Proceedings of the 4th International Conference on Mechatronics and Smart Systems
DOI: 10.54254/2755-2721/2026.33429

ST-AFL: Spatio-Temporal Adaptive Federated Learning for Traffic Flow Prediction

Xia Li

School of Computer and Artificial Intelligence, Beijing Wuzi University, Beijing, China
lx15079709917giselle@outlook.com

Abstract. Accurate traffic flow prediction is critical for Intelligent Transportation Systems (ITS), while centralized modeling faces privacy risks and data silos. Federated learning provides a privacy-preserving solution yet suffers from Non-IID data, spatio-temporal feature loss, and high communication overhead in real-world traffic scenarios. To address these issues, this paper proposes a Spatio-Temporal Adaptive Federated Learning (ST-AFL) framework. First, a spatio-temporal clustering strategy based on K-means groups clients with similar traffic patterns to alleviate data heterogeneity. Second, a lightweight Local ST-Model with Conv1d-GRU-FC structure is deployed on edge clients to capture spatial dependencies and temporal dynamics. Third, an adaptive layer-freezing and gradient sparsification mechanism is designed to reduce communication cost by uploading only unfrozen layer parameters. Comprehensive experiments on the Metr-LA dataset show that ST-AFL achieves a final MSE of 0.1491, significantly outperforming FedAvg (0.8860), while reducing communication overhead. ST-AFL effectively balances prediction accuracy, communication efficiency, and privacy protection, making it suitable for edge computing-based traffic prediction.

Keywords: Traffic Flow Prediction, Federated Learning, Spatio-Temporal Adaptive Learning

1. Introduction

Predicting the flow of traffic accurately is one of the major pillars of ITS that assist in effective traffic control, reduction of congestion, and facilitation of smart city functioning [1]. The mass of real-time traffic data is produced by distributed sensors and base stations due to the connection of vehicles and the use of edge computing. Nevertheless, traditional centralized machine learning poses significant privacy threats and data silos [2]. Federated Learning (FL) can train models in cooperation without exchanging raw data and hence balances privacy and model performance [1,3].

Although there has been development, there are still severe challenges in using FL to predict traffic. Older publications use vanilla FedAvg to train models with simple privacy preservation [1]. Later works aim at alleviating the phenomenon of data heterogeneity through clustering and personalized learning or lowering the communication costs by optimizing parameters and selective uploading [4,5]. Nevertheless, most methods do not take into consideration both the issues of data heterogeneity and communication efficiency. The straightforward averaging mechanism is incapable

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.33429

of maintaining intrinsic spatio-temporal correlations and thus cannot provide a sufficient amount of features [6,7]. In addition, the available literature primarily measures Mean Absolute Error (MAE) and Mean Squared Error (MSE) but does not consider deploying-related measurements like communication cost and model lightweightness [5,8].

The proposed ST-AFL framework in this paper can predict the traffic flow efficiently and with high precision. The Non-IID data and the preservation of regional spatiotemporal properties are ensured by a spatio-temporal clustering-based personalized mechanism [6,9]. Communication overhead is lowered through an adaptive layer-freezing and gradient sparsification approach that ensures stability [4,5]. Edge devices have been presented with a lightweight spatio-temporal model with a trade-off between accuracy and efficiency [10,11]. As experiments show, ST-AFL is more precise and efficient in communicating with other nodes than the state of the art baselines do when converging steadily in the presence of heterogeneity [3,4].

2. Method

2.1. Dataset preparation

The paper adopts the Metr-LA dataset [12], a real-life traffic speed dataset in HDF5 format. This includes 34,272 time-stamped records of 207 loop detector data on Los Angeles highways at a 5-minute interval. The objective of this is the mean traffic speed and the task is cast as a regression question to extrapolate the future based on past observations. Such a dataset is commonly employed in spatio-temporal prediction because it is authentic and representative.

Preprocessing is used to make sure that the data quality is maintained: missing values are replaced by the overall mean, outliers are trimmed by three standard deviations and StandardScaler is used to standardize features. To keep both computation and representativeness in balance, the dataset is reduced to 5,000 time steps and 20 road nodes as edge clients. The samples use 12 successive historical steps to predict the average speed of the next 3 steps with a sliding window, which is similar to real short-term traffic predictions requirements.

2.2. Spatio-Temporal Adaptive Federated Learning (ST-AFL)

The ST-AFL is suggested to solve three fundamental issues of federated traffic prediction which are Non-IID data distribution, spatio-temporal feature loss, and excessive communication overhead [1,2]. Its client-server architecture preserves privacy because it shares only model updates rather than raw data, which is required by ITS privacy [4,5]. The complete structure is depicted into Figure 1 where three main parts are described as follows:

To begin, a space-time clustering approach can be used to group clients having the same traffic behavior [6,9]. Time-averaged speed, peak time location and spatial correlation are the three features extracted per client (road node). Using these characteristics, K-means clusters 20 clients into 4 groups which decreases Non-IID drift and maintains regional spatio-temporal correlations that result in constant aggregation [1,6]. The Client Layer in Figure 1 shows how clients can be partitioned into four clusters with the same spatio-temporal characteristics.

