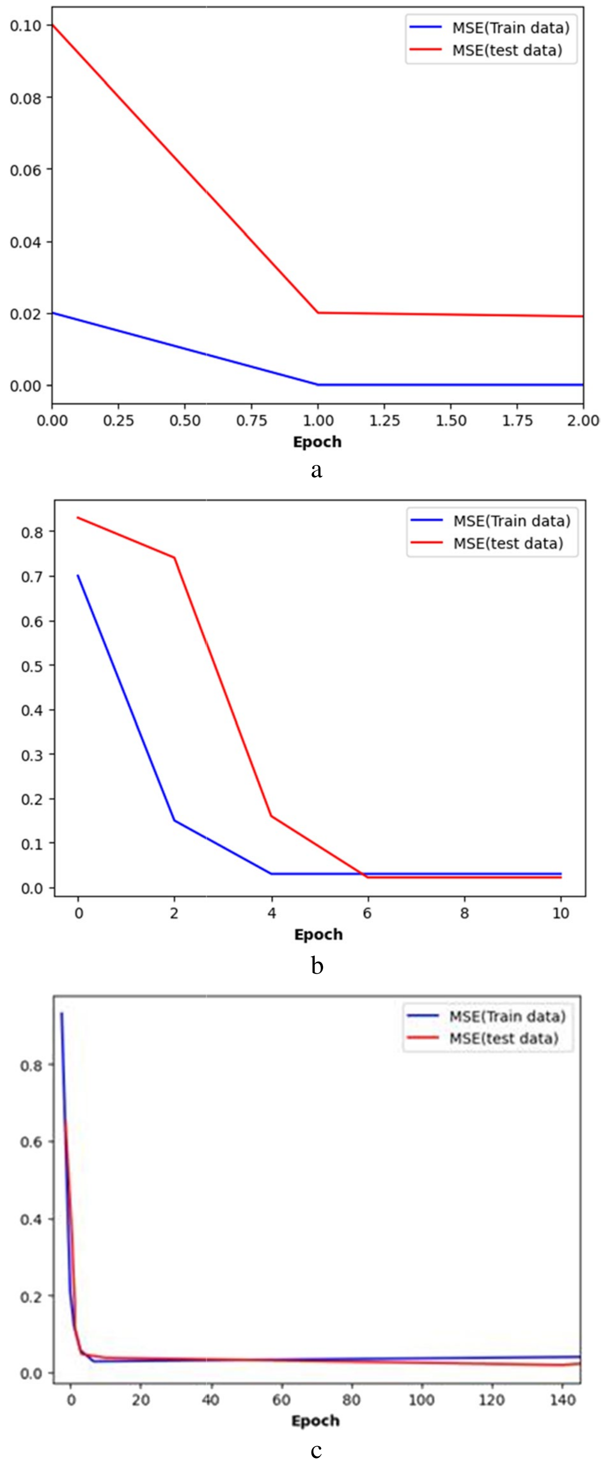


Fig. 15 Different methods were used to analyze the distributions of transmission delay for various sizes of V2V packets. (**a**) 10 V2V pairs. (**b**) 20 V2V pairs (**c**) 30 V2V pairs

using the FL technique for model training and the ensemble approach for integrating predictions from multiple FL models.

Compared to other existing techniques like VCP, AWCP, and AWCP-EWA, the proposed FLECO algorithm shows a high accuracy of approximately 97.5%, as depicted in the accuracy comparison chart of Fig. 12. The selection of the loss function in FL can vary based on the application's specific requirements and the stage of the FL process. The main objective is to achieve excellent model performance while maintaining the privacy and security of the decentralized data sources.

