2023 IEEE 26th International Conference on
Intelligent Transportation Systems (ITSC)
24-28 September 2023. Bilbao, Bizkaia, Spain

Driving Style Classification using Deep Temporal Clustering with Enhanced Explainability

Yuxiang Feng[1]*, *Member, IEEE*, Qiming Ye[2], Fahmy Adan[1], Luis Marques[1], Panagiotis Angeloudis[1]

2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC) | 979-8-3503-9946-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ITSC57777.2023.10421826

***Abstract*— The driving styles of human drivers exhibit a diverse range of observed driving patterns and manoeuvres, influenced by their habitual choice of driving behaviours. Understanding these driving styles is essential for enhancing road safety and reducing emissions. In this paper, we propose a deep temporal clustering approach to classify driving styles, providing enhanced explainability. A comprehensive car-following dataset was collected, incorporating extensive feature parameters. Subsequently, we developed a deep temporal clustering-based classification method that considers the variations in driving style within a single trip. The performance evaluation employed K-Shape clustering, and the significance of various features was assessed using SHAP values, enhancing the interpretability of the model. Our findings contribute to the advancement of driving style classification methods, promoting a deeper understanding of driving behaviours for improved road safety measures.**

I. Introduction

The rapid advancement of autonomous driving technologies is bringing us closer to the commercialization of autonomous vehicles (AVs). While ensuring the safety operation of AVs remains the primary focus of current research, it is also crucial to consider personalized AV control in future studies [1]. Ignoring the inter-personal differences in driving preferences can negatively impact passenger comfort and hinder public acceptance of these vehicles [1].

The first formal definition of driving style is ”the way individuals choose to drive or driving habits that have become established over a period of years” [2]. Despite extensive research on this topic, a unified definition of driving style has yet to be established within academic and professional traffic safety communities [3]. Generally, driving style refers to a driver's habitual choice of driving maneuvers, reflecting differences in cognition and action characteristics.

While some studies quantify driving style variations in terms of aggressiveness or violation levels, most classify it into distinct groups, namely aggressive, moderate, and conservative [4]. Aggressive driving style involves risky speeding, abrupt speed changes, harsh acceleration and deceleration, and improper lateral position maintenance. Conservative driving, on the other hand, encompasses opposite behaviors, while moderate driving serves as a reference for isolating other driving styles [4]. Numerous modern approaches have successfully classified these driving styles [5]–[10]. However, these approaches often assume that each participant's driving style remains consistent throughout a trip and overlook potential variations within a single trip. While this assumption may be suitable for overall evaluations, an alternative approach is to divide the entire trip into a set of maneuver segments, such as accelerating, braking and maintaining. Drivers exhibit aggressive driving style in accelerating manoeuvres does not necessarily imply they are aggressive in other manoeuvres [3]. Therefore, this paper proposes a manoeuvre-based driving style classification approach that aims to capture variations in each individual's driving patterns.

In recent years, deep learning-based methods have gained popularity in driving style classification studies due to their superior pattern recognition capabilities [11]. Although these methods achieve higher accuracy than conventional approaches, their ”black box” nature hinders explainability and makes it difficult to understand their underlying mechanisms [12]. To enhance the explainability and interpretability of classification results, Explainable Artificial Intelligence (XAI) techniques have been proposed. In this study, we propose fitting cluster labels with a supervised model and employing SHapley Additive exPlanations (SHAP) values to measure the importance of features in the obtained results [13]. This approach can improve our understanding of the classification results and recommend the best combination of feature parameters to represent these driving styles.

The primary contributions of this study are outlined as:

1) A deep clustering method considering temporal features is developed for driving style analysis.
2) The proposed approach emphasizes the driving style diversities observed in various driving maneuvers.
3) A comprehensive set of feature parameters is selected in this study. Their influence on classification results is analysed to improve explainability of the model.

This paper has been structured into five sections. Section II briefly reviews some relevant literature, both their classification method and feature parameters are evaluated. Section III explains the adopted methodology, consisting of a brief summary of the method and detailed explanation of each components. The obtained results and discussion of analysis are presented in section IV. Finally, section V discusses the findings and limitations of this study and suggests the potential for future works.

