



ARTICLE

A MAML-Based Federated Meta-Learning Framework for Wireless Network Traffic Prediction

Yuan Feng¹, Xiaotong Li², Xinhua Hu³, Jianwei Zhang⁴, Zengyu Cai^{2,*} and Liang Zhu²

¹School of Electronic Information, Zhengzhou University of Light Industry, Zhengzhou, China

²School of Computer Science and Artificial Intelligence, Zhengzhou University of Light Industry, Zhengzhou, China

³Xinyang Branch of China Mobile Communications Group Henan Co., Ltd., Xinyang, China

⁴School of Software Engineering, Zhengzhou University of Light Industry, Zhengzhou, China

*Corresponding Author: Zengyu Cai. Email: mailczy@163.com

Received: 25 June 2026; Accepted: 25 August 2026; Published: 22 October 2026

ABSTRACT: Wireless network traffic prediction enables operators to anticipate network trends, proactively develop network management strategies, and intelligently allocate network resources, thereby improving network service quality and enhancing users' internet experience. Existing centralized network traffic prediction methods require transmitting large volumes of traffic data, which consumes significant network resources, incurs extra communication costs, and faces difficulties in full data sharing due to privacy concerns. Federated learning is a distributed learning method for multi-client joint training without sharing between clients, which can effectively solve such problems. However, the network traffic prediction method based on federated learning has problems of low accuracy due to data heterogeneity and high communication resource consumption due to the large number of federated communication rounds. To address these issues, this paper combines model-agnostic meta-learning (MAML) and proposes a MAML-based federated meta-learning framework for wireless network traffic prediction, introducing the MAML meta-learning algorithm into the client training of federated learning to enhance the learning ability of the local model and train a global model that can quickly adapt to new network traffic patterns, thereby reducing the number of federated communication rounds and communication resource consumption; during the application phase, the global model is locally fine-tuned on a held-out adaptation set at each client to obtain high-precision personalized traffic prediction models. The central server adopts a mutual information (MI) based weighted aggregation scheme to improve global model generalization. Experimental evidence shows that this approach achieves lower errors and higher accuracy with far fewer rounds, thus markedly reducing the total communication overhead.

KEYWORDS: Wireless network traffic prediction; federated learning; meta-learning; MAML; mutual information

1 Introduction

With the rapid development and widespread application of wireless networks, the network infrastructure faces significant challenges in traffic processing, load balancing, and reliable management of heterogeneous devices [1]. Network traffic prediction technology, which forecasts future trends based on historical data, plays a crucial role in proactive resource scheduling and maintaining high network performance under heavy loads [2]. Existing network traffic prediction methods are dominated by centralized learning. However, this approach has several major drawbacks. Transmitting massive data to a central server consumes substantial network resources and increases communication overhead. It also impairs real-time performance. Most critically it raises severe data privacy and security.

Federated Learning (FL) emerges as a transformative distributed machine learning paradigm that addresses this challenge through decentralized collaboration [3]. This architecture fundamentally decouples data ownership from model training, thereby mitigating risks of privacy leakage, while still effectively harnessing distributed knowledge across the entire network. Federated learning protects user privacy and reduces data transfer in this way, which is especially suitable for scenarios with decentralized data and high privacy requirements. The coordinator builds the basic model and informs the participants about the basic structure and parameters of the model; each participant trains the model using local data and returns the results to the coordinator; the coordinator aggregates the models of each participant and builds a more accurate global model in order to improve the model performance and effect as a whole. Meanwhile, by integrating the model parameters of different nodes, federated learning can overcome the data island problem and improve the accuracy and reliability of prediction. However, for the wireless traffic prediction problem, the non-independent and identically distributed (non-IID) nature of traffic data presents significant data heterogeneity challenges for federated learning. Frequent client-server communication in federated learning raises network load and training time, so controlling overhead is essential for wireless traffic prediction.

To address data heterogeneity and high communication overhead in federated learning, we propose a MAML-based federated meta-learning framework for wireless network traffic prediction. At the client side, the MAML algorithm is adopted with a bi-level optimization strategy. The traffic data is divided into a support set and a query set. The support set is used to update model parameters, while the query set is used to evaluate model performance and further optimize the meta-parameters. This enables the model to adapt quickly to new tasks and traffic patterns, accelerating training convergence and reducing communication overhead. In the application phase, the global model is fine-tuned on a held-out adaptation set at each client before final prediction, thereby obtaining personalized models that better capture local traffic characteristics while ensuring that the final test set remains completely unseen. In addition, a weighted model aggregation based on mutual information is employed at the server, which assigns dynamic weights according to model importance and thereby enhances the generalization ability of the global model.

2 Related Work

2.1 The Current State of Wireless Network Traffic Prediction

Currently, deep learning has been increasingly applied in the field of network traffic prediction, becoming a mainstream method that significantly enhances the ability of network traffic prediction. Some researchers use linear models for prediction, such as Autoregressive Integrated Moving Average (ARIMA) model and its variants [4], but these models require smooth data, and the actual network traffic is characterized by suddenness, the accuracy of using these models is low. With the development of artificial intelligence, deep learning models have been used in a large number of traffic prediction tasks, Nie et al. [5] used a deep belief network model to predict network traffic, and Yang et al. [6] proposed a feed-forward neural network (ARIMA-BPNN) traffic prediction method based on simulated annealing (SA) optimization to obtain excellent prediction results. Recurrent Neural Network (RNN) can capture the temporal correlation of time series better, Fan and Ramakrishnan et al. used a variant of RNN, Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) for traffic prediction. Recurrent Unit for traffic prediction [7,8], which outperforms the RNN. Bi et al. [9] propose a traffic prediction model (ST-LSTM) that mixes Temporal Convolutional Network (TCN) and LSTM, where TCN is used to discover short-term dependencies in the data, and LSTM to capture long term dependencies and a filter is used to remove the data noise for more accurate prediction. Wireless networks have influences between network nodes and network traffic is characterized by spatiality, and many studies have been conducted to consider the spatial characteristics of traffic. Zhang et al. [10] used a densely connected convolutional neural network to capture the spatial

and temporal dependencies of network traffic, and Zhao, Vinchoff, and Li et al. [11–13] used a Graph Convolutional Network (GCN) to capture the spatial and temporal dependencies of network traffic. Geng et al. [14] proposed a traffic prediction model for temporal and spatial feature fusion, which extracts the temporal features of traffic using GRU based on self-attention mechanism, extracts the spatial features using GCN, and then fuses the features and generates the prediction results using Encoder-Decoder (EDC) structure. Andreoletti et al. [15] proposed a dynamic diffusion convolutional interaction graph neural network model, where the dynamic diffusion convolution module is used to capture spatial and dynamic features and the convolutional interaction model is used to extract temporal features.

Traditional centralized wireless traffic prediction methods require a central node to collect and process large volumes of traffic data from distributed clients. However, transmitting such raw data consumes significant network resources and time, often leading to congestion. Moreover, stringent privacy regulations make it difficult to gather all raw data from dispersed sources for model training.

With the maturity of edge computing, prediction and traffic management tasks can be deployed at the network edge for faster and more reliable distributed processing. Xiao et al. [16] proposed a distributed network traffic prediction method based on intelligent optimization to improve prediction accuracy in distributed environments. Federated learning, as a distributed machine learning framework, enables collaborative model training with edge clients while keeping data local, making it highly suitable for wireless network traffic prediction. Hu [17] proposed a base station network traffic prediction method based on distributed machine learning, which combines federated learning and LSTM with a DeepAR model to capture additional temporal features. However, traffic data from different base stations often follow non-independent and identically distributed (non-IID) patterns, posing significant data heterogeneity challenges to federated learning. Zhang et al. [18] proposed a Dual-Attention Federated Learning (FedDA) for wireless traffic prediction, which uses a dual-attention aggregation algorithm to improve the generalization of the global model. Lin et al. [19,20] proposed a distributed cloud-edge collaborative wireless communication traffic prediction framework based on cloud-edge collaboration, which utilizes a Jensen-Shannon divergence model and cooperative game theory to ensure coalition stability and enhance prediction accuracy. However, FL methods fail to capture individual traffic features, causing low accuracy and high overhead.

2.2 Federated Learning

Federated learning was originally proposed by a research team at Google in 2016 to address the privacy of personal data on the mobile device side. Federated learning belongs to a branch of distributed machine learning, and its core concept is to realize collaborative learning among multiple participants under the premise of safeguarding data security and privacy [21]. Since the original data is always kept locally and is not directly transmitted to other participants or central servers, the risk of data leakage and misuse can be avoided [22].

The classical federated learning framework is shown in Fig. 1. The central server first distributes the global model to participating clients. Each client then trains the model using its local dataset and uploads the updated local model to the server. The server aggregates the received local models to generate an updated global model, completing one communication round. This process iterates repeatedly to produce a well-generalized global model. Importantly, only model parameters are exchanged between the server and clients, effectively protecting data privacy.

