2025 International Conference on Advanced Machine Learning and Data Science (AMLDS) | 979-8-3315-1099-2/25/$31.00 ©2025 IEEE | DOI: 10.1109/AMLDS63918.2025.11159370

Integration of Explainable Artificial Intelligence and Multimodal Machine Learning for Drivers' Fitness

Mobyen Uddin Ahmed*, Arnab Barua, Mir Riyanul Islam, Shaibal Barua, Shahina Begum
School of Innovation, Design and Engineering, Mälardalen University, Västerås, Sweden
Email: {mobyen.uddin.ahmed, arnab.barua, mir.riyanul.islam, shaibal.barua, shahina.begum}@mdu.se
*Corresponding Author, Email: mobyen.uddin.ahmed@mdu.se, Phone: +46 21-10 73 69

***Abstract*—In order to fully utilize the data collected from different sources and transform them into valuable assets for the prediction and explanation of Artificial Intelligence (AI) systems, this paper introduces an approach that combines Multimodal Machine Learning (MML) with Explainable AI (XAI). The goal is to provide insights into a deeper understanding of driver performance in terms of mental fatigue in drivers. The performances of the drivers can be assessed primarily based on their mental fatigue levels. Detecting mental fatigue using MML with explainability remains a challenge, especially when heterogeneous data are collected in an unsupervised manner. This paper includes multiple modalities in primary prediction and explanation tasks, enabling Multimodal XAI (MXAI). The work used vehicular telemetry data collected from multiple driving scenarios in Spain and Italy, providing a rich multi-source dataset for MML model development. Here, MML integrates information fusion, co-learning, and reasoning to analyse multivariate unlabelled data for fatigue detection. It applied k-means clustering on these unlabelled data, followed by classification using Random Forest and XGBoost, effectively creating a semi-supervised learning approach. In this study, XAI is used to enhance the transparency and interpretability of the MML model. Here, the contribution of various parameters to fatigue classification was examined using SHAP–Shapley Additive Explanations. Hence, the work contributes to driver fitness using MML to improve model accuracy and robustness, as well as XAI for model interpretability and transparency in detecting fatigue-related patterns.**

***Index Terms*—Explainable AI, Multi-modal XAI, Multi-modal Machine Learning, Mental Fatigue.**

I. Introduction

Research indicates that drivers' fatigue can be assessed using multiple modalities, including vehicular telemetry and neurophysiological data [1]–[3]. However, labelling such data is challenging due to variations in driving behaviour, environmental conditions, and traffic dynamics [4], [5]. In the FitDrive[1] project, multimodal data, including drivers' neurophysiological signals, vehicular telemetry, and contextual information, are utilised for driver fitness, including mental fatigue classification using Multimodal Machine Learning (MML).

This study was supported by the following projects: 1) FitDrive, funded from the European Union's (EU) Horizon 2022 Research and Innovation programme, Grant Agreement No. 953432; 2) TRUSTY, financed by SESAR JU under the EU's Horizon 2022 Research and Innovation programme, Grant Agreement No. 101114838; 3) CPMXai, funded by the VINNOVA, Diary No. 2021-03679.

[1]https://www.fitdrive.eu/

MML has advanced rapidly, enabling the integration of diverse data sources like text, images, audio, and video [6], [7]. However, one of the major challenges in MML is the need for labelled data across all modalities [8]. In our previous studies ([3], [7]), MML demonstrated its effectiveness in fusing diverse data types to achieve accurate classification. By integrating multiple modalities, MML enhances model performance by using complementary information from different sources. The integration of Explainable AI (XAI) with MML presents a promising approach known as Multimodal XAI (MXAI). The authors in a survey [9] explore the evolution of MXAI across four areas—traditional Machine Learning (ML), deep learning, foundation models, and generative Large Language Models (LLM)—shedding light on its challenges and the path toward more transparent and trustworthy AI. Again, authors in [10], examine from a clinical standpoint, the challenges of XAI for multimodal and longitudinal datasets. The author in [11], uses multimodal data and interpretable ML to predict stress levels, highlighting the superiority of ensemble models and the role of Shapley Additive Explanations (SHAP) [12] based explainability in enhancing transparency for clinical decision-making.

In this paper, multiple modalities are first utilised by the primary prediction model for decision-making. Subsequently, these same modalities are leveraged to generate explanations for the model's behaviour, ensuring greater interpretability and transparency in multimodal AI systems [6]. It develops a comprehensive model for assessing driver performance concerning fatigue effects by using multivariate data analytics, MML and MXAI. The proposed model incorporates heterogeneous data sources, including biomedical signals, in-vehicle metrics, and contextual driving information, to provide a more holistic and reliable approach to fatigue detection. It includes k-means clustering and ensemble learning techniques. Then, it compared Extreme Gradient Boosting (XGBoost) with Random Forest (RF) to achieve high accuracy in fatigue classification. Additionally, SHAP [12] was employed for an in-depth interpretability analysis, revealing the contribution of various parameters to fatigue classification. This explainability analysis provides critical insights into the most significant indicators of driver fatigue, offering a transparent and interpretable framework.

