Assessing the Effect of Sensor Data Quality and Origin on Driver Behavior Modeling

DANUT DRAGOS DAMIAN[1], FELICIA ANISOARA MICHIS[2], LUMINITA MORARU[1,3]

[1] The Modelling & Simulation Laboratory, Dunarea de Jos University of Galati, Galati, ROMANIA

[2] Emil Racovita High School of Galati, Galati, ROMANIA

[3] Department of Physics, School of Science and Technology, Sefako Makgatho Health Sciences University, Pretoria, SOUTH AFRICA

Abstract: Although several public datasets for driver behaviour analysis are publicly available, inconsistencies in their formats hinder objective comparisons of machine learning and deep learning classification models. Such comparisons are essential for evaluating model performance under realistic driving conditions. To address this limitation, we present a benchmark study that investigates the influence of sensor data quality and source variability on the performance of driver behaviour classification models. Raw inertial measurement unit (IMU) data were analyzed from a publicly available driving dataset (Shardul, 2021, with 14,250 samples) and our proprietary smaller dataset, Drive2025 (containing 6,375 samples), both of which were collected under similar experimental conditions. Also, a combined dataset was built. The classification was performed on sequences of statistical feature vectors describing the dynamic behaviour of the vehicle: mean, variance, standard deviation, skewness, and kurtosis. For classification, the Random Forest (RF) and Support Vector Machine (SVM) algorithms were implemented as representative machine learning models. Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) architectures were used as deep learning counterparts. The experiments demonstrated the impact of data quality and origin on the performance of driver behaviour classification models. The CNN and LSTM models remain the most robust and stable, achieving accuracies of 0.80/0.81 and F1-scores of 0.85/0.84 on the proprietary dataset, and accuracies of 0.83/0.84 with F1-scores of 0.87/0.88 on the public dataset. On the combined dataset, they reached 0.82/0.85 accuracy and 0.84/0.84 F1-score, confirming strong generalization ability. The RF and SVM models showed better performance on the Mendeley dataset, with a moderate drop on the proprietary dataset due to natural noise and data variability. CNN and LSTM have considerable potential for improvement through appropriate filtering and preprocessing. These steps could significantly boost accuracy and prediction stability in real driving scenarios.

Keywords: driver behaviour, IMU signals, Random Forest, CNN, LSTM

Received: May 17, 2025. Revised: July 25, 2025. Accepted: August 17, 2025. Published: January 7, 2026.

1 Introduction

Key component of modern road safety is responsible Adriving, which is characterized by anticipating, adhering to, and being vigilant during travel. Despite the rapid advancements in intelligent vehicle technologies, such as automated emergency response systems, speed-adaptive control systems, and advanced driver assistance systems, human intervention remains crucial in preventing traffic accidents. The International Transport Forum's annual report on road safety for 2024 revealed that traffic accidents resulted in nearly 1.3 million deaths globally in 2023. This statistic underscores the fact that traffic remains one of the leading causes of death globally [1]. The Directorate-General for Mobility and Transport of the

EU (DG MOVE) estimates that there will be a 3% decrease in casualties at the European level,

with roughly 19,800 fewer deaths until 2024. The DG MOVE stresses the ongoing importance of using efficient preventative traffic education and driver behaviour monitoring strategies to effectively manage and reduce road fatalities [2]. To achieve the goal of zero victims by 2050, the EU's Vision Zero initiative and the EU Road Safety Policy Framework 2021–2030 seek to reduce the number of fatalities and serious injuries by 50% by 2030 through integrated care that includes education, the implementation of laws, infrastructure development, and the use of intelligent vehicle assistance systems [3, 4].