Recent studies have further explored federated learning in dynamic edge environments, including resource allocation for federated self-supervised learning in the Internet of Vehicles [23] and large language model-based task offloading in digital twin edge computing networks [24].

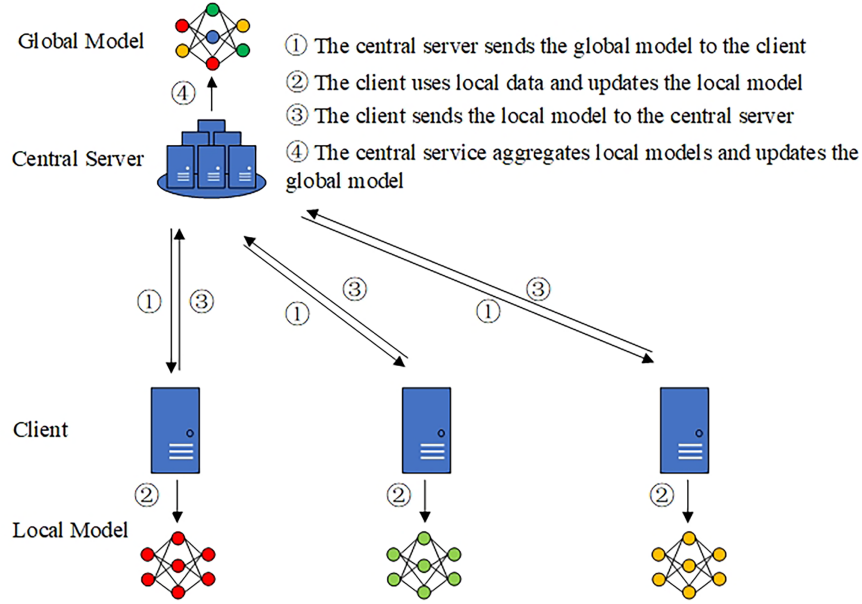


Figure 1: Classical architecture of federated learning.

2.3 Meta-Learning

Meta-learning, also known as “learning to learn,” is a subfield of machine learning. Its primary goal is to enable models to learn new tasks efficiently with limited data and few training steps, addressing the heavy data and time requirements of traditional methods. By training across multiple tasks, meta-learning helps models capture shared knowledge while adapting rapidly to unseen tasks. The most prominent optimization-based approach is Model-Agnostic Meta-Learning (MAML) [25], which uses gradient descent to train a model that can quickly adapt to new tasks with only one or two gradient updates. Specifically, each task is split into a support set and a query set. The model first performs parameter updates on the support set, then optimizes the meta-objective on the query set to improve generalization. This process allows rapid adaptation to new tasks via minimal gradient steps, as illustrated in Fig. 2. Overall, MAML enhances model adaptability and training efficiency for diverse tasks.

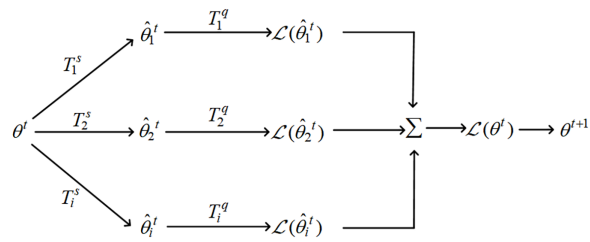


Figure 2: Logic diagram of the MAML meta-learning client algorithm.

2.4 Federated Meta-Learning

Federated meta-learning combines the strengths of federated learning and meta-learning to enable personalized learning and rapid model adaptation in distributed environments. While federated learning allows multiple participants to collaboratively train a model without sharing raw data, meta-learning optimizes the learning process itself by extracting transferable knowledge across tasks [26]. This approach is particularly

effective under federated learning, where data heterogeneity often causes the global model to underperform on specific clients. By treating each client as an independent task, federated meta-learning produces an initialized model that can quickly adapt to local data distributions with only a few gradient descent steps. Chen et al. [27] introduced model-agnostic meta-learning into federated learning and proposed a federated meta-learning algorithm that replaces the global model with a meta-learner. It splits local datasets into support and query sets, trains personalized models on the support set, and optimizes generalization on the query set to obtain a well-initialized shared global model. Fallah et al. [28] proposed a Personalized Federated Averaging (Per-FedAvg), which leverages the core idea of MAML to enable fast adaptation of the shared global model for different clients, thereby accelerating convergence. Overall, the fast adaptation capability of meta-learning effectively addresses system heterogeneity and non-IID data challenges in federated learning, significantly improving model performance and generalization across diverse data distributions.

Although FedMeta and Per-FedAvg have successfully introduced MAML-style bi-level optimization into the federated learning framework, they were primarily designed for image classification or simple regression tasks with homogeneous model architectures. In contrast, MAML-Federated Mutual Information (FedMI) is specifically tailored to the wireless traffic prediction scenario and differs from these methods in three key aspects.

- (1) Task construction. We construct dual-channel network traffic sequences (temporal adjacency features and periodic features) via sliding windows and treat consecutive time windows as distinct meta-tasks, enabling the model to capture the multi-scale temporal dynamics inherent in cellular traffic.
- (2) Personalization strategy. After the global model is obtained, we perform an explicit lightweight fine-tuning step on a held-out adaptation set at each client before final prediction. This post-distribution personalization stage is absent in both FedMeta and Per-FedAvg.
- (3) Aggregation mechanism. Instead of simple averaging or purely gradient-based personalization, the server employs a mutual-information-weighted aggregation that dynamically assigns higher weights to more informative local models, further enhancing generalization under severe non-IID traffic distributions.

These domain-specific designs collectively constitute the incremental contribution of the present work beyond existing federated meta-learning frameworks.

2.5 Long Short-Term Memory

Long Short-Term Memory is a specialized recurrent neural network widely used for sequence modeling. By incorporating memory cells and three gating units, it effectively captures long-term temporal dependencies while mitigating the vanishing gradient problem. This structure enables LSTM to model complex, nonlinear, and non-stationary time series data with strong expressive power. In wireless network traffic prediction, LSTM excels at extracting spatiotemporal features from historical traffic sequences. When combined with federated learning, it serves as an effective base model for distributed training across multiple clients.

3 Method

To address the low prediction accuracy caused by data heterogeneity and the high communication overhead in federated learning, this paper proposes a MAML-based federated meta-learning framework for wireless network traffic prediction. We first introduce the MAML-FedMI framework and its prediction objective. At the client side, the MAML algorithm is employed for model training, performing parameter updates on the support set and meta-optimization on the query set through a two-layer mechanism.

This enables the global model to adapt quickly to new traffic patterns while significantly reducing the number of communication rounds and resource consumption. Before local prediction, the global model is personalized and fine-tuned on local data to improve accuracy. In addition, a meta-learning-friendly preprocessing method is designed, which constructs dual-channel traffic sequences via sliding windows and splits the dataset into multiple tasks, each containing a support set and a query set. Finally, a weighted model aggregation algorithm based on mutual information is presented at the server to further enhance the generalization performance of the global model.

In this work, each base station is treated as an independent client that owns its local traffic data and participates in the federated training process. For clarity, we hereafter refer to these participating nodes uniformly as “clients”.

3.1 MAML-FedMI Wireless Network Traffic Prediction Framework

In this framework, each client utilizes the MAML meta-learning approach to train the model on the local training set. The MAML-FedMI wireless network traffic prediction framework is shown in Fig. 3.