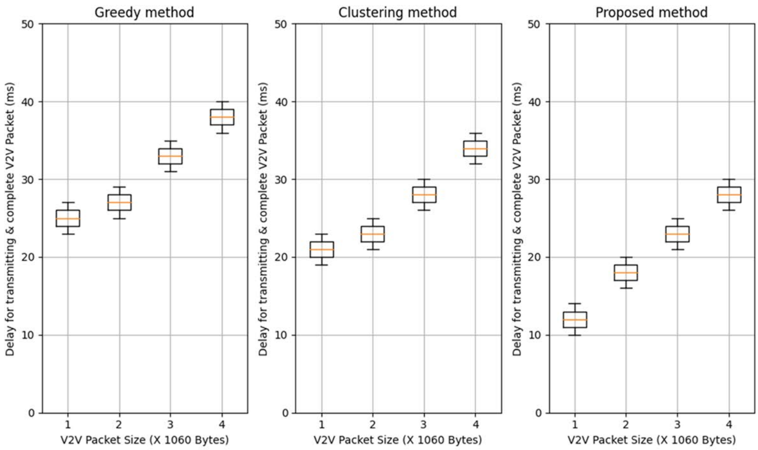
As the iteration increases, Fig. 13 demonstrates the variations in the loss function for all traffic, covering training data from both weekdays and weekends, across different maximum prediction timeframes. The decrease in loss functions for all three types of traffic data with an increase in communication rounds ensures the efficacy of the proposed multi-task FL framework. Additionally, it can be observed that the loss function for a shorter maximum prediction horizon is lower than the one for a longer maximum prediction timeframe. The decline in prediction performance can be attributed to the fact that, with a larger prediction horizon, a longer traffic speed series needs to be predicted, including traffic conditions in the future.

5.6 Mean square error of the existing model and the proposed model

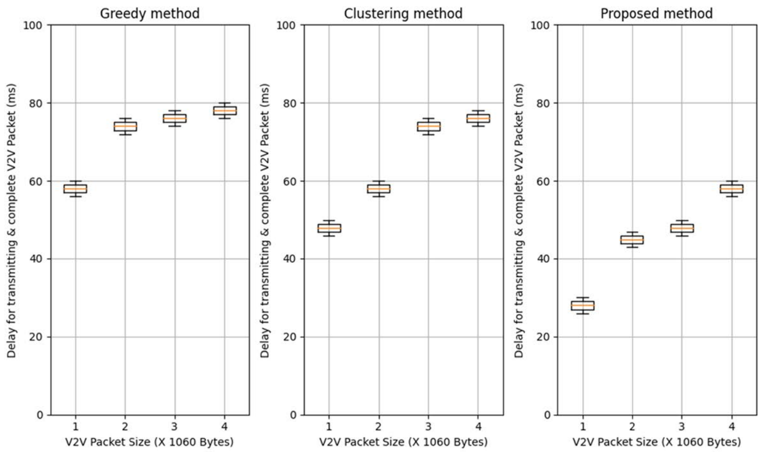
The accuracy of the communication system in vehicular communication is often evaluated using the Mean Square Error (MSE) metric. Vehicular communication refers to the transmission of information between vehicles and other entities to support applications such as safety and traffic management. The MSE metric in vehicular communication is used to determine the average squared difference between the transmitted and received data. The calculation involves finding the difference between the actual received signal and the expected received signal, squaring the difference, and averaging it over all the received signals. By measuring the amount of interference in the communication system, the MSE metric plays a critical role in the evaluation process. Achieving lower MSE values is critical for enhancing the reliability and effectiveness of vehicular communication systems as higher MSE values indicate a greater likelihood of errors and reduced accuracy of the received signal. To calculate MSE, it is necessary to train the model on a set of training data and then use it to generate predictions on a set of test data. Figure 14a illustrates the MSE of the data used for training and testing for the existing methodology.

In measuring the performance of a model in vehicular communication, the MSE for the training data determines how well the model fits it, while the MSE for test data evaluates how well the model generalizes to new, unseen data. A low MSE for both sets of data represents accurate prediction and a well-performing model. On the other hand, if the MSE for test data is considerably higher than that of the training data, it could imply that the model is overfitting to the training data, which leads to poor generalization of new data.

