Late Fusion-Based Multimodal Machine Learning for Driver Fatigue Detection in Emergency Response Driving Scenarios

Zulfa Sammadile[1,2], Deogratias Shidende[3], Bonny Mgawe[1], Sabine Moebs[3]

[1] The Nelson Mandela African Institution of Science and Technology, P.O. Box 447, Arusha, Tanzania
[2] Ardhi University, P.O. Box 35176, Dar es Salaam, Tanzania
[3] The Baden-Württemberg Cooperative State University, Marienstraße 20, 89518, Heidenheim, Germany
zulfasammadile@gmail.com

Abstract. Globally, road accidents are among the leading causes of death and injury, with driver fatigue being one of the main contributors. Driver fatigue is especially critical in emergency response driving, where factors such as long and overnight shifts are common. This study presents a multimodal driver fatigue detection Machine Learning (ML) model that utilizes a late fusion approach. A range of behavioral and physiological data was used to train base learners with various ML models, and their prediction probabilities were used to train four decision tree-based metaclassifiers (RF, XGB, LGB, and HGB) within a stacking late fusion framework. For a broader exploration, ensemble models of these metaclassifiers were trained using simple average, weighted average, majority voting, dynamic weighted average, and best confidence selection methods. Among the metaclassifiers, XGB and dynamic weighted average achieved the highest accuracies of 86.83% and 87.17% respectively, with XGB also yielding the lowest FNR, highlighting its inherent capability in handling heterogeneous and diverse features. To simulate real-world emergency driving conditions, where constant sensor inputs may be intermittent, missing modalities were simulated in Google Colab during training of the metaclassifiers. Later, the robustness of the metaclassifiers was analyzed across various missing modality scenarios, where up to three sensor inputs could be unavailable. In this analysis, XGB achieved the highest stability with the lowest SD of 0.039 across all missing modality scenarios, demonstrating its ability to maintain reliable driver fatigue detection performance even when some sensor inputs are missing.

Keywords: Missing modalities, Ensemble Learning, Emergency Response Driving, Drowsiness Detection, Stacking.

1 Introduction

Globally, road accidents claim approximately 1.19 million lives and result in up to 50 million non-fatal injuries annually, with driver fatigue accounting for at least 15% of

S. Moebs et al. (eds.), *Proceedings of the FIREtalk Conference - Research on FIRE! (research-on-fire 2025)*,
Advances in Social Science, Education and Humanities Research 1018,
https://doi.org/10.2991/978-94-6239-705-7_8

serious crashes [1, 2]. Given the lack of a definitive post-accident measure for fatigue, these statistics are likely underreported [3]. Studies have shown that driver fatigue increases the risk of road crashes by 29% [4], and acute sleep deprivation of at least 3 hours can increase the risk by up to 11.5 times [5]. This issue is particularly critical in emergency response driving, where personnel must maintain high performance under high-stress conditions, long shifts, and in rapidly evolving environments [6–8].

Driver fatigue is a state of mental or physical exhaustion that impairs a driver's ability by slowing reaction times, impairing decision-making, and reducing concentration [9]. It is often categorized into sleep-related (SR) fatigue from sleep debt and task-related (TR) fatigue, from physical exertion, cognitive overload, or cognitive underload [10]. Many drivers are unaware of their fatigued state, and timely intervention could prevent up to 90% of road accidents [3, 6], highlighting the need for effective driver fatigue detection systems.

In recent years, Machine Learning (ML) has shown promise in detecting driver fatigue using behavioral, physiological, and vehicle kinematics data [11, 12]. Most studies primarily employed unimodal approaches, which rely on a single data modality and thus raise concerns about reliability [12–14]. Consequently, multimodal approaches emerged, fusing multiple data modalities to improve driver fatigue detection [11, 15, 16]. However, a notable gap remains, particularly with missing data modalities, where sensor inputs from one or more channels could be unavailable or incomplete.

This gap is often overlooked because current research relies on controlled settings where sensors are always available, favoring early and intermediate fusion techniques [11, 15, 16]. However, these techniques are susceptible to performance degradation when sensor inputs are unreliable, making them unsuitable for emergency response driving, where full sensor availability is not guaranteed [17, 18]. To address this gap, we developed a late fusion-based multimodal driver fatigue detection model. By training modalities (i.e., base learners) independently and combining their outputs at the decision level (i.e., through metaclassifiers) [19], our architecture inherently maintains robustness even when specific data streams are incomplete. Moreover, late fusion mitigates the curse of dimensionality, reduces overfitting, and improves generalization [17, 20, 21]. To overcome late fusion's common sensitivity to poor aggregation, we implemented stacking, where base learner predictions are used to train metaclassifiers [22–24].

Our study trained a binary classification ML model using behavioral data: eye-tracking and yawn data, and physiological data: electroencephalogram (EEG), electrocardiogram (ECG), electrodermal activity (EDA), and respiration rate data (RESP). Base learners included Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR), Gradient Boosting (GB), Neural Networks (NN), and Convolutional Neural Network (CNN). Their prediction probabilities, infused with 30% missing modalities (NaN sensor readings), were used to train four tree-based metaclassifiers: RF, HighGradientBoosting (HGB), Extreme Gradient Boosting (XGB), and Light Gradient Boosting Machine (LGB). We also developed five ensemble models of these metaclassifiers to assess whether combining them would improve performance.