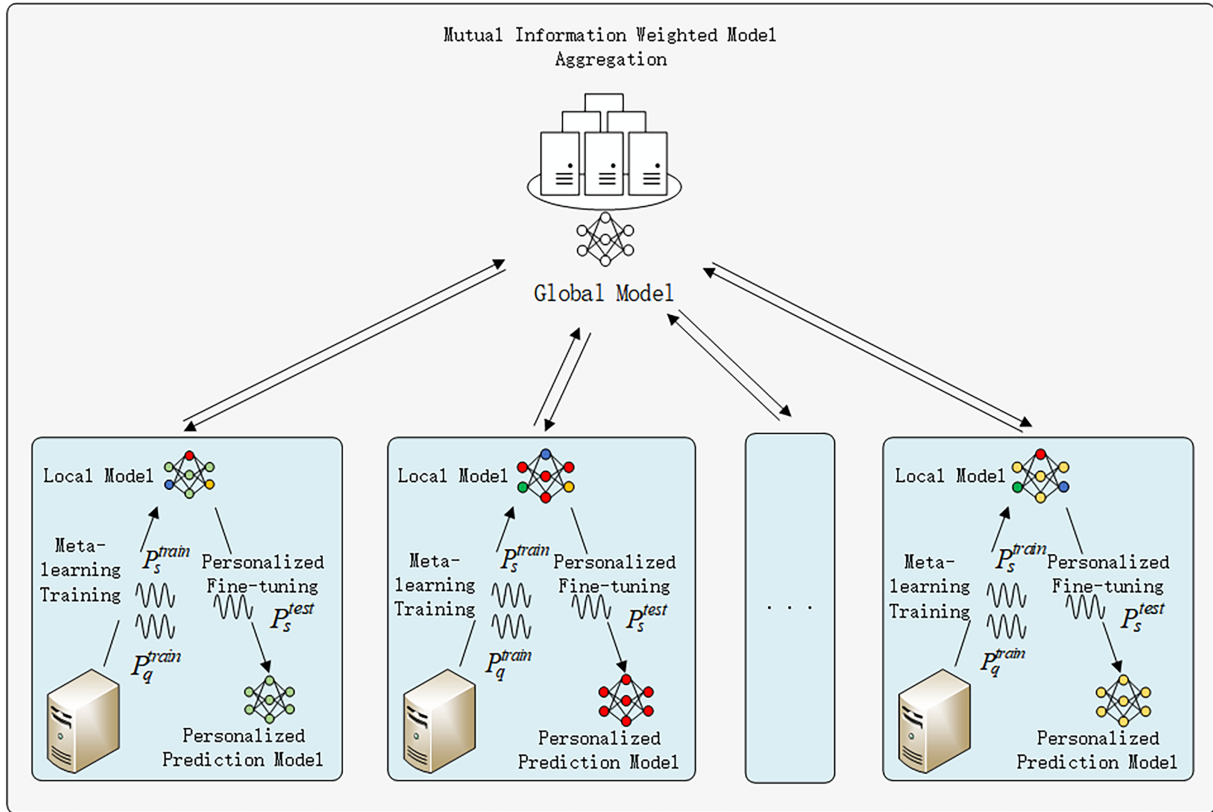


Figure 3: MAML-FedMI wireless network traffic prediction framework.

Firstly, the model update is performed on the support set, and subsequently, the optimization is performed on the query set, which enhances the model's ability to adapt to new tasks through this two-tier updating mechanism, thus obtaining a global model capable of quickly learning new traffic patterns. The central server employs a weighted model aggregation algorithm based on mutual information to further enhance the generalization performance of the model. After multiple rounds of federated learning, a predictive model that incorporates the global knowledge base is generated. When applying this model to the

traffic prediction task of clients, it is first fine-tuned by a local adaptation set to fully adapt to the personalized characteristics of local traffic, and finally obtains a personalized network traffic prediction model for clients.

Typically, the goal of a federated learning network traffic prediction method is to obtain a global model that is used to predict the traffic of K clients and minimize the average error, which can be described as follows

$$\arg \min_{\theta} \frac{1}{K} \sum_{k=1}^K \mathcal{L}(\theta, P_k), \quad (1)$$

where $\mathcal{L}(\theta; P_k)$ denotes the loss value predicted by the global model on the dataset of the first k client. In this paper, we introduce a meta-learning approach into federated learning for wireless network traffic prediction. The goal is to train a global model that can quickly adapt to new traffic prediction tasks. By performing simple fine-tuning of this global model on the client side, a highly accurate personalized prediction model can be obtained. Thus, the objective of the proposed MAML-based federated meta-learning framework can be expressed as follows

$$\arg \min_{\theta} \frac{1}{K} \sum_{k=1}^K \mathcal{L}\left(\theta - \alpha \nabla_{\theta} \sum_{i=1}^I \mathcal{L}(\theta_{k,i}, P_k^s), P_k^q\right), \quad (2)$$

where, P_k^s and P_k^q denote the support set and query set of client k , respectively; $\theta_{k,i}$ represents the model parameters of the i -th meta-task on client k ; α is the inner-loop learning rate; and $\nabla_{\theta} \sum_{i=1}^I \mathcal{L}(\theta_{k,i}, P_k^s)$ is the sum of gradients computed on the I support-set tasks. The term $\theta - \alpha \nabla_{\theta} \sum_{i=1}^I \mathcal{L}(\theta_{k,i}, P_k^s)$ therefore stands for the intermediate parameters obtained after the inner-loop updates. These intermediate parameters are subsequently evaluated on the query set P_k^q to produce the meta-objective, enabling the global model to adapt rapidly to new traffic patterns.

3.2 MAML-FedMI Wireless Network Traffic Prediction Process

The prediction flow of the MAML-FedMI wireless network traffic prediction method is shown in Algorithm 1. At the beginning of each round of federated learning, a group of clients $S = \max(K \cdot \delta, 1)$ is randomly selected as the participating objects of this round of federated learning, where δ is a hyperparameter that limits the proportion of clients selected in each round.

Algorithm 1: MAML-based federated meta-learning framework for network traffic prediction process

Input: federated communication rounds T , support set P_k^s , query set P_k^q , ratio of selected clients δ , learning rates α and β

Output: generalized global model θ_G

```

1:   for iteration rounds  $t = 1, 2, \dots, T$  do
2:       Randomly sample a set of clients  $S_t = \max(K \cdot \delta, 1)$ 
3:       for each client  $k \in S_t$  do
4:           Load the global model:  $\theta_{k,0}^t = \theta_G^t$ 
5:           Sample  $I$  tasks  $T_{k,i}^s$  from the support set  $P_k^s$ 
6:           for each task  $i = 1, 2, \dots, I$  do
7:               Inner-loop update:  $\hat{\theta}_{k,i}^t = \theta_{k,0}^t - \alpha \nabla_{\theta^t} \mathcal{L}(\theta_{k,0}^t; T_{k,i}^s)$ 
8:           end for

```

(Continued)

Algorithm 1 (continued)

```

9:          Sample  $I$  tasks  $T_{k,i}^q$  from the query set  $P_k^q$ 
10:         Outer-loop (meta) update:  $\theta_k^{t+1} = \theta_k^t - \beta \nabla_{\theta^t} \sum_{i=1}^I \mathcal{L}(\hat{\theta}_{k,i}^t; T_{k,i}^q)$ 
11:         end for
12:         Weighted MI aggregation to obtain the global model  $\theta_G^{t+1}$ 
13:         end for
14:         Personalized fine-tuning on the adaptation set of each client (performed only once after all
            federated rounds finish):  $\theta'_k = \theta - \alpha \nabla_{\theta} \mathcal{L}(f_{\theta}, D_k^{adapt})$ 
            (The adaptation set  $D_k^{adapt}$  is completely disjoint from both the meta-training data and the final
            test set; final evaluation is performed exclusively on the unseen test set.)

```

For each client $k \in S_t$, the global model parameters for the current federated learning round are first loaded and a batch of tasks $T_{k,i}^s$ are extracted from the support set P_k^s and the first (inner-loop) update is performed on each task to obtain the intermediate model parameters, which are updated with the following formula. In this work, we adopt the first-order approximation of MAML (FOMAML), which omits the second-order Hessian-vector products to reduce computational and memory overhead.

$$\hat{\theta}_{k,i}^t = \theta_{k,0}^t - \alpha \nabla_{\theta^t} \mathcal{L}(\theta_{k,0}^t, T_{k,i}^s), \quad (3)$$

where $\hat{\theta}_{k,i}^t$ denotes the intermediate parameters after the inner-loop update on the i -th support-set task, $\theta_{k,0}^t$ is the model parameters received from the server at the beginning of the current round, α is the inner-loop learning rate, and $T_{k,i}^s$ is the i -th support-set task of client k . Next, a batch of tasks $T_{k,i}^q$ is extracted from the query set P_k^q . The model is further optimized on these query tasks to improve generalization, yielding the outer-loop (meta) update:

$$\theta_k^{t+1} = \theta_k^t - \beta \nabla_{\theta^t} \sum_{i=1}^I \mathcal{L}(\hat{\theta}_{k,i}^t, T_{k,i}^q), \quad (4)$$

where β is the outer-loop (meta) learning rate, and $\nabla_{\theta^t} \sum_{i=1}^I \mathcal{L}(\hat{\theta}_{k,i}^t, T_{k,i}^q)$ is the sum of the query-set gradients computed with the intermediate parameters $\hat{\theta}_{k,i}^t$. This bi-level update enables the local model to adapt quickly to new traffic patterns.

During federated learning, after each client uploads local model parameters, the central server integrates these parameters using a weighted aggregation algorithm based on mutual information to generate a global model for the current round. This global model has excellent generalization performance and can quickly adapt to the new client data tasks added in the next round of federated learning. With this federated meta-learning approach, the global prediction model is able to achieve highly accurate predictions in fewer communication rounds, thus significantly reducing the communication overhead during federation learning.