II. Materials

The primary objective of the experiment was conducted through two studies: 1) Cycle 1 (C1) using simulator-road driving and 2) Cycle 2 (C2) using real-road driving, focused on gathering behavioural and neurophysiological data to study fatigue while driving. According to the existing literature [13], [14], there are specific times of the day when the likelihood of experiencing fatigue is higher. Driver fatigue annotation in the alignment process considered the results of neurophysiological data analysis. The complementary neurophysiological dataset, including its characteristics, signal modalities, and preprocessing pipeline, was developed and reported by another project partner in a previous paper [15]

TABLE I
LIST OF THE SELECTED PARAMETERS IN CYCLE 1.

Sl	Parameters	Descriptions	Unit
1	Yaw Rate Ext Sns	Speed of the yaw in the Y axis.	deg/s
2	Speed Lateral	Velocity along the vehicle's X axis.	m/s
3	Acceleration Lateral	Acceleration along the vehicle's X axis.	m/s^2
4	Steering Wheel Angle	The angle of the vehicle's steering wheel.	deg
5	Vehicle Velocity	The velocity of the vehicle.	m/s
6	Speed Forward	Velocity along the vehicle's Z axis.	m/s
7	Vert Accel Ext Sns	Acceleration along the world's Y axis.	m/s^2
8	Pitch Ext Sns	Angle of the pitch in X axis.	deg
9	Pitch Rate Ext Sns	Speed of the pitch in the X axis.	deg/s
10	Vert Vel Ext Sns	Velocity along the world's Y axis.	m/s
11	Roll Rate Ext Sns	The roll rate of a vehicle around its longitudinal axis.	deg/s
12	Acceleration Forward	Acceleration along the vehicle's Z axis.	m/s^2
13	Acceleration Pedal Position	Position of the acceleration pedal.	%
14	Long Vel Ext Sns	The longitudinal velocity of a vehicle.	m/s
15	Lat Vel Ext Sns	The lateral (sideways) velocity of a vehicle is the speed at which the vehicle moves to the left or right relative to its forward direction.	m/s
16	Yaw Ext Sns	The rotation of the vehicle around its vertical axis.	deg/s
17	Roll Ext Sns	The rotation around the vehicle's longitudinal axis.	deg/s
18	Lat Accel Ext Sns	The rate of velocity changes in a direction perpendicular to the direction of travel.	m/s^2
19	Long Accel Ext Sns	The rate of velocity changes along the direction of the vehicle's travel.	m/s^2

A. Cycle 1 Dataset

The C1 dataset was collected using a simulator with a simulated environment where two driving routes were used, one based on roads in Rome, Italy (C1.1), and the other on roads in León, Spain (C1.2). After screening based on selection criteria, 34 participants were finalised, with 17 participants from each country. During the C1 experiment, participants were instructed to drive in both challenging and monotonous environments. The dataset includes vehicular telemetry data that is 48 signals in total, where the signals containing signals like GPS, Timestamp, Inertial Measurement Unit (IMU) sensory data to measure Accelerometer – Measures linear acceleration (*e.g.*, changes in speed or direction) and Gyroscope – Measures angular velocity (rotation rates around different axes)., and other environment-related signals. Out of 48 signals, 19 were selected for further scrutiny. Twenty-five signals with binary values (zero and one) were omitted to prevent overfitting. Timestamps were used for sorting and filtering data chronologically and identifying anomalies or outliers, but were not utilised for analysis. Additionally, GPS-related signals were excluded to improve the model's overall applicability. The list of the 19 selected signals, along with their definitions, is presented in Table I.

B. Cycle 2 Dataset

The C2 data was collected from real-time driving. Like C1, C2 data were obtained from two experiments: one from Rome, Italy (C2.1), and another from León, Spain (C2.2). In C2.1, 19 signals were collected from 9 participants. Out of these 19 signals, 11 were selected based on their relevance in detecting fatigue. For example, changes in acceleration and speed often reflect variations in driver alertness levels, while fluctuations in engine load and throttle position can indicate inconsistent vehicle control. These specific patterns are crucial for accurately identifying signs of driver fatigue. Table II presents a list of the 11 selected signals, along with their definitions. In the C2.2 experiments, 12 participants took part, and data on available signals were collected from real driving. Like C2.1, the relevance of collected signals is analysed, and 15 signals related to fatigue were considered during the analyses. The list of signals is provided in Table III.

TABLE II
LIST OF THE SELECTED PARAMETERS IN CYCLE 2.1.

Sl	Parameter	Description	Unit
1	acc_x	Acceleration along the x-axis.	m/s^2
2	acc_y	Acceleration along the y-axis.	m/s^2
3	acc_z	Acceleration along the z-axis.	m/s^2
4	accelerator_pos_d	Position of the accelerator pedal.	%
5	accelerator_pos_e	Another sensor's reading of the accelerator pedal position.	%
6	engine_load	The current engine load compared to the maximum possible load.	%
7	pos_lat	Latitude position of the vehicle.	deg
8	pos_long	Longitude position of the vehicle.	deg
9	rpm	Revolutions per minute (RPM) of the engine	–
10	speed	Speed of the vehicle.	km/h
11	throttle_pos	Position of the throttle.	%

III. Approach and Methods

The structured approach and methodology presented in Fig. 1 provide a comprehensive framework for handling unlabelled

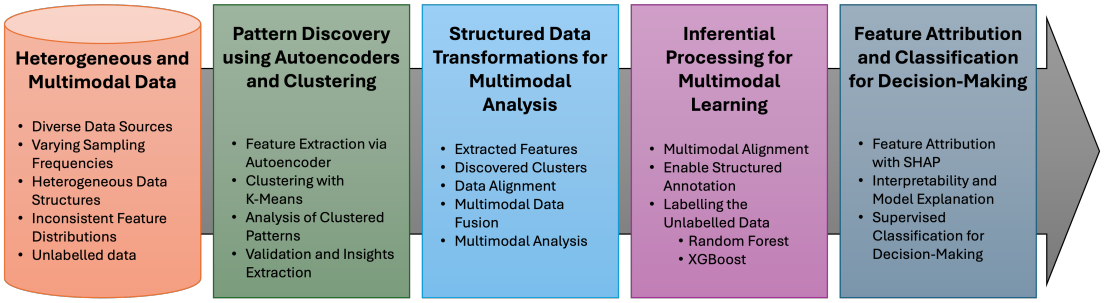

Fig. 1. Overall approach of integrating XAI and MML.