Next, a light weight spatio-temporal model is constructed to be used on resource-limited edge devices [3,10]. The model comprises of a 1D convolutional layer that can be used to obtain local spatial relationships, a GRU layer that can be used to obtain long-range temporal relationships, and a fully connected layer that can be used to create speed estimates. This architecture is represented by

DOI: 10.54254/2755-2721/2026.33429

the Local ST-Model (Conv1d -> GRU -> FC) in Figure 1 as it is an efficient and accurate solution that can be deployed at the edge and provides essential spatio-temporal characteristics [7,11].

The third point here is that an adaptive layer-freezing and gradient sparsification approach decreases the communication load [4,5,8]. Combined with the Local ST-Model in the Client Layer, every client computes layer-wise gradient magnitudes in local training. Frozen layers are those with gradient less than the threshold, and only unfrozen layer parameters are transmitted. This sparse upload as indicated by the arrow in Figure 1 minimizes duplicate transmission. The server does clustering-based aggregation: it starts by averaging intra-cluster parameters followed by fusion of cluster-level outputs to update the global model to the local spatio-temporal properties [1,2,13]. Lastly, the latest global model is sent to all the clients to be used in the next round which constitutes the entire federated process.

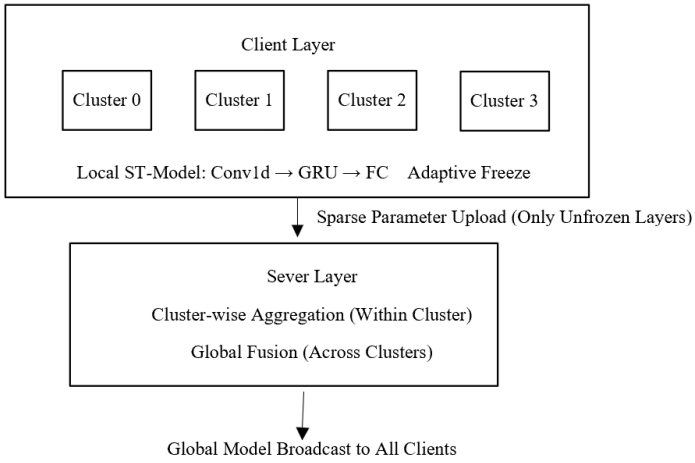


Figure 1. Framework of the proposed Spatio-Temporal Adaptive Federated Learning (ST-AFL) (picture credit: original)

2.3. Experimental hyperparameter configuration

All of the experiments are done on PyTorch 2.10.2 using the same hyperparameters to make the comparison of ST-AFL and FedAvg fair. Important parameters: sequence length=12, prediction length=3, communication rounds=20, local epochs=5, learning rate=0.001, clusters=4, freeze threshold=0.005. Adam optimizer with weight decay 1e-4 is employed to avoid overfitting [1,3]. Loss function is MSE [1,3]. Measure metrics are MAE, MSE (accuracy), and communication cost (KB, efficiency) [3,4,8].

3. Result and discussion

3.1. Experimental results

Comparative tests between ST-AFL and the classical FedAvg are performed under the same setup to test the effectiveness of ST-AFL. The results are presented in Figure 2, namely (left) training loss (MSE per round) and (right) communication cost (KB per round) which indicates that there are differences in the levels of accuracy and efficiency.

DOI: 10.54254/2755-2721/2026.33429

3.1.1. Prediction accuracy

Figure 2 (left) shows that ST-AFL is much faster in convergence and more accurate in final accuracy than FedAvg. Starting at 0.9944 MSE, ST-AFL rapidly drops to 0.4791 (Rd 5), 0.2187 (Rd 10), and converges to 0.1491 (Rd 20). Conversely, FedAvg has a serious problem with Non-IID drift, beginning at 1.1959 and gradually decreasing to 1.1213 (Rd 5), 1.0388 (Rd 10), and 0.8860 (Rd 20). The end MSE of FedAvg is 5.94-fold of ST-AFL which confirms that clustering and cluster-wise aggregation can help alleviate the effects of Non-IID and a small model can be used to reliably capture spatio-temporal correlations [1,6,11].

3.1.2. Communication overhead optimization

The figure on the right in Figure 2 illustrates that FedAvg requires full-parameter uploading and has a constant 16.02 KB cost per round, which leads to significant redundancy in edge devices. Conversely, ST-AFL adopts adaptive layer-freezing and sparse uploading to lower the overhead dynamically. ST-AFL stays at the same 16.02 KB in the initial 5 rounds, but it slowly decreases when the layers converge, down to 15.92 KB (Round 10), 15.79 KB (Round 19) and 15.19 KB (Round 20). The mean cost of 20 rounds is 15.83 KB which is 1.18 percent less compared to FedAvg. Such a decrease is done without loss of quality and thus ST-AFL can be used in edge computing conditions [3,4,8].