II. Literature review

The literature on driving style classification is extensive, as evidenced by various studies [5]–[10]. Table. I provides

[1]Y. Feng, F. Adan, L. Marques and P. Angeloudis are with the Transport Systems and Logistics Laboratory, Department of Civil and Environmental Engineering, Imperial College London, London SW7 2AZ, U.K.
[2]Q. Ye is with Future Cities Lab Global, Singapore Hub, Singapore-ETH Centre at CREATE, 1 Create Way, 138602, Singapore.
* Corresponding author: Yuxiang Feng (email: y.feng19@imperial.ac.uk)

TABLE I

SUMMARY OF DRIVING STYLE CLASSIFICATION STUDIES

Author	Year	Method	Feature Parameters						
			Acc.	Speed	Gap	Time Gap	RPM	Pedal	Steering
Albers and Albrecht [10]	2005	Self-organizing map	✓	✓			✓	✓	
Constantinescu et al. [8]	2010	K-means, hierarchical clustering	✓	✓					
Aljaafreh et al. [5]	2012	Fuzzy logic	✓	✓					
Al-Din et al. [6]	2013	Fuzzy logic	✓	✓	✓				
Dörr et al. [7]	2014	Fuzzy logic	✓	✓		✓			
Vaitks et al. [9]	2014	K-nearest neighbors	✓						
Wang et al. [14]	2017	semisupervised SVM		✓				✓	
Feng et al. [3]	2018	Support Vector Clustering		✓	✓		✓	✓	
Li et al. [11]	2019	CNN, LSTM	✓	✓				✓	✓
Guo et al. [15]	2021	Voting Decision tree+SVM+KNN	✓	✓				✓	
Zheng et al. [16]	2022	K-means in frequency domain	✓	✓	✓	✓			
Li et al. [17]	2022	Similarity-based clustering with DTW	✓	✓					

an overview of the methods used to differentiate driving styles, including both supervised [9], [11], [14], [15] and unsupervised methods [3], [8], [10], [16], [17]. Supervised classification methods require labeled training data to train the classifier, often achieved through experts' knowledge. On the other hand, unsupervised methods can be applied directly to unlabeled data. While the results are more objective, interpreting the outcomes can be challenging. High-dimensional driving data can pose difficulties in achieving optimal clustering performance. Existing studies demonstrate the effectiveness of clustering algorithms by comparing the results with expert knowledge [7] or by evaluating the data characteristics within each cluster and linking them to corresponding driving style groups [3], [16], [17].

Although supervised classification methods offer better interpretability, subjective labels based on experts' knowledge can introduce errors in data labeling, potentially failing to accurately represent the underlying driving style. Additionally, manually labeling driving data can be time-consuming and require substantial effort when dealing with a large number of participants. Consequently, unsupervised clustering methods present a more promising solution for driving style classification. Furthermore, as driving data is typically represented as time series, capturing and integrating the temporal correlations of the data is beneficial for accurate classification [11]. However, most existing studies treat driving data as independent samples and overlook temporal features.

The feature parameters utilized in previous studies are also summarized in Table. I. These parameters can be broadly categorized into three groups: vehicle-related, cognition-related, and operation-related. Vehicle-related parameters encompass vehicle kinematics, such as acceleration and vehicle speed, which are commonly employed in existing studies. Additionally, cognition-related parameters such as headway gap and time gap, which have distinct interpretations among different driving style groups, are incorporated to capture the driver's cognition characteristics [3], [6], [7], [16]. Lastly, operation-related parameters, including engine RPM, pedal opening, and steering wheel rotation, serve as indicators of drivers' operational characteristics [3], [10], [11], [14], [15]. Different combinations of these feature parameters have been used in existing studies. However, there is a lack of evaluation regarding the significance of these parameters and their contributions to the driving style classification results.