As road systems become increasingly automated, adaptive control systems and advanced driver monitoring are essential to ensure transportation safety and longevity. To construct Intelligent Transportation Systems (ITS), build predictive safety models, and apply personalised automobile management tactics, we must first understand individual driving behaviours. In previous studies, numerous computational methods from machine learning and deep learning areas were developed in driving behavior recognition. Deep learning algorithms have several advantages, but their ability to eliminate the need for intricate signal preprocessing and automatically extract features is particularly notable. Zhao et al. [5] built a CNN–BiLSTM–Attention (AM) model that can determine when a driver isn't paying attention. It revealed four tendencies that most drivers have with an accuracy of almost 98%. Using added Gaussian noise and the StateFarm dataset, the accuracy improved to 99.68%. This demonstrates the effectiveness of the attention mechanism in uncovering connections between drivers' operational data that are influenced by both time and location. Chen et al. [6] ran a comparable study and came up with the MCT–CNN–LSTM strategy. It was able to acquire a 97.3% success rate when looking at how people drive in real time with data from wireless sensors by using convolutional feature extraction and sequence modeling. Sun et al. [7] employed a CNN–BiLSTM model to find hard-to-see driving patterns. They could detect most things like straight-line travel lane changes, slowing down, and turning about 98% of the time. The research found that adding temporal convolution layers greatly improved the detection of sudden movements and unusual transitions. Mobini Seraji et al. [8] explored supervised and unsupervised techniques, including SVM, KNN, k-means, fuzzy c-means, and DBSCAN to analyse driver behaviour, predict fuel consumption, and enhance car safety. Additionally, the study examined mixed deep learning models employed in eco-driving vehicles and smart transportation systems. Garefalakis et al. [9] studied standard classifiers and ensemble methods such as Random Forest (RF), AdaBoost, and multilayer perceptrons in simulated and real-world driving scenarios. The RF model emerged as a strong performer with an accuracy of 84% in a controlled environment and 75% in the naturalistic driving study. In both scenarios, RF offers a balanced approach between precision and recall. SVMs outperformed in capturing true positive instances in both datasets, but show lower accuracy (68.67%) and F1-score (53.22%), suggesting a trade-off with precision. Roussou et al. [10] conducted studies on safe driving behaviour and evaluated the performance of LSTM networks and feedforward neural networks (ANNs) in predicting changes in driver time dynamics. They discovered that recurrent designs excel at capturing long-term dependencies, which are crucial for accurately identifying behaviour. LSTM models are adept at capturing temporal dependencies within data. In contrast, ANN models employ a feed-forward architecture that disregards the temporal aspect. The LSTM model specifically capitalises on the sequential nature of the data, which may enhance prediction performance. Hou et al. [11] explored the ways deep learning models can be applied to identify hazardous driving behaviour. To gather information, they investigated four methods: surveys, data from automobiles, eye tracking, and physiological sensing. They also analysed the efficacy of deep learning models, specifically DBN, CNN, and RNN, in identifying risky driving habits. All models achieved an overall classification recognition accuracy exceeding 80%. Shirole et al. [12] emphasised the importance of combining multiple data sources, including automotive telemetry, inertial sensors, and external data, to create comprehensive profiles of drivers' behaviour. They highlighted the critical need for generalisation across different vehicle types and

driving conditions, as this could impact prediction accuracy. Mei et al. [13] integrated driving style recognition with speed planning, covering data collection, preprocessing, and classification techniques. The paper focused on mixed machine learning models for driving behavior classification, considering both long-term and short-term factors, data processing techniques, and evaluation metrics. It also covered unsupervised rule-based and learning-based algorithms, along with evaluation indicators such as time efficiency and accuracy. Through the utilization of the Bi-LSTM model in conjunction with an attention mechanism that is founded on dilated convolutional neural networks (ID-CNN), Wang and Yao [13] addressed more effective methods for locating distracted drivers. The approach focused on the most important features, eliminated duplicate data, and facilitated the extraction of features of different sizes. Applying the model to the StateFarm dataset yielded an accuracy of 95.84% while on the Drive&Act-Distracted dataset, it achieved an accuracy of 97.89%. As a result, combining dilated convolutions with attention processes makes it easier and more accurate to identify drivers who are distracted while operating their vehicles.

This paper presents a benchmark study investigating the influence of sensor data quality and source variability on driver behaviour classification models. A new dataset is generated using real vehicle experiments, which face challenges in controlling environmental and external variables. This approach will enhance the robustness of the drivers' behaviour classification.

The main contributions of this paper are as follows: (1) The Drive2025 dataset, containing 6.375 samples, was generated using data collected from IMU sensors placed on two cars. (2) A set of fundamental statistical features describing the vehicle's dynamic behaviour, including mean, variance, standard deviation, skewness, and kurtosis, was directly computed. (3) To ensure the reliability and validity of the data, real vehicle experiment data were integrated with data from a publicly available driving dataset (Shardul, 2021, with 14.250 samples).

2 Method

This paper aims to conduct a detailed analysis of data recorded by IMU sensors to recognize different driving behaviors and classify them into two categories: normal and aggressive. Figure 1 illustrates the complete IMU-based workflow for driver behavior classification. IMU signals are first segmented using a sliding window mechanism. For each segment, statistical descriptors are computed, forming compact feature vectors. These vectors are then provided as input to both machine learning and deep learning classifiers, enabling a unified benchmark analysis focused on data quality and source variability.

