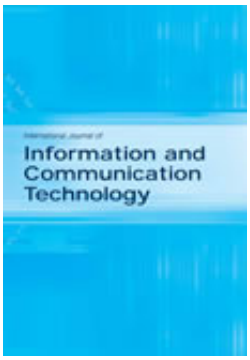


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

A data compliance sharing algorithm for intelligent connected vehicles empowered by federated learning

Bei Zhou

DOI: 10.1504/IJICT.2025.10073652

Article History:

Received:	02 July 2025
Last revised:	14 August 2025
Accepted:	16 August 2025
Published online:	10 October 2025

A data compliance sharing algorithm for intelligent connected vehicles empowered by federated learning

Bei Zhou

School of Economic Law,
Southwest University of Political Science and Law,
Chongqing, 401120, China
Email: 18315199916@163.com

Abstract: With the rapid development of intelligent connected vehicle (ICV) technology, massive amounts of vehicle data have become an important resource for advancing intelligent mobility and autonomous driving technologies. However, the sharing of these data involves significant privacy leakage risks and compliance challenges, especially in multi-party collaboration scenarios. To this end, this paper proposes a federated learning (FL) framework integrating differential privacy and Paillier homomorphic encryption for intelligent connected vehicle (ICV) data sharing. The architecture integrates differential privacy (DP) and homomorphic encryption (HE) through four functional layers. The four-layer architecture achieves 93.5% model accuracy with 79% lower privacy leakage risk (0.21 vs. 0.98 baseline) on 50,000 driving scenarios. Momentum-accelerated averaging and 8-bit gradient quantisation reduce bandwidth consumption by 62%. Experimental validation demonstrates superior privacy-utility balance compared to standalone DP (0.6 risk) and HE (0.5 risk) implementations.

Keywords: intelligent connected vehicle; ICV; federated learning; FL; data compliance sharing; privacy protection.

Reference to this paper should be made as follows: Zhou, B. (2025) 'A data compliance sharing algorithm for intelligent connected vehicles empowered by federated learning', *Int. J. Information and Communication Technology*, Vol. 26, No. 36, pp.42–57.

Biographical notes: Bei Zhou received her Master's degree from the Southwest University of Political Science and Law in 2018. She is currently a PhD candidate at the Southwest University of Political Science and Law. Her research interests include data governance in economic law and compliance governance.

1 Introduction

With the continuous progress of information technology, communication technology and artificial intelligence (AI) technology, intelligent connected vehicle (ICV), as an important part of the future transport system, has gradually become a key direction for the development of the global automotive industry (Zhang et al., 2025). In-depth mining and analysis of these data can help enhance vehicle intelligence, improve traffic management and optimise driving experience. However, the amount of data generated by

ICVs is huge and complex, and how to efficiently and securely manage and share such data, especially in terms of privacy protection and compliance, has become a core issue that restricts their widespread application.

Currently, the traditional data-sharing model faces enormous challenges. Firstly, this model not only has high transmission and storage costs but also has a very high risk of privacy leakage. In ICV use cases, the data typically includes sensitive information about users, such as their personal privacy, vehicle tracks, and other private information. If this data is leaked or misused, it might have major negative effects on the users and society as a whole. To cope with these problems, more and more research and applications have begun to explore the protection of data privacy through distributed learning approaches, among which federated learning (FL), as an emerging distributed learning framework, provides a practical solution for data privacy protection (Li et al., 2021). This approach not only protects user privacy but also reduces bandwidth consumption and computational overhead of data transmission and improves the efficiency of data sharing.

However, despite the unique advantages of FL in terms of data privacy protection, its application in ICV still faces many challenges, especially in ensuring data sharing compliance. How to ensure that the data sharing and model training processes conducted under the framework of FL are compliant with these laws and regulations has become an important issue that needs to be addressed. Therefore, on the basis of FL, how to establish a compliance assurance mechanism that can guarantee data privacy and meet the regulatory requirements has become the core issue of ICV data sharing.

This study suggests a design plan for an ICV data compliance sharing algorithm based on FL to deal with this problem. The algorithm will use the best parts of FL to come up with a way to secure people's privacy while still making sure that the data sharing procedure follows all the rules and laws on data protection. This paper's goal is to help ICV develop in a smart way by providing theoretical support and practical advice. It also wants to help build a traffic management system that is safer and smarter.

2 Federated learning

FL is a new way of doing distributed machine learning that lets many people train locally without sharing raw data. This way, it tries to reduce the risk of data privacy breaches. FL is different from traditional centralised learning methods because it does not store all the data on a central server. Instead, it trains models on the devices of the participants and sends the parameters or gradients of the local models to a central server, where they are combined to create a global model. This method keeps the raw data from leaving the location, which preserves privacy a lot, and lets several people work together to make the model operate better.

In the FL workflow, the first step is for the central server to setup a global model and provide the first parameters of that model to each participant. Everyone who takes part (like a smart device or an edge computing node) gets the global model and uses their own data to train it (Chen and Ran, 2019). At this point, the local devices do not have to share or upload any raw data. They just need to train the model on their own and do it repeatedly. Each person uses their own local data to train the global model. Because the data is spread out in different ways, each participant will train their model in a different way, and the trained local model will likewise be distinct. During this procedure, the

participants commonly use optimisation methods like SGD or Adam to change the model's weight parameters (Reyad et al., 2023). It is important to note that the training of these local models is only based on the participant's own data and does not entail storing or uploading data in a central location.

