**6G Network Slicing and Traffic Optimization Based on Federated Learning**

**Adel Alkhalil1, \*,Mohammed Altamimi1, Abdulaziz Aljarwan1, Magdy Abdelrhman2, Yaser Altameemi3 and and Aakash Ahmad4**

1College of Computer Science and Engineering, University of Ha'il, Ha'il, 81481 Saudi Arabia; [a.alkalel@uoh.edu.sa](mailto:a.alkalel@uoh.edu.sa)*,* [a.aljrwan@uoh.edu.sa](mailto:a.aljrwan@UOH.EDU.SA), [mh.altamimi@uoh.edu.sa](mailto:mh.altamimi@uoh.edu.sa)

2 Applied College, University of Ha'il, Ha'il, Saudi Arabia, 81481, [ma.abdallah@uoh.edu.sa](mailto:ma.abdallah@uoh.edu.sa)

3 University of Ha'il, Ha'il, 81481 Saudi Arabia; y.albakry@uoh.edu.sa

4  School of Computing and Communications, Lancaster University Leipzig, Germany; [a.ahmad13@lancaster.ac.uk](mailto:a.ahmad13@lancaster.ac.uk)

\*Corresponding Author: Adel Alkhalil, Email: [a.alkalel@uoh.edu.sa](mailto:a.alkalel@uoh.edu.sa)

**Abstract:** The prominent feature ofautonomous vehicles is collecting real-time data in the form of road images, video through on-board sensors and cameras. Such data is then deployed to optimize the vehicular traffic network. This paper proposes a novel framework for traffic data optimization and network slicing in 6G. The main idea is automatically get the training sample from the global model. Higher sample learning accuracy is improved by deploying knowledge distillation-based training mechanism. The traffic visual data privacy is preserved using adaptive differential method. Experimentations are performed using vehicle and other datasets. Simulations results show that the proposed method has superior performance as compared with existing methods.

**Keywords:** 6G, network slicing, IoV, traffic data analysis, AI.

**1 Introduction**

The Internet of Vehicles (IoV) is a distributed vehicle network that is one of the key technologies for realizing intelligent transportation systems and autonomous driving [1]. By connecting vehicles to the Internet according to commonly established standards, the IoV allows data exchange between vehicles and between vehicles and transportation systems. For example, vehicle-to-cloud communication (V2C) enables the sharing of traffic data between different vehicles. Based on the traffic data shared by vehicles, artificial intelligence (AI) models can pool learning experiences from various vehicles, thereby improving the intelligence level of the entire autonomous driving system. The synergy between the IoV and AI enables autonomous vehicles to better cope with complex traffic conditions, improve driving safety and efficiency, and continuously promote the advancement of autonomous driving technology.

Intelligent traffic recognition model is one of the key components for autonomous vehicles to achieve safe, compliant and efficient driving [2-3]. Various types of traffic data can be collected to train the recognition model. Based on the accurate recognition and interpretation of traffic conditions by the model, autonomous vehicles can better integrate into the real road traffic environment and provide a safer travel experience for drivers and pedestrians. However, the IoV data, such as traffic sign data, has visual color difference, distortion and damage, which will cause the recognition accuracy of the current model to decrease. To solve this problem, the visual data such as real-time images and videos of the road can be captured by the on-board camera of the autonomous vehicle, and continuously shared to achieve dynamic update of the recognition model and improve the recognition accuracy of the model.

Optimizing intelligent recognition models based on continuous sharing of IoV data faces problems such as inefficient sample collection, catastrophic forgetting, and privacy leakage. Specifically, due to the limited memory space of the vehicle, a “circular storage” strategy is usually adopted for data management, that is, covering the oldest old data to ensure that the storage space is always occupied by the latest data. If the self-driving car indiscriminately collects all traffic data encountered on the road, a large amount of traffic data will be repeatedly collected, resulting in a small proportion of samples that are conducive to improving the generalization ability of the model within the limited storage space. In addition, if the collected samples are manually labeled, the collection efficiency will be significantly reduced.

If the collected IoV data is sent directly to a server or a third-party organization, it is easy to cause privacy leakage because it can be associated with the travel routes of drivers or passengers. For this reason, federated learning (FL) is often used to collaboratively train IoV data from different vehicles without leaving the vehicle. However, during the training process of FL, cyclic storage will cause the sample data on the vehicle to change dynamically, and the model training based on new data will forget the knowledge learned by the original model, leading to catastrophic forgetting. This makes it impossible for the new model to learn the knowledge corresponding to the local old data and the old knowledge from the global model. In addition, although FL can avoid direct leakage of original data on the vehicle, attackers can still infer sensitive information of vehicle users from model parameters through model inversion attacks, inference attacks, etc.

In order to optimize the intelligent recognition model in real-time and enhance the efficiency, accuracy and security of model training, this paper combines knowledge distillation, differential privacy and other techniques to design a secure and efficient FL architecture for continuous sharing of IoV traffic data. This architecture can support vehicles to dynamically update traffic recognition models, ensure that the recognition model has a high generalization ability, assist vehicles to accurately identify various types of traffic information, improve the safety and reliability of autonomous driving, and provide support for the further development of autonomous driving technology. The main contributions of this paper include three aspects:

1. An efficient and secure FL architecture is proposed to improve the accuracy and privacy security of intelligent traffic recognition model training, and an efficient sample collection method is designed based on the global model to avoid the collection of invalid samples, thereby improving the storage space utilization and model learning efficiency.
2. A dual knowledge distillation training method is designed on the local side to effectively avoid the catastrophic forgetting problem caused by the dynamic changes of vehicle data and ensure that every sample collected by the vehicle can participate in the learning of the model to improve the accuracy of the training model.
3. An adaptive differential privacy (DP) budget allocation algorithm is designed to provide customer-level privacy protection strength for the training model parameters of each vehicle. It adaptively allocates a reasonable privacy budget to each model parameter to reduce the negative impact of differential privacy noise on the availability of model parameters, thereby ensuring that the global model has higher recognition accuracy.

The automatic collection method samples based on the trained model, which improves the sampling efficiency, and the samples obtained based on this collection method can improve the convergence speed of model training. In addition, the knowledge distillation method is one of the most effective methods to solve the catastrophic forgetting problem. The dual knowledge distillation training algorithm proposed in this paper has a significant effect on improving model accuracy. Finally, the choice of privacy protection based on DP method can effectively avoid excessive communication and computing overhead, and reduce the problem of model accuracy degradation caused by differential noise. Among them, the automatic sampling method and the dual knowledge distillation training algorithm are combined to overcome the negative impact of limited vehicle storage on shared data learning. The adaptive differential privacy strategy acts on the learning process to prevent the model training parameters from leaking original data information. These three techniques are related to each other and work together to achieve safe and efficient federated learning for continuous sharing of IoV data.

