TELKOMNIKA Telecommunication, Computing, Electronics and Control
Vol. 24, No. 4, August 2026, pp. 1168∼1176
ISSN: 1693-6930, DOI: 10.12928/TELKOMNIKA.v24i4.27790

Behavioral fingerprints: driver profiling using transformer models on next generation simulation trajectory data

Mohamed Laamimach[1], Mghari Mohammed[2], Aziz Mabrouk[1]
[1]Information Security Intelligent Systems and Applications (ISISA), Faculty of Sciences, Abdelmalek Essaadi University, Tetouan, Morocco
[2]Department of Computer Science, Faculty of Sciences, Abdelmalek Essaadi University, Tetouan, Morocco

Article Info

Article history:

Received Jan 9, 2026
Revised Apr 5, 2026
Accepted May 25, 2026

Keywords:

Driver behavior
Driver profiling
Intelligent transportation systems
Microscopic traffic analysis
Next generation simulation data
Transformer models

ABSTRACT

Characterizing individual driver behavior is essential for advancing intelligent transportation systems (ITS) and autonomous vehicle safety. While deep learning models excel at macroscopic traffic prediction, individual driving styles are often aggregated away. This paper addresses this gap by proposing a novel, weakly supervised transformer framework for driver behavior profiling using high-resolution next generation simulation (NGSIM) US-101 trajectory data. We extract microscopic behavioral features including acceleration, lane change dynamics, and headway management from 30-second observation segments. A transformer encoder learns complex temporal dependencies to classify drivers into 'aggressive' and 'normal' profiles, achieving a 97% F1-score on proxy-labeled segments. Crucially, these "proxy labels" are derived from heuristic statistics, meaning the model is trained to learn the mapping from sequences to these behavioral indicators rather than identifying objective aggression. Our methodology enables the creation of precise "behavioral fingerprints" that capture individual driving nuances. These insights are vital for developing adaptive ITS that anticipate traffic stability issues and enhance autonomous vehicle safety by predicting human intent.

Corresponding Author:

Mohamed Laamimach
Information Security Intelligent Systems and Applications (ISISA), Faculty of Sciences
Abdelmalek Essaadi University
Tetouan, Morocco
Email: mohamed.laamimach@etu.uae.ac.ma

1. INTRODUCTION

Metropolitan traffic congestion continues to impose escalating economic and environmental burdens on urban populations worldwide [1]. Precise forecasting of traffic dynamics serves as a fundamental enabler for contemporary intelligent transportation systems (ITS), supporting proactive interventions including dynamic route optimization, ramp metering control, and adaptive signal coordination [2]. The efficacy of these technological solutions depend critically on characterizing the diverse behavioral patterns of individual drivers who collectively generate observable traffic phenomena. This study is among the first to apply weakly supervised transformers to next generation simulation (NGSIM) trajectory data for driver profiling, bridging the gap between macroscopic flow and microscopic behavior.

The evolution of deep learning architectures, especially recurrent structures such as long short-term memory (LSTM) networks [3] and transformers [4], have greatly improved the accuracy of traffic forecasting.

While these models excel at predicting macroscopic states, applying them to profile individual driver behavior presents a significant challenge: the scarcity of ground truth data. Unlike macroscopic flow (which can be measured by sensors), specific driving styles like ”aggressive” or ”normal” are subjective and difficult to annotate at scale. Research has shown that heuristic rules often fail to capture the temporal nuances of driving; for instance, rule-based systems can have misclassification rates exceeding 15% when ignoring context [5].

Consequently, research often relies on heuristic rules derived from traffic flow theory to categorize drivers. However, reliance on static rules ignores the temporal context of driving: a sudden brake in a traffic jam is normal, whereas the same action on a free-flow highway is aggressive. Static rules often fail to capture this distinction. Current macroscopic models aggregate individual actions, ignoring the microscopic signatures that define traffic stability. To address this, we adopt a weak supervision strategy: we generate “proxy labels” using established heuristics (e.g., high acceleration variance). We then task deep learning models with learning the mapping from raw trajectory sequences to these labels.