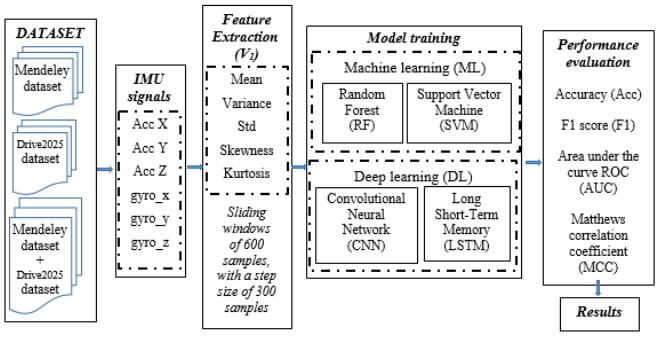

Fig. 1 IMU-based workflow for the classification of driver behavior

The methodological workflow includes the following stages: (1) Loading raw IMU data from CSV files; (2) Computing statistical features for each IMU channel; (3) Building and refining the feature vectors (V_1); (4) Standardizing features through normalization and splitting the dataset; (5) Training the four classification models (RF, SVM, CNN, LSTM); (6) Performance evaluation based on Accuracy, F1-score, AUC, and Matthews Correlation Coefficient (MCC) metrics; (7) Presenting the final results.

The experiment utilized the Shardul dataset (containing 14,250 samples), which was published on Mendeley Data [15]. Data were collected from IMU sensors that record the vehicle's linear accelerations and angular velocities. Our proprietary smaller dataset, Drive2025, contains 6,375 samples collected under similar experimental conditions. Both datasets involve urban and extra-urban routes, with normal and aggressive driving behaviors. To test cross-domain

consistency and model generalization capability, a combined dataset was also generated by merging data from both sources. Linear accelerations and angular velocities measured along the orthogonal axes Acc X, Acc Y, Acc Z, gyro_x, gyro_y, and gyro_z were collected. For each IMU signal, a set of fundamental statistical features, including mean, variance, standard deviation, skewness, and kurtosis, is computed. They were computed over sliding windows of 600 samples with a step size of 300 samples, ensuring 50% overlap between consecutive segments. Four classification models were implemented to assess the predictive capabilities of different learning paradigms. The Random Forest (RF) and Support Vector Machine (SVM) are representative machine learning algorithms operating in the statistical feature space. The Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) are deep learning architectures. CNN and LSTM architectures are trained on sequences of statistical feature vectors. This design choice ensures a fair and controlled comparison with traditional machine learning models operating in the same feature space, while still allowing LSTM to capture temporal dependencies across consecutive windows.

Each classifier was trained to distinguish between normal and aggressive driving behaviors based on patterns extracted from IMU signals.

2.1 Statistical Feature Extraction from IMU Signals

To describe driver dynamics, five key statistical features are computed from data collected from IMU sensors: mean (μ), variance (σ^2), standard deviation, skewness (skew), and kurtosis (kurt). Let a_i denote the i^{th} sample of the IMU signal, N is the total number of samples in the analyzed window, and E[·] is the expectation operator (mean of all samples). These features were calculated over each imposed window of the IMU time series, as follows [16].

Mean (μ)

$$\mu = \frac{1}{N}\sum_j a_j \quad (1)$$

Variance (Var)

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(a_i - \mu)^2 \quad (2)$$

Standard Deviation (STD)

$$STD = \sqrt{\sigma} \quad (3)$$

Skewness (Skew)

$$skew = \frac{E[(a-\mu)]^3}{\sigma^3} \quad (4)$$

Kurtosis (Kurt)

$$kurt = \frac{E[(a-\mu)]^4}{\sigma^4} \quad (5)$$

2.2 Classifier Algorithm

Random Forest (RF) - builds an ensemble of T decision trees trained on random subsets of the data and features. For a given input x, the final class prediction is obtained through majority voting [17]:

$$\hat{y} = mode\{h_1(x), h_2(x), ..h_T(x)\} \quad (6)$$

Each decision tree, denoted as $h_i(x)$, partitions the feature space by minimizing the Gini impurity:

$$G = 1 - \sum_{k=1}^{k} p_k^2 \quad (7)$$

Where p_k is the proportion of samples belonging to the class k in a node. The randomization in both feature selection and data sampling reduces correlation between trees and improves generalization.

Support Vector Machine (SVM) - aims to find an optimal separating hyperplane that maximizes the margin between two classes. The optimization problem

is formulated as [18]:

$$\min_{w,b} \frac{1}{2}\|w\|^2 \text{ s.t.} y_i(w^T x_i + b) \geq 1, \forall i$$
(8)

For nonlinear separations, a kernel function $K(x_i, x_j)$ is introduced to map the data into a higher-dimensional feature space as follows:

$$K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$$
(9)

Common kernels include the Radial Basis Function (RBF) is defined as $K(x_i, x_j) = \exp(-\Upsilon\|x_i - x_j\|)^2$ and provides nonlinear decision boundaries.