Thus, improvements are needed in existing clustering methods to better incorporate the temporal correlations of high-dimensional driving data and enhance the explainability of the obtained results. As a pioneering study, we propose a deep temporal clustering methods for driving style classification that addresses the variations within a trip. The influence of a more comprehensive set of feature parameters is also evaluated to better interpret the classification results.

III. METHODS

A. Model Summary

This study focuses on the classification of driving style variations using a temporal sequence of driving data, denoted as $x_i = (acceleration, vehicle, gap, rpm, throttle, brake, steering)_i$, where $i \in 1, 2, \dots, n$. The objective is to classify the sequence into three clusters, each represented by a centroid μ_j, where $j \in 1, 2, 3$.

The proposed methodology consists of three phases. In the first phase, individual trips are segmented into three distinct categories of driving manoeuvres: acceleration, braking, and maintaining. Manoeuvres belonging to the same category are then grouped together to create coherent assemblies. Subsequently, in the second phase, a temporal Autoencoder is trained to learn an effective latent representation for each assembly of manoeuvres. This representation has a significantly smaller dimension than the original data, which aids in the classification process. Finally, an iterative clustering technique is employed to minimize the Kullback-Leibler (KL) divergence between the probability distribution based on the centroid positions and the auxiliary target distribution. The pseudocode of the proposed method is outlined in Algorithm.1.

B. Driving Manoeuvre Detection

In this study, three categories of driving manoeuvres, namely acceleration, braking, and maintaining, are considered. These manoeuvres are primarily characterized by

Algorithm 1 Pseudocode of the proposed method

Input: dataset $\mathcal{D} = \{x_i\}$, number of desired clusters K
Output: set of clusters C

Split dataset into small time sequences $N \leftarrow \mathcal{D}$
Create assemblies of three driving manoeuvres $E \leftarrow N$
for $e \in E$ **do**
    Pre-train temporal autoencoder
    for $x_i \in e$ **do**
        Convert x_i to latent representation z_i
        Initialise cluster centroids $\mu_j, j \in K$
        while not converged **do**
            Soft assignment $C_j \leftarrow z_i$, see Eq. 4
            Compute KL divergence, see Eq. 6
            Update cluster centroids $\mu_j, j \in K$
        end while
    end for
end for

different tendencies in vehicle speed changes, such as ascending, descending, or remaining approximately constant. Therefore, vehicle speed is utilized as the key feature parameter to differentiate between the various driving manoeuvres. To facilitate manoeuvre detection, each trip is divided into a series of 1-second sequences, which are subsequently combined.

In order to capture the diverse patterns present in the temporal sequence data, dynamic time warping (DTW) is employed as the similarity metric for clustering. When comparing two driving sequences, denoted as $X = x_1, x_2, \ldots, x_n$ and $Y = y_1, y_2, \ldots, y_n$, DTW identifies the path that minimizes the Euclidean distance between the sequences as,

$$W^* = argmin\left(\sqrt{\sum_{k=1}^{K}(x_i - y_j)^2}\right) \tag{1}$$

where W^* is the optimal warping path and $K \in [n, 2n-1]$ is the number of elements in the optimal path.

The cumulative cost matrix $\gamma(i,j)$ can be computed as,

$$\begin{aligned}\gamma(i,j) &= d(x_i, y_j) + \\ &\quad min(\gamma(i-1, j-1), \gamma(i-1, j), \gamma(i, j-1))\end{aligned} \tag{2}$$

where $d(x_i, y_j)$ is the distance measurement.

Using DTW as the similarity metric, K-means clustering is subsequently employed to isolate distinct driving manoeuvres. Within each trip, driving manoeuvres belonging to the same cluster are combined. To assign labels to these manoeuvres, the average speed change within each group is utilized. Specifically, the manoeuvre with the highest speed increase is labeled as an accelerating manoeuvre, the manoeuvre with the greatest speed decrease is labeled as a braking manoeuvre, and the remaining manoeuvre is labeled as a maintaining manoeuvre.

