JOURNAL OF SOFT COMPUTING AND DATA MINING
e-ISSN: 2716-621X

JSCDM

Vol. 6 No. 3 (2025) 416-428
https://publisher.uthm.edu.my/ojs/index.php/jscdm

Homomorphic Encryption-Enabled Federated Transfer Learning for Privacy-Preserving of Internet of Vehicles

Yousif Khalid Yousif [1], Shahad jasim hasan[2], Ihab Yasien Mahmood[3] , Ali Q Saeed[1], Mohammed Tanash[4]*, Tarik AbuAin[5], Waleed AbdelKarim Abuain[6]

[1] *Department of Cloud Computing and IoT Techniques Engineering, Technical Engineering College for Computer and AI, Northern Technical University, Mosul, 41000, Nineveh, IRAQ*
[2] *University Presidency-Electronic Computer Center, University of Babylon, Babylon, IRAQ*
[3] *Computer Science Department, College of Education for Pure Sciences, University of Mosul, Mosul, IRAQ*
[4] *Holland Computing Centre, University of Nebraska-Lincoln, Lincoln, USA*
[5] *College of Computing and Informatics, Saudi Electronic,*
University, Riyadh 11673, KINGDOM OF SAUDI ARABIA
[6] *College of Science and Computer Engineering, Yanbu,*
Taibah University, Yanbu, Al-Madinah Al-Munawwarah, 42353, KINGDOM OF SAUDI ARABIA

*Corresponding Author: mtanash2@unl.edu
DOI: https://doi.org/10.30880/jscdm.2025.06.03.026

Article Info

Received: 3 July 2025
Accepted: 15 December 2025
Available online: 30 December 2025

Keywords

Federated transfer learning; homomorphic encryption, IoV, privacy preservation, heterogeneous data fusion, domain adaptation

Abstract

The swift incorporation of intelligent cars into the Internet of Vehicles (IoV) has contributed to the creation of vast quantities of various and sensitive data, which has caused serious concerns about the privacy and security of data. The conventional centralized learning models tend to face major difficulties associated with loss of privacy, transmission delay, and data handling. In the meantime, the current Federated Learning (FL) solutions are often unable to be flexible enough when knowledge is transferred into different heterogeneous vehicular settings. In order to overcome these drawbacks, this paper introduces a Privacy-Preserving Federated Transfer Learning (PP-FTL) model that can smoothly combine transfer learning with secure federated model updates based on homomorphic encryption. PP-FTL model specifically targets data fusion in IoV networks to provide the overall privacy protection and allow the cross-vehicle collaborative learning, even when there is heterogeneous and distributed data. The model uses an aggregation of encrypted weights and a knowledge distillation process that enables successful adaptation between vehicles. Experimental tests based on real-life vehicular network data prove that the suggested PP-FTL model outperforms the baseline FL models in terms of classification, speed of convergence, and preservation of privacy. System attains an average accuracy of 95.7%, a 30% reduction in overheads of communication, and no data leakage, which proves the appropriateness of the system in real-time applications and privacy-conscious IoV applications.

1. Introduction

Internet of Vehicles (IoV) is an essential part of Intelligent Transportation Systems (ITS) of recent times, and a smooth way of connecting vehicles with roadside infrastructure and cloud platforms using the tools of real-time communication and connectivity. The advent of intelligent sensors, 5G networks, and edge computing has enabled vehicles to keep on producing an enormous volume of diverse data feeds, such as traffic dynamics, vehicle telemetry, environmental metrics, and driver behavioral patterns [1, 2]. Such sources of data are priceless when creating a Machine Learning (ML) model to resolve essential issues, like traffic optimization, predicting accidents, self-driving, and the health of the vehicle. Nevertheless, the sensitivity of such information raises grave privacy, data protection, and regulatory compliance issues, especially when it comes to models such as the General Data Protection Regulation (GDPR) [3, 4].

In order to reduce those problems, FL is one of the proposed decentralized learning paradigms through which collaborative model training is achievable without having to transfer raw data to a central server. Rather, the local models are trained on the local data of the individual participating nodes, which are a vehicle or a roadside unit, and exchange only encrypted model updates with a central aggregator [5, 6]. Although the given approach increases the level of privacy protection, there are certain challenges in the realm of the IoV. The first restriction is the non-independence and identically distributed (non-IID) property of vehicular data, which is caused by the differences in geographic position, traffic density, hardware settings, and driving behavior, and may be a hindrance to model performance and convergence stability [7, 8]. More to the point, the security vulnerabilities have not yet been fully resolved because aggressive parties, or predatory aggregators, can still distill sensitive data out of shared model parameters [9, 10]. Moreover, the conventional FL models presuppose that all clients learn the same global goal, which is difficult to expect in the diverse vehicular conditions.

In the suggested model, every client vehicle initially trains local models on its own data, and thus keeps data privacy at the edge. Homomorphic Encryption (HE) is then used to encrypt the trained model parameters to provide data confidentiality when sending over the network. The encrypted updates are then sent to a federated aggregator, which then does secure aggregation but not decryption. The output is the final result that is an excellent global model integrating distributed vehicular intelligence without violating privacy, and provides secure, adaptive learning over the IoV ecosystem. This global model is further optimized with the help of TL methods to promote its generalization to clients with different data distributions. Lastly, the revised model is shared with the clients, allowing for a process of constant learning and adaptation in a privacy-aware and collaborative manner.

To overcome such limitations, TL has been pursued as a process to improve model generalization across one or more domains by leveraging knowledge from one domain (source) to another (target) [11, 12]. Federated TL (FTL) enables participation by participants with varied feature spaces or distributions in the FL context. Nonetheless, even the application of FTL does not eliminate the threats to the model parameters to the extent that they are not sufficiently safeguarded. To establish a system that supports the secure, accurate, and adaptive learning process for heterogeneous IoV devices, this paper proposes a new model, Privacy-Preserving Federated TL (PP-FTL), that integrates the merits of FL, TL, and Homomorphic Encryption (HE). The fundamental contributions of this work consist of:

- Safe knowledge sharing: Homomorphic encryption is utilized in order to protect parameter communication such that aggregate gradients can be computed without access to encrypted data.
- Domain adaptation: The TL module uses knowledge distillation to reduce the gap in distributions between nodes that have dissimilar feature spaces and data distributions.
- Reduced overhead: The proposed IoV architecture is optimized to minimize communication cost and latency, making it feasible for real-time vehicular deployment.