In this paper, we aim to bridge the gap between static rules and dynamic profiling. Building on our previous work [6], we propose a weakly supervised framework using a transformer-based model on the NGSIM US-101 dataset. While the labels are derived from rules, training a transformer on them allows us to investigate whether deep sequence models can effectively encode the temporal precursors that correlate with these rules. This aligns with the scope of microscopic traffic modeling and ITS safety applications relevant to the computing and control domain. Our contributions are: (i) we formulate a weak supervision pipeline that utilizes heuristic traffic rules to generate proxy labels, enabling the training of deep models without manual annotation; (ii) we design a transformer encoder architecture to capture long-range temporal dependencies in 30-second trajectory segments; and (iii) we conduct a comparative analysis against LSTM baseline and non-temporal baselines, demonstrating that the transformer architecture is significantly more effective at approximating these behavioral rules from raw kinematic data. The remainder of this paper is organized as follows: section 2 reviews related work, section 3 details our methodology, section 4 presents the experimental results, and section 5 concludes with future work.

2. RELATED WORK

Our research is situated at the intersection of three primary domains: the application of transformer models to traffic forecasting, the integration of multi-source data, and the use of microscopic behavioral modeling. For this paper, we specifically focus on leveraging microscopic behavioral modeling within a transformer framework for driver profiling.

2.1. Transformer models for traffic forecasting and sequence analysis

The application of deep learning to traffic forecasting has evolved significantly from classical time-series models like autoregressive integrated moving average (ARIMA) [7] to architectures capable of capturing complex non-linear dependencies. While early successes were achieved with convolutional neural networks (CNNs) for spatial features [8] and LSTMs for temporal patterns [9], the transformer architecture [4] has recently emerged as the state-of-the-art due to its superior ability to model long-range dependencies via self-attention.

The transformer paradigm has been rapidly adopted in traffic prediction. Xu *et al.* [10] proposed spatial-temporal transformer networks (STTN), and Jiang *et al.* [11] introduced PDFormer. Other sophisticated architectures have further pushed the boundaries by integrating self-attention with graph-based methods, such as a transformer-based dynamic multi-graph convolutional network (TDMGCN) [12] and a transformer-based spatiotemporal graph diffusion convolution network (TSGDC) [13]. While these models showcase the power of fusing spatial and temporal data within a transformer framework, they typically operate on macroscopic traffic features or aggregate individual vehicle data for system-level predictions. More recently, transformer-based models incorporating spatial masks derived from road network topology, such as Trafficformer [14], have improved short-term speed forecasting from loop-detector data, while multi-source frameworks such as multi-source data traffic prediction (MDTP) [15] and hierarchical multi-source network (HiMSNet) [16] fuse trajectory or drone-based observations with fixed-sensor data to improve prediction robustness, extending a long-standing tradition of sensor data fusion in ITS [17]. Generative approaches based on diffusion models, such as traffic scenario generation diffusion (TSGDiff) [18], and generic interpretable architectures for multi-horizon forecasting, such as the temporal fusion transformer (TFT) [19], further illustrate the breadth of attention-based methods being explored for traffic-related sequence modeling. Our current work distinguishes itself by applying transformers directly to individual, microscopic behavioral sequences for the purpose of the driver themselves, rather than predicting traffic state. Even so, existing models often struggle with the scarcity of individual-level ground truth, which we address via weak supervision.

2.2. Microscopic behavioral modeling and driver profiling

The causal link between microscopic driver behavior and macroscopic traffic phenomena is a foundational concept in traffic engineering, established primarily through simulation [5], [20]. Simulators like simulation of urban mobility (SUMO) [21] and CityFlowER [22] are widely used to model how individual vehicle decisions aggregate into observable outcomes.

Driver profiling, or the characterization of individual driving styles, has gained importance for applications ranging from personalized insurance to autonomous vehicle interaction. Traditional approaches often rely on statistical analysis of driving data or simpler machine learning models to identify patterns in speed, acceleration, and braking [23]. More recent work has explored deep learning for driver identification or style classification, often using aggregated telematics data. Even so, using high-resolution trajectory data like NGSIM for fine-grained behavioral profiling with advanced sequence models like transformers is still an evolving area. Our work aims to extract richer, dynamic ”behavioral fingerprints” by leveraging the detailed temporal evolution of microscopic features.

The NGSIM program represents a pinnacle of high-resolution, multi-vehicle trajectory data collection. It provides an unparalleled microscopic view of traffic, enabling the detailed analysis of individual driver actions, inter-vehicle dynamics, and their evolution over time. This richness goes beyond what can be captured by traditional loop detectors or probe vehicles. Our research leverages this detailed data to create the fine-grained feature sets required for robust driver profiling, moving beyond aggregate metrics to the underlying behavioral signals.