**2 Related Work**

In the IoV, vehicles can collect various types of traffic and vehicle behavior data through on-board equipment. These data can be used to improve driving safety, autonomous driving and other applications. However, transmitting this data directly to a central server for training may involve privacy issues and high-cost data transmission. FL, as a privacy-safe distributed machine learning (ML) technology, was proposed by [4]. It can ensure that the original data of the vehicle does not leave the vehicle, and only the model parameters are shared between vehicles, thereby protecting user privacy and reducing the need for data transmission. The authors in [5] proposed a communication-efficient FL method FedAvg, which further reduced the communication cost of model parameter transmission. Subsequently, some high-performance federated learning schemes that can be used to train non-independent and identically distributed data have been proposed in [6-8].

At present, FL has been widely used in intelligent transportation systems, IoV and other fields [9-10]. The authors in [11] designed an efficient FL scheme for traffic sign recognition by combining pulse neural networks. The authors in [12] proposed a FL method for enhancing image recognition of autonomous vehicles. In addition, some high-performance federated learning schemes for IoV have also been proposed. The authors in [13] proposed a semi-synchronous federated learning protocol with dynamic aggregation to improve the learning efficiency of machine learning models in IoV. The authors in [14] designed a 6G IoV two-layer FL based on heterogeneous model aggregation, aiming to improve the convergence speed and accuracy of the model. The authors in [15] designed a robust hierarchical FL for IoV to prevent poisoning attacks from malicious clients. However, these schemes do not consider the catastrophic forgetting problem caused by dynamic changes in vehicle data and the privacy leakage problem caused by model parameters.

Catastrophic forgetting refers to the phenomenon that when learning new tasks or new data in a neural network or ML model, the performance of the old tasks or data that have been learned will be significantly reduced. In order to solve the problem of catastrophic forgetting, the authors in [16] proposed learning without forgetting method, which uses knowledge distillation to generate data and labels of old tasks to transfer the knowledge of old tasks to the new model. The authors in [17] proposed the elastic weight sharing method, which protects the parameters of the learned tasks from the interference of new task learning by introducing regularization terms to the model parameters. The authors in [18] designed an incremental classifier and representation learning based on data replay. This scheme uses a storage buffer to store samples of old tasks and uses the samples of old tasks together with the samples of new tasks for training to alleviate the problem of catastrophic forgetting. However, the schemes described in references [16-18] cannot be directly used to solve the catastrophic forgetting problem caused by the dynamic changes of vehicle data, which causes the local model and the global model to exist at the same time.

In the FL architecture, the model parameters transmitted by the local client may be attacked by an honest but curious server, such as model inversion attack [19], inference attack [20-21], etc., which may lead to the leakage of private information [22]. In order to enhance the privacy protection capability of FL and prevent model parameters from leaking user private information, many privacy-preserving FL schemes have been proposed. The most well-liked and effective methods among them are those based on blockchain [25], homomorphic encryption (HE) [24], secure multi-party computation (SMC) [23], and differential privacy [26-28]. However, the communication cost of the SMC-based scheme is high. The computational overhead of the HE and blockchain-based schemes is too large. The differential privacy-based scheme has obvious advantages in communication and computational overhead, but the added differential privacy noise will disturb the true value of the model parameters, and the noise will be superimposed as the model training progresses, affecting the model convergence or the accuracy of the model training. To this end, designing a FL architecture that takes into account both strong privacy protection and high accuracy is a difficult problem that needs to be solved urgently.

Table 1 summarizes the advantages and disadvantages of the above-mentioned related work in terms of privacy protection, learning efficiency, and model accuracy. Through comparison, it can be seen that designing a FL architecture that takes into account strong privacy protection, efficient learning, and high accuracy is a difficult problem that needs to be solved urgently. The proposed approach is committed to solving this problem.

**Table 1:** Summary of related works.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Performance metric | Literature | | | | Proposed |
| [11 ~ 15] | [16~18] | [23~25] | [26~28] |
| Strong privacy protection | X | X |  |  |  |
| Efficient learning |  |  | X |  |  |
| High accuracy | X |  |  | X |  |

**3 Fundamental Concepts**

In this section, this article mainly introduces the definitions and basic knowledge related to FL, differential privacy, and distilled knowledge.

**3.1 Fundamental Concept of Federated Learning**

FL is a kind of distributed machine learning in which a central server coordinates the collaborative training of a model by several dispersed clients [29]. With FL, the client only has to upload the model parameters—such as gradients or model weights—to the central server; the original data set is kept on the client's local device. Generally speaking, federated learning may be broken down into several training cycles. The server gives the client the starting model or global model update at the beginning of each training round. Then each client uses its own data to train the model locally. Finally, the trained model update is uploaded to the server for aggregation to obtain a new global model. The above operations are repeated many times until the model converges, and the training is completed.

Assume there are clients, and the sample index set is , where the set stores the sample index of client . Let denotes the -th client sample, thus, the owned samples by such client can be expressed as , and the number of samples of client is expressed as . The loss function corresponding to a single sample is usually defined as .

In federated learning, the optimization model can be expressed as:

(1)

Where represents the total number of samples in training, and:

(2)

The Federated Average Algorithm (Fed-Avg) aims to reduce the number of communication rounds required for model training by using the client to perform more calculations, thereby improving training efficiency. In FedAvg, the client uploads model weights rather than model gradients to the central server. After receiving the locally uploaded model weights, the server performs weighted aggregation, i.e.:

(3)

Determine the global model weight of the new round , until the model converges. Currently, FedAvg has been widely used in distributed learning as an efficient and robust technology. The federated learning architecture designed in this paper adopts the update method of FedAvg.

**3.2 Differential Privacy**

The authors in [30] proposed a differential privacy protection model. Differential privacy can resist knowledge attacks from any background and can achieve quantitative analysis of the privacy protection level through privacy budget. The differential privacy protection model was originally applied in the field of data statistical query to protect the privacy information of users during query. After expansion, it was widely used in the fields of data publishing and data mining. Moreover, interactive statistical query and non-interactive data publishing can both use differential privacy models to protect privacy information.