After each participant has finished their local training, they communicate changes to their local models (usually adjustments to the gradient or weight) to the central server. The central server does not get to the raw data directly. Instead, each participant sends it parameter updates, and the server combines them. Simple mean aggregation and weighted aggregation are two common ways to combine data. Weighted aggregation lets you give more weight to updates based on how much training data each participant has (Deng et al., 2022). This way, the central server may get a global model that includes the information from all the participants. Each participant gets back the aggregated global model, which they use to continue training in their own area based on the changes to the global model. This method is repeated several times until the global model converges or meets performance standards that have already been set. Each round of iteration makes each participant's contribution to the global model stronger (Wen et al., 2023). This lets the model be optimised and enhanced over the data distribution of all participants.

There are several ways that FL's main features and benefits show up. First, FL does an excellent job of protecting people's privacy by prohibiting data from leaving the area. In traditional centralised learning, data has to be sent to the server from a single location, which could lead to data leaks and misuse. FL does not do this, and it can be paired with encryption to make data further safer. FL can also cut down on bandwidth use from data transfer and make computing more efficient by having each device do its own calculations.

FL is very scalable and flexible because it is not centralised. As new devices join the network, FL can easily add more computer nodes and spread computing jobs across all of them. This takes some of the load off the central server. FL can also let edge computing devices join in on the model training, which makes the system even faster and more powerful (Wang et al., 2019). FL's flexibility lets it meet a wide range of needs in real-world situations. For instance, in smart homes, several devices can work together to train home control models without revealing private home data. In the medical arena, several hospitals can work together to train illness diagnosis models under the framework of FL without giving away patients' personal information.

FL, on the other hand, has a lot of problems to deal with. The data from the participants may come from diverse devices or sources, which can make the data less stable while training a global model. Variable devices have variable amounts of computational power, network access, and storage space, which makes FL harder to implement (Imteaj et al., 2021). FL's training process also usually needs several rounds of model updating and parameter synchronisation. Each round of communication and aggregation can cause delays, especially when there are a lot of devices involved. This makes it hard for FL to figure out how to do model synchronisation and aggregation quickly.

As technology keeps getting better, FL's potential uses in many areas are still wide, especially in situations where privacy is very important. FL is predicted to become a vital technique for processing data in the future.

3 ICV data sharing technology

As ICVs grow quickly, collecting, sending, and sharing vehicle data has become a key aspect of new ideas in the car business. ICVs collect a lot of information when driving, such as the condition of the vehicle, sensor data, real-time road conditions, driving behaviour, traffic information, and more (Terzi et al., 2018). These data can be used to make vehicles smarter and improve driving experience. They can also help with the development of urban traffic management, intelligent transportation systems (ITS), and self-driving cars. But existing technology and rules still find it hard to figure out how to communicate this information in a way that is both efficient and compliant, especially when it comes to keeping data private and safe.

There are a number of parts to ICV's data sharing system. The first step is to gather and process data. ICVs have a lot of sensors, like cameras, radar, LIDAR, GPS, and more, that keep an eye on the vehicle and the area around it and collect information about changes in both in real-time. There is a lot of different data, like picture data, sensor signals, vehicle position, speed, and more. One of the most important technologies in the data sharing system is how to make sure that the data is real and complete while also processing and storing it quickly (George and George, 2022). The second is sending and receiving data. ICVs commonly send data over communication networks that are built within the vehicle. Communication between vehicles (V2V), vehicles and infrastructure (V2I), and vehicles and cloud platforms become the major means for ICVs to share data. Because of the unique features of ICV, the need for real-time and low-latency communication is very high, which makes it hard to make sure that communication stays stable and fast during data transmission. Also, as the amount of data grows, it is still important to find ways to boost network capacity and transmission speed, as well as lower the time it takes for data to go through when there is network congestion or signal loss.

Also, ICV data exchange systems need to work on problems with data standardisation and interoperability. It has become a technical challenge that needs to be solved quickly to figure out how to make sure that data can be easily shared and connected between different vehicles and between vehicles and infrastructure. This is because ICVs from different manufacturers and models may have different data formats, interface standards, and so on. To make data work together, the industry has suggested some standard solutions based on vehicle information exchange protocols (like V2X communication protocols) and unified data formats (Ameen et al., 2019). These are the basic building blocks for sharing data across platforms and making systems work together.

In recent years, with the development of ICV, research addressing data sharing has made significant progress. In terms of data privacy protection, many studies focus on how to achieve data sharing while protecting user privacy (Gupta et al., 2022). Meanwhile, data encryption technology is also a common privacy protection means, and many scholars have proposed encryption schemes based on homomorphic encryption (HE), secret sharing and other methods to ensure security during data transmission. In terms of data transmission and communication efficiency, current research focuses on how to improve the data transmission rate and reliability between V2V and V2I (Malik et al., 2019). In addition, research on the standardisation of in-vehicle communication protocols in data sharing is also progressing, with researchers committed to proposing more efficient communication protocols to enhance data interoperability between

different vendors and different devices. For instance, the International Telecommunication Union (ITU) and the European Telecommunications Standards Institute (ETSI) have suggested standardisation plans for V2X communications, internet of vehicles (IoT) platforms, data formats, and interfaces. The goal is to make it easier for data to work with products from different vendors and countries. In terms of data standardisation, the industry is actively promoting the standardisation of ICV data sharing. The development of these standards will help solve the compatibility and interoperability problems that currently exist in the process of cross-platform data sharing.

Overall, research on ICV data sharing technology has come a long way in terms of protecting privacy, making communication more efficient, standardising data, and making it more secure. However, there are still many technical and policy issues that need to be worked out. As technologies like 5G, blockchain, and AI continue to get better, ICV data exchange will become more and more vital in the future for things like smart transportation and self-driving cars.