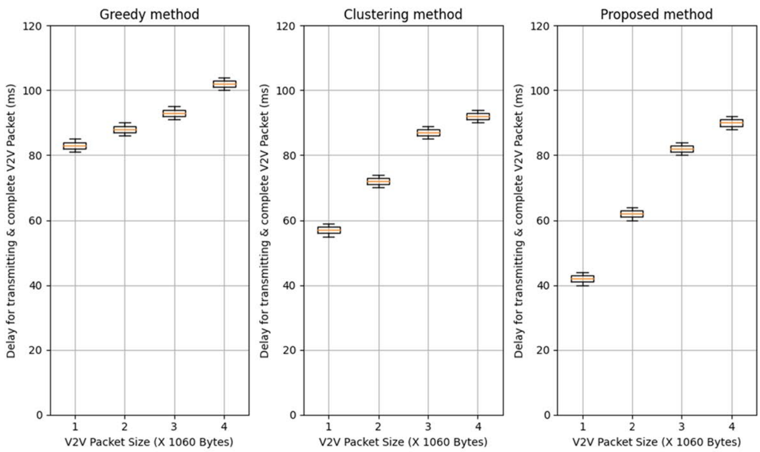
Measuring the MSE of a Fed-NTP model would assess the model's ability to generalize to new, unseen data by evaluating its performance on the test data. When the MSE is lower, it indicates that the model is capable of making more precise predictions, while a higher MSE may indicate that the model has limitations in generalizing to new data. Figure 14b


Greedy method
Clustering method
Proposed method
Delay for transmitting & complete V2V Packet (ms)
V2V Packet Size (X 1060 Bytes)

(a)


Greedy method
Clustering method
Proposed method
Delay for transmitting & complete V2V Packet (ms)
V2V Packet Size (X 1060 Bytes)

(b)


Greedy method
Clustering method
Proposed method
Delay for transmitting & complete V2V Packet (ms)
V2V Packet Size (X 1060 Bytes)

(c)

demonstrates that the MSE for the test data is notably higher than the MSE for the training data, indicating that the model is overfitting to the training data and not generalizing well to new data.

The federated learning approach distributes the training data across multiple devices, avoiding the need for centralized data storage. The model is trained collaboratively through local updates from individual devices and global updates from a central server. The final goal is to develop a model that can precisely forecast the target variable on new and unseen data, despite the distributed training data. Figure 14c shows that the testing and training data are similar, indicating that the model is stable and has a low MSE. Therefore, the evaluation of the proposed model suggests that it has high accuracy, low latency, and reliability, with a low loss metric.

In addition, Fig. 15 illustrates that when the number of V2V pairs remains constant, the transmission delays for all methods rise as the packet size increases. This is because larger V2V packets require more time for transmission and higher transmit power for V2V communication. Figure 15 presents multiple observations that serve to validate the advantages of our proposed method. The suggested technique has two distinct advantages: 1) its average transmission delay (represented by the colored horizontal line) is shorter than that of the other two ways; and 2) its transmission delay variations are smaller across all V2V pairs, as seen by the boxplots' vertically narrower gaps. The proposed algorithm is designed to handle different packet sizes effectively.

Furthermore, it can be observed from Fig. 15 that as the packet size increases, the transmission delays for all methods increase, while keeping the number of V2V pairs constant. The reason for this is that when V2V packets are larger, it takes longer to transmit them and requires a higher transmit power for V2V communication. The advantages of our proposed method are supported by multiple observations, as shown in Fig. 15. There are two main benefits to using the suggested technique. Firstly, the average transmission delay is shorter compared to the other two methods. This can be observed by looking at the colored horizontal line. Secondly, the transmission delay variations are smaller for all V2V pairs. This can be seen by the narrower gaps in the boxplots. The algorithm that has been suggested is specifically created to efficiently manage various sizes of packets.

5.7 Link stability analysis

The observed packets transmitted per second are depicted in Fig. 16, specifically in relation to the intersection (Int) and independent vehicles (ind). Consideration is given to the intersection of a highway and a road intersecting. The data indicates that when the link stability is high, the quality of packet transmission improves, but when the link stability is low, the quality of packet transmission decreases.

5.8 Impact of vehicle density

Figures 17, 18, and 19 are used to analyze the impact of vehicle density on message delivery ratio, average delay, and outage time. The proposed method has a higher message delivery ratio compared to the existing methods, as observed in Fig. 17. The probability of establishing a connection is an essential requirement for data or message delivery. The probability of establishing a connection is low when approaching an intersection with a vehicle coming from the opposite direction. As a result, the level of congestion increases.