TABLE III
LIST OF THE SELECTED PARAMETERS IN CYCLE 2.2.

Sl	Parameters	Descriptions	Unit
1	Wheel-Based Vehicle Speed	The speed of the vehicle based on wheel rotation.	km/h
2	Brake Switch	Indicates the pressed status (on/off) of the brake pedal.	–
3	Clutch Switch	Indicates the pressed status (on/off) of the clutch pedal.	–
4	Actual Engine Percent Torque	Engine torque output compared to the maximum possible torque.	%
5	Steering Wheel Angle	The angle of the vehicle's steering wheel.	deg
6	Lateral Acceleration	The vehicle's acceleration in the lateral direction.	m/s^2
7	Tachograph Vehicle Speed	The vehicle speed recorded by the tachograph.	km/h
8	Accelerator Pedal Position 1	The position of the accelerator pedal.	%
9	Engine Percent Load At Current Speed	Engine load at the current speed compared to the maximum possible load.	%
10	Brake Pedal Position	Position of the brake pedal with respect to its total possible travel.	%
11	Position Longitude	The longitude position of the vehicle.	deg
12	Position Latitude	The latitude position of the vehicle.	deg
13	Yaw Rate	The rotation of the vehicle around its vertical axis.	deg/s
14	Engine Speed	Revolutions per minute (RPM) of the engine.	–
15	Longitudinal Acceleration	The vehicle's acceleration in the longitudinal direction.	m/s^2

multimodal data. By integrating feature extraction, clustering, multimodal alignment, data fusion, and explainable classification, the approach ensures effective data transformation, improved interpretability, and robust decision-making.

Step 1. Heterogeneous and Multimodal Data: The datasets in this study exhibit multimodal characteristics due to diverse data sources, varying sampling frequencies, heterogeneous data structures, and inconsistent feature distributions. The vehicular data sets (C1.1, C1.2, C2.1, and C2.2) capture various aspects of vehicle dynamics, introducing challenges related to multimodality, including differences in sources, frequencies, structures, and distributions. Additionally, these datasets lack annotations, making interpretation more complex. A neurophysiological dataset serves as a supporting modality, providing insights into driver states. The combination of vehicle-related parameters and neurophysiological signals forms a complex multimodal dataset, where each modality presents unique challenges in terms of alignment, fusion, and analysis.

Step 2. Pattern Discovery using Autoencoders and Clustering: Since the vehicular dataset lacks predefined labels, unsupervised learning techniques are applied to uncover meaningful patterns. The process begins with feature extraction using an autoencoder, by training the autoencoder to reconstruct input data, it captures essential patterns while reducing redundancy, resulting in a lower-dimensional representation that preserves relevant information for further analysis. Once the features are extracted, k-means clustering is applied to identify potential behavioural patterns within the dataset. This method groups similar instances, allowing for the exploration of whether natural clusters correspond to vehicle dynamics patterns. To ensure meaningful segmentation, validation and insights extraction are performed, examining the consistency and coherence of the clusters. The combination of autoencoder-based feature extraction and k-means clustering provides an effective approach for analysing the dataset without requiring labelled information.

Step 3. Structured Data Transformations for Multimodal Analysis: Intermediate conclusions from data processing result in structured outputs that contribute to the final analysis. These structured transformations include extracted features that capture essential patterns while removing redundancy, clustered data that groups features into meaningful patterns, aligned data that standardises datasets for cross-modal comparisons, and fused data that merges multimodal datasets to enhance contextual understanding. Feature extraction condenses the original data into a more structured form, ensuring that important relationships among parameters are retained. Clustering helps reveal potential behavioural patterns, while data alignment ensures that vehicular and neurophysiological data are synchronised for meaningful comparisons. The fusion of multimodal datasets facilitates deeper pattern recognition and provides a comprehensive analysis of relationships between vehicle behaviour and external influences. By transforming raw, heterogeneous data into structured and meaningful outputs, this step lays the foundation for data-driven decision-making and system optimisation. The multimodal-alignment

procedure with details is presented in our previous paper [3].

Step 4. Inferential Data Processing for Multimodal Learning: Inferential processing involves representation techniques that transform raw data into structured insights. The first process, multimodal alignment, establishes correspondences between heterogeneous datasets such as vehicular and neurophysiological data [6]. Once the datasets are aligned, data fusion is applied to merge structured information from different sources. Instead of treating each dataset separately, this approach integrates multiple perspectives, ensuring a more comprehensive understanding of vehicle behaviour. The final process involves labelling unlabelled data using machine learning models. Since the vehicular dataset lacks ground truth labels, supervised learning approaches such as RF and XGBoost are applied to infer labels based on learned patterns. For the RF and XGBoost, the study relied on widely accepted "off-the-shelf" configurations; the details are in our previous paper [3]. These models leverage labelled subsets to classify new, unseen samples, generating structured annotations that enable future supervised learning applications. By applying trained models to derive new insights, this step refines the dataset into a more meaningful form for downstream analysis.