4 System design of algorithms

This study creates a data sharing algorithm based on FL technology to meet ICV's privacy protection and compliance needs during the data sharing process. The algorithm makes sure that data protection laws and regulations are followed around the world while keeping data private using a multi-layered technical architecture (see Figure 1).

4.1 Data privacy protection layer

One of the main goals of the ICV data sharing procedure is to make sure that users' privacy is protected. The data privacy protection layer uses effective privacy protection methods to keep any sensitive information from being shared when data is being processed and sent. Two methods, differential privacy (DP) and HE, are used for this.

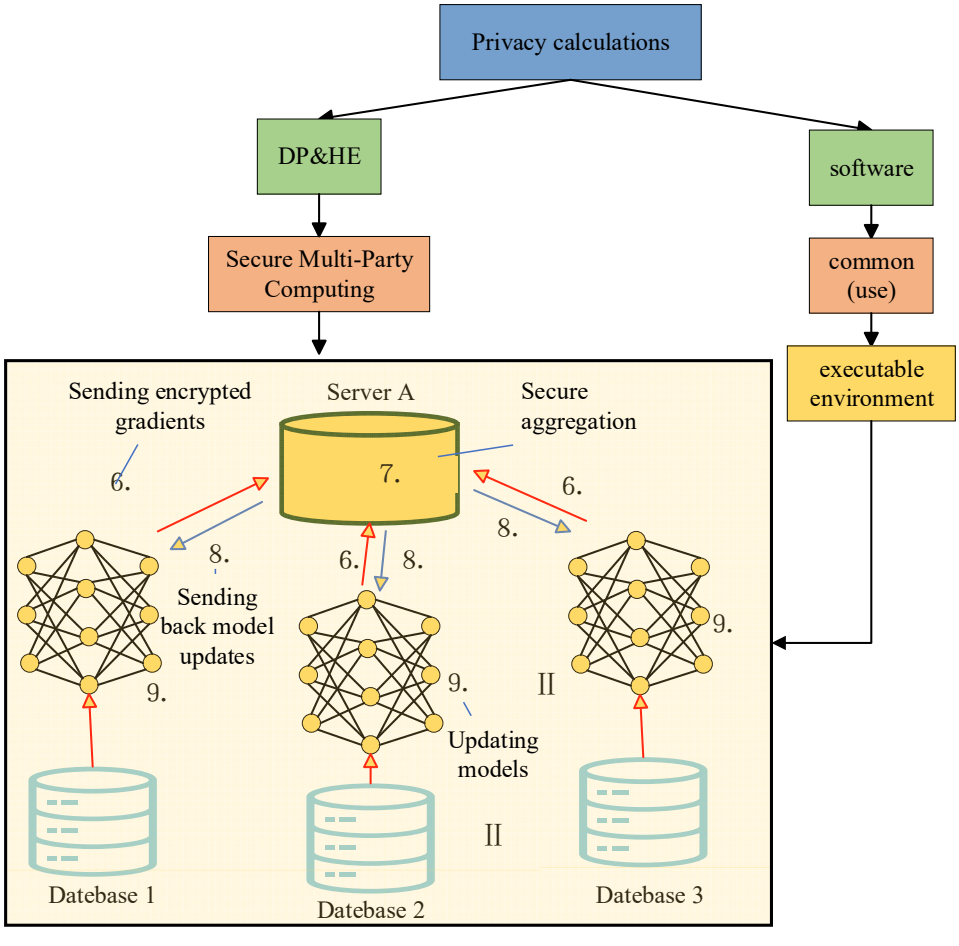
DP makes sure that no personal information can be guessed when the data is combined by changing the dataset. DP processes all of the sensitive information before the vehicle data is uploaded to preserve people's privacy during the statistical analysis procedure. The main idea behind DP is that the outcomes of processing two datasets D and D' that are next to each other are virtually the same (Tu et al., 2020). The privacy budget ϵ determines how strong this privacy protection is. The mathematical way to write DP is:

$$\Pr\left[A(D) \in S\right] \leq e^{\delta} \cdot \Pr\left[A(D') \in S\right] \tag{1}$$

where A is the DP method, D and D' are two datasets that are next to each other, S is the collection of results of an event, and ϵ is the privacy budget. The calculation indicates that selecting the right privacy budget can successfully regulate the level of disturbance, which lowers the danger of leaking private information about individuals.

In the meantime, the HE technique is added to the data processing session to make the privacy protection of the data even better. HE lets us work with encrypted data without having to decrypt it (Doan et al., 2023). The attacker cannot get the original information back from the encrypted data, even if they intercept the transmission. In the FL framework, vehicles submit encrypted updates of their local models to a central server.

Figure 1 Design of ICV data compliance sharing algorithm (see online version for colours)



The computational principle of HE is embodied in an additive operation with the formula:

$$D\left(\varepsilon(a)\oplus\varepsilon(b)\right)=a+b \tag{2}$$

where a and b are the original data, after encryption, the server can only operate on the encrypted data, and cannot directly obtain the real values of a and b.

In addition, encrypted aggregation is used as an important technique to protect the privacy of model updates in FL. The mathematical formulation of encryption aggregation is as follows:

$$D\left(\sum_{i=1}^{n}\varepsilon\left(x_i\right)\right)=\sum_{i=1}^{n}x_i \tag{3}$$

where $E(x_i)$ is the encrypted data of each participant. The central server does the aggregation operation on the encrypted data right after it gets it, and then it gets the right

result by decrypting it. This method keeps people from directly accessing the original data, which makes data privacy protection even stronger.

This layer protects the privacy of ICV in data sharing and FL by using both DP and HE approaches together. The combination of these methods lets the algorithm be very efficient while avoiding privacy leaks and data exploitation.