Section 2 presents a review of related research in FL, TL, and privacy-preserving techniques in IoV. Section 3 outlines the problem formulation and research objectives. Section 3 details the proposed PP-FTL model along with its architectural components, including the mathematical model. Section 4 presents the simulation results and comparative analysis. Finally, Section 5 concludes the paper with future directions.

2. Related Works

The recent improvements in FL, TL, and privacy-saving methods have now broadened the opportunities for smart processing of data in heterogeneous IoV environments. The classic centralized ML models are effective in a small application but extremely threatening to privacy, and their scaling ability is limited to large vehicular systems. Meanwhile, FL has become a decentralized model of learning so that distributed clients can collectively learn models without sharing raw data, thereby promoting greater data confidentiality and user privacy [1, 13].

Although these advantages exist, FL faces significant challenges in handling data heterogeneity, especially when clients vary in data distribution, including their feature spaces or learning objectives. This is more pronounced in the case of IoV, where data gathered by vehicles vary depending on geographic, environmental, and contextual considerations [14, 15]. TL offers an approach that shows promise by transferring knowledge into a similar yet non-identical domain, thereby enhancing generalization and flexibility [16]. The combination of TL with FL has demonstrated positive outcomes with respect to cross-domain learning, despite the issues surrounding data security, communication performance, and trust handling [6].

Encryption protocols like DP [17] and HE have been extensively used to improve privacy. Remarkably, HE supports arithmetic operations on encrypted messages, and thus safe aggregation is possible without sensitive data being exposed [7, 18]. More recent literature (e.g., [10]) has also shown that in applications like healthcare and finance, HE can be used to create privacy-sensitive distributed intelligence systems in combination with FL. Nonetheless, the use of such methods in the IoV is difficult because of latency, processing delays, and structural complexity.

There are a number of high-profile studies that have contributed to this field. As an example, [11] proposed a hybrid FL model with partial HE of smart healthcare, and [12] suggested an adaptive meta-learning model of knowledge transfer in federated settings. All concurrently, these attempts point to the rising possibilities of combining FL, TL, and encryption schemes in order to realize secure, adaptive, and scalable learning in smart vehicle systems. However, both lacked robust mechanisms for handling highly heterogeneous vehicular data. In contrast, our proposed model addresses this limitation by integrating privacy-preserving HE mechanisms with domain-adaptive transfer learning under a unified federated architecture, specifically designed for heterogeneous IoV contexts. Table 1 summarizes a comparative analysis of selected recent works focusing on FL, TL, HE, and their combinations in privacy-aware applications, highlighting the gaps addressed by our proposed model.

Table 1 *Comparison of existing works on FL, TL, and HE in privacy-preserving systems*

Ref.	Methodology	Use Case	Data Heterogeneity Handling	Privacy Technique	Limitations
[11]	FL with partial HE	Smart Healthcare	Limited	Partial Homomorphic- Enc.	High computation,
[12]	No domain transfer	Edge Computing	Moderate	No encryption	Lack of privacy protection
[19]	FL + TL via meta-learning	Finance Sector	None	Full Homomorphic Enc.	No domain adaptation
[20]	HE-enabled federated averaging	Smart Grid	Partial	DP	Accuracy loss with noise injection
This work	Multi-task FL	IoV	High	Full Homomorphic Enc.	Secure, adaptive, high accuracy

This has been recently studied by researchers who have proposed to use federated learning in vehicular networks and eliminate the issue of data privacy that is related to sensor-rich intelligent transportation systems. A prominent contribution is by Lu et al. [21], who suggested a hierarchical federated learning architecture of autonomous driving, where every vehicle trains on LiDAR and camera data and sends encrypted updates of their models to a global aggregator. This paper has shown that federated learning can significantly reduce privacy leakage, with a high detection accuracy on road events. However, the suggested framework lacked the use of homomorphic encryption in the process of model aggregation and made the system susceptible to inference attacks at the server-side, especially in highly heterogeneous IoV settings where non-IID data distributions only increase the chance of reconstruction attacks.

At the same time, homomorphic encryption has been actively embraced to improve the secure sharing of models in the distributed vehicular system. A framework of collaborative learning, in which vehicles broadcast encrypted gradient updates to predict traffic incidents, was presented by Zhang et al. [22], which is also HE-enabled. Their design was effective in withholding access to raw vehicle properties, such as CANBus signals and visual descriptors, and, as such, demonstrates the potential of HE for privacy-preserving analytics. Nonetheless, the test was conducted on stationary smart city sensors rather than on multi-modal and mobile IoV clients, and the system did not have a way to address non-IID sensor imbalance among vehicles. These weaknesses in managing data heterogeneity highlight the need for HE-federated transfer learning that can effectively learn across different configurations of vehicle sensors.

3. Methodology

The proposed model integrates FL, TL, and HE to address the aforementioned challenges in IoV. In the IoV architecture, multiple vehicular clients train local models using their private data without sharing it with a centralized server. These clients may have distinct data distributions and feature spaces due to domain variability. To enable inter-client knowledge sharing, we apply TL using domain adaptation techniques that minimize distribution shifts across different client datasets. The architectural flow of the proposed Privacy-Preserving Federated Transfer Learning (PP-FTL) model for heterogeneous IoV environments is illustrated in Figure 1.