Step 5. Feature Attribution-based Explanation and Classification for Decision-Making: The final stage involves two critical processes: generating Shapley values for feature attribution using SHAP and using labelled data for classification and decision-making. To enhance model interpretability, the SHAP technique is applied, which quantifies the influence of individual features on model predictions. SHAP is grounded in cooperative game theory, ensuring a fair distribution of feature importance across all possible input variations. This method improves model transparency, allowing for a better understanding of which features drive predictions and whether the model relies on meaningful patterns. Once labelled data is available, it is used to train supervised classification models, such as RF and XGBoost, to distinguish different behavioural states within the dataset. The structured annotations generated in previous steps provide a solid foundation for optimising model performance in fatigue detection and vehicle behaviour analysis. Finally, all processed data—including extracted features, clustered patterns, and fused multimodal data—are integrated into a final classification pipeline. This ensures that all inferential steps contribute to a structured and well-prepared dataset, enabling the development of a robust predictive model for practical deployment and real-world applications.

IV. Experimental Results

For the experimental work to detect driver mental fatigue, first, an unsupervised ML algorithm, *i.e.*, the k-means clustering algorithm, is considered, then two supervised ML algorithms, such as RF and XGBoost, are considered. Finally, it explores the contributing parameters using SHAP for fatigue classification.

A. Pattern Discovery using Clustering

The optimal number of clusters, denoted by k, was determined through the Elbow method, which suggested that $k = 3$ provided the best balance between within-cluster variance and the number of clusters. The results were visualised using t-SNE (t-distributed Stochastic Neighbours Embedding), a widely used technique for reducing the dimensionality of high-dimensional data to create a two-dimensional scatter plot. Fig. 2 and 3 display the t-SNE visualisations for the C1.1 and C1.2 datasets, respectively. According to the figures it is suggested that while both datasets exhibit clear cluster separations, there are notable variations in the underlying data patterns, which may be due to differences in the features or characteristics captured in each subset.

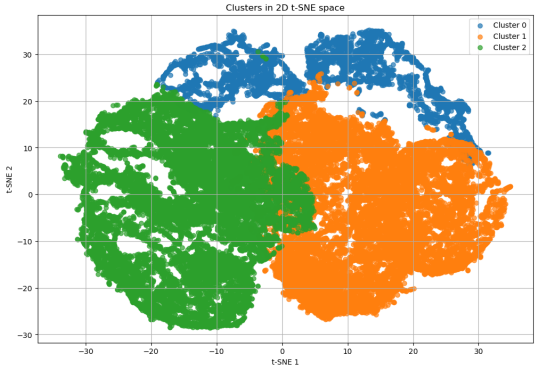

Fig. 2. Clusters in 2D t-SNE space built using extracted features of C1.1.

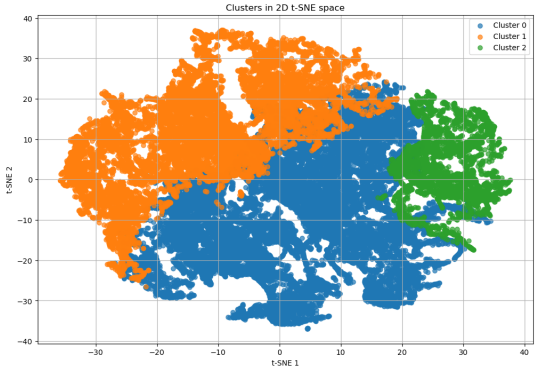

Fig. 3. Clusters in 2D t-SNE space built using extracted features of C1.2.

Similarly, Fig. 4 and 5 present the t-SNE visualisations for the C2.1 and C2.2 datasets. As with the previous datasets, both C2.1 and C2.2 exhibit well-defined cluster separations. However, there are observable differences in the cluster distributions between C2.1 and C2.2, indicating that these two datasets also contain distinct data patterns. These variations suggest that the data in C2.1 and C2.2 might have been influenced by different factors or characteristics, which are reflected in the way the clusters are distributed. In summary, the t-SNE visualisations for all four datasets (C1.1, C1.2, C2.1, and C2.2) reveal clear cluster separations, but the observed differences in the cluster distributions between corresponding datasets (*e.g.*, C1.1 vs. C1.2 and C2.1 vs. C2.2) highlight the inherent variations in the data patterns.

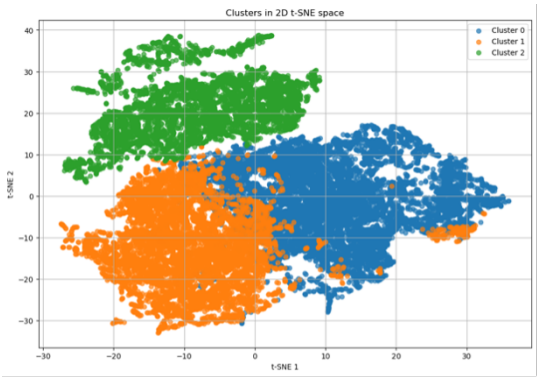

Fig. 4. Clusters in 2D t-SNE space built using extracted features of C2.1.