4.2 FL layer

The last layer uses a number of methods to protect the privacy and security of ICVs while they share data. In this layer, it also talks about how to do collaborative learning well with a lot of people while keeping data safe via a decentralised model training strategy. The main idea behind FL is to let each participant (like various cars) train the global model jointly without giving each other the raw data. Each participant does data processing and model training on their own computer and only sends updates of the local model (not raw data) to the central server. This keeps sensitive information from getting out.

Federated averaging (FedAvg) is the method utilised in this paper to combine models. In FedAvg, the central server collects all the local model updates from all the participants and uses them to figure out the global model update (Collins et al., 2022). The formula for calculating the global model parameters is:

$$w_{global} = \frac{1}{N}\sum_{i=1}^{N} \alpha_i w_i \tag{4}$$

where N is the total number of participants, w_{global} is the aggregation result of the global model, w_i is the model update of the i^{th} participant, and α_i is the weight of that participant.

In the study of this paper, another important optimisation step is the local update. The update formula for the global model is:

$$w_{global}^{(t+1)} = w_{global}^{(t)} + \sum_{i=1}^{N} \alpha_i \left(w_i^{(T_i)} - w_{global}^{(t)} \right) \tag{5}$$

where $w_{global}^{(t)}$ denotes the parameters of the t^{th} round of global model, $w_{global}^{(t+1)}$ is the $t + 1$st round of global model update, $w_i^{(T_i)}$ is the local update model, and T_i is the number of local training rounds for the i^{th} participant.

In addition, in order to further reduce the communication cost, asynchronous updating strategy is also applied in this study. The formula for asynchronous updating is:

$$w_{global}^{(t+1)} = w_{global}^{(t)} + \sum_{i=1}^{N} \alpha_i \left(w_i^{(t)} - w_{global}^{(t)} \right) \tag{6}$$

4.3 Data transmission and communication layer

This algorithm uses quantisation and sparsification techniques to reduce the amount of data uploaded and downloaded in each round of communication. The quantisation process can be represented as:

$$\hat{w}_i = Quantise(w_i) \tag{7}$$

where w_i is the local model update for the i^{th} participant and $\hat{w}_i$ is the quantisation parameter resulting from the quantisation operation.

If the model update is w_i, the sparse update comes from the sparsification process, which can be written arithmetically:

$$w_i^{sp} = Sparse(w_i, \theta) \tag{8}$$

where θ is the sparsification criterion and only parameters above this threshold are retained. This method makes the communication of the system simpler and more efficient by reducing the amount of data sent.

To simplify communication, we use the gradient compression protocol (GCP), which reduces the amount of communication required for each update (Luo et al., 2021). If we assume that the model parameters are updated to $w_{global}^{(t)}$ at round t, then the model parameters can be sent using GCP as follows:

$$\Delta w_{global}^{(t)} = Compress\left(w_{global}^{(t)} - w_{global}^{(t-1)}\right) \tag{9}$$

where $\Delta w_{global}^{(t)}$ is the gradient update of the current round, the compressed parameters are transmitted less, effectively reducing the bandwidth requirement.

If we say that the communication delay between a participant i in layer l and other nodes is $D_{l,i}$, then optimising the communication pathways between and within layers can cut down on the network's overall transmission latency by a lot.

In short, this layer not only cuts down on communication overhead, but it also speeds up model training by using methods like data compression, optimising transmission protocols, asynchronous communication, and optimising network architecture. In ICV systems, the FL framework says that making sure communication is clear is the most important thing for effective collaborative learning. These optimisation tactics make the whole training process work better and respond faster in real-time, all while keeping data private.

4.4 Data compliance validation layer

In the ICV system, this layer's main job is to make sure that data sharing and use precisely follow all relevant laws and rules. In this study, we use a role-based access control (RBAC) approach to carefully manage who has access to the data based on the roles and permissions of each participant (Cruz et al., 2018). Specifically, if the i^{th} participant has the function of R_i and the access rights of R_i, the RBAC mechanism used for participant access control can be written as:

$$Access(R_i, P_i) \rightarrow w_i^{allowed} \tag{10}$$

where $Access(R_i, P_i)$ is the control function for the model updates that can be accessed. This manner, the system can make sure that only data and model updates that are compliant may be utilised for training. This stops data from being misused and stops problems with compliance.

Another important way to make sure that data is handled and shared in accordance with data protection rules is through compliance auditing. The compliance auditing mechanism may keep track of every data processing procedure and every model update operation by monitoring and recording all the system's data processing activities in real-time (Pasquier et al., 2018). This ensures that all operations are done within the limits of legal authorisation. The auditing system can find and report any violations of compliance rules quickly, so that they can be used as evidence in a later legal review. If the system finishes processing and sending data in round t, the audit log L_t can be written as:

$$L_t = \left\{ Timestamp(t),\ Action\left(A_t\right),\ Actor\left(P_t\right),\ Outcome\left(O_t\right) \right\} \tag{11}$$

where $Timestamp(t)$ is the time of the audit event, $Action(A_t)$ is the action that was taken, $Actor(P_t)$ is the person who did the action, and $Outcome(O_t)$ is the result of the action. This auditing solution makes sure that every data processing and model update is traceable and compliant, and it gives the system full transparency.