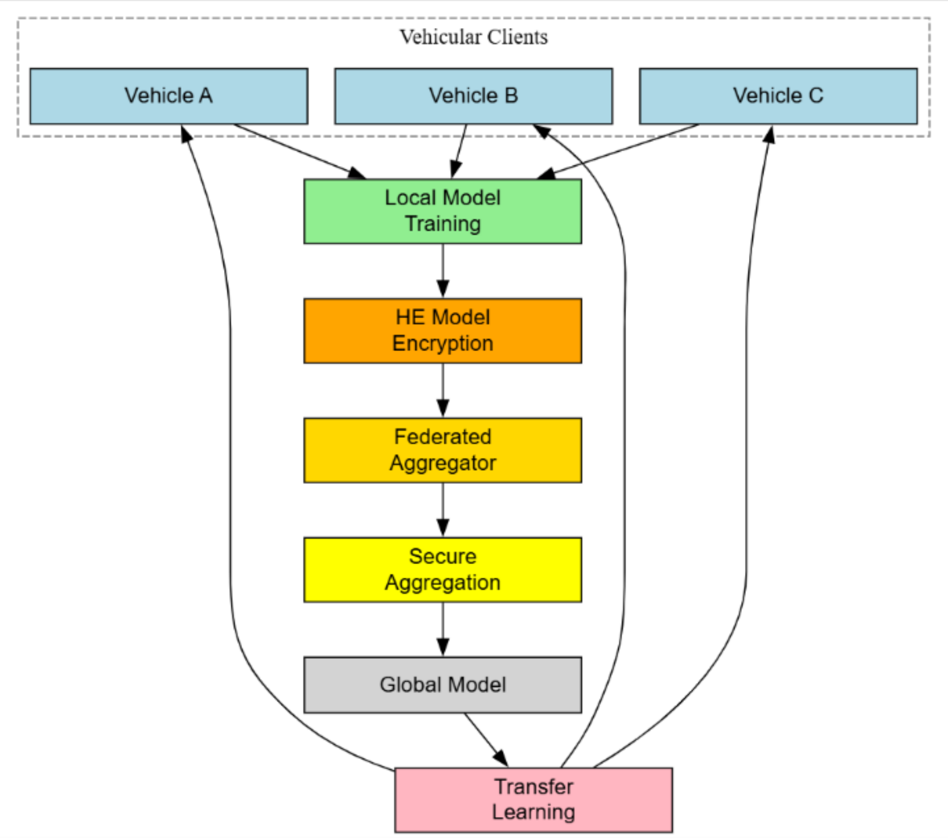

Fig. 1 *Compact architecture of the proposed PP-FTL model for IoV*

The IoV-based PP-FTL model offers an efficient and secure model of collaborative learning among connected vehicles without sharing sensitive data. As illustrated in Fig. 1, the model starts at the vehicular client layer, where vehicles (B, C, and A) gather the various local driving data using the onboard sensors and edge modules. The local model training of each vehicle is done independently in order to learn regularities, like traffic prediction, anomaly detection, or driver behaviour, and no raw data is sent to the central server. This decentralized model reduces communication load and enables efficient exploitation of heterogeneous data distribution across different automobiles.

As shown in Figure 1, the model parameters of each vehicle are transmitted and aggregated using the model parameters of both vehicles, with each vehicle's model parameters encrypted with an HE that guarantees secure model weights or gradients. These encrypted updates are forwarded to the federated aggregator, which aggregates them without decryption. Because HE is a mathematical machine that can perform calculations on encrypted data, it enables the system to combine model updates safely without revealing sensitive information. This process is further enhanced by the fact that the contribution of any single client cannot be reverse-engineered or compromised, ensuring end-to-end confidentiality even in an untrusted network by simply including a secure aggregation protocol.

After the safe aggregation, the new global model is released to the rest of the vehicular clients. TL is implemented as a complementary mechanism to increase adaptability across contrasting vehicular contexts and heterogeneous data distributions. TL also allows the global model to transfer learned features and patterns across domains, including urban-to-rural environments, thereby enhancing its generalization and ability to operate effectively. FL, HE, and TL integration, therefore, creates a full-fledged privacy-guaranteeing learning scheme that provides safe knowledge sharing, scalability, and reliability in the IoV ecosystem.

Mathematically, each client $i \in \{1,2,\ldots,N\}$ possesses a local dataset $Di = \{x_{ij}, y_{ij}\}$, where x_i denotes input features and $y_i \in$R the corresponding labels. The regional model $fi(.)$ is trained by minimizing the empirical loss:

$$\mathcal{L}_i(\theta_i) = \frac{1}{n_i}\sum_{j=1}^{n_i} \mathrm{l}\left(f_i\left(x_i^j;\theta_i\right), y_i^j\right) \qquad (1)$$

where $\ell(\cdot)$ is a loss function such as Mean Squared Error (MSE) or cross-entropy. After local training, each client encrypts its model update $\Delta\theta_i$ using a homomorphic encryption scheme $E\ (\cdot)$, such that the server receives only $E(\Delta\theta_i)$. Aggregation at the server is done in the encrypted domain using:

$$\mathcal{E}(\bar{\theta}) = \frac{1}{N}\sum_{i=1}^{N}\mathcal{E}(\Delta\theta_i) \tag{2}$$

The homomorphic property ensures:

$$\mathcal{E}(a) + \mathcal{E}(b) = \mathcal{E}(a+b), \quad \forall a, b \in R \tag{3}$$

Once aggregated, the server sends back the global encrypted model $\mathcal{E}(\bar{\theta})$ to the clients, who decrypt and apply the updated model. To handle domain heterogeneity, we employ feature alignment using Maximum Mean Discrepancy (MMD) minimization, expressed as:

$$\text{MMD}^2(\mathcal{D}_s, \mathcal{D}_t) = |\frac{1}{n_s}\sum_{i=1}^{n_s}\phi(x_i^s) - \frac{1}{n_t}\sum_{j=1}^{n_t}\phi(x_j^t)|^2 \tag{4}$$

where D_s and D_t are the source and target distributions, and ϕ is the kernel mapping function. The goal is to minimize MMD alongside local loss:

$$\mathcal{L}ttl = \mathcal{L}_i(\theta_i) + \lambda \cdot MMD^2(\mathcal{D}_s, \mathcal{D}_t) \tag{5}$$

where λ is a balancing parameter.