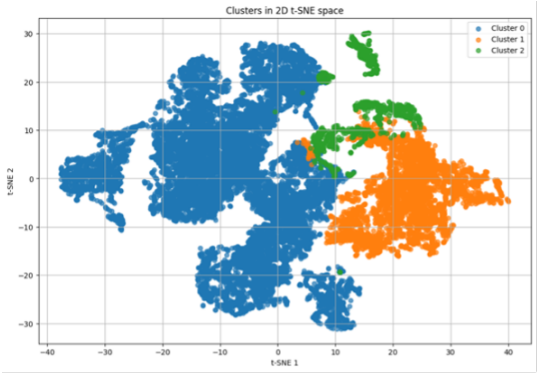

Fig. 5. Clusters in 2D t-SNE space built using extracted features of C2.2.

B. Multimodal Learning and Classification

The dataset used in this study consists of multiple subsets, with each subset denoted by a specific identifier (C1, C2.1, and C2.2). To conduct the multimodal learning and classify the labelled dataset, we consider 5055 samples out of 85975 samples, and the details can be found in our previous paper [3]. Here, the process of assigning these labels can be found in the following article [16]. Thus, the final training dataset was obtained considering the vehicular parameters and labels obtained from the process mentioned above. Each of these subsets was analysed using two different ML algorithms—RF and XGBoost to evaluate their predictive performance in terms of accuracy and F_1 score. The data was split into a train and test set with an 80% and 20% distribution, respectively. When splitting the dataset into train and test, a chronological approach was used due to the related time stamps of each sample, which were dropped during analysis. The results of the classification are presented in Table IV. For the combined C1 dataset (which merges C1.1 and C1.2), both machine learning models performed excellently. RF achieved an accuracy of 98%, with a corresponding F_1 score of 0.98, indicating high precision and recall in the classification task. XGBoost, although slightly lower in performance, still demonstrated strong results with a 97% accuracy and an F_1 score of 0.97.

For the C2.1 dataset, which consists of a larger number of samples (3,645), both models continued to perform well, though there was a slight shift in performance. The XGBoost model achieved the highest test accuracy at 98%, along with an F_1 score of 0.98, making it the most effective model for this dataset. On the other hand, RF also performed admirably, achieving a test accuracy of 97% and an F_1 score of 0.97. This indicates that while XGBoost was marginally better in classification performance, RF still provided reliable results. The C2.2 dataset showed a notable difference in performance compared to the previous datasets. RF achieved a relatively lower accuracy of 80%, with an F_1 score of 0.80, indicating some challenges in correctly identifying the target class. However, XGBoost demonstrated a better performance for this dataset, with a higher test accuracy of 86% and an F_1 score of 0.86.

TABLE IV
CLASSIFICATION RESULTS OF C1 AND C2 DATASETS.

Dataset	Number of Samples	Method	Test Accuracy	F_1 score
C1 (combined C1.1 and C1.2)	1011	RF	**98%**	**0.98**
		XGBoost	97%	0.97
C2.1	3645	RF	97%	0.97
		XGBoost	**98%**	**0.98**
C2.2	3240	RF	80%	0.80
		XGBoost	**86%**	**0.86**

TABLE V
SUMMARY OF PARAMETER RANKING BY XGBOOST USING SHAP.

Sl	Parameter	Ranking on C1.1 Dataset	Ranking on C1.2 Dataset
1	Accelerator Pedal Position	1	3
2	LatAccelExtSns	8	1
3	LatVelExtSns	4	5
4	LongVelExtSns	3	4
5	RollExtSns	–	8
6	Speed Lateral	6	7
7	Steering Wheel Angle	5	9
8	Vehicle Velocity	7	–
9	YawExtSns	2	6
10	YawRateExtSns	9	2

C. Feature Attribution based Explanation

To assess the contributions of different parameters in fatigue classification, SHAP [12] was applied to the trained XGBoost model across all experimental cycles, including C1.1, C1.2, C2.1, and C2.2. Specifically, the SHAP values, the comparison tables, the Beeswarm summary plot, the Cohort bar plot, and the Force plots were used to illustrate the impact of each parameter. It's important to note that the SHAP values do not quantify fatigue directly but instead describe the influence of each parameter on the trained classifier's inference mechanism, in this case, the XGBoost model. Table V presents the comparison for the C1.1 and C1.2 datasets, respectively, allowing for a comparison of how feature contributions differ across these subsets. The ranking of parameters across the C1.1 and C1.2 datasets reveals key differences in their influence on the classification process. In the C1.1 dataset, the most

significant parameter is *Acceleration Pedal Position*, followed by *YawExtSns* and *LongVelExtSns*, indicating that acceleration and yaw-related features play a critical role in fatigue classification. In contrast, in the C1.2 dataset, the most influential parameter is *LatAccelExtSns*, while *YawRateExtSns* and *Acceleration Pedal Position* rank second and third, respectively. Comparing both datasets, *YawExtSns* remains an essential factor in both cases, albeit with a lower ranking in C1.2. Additionally, *LatAccelExtSns*, which ranks eighth in C1.1, emerges as the most significant feature in C1.2, suggesting differences in dataset characteristics. *RollExtSns* and *Vehicle Velocity* are absent (presented by "-" in the Table V) in one of the datasets, further indicating that specific features may be dataset dependent.