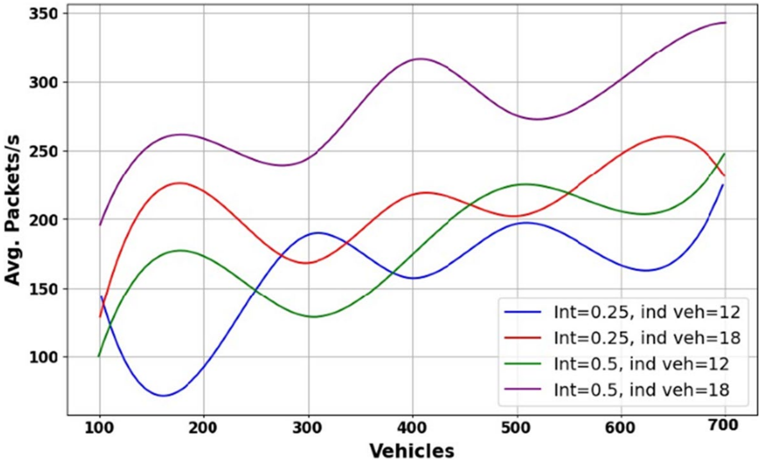


Fig. 16 Analysis of Intersection and Independent Vehicle for Avg. Packets/Sec

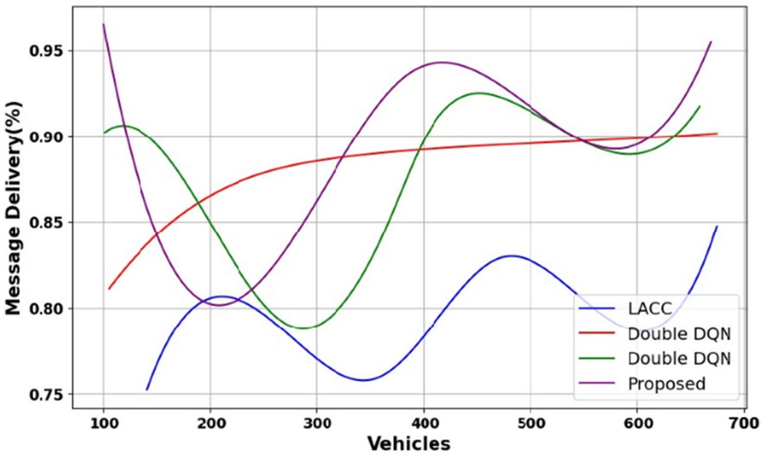


Fig. 17 Message Delivery versus Vehicles

The delay experienced in the network is influenced by the density of vehicles and how it affects message transmission. In this particular situation, two factors contribute to the delay: congestion and paused transmission. Congestion occurs when there is an excessive amount of data being transmitted, leading to a backlog and slower processing. Additionally, when the system is unable to handle the load, it can result in outages and further exacerbate the delay. Additionally, the contention window may be decreased in certain transmissions to address issues related to transmission losses. This opportunity allows for the transmission of a smaller amount of data or message. Delay is increased when the transmitting vehicle retains the messages that fall within the interval of "t_o" or "t_c—TTL_l". The proposed approach for linear optimization of metaheuristics involves filtering vehicles based on link stability and service capacity during data dissemination. In addition to these advantages, neighbor selection is restricted using metaheuristic routing, reducing the delay. The