3. METHODOLOGY

Our methodology is designed to process raw NGSIM trajectory data into a supervised learning problem for driver profiling. The process consists of four main stages: (i) data selection and preprocessing; (ii) feature engineering from individual driver trajectories; (iii) heuristic-based labeling of driver profiles; and (iv) the architectural specification of our transformer-based profiling model.

3.1. NGSIM dataset

Developed by the Federal Highway Administration, the NGSIM program provides essential high-resolution trajectory data that enables microscopic traffic analysis. For this investigation, we employ the US-101 (Lankershim Boulevard) dataset, which documents real-world vehicle trajectories at 0.1-second (10 Hz) temporal granularity. It captures precise details such as vehicle locations, speeds, accelerations, and inter-vehicle relationships for individual vehicles operating on an urban arterial environment with multiple lanes and intersections. The raw dataset initially comprises approximately 11.8 million rows and 25 columns of trajectory data. The dataset is characterized by:

- High resolution: vehicle locations are recorded every one-tenth of a second (10 Hz), allowing for precise capture of instantaneous speeds, accelerations, and subtle changes in driving behavior.
- Microscopic detail: each row of the dataset corresponds to a single state of the vehicle at a specific ‘Frame_ID‘ (timestamp), including its unique ‘Vehicle_ID‘. Key columns include ‘Global_Time‘ (absolute timestamp), ‘Local_X‘ and ‘Local_Y‘ (lateral and longitudinal), ‘v_Vel‘ (velocity), ‘v_Acc‘ (acceleration), ‘Lane_ID‘ (current lane), ‘Preceding‘ (ID of the vehicle ahead in the same lane), ‘Space_Headway‘, and ‘Time_Headway‘ (spacing and time to the preceding vehicle). ‘v_Class‘ categorizes vehicles into motorcycles, autos, or trucks.
- Scale: the raw dataset comprises approximately 11.8 million rows and 25 columns, representing hundreds of hours of aggregated vehicle movement and thousands of individual vehicle trajectories. This rich, high-volume data provides an extensive basis for detailed behavioral analysis.

A representative sample of the raw NGSIM trajectory data is presented in Table 1, illustrating the granular nature of the information available for each vehicle at a given timestep.

Table 1. Sample raw NGSIM trajectory data

Veh_ID	Frame	Global_Time	v_Vel	v_Acc	Lane	Prec	Headway	v_Class
515	100	1118184010000	58.7	0.3	3	511	2.15	2
515	101	1118184010100	58.9	0.2	3	511	2.14	2
515	102	1118184010200	59.1	0.2	3	511	2.13	2
515	103	1118184010300	59.0	-0.1	3	511	2.16	2
516	100	1118184010000	45.2	1.5	2	513	1.80	2
516	101	1118184010100	45.4	0.8	2	513	1.78	2
517	100	1118184010000	62.1	0.0	4	0	999.9	2
517	101	1118184010100	62.1	0.0	4	0	999.9	2

3.2. Initial filtering and cleaning

To create a focused and reliable dataset for the profiling of standard passenger vehicle behaviors, the following initial processing steps were meticulously applied:

- Vehicle class filtering: only vehicles categorized with a ‘v_Class‘ equal to 2 (automobiles) were retained, excluding motorcycles and trucks. This reduced the dataset to approximately 11,473,680 rows and 25 columns.
- Trajectory length filtering: to ensure that each observed driver segment provides sufficient information to characterize a driving style, vehicles with ‘Total_Frames‘ less than 300 (equivalent to 30 seconds of continuous data at 10 Hz) were removed. This refinement ensured meaningful observation lengths, resulting in a dataset of approximately 11,472,258 rows and 3,211 unique ‘Vehicle_ID‘s.
- Erratic data removal: data points containing physically unrealistic or erroneous ‘v_Acc‘ (instantaneous acceleration) or ‘v_Vel‘ (instantaneous velocity) values were filtered out. Thresholds were set to retain ‘v_Acc‘ values between -35.0 and 35.0 ft/s^2 and ‘v_Vel‘ values between 0.0 and 140.0 ft/s (approximately 95 mph). This crucial step addressed potential sensor noise or data transcription errors. Importantly, this filtering step did not significantly alter the total number of rows or unique vehicles, indicating a relatively clean initial raw dataset after the first two filtering steps.