C. Latent Representation

Following the detection of driving manoeuvres, segments representing the same type of manoeuvre are extracted and combined to form distinct driving manoeuvre groups. Each group comprises a large set of 5×7 data sequences, where 5 denotes the temporal length ($1s$) and 7 represents the dimension of input features. To learn the latent representation of each group, a customised temporal autoencoder is employed. It consists of an encoder and a decoder, both comprise a multi-layer perceptron (MLP) and two bidirectional LSTM (BiLSTM) layers. The MLP consists of three fully connected layers and is used to reduce the dimension of the feature parameters, while the two BiLSTM layers capture temporal changes in both forward and backward directions. By leveraging this architecture, the latent representation incorporates temporal information from all feature parameters at a reduced dimension, facilitating subsequent clustering tasks.

To ensure the accuracy of the latent features in representing the original data, separate autoencoders are trained for each driving manoeuvre group. The mean squared error (MSE) is employed as the loss function to evaluate the dissimilarity between the original (x_i) and reconstructed features ($\hat{x}_i$), as depicted in Equation 3. Once the predetermined MSE criterion (0.01) is met, the decoder is discarded, and only the encoder is used to map input features into the latent representation for temporal clustering.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2 \tag{3}$$

D. Temporal Clustering

After obtaining the mapping from input feature parameters to latent representation, the cluster centroids are initialized by performing K-means clustering on the latent representation, with the default number of clusters set to three. The clustering process is then trained using an unsupervised algorithm that iterates between two steps.

In the first step, a soft assignment of input x_i is computed by ranking the distances between the centroid μ_j and its corresponding latent representation z_i. The second step involves refining the cluster centroids using a loss function that maximizes the assignment confidence by employing an auxiliary target distribution. The aim is to enhance the clustering performance by encouraging high-confidence assignments.

This iterative process continues until a convergence criterion is met, indicating that the clustering algorithm has reached a stable state. Consistent with previous studies [18], [19], the similarity between the latent representation z_i and centroid μ_j is measured using the student's t-distribution kernel. In an unsupervised setting, the probability of assigning sample i to cluster j can be denoted as,

$$q_{ij} = \frac{\sum_{j'}\left(1 + \|(z_i - \mu_{j'}\|^2\right)}{1 + \|(z_i - \mu_j\|^2} \tag{4}$$

To optimize the cluster centroids, the model is trained by minimizing the KL divergence loss between q_{ij} and the target distribution p_{ij}. Following the approach suggested by Xie et al. [18], the target distribution can be computed as,

$$p_{ij} = \frac{q_{ij}^2 / f_j}{\sum_{j'} q_{ij'}^2 / f_{j'}} \tag{5}$$

where $f_j = \sum_i q_{ij}$ are soft cluster frequencies.

The KL divergence loss can be defined as,

$$L = KL(P||Q) = \sum_{i=1}^{n} \sum_{j=1}^{3} p_{ij} \log \frac{p_{ij}}{q_{ij}} \tag{6}$$

where n is the number of samples in dataset and 3 is the number of clusters.

The cluster centroids $\{\mu_j\}_{j=1}^{3}$ is optimised using Stochastic Gradient Descent (SGD). The gradient to update cluster centroid μ_j can be denoted as,

$$\begin{aligned} \frac{\partial L}{\partial \mu_j} = & -2 \sum_i \left(1 + \|z_i - \mu_j\|^2\right)^{-1} \\ & \times (p_{ij} - q_{ij})(z_i - \mu_j) \end{aligned} \tag{7}$$

E. Model Explainability

Model explainability pertains to the process of elucidating the outputs generated by machine learning models by examining how and which features impact the model's final output. Enhancing explainability in driving style classification can provide insights into the significance of different feature parameters and aid in identifying a more effective subset of features. Among the existing methods, the SHAP (SHapley Additive exPlanations) framework is a widely used local diagnostic method that quantifies the marginal contribution and importance of each feature [13]. The SHAP value, denoted as ϕ_j, for a specific feature j can be defined as,

$$\begin{aligned} \phi_j = & \frac{1}{|N|!} \sum_{S \subseteq N \setminus \{j\}} |S|!(|N| - |S| - 1)! \\ & [f(S \cup \{j\}) - f(S)] \end{aligned} \tag{8}$$

where $| \bullet |$ is the number of elements in the set, N is the original feature set, S represents any feature subset in N, $N \setminus \{j\}$ denotes a subset of all elements in the sequence before feature j. $f(S)$ is the output of the model of feature subset S, and $f(S \cup \{j\}) - f(S)$ is the cumulative contribution value of feature j.