Convolutional Neural Network (CNN) - In a CNN, the convolutional layer applies a set of learnable filters (kernels) to the input signal or image. For a 1D convolution (e.g., IMU signal) [19]:

$$s(t) = (x * w)(t) = \sum_{r=0}^{k-1} x(t+r) \cdot w(r)$$
(10) Each convolutional operation is followed by an activation function (e.g., ReLU) and pooling to reduce dimensionality:

$$P(i) = \max_{t \in R(i)} 8(t)$$
(11)

The final layers are fully connected and output the probability distribution over classes through the softmax function:

$$P(y = k|x) = \frac{e^{z_k}}{\sum_{j=1}^{K} e^{z_j}}$$
(12)

CNNs automatically learn hierarchical spatial representations from sensor data.

Long Short-Term Memory (LSTM) networks - extend recurrent neural networks by incorporating memory cells and gating mechanisms to preserve long-term dependencies [20]. For each time step t:

$$f_t = \sigma\left(W_f\left[h_{t-1}, x_t\right] + b_f\right) \text{ (forget gate)}$$
(13)

$$i_t = \sigma\left(W_i\left[h_{t-1}, x_t\right] + b_i\right) \text{ (input gate)}$$
(14)

$$\tilde{C}_t = \tanh\left(W_C\left[h_{t-1}, x_t\right] + b_C\right) \text{ (candidate cell state)}$$
(15)

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \text{ (cell state update)}$$
(16)

Where W_f represents the weight matrix associated with the forget gate, $\left[h_{t-1}, x_t\right]$ denotes the concatenation of the current input and the previous hidden state, b_f is the bias with the forget gate, and σ denotes the sigmoid activation function. $\cdot$ denotes element-wise multiplication, tanh is an activation function, and $i_t \cdot C_t$ that represents the new candidate values scaled by how much we decided to update each state value.

On the other hand,

$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right) \text{ (output gate)}$$
(17)

$$h_t = o_t * \tanh\left(C_t\right) \text{ (hidden state)}$$
(18)

Where o_t is the output gate activation and C_t is the current cell state.

2.3 Evaluating performance

Accuracy (Acc), F1 score (F1), area under the curve ROC (AUC), and Matthews Correlation Coefficient (MCC) are the metrics used to evaluate the performance of each classifier[21]. Let TP, TN, FP, and FN denote the number of true positives, true negatives, false positives, and false negatives, respectively, in the confusion matrix.

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}} \quad (19)$$

Acc gives a basic idea of how well a classifier works, but it might be misleading when working with datasets that aren't balanced [22]. It's also important to consider the F1-score which offers a clearer picture of the model's true performance. [23].

$$\text{F1} = \frac{2\left(\text{Precision x Recall}\right)}{\text{Precision} + \text{Recall}} \quad (20)$$

Where, $\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$ and $\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$.

The F1-score is a way to find out how well a machine learning model performs classification. A higher F1-score, which ranges from 0 to 1, means greater overall performance.

Area under the curve ROC (AUC) [24]:

$$\text{AUC} = \int_0^1 \text{TPR}\left(\text{FPR}\right)\text{d}\left(\text{FPR}\right) \quad (21)$$

Where, $\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}$ and $\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}$.

An AUC close to 1 indicates excellent separability, meaning the model can effectively distinguish between classes, whereas an AUC of 0.5 implies random performance [25]. In driver behavior modeling, AUC is widely used because it is threshold-independent and robust to class imbalance [26].

$$\text{MCC} = \frac{\left(\text{TPx TN}\right) - \left(\text{FPxFN}\right)}{\sqrt{(\text{TP} + \text{FP}(\left(\text{TP} + \text{FN}\right)\left(\text{TN} + \text{FP}\right)\left(\text{TN} + \text{FN}\right)}} \quad (22)$$

MCC is useful for imbalanced datasets since it considers both positive and negative classes. It might be anywhere from -1 to +1. A +1 means the prediction was perfect, a 0 means it was no better than random, and a -1 means that the prediction and observation were completely different.

This study focuses on a controlled benchmark analysis to determine how sensor data quality and origin affect classification performance across various learning paradigms, rather than novelty in algorithms.

3 Results and Discussions

A sliding window containing 600 samples with a step size of 300 samples was used to generate the dataset. This ensured a 50% overlap between consecutive segments. This choice was guided by the following rule: the window and segment sizes were increased until the error approximation reached its minimum. The proprietary Drive2025 dataset covers realistic driving conditions like road imperfections, vibrations, and uncontrollable environmental factors. Although the dataset has a small sample size, we chose it because our study prioritises resilience over absolute accuracy.

