Information Theory-Based Feature Extraction for Transportation Mode Detection using Federated Ensemble Learning

Edvar M. S. Neto[1], David H. S. Lima [1,2], Douglas L. L. Moura[1,3], Andre L. L. Aquino[1]

[1]Orion Lab. - Computer Institute – Federal University of Alagoas
Maceió – AL – Brazil

[2]Federal Institute of Alagoas – Rio Largo - AL – Brazil

[3]Department of Computer Science - Federal University of Minas Gerais
Belo Horizonte - MG - Brazil

{edvar.neto, alla, douglas.moura}@orion.ufal.br, david.lima@ifal.edu.br

Abstract. *Transportation mode detection is a key enabler of intelligent mobility systems, supporting real-time traffic management, personalized travel services, and urban planning. However, achieving high accuracy while balancing computational efficiency and user privacy remains a critical challenge, known as the efficiency-privacy-accuracy trilemma. This work introduces a novel information-theoretic approach integrating entropy and complexity metrics into centralized and federated learning architectures. Our approach addresses performance issues in traditional federated learning by leveraging the structural properties and patterns within time-series data. The results demonstrate that the proposed approach outperforms conventional methods in MLP, Federated MLP, and federated ensemble learning, achieving accuracy improvements of 3.8%, 4.31%, and 5.54%, respectively. These results enable practical applications such as privacy-preserving transportation services and data-efficient urban planning.*

1. Introduction

The rapid development of communication technologies, sensor technologies, and the Internet of Things (IoT) has significantly impacted Intelligent Transportation Systems (ITS), enabling data-driven approaches to enhance efficiency, safety, and privacy protection [Zhang et al. 2024]. In this context, Transport Mode Detection (TMD) plays a key role in this transformation by utilizing smartphone sensors, such as accelerometers, gyroscopes, and GPS, to capture rich movement data and predict a person's mode of travel, such as walking, cycling, driving, or using public transportation. TMD provides valuable insights into human mobility patterns and enables context-aware services, improving traffic management, device profiling, route optimization, targeted advertising, healthcare, and traveling support [Qureshi and Abdullah 2013].

Traditional solutions process sensor data on remote servers to classify transportation modes. Although smartphones serve as a rich source of movement data for TMD, the sensitive nature of these data raises privacy concerns, as they can reveal personal information, habits, and routines. To mitigate these concerns, the researchers implement privacy-preserving techniques, such as data anonymization, encryption, and local processing. One of the most effective approaches is Federated Learning (FL) [Yurdem et al. 2024], which enables decentralized model training directly on users' devices, eliminating the need to transmit raw sensor data to central servers. This method enhances privacy by keeping data local while allowing collaborative learning across multiple devices. FL is particularly suitable for the TMD problem for three main reasons: (i) it preserves user privacy by keeping sensitive location and movement data on devices; (ii) it minimizes bandwidth consumption by transmitting only model parameters instead of raw sensor data; and (iii) it allows local personalization, adapting to each user's particularities without compromising the global model quality. However, FL's effectiveness critically depends on the quality of locally extracted features. In this context, our information theory-based approach complements FL by providing more discriminative and compact representations of sensor data that preserve information integrity even in distributed learning environments with data heterogeneity.

Despite its proven benefits, FL often suffers from accuracy loss due to various factors, such as data and device heterogeneity, sensor noise, and synchronization issues. Although several studies in the literature address this problem in the context of TMD, the most effective solution relies on only statistical features from sensor data without fully exploring the complexity and information structure embedded in movement patterns [Alam et al. 2023b]. Given these limitations, the central research question addressed by this work is:

> *"How to extract discriminative features from sensor data to maximize the accuracy for TMD without exposing the data to third parties?"*

Previous works have explored Information Theory concepts to enhance feature extraction from vehicular sensor data, with results showing a significant increase in accuracy compared to traditional methods [Santos et al. 2024]. Consequently, information theory-based metrics present a promising alternative for generating more discriminative features for TMD and capturing the structural properties and patterns within time-series data.