Another key technique in the compliance validation layer is privacy-preserving protocols. By following tight privacy rules, it is possible to keep data encrypted even when models are updated, which keeps others from getting access to it and sharing it. For instance, once the data is encrypted using the HE method, it stays encrypted during model training, and the decryption operation can only happen when the right conditions are met. If the i^{th} participant's encrypted data is $E(w_i)$ and the operation to decrypt it is $D(E(w_i))$, HE can be written as:

$$E\left(w_i\right) \to D\left(E\left(w_i\right)\right) = w_i \tag{12}$$

This encryption method keeps data protected even when it is being sent or combined with other data, which helps safeguard privacy even more.

In short, the data compliance verification layer makes sure that the ICV system always follows the rules and laws when sharing data and training models under the FL framework. It does this by using data access control, compliance auditing, and privacy protection protocols. These methods can help lower the legal risks that come from data privacy leaks or compliance problems. This makes the system safer, more open, and more reliable.

4.5 Feedback and optimisation layer

This is done to make the whole system work better and more efficiently. In the FL framework, each participant changes the model parameters on their own, but the performance of the overall system is typically affected by things like how often the model is updated, how many people are involved, and how much bandwidth the network has.

One of the most important strategies in the optimisation layer is the model update strategy. Because each participant is trained in the FL framework using data from their own area, the model updates that come out are not all the same quality or stability. So, one essential way to make the system work better is to optimise the model update approach, notably how to change the weights and how often to update them based on how good the local model updates are. We use an adaptive updating technique in this research that changes the weight of each participant in the global model based on how good its

model update is (Davis et al., 2019). For the model update w_i of the i^{th} participant, its update weight α_i can be written as:

$$\alpha_i = \frac{\left\| w_i - w_{global} \right\|}{\sum_{j=1}^{N} \left\| w_j - w_{global} \right\|} \tag{13}$$

where w_{global} stands for the global model's parameters and $||w_i - w_{global}||$ represents the difference between the i^{th} participant's model update and the global model. A bigger difference suggests that this participant's update is better, and the weight goes up.

In this paper, momentum optimisation (MO) method is used, which adjusts the current gradient by introducing a weighted average of previous gradients to accelerate convergence and mitigate oscillations during training (Teo and Nguyen, 2024). The update rule for MO is:

$$v_i^{(t)} = \beta v_i^{(t-1)} + (1-\beta)\nabla f\left(w_i\right) \tag{14}$$

$$w_i^{(t)} = w_i^{(t-1)} - \mu v_i^{(t)} \tag{15}$$

where $v_i^{(t)}$ is the momentum of the i^{th} participant at round t, β is the momentum factor, $\nabla f(w_i)$ is the gradient, and η is the learning rate.

If the communication frequency for each round is f_t, the following equation can be used to show how to change the communication frequency:

$$f_t = f_0 \times \left(1 + \alpha \cdot \frac{Bandwidth}{Max\ Bandwidth}\right) \tag{16}$$

where f_0 is the original frequency and α is the adjustment factor. This adaptive tuning technique helps to optimise the communication overheads under different network conditions, resulting in better system performance.

5 Experimental results and analyses

5.1 Datasets

In order to validate the FL-driven ICV data compliance sharing algorithm proposed in this paper, two datasets were self-constructed in this study, namely ICV Driving Behaviour and Traffic Environment Dataset (ICV-Behaviour-Env-Dataset) and ICV Sensor and Environmental Dataset (ICV-Sensor-Env-Dataset).

Tables 1 and 2 show information about the two datasets.

These two datasets were made with full attention to the needs of data privacy, data diversity, and data compliance in FL. Not only can each dataset be used for things like analysing driving behaviour and sensing the environment, but they can also be used to successfully test the algorithms for compliance sharing in FL. Using these two datasets in experiments can show that FL can be used in ICV, especially to safeguard privacy and share compliance information.

Table 1 Information about ICV-Behaviour-Env-dataset

Data types	Speed, acceleration, steering angle, braking behaviour, traffic signs, pedestrians, other vehicles
Data source	Simulation platform (CARLA), virtual vehicle sensors, traffic environment simulation
Dataset size	50,000 driving scenes, covering urban roads, highways, rainy, and foggy conditions
Number of participants	30 virtual vehicles, representing different brands and sensor configurations
Compliance requirements	Data minimisation principle, data anonymisation, compliance with GDPR and CPLA
Application scenarios	Driving behaviour analysis, autonomous driving decision-making, traffic environment perception
Data labels	Driving behaviour, traffic signs, pedestrians, other vehicles

Table 2 Information about ICV-Sensor-Env-dataset

Data types	LiDAR point cloud data, radar data, traffic signs, lane markings, pedestrians, other vehicles
Data source	Simulation platform (SUMO), virtual sensor data, traffic environment simulation
Dataset size	20,000 environmental scenes, covering different roads, weather conditions, and time periods
Number of participants	40 virtual vehicles, representing different cities, regions, and traffic conditions
Compliance requirements	Compliance with privacy protection regulations, data anonymisation
Application scenarios	Object detection, environment perception, autonomous driving system training and testing
Data labels	Traffic signs, lane markings, pedestrians, other vehicles, road obstacles