Let denotes a set of datasets, and a relation . Then for any two adjacent data sets and , and any output , the following conditions are met:

(4)

Where is the privacy budget. That is, the smaller is, the stronger the privacy protection.

In federated learning, and can be defined as two different types of adjacent datasets, namely, sample-level adjacent datasets [27, 31-32] and client-level adjacent datasets [33-35].

1. Sample-level adjacent datasets. Assume that and are two datasets containing training samples. If and differ in only one sample, they are called sample-level adjacent datasets.
2. Client-level adjacent dataset. Assume that and are datasets of two training samples, each of which belongs to a unique client. If and differ by only one client-level sample, then and are client-level adjacent.

For example, suppose and are two datasets. If denotes one sample, then and are sample-level adjacent. If it is a sample of client, then and are client level adjacent.

Let query function to preserve privacy and the corresponding sensitivity can be calculated as . This paper adopts Gaussian noise mechanism as an approximate method to provide privacy protection, which can be expressed as , where denotes the noise matrix.

In addition, differential privacy also has the important property of sequential combination.

**Theorem 1:** Sequential combination [36]. Let mechanisms satisfy -DP respectively. Then, the mechanism formed by satisfies:

-DP (5)

**3.3 Knowledge Distillation**

Knowledge distillation [16, 37] aims to transfer the knowledge of the teacher model to the student model. It is a method that can be used to solve the problems of transfer learning and forgetting. The optimization model of the distillation network can be expressed as:

(6)

Among them, and represents the soft labels predicted by the teacher model and the soft labels predicted by the student model, respectively, represents the output of the student model, and represents the hard label, which is the true output of the training sample. The loss function consists of two parts: distillation loss and cross entropy loss . Distillation loss refers to the difference between the predictions of the student model and the predictions of the teacher model. The cross entropy loss is the difference between the student model prediction and the hard label. The distillation loss can be converted into the cross entropy loss between the soft label and the soft label . Soft labels usually refer to the probability distribution of categories, and in neural networks they are mostly Softmax outputs processed based on the temperature parameter . Assuming that the Softmax output of the student model is , , then:

(7)

Cross entropy is a measure of the difference between two probability distributions, and is often used to measure the difference between the predicted distribution of a model and the true distribution. Assuming there are two probability distributions and , the cross entropy calculation is expressed as:

(8)

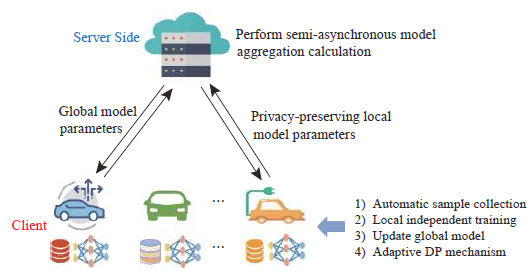
**4 Methods/Experimental**

This section introduces the proposed federated learning system model for dynamic update of traffic information recognition, the automatic sampling method based on the global model, the local training algorithm based on double knowledge distillation, and the adaptive differential privacy strategy.

**4.1 System Model**

This section introduces the overall architecture of the federated learning model based on differential privacy technology and the implementation process of one round of federated training.

In order to improve the generalization ability of the traffic information recognition model and ensure the privacy and security of training data, this paper proposes a secure and efficient federated learning architecture. As shown in Figure 1, the proposed framework mainly consists of a central server and multiple clients. Among them, the server is assumed to be honest but curious, that is, it honestly abides by the federated learning aggregation protocol, but is curious about the sensitive information of the client and steals the user's privacy information through various attacks. In real scenarios, the server may be a third-party organization or enterprise. The main goal is to use client data to improve the traffic information recognition model to improve the performance of its products (such as self-driving cars, recognition systems, etc.). The clients participating in the training are mainly various types of smart vehicles on the road. In this paper, it is assumed that the client is an honest node, which is mainly responsible for selectively collecting traffic information samples, completing the training and updating of the recognition model, and performing differential privacy processing on the updated model parameters.



**Figure 1:** Proposed system model.

During the training process, the server performs semi-asynchronous model parameter aggregation calculation. That is, the aggregation operation is performed only when the server receives a preset number of local model parameters. This aggregation method can avoid the inefficiency caused by synchronous aggregation and the parameter outdated problem caused by asynchronous aggregation. Before the federated training, it is assumed that the server obtains an initial global model based on the public traffic data sample training. Algorithm 1 gives the specific implementation process of the proposed scheme to complete one round of federated learning training process.

|  |
| --- |
| **Algorithm 1:** Federated learning training |
| **Input:** Current data of client , the previous round global model parameter , global model parameters  **Output:** Global model parameters  **Initialization:** Number of clients is and semi-asynchronous aggregation threshold   1. Collect new traffic data samples using previous round of global model 2. Based on the local model parameters and , local independent training is completed to obtain 3. Based on , , training is performed, is updated, and the model parameters are obtained 4. Perform adaptive differential privacy processing: 5. Upload to the server by the client 6. Determine new global model parameter from local model parameters aggregation |

Algorithm 1 mainly includes six steps:

1) The local client downloads the current global model and collects and saves samples based on the global model;

2) The local model is trained separately based on the collected data until the local model converges;

3) The global model is trained based on the local model and the sample data saved by the current client;

4) According to the adaptive differential privacy strategy, the updated global model parameters are privacy processed;

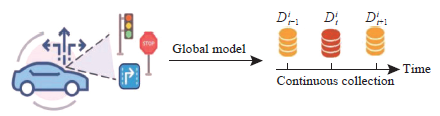
5) The privacy-protected global model parameters are uploaded to the server;

6) After receiving the agreed number of client model parameters, the server performs aggregation calculation to obtain a new global model.

Repeat the above 6 steps to dynamically update the traffic information recognition model, improve the generalization ability of the recognition model, ensure more accurate recognition of various types of traffic, and optimize autonomous driving services.

**4.2 Automatic Acquisition Based on Global Model**

The local vehicle downloads the global model of the server and uses the global model to take photos and identify traffic targets on the road. If the maximum probability value of the model's Softmax output is less than *a* or greater than *b*, no collection is performed. The lower threshold *a* is to limit the error rate of the automatic label setting during collection and reduce the negative impact of erroneous samples on model learning. The upper threshold *b* is to avoid wasting resources to repeat the learning of knowledge that has been mastered well. If the maximum probability value of the model's Softmax output is within the preset threshold range, i.e. (*a*, *b*), the image is collected, where the image is the feature of the sample, and the output corresponding to the maximum probability is the label of the sample, which is saved in the vehicle memory as the latest sample, and the earliest saved sample in the memory is deleted. In each sample collection process, the local vehicle operates based on the latest global model currently downloaded, and sample collection can be continued, as shown in Figure 2. Assuming that the currently collected samples are represented as , when the newly collected data completely covers , the new sample set is represented as .