3.3. Driver observation segment extraction

To capture granular snapshots of individual driver behavior for subsequent profiling, the cleaned vehicle trajectories were segmented into overlapping observation windows. This approach maximizes the number of training samples and allows the model to learn from the temporal evolution of driving styles.

- Segment length: each extracted segment consists of 300 consecutive frames, representing 30 seconds of continuous driving data. This duration was selected as a balanced timeframe, long enough to capture characteristic behavioral patterns (e.g., a few lane changes, varied acceleration patterns) but short enough to represent a consistent driving moment.
- Step size: segments were extracted using a sliding window approach with a ’step_size’ of 50 frames (equivalent to 5 seconds). This overlap between consecutive segments for the same vehicle substantially increases the number of available samples for training the deep learning model.
- Extraction process: the process involved iterating through each unique ‘Vehicle_ID‘ in the cleaned DataFrame. For each vehicle, continuous sub-trajectories of 300 frames were extracted, advancing by 50 frames at a time. A strict check ensured that only complete 300-frame segments were retained. This procedure generated a total of 211,824 driver observation segments, each consisting of a DataFrame retaining the original 25 NGSIM columns for the 30-second window.

3.4. Microscopic behavioral feature engineering

For each of the 211,824 extracted segments, a comprehensive set of 27 features was engineered to create a ”behavioral fingerprint”.These features were selected because they capture the fundamental dimensions of longitudinal control (acceleration), lateral maneuvering (lane changes), and car-following interactions (headway), which are the primary indicators of driving. Let S_k denote the k-th segment consisting of $N_k = 300$ frames.

- Acceleration dynamics (8 features): these characterize longitudinal control. We compute standard statistical moments (mean μ_a, variance σ_a^2, standard deviation σ_a, max, min) for the instantaneous acceleration v_Acc. Additionally, we derive:
  a) Skewness (γ_1): reflects the asymmetry of acceleration, indicating a tendency towards sharp braking or

rapid acceleration.

$$\gamma_1 = \frac{1}{N_k}\sum_{j=1}^{N_k}\left(\frac{v\_Acc_j - \mu_a}{\sigma_a}\right)^3 \tag{1}$$

b) High-intensity thresholds: the percentage of time spent accelerating $> 5ft/s^2$ ($P_{>a_{th}}$) or decelerating $< -5ft/s^2$ ($P_{<d_{th}}$).

$$P_{>a_{th}} = \frac{1}{N_k}\sum_{j=1}^{N_k}\mathbb{I}(v\_Acc_j > 5.0) \times 100\% \tag{2}$$

- Speed metrics (4 features): includes mean, variance, and standard deviation of velocity. We also compute the coefficient of variation (CV_v) to normalize speed variability relative to the mean speed:

$$CV_v = \frac{\sigma_v}{\mu_v}, \quad \text{if } \mu_v \neq 0 \tag{3}$$

- Lane changing behavior (3 features): captures lateral maneuvering.
  a) Lane change rate (R_{LC}): normalized frequency of lane changes per minute.

$$R_{LC} = \frac{\text{Count of lane changes}}{0.5 \text{ min}} \tag{4}$$

  b) Lateral deviation (σ_{lat}): standard deviation of the vehicle's local X position within the lane, serving as a proxy for lane-keeping stability.
- Headway management (9 features): describes car-following interaction. We compute statistics (mean, variance, min) for both space headway (μ_{h_s}) and time headway (μ_{h_t}), calculated exclusively on frames where a preceding vehicle is present (F_k).
  a) Critical headway percentage: the proportion of following time spent with a time headway below 1.0 second (indicating tailgating).

$$P_{<h_{crit}} = \frac{1}{|F_k|}\sum_{j\in F_k}\mathbb{I}(Time\_Headway_j < 1.0) \times 100\% \tag{5}$$

- Contextual features (3 features): includes vehicle physical dimensions (length and width) and normalized time of day (scaled to $[0, 1]$) to account for traffic density variations associated with peak/off-peak hours.