After all federated communication rounds are completed, the final global model is personalized and fine-tuned on the held-out adaptation set of each client. This post-training personalization step is performed only once and is completely separated from the federated training loop. The adaptation set is disjoint from both the meta-training data and the final test set, ensuring that the final evaluation remains unbiased. The process of fine-tuning is as follows

$$\theta'_k = \theta - \alpha \nabla_{\theta} \mathcal{L}(f_{\theta}, D_k^{adapt}) \quad (5)$$

where θ is the final global model obtained after federated training, D_k^{adapt} is the held-out adaptation set of clients k , and θ'_k is the resulting personalized model. Personalized fine-tuning is executed only once after the completion of all communication rounds. The final performance evaluation is conducted exclusively on the completely unseen final test set.

The personalized model fine-tuned using the adaptation set can more accurately capture specific patterns of local network traffic, achieve higher prediction accuracy, and effectively solve the data heterogeneity problem of federated learning traffic prediction.

3.3 Data Preprocessing and Local Prediction Modeling

The federated learning network traffic prediction method is based on distributed training of traffic data from multiple clients, assuming that there are K clients participating in federated learning, such that d_n^k represents the network traffic value of the k client in the n time slot, and the network traffic prediction is to predict the traffic value $\{d_{n-1}^k, d_{n-2}^k, \dots, d_{n-m}^k\}$ in the future time based on the network traffic value in the past time.

Therefore, so that the traffic value of the past neighboring m time of the client k in the n time slot is used as the input of the traffic prediction task x_n^k , and d_n^k represents the output of the traffic prediction task y_n^k . In order to better mine the data characteristics of network traffic, the periodic traffic data is added into the training process of the model, $x_n^k = \{d_{n-1}^k, d_{n-2}^k, \dots, d_{n-m}^k; d_{n-q}^k, d_{n-2q}^k, \dots, d_{n-pq}^k\}$, where q is the number of cycles and p is the number of adjacent cycles. This is done by sampling from the dataset using a sliding window as shown in Fig. 4. The neighboring time windows in one cycle to sample the traffic value of the current timestamp n and the value of the previous m time point, and the period sliding window samples the traffic value of the same timestamp n in the neighboring p cycles, thus constructing a dual-channel network traffic sequence dataset.

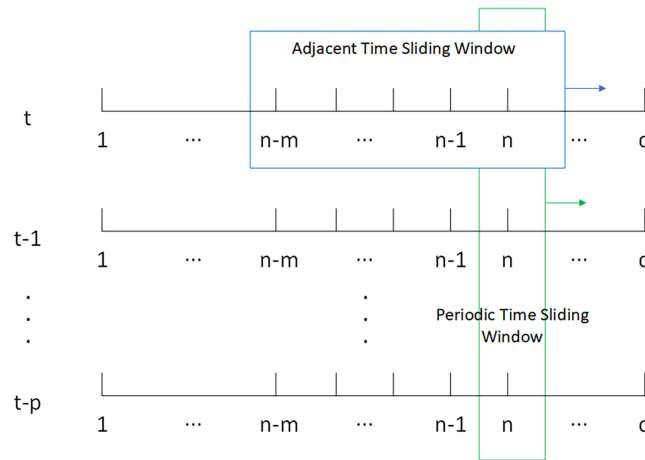


Figure 4: Dual-channel network traffic sequence construction.

During the training process of federated learning, in order to reduce communication resource consumption, it is necessary for federated learning to be able to train a good global model with a relatively small number of communication rounds. Clients can adopt the MAML meta-learning method to enhance the generalization ability of the model, so that the aggregated global model can quickly adapt to new traffic

prediction tasks. Before constructing the meta-tasks, the local data of each client are chronologically split into three disjoint parts: a meta-training set (70%), an adaptation set (10%), and a final test set (20%). The meta-training set is further divided into support and query sets for the bi-level optimization, while the adaptation set is used only for the post-training personalization step described in [Section 3.2](#). The final test set remains completely unseen until evaluation.

Specifically, the meta-training data P_k^{train} of the k -th client are converted into a sequence of meta-tasks via a sliding-window strategy. Given a window size ω and stride s , each window forms the input sequence of one meta-task $T_i = \{S_i, Q_i\}$. For every meta-task, the samples inside the window are randomly partitioned into a support set S_i (60%, corresponding to P_k^s) and a query set Q_i (40%, corresponding to P_k^q). The model first performs the inner-loop gradient update on the support set S_i to obtain the intermediate parameters $\hat{\theta}_k$, and then computes the meta-gradient on the query set Q_i for the outer-loop optimization, yielding a local traffic prediction model with strong generalization ability. Consecutive windows may retain a small overlap (controlled by stride s) to increase the number of tasks while preserving temporal continuity.

3.4 Weighted Model Aggregation Algorithm Based on Mutual Information

In federated learning (FL), the central server's model aggregation algorithm plays a decisive role in the generalization performance of the global model. Its core goal is to produce a model with strong generalization across all clients by balancing personalized and shared features to address data heterogeneity. While Federated Averaging (FedAvg) [29] often suffers from weight dispersion and performs poorly on non-IID data, similarity-based aggregation methods have gained increasing attention [30]. Building on this, we propose a weighted model aggregation based on mutual information. The key idea is to compute mutual information between client models to quantify their correlations and information content, thereby assigning dynamic weights for more accurate aggregation. This approach mitigates the negative impact of data heterogeneity and further enhances the generalization ability of the global model.

Direct estimation of mutual information between two high-dimensional neural network parameter tensors is computationally intractable. Following common practice in federated learning literature and our previous work, we first extract the weight vector of the final linear prediction layer of each local model (which has a much lower dimension and captures the most discriminative traffic patterns). Mutual information between these vectors is then estimated using a k -nearest-neighbor (k -NN) based estimator. The resulting MI values are normalized via the SoftMax function to obtain the aggregation weights.

The specific steps of the algorithm are, when the central server receives the information of the selected K input model parameters in the t round of federated communication, it calculates the mutual information between different input model parameters. And reassign weights for each input model based on the mutual information similarity to quantify the influence coefficients of each input on the aggregated model, for the client k , its weighted model as follows

$$\tilde{\theta}_k^{t+1} = \sum_{r=1}^K \tilde{m}_{k,r} \theta_r^{t+1} \quad (6)$$

where $\tilde{m}_{k,r}$ is the SoftMax version of the mutual information of the models k and r , and $\tilde{\theta}_k^{t+1}$ is the weighted model parameters. The weighted model parameters are then aggregated to obtain the aggregated model.

$$\theta^{t+1} = \frac{1}{K} \sum_{k=1}^K \tilde{\theta}_k^{t+1} \quad (7)$$

The global model obtained by the weighted model aggregation algorithm takes into account the importance of different input models when they are aggregated, so that the aggregated model is not overly biased and its generalization ability is enhanced.

3.5 Computational and Communication Complexity Analysis

While a rigorous convergence analysis of MAML-based bi-level optimization combined with mutual-information-weighted aggregation under non-IID wireless traffic remains an open theoretical question, we provide a quantitative examination of the computational and communication complexity of MAML-FedMI to better understand its practical efficiency.

In our framework, the bi-level optimization is implemented using the FOMAML, which replaces the expensive second-order Hessian-vector products with first-order gradients. Let $|\theta|$ denote the number of model parameters, K the number of clients selected in each round, and R the number of communication rounds.

Computational complexity: Compared with standard FedAvg, the dominant extra cost comes from the inner-loop updates performed on the support set. With the first-order approximation, the computational cost per client is approximately twice that of a conventional local SGD update. This overhead remains acceptable for modern edge devices.

Communication complexity: The communication cost of each round is $O(K \cdot |\theta|)$, identical to FedAvg, FedMI, and Federated Mutual Information Clustering (FedMIC). However, MAML-FedMI reaches the target prediction accuracy in significantly fewer communication rounds than the competing methods. Consequently, the total communication volume, measured by $R \times K \times |\theta|$, is substantially reduced. This reduction in total transmitted parameters constitutes the primary practical advantage of the proposed method in bandwidth-constrained wireless networks.

4 Experiment and Result Analysis

4.1 Experimental Dataset

The Telecom Italia dataset is a widely recognized open dataset for network traffic prediction. It records communication activities from 01 November 2013 to 01 January 2014 in Milan and the province of Trento, Italy, at 10-min intervals. The study area is divided into 100×100 grids, each approximately 235×235 square meters. Multidimensional data, including Internet usage, Call, and SMS, were extracted for each grid. In this paper, we use the data from Trento province, which contains 6259 rasters. Each raster is treated as a client in federated learning. For each experiment, 100 grids are randomly selected using a fixed random seed to ensure reproducibility. We extract 8 consecutive weeks of traffic data. These data are chronologically split into three disjoint parts: a meta-training set (70%, further divided into support and query sets for MAML bi-level optimization), a held-out adaptation set (10%, used exclusively for personalization fine-tuning), and a final test set (20%, completely unseen until evaluation). Missing values are filled with zero. Each selected time series is normalized independently using Z-score standardization, where the mean and standard deviation are computed solely on the meta-training portion.