5.2 Effectiveness of privacy protection in ICV data compliance sharing

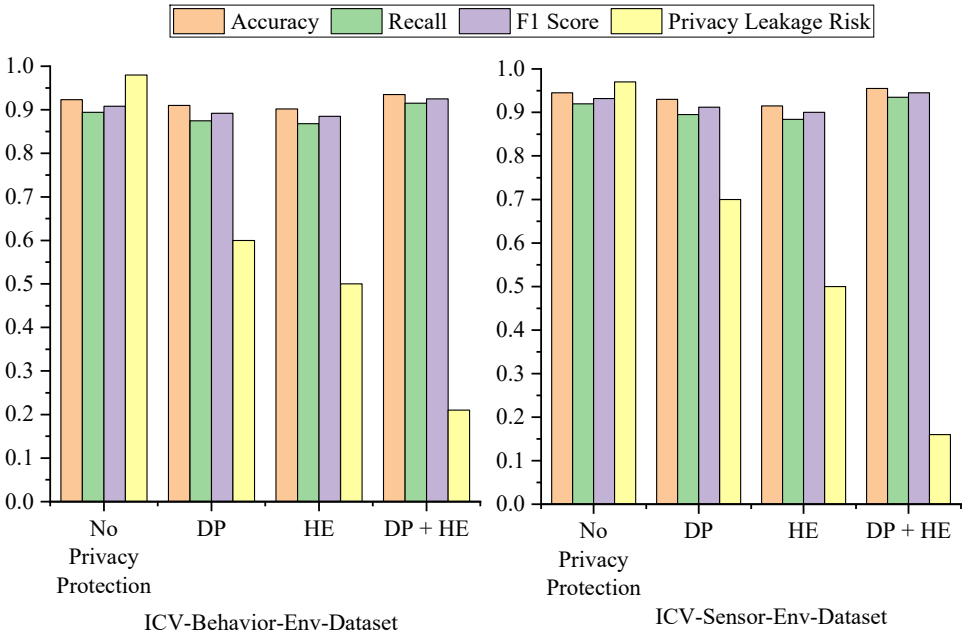
In order to validate the FL-driven ICV data-compliant sharing algorithm proposed in this paper, the goal of this study is to understand whether the privacy-preserving features in the FL-based ICV data-compliant sharing algorithm are effective. We tested several ICV datasets to see how well these privacy-preserving strategies protect people's privacy when they share data.

The experiments use two home-made ICV datasets: the ICV-Behaviour-Env-Dataset and the ICV-Sensor-Env-Dataset. The ICV-Behaviour-Env-Dataset simulates a variety of driving behaviours including speed, acceleration, and steering angle. Its main purpose is to test how effective the privacy-preserving approach is when analysing driving behaviour. The ICV-Sensor-Env-Dataset, on the other hand, contains a large amount of sensor data, such as LiDAR point clouds and radar data. These data are used to test the effectiveness of privacy preserving methods in detecting the environment and finding targets.

DP and HE techniques are used to encrypt each participant's data so that sensitive information would not leak out even if the data is shared. This experiment looks at a few

important metrics: accuracy, recall, F1-score, and the danger of privacy leaking. Figure 2 displays the results of the experiment on the two datasets. It illustrates how well the model works and how well it protects privacy with different privacy protection methods.

Figure 2 Effectiveness of FL-based privacy protection (see online version for colours)



Without privacy protection, the ICV-Behaviour-Env-Dataset has the greatest results: an accuracy of 0.923, a recall of 0.894, an F1-score of 0.908, and a privacy leakage risk of 0.98. This means that the algorithms can get things right most of the time, but they do not safeguard privacy very well and there is a very high danger of data leaking.

The risk of privacy leaks goes down a lot to 0.6 if DP is included, however the performance also goes down a little. The accuracy goes down to 0.910, the recall goes down to 0.875, and the F1-score is 0.892. DP protects privacy, but its methods, including noise injection, lower the algorithm's accuracy and recall, therefore the performance loss is unavoidable. But when the danger of privacy leaks goes down, the protection gets better, which is good for situations where privacy is very important, but some performance loss is acceptable.

Next, the danger of privacy leaking goes down much further with HE, to 0.5, with an accuracy of 0.902, a recall of 0.868, and an F1-score of 0.885. HE protects privacy more than DP since it lets you do calculations on encrypted data, which keeps the data private. Still, HE is a good solution for situations where privacy is absolutely important.

The combination of DP and HE reduces the danger of privacy leaks to a minimum (0.21) and balances performance perfectly. The F1-score is 0.925, the recall is 0.915, and the accuracy is 0.935. The combo scenario protects privacy better than previous situations that use DP or HE alone, with nearly no cost in performance. The combination of DP and HE offers the best performance and the strongest privacy protection.

Without privacy protection, the performance in the ICV-Sensor-Env-Dataset stays the same, with an accuracy of 0.945, a recall of 0.920, an F1-score of 0.932, and a risk of privacy leakage of 0.97. Even while the algorithm works quite well in this situation, it does not ensure privacy at all, thus it cannot be used in real life.

The risk of privacy leaks goes down to 0.7, the accuracy goes up to 0.930, the recall goes up to 0.895, and the F1-score goes up to 0.912 once DP is added. The algorithm does not work as well when privacy protection is turned on. This is mostly because the noise injection process of DP makes it lose some information, which lowers the algorithm's precision and recall. However, privacy protection is greatly improved and works well in situations when privacy and compliance are very important.

With HE, the danger of privacy loss goes down to 0.5, the accuracy goes up to 0.915, the recall goes up to 0.884, and the F1-score goes up to 0.900. Even if HE offers better privacy protection, the speed is still worse because the encryption computation is so complicated. At this stage, the algorithm's accuracy and recall are still not as good as they were before the encryption.

Lastly, the method works best when DP and HE are combined, and the probability of privacy leaking is at its lowest (0.16). The F1-score is 0.945, the recall is 0.935, and the accuracy is 0.955. With the highest accuracy, recall, and F1-score, this combination makes sure that the algorithm works as well as possible while keeping data private.