**Figure 2:** Vehicles feature collection using automatic global model.

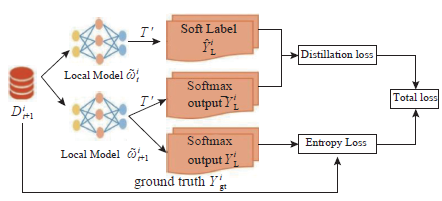
The samples collected in this way can improve the generalization ability of the model, so that the model can more accurately identify different types of traffic objects after training. At the same time, it avoids repeated collection of samples corresponding to the knowledge that has been learned well. In addition, the samples are automatically collected based on the global model, which significantly reduces the workload of manually setting labels for samples. As the recognition rate of the global model gradually increases, the samples that can be used to optimize the model will also be reduced accordingly, further reducing the workload of sample collection.

**4.3 Local Training Algorithm for Dual Knowledge Distillation**

In order to ensure that all samples collected by local vehicles participate in model learning, this paper designs a local training algorithm based on dual knowledge distillation, as shown in Algorithm 1. Algorithm 1 mainly includes two parts: local independent training and global model update, which are completed on the client. Local independent training saves the knowledge corresponding to the samples covered during the collection process in the local model. Global model update integrates the knowledge learned by the local model and the currently collected data into the global model.

4.3.1 Local independent training

This paper designs a method based on knowledge distillation to deal with the catastrophic forgetting problem caused by dynamic changes in data. In order to prevent the previous data from being forgotten, the learned knowledge is stored in the model, and in the subsequent training, the is used as a teacher model to pass the learned knowledge to the student model , as shown in Figure 3. This method ensures that both previous samples and current samples participate in the learning of the student model .



**Figure 3:** Local independent training for dynamic changes in data.

The specific implementation is as follows:

1. Initialization. The initial local model is the global model that the local vehicle first downloads from the server. When the vehicle memory is filled with the first collected samples, fine-tuning is performed on the initial local model until the model converges to obtain a new model. The model parameters are expressed as .
2. Dynamic update of local model. As described in first step of Algorithm 2, when the newly collected local samples completely cover the samples in the previous round of local model training and the global model parameters are not received at this time, the local model is updated and trained. The updating process is shown in Figure 3. First, the sample data generates a soft label based on the local model . Secondly, the local model generates the distilled Softmax output and the regular Softmax output based on the data . The total loss function of model training is the sum of distillation loss and cross entropy loss, so the optimization model can be expressed as:

(9)

Among them, is the true label groudtruth corresponding to the sample , is the distillation loss, is the cross entropy loss, and the function represents the cross entropy calculation. When the model converges, the local model update ends, as shown in step 2 of Algorithm 2.

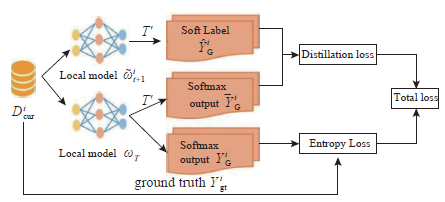
4.3.2 Global model update on local side

As described in step 3 of Algorithm 2, when the local end receives the global model parameter , it updates the model based on the current local model and the sample . The specific process is shown in Figure 4. Combined with the knowledge distillation method, the knowledge learned by the local model is saved while fine-tuning the global model. The optimized model can be expressed as:

(10)

The global model is updated by referring to the FedAvg scheme, and the local end adopts the small batch gradient descent update method, see steps 4~6 of Algorithm 2. In order to facilitate the subsequent privacy protection processing, the gradient obtained by training is clipped, as shown in step 13 of Algorithm 2. Finally, the updated is calculated, denoted as , as shown in line 14 of Algorithm 2.

|  |
| --- |
| **Algorithm 2:** Local training for dual knowledge distillation |
| **Input:** , global model , model parameters wit of  **Output:** Global model parameter  **Initialization:** The upper limit of client storage samples is *S*, and the gradient clipping threshold is *C*.   1. When client does not receive 2. until converges, 3. When the client downloads , the sample currently saved on the local end is represented by 4. is divided into batches 5. , and 6. for number of local iterations: 1 to do 7. for batch do 8. end for 9. end for 10. return |



**Figure 4:** Optimized global model with updates on local side.

**4.4 Adaptive Differential Privacy Algorithm**

The adaptive differential privacy strategy ensures strong privacy protection of client data while minimizing the impact of differential noise on global model recognition accuracy. In order to do this, this work creates a model parameter permutation rule to add adaptive differential noise to the model parameter and provides client-level privacy protection methods. The specific mechanism is shown in Algorithm 3. In Algorithm 3, steps 1 and 2 implement the client-level differential privacy setting, calculating the sensitivity corresponding to the current global training round number and the noise scale to be added. Step 3 is the designed permutation rule to select the model parameter that is close to the model parameter and does not leak the customer's privacy as the final value uploaded to the server. The implementation process of customer-level privacy protection and replacement rules will be introduced in detail below.

|  |
| --- |
| **Algorithm 3:** Adaptive differential privacy processing |
| **Input:** , , , learning rate  **Output:** Global model parameters   1. Calculate the global sensitivity at the *T*th round of training: 2. Calculating the noise scale 3. if 4. if 5. else 6. end if 7. else 8. end if 9. return |

4.4.1 Privacy protection on client-level

This paper applies differential privacy by using the concept of customer adjacent datasets to model training. The query function *Q* of client-level differential privacy is defined as the model update training procedure within a single global training round which is expressed as:

(11)

The privacy sensitivity calculation corresponding to the query function is expressed as:

(12)

Among them, is the set of current samples of the client in the *T*th round of global model training, and and represent samples of different clients. In order to quantify the sensitivity, this paper clips the gradient obtained by training, and the gradient clipping threshold is *C*, as shown in step 7 of Algorithm 2. According to the above settings, the sensitivity in the *T*th round of global training can be calculated.

4.4.2 Model parameter substitution rules

In the step of adding differential noise, this paper adopts an adaptive strategy [38]. The local client adaptively chooses to upload one of the three model parameters after differential privacy processing, the model parameters uploaded locally in the previous round, or the current global model parameters to the server. Since the model after differential privacy processing cannot restore the original information, and the model parameters uploaded locally in the previous round and the current global model parameters are both known to be privacy-safe. Therefore, the adaptive strategy designed in this paper ensures that the uploaded model parameters will not cause privacy leakage. Selecting the model parameters that are most conducive to model training is the goal that the designed replacement rule needs to achieve.