Algorithm 1 shows the input, output, and process steps of the PP-FTL model for privacy-preserving FTL in IoV. The PP-FTL algorithm is a synchronized round algorithm, in which all vehicles within the IoV ecosystem train a Deep Learning (DL) model locally using sensor measurements, e.g., speed, route, and environmental measurements. Vehicles do not transfer raw data; instead, they homomorphically encrypt their trained model parameters to make IoV-related decisions and transmit them to a central aggregator. The aggregator is used to perform safe calculations on encrypted data to produce a global model without decrypting the individual data. TL is then used to adapt this global model to diverse vehicular environments (e.g., highway vs. city driving). The last step is the redistribution of the refined model to every vehicle, which ensures continuous learning, privacy protection, and flexibility across heterogeneous driving scenarios.

```
Algorithm 1: PP-FTL in IoV
Input: Local vehicular datasets $D_i$, learning rate $\eta$, number of communication rounds $R$;
Output: Global Transfer-Learned Model $M_g$;
Start-process:
    1. Initialization: The central server initializes a global model $M_g^0$ and distributes it to all participating vehicles $V_i$;
    2. Local Training: Each vehicle $V_i$trains $M_g^{r-1}$on its local dataset $D_i$to obtain an updated model $M_i^r$;
    3. Homomorphic Encryption: Each vehicle encrypts its trained model parameters $Enc(M_i^r)$using an HE scheme;
    4. Federated Aggregation: The encrypted updates $Enc(M_i^r)$are transmitted to the aggregator, which performs homomorphic summation $Enc(M_g^r) = \sum Enc(M_i^r)/n$without decryption;
    5. Secure Aggregation: The aggregated encrypted model is decrypted securely by the trusted server to obtain the global model $M_g^r$;
    6. TL Adaptation: The global model $M_g^r$undergoes fine-tuning through TL to adapt its knowledge to different vehicular environments;
    7. Iteration: Steps 2–6 are repeated for $R$rounds until convergence;
End-process.
```

The proposed PP-FTL model combines three effective options, such as FL, homomorphic encryption, and transfer learning, to allow safe, adaptable, and smart data interaction within the IoV. Its significant attributes are decentralized local training, in which the sharing of raw data is eliminated, encrypted model transmission, and cross-domain flexibility that enables knowledge transfer across heterogeneous vehicular settings. The model is applicable to ITS applications such as traffic flow prediction, accident detection, route optimization, and vehicle anomaly monitoring, where real-time learning and data confidentiality are imperative. The model will enable vehicles to cooperatively train models without sacrificing privacy, thereby improving system intelligence, data privacy, and model generalization, resulting in quicker decision-making, reduced communication overhead, and overall efficiency, safety, and resilience in contemporary ITS.

4. Simulation and Results

To determine the effectiveness of the presented privacy-saving federated TL system augmented with HE for heterogeneous data fusion in the IoV, extensive simulations were conducted using a synthetically generated dataset. We offer a combination of a virtual world based on the CARLA platform to run high-fidelity vehicle and sensor simulation. It is integrated with the SUMO traffic simulator to recreate large-scale traffic flow and mobility control, and is managed using ROS, which serves as the middleware coordinating LiDAR, RGB camera, and CAN-Bus message streams. The client-server framework of CARLA allows creating various heterogeneous clients, each of which is a simulated instance of a real IoV with a specific sensor setup, a variety of feature dimensions, and a certain environmental scenario. The simulation design's main components for evaluating the PP-FTL model are shown in Figure 2.

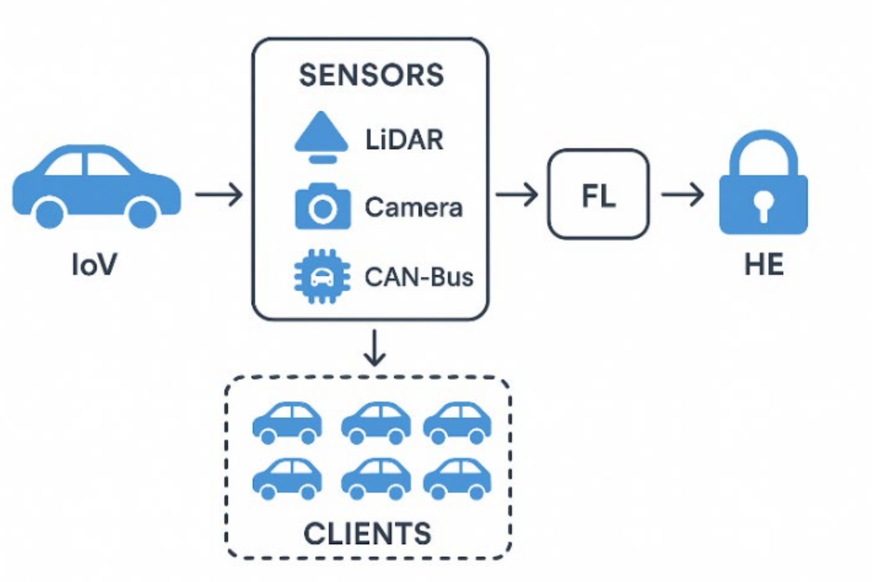

Fig. 2 *Simulation design*

SUMO simulator provides realistic microscopic mobility behaviour for the IoV environment. Thus, a faithful reenactment of traffic congestion, intersection conflicts, and dynamic route-changing events can be reenacted, which, in its turn, affects the distribution of labels (normal versus incident). Information about each simulated vehicle is recorded via ROS, preprocessed, and sent to the federated learning pipeline, where FTL with HE using the Microsoft SEAL or TenSEAL libraries updates encrypted models, keeping the original sensor data confidential. The specified configuration incorporates actual IoV heterogeneity, environmental bias, sample volume dissimilarity, and non-IID label distribution, making it applicable to assessing privacy-preserving FTL algorithms in a multi-client vehicular setting. Table 2 shows the simulation-related settings.

Table 2 *Simulation setting*

Component	Specification	Purpose
CPU	Intel Xeon / AMD EPYC, 16–32 cores	Runs multiple simulation clients and CARLA server
GPU	NVIDIA RTX 3080/4090 or A100	Accelerates rendering, sensor simulation, and DL training
RAM	64–256 GB DDR4/DDR5	Required for LiDAR and image-based simulation workloads
Storage	1–2 TB NVMe SSD	Stores sensor logs, LiDAR frames, and simulation scenarios
Network	1–10 Gbps LAN with added latency simulation	Emulates IoV communication delays and packet loss
Operating System	Ubuntu 20.04/22.04 LTS	Compatible with CARLA, ROS, PyTorch, and SUMO
Simulation Engine	CARLA Simulator (Unreal Engine)	Generates LiDAR, camera streams, and vehicle behavior
Traffic Simulator	SUMO + CARLA-SUMO co-simulation	Realistic traffic density, routing, and mobility
Middleware	ROS Noetic / ROS2 Foxy	Synchronizes LiDAR, camera, CAN-Bus, and publishes topics
DL	PyTorch or TensorFlow	Model training and federated transfer learning
FL	Flower or TensorFlow Federated (TFF)	Multi-client FL orchestration
HE	Microsoft SEAL / TenSEAL	Secure, encrypted aggregation of model updates
Data Tools	Pandas, NumPy, OpenCV, PCL	Preprocessing of multi-sensor IoV data

The simulation involves the presence of six client nodes (Client 1-Client 6) that are modelled to their real-life (i.e., heterogeneity generally found in an IoV setting), as shown in Table 3. The change in feature dimensions is due to differences in sensor types and deployments (LiDAR, camera, and CAN bus sensors), and the change in label distributions is intended to explain environmental bias and the imbalance in incident label distributions. This natural heterogeneity creates major problems for traditional FL algorithms, making it important to consider a domain-adaptive TL model that can successfully transfer and generalize knowledge across various vehicular situations.