This work explores the application of Information Theory-based metrics for improving TMD. By combining the FL technique with Information Theory metrics, we propose a novel approach that enhances feature extraction using **Permutation Entropy** (H) [Bandt and Pompe 2002] and **Statistical Complexity** (C) [Rosso et al. 2007] to improve the accuracy of transportation mode detection. H quantifies the randomness in sensor data by analyzing the relative ordering of values, while C captures structural richness by integrating entropy with disequilibrium measures. This combination enables extracting more discriminative features while preserving privacy within the FL framework. Using a real-world dataset, we evaluated three learning methods (Multilayer Perceptron - MLP, federated MLP, and ensemble FL). The experimental results showed that the proposed solution improves accuracy by up to 4.2% compared to the traditional solution.

The remainder of this paper is structured as follows: Section 2 provides a comprehensive review of the evolution of TMD, highlighting key advancements and existing challenges in the field. Section 3 introduces our information theory-based approach, detailing the theoretical foundation of information-theoretic metrics, their integration into learning architectures Section 4 presents and analyzes the results obtained from the three training methodologies, offering insights into the performance improvements achieved.

Finally, Section 5 concludes the paper by summarizing our contributions, discussing limitations, and outlining potential directions for future research.

2. Related Work

TMD has become essential for ITS due to its ability to provide context information, enabling the delivery of appropriate services based on user needs. With advances in sensor technology, mobile phones have become central in capturing vast amounts of context information, making machine learning essential to analyze these data and enhancing transportation mode detection. Consequently, various innovative approaches have emerged, with researchers exploring different machine-learning techniques to improve accuracy, efficiency, and privacy.

[Xiao et al. 2017] demonstrated that ensemble methods could achieve impressive results using only GPS trajectory data, with their XGBoost implementation reaching 90.77% accuracy on the GEOLIFE dataset. While effective for accuracy, these approaches required raw data transmission to central servers, creating significant privacy vulnerabilities. Building on this foundation, [Alam et al. 2023a] enhanced detection robustness through sophisticated feature engineering and convolutional neural networks that could handle sensor noise and inconsistent sampling frequencies. Their feature cloning and fusion techniques improved classification performance substantially but introduced computational complexity that limited practical implementation on resource-constrained devices.

As privacy concerns grew and the need for real-time processing increased, researchers began exploring edge computing architectures. For example, [Yan and Qin 2020] demonstrated that processing traffic data locally at intersections rather than transmitting it to centralized servers could improve transportation efficiency by approximately 20% and urban security by 35%. These findings validated edge computing's potential, though they primarily addressed system-level traffic management rather than individual mobility patterns. The transition to edge computing marked an essential step toward privacy-preserving TMD but left unresolved challenges regarding computational efficiency on heterogeneous devices.

The emergence of FL offered a promising solution to the privacy-efficiency dilemma by enabling model training without centralizing sensitive data. [Zhang et al. 2021] demonstrated FL's capabilities in vehicular networks, showing a 75% acceleration in model training speed and a 25% reduction in communication overhead compared to centralized methods. Expanding this paradigm, [Elbir et al. 2022] proposed hybrid federated-centralized frameworks that balance computational demands and communication costs across heterogeneous networks. However, their work didn't specifically address the computational optimization needed for resource-limited edge devices in transportation contexts.

[Alam et al. 2023b] employs an ensemble technique with majority voting within a federated framework, where two machine learning models, Random Forest, and XGBoost, are combined with an MLP. In this approach, whenever the MLP makes an error, it is corrected by Random Forest and XGBoost through majority voting. Extensive testing on the TMD dataset demonstrates that our FedEL model outperforms most centralized and decentralized models. While effectively preserving privacy, this approach loses accuracy in larger windows, with increased temporal context and a higher probability of transitions between different modes.

Unlike previous work, we leverage Permutation Entropy and Statistical Complexity to optimize the data representation. When combined with FL, this information-theoretic approach enhances privacy by minimizing data exposure, improves communication efficiency through reduced transmission, and boosts detection accuracy by utilizing enriched features.

3. Information Theory Approach

The solution presented in this work aims to enhance TMD by integrating concepts from information theory into feature extraction from smartphone sensor data using FEL architecture. The following sections describe the FEL architecture and the proposed method for feature extraction in detail.