4.2 Experimental Setup and Baseline Method

Experimental setup. The experiments were conducted in a Python 3.9 environment using the PyTorch deep learning framework to simulate the federated learning setting. The hardware configuration included an Intel Core i7-10700 CPU, an NVIDIA GeForce GTX 1650 GPU, and 32 GB of RAM, which provided sufficient computational resources for efficient training and evaluation.

Federated learning training. For each experiment, traffic data from 100 randomly selected clients were used, covering 8 weeks in total. These data are chronologically split into a meta-training set (70%), a held-out adaptation set (10%), and a final test set (20%), consistent with the protocol defined in Sections 3.3 and 4.1. Two-channel training samples are constructed using a sliding window with both time-channel and period-channel dependency lengths set to 3 and a stride of 1. A total of 100 federated communication rounds are performed, with a client participation ratio of 0.1 and a meta-learning rate of 0.01. At each client, the local LSTM model (with 64 hidden units and two LSTM layers) is trained using stochastic gradient descent with a learning rate of 0.01 and a batch size of 20. All experimental results in Tables 1 and 2 are reported as mean $\pm$ standard deviation over three independent runs with different random seeds.

Table 1: Comparison of predictive performance of different methods in terms of MSE and MAE.

Methodologies	MSE			MAE		
	Internet	Call	SMS	Internet	Call	SMS
SVR	6.451 $\pm$ 0.187	2.478 $\pm$ 0.092	6.192 $\pm$ 0.203	1.289 $\pm$ 0.041	0.982 $\pm$ 0.035	1.527 $\pm$ 0.048
LSTM	4.557 $\pm$ 0.134	1.333 $\pm$ 0.051	4.318 $\pm$ 0.129	1.087 $\pm$ 0.033	0.531 $\pm$ 0.021	1.028 $\pm$ 0.037
FedAvg	4.254 $\pm$ 0.118	1.403 $\pm$ 0.047	2.236 $\pm$ 0.071	0.988 $\pm$ 0.029	0.493 $\pm$ 0.018	0.743 $\pm$ 0.024
FedMI	4.013 $\pm$ 0.109	1.386 $\pm$ 0.044	2.217 $\pm$ 0.068	0.968 $\pm$ 0.027	0.491 $\pm$ 0.017	0.748 $\pm$ 0.023
FedDA	3.149 $\pm$ 0.089	0.926 $\pm$ 0.036	2.189 $\pm$ 0.062	0.829 $\pm$ 0.025	0.438 $\pm$ 0.015	0.748 $\pm$ 0.022
FedMIC	2.542 $\pm$ 0.074	0.877 $\pm$ 0.031	2.092 $\pm$ 0.055	0.752 $\pm$ 0.021	0.418 $\pm$ 0.014	0.736 $\pm$ 0.019
MAML-FedMI	2.230 $\pm$ 0.061	0.839 $\pm$ 0.027	2.046 $\pm$ 0.049	0.711 $\pm$ 0.018	0.402 $\pm$ 0.012	0.734 $\pm$ 0.017

Table 2: Ablation study of different components.

Methodologies	MSE			MAE		
	Internet	Call	SMS	Internet	Call	SMS
FedAvg	4.254 $\pm$ 0.118	1.403 $\pm$ 0.047	2.236 $\pm$ 0.071	0.988 $\pm$ 0.029	0.493 $\pm$ 0.018	0.743 $\pm$ 0.024
FedAvg + FT	3.612 $\pm$ 0.095	1.215 $\pm$ 0.041	2.148 $\pm$ 0.058	0.891 $\pm$ 0.027	0.461 $\pm$ 0.016	0.721 $\pm$ 0.020
FedMI	4.013 $\pm$ 0.109	1.386 $\pm$ 0.044	2.217 $\pm$ 0.068	0.968 $\pm$ 0.027	0.491 $\pm$ 0.017	0.748 $\pm$ 0.023
MAML + FedAvg	2.687 $\pm$ 0.078	0.975 $\pm$ 0.035	2.103 $\pm$ 0.052	0.782 $\pm$ 0.024	0.428 $\pm$ 0.014	0.712 $\pm$ 0.018
MAML-FedMI (w/o dual)	2.514 $\pm$ 0.071	0.912 $\pm$ 0.032	2.087 $\pm$ 0.049	0.751 $\pm$ 0.022	0.419 $\pm$ 0.013	0.701 $\pm$ 0.017
MAML-FedMI (Full)	2.230 $\pm$ 0.061	0.839 $\pm$ 0.027	2.046 $\pm$ 0.049	0.7112 $\pm$ 0.018	0.402 $\pm$ 0.012	0.734 $\pm$ 0.017

To verify its superiority, five methods are selected for comparison against MAML-FedMI. All experimental results in Table 1 are reported as mean $\pm$ standard deviation over three independent runs with different random seeds to ensure statistical reliability.

- (1) SVR [31]. Support Vector Regression (SVR) is a traditional machine learning model based on Support Vector Machines. It can effectively handle high-dimensional data, shows strong robustness to noise and outliers, and addresses nonlinear relationships through kernel functions. SVR has been successfully applied to traffic prediction.
- (2) LSTM [7]. LSTM is a deep learning model commonly used to capture time series features to solve the gradient vanishing and gradient explosion problems of traditional RNNs when dealing with long sequences, and has been widely used in network traffic prediction.
- (3) FedAvg [29]. FedAvg is the classical federated learning method, which implements mean aggregation of model parameters during model aggregation.

- (4) FedMI. FedMI is a federated learning traffic prediction method, which implements weighted model aggregation based on mutual information by considering different aggregation weights given to the importance of the model at the time of model aggregation.
- (5) FedDA [18]. FedDA uses data sharing and K-Means algorithm for clustering clients using two-layer attention mechanism aggregation model, the performance of FedDA transforms with the parameter of proportion of data contribution, and this experiment employs it at 10% data sharing rate setting to compare with the proposed method.
- (6) FedMIC [32]. FedMIC is a network traffic prediction method based on mutual information clustering, where the client uses the mutual information of the models to evaluate the similarity and thus achieve clustering.

To evaluate the impact of different learning paradigms on wireless traffic prediction, two baselines were established: one fully distributed, where each node independently trains SVR and LSTM models using only its local data without any interaction among models; the other is federated learning-based, comprising four variants. Among them, FedAvg and FedMI perform global aggregation at a central server without client grouping, while FedDA and FedMIC employ differentiated clustering strategies for local aggregation. To ensure comparability, all federated methods use the same LSTM network architecture and hyperparameters, eliminating potential interference caused by differences in model structure.

Note that representative federated meta-learning methods such as FedMeta and Per-FedAvg are not included in the current experimental comparison. These methods were originally designed for image classification with homogeneous architectures; directly porting them to dual-channel traffic sequences without extensive task re-engineering may lead to unfair comparisons. A systematic comparison under a unified implementation framework is planned as future work.

4.3 Evaluation Metrics

In wireless traffic prediction tasks, the main issue of prediction performance is the error between the true observed and predicted values, and the prediction performance of these algorithms is usually evaluated using Mean Square Error (*MSE*), Mean Absolute Error (*MAE*), and *R*-squared score.

MSE measures the average of the squares of the differences between the predicted values and the true values. *MSE* is very sensitive to large errors, so it is a good indicator of how well the model performs in predicting extreme errors. The formula is as follows, where y^k is the 7th real observation of k and $\hat{y}^k$ is the corresponding predicted value.

$$MSE = \frac{1}{K} \sum_{k=1}^K (\hat{y}^k - y^k)^2 \quad (8)$$

The *MAE* measures the average absolute difference between the predicted and true values. It directly reflects the true level of prediction error. Unlike *MSE*, *MAE* assigns equal weight to both large and small errors, making it less sensitive to outliers. Its calculation formula is as follows

$$MAE = \frac{1}{K} \sum_{k=1}^K |\hat{y}^k - y^k| \quad (9)$$

The *R*-squared score measures how well the model fits the observed data. It represents the proportion of the total variance in the data that is explained by the model. *R*-squared ranges from 0 to 1, with values closer to 1 indicating better model fit and stronger explanatory power. Its calculation formula is as follows

$$R^2 = 1 - \frac{\sum_{k=1}^K (\hat{y}^k - y^k)^2}{\sum_{k=1}^K (\hat{y}^k - \bar{y})^2} \quad (10)$$

In summary, smaller values of both *MSE* and *MAE* indicate higher prediction accuracy. *R*-squared normalizes model performance to the range [0, 1], with higher values signifying better fit and greater predictive accuracy.