3.4.1. Handling missing headway data

In free-flow conditions, headway values are undefined. To maintain numerical consistency, we impute these missing values with semantic constants: mean/variance are set to 1000.0, and the critical headway percentage is set to 0.0. While this introduces a potential bias by treating "no vehicle" as "very far vehicle," it is necessary for the attention mechanism to process the full sequence and has been justified in similar time-series tasks.

3.5. Heuristic proxy label generation (weak supervision)

Due to the absence of human-annotated labels, we employ a heuristic labeling strategy. We define two behavioral profiles based on statistical thresholds:

- Aggressive: a segment is labeled 'aggressive' if it falls into the top 20^{th} percentile for acceleration variance OR lane change count, or the bottom 20^{th} percentile for mean time headway.
- Normal: all other segments are labeled 'normal'.

To ensure the model does not simply replicate a deterministic rule, we enforce a strict decoupling between the spatial-temporal scope of the inputs and the labels. The proxy labels are global—derived from statistical thresholds calculated over the entire continuous 30-second segment (300 frames). In contrast, the transformer operates on a local temporal sequence, processing the data as 12 sequential sub-windows (12×27 features). Because the model is never fed the global 30-second summary statistics directly, it cannot trivially bypass the learning process. Instead, it is forced to utilize its multi-head self-attention mechanism to aggregate local temporal dynamics and approximate the global behavioral boundary.

3.6. Transformer-based profiling model

We propose a transformer-based sequence model to learn these profiles. The architecture (Figure 1) consists of:
- Input embedding: the 27-feature vectors are projected into a 64-dimensional latent space.
- Positional encoding: added to retain the temporal order of the 12 sub-windows.
- Transformer encoders: two stacked blocks with multi-head self-attention (4 heads) to capture long-range dependencies.
- Classification head: a global average pooling layer followed by a dense Softmax layer to predict the class probabilities (aggressive vs. normal).
- Training: utilized the Adam optimizer with a learning rate of 1e-4, a dropout rate of 0.2, and was conducted over 50 epochs.

The input trajectory sequence (12×27) is projected via an embedding layer and processed by two stacked transformer encoder blocks. These blocks utilize multi-head self-attention to capture temporal dependencies. The resulting features are condensed via global pooling and passed to a multi-layer perceptron (MLP)/Softmax classifier to predict the heuristic behavioral profile (aggressive vs. normal).

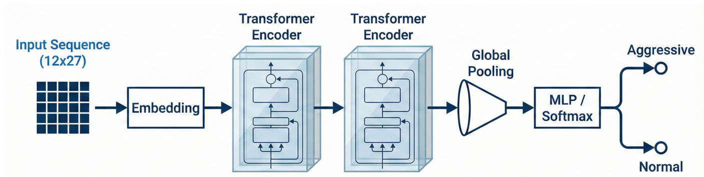

Figure 1. Architecture of the weakly supervised profiling model

4. RESULTS AND DISCUSSION

4.1. Baseline comparison

Four models were evaluated under identical 5-fold cross-validation conditions: logistic regression (LR), random forest (RF), LSTM, and the proposed transformer. Cross-validation was applied to ensure robustness against overfitting to the heuristic proxy labels. Results are reported in Table 2 in terms of accuracy, precision, recall, and F1-score.

Table 2. Comparative performance: transformer vs. baselines

Model	Accuracy	Precision	Recall	F1-score
LR	0.81	0.80	0.81	0.80
RF	0.88	0.89	0.87	0.88
LSTM (baseline)	0.86	0.85	0.86	0.85
Transformer (ours)	0.97	0.97	0.96	0.97

The transformer achieves an F1-score of 0.97, the highest across all models. The LR baseline performs weakest (F1 = 0.80), confirming that the classification boundary between aggressive and normal profiles is inherently non-linear and cannot be resolved by a simple decision function. The RF model improves substantially over LR (F1 = 0.88), demonstrating that ensemble feature interactions capture meaningful behavioral structure. However, because RF operates on independently drawn sub-windows rather than the full temporal sequence, it is unable to capture the ordering and dynamics of driving events. This aligns with the well-established view that driving style is a fundamentally temporal phenomenon: a sudden deceleration in a traffic jam carries an entirely different behavioral meaning than the same action on a free-flow highway [5], [24]. The LSTM baseline (F1 = 0.85) underperforms the RF, likely because recurrent processing is sensitive to the gradient instability that arises when modeling dependencies across long sequences of sub-windows. The transformer overcomes this limitation through its multi-head self-attention mechanism, which assigns attention weights across all 12 sub-windows simultaneously and is therefore better suited to capturing long-range dependencies within the 30-second observation segment [4]. This superiority of self-attention over recurrence at the microscopic, individual-vehicle level is consistent with findings from macroscopic traffic forecasting studies [11], [25].