3.1. Federated Ensemble Learning (FEL)

The FEL approach adopted in this work follows the model established by [Alam et al. 2023b]. They collect the data from smartphones deployed in various modes of transportation, including buses, trains, cars, and pedestrians. Roadside units (RSUs) initially serve as the primary data aggregation point, capturing information from these mobile devices. They transmit the collected data to Multi-access Edge Computing (MEC) servers, which train local models at the network edge. Instead of sharing raw data, MEC servers transmit only the locally trained model parameters to the central cloud, aggregating them to update a global model. This approach reduces communication overhead, enhances privacy, and enables near-real-time processing in vehicular networks.

Figure 1 illustrates the workflow of FEL for TMD. Initially, Each client device trains a local model on its own data at the MEC servers, ensuring that raw data remains on the client and only model parameters are shared. After that, to enhance performance compared to traditional FL, the locally trained models are aggregated using a federated ensemble method, which combines multiple models (such as MLP, Random Forest - RF, and XGBoost - XGB) to improve robustness and accuracy. The neural network updates both local and global model weights, where it computes the global model weights according to the FedAvg technique

$$w_{p+1} = \sum_{l=1}^{L} \frac{s_l \, w_{p+1}^{l}}{s}, \tag{1}$$

where $p+1$ represents the updated global model weight parameter, L is the total number of participants, s_l is the number of samples from participant l, s denotes the total number of samples from all participants, and w_{p+1}^{l} is the local model weight parameter of the participant l in iteration $p+1$. Finally, majority voting determines the final prediction by selecting the class with the highest votes among the three models.

RF and XGF are not neural networks, so they have no assigned weights. Instead, they fit data locally for each client and transfer their current decision trees to the global model. This hybrid ensembling combines a neural network with two non-neural models by aggregating weights and decision trees. This approach enhances privacy, reduces

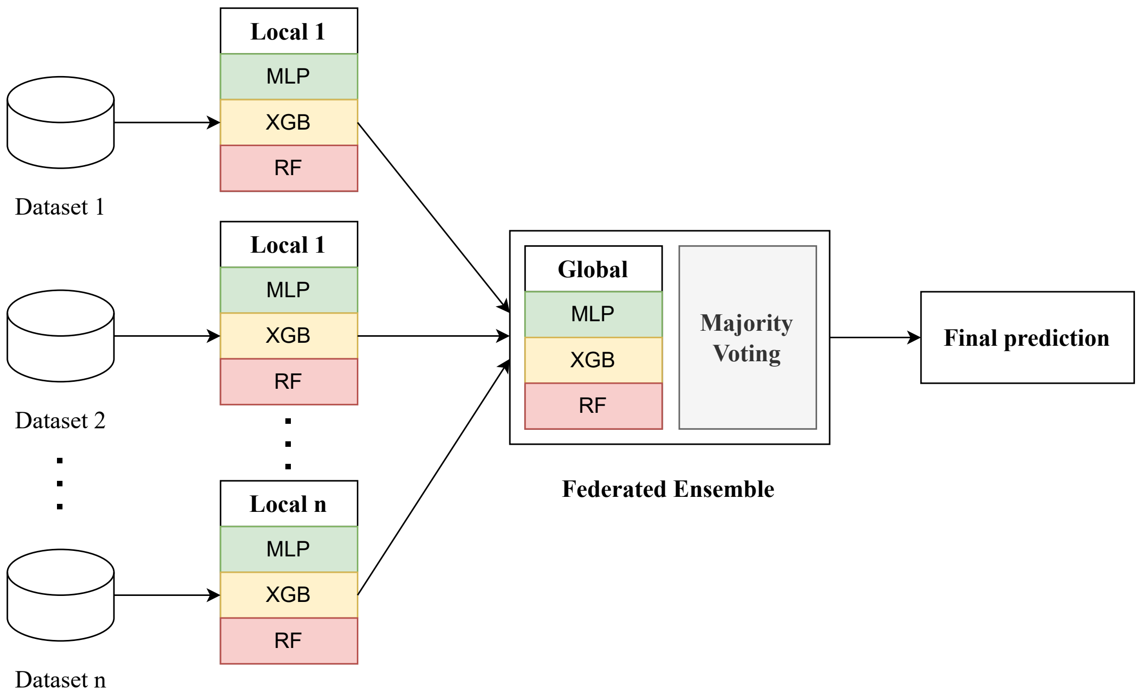

Figure 1. Federated-ensemble learning approach for TMD.

latency, and optimizes computational resources. By leveraging edge computing for local data processing and storage, the architecture supports efficient and privacy-preserving federated learning in vehicular networks.