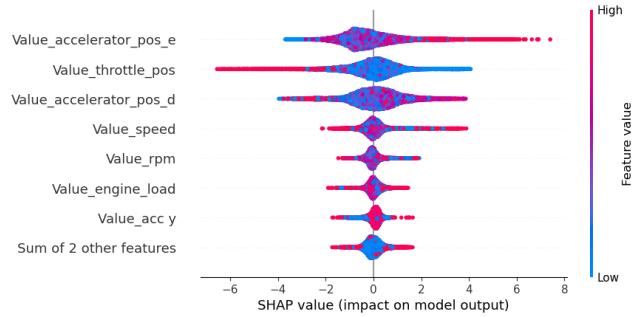

Fig. 6. Beeswarm summary plot from SHAP presenting the impact of top parameters on the fatigue classification by the XGBoost model with C2.1.

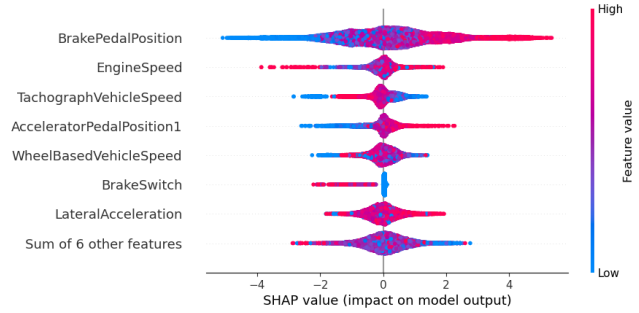

Fig. 7. Beeswarm summary plot from SHAP presenting the impact of top parameters on the fatigue classification by the XGBoost model with C2.2.

An analysis using Beeswarm plots was conducted for the datasets C2.1 and C2.2, with the results presented in the respective sub-figures of Fig. 6 and 7. Each dot in the Beeswarm plot represents a single sample for a specific feature, with its horizontal position reflecting the SHAP value for that feature. Dots cluster along each row to represent the density of values, and colour is used to show the original value of each feature, providing additional context to the parameter's role in the classification. For C2.1, *Value_accelerator_pos_e* is the dominant parameter, while for C2.2, *BrakePedalPosition* takes the lead. It is important to note that only the top influencing parameters are explicitly displayed in these plots.

The contributions of the parameters were also analysed for each fatigue class, specifically low and high, using SHAP

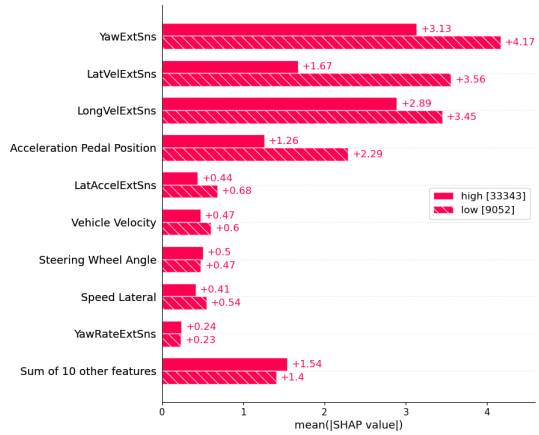

Fig. 8. Cohort bar plot presenting the importance values of top parameters in terms of SHAP values for the two classes of fatigue in C1.1.

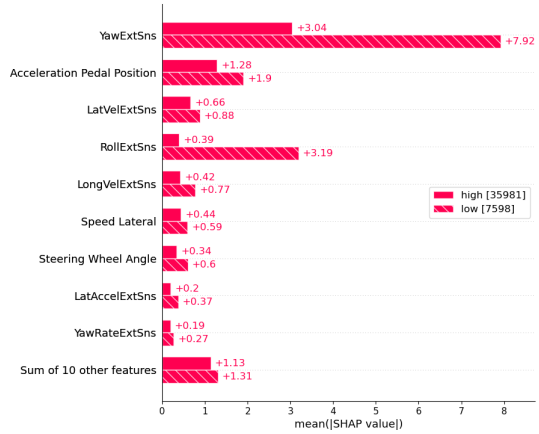

Fig. 9. Cohort bar plot presenting the importance values of top parameters in terms of SHAP values for the two classes of fatigue in C1.2.

cohort plots. Fig. 8 and 9 present the cohort plots for the C1.1 and C1.2 datasets, respectively, showing the same set of parameters identified in the corresponding Beeswarm plots. This visualisation allows for a clear comparison of feature contributions when analysing the datasets separately. The results indicate that in the C1.1 dataset, *YawExtSns*, *Acceleration Pedal Position*, and *LongVelExtSns* play a more significant role in fatigue classification. Notably, *Acceleration Pedal Position* emerges as a particularly important factor in C1.1, contributing more prominently than in C1.2 or the combined dataset.

Similar analyses for individual classes of fatigue were done for the C2.1 and C2.2 datasets, which are presented in Fig. 10 and 11, containing the same list of top parameters shown in the corresponding Cohort bar plots.