However, it is important to note that the SHAP framework can only be directly applied to supervised models. Hence, in this study, a LSTM classifier [11] is employed to train and fit the clustering labels. Subsequently, the trained classifier is utilized to compute the SHAP values, enabling the assessment of feature importance and contributions within the clustering process.

IV. Results and Discussion

The proposed model was implemented using PyTorch. The training and evaluation processes were conducted on a Desktop PC with the following specifications: CPU - AMD Ryzen 9 5900X 12-Core @ 3.70GHz, GPU - NVIDIA GeForce RTX 3070, and RAM - 32 GB.

For training the autoencoder, an ADAM optimiser was employed. In contrast, the clustering model was trained using the SGD optimiser. A learning rate of 0.01 and a batch size of 128 are used for both optimisations. These configurations were selected to optimize the training process and enhance the model's performance.

A. Data Collection

A total of 25 experienced drivers (13 males and 12 females) participated in the data collection. The participants had an average age of 25.1±3.4 years.

To capture driving behavior, three distinct car-following scenarios, each lasting 8 minutes, were created using a bespoke driving simulation software. The software was developed using the Unity3D game engine (version 2019.4.f1) and deployed on an XPI DS1 driving simulator, which provided the necessary hardware devices (surrounding screens, steering wheel with force feedback, pedals, gear shifter, etc.) for an immersive driving experience.

During the simulation, each participant was instructed to activate a simulated vehicle from a predefined starting position and drive in the simulated environment based on their usual driving habits. The driving simulator's sensors recorded various relevant information, including time, position, direction, speed, headway distance from the leading vehicle, engine rpm, throttle and brake position, steering wheel rotation, and speed of the leading vehicle. The collected driving data was sampled at a fixed frequency of 5 Hz and exported for subsequent post-processing and analysis. To ensure consistency, all feature parameters were normalized to the range of $[0, 1]$ per category for further analysis and usage in the proposed method.

B. Driving Manoeuvre Detection

The proposed driving manoeuvre detection approach relies on the distinct changing tendencies in vehicle speed. An example of the segmented results for a sample trip is depicted in Fig. 1.

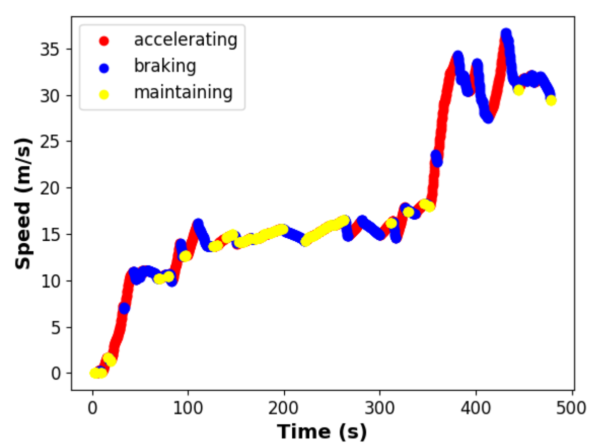

Fig. 1. Manoeuvre detection of a sample trip

The maximum and average vehicle speeds observed in this trip are 132 km/h and 64.5 km/h, respectively. Meanwhile, the maximum acceleration and deceleration values recorded are 1.42 m/s^2 and -1.87 m/s^2. The segmented trip comprises 244 seconds of accelerating manoeuvres, 169 seconds of braking manoeuvres, and 64 seconds of maintaining manoeuvres. After merging adjacent driving manoeuvres of the same type, the total accumulated numbers for each manoeuvre category are as follows: 52 for accelerating, 31 for braking, and 39 for maintaining. These numbers represent the combined occurrences of the respective driving manoeuvres observed in the sample trip.