3.2. Information Theory-Based Feature Extraction

We adopt information theory metrics to provide a more comprehensive representation of the structure inherent in the time-series data from sensors. Figure 2 illustrates extracting features from time series data.

First, the dataset is divided into fixed-size windows, with the window size parameter determining the length of each segment(Figure 2(a)). Within each window, we compare the values to identify their relative ordinal ranking (Figure 2(b)), generating a permutation pattern π_i (Figure 2(c)). The **embedding dimension** (d_x) parameter controls the number of consecutive data points considered to generate each pattern (in this case, $d_x = 4$). The frequency distribution of these patterns across the entire time series provides the probability values $p(\pi_i)$ used in the entropy calculation (Figure 2(d)). Finally, we compute two key information-theoretic features, H and C, defined as follows:

In our approach, the **Permutation Entropy** (H) is a measure of the randomness in time series data, which quantifies the signal's complexity based on the data's ordinal patterns rather than the absolute values. Formally,

$$H(d) = -\sum_{i=1}^{d!} p(\pi_i) \log p(\pi_i), \tag{2}$$

where $p(\pi_i)$ represents the probability of observing the ordinal pattern π_i in a time series of embedding dimension d. We normalize this value by dividing it by $\log(d!)$.

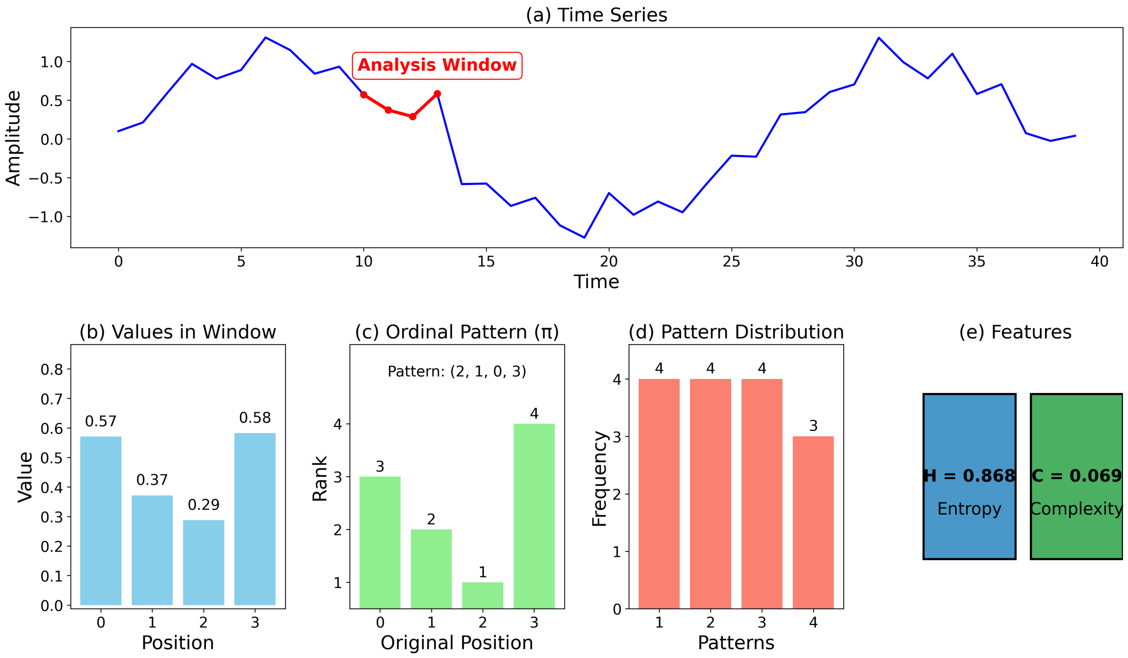

Figure 2. Computation of permutation entropy and statistical complexity from time series data.

Complementary, **Statistical Complexity** (C) quantifies the organizational structure within time series data by combining entropy with a measure of disequilibrium. C provides additional insights into the complexity of a system, complementing H. Formally

$$C_{JS}[P] = Q_{JS}[P, P_e] \cdot H_{norm}[P], \tag{3}$$

where $H_{norm}[P]$ is the normalized permutation entropy and $Q_{JS}[P, P_e]$ is the Jensen-Shannon divergence between the observed distribution P and the uniform distribution P_e.