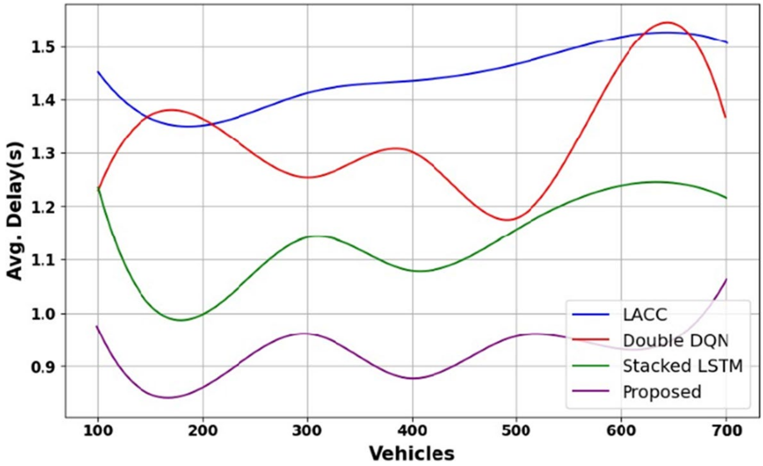


Fig. 18 Average Delay versus Vehicles

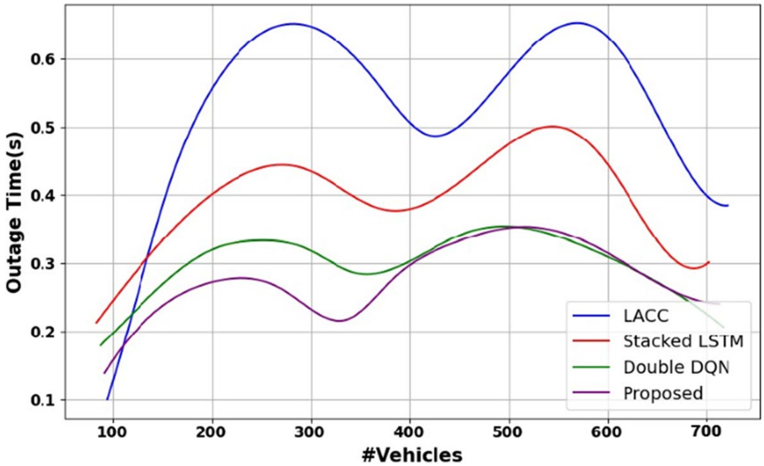


Fig. 19 Outage Time versus Vehicles

sequential operation reduces delay in FLECO. The neighbor selection process is similar in encountering any number of vehicles (Fig. 18).

Figure 19 illustrates the comparison of outage times between the proposed methods and the existing methods. The metric for measuring link stability is selected by considering the duration of outages. The estimated outage, which is calculated as (t_c—TTL_l), applies to vehicles that are traveling in the same lane and the same direction. On the other hand, the outage (t_c—TTL_l) refers to the longest duration of link disconnection at a point where lines intersect.

Analyzing the Dissemination vs. Message Delivery graph, which compares the performance of the proposed model against LACC, Double DQN, and Stacked LSTM, it is evident that the proposed model stands out as efficient and effective in disseminating messages. Figure 20 illustrates that as the number of disseminated messages increases from 100 to 700, the proposed model consistently achieves a higher message delivery rate compared

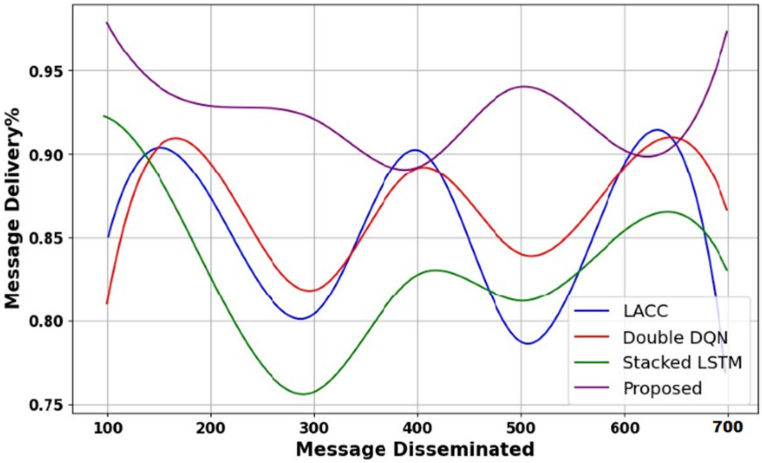


Fig. 20 Message Delivery versus Message Disseminated

to the other techniques—LACC, Double DQN, and Stacked LSTM. This demonstrates the superior efficiency and effectiveness of the proposed model in delivering messages across the varying dissemination range. The proposed model showcases a robust ability to disseminate a higher number of messages while maintaining a commendable message delivery rate. This efficiency is particularly vital in vehicular communication, where timely and reliable message delivery is crucial for effective coordination and safety on the road.

5.9 Cost of computation

Computation cost encompasses both the latency of computation and the latency of vehicle communication. The time it takes to form groups, have vehicles join them, and perform other necessary computations for implementing privacy schemes contributes to the computation latency. The time it takes for vehicles to communicate with each other to establish a privacy protection scheme for vehicle behavior is known as communication latency. Table 2 displays the average communication latency at various levels of traffic densities. The results in computation and communication latency are improved by our proposed scheme.