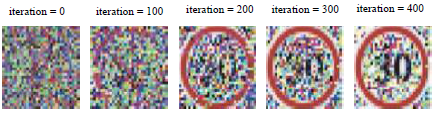
To achieve this goal, this paper uploads the model parameters of the three types of model parameters that are closest to the model parameters obtained by local training to the server to ensure the accuracy of the global model. The specific implementation process is shown in Algorithm 3. Taking as an example, if any difference between and or is less than the deviation caused by adding differential noise to , the model parameter finally uploaded to the server is the one that is closer to between and . Otherwise, the model parameter after differential privacy processing is uploaded. The specific process of replacing rules can be found in step of Algorithm 3. In addition, due to the inconsistency between the distribution of differential privacy noise and model parameters, a hyperparameter is set during the gap comparison process to balance the differences between different distributions.

**5 Traffic Data Security Analysis**

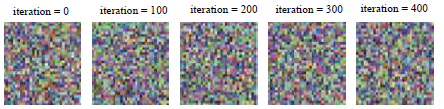
In federated learning, the client will pass the model parameters to the server or a third-party organization for aggregation operations. If the server is honest and curious, it may mine the original data information through methods such as reversal attacks. To verify the effectiveness of the differential privacy protection strategy proposed in this paper, the model parameters without privacy protection and the model parameters processed by the privacy strategy in this paper are reversed. The experimental results are shown in Figure 5.

Figure 5 (a) shows that the model parameters without privacy protection can restore the original information of the sample through inversion attack. Figure 5 (b) shows that the model parameters processed by the privacy policy in this paper can effectively resist the inversion attack without revealing the original sample information. Therefore, the proposed method provides privacy protection for client data. In addition, through theoretical analysis, it can be obtained that the privacy protection strategy designed in this paper can achieve customer-level privacy protection. The specific analysis is as follows:

Assume that represents an-dimensional query function and its sensitivity is . The Gaussian noise mechanism adds noise of scale to the output of each client (i.e., model weight), satisfying .



(a)



(b)

**Figure 5:** Reversal attack illustration in traffic sign recognition. (a) without privacy protection; (b) with differential privacy protection.

**Theorem 2:** Assuming that the privacy budget is an arbitrary value, when and , then the proposed method satisfies the client-level -DP. Among them, the condition is the upper limit constraint of the constant , , , represents the probability of each client, denotes the entire global training rounds, and denotes the number of rounds with privacy budget.

Proof: In the proposed scheme, the client training process of each user is the query function . Then, in the *T*th round of global model training, the query function of client is represented by , and the output of the query function is the model weight obtained by the client based on local data training. For this purpose, the Gaussian noise mechanism will return .

Therefore, the sensitivity of the query function can be expressed as:

(13)

Since is the gradient obtained after gradient clipping, , where is the gradient clipping threshold. Therefore, we can calculate . The noise scale , then the privacy loss can be calculated as:

(14)

Where represents the output of the model and is the difference between and . Therefore, the privacy loss can also be expressed as:

(15)

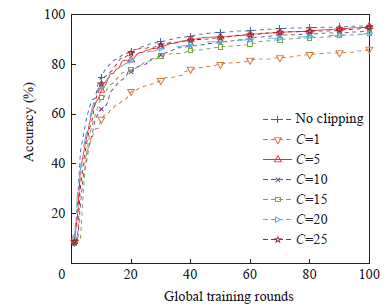
According to Theorem A.1.9 in [30], when , the mechanism satisfies -DP, where and .

Assume that the Gaussian noise mechanism is , which represents the entire global training process, and is the global training of the *T*th round. The query function of each round is only related to the previous global model weight and its local model . Since the global model weight is different in each round, are independent of each other. Among them, the global model weight is public. If we assume that the attacker knows the information , and allocates the same privacy preset and in each round, and the customer is aware that the amount of training sessions allotted for the privacy budget is , then according to Theorem 1, the Gaussian noise mechanism satisfies -DP, and Theorem 2 holds. In the proposed scheme, the local model is private. For this reason, the proposed method can actually provide slightly higher privacy protection than -DP.

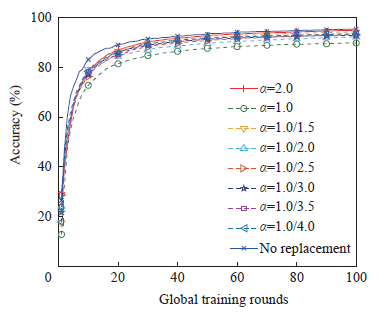
**6 Results and Discussion**

This paper mainly conducts performance evaluation experiments on the proposed scheme on the German traffic sign recognition benchmark (GTSRB). The GTSRB database has more than 50,000 images, covering 43 different categories of traffic sign images. The same category contains traffic signs with different shapes, large color differences, and blurred icons. In the experiment, the size of each image is set to 30×30×3, and each client is randomly assigned the same number of training samples. The convolutional neural network (CNN) is used to train the traffic sign recognition model. In addition, the generality of the main algorithm was verified based on the vehicle recognition dataset, which contains 1, 600 vehicle images and 10 different vehicle categories. The experiment is mainly based on the federated learning architecture for simulation. The client uses the small batch stochastic gradient descent training method, the batch size is set to 50, and the number of local iterations is 10. The experiment uses Python 3.10 programming language, and the experimental environment is configured with RTX 3070Ti, 32GB memory, and Windows 11 operating system.

First, the two important hyperparameters involved in the proposed scheme (the gradient clipping threshold C and the hyperparameters in the model parameter replacement rule) were experimentally set. As shown in Figure 6(a), when = 5, the training accuracy is close to that obtained by model training without gradient clipping. During differential privacy processing, the smaller the threshold , the less differential noise is added, which is more beneficial to ensure high model accuracy. Therefore, in the proposed scheme, the gradient clipping threshold = 5 is set. In order to reduce the negative impact of model parameter replacement on model training accuracy, it is crucial to set reasonable hyperparameters. As shown in Figure 6 (b), when =2, the model training is least affected. Therefore, in the subsequent experiments, the hyperparameter = 2 is set.