5.3 Effectiveness of FL compliance validation in ICV data sharing

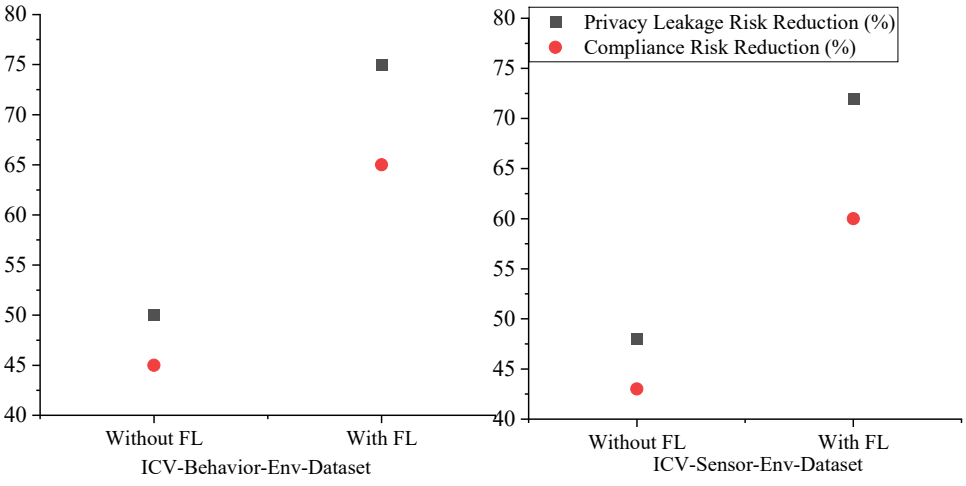
Experiment 2 looks at how FL affects ICV data compliance sharing, notably how it affects compliance validation. The experimental design used two situations for the comparison: one used FL to share data compliance, and the other used the traditional non-FL approach, which is direct centralised data processing. Both models use the same methods to safeguard data privacy, and they both use a mix of DP and HE methods to do so. In the FL scheme, each participant (like different ICV vehicles) trains the local model and then updates it using the FL aggregation mechanism. In the non-FL scheme, on the other hand, each participant sends their data to a central server for processing, which usually means more data transmission and privacy risks.

Figure 3 shows the outcomes of the experiment.

Experiment 2 clearly shows how FL affects the sharing of ICV data compliance. The major goal of the experiment is to see how the risk of privacy leaks and compliance changes when FL is and is not used. The experiments are done on two datasets: the ICV-Behaviour-Env-Dataset and the I ICV-Sensor-Env-Dataset *t*. The results show that the scenario using FL does much better at protecting privacy and following the rules than the traditional non-FL model.

When FL is not utilised, the danger of privacy leakage in the ICV-Behaviour-Env-Dataset is cut in half, and the risk of compliance is cut by 45%. However, when FL was added, the probability of privacy leaks went down by 75% and the risk of not following the rules went down by 65%. This result reveals that adding FL greatly improves the privacy protection of data while it is being shared, and the risk of non-compliance is well managed. The distributed model training method used by FL technology keeps data from being directly centralised, which lowers the risk of privacy leaks.

Figure 3 FL-driven compliance effects (see online version for colours)



Also, when FL is not utilised, the danger of privacy leaks is 48% lower and the risk of compliance is 43% lower on the ICV-Sensor-Env-Dataset. With FL, the risk of a privacy breach goes down by 72% and the danger of not following the rules goes down by 60%. Both solutions improved compliance and privacy protection to some extent, but the one that used FL did far better on both counts. This supports the idea that FL is useful for sharing ICV data compliance, specifically its big role in making privacy protection more effective and lowering compliance risk.

On balance, FL has significant advantages over the traditional non-FL model in terms of privacy protection effectiveness and compliance risk control in ICV data compliance sharing. After using FL, the privacy leakage risk and compliance risk are significantly reduced, especially in the control of compliance risk, the FL scheme shows stronger advantages. This shows that FL can not only effectively ensure data privacy, but also provide more accurate guarantees in compliance verification, driving further improvements in data sharing security and compliance in the ICV industry.

6 Conclusions

This paper explores the effectiveness of applying FL technology for data sharing and compliance verification in ICV environments. By analysing the advantages of FL technology, an ICV data compliance sharing algorithm incorporating FL is proposed, which not only guarantees the security of data privacy, but also effectively improves the accuracy of compliance verification.

However, this study also has certain limitations. First, the algorithm described in this paper works well in theory and in tests, but we still need to investigate how complicated and expensive it would be to use it in real life. In the real world of ICV, things like the number of people involved, the state of the network, and the processing power of the devices can all have an effect on how well the FL algorithm works. Second, this study is mostly about privacy protection and compliance verification, but the ICV field includes a wide range of technologies and standards that are hard to understand. Finally, even

though two different datasets were used for the studies, these datasets are still not very big or complicated.

In the future, research can move in the following directions. First, the FL algorithm can be improved even more to make it work better and be more stable when there is more than one participant. In particular, the goal is to find ways to improve the training process of the FL model and cut down on communication overhead during the model aggregation process when there are a lot of participants and a lot of data. Second, as ICV technology continues to improve, so do the rules and standards for sharing data in compliance. In the future, it will be important to think about how to keep the compliance validation mechanism of FL models up to date and flexible in a regulatory environment that is always changing. Also, using more advanced privacy-preserving technologies, such the combination of FL and blockchain, could lead to more creative ways to share data in ICVs.

To sum up, this article looks into the FL-based data compliance sharing algorithm for ICV, shows how useful this technology may be for protecting privacy and checking compliance, and comes up with new ideas and methods for sharing data in the ICV field. There are still some problems to solve, but as technology keeps getting better, FL and its related techniques will become more and more vital for sharing ICV data and protecting privacy.

Declarations

All authors declare that they have no conflicts of interest.