The Mendeley public dataset has 46 segments for analysis, the Drive2025 proprietary dataset has 19 segments, and the combined dataset has 67 segments. Each segment contains 36 statistical descriptors extracted from six IMU channels (Acc X/Y/Z, gyro_x/y/z). These descriptors cover measures of amplitude, variability, and distribution asymmetry. The dataset is divided into 80% training and 20% testing. All models were trained to discriminate between normal (0) and aggressive (1) driving behaviours.

The hardware environment consisted of an Intel(R) Core(TM) i3-4030U CPU running at 1.90 GHz and 8 GB of RAM, running on Windows 10 Pro.

Boxplots were created for each dataset to evaluate the descriptive and classification capabilities of each feature. (Figs. 2-4). These results clearly demonstrate the impact of sensor data quality and source variability on the performance of driver behaviour classification models. Distinct clusters of descriptors separate the normal (0) and aggressive (1) classes.

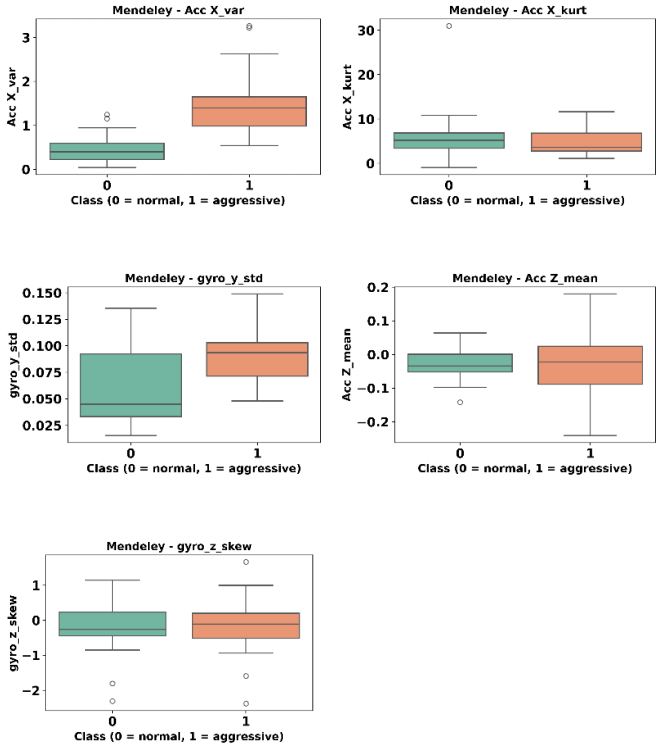

Fig 2. Boxplots of statistical descriptors for the Mendeley database of normal (0) and aggressive (1) driving behaviours during testing.

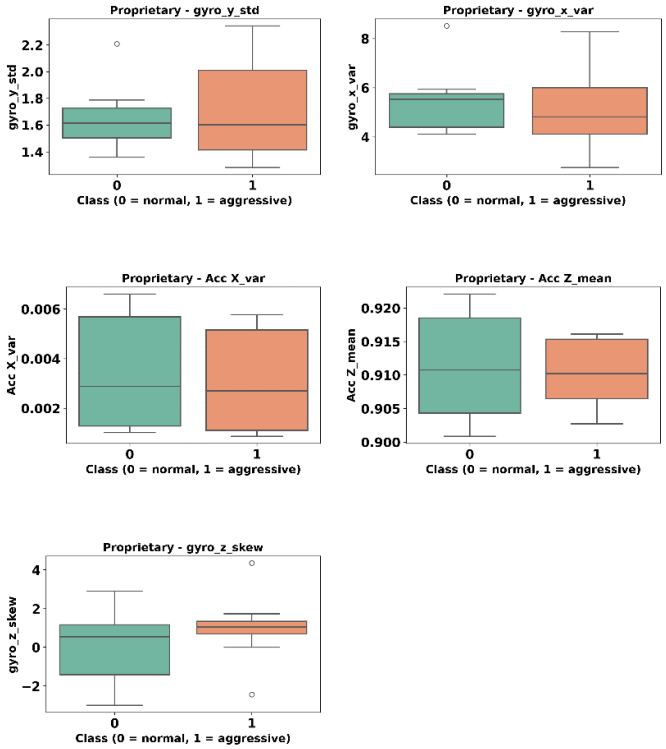

Fig 3. Boxplots of statistical descriptors for the Drive2025 database of normal (0) and aggressive (1) driving behaviours during testing.