4.4 Prediction Results and Comparative Analysis

Prediction error analysis. the experimental results are shown in [Table 1](#), compared with FedDA, MAML-FedMI has 29.2%, 9.4%, and 6.5% MSE gain and 14.2%, 8.2%, and 1.9% MAE gain for Internet, Call, and SMS data, respectively; compared with FedMIC, MAML-FedMI has 12.3%, 4.4%, and 2.3% MSE gain and 5.2%, 3.9%, and 0.4% MAE gain for Internet, Call, and SMS data have 12.3%, 4.3%, and 2.3% MSE gains and 5.5%, 3.9%, and 0.4% MAE gains, respectively. The global model trained by MAML-FedMI is personalized and fine-tuned before it is used for client prediction, and the resulting personalized prediction model has a high accuracy in predicting local traffic and lower error, thus the proposed method has better prediction results compared to other federated learning methods without personalized fine-tuning. The improvements remain stable across different random seeds, as indicated by the relatively small standard deviations.

Comparison of fully distributed learning strategies. Traditional machine learning models such as SVR struggle to capture the complex temporal dependencies and nonlinear characteristics of traffic data, resulting in significantly lower accuracy than deep learning models like LSTM. Moreover, fully distributed strategies train models independently at each client using only local data, lacking cross-regional knowledge sharing. In contrast, the federated learning framework enables deep fusion of knowledge across regions while preserving data privacy, thereby substantially improving prediction accuracy and robustness.

Comparison of federated learning strategies. Compared with FedAvg, the mutual information-based weighted aggregation in FedMI effectively enhances the generalization of the global model. Building upon this, MAML-FedMI further adopts MAML for local training, achieving bi-level optimization through support and query sets. This approach deeply extracts personalized traffic features and improves global model generalization. Consequently, MAML-FedMI significantly outperforms FedMI, FedDA, and FedMIC. The key advantage is that, unlike other methods that directly apply a shared global model, MAML-FedMI performs local personalized fine-tuning, making the final model better suited to each client's specific traffic patterns and thus achieving lower prediction errors.

Prediction results display analysis. in order to fully verify the superiority of the proposed method, the traffic data of a certain client is randomly selected in the experiment, and the prediction results of different methods on this grid are plotted in comparison, as shown in [Fig. 5](#). The experimental results show that the MAML-based federated meta-learning method exhibits significant prediction advantages, and is significantly better than the FedDA and FedMIC methods in terms of accuracy. Specifically, although FedDA and FedMIC mine the common traffic features of similar clients through clustering strategy, which improves the prediction performance to a certain extent, the federated meta-learning method realizes the personalized fine-tuning of the client traffic by strengthening the client's self-learning capability, so that the model can better adapt to the local traffic features, thus obtaining the prediction results that are more closely related to the real values.

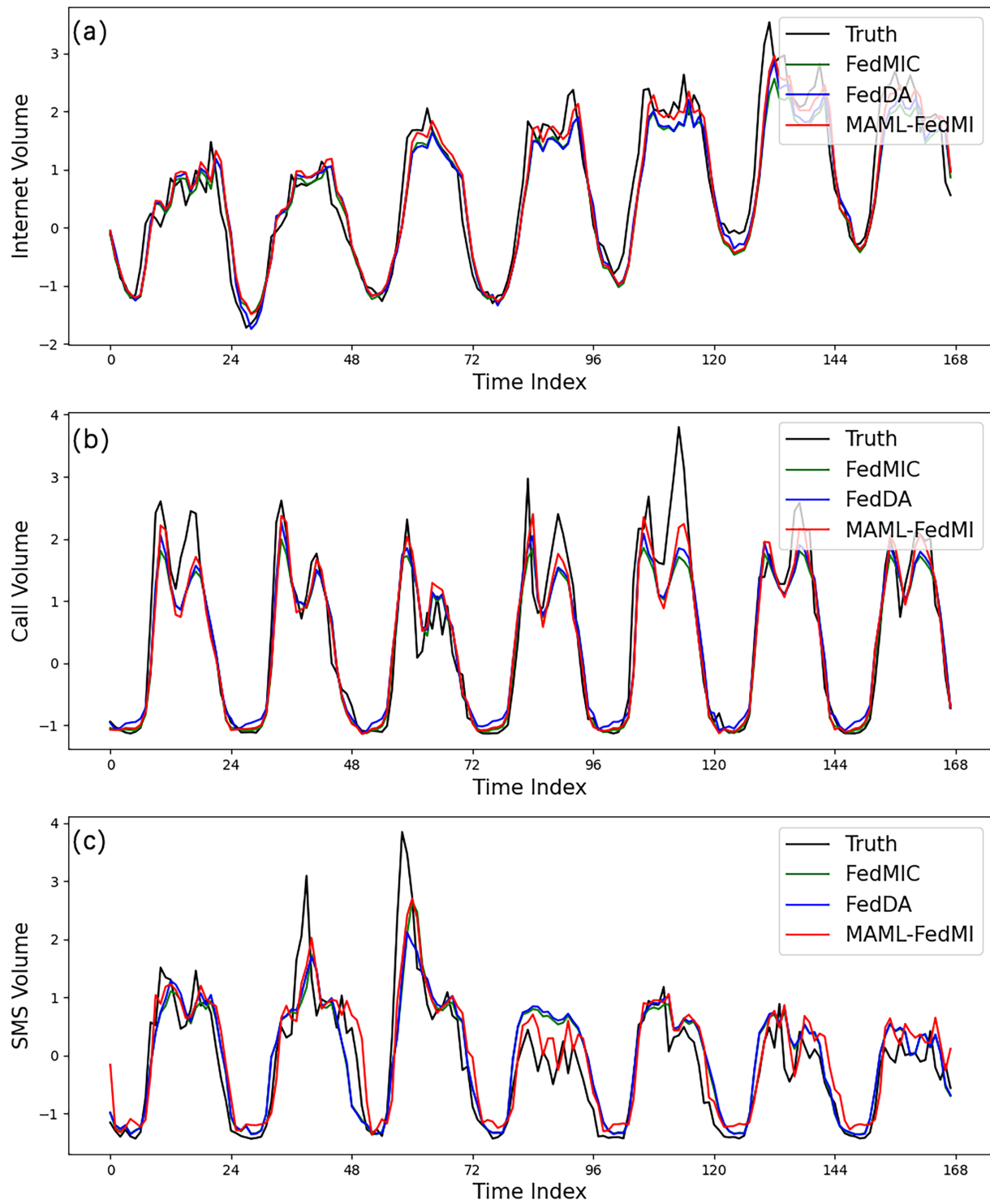


Figure 5: Comparison of predicted and true values of different methods. (a) Internet traffic; (b) call traffic; (c) SMS traffic.

Cumulative distribution of error analysis. Fig. 6 shows the cumulative distribution function of the absolute errors obtained from the prediction of the traffic data of the selected clients. Overall, the prediction

error of the traffic of the 100 clients selected in the experiment is basically below 2.0, which indicates that the proposed network traffic prediction method can predict more accurately in the vast majority of client traffic prediction. On the final test set, the proposed method achieves lower prediction errors for most clients than the other two methods, confirming its superior performance.

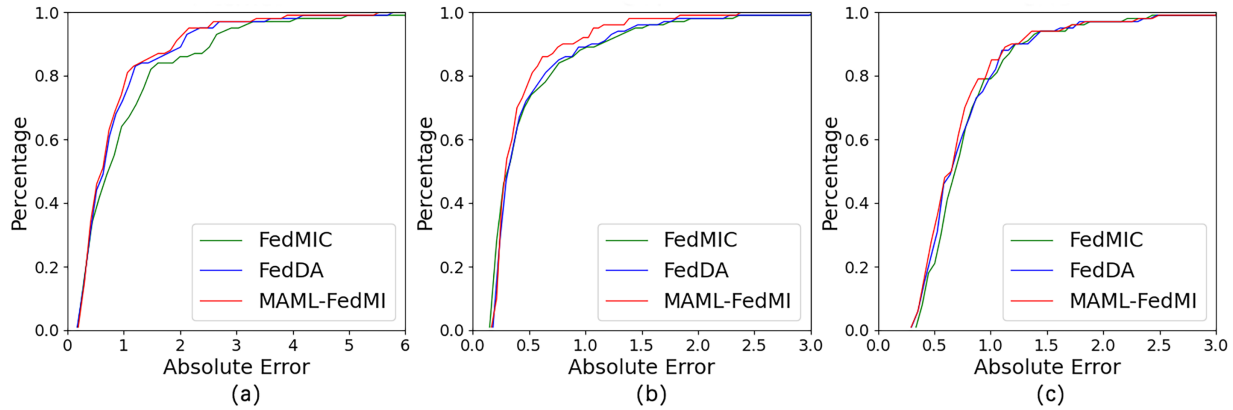


Figure 6: Comparison results of CDF for different methods. (a) Internet traffic; (b) call traffic; (c) SMS traffic.