(a)

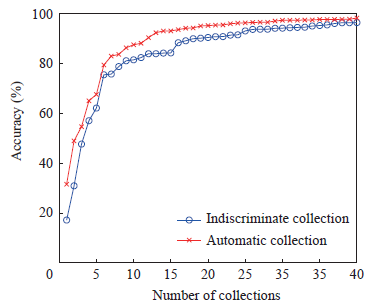


(b)

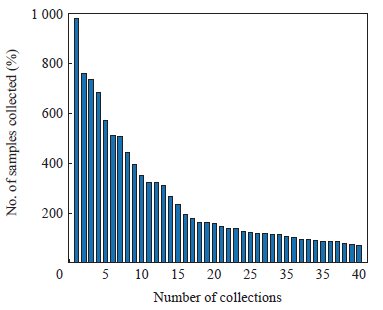
**Figure 6:** Accuracy comparison of algorithm for traffic data. (a) Gradient clipping threshold ; (b) Hyperparameter .

In order to verify the efficiency of the automatic acquisition method based on the global model, the effects of the automatic acquisition method proposed in this paper and the indiscriminate acquisition method on model training were compared. In this comparative experiment, the maximum probability value threshold range of the model Softmax output was set to [0.5, 0.9].

As shown in Figure 7 (a), the automatic collection method collects traffic sign samples that have not been learned well in a targeted manner, which can make the model more accurate and improve the convergence speed of the model. The indiscriminate collection method has the problem of repeatedly collecting learned traffic sign samples, which will reduce the generalization performance of the model. It can be seen that the method of setting the threshold ensures the effectiveness of the collection. In addition, the experiment selected 1,000 traffic sign samples to simulate the traffic sign sample set that can be collected. If indiscriminate collection is adopted, it means that any traffic sign encountered by the vehicle during operation will be collected, which is easy to cause a waste of storage space. The automatic collection method based on the global model proposed in this paper, as the generalization ability of the model is improved, the samples that can be collected will become fewer and fewer, as shown in Figure 7 (b). This method effectively releases the vehicle storage space, does not need to collect redundant and repeated samples for model training, and further optimizes the vehicle storage and computing resources.



(a)

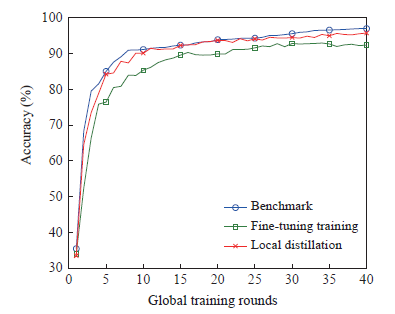


(b)

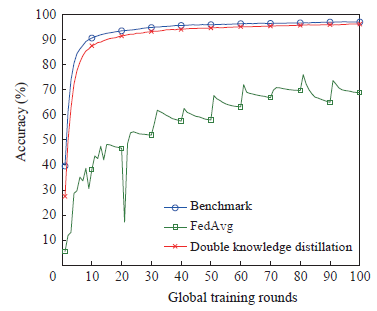
**Figure 7:** Traffic data recognition performance evaluation of proposed algorithm under GTSRB dataset. (a) collection methods accuracy; (b) changes in the number of samples that can be collected.

The performance of the dual knowledge distillation algorithm proposed in this paper is evaluated based on the GTSRB and vehicle recognition datasets. In order to better reflect the feasibility and superiority of the proposed algorithm, experimental comparisons are completed based on different hypothetical scenarios.

In order to verify the performance of local independent training, an experimental verification is conducted for a single local user. In this experiment, in order to simulate the dynamic change of user local data, 775 sample data are randomly selected for each round of training. The baseline algorithm, fine-tuning training and the local independent training method (local distillation) proposed in this paper are compared. The comparison baseline is trained based on all selected data, while the fine-tuning training directly updates the model based on the user's current data in each round. The experimental results are shown in Figure 8(a). The knowledge learned by the local training method based on distilled knowledge is close to that of the baseline scheme, and is higher than the accuracy based on fine-tuning training. This shows that the local model trained based on distilled knowledge can learn more sample information and can effectively alleviate the impact of data changes on model training. In addition, based on the federated learning architecture, the catastrophic forgetting problem caused by dynamic changes in data and the effect of the dual knowledge distillation algorithm proposed in this paper in dealing with the catastrophic forgetting problem are verified. In this experiment, 20 users jointly participated in the training, and the number of training samples for each user in each round was 500. Due to the limited storage space of the vehicle, new data will overwrite old data. To simulate this scenario, assume that every 10 rounds of training, the user sample data is completely replaced by 500 new sample data that are not repeated before. The experimental results are shown in Figure 8(b). The traditional FedAvg algorithm is prone to catastrophic forgetting when processing dynamic data, resulting in a decrease in model accuracy. The training algorithm based on dual knowledge distillation proposed in this paper can effectively solve the catastrophic forgetting problem and obtain higher training accuracy.



(a)

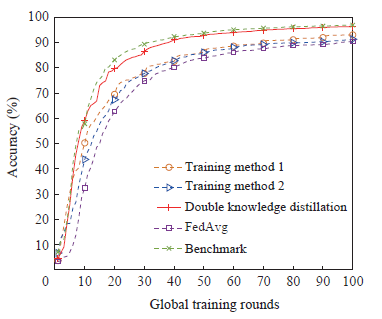


(b)

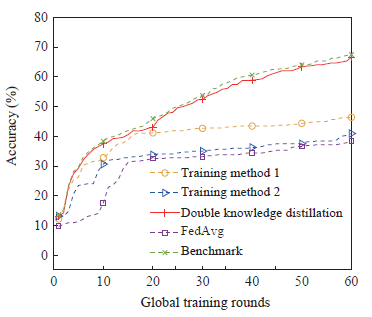
**Figure 8:** Accuracy comparison of algorithms. (a) Local distillation; (b) Double knowledge distillation.

The dual knowledge distillation algorithm proposed in this paper mainly consists of two parts: local independent training and global model update. In order to verify the rationality and necessity of each component, relevant comparative experiments were conducted based on the GTSRB dataset and the vehicle recognition dataset.