4.2. Classification performance and behavioral fingerprinting

The confusion matrix obtained on the held-out test set is shown in Figure 2. Of the total 10,000 test samples, 9,675 were correctly classified and 325 were misclassified, giving an overall accuracy of 96.75%. The false-negative rate (aggressive segments predicted as normal) is notably lower than the false-positive rate, indicating that the model is conservative in assigning aggressive labels, which is a desirable property for safety-critical ITS applications.

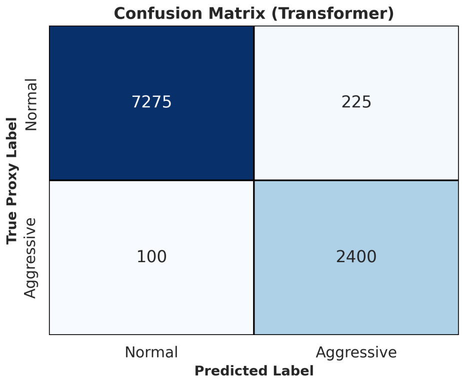

Figure 2. Confusion matrix (transformer)

A closer analysis of the misclassified samples reveals that errors concentrate at the boundary between the two profiles: segments where a driver's kinematic behaviour fluctuates around the 20th/80th percentile thresholds used for proxy labeling. This is an expected outcome of the weak supervision strategy and does not indicate a model failure. On the contrary, it reveals an important property of the learned representation: rather than replicating the deterministic heuristic rule, the transformer has learned a smoother, probabilistic decision boundary that better reflects the continuum of real driving behaviour [5]. A driver whose acceleration variance occasionally crosses the labeling threshold is genuinely ambiguous, and the model's uncertainty in such cases is well-founded. This finding mirrors the distinction observed in macroscopic traffic modeling between crisp rule-based classification and the more nuanced patterns captured by deep sequence models [24].

4.3. Generalization and implications for ITS

To assess whether the learned behavioral signatures are specific to the US-101 corridor or reflect general human driving characteristics, we applied the trained model without retraining to trajectory data from the geographically distinct NGSIM I-80 dataset. Classification accuracy on this held-out dataset was 94%, a reduction of less than three percentage points relative to the in-distribution test set. This degree of cross-dataset transferability suggests that the microscopic behavioral features acceleration dynamics, lane-change rate, and headway management encode driving patterns that are relatively stable across different road environments, consistent with evidence from drone-based studies linking microscopic driving signatures to macroscopic flow outcomes across diverse urban settings [24]. The ability to generalize without retraining is practically important: it indicates that a model trained on one high-resolution dataset could be deployed for driver profiling in environments where new labeled trajectory data are unavailable, supporting scalable and personalized ITS applications such as adaptive warning systems and autonomous vehicle interaction [26].

5. CONCLUSION

This work introduces a weakly supervised transformer-based framework for driver behavior profiling, leveraging NGSIM trajectory data to create distinctive behavioral signatures. By engineering a rich set of microscopic behavioral features and applying a heuristic proxy labeling strategy, we successfully trained a model to classify drivers into 'aggressive' and 'normal' profiles with high fidelity. Crucially, our comparative analysis demonstrates that the transformer architecture significantly outperforms both recurrent LSTM and non-temporal baselines. This finding confirms that the self-attention mechanism is uniquely suited for capturing the long-range temporal dependencies that characterize driving styles. This work demonstrates the feasibility of creating detailed "behavioral fingerprints" from raw trajectory data, moving beyond macroscopic traffic analysis to understand the individual agents. These profiles have significant potential for ITS applications, including enhancing autonomous vehicle perception by predicting human intent. The primary limitation of this study is the reliance on heuristic-based labels, which act as noisy proxies for ground truth. Additionally, the NGSIM US-101 data 2005 is relatively old; shifts in traffic composition and vehicle technology may affect generalization to modern fleets. Future work will explore unsupervised learning and clustering techniques to discover these profiles organically from the data, potentially revealing more nuanced behavioral categories, and integrate explainable artificial intelligence (AI) to interpret the learned fingerprints.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Mohamed Laamimach	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓				
Mghari Mohammed		✓				✓			✓	✓	✓	✓	✓	
Aziz Mabrouk	✓		✓	✓	✓					✓		✓	✓	