Fig. 7 illustrates the comparison of the prediction accuracy of different algorithms during the increase of federated communication rounds, where the prediction accuracy is measured by the R-square score. From the figure, it can be observed that as the number of federated communication rounds increments, the clients gradually learn and master the characteristics of the local traffic in depth, while the central server continues to aggregate the models from the clients around the world, resulting in an ever-expanding knowledge base of the global model. For Internet traffic, to achieve more than 70% accuracy, FedDA needs 31 rounds, FedMIC needs 18 rounds, and MAML-FedMI needs only 7 rounds; for Call traffic, to achieve more than 70% accuracy, FedDA needs 58 rounds, FedMIC needs 10 rounds, and MAML-FedMI needs only 4 rounds; for SMS traffic, to achieve more than 65% accuracy, FedDA requires 17 rounds, FedMIC requires 40 rounds and MAML-FedMI requires 14 rounds. It can be seen that the proposed MAML-based federated meta-learning algorithm achieves higher prediction accuracy in federated traffic in fewer traffic rounds, which is due to the fact that the meta-learning method introduced by MAML-FedMI enhances the learning capability of the local model and gives the model a strong generalization performance to adapt to the new traffic patterns quickly. Therefore, if the federated learning traffic prediction method is made to achieve the same accuracy, the proposed federated meta-learning method has the fewest number of federated communication rounds, which can effectively reduce the consumption of communication resources.

The MAML-based federated meta-learning algorithm achieves 73.8%, 78.2%, and 69.7% accuracy for Internet, Call, and SMS traffic within 100 FL rounds, outperforming two other algorithms, which confirms that personalized fine-tuned local models improve prediction accuracy.

The present study focuses on single-step prediction using the Telecom Italia Trento dataset. This choice is motivated by the fact that short-term (single-step) traffic forecasting is the most common and practically urgent requirement for real-time network resource allocation and load balancing. Within this setting, we conducted complete experiments on three traffic types—Internet, Call, and SMS. As reported in Table 1 and Fig. 5, these three types exhibit clearly different prediction errors and require substantially different numbers of communication rounds to reach the same accuracy level. These differences indicate that the proposed method has been evaluated under traffic patterns with distinct statistical behaviors (different degrees of burstiness and temporal dependency), even though a more fine-grained, artificially controlled

heterogeneity setting is not included. Regarding multi-step prediction and systematic evaluation under varying levels of data heterogeneity, we acknowledge that they are not covered in the current experiments.

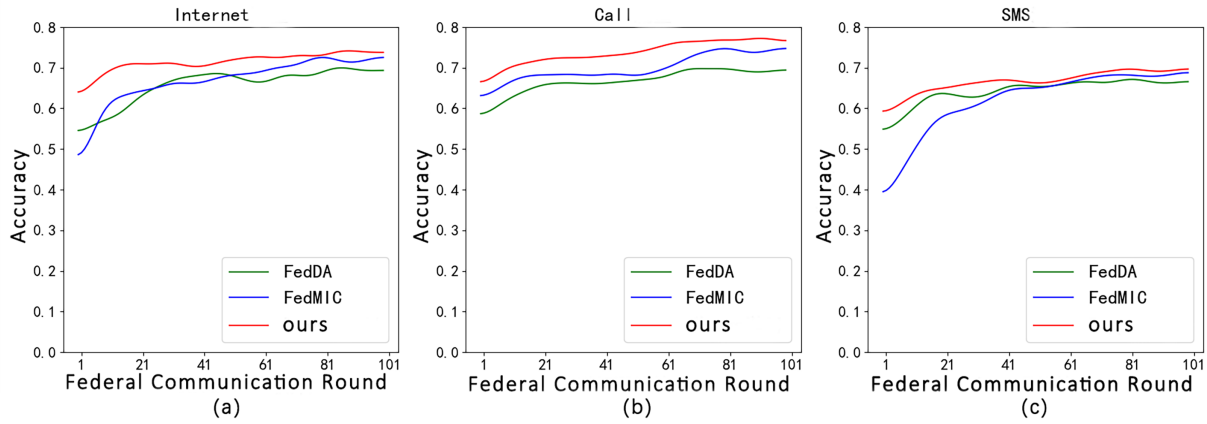


Figure 7: Effect of communication rounds on accuracy rate. (a) Internet traffic; (b) call traffic; (c) SMS traffic.

4.5 Ablation Study

All ablation results are also averaged over three independent runs with different random seeds and reported as mean $\pm$ standard deviation. To further isolate and quantify the contribution of each core component in MAML-FedMI, we conduct a systematic ablation study. Specifically, we examine four key design choices: (1) MAML-based local training, (2) mutual-information-weighted model aggregation, (3) dual-channel temporal preprocessing, and (4) personalized fine-tuning on the held-out adaptation set. By progressively enabling or disabling these modules while keeping all other settings identical, we obtain six variants, as summarized below:

FedAvg: Standard federated averaging with dual-channel input and no personalized fine-tuning.

FedAvg + FT: FedAvg with additional personalized fine-tuning on the adaptation set.

FedMI: Mutual-information-weighted aggregation with dual-channel input, without MAML or fine-tuning.

MAML + FedAvg: MAML-based local training combined with simple averaging and personalized fine-tuning.

MAML-FedMI (w/o dual): Full framework except that the dual-channel construction is replaced by a single temporal channel.

MAML-FedMI (Full): The complete proposed method.

All variants are evaluated under the same data-splitting protocol (meta-training 70%, adaptation 10%, final test 20%) and the same hyper-parameters. Results are averaged over three independent runs with different random seeds and reported as mean $\pm$ standard deviation in [Table 2](#).

From the results, several observations can be drawn. First, the full MAML-FedMI achieves the lowest MSE across all three traffic types and the lowest MAE on Internet and Call traffic, demonstrating clear overall advantages of the joint design. Second, replacing mutual-information-weighted aggregation with simple averaging (MAML + FedAvg) leads to a noticeable performance degradation on most metrics, indicating that the MI-based weighting effectively mitigates the adverse effects of non-IID traffic distributions. Third, removing the dual-channel construction increases MSE on all traffic types. For SMS MAE, the dual-channel variant shows a marginally higher error. This suggests that the periodic features, while beneficial for Internet

and Call traffic, provide limited additional value (or even mild noise) for the more bursty SMS traffic. Fourth, personalized fine-tuning on the held-out adaptation set brings consistent gains, as evidenced by the improvement of FedAvg + FT over plain FedAvg. Overall, although a single metric on SMS exhibits a small deviation, the full model remains the most balanced and robust solution across the three traffic types, and the synergistic combination of MAML-based adaptation and MI-weighted aggregation constitutes the primary source of the observed advantages.

Overall, the ablation study verifies that the key components of MAML-FedMI, especially MAML-based local training and mutual-information aggregation, contribute positively to the final performance on the majority of metrics, and that their integration is essential for achieving accurate and communication-efficient wireless traffic prediction under heterogeneous federated settings.

5 Conclusion

Regarding the challenges of data privacy, heterogeneity, and high communication overhead in wireless network traffic prediction, this paper proposes a MAML-based federated meta-learning framework for wireless network traffic prediction. The core innovation lies in integrating meta-learning and a mutual information-based weighted model aggregation algorithm into the federated learning paradigm. Specifically, the MAML algorithm equips local models with the ability to rapidly adapt to new traffic patterns through a bi-level optimization mechanism on support and query sets. This significantly accelerates convergence and reduces communication costs. Meanwhile, the weighted model aggregation algorithm based on mutual information assigns appropriate weights to different client models according to their importance, thereby enhancing the generalization ability of the global model and improving overall training efficiency. Furthermore, personalized fine-tuning of the global model on adaptation set is performed before deployment, enabling the model to better capture client-specific traffic characteristics and achieve higher prediction accuracy. Compared with existing wireless network traffic prediction methods, the proposed MAML-FedMI approach fully leverages the advantages of meta-learning while effectively improving the performance and efficiency of federated learning through task partitioning and advanced aggregation strategies. This method offers a promising privacy-preserving solution for wireless traffic prediction. While the proposed method achieves strong empirical performance, a rigorous theoretical convergence guarantee of the combined MAML-based bi-level optimization and mutual-information-weighted aggregation under highly non-IID wireless traffic remains an open challenge. A formal theoretical analysis is left as an important direction for future work. Future work will capture data nuances, improve generalization across environments, and ensure robustness in diverse scenarios.

Acknowledgement: The authors would like to thank the experimental team for their strong cooperation and the tutor for their full support in the research process.

Funding Statement: This research is partially supported by the National Natural Science Foundation of China (62473341), Key Research and Development Special Project of Henan Province (252102210166, 252102211086, 252102210060).