The contribution of parameters to the classifier model's decisions for individual samples is further explored using SHAP force plots. Fig. 12 and 13 illustrate the force plots for two specific samples: one with low fatigue and one with high fatigue. In a force plot, the bolded value represents the collective SHAP value derived from all the parameters for either low or high fatigue. The base value, which is the average of SHAP values across all samples, acts as a reference

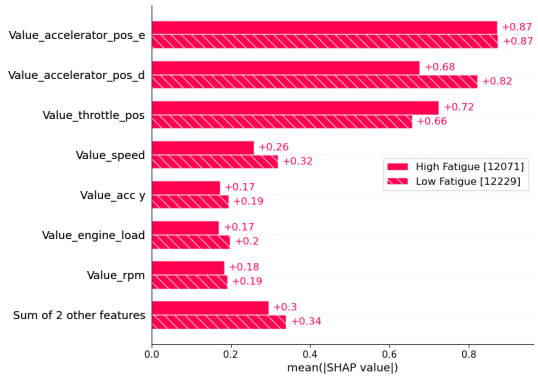

Fig. 10. Cohort bar plot presenting the importance values of top parameters in terms of SHAP values for the two classes of fatigue in C2.1.

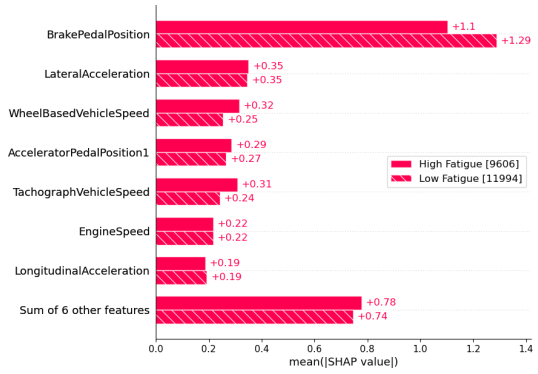

Fig. 11. Cohort bar plot presenting the importance values of top parameters in terms of SHAP values for the two classes of fatigue in C2.2.

point. If the collective SHAP value for a specific sample is below the base value, the model classifies the sample as low fatigue. Conversely, if the collective SHAP value is above the base value, the sample is classified as high fatigue. At the bottom of the force plot, the parameters are shown alongside their respective values, indicating the contribution of each parameter to the final prediction. The red and blue colour bars represent the extent to which each parameter either increases (red) or decreases (blue) the final SHAP value for that sample. Additionally, the width of the colour bars indicates the degree of influence each parameter has on the final classification decision, with wider bars signifying a stronger influence. The values depicted in these force plots are specific to the trained XGBoost classifier for the C1 dataset.

V. Summary and Conclusions

In summary, by integrating multimodal learning techniques, alignment strategies, and advanced feature extraction methods, the study significantly enhances both the accuracy and interpretability of driver fatigue classification. The role of precise alignment techniques in mitigating inconsistencies between different data modalities cannot be overstated. SHAP results provide valuable insights into the influence of various driving patterns on fatigue detection. This is also true as the literature also influences, for instance, research has demonstrated that drivers experiencing fatigue tend to exhibit reduced speed variability, longer reaction times, and less frequent lane corrections, leading to more consistent lateral positions [17]. Additionally, studies have suggested that fatigue is associated with diminished attention, which often manifests as erratic speed control or drifting within the lane [18]. These studies have shown that certain parameters for driving behaviours, such as *YawExtSns*, *Acceleration Pedal Position*, *LongVelExtSns*, *BrakePedalPosition* and *Speed*, can be significant indicators in the classification of driver fatigue. These results align with previous literature, reinforcing the importance of these driving behaviours as key indicators of fatigue.

Thus, this study, part of the FitDrive project, aims to develop a data-driven decision support system for identifying driver mental fatigue, providing detailed explanations for its predictions through the integration of MML and XAI. The key contributions and findings of this research are:

- *Representation of Diverse Data*: The study implemented advanced feature extraction techniques to create a new representation of multimodal data, which is crucial for accurately assessing driver fitness. This data included multiple sources, such as biometric data, vehicle telemetry, and environmental context, providing a more comprehensive view of a driver's mental state.
- *Alignment of True Labels*: A significant challenge in fatigue classification is the alignment of true fatigue labels with unlabelled data. By aligning these true labels effectively, the study ensured consistency and reliability in the labelling process, thus enhancing the robustness of the models.
- *Fusion of Features*: The study successfully integrated diverse feature sets into a unified dataset, enabling the models to make more accurate classifications of fatigue. By combining features from multiple modalities (*e.g.*, biometric signals, vehicle telemetry, contextual data), the model could consider a broader range of factors, improving its classification capabilities.
- *Fatigue Classification*: The developed models demonstrated high efficacy in classifying driver fatigue, illustrating the success of the data-driven approach. This not only validated the overall framework but also proved the viability of using multimodal data for fatigue detection, setting a foundation for future applications in driver safety and health monitoring.
- *Parameter Contribution Analysis*: A detailed SHAP [12] based analysis was conducted to assess the contributions of various parameters to the classification of driver fatigue. This analysis identified the most significant indicators, such as specific vehicle metrics and physiological signals, providing valuable insights into the drivers' behaviour and mental state.

Acknowledgment

The authors would like to express their sincere gratitude to all the members of the FitDrive project for their invaluable support in data collection, study protocol development, monitoring, and other essential tasks.