C	: **C**onceptualization	I	: **I**nvestigation	Vi	: **Vi**sualization
M	: **M**ethodology	R	: **R**esources	Su	: **Su**pervision
So	: **So**ftware	D	: **D**ata Curation	P	: **P**roject Administration
Va	: **Va**lidation	O	: Writing - **O**riginal Draft	Fu	: **Fu**nding Acquisition
Fo	: **Fo**rmal Analysis	E	: Writing - Review & **E**diting		

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The NGSIM data is available at https://ops.fhwa.dot.gov/trafficanalysistools/ngsim.htm.

REFERENCES

[1] D. Schrank, B. Eisele, and T. Lomax, "*2021 Urban Mobility Report*," Texas A&M Transportation Institute, 2021.

[2] S. An, B.-H. Lee, and D.-R. Shin, "A Survey of Intelligent Transportation Systems," in *2011 Third International Conference on Computational Intelligence, Communication Systems and Networks*, IEEE, Jul. 2011, pp. 332–337. doi: 10.1109/CICSyN.2011.76.

[3] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[4] A. Vaswani *et al.*, "Attention Is All You Need," *arXiv*, pp. 1–9, Aug. 2023, [Online]. Available: http://arxiv.org/abs/1706.03762.

[5] M. Treiber and A. Kesting, *Traffic Flow Dynamics*. Cham: Springer Nature Switzerland, 2025. doi: 10.1007/978-3-031-93922-8.

[6] M. Laamimach and A. Mabrouk, "Urban Congestion Prediction Based on Driver Behavior: Hybrid CNN-LSTM Model with Attention Mechanism," *Lecture Notes in Networks and Systems*, vol. 1924, pp. 491–502, 2026, doi: 10.1007/978-3-032-22567-2_43.

[7] B. M. Williams and L. A. Hoel, "Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results," *Journal of Transportation Engineering*, vol. 129, no. 6, pp. 664–672, Nov. 2003, doi: 10.1061/(ASCE)0733-947X(2003)129:6(664).

[8] J. Zhang, Y. Zheng, and D. Qi, "Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, Feb. 2017, doi: 10.1609/aaai.v31i1.10735.

[9] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transportation Research Part C: Emerging Technologies*, vol. 54, pp. 187–197, 2015, doi: 10.1016/j.trc.2015.03.014.

[10] M. Xu *et al.*, "Spatial-Temporal Transformer Networks for Traffic Flow Forecasting," *arxiv,* Mar. 2021.

[11] J. Jiang, C. Han, W. X. Zhao, and J. Wang, "PDFormer: Propagation Delay-Aware Dynamic Long-Range Transformer for Traffic Flow Prediction," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 4, pp. 4365–4373, Jun. 2023, doi: 10.1609/aaai.v37i4.25556.

[12] J. Zhang, Y. Yang, X. Wu, and S. Li, "Spatio-temporal transformer and graph convolutional networks based traffic flow prediction," *Scientific Reports*, vol. 15, p. 24299, Jul. 2025, doi: 10.1038/s41598-025-10287-5.

[13] S. Wei, Y. Yang, D. Liu, K. Deng, and C. Wang, "Transformer-Based Spatiotemporal Graph Diffusion Convolution Network for Traffic Flow Forecasting," *Electronics*, vol. 13, no. 16, p. 3151, Aug. 2024, doi: 10.3390/electronics13163151.

[14] A. Chang, Y. Ji, and Y. Bie, "Transformer-based short-term traffic forecasting model considering traffic spatiotemporal correlation," *Frontiers in Neurorobotics*, vol. 19, p. 1527908, Jan. 2025, doi: 10.3389/fnbot.2025.1527908.

[15] Z. Fang, L. Pan, L. Chen, Y. Du, and Y. Gao, "Mdtp: A multi-source deep traffic prediction framework over spatio-temporal trajectory data," *Proceedings of the VLDB Endowment*, vol. 14, no. 8, pp. 1289–1297, Apr. 2021, doi: 10.14778/3457390.3457394.