Table 2 Communication costs at a different number of vehicles

Number of Vehicles	Average communication cost (ms)			
	Proposed	LACC	Stacked LSTM	Double DQN
100	21.32	32.67	45.12	55.61
200	23.56	25.45	34.65	45.62
300	31.35	35.31	43.12	64.32
400	43.12	55.34	60.45	75.18
500	56.32	63.24	68.23	79.34
600	67.25	75.64	89.76	101.25
700	70.34	80.23	97.45	120

6 Conclusion

The implementation of federated learning in vehicular communication offers a promising solution to enhance the performance of ITS and provide better services to drivers and passengers. By facilitating the exchange of data and collaborative training of machine learning models among vehicles, this technology ensures the privacy of sensitive information. The application of FL in vehicular communication can enhance the accuracy of predictive models used in ITS, including traffic flow prediction and vehicle routing. This, in turn, can minimize traffic congestion, improve safety, and optimize the efficiency of transportation systems. In addition, the approach of federated learning can tackle concerns regarding privacy and security associated with the sharing of sensitive information in vehicular communication. The decentralized nature of federated learning, where data is stored locally on each vehicle and only model updates are exchanged, minimizes the chances of data exposure and malicious attacks.

Data availability No data generated during this study.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Li Y, Tao X, Zhang X, Liu J, Xu J (2021) Privacy-preserved federated learning for autonomous driving. IEEE Trans Intell Transp Syst 23(7):8423–8434
2. Yang H, Zhao J, Xiong Z, Lam KY, Sun S, Xiao L (2021) Privacy-preserving federated learning for UAV-enabled networks: Learning-based joint scheduling and resource management. IEEE J Sel Areas Commun 39(10):3144–3159
3. Wei S, Zou Y, Zhang X, Zhang T, Li X (2019) An integrated longitudinal and lateral vehicle following control system with radar and vehicle-to-vehicle communication. IEEE Trans Veh Technol 68(2):1116–1127
4. Sattler F, Wiedemann S, Müller KR, Samek W (2019) Robust and communication-efficient federated learning from non-iid data. IEEE Trans Neural Netw Learning Syst 31(9):3400–3413
5. Guo C, Liang L, Li GY (2019) Resource allocation for low-latency vehicular communications: An effective capacity perspective. IEEE J Sel Areas Commun 37(4):905–917
6. Singh S, Saraswat A, Yadav S (2019) Traffic simulation integration using SUMO simulator. Int J Sci Res Rev 7(2):22–26
7. Bagheri M, Bartin B, Ozbay K (2022) Simulation of vehicles' gap acceptance decision at unsignalized intersections using SUMO. Procedia Comput Sci 201:321–329
8. Lyu F, Zhu H, Cheng N, Zhou H, Xu W, Li M, Shen X (2019) Characterizing urban vehicle-to-vehicle communications for reliable safety applications. IEEE Trans Intell Transp Syst 21(6):2586–2602
9. Manivannan D, Moni SS, Zeadally S (2020) Secure authentication and privacy-preserving techniques in Vehicular Ad-hoc NETworks (VANETs). Vehicular Communications 25:100247
10. Song HM, Woo J, Kim HK (2020) In-vehicle network intrusion detection using deep convolutional neural network. Veh Commun 21:100198
11. Mirsadeghi F, Rafsanjani MK, Gupta BB (2021) A trust infrastructure based authentication method for clustered vehicular ad hoc networks. Peer-to-Peer Netw App 14:2537–2553
12. Lakshmanaprabu SK, Shankar K, Rani SS, Abdulhay E, Arunkumar N, Ramirez G, Uthayakumar J (2019) An effect of big data technology with ant colony optimization based routing in vehicular ad hoc networks: Towards smart cities. J Clean Prod 217:584–593
13. Ai Y, Felipe AP, Kong L, Cheffena M, Chatzinotas S, Ottersten B (2021) Secure vehicular communications through reconfigurable intelligent surfaces. IEEE Trans Veh Technol 70(7):7272–7276