The experiment compared five different federated training methods: Training method 1, that is, local independent training combined with global model update not based on distillation. Training method 2, that is, no local independent training combined with global model update based on distillation. Local training method of dual knowledge distillation. FedAvg training method. Comparison benchmark scheme, that is, model training based on all existing data. In the experiment based on GTSRB, the number of users participating in the training is 20, the maximum number of stored samples for each user is 100, and 10 new sample data are collected in each round of training. Figure 9 (a) shows the experimental comparison results. The model accuracy obtained based on the dual knowledge distillation algorithm is close to the comparison benchmark and higher than the model accuracy obtained by other training methods. In the experiment based on the vehicle recognition dataset, due to the small amount of data, the number of users participating in the training was set to 10, the maximum number of samples stored for each user was 20, and 2 new sample data were collected in each round of training. Figure 9 (b) is the experimental comparison result, which also shows that the dual knowledge distillation algorithm performs well, and the local independent training and global update method based on distilled knowledge in the algorithm are helpful to solve the catastrophic forgetting problem and ensure high model accuracy.



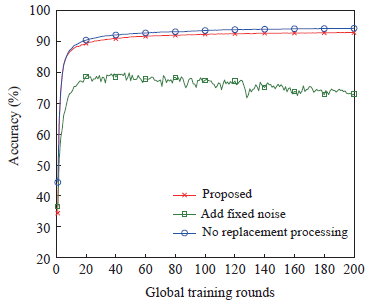
(a)



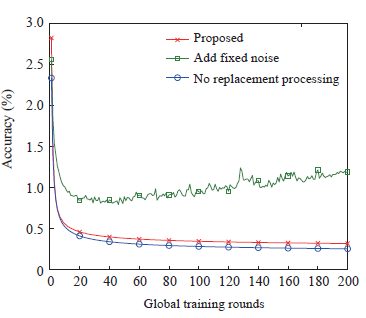
(b)

**Figure 9:** Accuracy comparison of different federated learning algorithms. (a) GTSRB dataset; (b) Vehicle recognition dataset.

Different differential privacy protection strategies are compared, including adding fixed-size differential noise to each round of training, adding variable privacy noise without model replacement, and adaptive differential privacy in the proposed scheme. As shown in Figure 10, when adding fixed noise to each round, as the training progresses, the noise will be superimposed and the model will not converge. In order to ensure model convergence, the added noise is required to be gradually reduced during training. To this end, the noise size in this paper is set to , where , is the initial learning rate, is a constant, is the number of global training rounds, = 4, = 5. As can be seen from Section 5, the fewer rounds of privacy budget are allocated, the stronger the differential privacy protection provided.



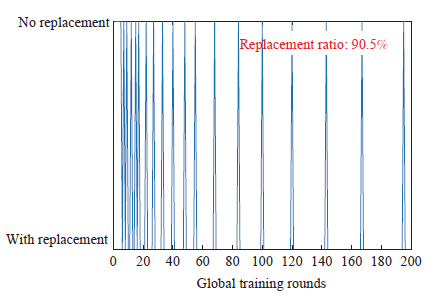
(a)



(b)

**Figure 10:** Performance comparison of proposed and existing algorithms. (a) accuracy; (b) information loss.

In the experiment on the model parameter replacement rate, the model was trained for 200 rounds. When model replacement occurred, it was marked as 0, and when no replacement occurred, it was marked as 1. Figure 11 shows the situation of model replacement in 200 rounds of training.



**Figure 11:** Comparison of model occurrence with and without replacement.

Among them, the number of model replacements reached 181 times, accounting for 90.5%. The number of rounds participating in budget allocation was greatly reduced, and the privacy protection strength was significantly improved. Combined with Figures 10 and 11, it can be concluded that the proposed scheme designed in this paper can ensure model convergence, and only sacrifice a small amount of model accuracy while significantly improving the privacy protection strength.

The proposed scheme can effectively deal with the problems of inefficient collection and catastrophic forgetting in the continuous sharing of Internet of Vehicles data. The proposed dual knowledge distillation training algorithm ensures the accuracy of the recognition model, and the adaptive differential privacy strategy significantly improves the privacy protection strength. It provides a safe and efficient training method for distributed model learning in IoV, and has a positive effect on promoting the development of autonomous driving, smart IoV and other fields.

**7 Conclusion**

This paper aims to solve the problems of inefficient collection, catastrophic forgetting and privacy leakage in the continuous sharing of Internet of Vehicles data, and designs a safe and efficient federated learning scheme. In this scheme, an automatic collection method based on a global model, a training algorithm based on double knowledge distillation and an adaptive differential privacy strategy are proposed to ensure that IoV data can be safely and continuously used for artificial intelligence tasks such as traffic information recognition, and high model accuracy can be obtained through training. Finally, the performance of the proposed scheme is evaluated through security analysis and experimental evaluation. The experimental results show that the proposed scheme performs well in processing IoV data with limited storage space, dynamic data changes and privacy requirements.