Table 3 *Synthetic dataset used for simulation*

Client ID	Feature Dimension	Number of Samples	Label Distribution
Client 1	20	1000	Balanced
Client 2	25	1200	Skewed
Client 3	18	950	Uniform
Client 4	22	1100	Normal
Client 5	30	1050	Skewed
Client 6	15	980	Balanced

The dataset that is used in this study is a completely synthetic and simulation-based set that is designed to simulate real operating conditions of the IoV. The dataset represents a large scale of vehicular conditions and sensor arrangements spread out among several client nodes, as shown in Table 4. Every node denotes a simulated vehicle or a group of vehicles with various sensors that produce non-IID-based data, which exhibit variance in both dimensions of features as well as distribution of labels. It was run with data produced with many virtual vehicles with a different set of LiDAR, RGB camera, and CAN-Bus sensors, thus forming heterogeneous feature spaces that reflect differences in sensor hardware and sampling properties that can be seen in the field. All vehicles generate a composite feature representation that includes spatial LiDAR features, visual features based on camera frames, and vehicle state features based on CAN-Bus messages. Since all simulated client nodes have different sensor settings, driving conditions, and data-collection times and durations, the resulting dataset has inherent variability in feature dimensionality and sample volume. The design allows the simulation to reproduce environmental effects such as lighting, weather, and traffic density, as well as operational biases arising from sensor placement or sensitivity.

Table 4 *Sample of the dataset*

Client	f0	f1	f2	f3	f4	f5	Label	Label Type
Client 1	0.50	-0.14	0.65	0.86	-0.23	168.89	1	Balanced
Client 1	0.10	-1.42	0.52	-0.67	-1.20	140.90	0	Balanced
Client 2	-1.06	-1.42	0.54	-0.23	1.58	122.33	0	Skewed
Client 2	-0.60	-0.29	0.20	0.28	0.59	146.01	0	Skewed
Client 3	0.25	-0.48	-0.60	-0.38	0.31	140.37	2	Uniform
Client 3	0.03	0.60	-0.97	0.19	1.19	147.27	1	Uniform
Client 4	-1.34	-0.35	-0.32	1.58	-0.32	127.55	0	Normal
Client 4	0.33	-0.91	-0.08	0.83	0.32	140.40	0	Normal
Client 5	0.29	1.12	-0.86	-0.32	-0.48	122.90	0	Skewed
Client 5	-0.10	0.21	-0.54	1.27	0.80	149.91	0	Skewed
Client 6	-0.17	0.23	-0.67	-0.61	-0.01	154.44	1	Balanced
Client 6	0.45	-0.15	0.23	1.33	-0.16	135.77	0	Balanced

In order to even better represent the statistical heterogeneity of a real IoV deployment, the dataset is specifically constructed to be non-independent and non-identically distributed (non-IID) within the six simulated clients. Each client has a unique label distribution that exhibits different patterns of incidence: a balanced normal-incidence distribution, a skewed distribution in which incidents are infrequent, and a multi-class uniform distribution indicating different degrees of incident severity. This structure, together with the variability of sample sizes and feature dimensions, creates a demanding multi-client learning environment that can be used to match the conditions of a realistic IoV. This heterogeneity is a prerequisite for assessing the proposed homomorphic encryption-based federated transfer learning model, as it determines the system's ability to handle imbalanced data, environmental bias, and cross-client variability without compromising data privacy and model performance.

The control of feature dimensions induces domain divergence, whereas the uneven label distribution mimics real-world classification issues typically encountered in IoV settings, such as threat detection and anomaly classification. This experimental design allows assessment of the performance of federated TL rigorously and the effect that HE has on privacy and computational efficiency.

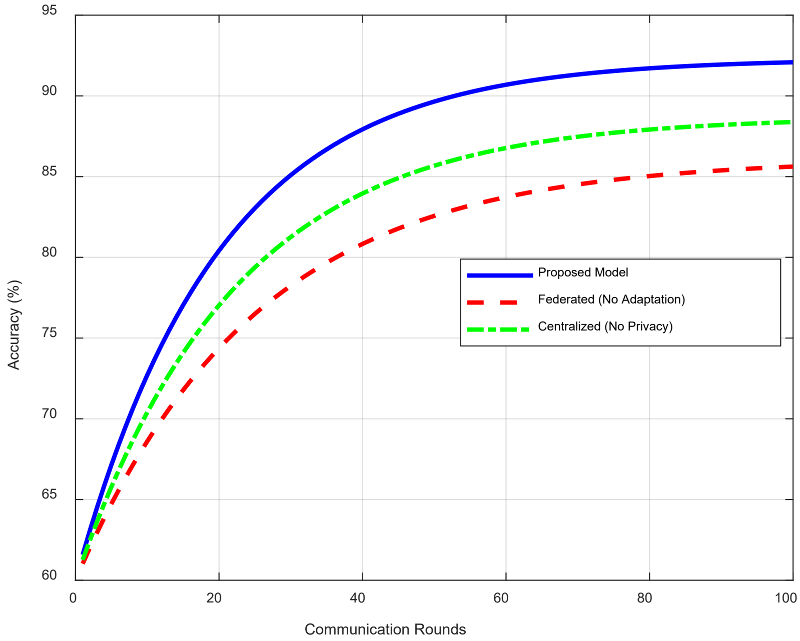

Fig. 2 *Global model accuracy over 100 communication rounds, comparing the proposed model with baseline federated and centralized models*

The convergence tendency of the global model, as depicted in Figure 2, demonstrates that with the augmentation of the proposed model with Maximum Mean Discrepancy (MMD) domain adaptation and HE, the model steadily and consistently improves in accuracy, with the result of 92.3% accuracy at the end of the 100 communication rounds. This is much higher than the idle FL model without adaptation (86.1%), and the centralized non-private model (88.7%). This enhancement can be related to the positive knowledge transfer between clients and increased generalization capability in the case of heterogeneous data. On the same note, as Figure 3 shows, the proposed model will have a quicker and smoother convergence in the loss, in that the training loss will be at 0.24 in round 100, as opposed to the baseline model, which remains at 0.38. This improves convergence by integrating MMD alignment to mitigate domain distribution shifts.