Figure 2 illustrates the discrimination value of statistical descriptors for the Mendeley database. Unlike variance and standard deviation, skewness and kurtosis reveal asymmetry and impulsive driving indicative of aggressive behaviour. The Acc_X_var and Acc_X_kurt data highlight strong longitudinal variations in acceleration, specific to braking and rapid acceleration phases. Gyro_Y_std shows high dispersion for aggressive behaviors, reflecting the lateral instability of the vehicle. Acc_Z_mean captures differences in vertical oscillation caused by road irregularities, while Gyro_Z_skew indicates a pronounced asymmetry in rotational movements around the vertical axis. The distributions are compact and coherent, and the differences between the medians confirm that longitudinal and lateral dynamics are the main factors defining aggressive behavior in this dataset.

Figure 3 presents boxplots of statistical descriptors for the proprietary Drive2025 set. This data set was recorded under real traffic conditions, the variability of the signals is higher, and the influence of external factors (surface type, vibrations, driver–vehicle interaction) is visible in the gyroscopic characteristics. The Gyro_Y_std and Gyro_X_var descriptors clearly separate classes, suggesting sudden steering movements and lateral instability. Acc_X_var and Acc_Z_mean reveal moderate but consistent differences between smooth and aggressive manoeuvres, while Gyro_Z_skew demonstrates extreme values and pronounced asymmetry, typical of irregular movements and strong vibrations. In real driving, gyroscopic descriptors are the most sensitive to aggressive behaviours, while acceleration characteristics remain relevant for identifying longitudinal variations.

Figure 4 presents boxplots illustrating statistical descriptors for combined datasets.

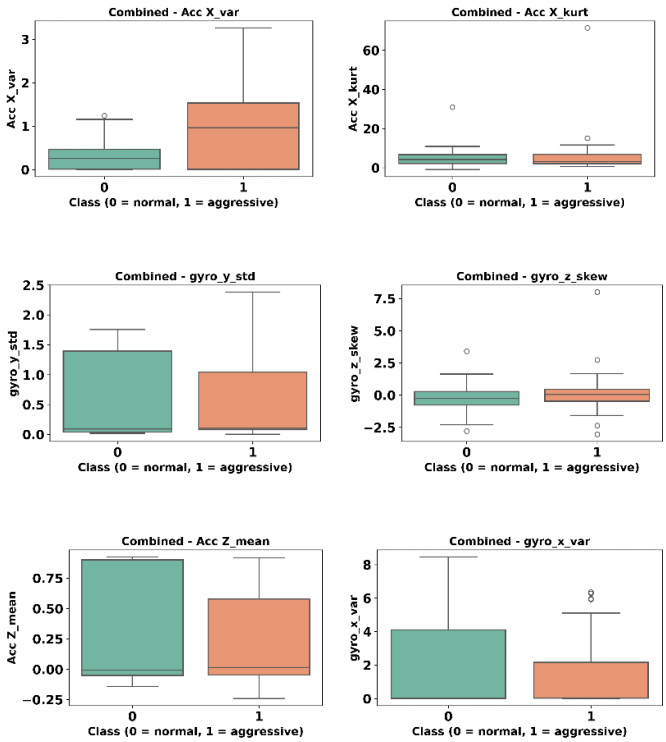

Fig 4. Boxplots of statistical descriptors for the combined database of normal (0) and aggressive (1) driving behaviours during testing.

When the ordered structure of the Mendeley data was combined with the realistic complexity of the Drive2025 dataset, a high stability of the discriminant features was observed. The stability of Acc_X_var, Acc_X_kurt, Gyro_Y_std, Gyro_Z_skew, Acc_Z_mean, and Gyro_X_var descriptors was evident in their ability to separate between classes regardless of the data's origin. This stability demonstrates that they capture fundamental properties of vehicle dynamics, independent of factors such as data source, road type, or recording conditions. These results confirm that the analysed features are domain-invariant descriptors. This makes them useful for building robust models to recognise aggressive driving behaviour. Furthermore, the Acc_X_var and Gyro_Y_std features consistently emerge as major discriminant factors. This suggests that longitudinal and lateral manoeuvres are the most pertinent for detecting aggressive behaviour. The stability of these descriptors within the combined set reinforces their potential for use in generalisable classification models and mobile driving style monitoring applications.

Table 1 compares the results provided by classifiers in terms of performance measures. The results demonstrate that the LSTM model consistently achieved the highest overall performance across all datasets, with the CNN model performing closely behind.