[16] W. Xiong, R. Fonod, A. Alahi, and N. Geroliminis, "Multi-Source Urban Traffic Flow Forecasting with Drone and Loop Detector Data," *arXiv:2501.03492*, 2025, doi: 10.48550/arXiv.2501.03492.

[17] N.-E. El Faouzi, H. Leung, and A. Kurian, "Data fusion in intelligent transportation systems: Progress and challenges – A survey," *Information Fusion*, vol. 12, no. 1, pp. 4–10, 2011, doi: 10.1016/j.inffus.2010.06.001.

[18] H. Zhang, H. Dong, and Z. Yang, "TSGDiff: Traffic state generative diffusion model using multi-source information fusion," *Transportation Research Part C: Emerging Technologies*, vol. 174, p. 105081, May 2025, doi: 10.1016/j.trc.2025.105081.

[19] B. Lim, S. Ö. Arık, N. Loeff, and T. Pfister, "Temporal Fusion Transformers for interpretable multi-horizon time series forecasting," *International Journal of Forecasting*, vol. 37, no. 4, pp. 1748–1764, Oct. 2021, doi: 10.1016/j.ijforecast.2021.03.012.

[20] B. S. Kerner, The Physics of Traffic, vol. 58, no. 11. in *Understanding Complex Systems*, vol. 58. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004. doi: 10.1007/978-3-540-40986-1.

[21] P. A. Lopez *et al.*, "Microscopic Traffic Simulation using SUMO," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, Nov. 2018, pp. 2575–2582. doi: 10.1109/ITSC.2018.8569938.

[22] L. Da, C. Chu, W. Zhang, and H. Wei, "CityFlowER: An Efficient and Realistic Traffic Simulator with Embedded Machine Learning Models," *arXiv:2402.06127*, 2024, doi: 10.48550/arXiv.2402.06127.

[23] G. A. M. Meiring and H. C. Myburgh, "A Review of Intelligent Driving Style Analysis Systems and Related Artificial Intelligence Algorithms," *Sensors*, vol. 15, no. 12, pp. 30653–30682, Dec. 2015, doi: 10.3390/s151229822.

[24] E. Mantouka, P. Fafoutellis, and E. I. Vlahogianni, "From Microscopic Driving Behavior to Macroscopic Traffic Dynamics Using Drone Observations," *Data Science for Transportation*, vol. 8, no. 1, p. 9, Apr. 2026, doi: 10.1007/s42421-026-00152-w.

[25] S. Liu and X. Wang, "An improved transformer based traffic flow prediction model," *Scientific Reports*, vol. 15, no. 1, p. 8284, Mar. 2025, doi: 10.1038/s41598-025-92425-7.

[26] F. A. Sellal, A. Ayoub Bellachia, M. M. Dif, E. De Rautlin De La Roy, M. A. Bouchiha and Y. Ghamri-Doudane, "XAI-Driven Machine Learning System for Driving Style Recognition and Personalized Recommendations," *2025 IEEE 102nd Vehicular Technology Conference (VTC2025-Fall)*, 2025, pp. 1-7, doi: 10.1109/VTC2025-Fall65116.2025.11310755.

BIOGRAPHIES OF AUTHORS

Mohamed Laamimach is a researcher at the Information Security Intelligent Systems and Applications (ISISA) lab. His research interests include deep learning, trajectory analysis, and intelligent transportation systems. He can be contacted at email: mohamed.laamimach@etu.uae.ac.ma.

Mghari Mohammed is a Professor Assistant of Computer Sciences at Abdelmalek Essaadi University, Morocco. Where he received the Ph.D. degree in AI and Computer Science from the Abdelmalek Essaâdi University in 2024. His work specifically involved applying advanced transformer models to low-resource languages, achieving state-of-the-art performance on a specific task. He can be contacted at email: mohammed.mghari@uae.ac.ma.

Aziz Mabrouk is a Full Professor at Abdelmalek Essaâdi University, Tétouan, Morocco. His research focuses on decision support systems, spatial data modeling, and graph-based approaches for complex systems. His scientific contributions address transport and mobility planning, trajectory analysis, voronoi diagrams, and spatial optimization, with applications to intelligent transportation systems and urban decision-making. He can be contacted at email: amabrouk@uae.ac.ma.