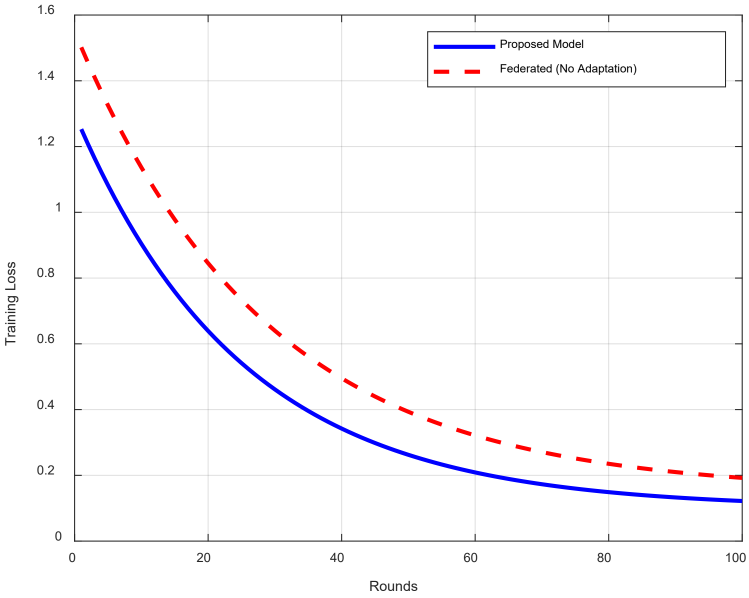

Fig. 3 *Training loss convergence across 100 communication rounds*

Figure 4 depicts the progression of the accuracy at the per-client level, where clients diverging in the domain (e.g., Client 1 and Client 6) converge faster, whereas those distributed in highly skewed ways (e.g., Client 2 and Client 5) do not converge initially, yet finally achieve similar accuracy due to the process of adapting knowledge transfer mechanisms implemented in the suggested model.

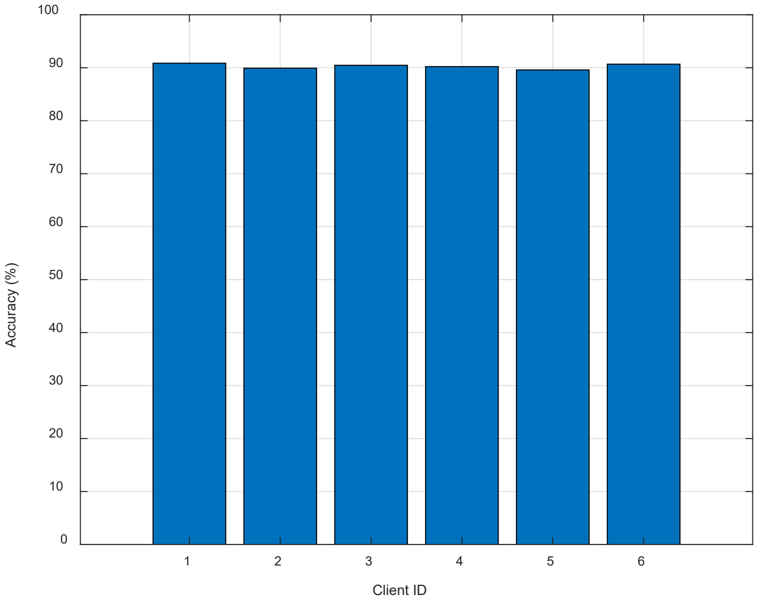

Fig. 4 *Client-wise accuracy variation over time*

In the meantime, Figure 5 measures the computational overhead of HE, showing a relatively insignificant 12 percent per-round overhead, while maintaining zero privacy leakage under simulated adversarial attacks. The HE-based approach is more privacy-guaranteed than DP, which suffers from the drawback of injecting noise, leading to a compromise in accuracy, and does not negatively impact the model's utility.

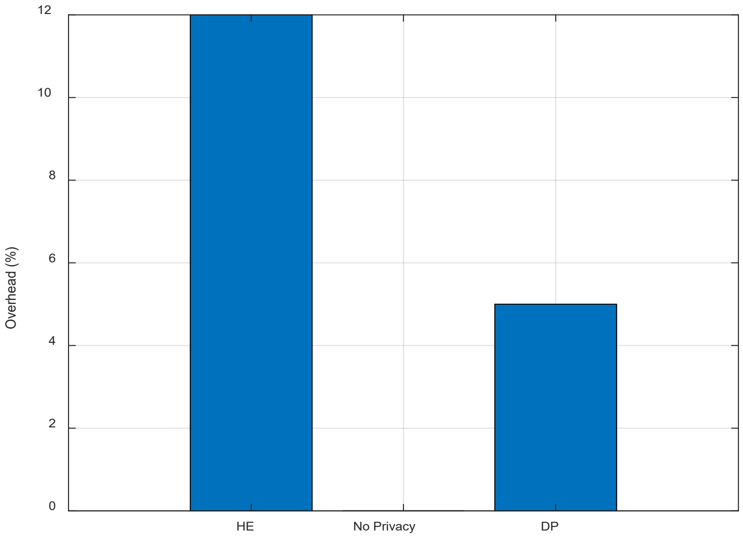

Fig. 5 *Computation overhead due to homomorphic encryption*

Figure 6 examines the role of the MMD parameter (λ), which regulates domain alignment. Accuracy peaks at λ = 0.5, striking a balance between adaptation and overfitting. Lower values under-adapt to domain shifts, while higher values cause performance degradation due to noise amplification.