Table 1. Average performance metrics for the selected classifiers: RF, SVM, CNN, and LSTM

Dataset	Model	Accuracy	F1-score	AUC	MCC
Mendeley (46 windows)	RF	0.86	0.85	0.88	0.81
	SVM	0.84	0.83	0.86	0.78
	CNN	0.88	0.87	0.91	0.84
	LSTM	**0.89**	**0.88**	**0.92**	**0.86**
Drive2025 (19 windows)	RF	0.79	0.78	0.81	0.73
	SVM	0.77	0.76	0.79	0.70
	CNN	0.81	0.80	0.83	0.77
	LSTM	**0.82**	**0.81**	**0.84**	**0.78**
Combined (67 windows)	RF	0.83	0.82	0.85	0.79
	SVM	0.82	0.81	0.84	0.78
	CNN	0.85	0.84	0.87	0.81
	LSTM	**0.86**	**0.85**	**0.88**	**0.82**

These results confirm that deep learning models exhibit higher robustness to data variability, particularly in noisy real-world conditions. In contrast, traditional machine learning classifiers are more susceptible to fluctuations in sensor quality. This explains their reduced performance on the proprietary dataset.

Both deep learning architectures exhibited significant robustness, consistently achieving stable F1-scores and MCC values despite a reduction in the number of samples. This highlights their ability to capture temporal dependencies and spatial patterns in the IMU signal sequences, which is crucial for differentiating between smooth and abrupt manoeuvres. The RF and SVM classifiers performed less well than the deep learning classifiers but showed better performance on the Mendeley dataset, with a moderate drop on the proprietary dataset due to natural noise and data variability. Among them, RF achieved an AUC closer to values reached by deep learning classifiers. This indicated a good balance between sensitivity and specificity, suggesting the RF model is strong and practically feasible.

While this approach has potential, it does come with some limitations and challenges. Firstly, it introduces a slight computational complexity due to the use of sliding windows and feature extraction. Secondly, achieving real-time performance is a significant challenge, especially when demonstrating high classification accuracy in real-time driving environments.

4 Conclusions

This study reveals that data quality, sensor reliability, and feature representativeness impact the classification of driving behaviour. The results demonstrate that data quality and source directly influence model performance. Notably, Mendeley achieved the highest accuracy (with an LSTM model achieving 0.89 accuracy), thanks to its consistent sampling and low sensor noise. In contrast, the Drive2025 model struggles with data variability and signal fluctuations caused by uneven roads, steep curves, and natural noise. These findings suggest that resilient temporal models like LSTM can effectively predict driver behaviour in noisy and uncertain environments.

Since the combined dataset improved generalisation accuracy and reduced overfitting in smaller datasets, future research should consider using data augmentation or synthetic data synthesis techniques, such as GANs.

References

[1] International Transport Forum (ITF). (2024). Road Safety Annual Report 2024. OECD Publishing, Paris. https://www.itf-oecd.org/sites/default/files/docs/irtad-road-safety-annual-report-2024.pdf

[2] European Commission, Directorate-General for Mobility and Transport (DG MOVE). (2025). EU road fatalities drop 3% in 2024, but progress remains slow. Brussels. https://transport.ec.europa.eu/news-events/news/eu-road-fatalities-drop-3-2024-progress-remains-slow-2025-03-18_en

[3] European Commission. (2021). EU Road Safety Policy Framework 2021–2030: Next steps towards "Vision Zero". Publications Office of the European Union. https://www.krbrd.gov.pl/wp-content/uploads/2022/05/EU-Road-Safety-Policy-Framework-2021-2030.pdf

[4] European Climate, Infrastructure and Environment Executive Agency (CINEA). (2023). EU Road Safety: Towards Vision Zero. https://cinea.ec.europa.eu/publications/digital-publications/eu-road-safety-towards-vision-zero_en

[5] D. Zhao, H. Li, Z. Fu, B. Ma, F. Zhou, C. Liu, and W. He, "A novel method for distracted driving behaviors recognition with hybrid CNN-BiLSTM-AM model", Complex & Intelligent Systems, vol. 11, no. 357, pp. 1–17, 2025. https://doi.org/10.1007/s40747-025-01983-w

[6] K. Chen, Y. Zhang, L. Wang, H. Liu, and X. Zhao, "MCT-CNN-LSTM: A driver behavior wireless perception", Sensors, vol. 25, no. 4, pp. 1–18, 2025. https://doi.org/10.3390/s25072268

[7] G. Sun, H. Zhang, L. Zhong, and Q. Li, "Identification of driving behavior in continuous diverging sections of expressway system interchange based on CNN-BiLSTM", Scientific Reports, vol. 15, no. 1, art. no. 10631, pp. 1–14, 2025. https://doi.org/10.1038/s41598-025-94000-6

[8] M. H. Mobini Seraji, S. Shaffiee Haghshenas, V. Simic, D. Pamucar, G. Guido, and V. Astarita, "A state-of-the-art review on machine learning techniques for driving behavior analysis: clustering and classification approaches", Complex & Intelligent Systems, vol. 11, art. no. 386, 2025. https://doi.org/10.1007/s40747-025-01988-5