From Fig. 1, it can be observed that all three driving manoeuvres are evenly distributed throughout the trip, with minimal abrupt changes between manoeuvres. However, a few abrupt changes are noticed, which are primarily attributed to maintaining a safe headway distance with the leading vehicle or adjusting to the curvature of the road. These observations were made during the experiment and highlight the real-world scenarios encountered during driving.

C. Driving Style Classification

As the proposed method evaluates driving style in small time sequences, it is difficult to acquire ground truth labels for evaluation. Therefore, the statistics of the clustering results (Table. II) are used for performance evaluation.

To label each cluster, the distinctive behaviors associated with specific driving styles, such as risky speeding and abrupt speed changes for aggressive driving, were considered. Furthermore, the ordering of clusters was rearranged to enhance visual interpretation. As a result, clusters 1, 2, and 3 align more closely with aggressive, moderate, and conservative driving, respectively. A minor improvement in Silhouette Coefficient is obtained with the proposed method (0.642 compared to 0.608 for K-Shape). Meanwhile, the proposed method also exhibited more representative behaviors compared to K-Shape. For instance, in the accelerating maneuver, cluster 1 (representing aggressive driving) of the proposed method exhibited the highest values across all features except brake, whereas the K-Shape cluster only demonstrated the highest values in velocity and throttle. Similar trends were observed in other maneuvers as well. Statistics of the proposed method are more aligned with the representative behaviours of each driving style. These findings indicate that the driving style clusters obtained by the proposed method outperform those obtained by K-Shape, as they capture more nuanced and representative driving behaviors.

To illustrate the driving style variation, the classification result of a selected participant is shown in Fig. 2.

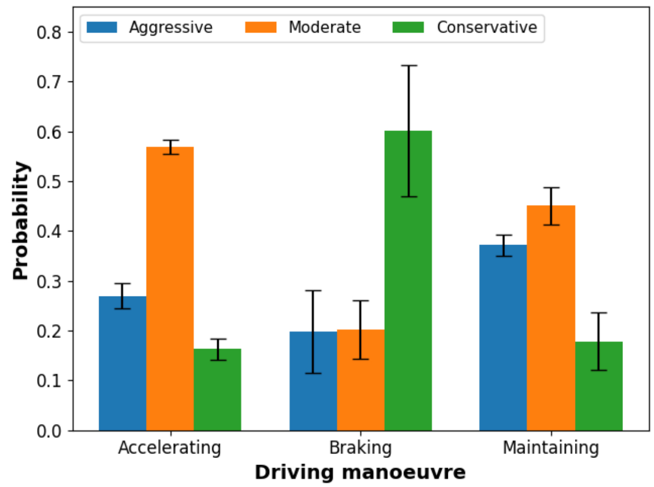

Fig. 2. Classification result of a sample driver

The results are the average of three trips, with the error bars denoting the standard deviation. It is evident that the driving style significantly varies across different manoeuvres. Specifically, this participant demonstrates a moderate driving style in both accelerating and maintaining manoeuvres, while adopting a more conservative approach during braking events. Furthermore, as shown with the standard deviation, the consistency of the driving style is apparent within each manoeuvre across the different trips, indicating that this participant maintains a consistent driving style. These findings emphasize the importance of distinguishing driving style