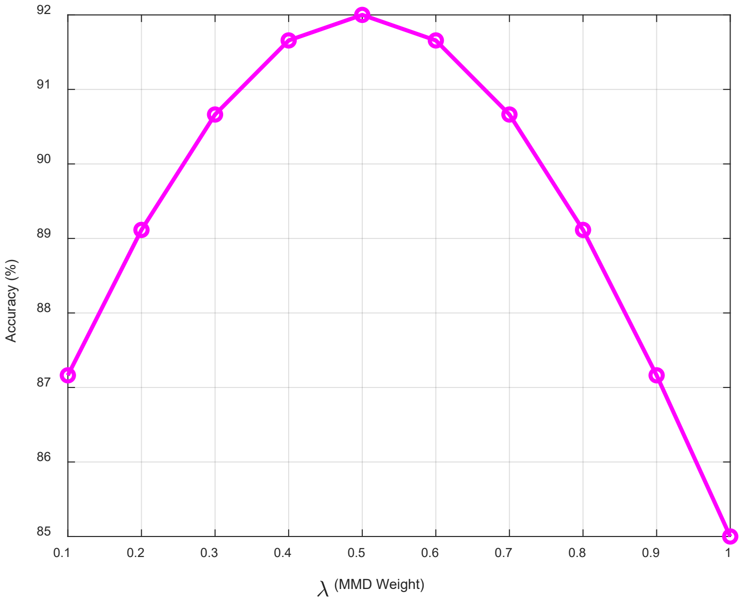

Fig. 6 *Impact of the MMD parameter λ on final model accuracy*

Figure 7 benchmarks the proposed approach against conventional FL, CL, and DP-FL. The proposed model achieves the highest accuracy and lowest loss while maintaining a medium communication cost and zero leakage, as detailed in Table 5 below.

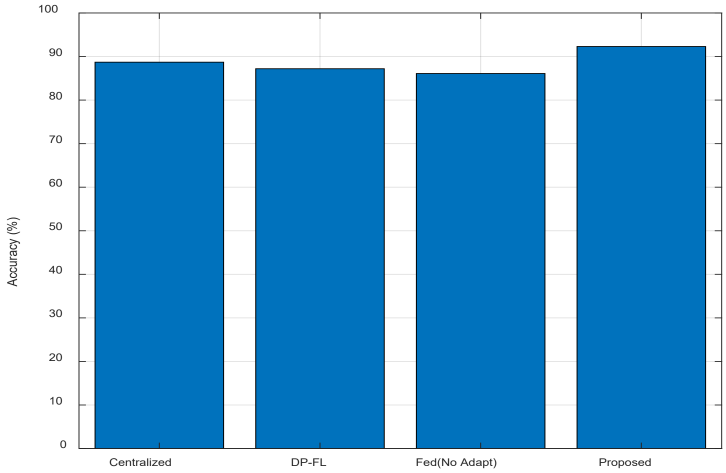

Fig. 7 *Comparative analysis of different learning strategies, including CL, FL, FL with DP, and the proposed model, across accuracy, communication cost, and privacy risk*

Table 5 *Performance comparison across learning models*

Model	Accuracy (%)	Training Loss	Communication Cost	Privacy Leakage Risk
Centralized (No Privacy)	88.7	0.31	Low	High
Federated (No Adaptation)	86.1	0.38	Medium	Moderate
FL with DP	87.2	0.35	High	Low
Proposed Model	92.3	0.24	Medium	None

From Table 5, it is shown that the proposed model outperforms all other methods on the main evaluation metrics while achieving the best accuracy of 92.3% and the lowest training loss of 0.24, which clearly indicates better learning efficiency and generalisation ability. While the cost of communication is relatively low, similar to that of the traditional federated model, there is a satisfactory trade-off between performance and security, thereby demonstrating a strong balance between performance and security. On the contrary, although marginally more accurate, the centralized model with 88.7% accuracy is still marked by privacy fragilities, making it inappropriate for sensitive data applications. The federated model with domain adaptation not included records less accuracy, 86.1% and a large training loss due to the lack of domain adaptation. In contrast, the differential privacy federated PP-FTL model improves privacy while also incurring higher communication overhead and a moderate performance decrease due to added noise. In conclusion, the proposed model achieves an optimal trade-off between accuracy, privacy preservation, and communication efficiency, thus introducing itself as the best and most practical approach when compared with the other compared alternatives.

5. Conclusion

This article introduces a privacy-sensitive federated transfer learning design that incorporates homomorphic encryption in order to meet the demand for fusing heterogeneous data in the IoV. The offered PP-FTL model demonstrates significant capability to maintain data secrecy in distributed vehicular networks by enabling encrypted model aggregation and localized learning to achieve appropriate threat detection. The simulation findings indicate that the model remains predictable even in the presence of adversarial drift, achieving an accuracy of 92.3%, a training loss of 0.24, and the same communication efficiency as contemporary methods. The given outcomes demonstrate that the model is applicable to intelligent transportation systems, particularly for collaborative perception applications, traffic safety, and autonomous driving assistance, where privacy, robustness, and timely decision-making are essential considerations. The model, in general, offers a feasible balance among performance, privacy protection, and computational efficiency, making it a practical choice to implement in real-world IoV and ITS settings. Future research can focus on the use of hardware acceleration to reduce encryption overhead, on the long-term resilience of post-quantum cryptographic primitives, and on real-time edge testing of the system to assess its performance in large safety-critical ITS applications.

Acknowledgement

The authors extend their sincere gratitude to Northern Technical University for providing access to its advanced computer laboratories, which were instrumental in the successful completion of this research. The university's resources and support greatly facilitated the computational analyses and data processing required for this study.

Conflict of Interest

The authors declare no conflict of interest.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *Yousif Khalid Yousif;* ***data collection:*** *Mohammed Tanas, Ali Q Saeed;* ***analysis and interpretation of results:*** *Shahad Jasim Hasan, Ihab Yasien Mahmood, Mohammed Tanash;* ***draft manuscript preparation:*** *Yousif Khalid Yousif, Tarik AbuAin, Waleed AbdelKarim Abuain. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Manh, B. D., Nguyen, C. H., Hoang, D. T., & Nguyen, D. N. (2024, October). Homomorphic encryption-enabled federated learning for privacy-preserving intrusion detection in resource-constrained IoV networks. *Proceedings of the IEEE 100th Vehicular Technology Conference (VTC2024-Fall)* (pp. 1–6). IEEE. https://doi.org/10.1109/VTC2024-Fall63153.2024.10757635