From each sensor, we extract six features: mean, standard deviation, minimum, maximum, entropy, and complexity. These features are crucial for capturing the essential statistical characteristics and the more complex dynamics of the sensor data. Together, they comprehensively represent the sensor's behavior over time.

4. Evaluations and Discussion

We performed the experiments on a computer Intel Core i5-10300H processor (2.50GHz) and 16 GB of RAM. We use a real-world dataset, referred to as the TDM dataset [Carpineti et al. 2018], comprising sensor readings from 14 distinct smartphone sensors, including accelerometer, gyroscope, magnetic field, and rotation vectors, collected during various transportation activities across multiple subjects. The dataset includes five transportation modes: Car (20.03%), Bus (20.00%), Still (20.00%), Train (20.00%), and Walking (19.97%), representing common daily mobility patterns. To prepare the data, we segmented the time series into 10-second windows and extracted features following the method described in Section 3.2, resulting in a data set of dimension 3.250×101.

Our system model incorporates a local data preprocessing stage, where five edge devices (i.e., smartphones) compute both traditional statistical features and information-theoretic features. Each client trains a local model on this enhanced feature set without sharing raw data. We transmit only model parameters to a central server and aggregate them into a global model before redistributing them to edge devices for deployment. We evaluated our approach using three distinct learning paradigms [Alam et al. 2023b]: **Traditional MLP** with a neural network with four layers (input, two hidden, output), where each hidden layer has 256 neurons, ReLU activation functions, and dropout regularization (rate = 0.40). This scheme follows a centralized approach and does not employ federated learning. **Federated MLP** follows the same neural network architecture as the traditional MLP but is trained using the FedAvg technique. **Federated Ensemble (FedEL)** [Alam et al. 2023b] combining federated versions of XGBoost, Random Forest, and MLP models through majority voting.

We applied different data partitioning strategies based on the learning approach. For the Traditional MLP, we divided the preprocessed dataset into three sets: 64% training, 16% validation, and 20% testing, stratified to maintain class distribution. For the federated learning approaches, we split the dataset into 80% training and 20% testing sets, then distributed the training data across five simulated clients with randomly shuffled and sharded data to simulate realistic federated environments. Table 1 lists the simulation parameters analyzed in this evaluation based on existing literature. We compared our approach with the one proposed by [Alam et al. 2023b], which used only traditional statistical features. The evaluation metrics used in this work were accuracy, precision, recall, and F1-score.

Table 1. Simulation parameters.

Parameter	Value
Number of layers	2
Neurons per layer	256
Activation Function	ReLU
Regularization	Dropout (rate = 0.40)
Batch Size	$\{8, 16, 32, \ldots, 256\}$
Epochs	$\{60, 100, 200\}$
Window size	10 sec
Embedding dimension (d_x)	$\{3, 4, 5\}$

Different transportation modes exhibit distinctive information-theoretic signatures due to their inherent motion patterns. Due to its rhythmic nature, walking produces regular acceleration patterns with low entropy but moderate complexity. In contrast, motorized transport like buses exhibits higher entropy from variable road conditions and high complexity, reflecting structured stop-and-go patterns. Train travel often shows moderate entropy with lower complexity due to constrained movement along tracks. These distinctive signatures provide valuable discriminative power beyond what traditional statistical features alone can capture.

Figure 3 presents a comparison between the traditional statistical approach and the information theory approach, both using 10-second windows. These results demonstrate that the addition of information-theoretic features substantially improves model performance across all architectures. The impact is particularly pronounced in the Federated Ensemble, where including information-theoretic features improves accuracy by 5.54% compared to the baseline. With extensive hyperparameter optimization, traditional MLP with traditional statistical features could reach 95.08% accuracy, suggesting that conventional statistical features can achieve comparable performance but require significantly more tuning effort.