TABLE II

PERFORMANCE EVALUATION

Manoeuvre	**Features**	**Proposed Method**			**K-Shape**		
		Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 1**	**Cluster 2**	**Cluster 3**
Accelerating	Velocity (m/s)	20.3±9.4	19.4±10.6	19.2±10.7	20.7±9.8	15.2±8.9	11.4±8.8
	Acceleration (m/s^2)	1.0±0.8	0.8±0.7	0.7±0.6	0.8±0.7	1.3±1.2	1.1±1.1
	RPM (r/min)	3596±1675	3588±1757	3435±1618	3665±1709	3445±1722	3789±1479
	Throttle (%)	72.4±25.2	66.8±21.1	65.6±19.2	70.5±22.8	69.0±21.8	57.5±23.0
	Brake (%)	0.1±2.3	0.1±1.7	0.2±1.5	0.1±2.0	0.1±1.3	0.3±2.2
Braking	Velocity (m/s)	20.8±10.2	18.8±10.4	18.2±10.1	20.0±9.5	17.6±9.9	19.4±10.3
	Acceleration (m/s^2)	-2.1±3.4	-2.1±2.5	-1.2±1.4	-4.9±3.9	-3.4±3.2	-1.4±2.3
	RPM (r/min)	3972±2338	3228±1688	3048±1615	3650±1704	3205±1911	3453±2011
	Throttle (%)	31.1±32.1	10.9±14.2	10.2±14.7	8.8±17.3	15.5±21.7	19.4±25.6
	Brake (%)	16.4±28.8	12.7±19.5	3.4±9.6	36.3±34.1	26.9±28.7	7.6±17.9
Maintaining	Velocity (m/s)	26.5±7.8	25.9±6.6	10.7±10.1	22.0±8.8	19.1±11.3	14.2±12.0
	Acceleration (m/s^2)	-0.1±0.6	-0.2±0.9	-0.4±1.5	-0.3±0.9	-0.9±2.3	-0.2±0.9
	RPM (r/min)	8268±729	6162±1077	1970±1477	5815±2741	5084±2941	3360±3005
	Throttle (%)	72.9±21.3	43.7±12.2	22.1±20.4	49.5±28.9	43.8±30.6	33.6±29.5
	Brake (%)	0.9±3.6	1.5±6.2	5.2±14.8	2.9±12.2	8.8±20.7	2.7±9.3

cyan, orange and green colors denote largest, medium and smallest values, respectively.

classification based on specific manoeuvres, as drivers can consistently exhibit different driving styles across distinct manoeuvres.

D. Result Interpretation

To improve the comprehension of the findings, an assessment was conducted to determine the contribution of feature parameters to the clustering outcomes. Specifically, focusing on the clustering results pertaining to accelerating manoeuvres, the SHAP values were visualized in Fig. 3. This visualisation combines the feature importance with their corresponding effects on the clustering results. The features are ranked based on their significance in influencing the outcomes. Notably, the analysis reveals that throttle and engine RPM emerge as the most influential features in this manoeuvre, whereas brake exhibits minimal impact. Moreover, most of the features (throttle, RPM, velocity, and acceleration) positively contribute to the clustering results, aligning with established knowledge and affirming the validity of the clustering outcomes. Additionally, the sequential order of feature parameters can guide future studies in identifying a more representative subset of features for driving style classification.

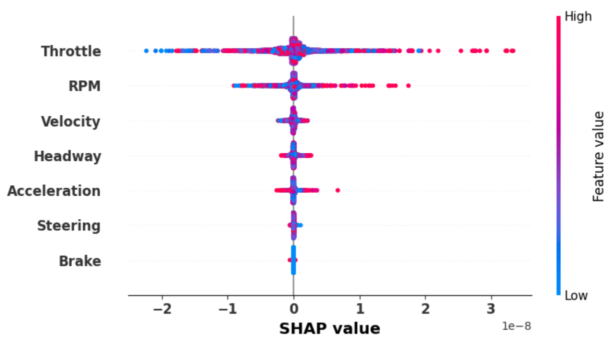

Fig. 3. SHAP summary

V. Conclusion

A novel approach for classifying driving styles with temporal features was proposed. Our model emphasises variations in individual's driving style in a single trip. The experimental data were collected using three bespoke car-following scenarios with 25 participants.

This paper's primary contribution is the introduction of a deep temporal clustering method for driving style classification. Instead of considering each individual's driving style as a static category, the proposed approach highlights the diversities observed in various driving maneuvers. Furthermore, an XAI method is incorporated to evaluate the the influence of feature parameters on the results and improve explainability of the proposed method. As part of future research, ablation studies will be conducted to explore the impact of various similarity metrics on enhancing the performance of the model.

Moreover, it is important to acknowledge that this investigation focuses on data obtained from driving simulation and primarily addresses factors pertaining to the driver's viewpoint. Future research endeavors will encompass real-world datasets and a broader range of external influential variables, such as diverse road conditions, weather conditions, and more intricate scenarios involving various types of road agents.