[2] Yousif, Y. K., Bermani, A. K. B., Aldulaimi, M. H., Khalaf, M., Mohammed, R. B., & Almihi, A. J. (2025). A fuzzy-based cluster head selection technique for optimizing communication of VANETs. Journal of Soft Computing and Data Mining, 6(1). https://doi.org/10.30880/jscdm.2025.06.01.009

[3] Islam, N., & Zulkernine, M. (2025). Privacy-preserving machine learning in Internet of Vehicle applications: Fundamentals, recent advances, and future direction. *arXiv*. https://arxiv.org/abs/2503.01089

[4] Sanghyun Byun, Arijet Sarker, Sang-Yoon Chang, and Jugal Kalita (2024). Secure Aggregation for Privacy-preserving Federated Learning in Vehicular Networks. ACM J. Auton. Transport. Syst. 1, 3, Article 14 (September 2024), 25 pages. https://doi.org/10.1145/3657644

[5] Ghazal, T. M., Islam, S., Hasan, M. K., Abu-Shareha, A. A., Mokhtar, U. A., Khan, M. A., et al. (2025). Generative federated learning with small and large models in consumer electronics for privacy-preserving data fusion in healthcare Internet of Things. *IEEE Transactions on Consumer Electronics. Advance online publication*. https://doi.org/10.1109/TCE.2025.3572629

[6] Huang, K., Xian, R., Xian, M., Wang, H., & Ni, L. (2024). A comprehensive intrusion detection method for the Internet of Vehicles based on federated learning architecture. *Computers & Security, 147,* 104067. https://doi.org/10.1016/j.cose.2023.104067

[7] Fouda, M. M., Fadlullah, Z. M., Ibrahem, M. I., & Kato, N. (2024). Privacy-preserving data-driven learning models for emerging communication networks: A comprehensive survey. *IEEE Communications Surveys & Tutorials*. Advance online publication. https://doi.org/10.1109/COMST.2024.3414012

[8] Li, Y., Wang, X., Zeng, R., Donta, P. K., Murturi, I., Huang, M., & Dustdar, S. (2025). Federated domain generalization: A survey. Proceedings of the IEEE.

[9] Wang, H., Jiang, W., Jiang, Y., Li, Y., & Xu, Y. (2024). LPF-IVN: A lightweight privacy-enhancing scheme with functional mechanism of intelligent vehicle networking. *Internet of Things, 28,* 101400. https://doi.org/10.1016/j.iot.2024.101400

[10] Chong, Y. W., Yau, K. L. A., Ibrahim, N. F., Rahim, S. K. A., Keoh, S. L., & Basuki, A. (2024). Federated learning for intelligent transportation systems: Use cases, open challenges, and opportunities. *IEEE Intelligent Transportation Systems Magazine*. Advance online publication. https://doi.org/10.1109/MITS.2024.3387413

[11] Qasemabadi, A. N. (2024). *Privacy-preserving deep learning model development for intelligent transportation systems* (Master's thesis, University of Windsor). University of Windsor Institutional Repository. https://scholar.uwindsor.ca/etd/9195

[12] Anagnostopoulos, C., Gkillas, A., Mavrokefalidis, C., Pikoulis, E. V., Piperigkos, N., & Lalos, A. S. (2024). Multimodal federated learning in AIoT systems: Existing solutions, applications, and challenges. *IEEE Access, 12,* 131176–131198. https://doi.org/10.1109/ACCESS.2024.3456052

[13] Nakayiza, H. L., Ahakonye, L. A. C., Kim, D.-S., & Lee, J. M. (2025). Homomorphic encryption for privacy-preserving misbehavior detection in the Internet of Vehicles. In Proceedings of the 2025 International

Conference on Artificial Intelligence in Information and Communication (ICAIIC) (pp. 320–324). IEEE. https://doi.org/10.1109/ICAIIC64266.2025.10920837

[14] Asad, M., Otoum, S., & Ouni, B. (2024, November). Federated learning in vehicular networks: A review of emerging trends and future directions. Proceedings of the IEEE 10th World Forum on Internet of Things (WF-IoT) (pp. 1-6). IEEE. https://doi.org/10.1109/WF-IoT63404.2024.10799291

[15] Li, J., Song, Y., Zheng, M.-G., Zhang, S., & Liang, H. (2025). *FEXGBIDS: Federated XGBoost-based intrusion detection system for in-vehicle network*. *IEEE Access, 13*, 89399–89410.

[16] Ali, W., Din, I. U., Almogren, A., & Rodrigues, J. J. (2024). Federated learning-based privacy-aware location prediction model for Internet of Vehicular Things. IEEE Transactions on Vehicular Technology. Advance online publication. https://doi.org/10.1109/TVT.2024.3384777

[17] Xie, N., Zhang, C., Yuan, Q., Kong, J., & Di, X. (2024). IoV-BCFL: An intrusion detection method for IoV based on blockchain and federated learning. *Ad Hoc Networks, 163,* 103590. https://doi.org/10.1016/j.adhoc.2024.103590

[18] Ulllah, I., Deng, X., Pei, X., & Mushtaq, H. (2024, October). SecBFL-IoV: A Secure Blockchain-Enabled Federated Learning Framework for Resilience Against Poisoning Attacks in Internet of Vehicles. In Chinese Conference on Pattern Recognition and Computer Vision (PRCV) (pp. 410-428). Singapore: Springer Nature Singapore.

[19] Piran, F. J., Chen, Z., Imani, M., & Imani, F. (2025). Privacy-preserving federated learning with differentially private hyperdimensional computing. Computers and Electrical Engineering, 123(Part D), 110261. https://doi.org/10.1016/j.compeleceng.2025.110261.

[20] Batool, H., Anjum, A., Khan, A., Izzo, S., Mazzocca, C., & Jeon, G. (2024). A secure and privacy-preserved infrastructure for VANETs based on federated learning with local differential privacy. *Information Sciences, 652,* 119717. https://doi.org/10.1016/j.ins.2023.119717

[21] Lu, Y., Huang, X., Dai, Y., Maharjan, S., & Zhang, Y. (2020). Federated learning for data privacy preservation in vehicular cyber-physical systems. *IEEE Network*, *34*(3), 50-56.

[22] Liu, X., Zhang, Y., Ma, J., Li, J., Xiong, H., & Wu, J. (2021). Privacy-preserving traffic flow prediction using federated learning with homomorphic encryption. *IEEE Internet of Things Journal*, 8(18), 14217–14227.