**Availability of Data and Materials:** Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

**Competing Interest:** The authors declare no competing interest regarding this study.

**Funding:** This study received no specific funding.

**References**

1. K. Sun and J. Xu, “Feature analysis of 5G traffic data based on visibility graph,” *Frontiers in Physics*, vol. 12, 2024. doi: 10.3389/fphy.2024.1477382
2. S. Paniego, E. Shinohara, and J. Canas, “Autonomous driving in traffic with end-to-end vision-based deep learning,” *Neurocomputing*, vol. 595, 127874, pp. 1-15, 2024.
3. Y. Yan, C. Deng, J. Ma, Y. Wang, Y. Li, “A traffic sign recognition method under complex illumination conditions,” *IEEE Access*, vol. 11, pp. 39185-39196, 2023.
4. O. Karachalios, A. Zafeiropoulos, K. Kontovasilis, and S. Papavassiliou, “Distributed machine learning and native AI enablers for end-to-end resource management in 6G,” *Electronics*, vol. 12, no. 18, pp. 1-16, 2023.
5. C. Wu, F. Wu, L. Lyu, Y. Huang, and X. Xie, “Communication-efficient federated learning via knowledge distillation,” *Nature Communications*, vol. 13, no. 32, pp. 1-18, 2022.
6. Y. Lv, H. Ding, H. Wu, Y. Zhao, and L. Zhang, “FedRDS: Federated learning on Non-IID data via regularization and data sharing,” *Applied Sciences*, vol. 13, no. 23, pp. 1-14, 2023.
7. S. Lo, Q. Lu, L. Zhu, H. Paik, X. Xu, and C. Wang, “Architectural patterns for the design of federated learning systems,” *Journal of Systems and Software*, vol. 191, 111357, pp. 1-16, 2022.
8. Y. Wang, D. Zhai, and Y. XIa, “RFVIR: A robust federated algorithm defending against Byzantine attacks,” *Information Fusion*, vol. 105, 102251, pp. 1145-1158, 2024.
9. M. Manias, and A. Shami, “Making a case for federated learning in the internet of vehicles and intelligent transportation systems,” *IEEE Network*, vol. 35, no. 3, pp. 88-94, 2021.
10. X. Ling, Z. Pengcheng, G. Jianping et al., “A survey of the social internet of vehicles: secure data issues, solutions, and federated learning,” *IEEE Intelligent Transportation Systems Magazine*, vol. 15, no. 2, pp. 70-84, 2022.
11. X. Kan, Z. Zhe, L. Bao et al., “Efficient federated learning with spike neural networks for traffic sign recognition,” *IEEE Transactions on Vehicular Technology*, vol. 71, no. 9, pp. 9980-9992, 2022.
12. B. Youssef, L. Alhmidi, Y. Bazi, and M. Zuair, “Federated learning approach for remote sensing scene classification,” *Remote Sensing*, vol. 16, no. 12, pp. 1-15, 2024.
13. L. Feiyuan, Y. Qinglin, L. Ruiqi et al., “Semi-synchronous federated learning protocol with dynamic aggregation in internet of vehicles,” *IEEE Transactions on Vehicular Technology*, vol. 71, no. 5, pp. 4677-4691, 2022.
14. Z. Xiaokang, L. Wei, S. Jinhua et al., “Two-layer federated learning with heterogeneous model aggregation for 6G supported internet of vehicles,” *IEEE Transactions on Vehicular Technology*, vol. 70, no. 6, pp. 5308-5317, 2021.
15. Z. Hongliang, Z. Yifeng, H. Hejiao et al., “Toward robust hierarchical federated learning in internet of vehicles,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 5, pp. 5600-5614, 2023.
16. L. Zhizhon, and D. Hoiem, “Learning without forgetting,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 12, pp. 2935-2947, 2017.
17. J. Kirkpatrick, R. Pascanu, N. Rabinowitz et al., “Overcoming catastrophic forgetting in neural networks,” *Proceedings of the National Academy of Sciences*, vol. 114, no. 13, pp. 3521-3526, 2017.
18. G. Ven, T. Tuytelaars, and A. Tolias, “Three types of incremental learning,” *Nature Machine Intelligence*, vol. 4, pp. 1185-1197, 2022.
19. H. Jeong, K. Chung, S. Hwang, and S. Son, “Targeted model inversion: distilling style encoded in predictions,” *Computers & Security*, vol. 144, 103967, pp. 513-524, 2024.
20. Z. Zhang, J. Ma, X. Ma, R. Yang, X. Wang, and J. Zhang, “Defending against membership inference attacks: RM learning is all you need,” *Information Sciences*, vol. 670, 120636, pp. 1475-1487, 2024.
21. S. Shin, M. Boyapati, K. Suo, K. Kang, and J. Son, “An empirical analysis of image augmentation against model inversion attack in federated learning,” *Cluster Computing*, vol. 26, pp. 349-366, 2022.
22. M. Afzal, A. Abdellatif, M. Zubair, M. Mehmood, and Y. Massoud, “Privacy and security in distributed learning: a review of challenges, solutions, and open research issues,” *IEEE Access*, vol. 11, pp. 114562-114581, 2023.
23. S. Lu, R. Li, W. Liu, C. Guan, and X. Yang, “Top-*k* sparsification with secure aggregation for privacy-preserving federated learning,” *Computers & Security*, vol. 124, 102993, pp. 345-356, 2023.
24. F. Chen, G. Yuanbo, W. Na et al., “Highly efficient federated learning with strong privacy preservation in cloud computing,” *Computers & Security*, vol. 96, 101889, 2020.
25. L. Yunlong, H. Xiaohong, Z. Ke et al., “Blockchain empowered asynchronous federated learning for secure data sharing in internet of vehicles,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 4, pp. 4298-4311, 2020.
26. Y. Zhu and W. Yan, “Traffic sign recognition based on deep learning,” *Multimedia Tools and Applications*, vol. 81, pp. 17779-17791, 2022.
27. E. Kim, and E. Lee, “Evaluating the impact of mobility on differentially private federated learning,” *Applied Sciences*, vol. 14, no. 12, pp. 1-16, 2024.
28. W. Kang, L. Jun, D. Ming et al., “Federated learning with differential privacy: algorithms and performance analysis,” *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3454-3469, 2020.
29. P. Kairouz, H. McMahan, B. Avent et al., “Advances and open problems in federated learning,” *Foundations and Trends in Machine Learning*, vol. 14, no. 1, pp. 1-210, 2021.
30. W. Dong, and X. Zhengquan, “Impact of inaccurate data on differential privacy,” *Computers & Security*, vol. 82, pp. 68-79, 2019.
31. K. Pan, Y. Ong, M. Gong, H. Li, A. Qin, and Y. Gao, “Differential privacy in deep learning: a literature survey,” *Neurocomputing*, vol. 589, 127663, pp. 812-823, 2024.
32. S. Tai, C. Dewi, R. Chen, Y. Liu, X. Jiang, and H. Yu, “Deep learning for traffic sign recognition based on spatial pyramid pooling with scale analysis,” *Applied Sciences*, vol. 10, no. 19, pp. 1-16, 2020.
33. L. Fang, B. Du, and C. Wu, “Differentially private recommender system with variational autoencoders,” *Knowledge-Based Systems*, vol. 250, 109044, pp. 1751-1763, 2022.
34. G. Andrew, O. Thakkar, B. McMahan et al., “Differentially private leaning with adaptive clipping,” *Advances in Neural Information Processing Systems*, vol. 34, pp. 17455-17466, 2021.
35. M. Asad, A. Moustafa and C. Yu, “A critical evaluation of privacy and security threats in federated learning,” *Sensors*, vol. 20, no. 24, pp. 1-14, 2020.
36. R. Valiente, D. Chan, A. Perry, J. Lampkins, et al., “Robust perception and visual understanding of traffic signs in the wild,” *IEEE Open Journal of Intelligent Transportation Systems*, vol. 4, pp. 611-625, 2023.
37. J. Guo, B. Yu, J. Maybank et al., “Knowledge distillation: a survey,” *International Journal of Computer Vision*, vol. 129, no. 6, pp. 1789-1819, 2021.
38. I. Ahmad, M. Adnan, N. Amin, A. Umer, A. Khurshid et al., “Adaptive and priority-based data aggregation and scheduling model for wireless sensor netowrk,” *Knowledge-Based Systems*, vol. 303, pp. 1517-1526, 2024.