References

[1] A. Rezaei and B. Caulfield, "Safety of autonomous vehicles: what are the insights from experienced industry professionals?" *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 81, pp. 472–489, 2021.

[2] J. Elander, R. West, and D. French, "Behavioral correlates of individual differences in road-traffic crash risk: An examination of methods and findings," *Psychological bulletin*, vol. 113, pp. 279–94, 04 1993.

[3] Y. Feng, S. Pickering, E. Chappell, P. Iravani, and C. Brace, "Driving style analysis by classifying real-world data with support vector clustering," in *2018 3rd IEEE International Conference on Intelligent Transportation Engineering (ICITE)*, 2018, pp. 264–268.

[4] 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. 30 653–30 682, 2015.

[5] A. Aljaafreh, N. Alshabatat, and M. S. Najim Al-Din, "Driving style recognition using fuzzy logic," in *2012 IEEE International Conference on Vehicular Electronics and Safety (ICVES 2012)*, 2012, pp. 460–463.

[6] M. Najim Al-Din, A. Aljaafreh, N. Albdour, and M. Saleh, "Driving styles recognition using decomposed fuzzy logic system," *International Journal of Electrical, Electronics and Computer Systems*, vol. 16, pp. 820–824, 09 2013.

[7] D. Dörr, D. Grabengiesser, and F. Gauterin, "Online driving style recognition using fuzzy logic," in *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 2014, pp. 1021–1026.

[8] Z. Constantinescu, C. Marinoiu, and M. Vladoiu, "Driving style analysis using data mining techniques," *Int. J. Comput. Commun. Control*, vol. 5, pp. 654–663, 2010.

[9] V. Vaitkus, P. Lengvenis, and G. Žylius, "Driving style classification using long-term accelerometer information," in *2014 19th International Conference on Methods and Models in Automation and Robotics (MMAR)*, 2014, pp. 641–644.

[10] A. Albers and M. Albrecht, "Method using a self-organizing map for driver classification as a precondition for customer oriented design," in *Proceedings of the 15th International Conference on Engineering Design ICED 05*, 2005.

[11] G. Li, F. Zhu, X. Qu, B. Cheng, S. Li, and P. Green, "Driving style classification based on driving operational pictures," *IEEE Access*, vol. 7, pp. 90 180–90 189, 2019.

[12] X. Peng, Y. Li, I. W. Tsang, H. Zhu, J. Lv, and J. T. Zhou, "Xai beyond classification: Interpretable neural clustering," *Journal of Machine Learning Research*, vol. 23, no. 6, pp. 1–28, 2022.

[13] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS'17. Red Hook, NY, USA: Curran Associates Inc., 2017, p. 4768–4777.

[14] W. Wang, J. Xi, A. Chong, and L. Li, "Driving style classification using a semisupervised support vector machine," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 5, pp. 650–660, 2017.

[15] Y. Guo, X. Wang, Y. Huang, and L. Xu, "Collaborative driving style classification method enabled by majority voting ensemble learning for enhancing classification performance," *PLOS ONE*, vol. 16, no. 7, pp. 1–21, 07 2021.

[16] X. Zheng, P. Yang, D. Duan, X. Cheng, and L. Yang, "Real-time driving style classification based on short-term observations," *IET Communications*, vol. 16, no. 12, pp. 1393–1402, 2022.

[17] X.-S. Li, X.-T. Cui, Y.-Y. Ren, and X.-L. Zheng, "Unsupervised driving style analysis based on driving maneuver intensity," *IEEE Access*, vol. 10, pp. 48 160–48 178, 2022.

[18] J. Xie, R. Girshick, and A. Farhadi, "Unsupervised deep embedding for clustering analysis," in *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*, ser. ICML'16. JMLR.org, 2016, p. 478–487.

[19] N. S. Madiraju, S. M. Sadat, D. Fisher, and H. Karimabadi, "Deep temporal clustering : Fully unsupervised learning of time-domain features," 2018.