References

Ameen, H.A., Zaidan, B.B., Zaidan, A.A., Saon, S., Nor, D.M., Malik, R.Q., Kareem, Z.H., Garfan, S., Zaidan, R.A. and Mohammed, A. (2019) 'A deep review and analysis of data exchange in vehicle-to-vehicle communications systems: coherent taxonomy, challenges, motivations, recommendations, substantial analysis and future directions', *IEEE Access*, Vol. 7, pp.158349–158378.

Chen, J. and Ran, X. (2019) 'Deep learning with edge computing: a review', *Proceedings of the IEEE*, Vol. 107, No. 8, pp.1655–1674.

Collins, L., Hassani, H., Mokhtari, A. and Shakkottai, S. (2022) 'Fedavg with fine tuning: local updates lead to representation learning', *Advances in Neural Information Processing Systems*, Vol. 35, pp.10572–10586.

Cruz, J.P., Kaji, Y. and Yanai, N. (2018) 'RBAC-SC: role-based access control using smart contract', *IEEE Access*, Vol. 6, pp.12240–12251.

Davis, S.E., Greevy Jr., R.A., Fonnesbeck, C., Lasko, T.A., Walsh, C.G. and Matheny, M.E. (2019) 'A nonparametric updating method to correct clinical prediction model drift', *Journal of the American Medical Informatics Association*, Vol. 26, No. 12, pp.1448–1457.

Deng, Y., Lyu, F., Ren, J., Chen, Y.-C., Yang, P., Zhou, Y. and Zhang, Y. (2022) 'Improving federated learning with quality-aware user incentive and auto-weighted model aggregation', *IEEE Transactions on Parallel and Distributed Systems*, Vol. 33, No. 12, pp.4515–4529.

Doan, T.V.T., Messai, M.-L., Gavin, G. and Darmont, J. (2023) 'A survey on implementations of homomorphic encryption schemes', *The Journal of Supercomputing*, Vol. 79, No. 13, pp.15098–15139.

George, A.S. and George, A.H. (2022) 'Data sharing made easy by technology trends: new data sharing and privacy preserving technologies that bring in a new era of data monetization', *Partners Universal International Research Journal*, Vol. 1, No. 3, pp.13–19.

Gupta, I., Singh, A.K., Lee, C-N. and Buyya, R. (2022) 'Secure data storage and sharing techniques for data protection in cloud environments: a systematic review, analysis, and future directions', *IEEE Access*, Vol. 10, pp.71247–71277.

Imteaj, A., Thakker, U., Wang, S., Li, J. and Amini, M.H. (2021) 'A survey on federated learning for resource-constrained IoT devices', *IEEE Internet of Things Journal*, Vol. 9, No. 1, pp.1–24.

Li, Q., Wen, Z., Wu, Z., Hu, S., Wang, N., Li, Y., Liu, X. and He, B. (2021) 'A survey on federated learning systems: vision, hype and reality for data privacy and protection', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 35, No. 4, pp.3347–3366.

Luo, P., Yu, F.R., Chen, J., Li, J. and Leung, V.C. (2021) 'A novel adaptive gradient compression scheme: reducing the communication overhead for distributed deep learning in the internet of things', *IEEE Internet of Things Journal*, Vol. 8, No. 14, pp.11476–11486.

Malik, R.Q., AlSattar, H.A., Ramli, K.N., Zaidan, B., Zaidan, A., Kareem, Z.H., Ameen, H.A., Garfan, S., Mohammed, A. and Zaidan, R.A. (2019) 'Mapping and deep analysis of vehicle-to-infrastructure communication systems: coherent taxonomy, datasets, evaluation and performance measurements, motivations, open challenges, recommendations, and methodological aspects', *IEEE Access*, Vol. 7, pp.126753–126772.

Pasquier, T., Singh, J., Powles, J., Eyers, D., Seltzer, M. and Bacon, J. (2018) 'Data provenance to audit compliance with privacy policy in the internet of things', *Personal and Ubiquitous Computing*, Vol. 22, pp.333–344.

Reyad, M., Sarhan, A.M. and Arafa, M. (2023) 'A modified Adam algorithm for deep neural network optimization', *Neural Computing and Applications*, Vol. 35, No. 23, pp.17095–17112.

Teo, R.S. and Nguyen, T.M. (2024) 'MomentumSMoe: integrating momentum into sparse mixture of experts', *Advances in Neural Information Processing Systems*, Vol. 37, pp.28965–29000.

Terzi, R., Sagiroglu, S. and Demirezen, M.U. (2018) 'Big data perspective for driver/driving behavior', *IEEE Intelligent Transportation Systems Magazine*, Vol. 12, No. 2, pp.20–35.

Tu, B., Zhou, C., He, D., Huang, S. and Plaza, A. (2020) 'Hyperspectral classification with noisy label detection via superpixel-to-pixel weighting distance', *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 58, No. 6, pp.4116–4131.

Wang, X., Han, Y., Wang, C., Zhao, Q., Chen, X. and Chen, M. (2019) 'In-edge AI: intelligentizing mobile edge computing, caching and communication by federated learning', *IEEE Network*, Vol. 33, No. 5, pp.156–165.

Wen, J., Zhang, Z., Lan, Y., Cui, Z., Cai, J. and Zhang, W. (2023) 'A survey on federated learning: challenges and applications', *International Journal of Machine Learning and Cybernetics*, Vol. 14, No. 2, pp.513–535.

Zhang, H., Qi, Y. and Zhang, G. (2025) 'Comparative analysis of intelligent connected vehicle industry in China, United States and European Union from technology lifecycle perspective', *Kybernetes*, Vol. 54, No. 2, pp.749–770.