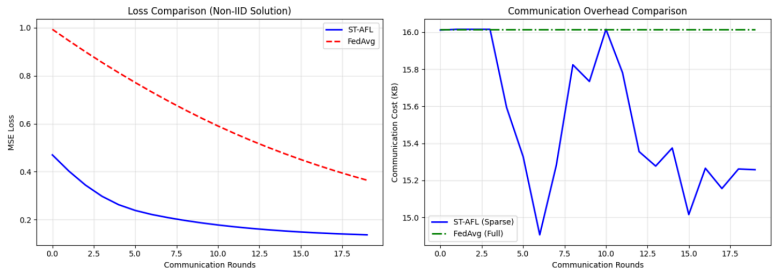


Figure 2. Experimental results: (left) loss comparison (right) communication overhead comparison (picture credit: original)

3.2. Discussion

The findings of the experiment prove that ST-AFL solves the three main issues of federated traffic prediction according to the design. Spatio-temporal clustering minimizes intra-cluster heterogeneity and cluster-wise aggregation alleviates model drift, which means that accuracy is significantly improved [6,9,13]. Adaptive layer-freezing is a useful tool to filter redundant updates and minimize the communication cost [3,4,8]. The Conv1d-GRU model can also be used to correctly represent traffic data in space and time [7,10,11].

The present study has three constraints: fixed K-means clusters cannot respond to changing traffic conditions; communication reduction may be optimized even better; the experiment is restricted to 20 clients and 5,000 steps, and it would need to be validated on a larger scale.

The future work will investigate the concept of online dynamic clustering to observe change in patterns in real time, to optimize the sparsification and threshold that can reduce communication even further, and to test scalability with larger datasets and more customers [3,9].

DOI: 10.54254/2755-2721/2026.33429

4. Conclusion

According to the introduction, the prediction of traffic flow has privacy threats and data silos and the implementation of federated learning suffers from the problem of Non-IID data, spatial-temporal feature loss and significant communication overhead. The paper suggests using ST-AFL as a solution that combines spatio-temporal clustering, a simple Conv1d-GRU-FC model, and adaptive layer-freezing to provide privacy-preserving, precise and communication-efficient traffic prediction. Experimental results on Metr-LA demonstrate that ST-AFL is significantly more effective in terms of prediction accuracy and lower communication overhead. The model balances well between model performance, efficiency, and privacy in edge-based intelligent transportation systems. The drawbacks are fixed clustering and slight reduction of communication. Dynamic clustering, better sparsification, and large-scale validation will be developed in future work to improve adaptability and scalability.

References

[1] Alsehaimi, B., Alzamzami, O., Alowidi, N., & Ali, M. (2026). Adaptive spatio-temporal federated learning to predict traffic flows: The framework and the aggregation methods assessment. Applied Sciences, 16(5), 2402.
[2] Wang, X., Chen, J., Chen, F., Jin, B., Zou, F., Liao, L., & Zeng, R. (2026). FedCTPM: A federated cross-regional collaboration learning model of traffic flow prediction. Expert Systems with Applications, 305, 130779.
[3] Zhao, P., Liao, Z., Zhao, Y., Xu, J., & Yi, A. (2025). A semi-asynchronous federated learning approach combining personalization and staleness awareness to predict traffic flow in dynamic Internet of Vehicles. The Journal of Supercomputing, 81(8), 1018.
[4] Li, K., Xu, Q., Wang, D., Nie, X., & Wang, H. (2026). A communication-efficient federated learning method for traffic flow prediction. Engineering Applications of Artificial Intelligence, 170, 114182.
[5] Zhu, K., Lu, M., Li, H., Xiong, N. N., & He, W. (2025). CMBA-FL: Communication-mitigated and blockchain-assisted federated learning on predicting traffic flow. Digital Communications and Networks, 11(3), 724–733.
[6] Feng, J., Du, C., & Mu, Q. (2024). Traffic flow prediction based on federated learning and spatio-temporal graph neural networks. ISPRS International Journal of Geo-Information, 13(6), 210.
[7] Zhang, Z., Zhao, Y., & Wang, J. (2025). Multi-scale temporal feature-enhanced federated learning structure in network traffic prediction. Computer Networks, 273, 111742.
[8] Gao, F., Zhang, C., Qiao, J., Li, K., & Cao, Y. (2024). Communication-efficient wireless traffic prediction with federated learning. Mathematics, 12(16), 2539.
[9] Que, C., & Khan, F. N. (2025). A scalable federated learning-based solution to accurately predict traffic in edge computing-enabled metro optical networks. Computers & Industrial Engineering, 203, 111004.
[10] Dai, G., Tang, J., Zeng, J., Hu, C., & Zhao, C. (2024). Prediction of traffic flow in road networks: Personalized federated learning approach with client reputation. Computers and Electrical Engineering, 120, 109678.
[11] Kaur, G., Grewal, S. K., & Jain, A. (2024). Federated learning based spatio-temporal framework of real-time traffic prediction. Wireless Personal Communications, 136(2), 849–865.
[12] He, H. (2025). California traffic network datasets: METR-LA, PEMS-BAY, PEMS04 and PEMS08 to analyze traffic speed and flow. IEEE Dataport.
[13] Zhang, J., Hu, X., Cai, Z., Zhu, L., & Feng, Y. (2023). Federated learning using mutual information clustering to predict wireless traffic. Electronics, 12(21).