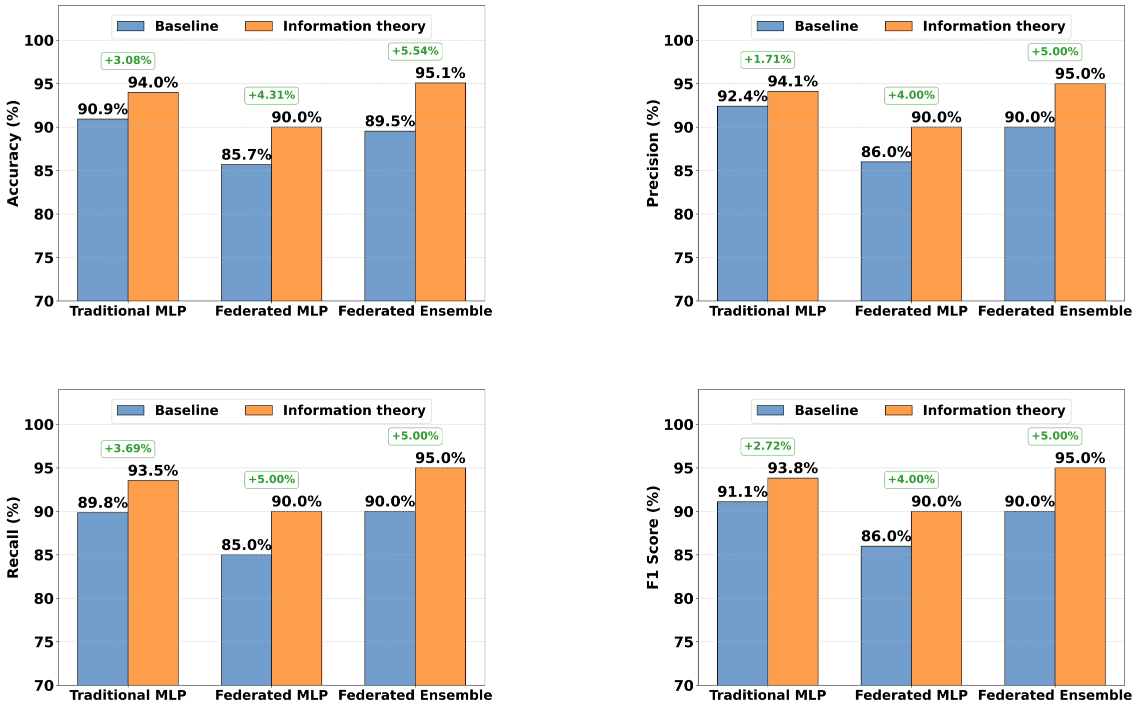

Figure 3. Comparison of model performance with and without information-theoretic features.

Our detailed parameter analysis reveals that the optimal embedding dimension (d_x) varies systematically across architectures. Traditional MLP benefits most from $d_x = 5$ configuration, suggesting it can effectively leverage longer-range patterns in sensor data when all data is centralized. In contrast, Federated MLP performs optimally with $d_x = 4$, while Federated Ensemble achieves peak performance with $d_x = 3$. This architectural sensitivity to embedding dimension likely stems from the fundamental differences in how each learning paradigm processes information patterns. In centralized learning, longer embedding dimensions can capture subtle temporal dependencies that might be fragmented in distributed settings, while ensemble methods can effectively combine simpler patterns extracted from distributed data.

We also observed that models incorporating information-theoretic features consistently perform better with smaller batch sizes (32-64) than baseline models (128). This observation suggests that entropy and complexity metrics introduce higher-variance gradients that benefit from more frequent updates, an essential consideration for hyperparameter optimization when integrating information-theoretic features. As shown in Figure 3, the most substantial improvement occurs in the Federated Ensemble architecture, demonstrating a significant performance increase of 5.54% with the incorporation of information-theoretic features.

Our results demonstrate that by leveraging information-theoretic features, we can effectively extract discriminative features from sensor data to maximize accuracy in TMD while preserving privacy. Specifically, we achieved up to a 5.54% accuracy improvement in federated settings, with Federated MLP reaching 90% accuracy and Federated Ensemble achieving 95.1%. These findings highlight that our approach successfully balances the efficiency-privacy-accuracy trilemma, offering a solution that enhances detection accuracy without exposing data to third parties, especially in privacy-preserving distributed learning contexts.

5. Conclusion

This research addresses the efficiency-privacy-accuracy trilemma in transportation mode detection within the context of evolving intelligent mobility systems. We demonstrated significant improvements across multiple learning architectures by introducing an information-theoretic approach leveraging entropy and complexity metrics. The federated paradigm ensures that raw sensor data remains on users' devices, addressing critical location privacy concerns while reducing bandwidth consumption. Our results show that the proposed approach outperforms traditional methods, achieving accuracy improvements of 3.8% in MLP, 4.31% in Federated MLP, and 5.54% in Federated Ensemble Learning.