14. Ali I, Lawrence T, Omala AA, Li F (2020) An efficient hybrid signcryption scheme with conditional privacy-preservation for heterogeneous vehicular communication in VANETs. IEEE Trans Veh Technol 69(10):11266–11280
15. Ali ES, Hasan MK, Hassan R, Saeed RA, Hassan MB, Islam S, ... Bevinakoppa S (2021) Machine learning technologies for secure vehicular communication in internet of vehicles: recent advances and applications. Secur Commun Netw 1–23
16. Ye H, Li GY, Juang BHF (2019) Deep reinforcement learning based resource allocation for V2V communications. IEEE Trans Veh Technol 68(4):3163–3173
17. He H, Wang Y, Han R, Han M, Bai Y, Liu Q (2021) An improved MPC-based energy management strategy for hybrid vehicles using V2V and V2I communications. Energy 225:120273
18. Tahir MN, Leviäkangas P, Katz M (2022) Connected vehicles: V2V and V2I road weather and traffic communication using cellular technologies. Sensors 22(3):1142
19. Bao X, Li H, Zhao G, Chang L, Zhou J, Li Y (2020) Efficient clustering V2V routing based on PSO in VANETs. Measurement 152:107306
20. Qin P, Fu Y, Feng X, Zhao X, Wang S, Zhou Z (2021) Energy-efficient resource allocation for parked-cars-based cellular-V2V heterogeneous networks. IEEE Internet Things J 9(4):3046–3061
21. Samarakoon S, Bennis M, Saad W, Debbah M (2019) Distributed federated learning for ultra-reliable low-latency vehicular communications. IEEE Trans Commun 68(2):1146–1159
22. Valayapalayam Kittusamy SR, Elhoseny M, Kathiresan S (2022) An enhanced whale optimization algorithm for vehicular communication networks. Int J Commun Syst 35(12):e3953
23. Sepasgozar SS, Pierre S (2022) Fed-NTP: A Federated Learning Algorithm for Network Traffic Prediction in VANET. IEEE Access 10:119607–119616
24. Arya M, Sastry H, Dewangan BK, Rahmani MKI, Bhatia S, Muzaffar AW, Bivi MA (2023) Intruder Detection in VANET Data Streams Using Federated Learning for Smart City Environments. Electronics 12(4):894
25. Taïk A, Mlika Z, Cherkaoui S (2022) Clustered vehicular federated learning: Process and optimization. IEEE Trans Intell Transp Syst 23(12):25371–25383
26. Yu Z, Hu J, Min G, Zhao Z, Miao W, Hossain MS (2020) Mobility-aware proactive edge caching for connected vehicles using federated learning. IEEE Trans Intell Transp Syst 22(8):5341–5351
27. Ge X (2019) Ultra-reliable low-latency communications in autonomous vehicular networks. IEEE Trans Veh Technol 68(5):5005–5016
28. Yang M, Ai B, He R, Chen L, Li X, Li J, Zhong Z (2019) A cluster-based three-dimensional channel model for vehicle-to-vehicle communications. IEEE Trans Veh Technol 68(6):5208–5220
29. Al-Shareeda MA, Anbar M, Manickam S, Yassin AA (2020) Vppcs: Vanet-based privacy-preserving communication scheme. IEEE Access 8:150914–150928
30. Zhang D, Zhang T, Liu X (2019) Novel self-adaptive routing service algorithm for application in VANET. Appl Intell 49:1866–1879
31. Yang Y, Gao Z, Ma Y, Cao B, He D (2020) Machine learning enabling analog beam selection for concurrent transmissions in millimeter-wave V2V communications. IEEE Trans Veh Technol 69(8):9185–9189
32. Zhao D, Qin H, Song B, Zhang Y, Du X, Guizani M (2020) A reinforcement learning method for joint mode selection and power adaptation in the V2V communication network in 5G. IEEE Trans Cogn Commun Netw 6(2):452–463

Authors and Affiliations

Ruchi Gupta[1] · **Tanweer Alam**[2]

Tanweer Alam
tanweer03@iu.edu.sa

[1] Faculty of Information Technology, Ajay Kumar Garg Engineering College Ghaziabad, Abdul Kalam Technical University, Lucknow, India

[2] Faculty of Computer and Information Systems, Islamic University of Madinah, Madinah, Saudi Arabia