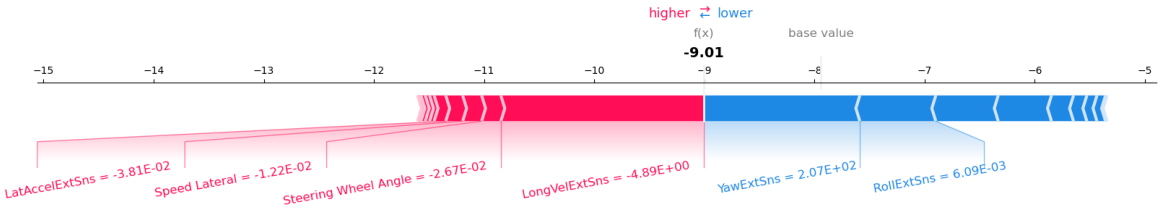

Fig. 12. SHAP Force plot showing the contribution of the parameters in the XGBoost classifier's decision for LOW fatigue.

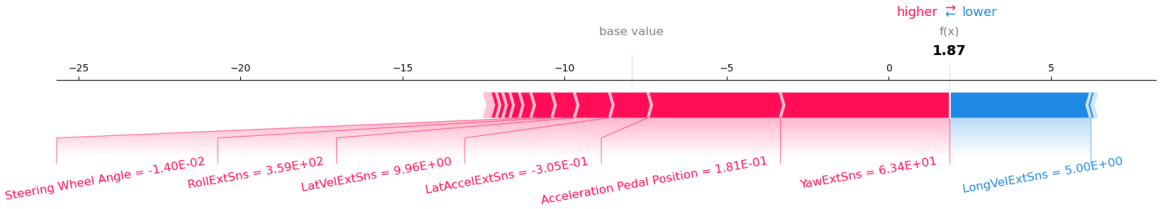

Fig. 13. SHAP Force plot showing the contribution of the parameters in the XGBoost classifier's decision for HIGH fatigue.

References

[1] J. Paxion, E. Galy, and C. Berthelon, "Mental Workload and Driving," *Front. Psychol.*, vol. 5, 2014.

[2] S. Luo *et al.*, "Effects of Distracting Behaviors on Driving Workload and Driving Performance in a City Scenario," *Int. J. Environ. Res. Public Health*, vol. 19, no. 22, p. 15 191, 2022.

[3] A. Barua *et al.*, "Second-Order Learning with Grounding Alignment: A Multimodal Reasoning Approach to Handle Unlabelled Data:" in *Proc. 16th Int. Conf. Agents Artif. Intell.*, 2024, pp. 561–572.

[4] H. Singh and A. Kathuria, "Analyzing Driver Behavior under Naturalistic Driving Conditions: A Review," *Accid. Anal. & Prev.*, vol. 150, p. 105 908, 2021.

[5] D. I. Tselentis and E. Papadimitriou, "Driver Profile and Driving Pattern Recognition for Road Safety Assessment: Main Challenges and Future Directions," *IEEE Open J. Intell. Transp. Syst.*, vol. 4, pp. 83–100, 2023.

[6] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal Machine Learning: A Survey and Taxonomy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 2, pp. 423–443, 2019.

[7] A. Barua, M. U. Ahmed, and S. Begum, "A Systematic Literature Review on Multimodal Machine Learning: Applications, Challenges, Gaps and Future Directions," *IEEE Access*, vol. 11, pp. 14 804–14 831, 2023.

[8] P. Lu *et al.*, "Learn to Explain: Multimodal Reasoning via Thought Chains for Science Question Answering," *Adv. Neural Inf. Process. Syst.*, vol. 35, 2022.

[9] S. Sun *et al.*, *A Review of Multimodal Explainable Artificial Intelligence: Past, Present and Future*, arXiv preprint, arXiv:2412.14056 [cs], 2024.

[10] A. Pahud de Mortanges *et al.*, "Orchestrating Explainable Artificial Intelligence for Multimodal and Longitudinal Data in Medical Imaging," *NPJ Digit. Med.*, vol. 7, no. 1, pp. 1–10, 2024.

[11] A. Destiny, "Leveraging Explainable AI and Multimodal Data for Stress Level Prediction in Mental Health Diagnostics," *Int. J. Res. Innov. Appl. Sci.*, vol. IX, no. XII, pp. 416–425, 2025.

[12] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Proc. of the 31st Int. Conf. on Neural Inf. Process. Syst.*, ser. NIPS'17, 2017, pp. 4768–4777.

[13] R. Fu, H. Wang, and W. Zhao, "Dynamic Driver Fatigue Detection using Hidden Markov Model in Real Driving Condition," *Expert Syst. with Appl.*, vol. 63, pp. 397–411, 2016.

[14] A. Joshi *et al.*, "In-the-wild Drowsiness Detection from Facial Expressions," in *Proc. IEEE Intell. Veh. Symp. (IV)*, 2020, pp. 207–212.

[15] A. Giorgi *et al.*, "Neurophysiological Mental Fatigue Assessment for Developing User-centered Artificial Intelligence as a Solution for Autonomous Driving," *Front. Neurorobot.*, vol. 17, p. 1 240 933, 2023.

[16] G. Di Flumeri *et al.*, "EEG-Based Index for Timely Detecting User's Drowsiness Occurrence in Automotive Applications," *Front. Hum. Neurosci.*, vol. 16, 2022.

[17] P. Jackson *et al.*, "Fatigue and Road Safety: A Critical Analysis of Recent Evidence," Department for Transport, London, Tech. Rep., 2011.

[18] H. P. Van Dongen *et al.*, "The Cumulative Cost of Additional Wakefulness: Dose-Response Effects on Neurobehavioral Functions and Sleep Physiology From Chronic Sleep Restriction and Total Sleep Deprivation," *Sleep*, vol. 26, no. 2, pp. 117–126, 2003.