Transportation service providers could implement this approach to offer improved mode detection without requiring continuous high-frequency data transmission, while urban planners could gather mobility patterns with stronger privacy guarantees. While our results are promising, several technical limitations exist. The information-theoretic features introduce computational complexity during feature extraction, potentially offsetting efficiency gains in resource-constrained environments. Our synchronous federated implementation also introduced training delays due to heterogeneous data distribution across clients.

Future research should investigate the impact of window size on transportation mode detection performance, examining the trade-off between longer windows (providing richer context) and the increased probability of capturing mode transitions within a single window. Studies should also incorporate additional performance metrics beyond accuracy, particularly bandwidth utilization and energy efficiency on mobile devices. Further directions include exploring asynchronous federated architectures to reduce communication delays, developing adaptive model compression techniques for intermittent connectivity scenarios, and evaluating the system's resilience against privacy attacks.

Acknowledgments

This study was partly financed by the Research Foundation of the State of Alagoas (FAPEAL) under grants E:60030.0000000352/2021 and the National Council for Scientific and Technological Development (CNPq) under grant 407515/2022-4.

References

Alam, M. G. R., Haque, M., Hassan, M. R., Huda, S., Hassan, M. M., Strickland, F. L., and AlQahtani, S. A. (2023a). Feature cloning and feature fusion based transportation mode detection using convolutional neural network. *IEEE Transactions on Intelligent Transportation Systems*, 24(4):4671–4681.

Alam, M. M., Ahmed, T., Hossain, M., Emo, M. H., Bidhan, M. K. I., Reza, M. T., Alam, M. G. R., Hassan, M. M., Pupo, F., and Fortino, G. (2023b). Federated ensemble-learning for transport mode detection in vehicular edge network. *Future Generation Computer Systems*, 149:89–104.

Bandt, C. and Pompe, B. (2002). Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.*, 88:174102.

Carpineti, C., Lomonaco, V., Bedogni, L., Felice, M. D., and Bononi, L. (2018). Custom dual transportation mode detection by smartphone devices exploiting sensor diversity. In *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pages 367–372.

Elbir, A. M., Soner, B., Çöleri, S., Gündüz, D., and Bennis, M. (2022). Federated learning in vehicular networks. In *2022 IEEE International Mediterranean Conference on Communications and Networking (MeditCom)*, pages 72–77.

Qureshi, K. N. and Abdullah, A. H. (2013). A survey on intelligent transportation systems. *Middle-East Journal of Scientific Research*, 15(5):629–642.

Rosso, O. A., Larrondo, H. A., Martin, M. T., Plastino, A., and Fuentes, M. A. (2007). Distinguishing noise from chaos. *Phys. Rev. Lett.*, 99:154102.

Santos, M., Santos, G., and Aquino, A. (2024). Identificação do comportamento de motoristas: Uma abordagem baseada em teoria da informação. In *Anais do XVI Simpósio Brasileiro de Computação Ubíqua e Pervasiva*, pages 31–40, Porto Alegre, RS, Brasil. SBC.

Xiao, Z., Wang, Y., Fu, K., and Wu, F. (2017). Identifying different transportation modes from trajectory data using tree-based ensemble classifiers. *ISPRS International Journal of Geo-Information*, 6(2).

Yan, G. and Qin, Q. (2020). Retracted: The application of edge computing technology in the collaborative optimization of intelligent transportation system based on information physical fusion. *IEEE Access*, 8:153264–153272.

Yurdem, B., Kuzlu, M., Gullu, M. K., Catak, F. O., and Tabassum, M. (2024). Federated learning: Overview, strategies, applications, tools and future directions. *Heliyon*, 10(19):e38137.

Zhang, H., Bosch, J., and Olsson, H. H. (2021). End-to-end federated learning for autonomous driving vehicles. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8.

Zhang, S., Li, J., Shi, L., Ding, M., Nguyen, D. C., Tan, W., Weng, J., and Han, Z. (2024). Federated learning in intelligent transportation systems: Recent applications and open problems. *IEEE Transactions on Intelligent Transportation Systems*, 25(5):3259–3285.