[9] T. Garefalakis, E. Michelaraki, S. Roussou, C. Katrakazas, T. Brijs and G. Yannis, "Predicting risky driving behavior with classification algorithms: results from a large-scale field-trial and simulator experiment", European Transport Research Review, vol. 16, art. no. 65, 2024. https://doi.org/10.1186/s12544-024-00691-9

[10] S. Roussou, E. Michelaraki, C. Katrakazas, A. P. Afghari, C. Al Haddad, M. R. Alam, C. Antoniou, E. Papadimitriou, T. Brijs and G. Yannis, "Unfolding the dynamics of driving behavior: a machine learning analysis from Germany and Belgium", European Transport Research Review, vol. 16, no. 1, art. no. 40, 2024. https://doi.org/10.1186/s12544-024-00655-z

[11] J. Hou, B. Zhang, Y. Zhong and W. He, "Research Progress of Dangerous Driving Behavior Recognition Methods Based on Deep Learning",

World Electric Vehicle Journal, vol. 16, no. 2, art. no. 62, 2025.
https://doi.org/10.3390/wevj16020062
[12] V. Shirole, A. K. Shahade, and P. V. Deshmukh, "A comprehensive review on data-driven driver behavior scoring in vehicles: technologies, challenges and future directions", Discover Artificial Intelligence, vol. 5, art. no. 26, 2025. https://doi.org/10.1007/s44163-025-00244-6
[13] P. Mei, H. R. Karimi, L. Ou, H. Xie, C. Zhan, G. Li, and S. Yang, "Driving style classification and recognition methods for connected vehicle control in intelligent transportation systems: A review", ISA Transactions, vol. 158, pp. 167–183, 2025. https://doi.org/10.1016/j.isatra.2025.01.033
[14] Z. Wang and L. Yao, "Recongnition of Distracted Driving Behavior Based on Improved Bi-LSTM Model and Attention Mechanism", in IEEE Access, vol. 12, pp. 67711-67725, 2024, https://doi.org/10.1109/ACCESS.2024.3399789.
[15] S. Nazirkar, "Phone sensor data while driving a car and normal or aggressive driving behaviour classification", Mendeley Data, Vol.1, 2021.
https://doi.org/10.17632/5stn873wft.
[16] E. Escobar-Linero, F. Luna-Perejón, L. Muñoz-Saavedra, J. L. Sevillano, and M. Domínguez-Morales, "On the feature extraction process in machine learning. An experimental study about guided versus non-guided process in falling detection systems", Engineering Applications of Artificial Intelligence, vol. 114, pp 105170, 2022.
https://doi.org/10.1016/j.engappai.2022.105170
[17] L.Breiman, "Random forests", Machine Learning, vol. 45, no.1, pp. 5–32, 2001.
http://dx.doi.org/10.1023/A:1010933404324
[18] C. Cortes, and V. Vapnik, "Support-vector networks", Machine Learning, vol. 20, pp. 273–297, 1995.
http://dx.doi.org/10.1007/BF00994018
[19] T. Cover, and P. Hart, "Nearest neighbor pattern classification", in IEEE Transactions on Information Theory, vol. 13, no. 1, pp. 21–27, 1967.
https://doi.org/10.1109/TIT.1967.1053964
[20] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to Forget: Continual Prediction with LSTM", Neural Computation, vol. 12, no. 10, pp. 2451–2471, 2000.
https://doi.org/10.1162/089976600300015015
[21] D. M. W. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation", Journal of Machine Learning Technologies, vol. 2, no.1, pp. 37–63, 2011.
https://doi.org/10.9735/2229-3981
[22] A. Tharwat, "Classification assessment methods", Applied Computing and Informatics, vol. 17, no. 1, pp. 168–192, 2021.
https://doi.org/10.1016/j.aci.2018.08.003
[23] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning", Scientific Reports, vol. 14, no. 6086, 2024.
https://doi.org/10.1038/s41598-024-56706-x.
[24] V. Škvára, T. Pevný, and V. Šmídl, "Is AUC the best measure for practical comparison of anomaly detectors?", arXiv preprint arXiv:2305.04754, *2023*.
https://doi.org/10.48550/arXiv.2305.04754
[25] K. Feng, H. Hong, K. Tang, and J. Wang, "Decision Making with Machine Learning and ROC Curves", arXiv preprint arXiv:1905.02810, 2019.
https://doi.org/10.48550/arXiv.1905.02810
[26] D. Chicco, and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation", BMC Genomics, vol. 21, art. no. 6, 2020. https://doi.org/10.1186/s12864-019-6413-7