Author Contributions: Study conception and design: Xiaotong Li, Zengyu Cai; data collection: Yuan Feng, Liang Zhu; analysis and interpretation of results: Xiaotong Li, Jianwei Zhang; draft manuscript preparation: Xiaotong Li; revision and supplementation of experiments: Xinhua Hu. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the Corresponding Author, Zengyu Cai, upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Letaief KB, Chen W, Shi Y, Zhang J, Zhang YA. The roadmap to 6G: AI empowered wireless networks. *IEEE Commun Mag.* 2019;57(8):84–90. doi:10.1109/MCOM.2019.1900271.
2. Jiang W. Cellular traffic prediction with machine learning: a survey. *Expert Syst Appl.* 2022;201(3):117163. doi:10.1016/j.eswa.2022.117163.
3. Yang Q, Liu Y, Chen T, Tong Y. Federated machine learning: concept and applications. *ACM Trans Intell Syst Technol.* 2019;10(2):12. doi:10.1145/3298981.
4. Yang Y, Geng S, Zhang B, Zhang J, Wang Z, Zhang Y, et al. Long term 5G network traffic forecasting via modeling non-stationarity with deep learning. *Commun Eng.* 2023;2(1):33. doi:10.1038/s44172-023-00081-4.
5. Nie L, Jiang D, Yu S, Song H. Network traffic prediction based on deep belief network in wireless mesh backbone networks. In: *Proceedings of the 2017 IEEE Wireless Communications and Networking Conference (WCNC)*; 2017 Mar 19–22; San Francisco, CA, USA. p. 1–5. doi:10.1109/WCNC.2017.7925498.
6. Yang H, Li X, Qiang W, Zhao Y, Zhang W, Tang C. A network traffic forecasting method based on SA optimized ARIMA-BP neural network. *Comput Netw.* 2021;193(3):108102. doi:10.1016/j.comnet.2021.108102.
7. Fan J, Mu D, Liu Y. Research on network traffic prediction model based on neural network. In: *Proceedings of the 2019 2nd International Conference on Information Systems and Computer Aided Education (ICISCAE)*; 2019 Sep 28–30; Dalian, China. p. 554–7. doi:10.1109/iciscae48440.2019.221694.
8. Ramakrishnan N, Soni T. Network traffic prediction using recurrent neural networks. In: *Proceedings of the 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*; 2018 Dec 17–20; Orlando, FL, USA. p. 187–93. doi:10.1109/ICMLA.2018.00035.
9. Bi J, Yuan H, Xu K, Ma H, Zhou M. Large-scale network traffic prediction with LSTM and temporal convolutional networks. In: *Proceedings of the 2022 International Conference on Robotics and Automation (ICRA)*; 2022 May 23–27; Philadelphia, PA, USA. p. 3865–70. doi:10.1109/ICRA46639.2022.9812427.
10. Zhang C, Zhang H, Yuan D, Zhang M. Citywide cellular traffic prediction based on densely connected convolutional neural networks. *IEEE Commun Lett.* 2018;22(8):1656–9. doi:10.1109/LCOMM.2018.2841832.
11. Zhao L, Song Y, Zhang C, Liu Y, Wang P, Lin T, et al. T-GCN: a temporal graph convolutional network for traffic prediction. *IEEE Trans Intell Transp Syst.* 2020;21(9):3848–58. doi:10.1109/TITS.2019.2935152.
12. Vinchoff C, Chung N, Gordon T, Lyford L, Aibin M. Traffic prediction in optical networks using graph convolutional generative adversarial networks. In: *Proceedings of the 2020 22nd International Conference on Transparent Optical Networks (ICTON)*; 2020 Jul 19–23; Bari, Italy. p. 1–4. doi:10.1109/icton51198.2020.9203477.
13. Zhou X, Zhang Y, Li Z, Wang X, Zhao J, Zhang Z. Large-scale cellular traffic prediction based on graph convolutional networks with transfer learning. *Neural Comput Appl.* 2022;34(7):5549–59. doi:10.1007/s00521-021-06708-x.
14. Geng X, Zhang F, Zhang M, Wang H. Traffic prediction based on spatio-temporal feature embedding fusion and gate operation optimization. *Neurocomputing.* 2025;657:131658. doi:10.1016/j.neucom.2025.131658.
15. Andreoletti D, Troia S, Musumeci F, Giordano S, Maier G, Tornatore M. Network traffic prediction based on diffusion convolutional recurrent neural networks. In: *Proceedings of the IEEE INFOCOM 2019—IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*; 2019 Apr 29–May 2; Paris, France. p. 246–51. doi:10.1109/infcomw.2019.8845132.
16. Xiao F, Zhao SS, Wang SH, Wang RC, Xu SY. Traffic prediction method used in distributed network based on intelligent optimization. *J Beijing Univ Posts Telecommun.* 2015;38(S1):45–8. (In Chinese). doi:10.13190/j.jbupt.2015.s1.011.
17. Hu YD. Research on base station network traffic prediction method based on distributed machine learning [master's thesis]. Chengdu, China: University of Electronic Science and Technology of China; 2022. (In Chinese). doi:10.27005/d.cnki.gdzku.2022.002491.

18. Zhang C, Dang S, Shihada B, Alouini MS. Dual attention-based federated learning for wireless traffic prediction. In: Proceedings of the IEEE INFOCOM 2021—IEEE Conference on Computer Communications; 2021 May 10–13; Vancouver, BC, Canada. p. 1–10. (In Chinese). doi:10.1109/INFOCOM42981.2021.9488883.
19. Lin S, Ma J, Zhuang B, Li Y, Li Z, Li T, et al. Wireless traffic prediction based on federated learning. J Comput Appl. 2023;43(6):1900–9. (In Chinese). doi:10.11772/j.issn.1001-9081.2022050721.
20. Lin SJ, Ma J, Li YY, Zhuang B, Li T, Li ZY, et al. Citywide wireless traffic prediction based on personalized federated learning. J Univ Electron Sci Technol China. 2023;52(1):64–73. (In Chinese). doi:10.12178/1001-0548.2022102.
21. Wen J, Zhang Z, Lan Y, Cui Z, Cai J, Zhang W. A survey on federated learning: challenges and applications. Int J Mach Learn Cybern. 2023;14(2):513–35. doi:10.1007/s13042-022-01647-y.
22. Mothukuri V, Parizi RM, Pouriyeh S, Huang Y, Dehghantanha A, Srivastava G. A survey on security and privacy of federated learning. Future Gener Comput Syst. 2021;115(4):619–40. doi:10.1016/j.future.2020.10.007.
23. Gu X, Wu Q, Fan P, Fan Q, Cheng N, Chen W, et al. DRL-based resource allocation for motion blur resistant federated self-supervised learning in IoV. IEEE Internet Things J. 2025;12(6):7067–85. doi:10.1109/JIOT.2024.3492326.
24. Wu Q, Xie Y, Fan P, Qin D, Wang K, Cheng N, et al. Large language model-based task offloading and resource allocation for digital twin edge computing networks. IEEE Trans Mob Comput. 2026;25(7):10290–301. doi:10.1109/TMC.2026.3664866.
25. Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. In: Proceedings of the 34th International Conference on Machine Learning (ICML); 2017 Aug 6–11; Sydney, Australia. p. 1126–35.
26. Liu X, Deng Y, Nallanathan A, Bennis M. Federated learning and meta learning: approaches, applications, and directions. IEEE Commun Surv Tutor. 2024;26(1):571–618. doi:10.1109/COMST.2023.3330910.
27. Chen F, Luo M, Dong Z, Li Z, He X. Federated meta-learning with fast convergence and efficient communication. arXiv:1802.07876. 2018.
28. Fallah A, Mokhtari A, Ozdaglar A. Personalized federated learning: a meta-learning approach. arXiv:2002.07948. 2020.
29. McMahan B, Moore E, Ramage D, Hampson S, Arcas BA. Communication-efficient learning of deep networks from decentralized data. In: Proceedings of the 20th International Conference on Artificial Intelligence and Statistics; 2017 Apr 20–22; Fort Lauderdale, FL, USA. p. 1273–82.
30. Li L, Zhao Y, Wang J, Zhang C. Wireless traffic prediction based on a gradient similarity federated aggregation algorithm. Appl Sci. 2023;13(6):4036. doi:10.3390/app13064036.
31. Feng H, Shu Y, Wang S, Ma M. SVM-based models for predicting WLAN traffic. In: Proceedings of the 2006 IEEE International Conference on Communications; 2006 Jun 11–15; Istanbul, Turkey. p. 597–602. doi:10.1109/ICC.2006.254860.
32. Zhang J, Hu X, Cai Z, Zhu L, Feng Y. Federated learning based on mutual information clustering for wireless traffic prediction. Electronics. 2023;12(21):4476. doi:10.3390/electronics12214476.