This research presents three main contributions to multi-modal driver fatigue detection. First, a robust late fusion ML model under missing modality conditions, with a particular focus on emergency response driving; second, a cross-dataset training methodology on non-synchronized, independently collected multimodal datasets without data loss, addressing the challenge of developing such ML models in the absence of synchronized datasets; third, an empirical evaluation of ensemble metaclassifiers in handling missing modalities. The following research questions guided the study:

1. How effective are physiological and behavioral data modalities individually, and how does their fusion enhance fatigue detection?
2. How can late fusion strategies be applied to maintain driver fatigue detection reliability and performance in emergency response driving, where constant availability of data streams is not guaranteed?

The remainder of the paper is structured as follows: Section 2 presents the literature review; Section 3 describes the materials and methods; Section 4 presents the results; Section 5 discusses the findings; and Section 6 concludes the study.

2 Literature Review

2.1 The Critical Problem of Driver Fatigue

Driver fatigue is a major global contributor to road accidents [25]. Drowsy driving is notoriously underreported; the National Highway Traffic Safety Administration (NHTSA) estimated 91,000 crashes in 2017, while the American Automobile Association suggests the actual number is closer to 320,000 annually [26, 27]. Unlike alcohol impairment, there is no objective "fatigue breathalyzer" for quantification, causing police reports to rely on subjective judgment [3, 28, 29].

Emergency response driving is a uniquely high-risk, high-stress environment [30, 31]. The need for abrupt, aggressive maneuvers renders vehicle kinematics-based detection methods unsuitable, as they often misinterpret intentional actions as fatigue, causing false alarms [32, 33]. In this high-stakes context, it is crucial to differentiate between sleep-related (SR) fatigue, a physiological state linked with sleep debt and circadian rhythms, and task-related (TR) fatigue, induced by the cognitive and physical demands of the driving task itself [10, 34]. These two types often have a cumulative effect in emergency response, where chronic SR fatigue is compounded by acute TR fatigue during high-stress calls [35, 36]. Therefore, a detection system must account for both the driver's baseline state and acute task-induced decrements.

2.2 Unimodal Fatigue Detection

Behavioral systems primarily use in-vehicle cameras to monitor physical fatigue cues, such as eye activity, facial expressions, and head movements [37, 38]. Key metrics include Percentage of Eye Closure (PERCLOS), blink rate, and yawning detection [39, 40]. These systems, often employing Deep Learning (DL) techniques such as CNNs, have demonstrated high accuracy, with some hybrid models achieving 97.83% [41, 42]. However, they are highly susceptible to environmental factors, such as poor lighting or occlusions from eyeglasses [39, 43]. A critical drawback is that behavioral signs typically appear late in the fatigue process, after alertness has already been impaired, rendering them inadequate for providing timely warnings [44, 45].

Physiological systems provide a direct means of assessing fatigue by monitoring the body's transition from alertness to drowsiness through wearable sensors. EEG is considered a gold standard for detecting microsleeps and can provide early warnings by monitoring subtle brainwave shifts [44, 45]. Other crucial signals include electrooculogram (EOG) for eye movements, ECG for heart rate (HR) and heart rate variability (HRV), skin temperature, RESP, and EDA, also sometimes referred to as galvanic skin response (GSR) for changes in the autonomic nervous system [46, 47]. Driver fatigue detection systems that employ such signals have demonstrated high accuracy, often exceeding 90% [43, 47]. However, their use is challenged by high inter- and intra-subject variability, non-specificity of physiological responses to fatigue, noise in the signals, and the lack of universally accepted guidelines for interpreting the measures [43].

Vehicle kinematics systems infer fatigue by analyzing vehicle operational parameters, such as steering wheel angle (SWA) and speed variations [32, 37]. Although applicable in normal driving scenarios, these systems are fundamentally unsuitable for emergency response driving [37]. The high-speed, non-linear maneuvers inherent to emergency driving, such as sudden braking and aggressive steering, are indistinguishable from fatigue-induced errors, leading to frequent false alarms [32].

2.3 Multimodal and Fusion Strategies

Due to the limitations of single modality systems, fusing data streams significantly improves the accuracy and reliability of fatigue detection [37, 48, 49]. By combining behavioral signs with physiological data, multimodal systems can detect fatigue more reliably than unimodal approaches [50, 51], as systems can continue to function even if one sensor fails [52]. This robustness is crucial in dynamic environments such as emergency response driving.

Data fusion can be achieved through early, intermediate, or late fusion strategies [53]. Early and intermediate fusion strategies merge data before or during the modeling stage, with early fusion combining raw or preprocessed features from different modalities, and intermediate fusion integrating features at the representation level after modality-specific feature extraction [11, 15, 16]. Although both approaches can potentially leverage complex interactions between modalities [54], they require all sensor data to be available and time-synchronized, conditions that are often unfeasible in real-world driving. A missing or noisy input can cause the model to fail, making it vulnerable to environmental factors like sun glare or sensor dropout [55, 56].

In contrast, late fusion processes each modality independently and combines the outputs at a higher decision level [57]. The key advantage is its flexibility and graceful degradation; if one sensor fails, the system can still operate using the remaining modalities [52, 58]. This makes late fusion ideal for unpredictable environments such as emergency response driving, as it is robust to missing or noisy data [59]. Late fusion also allows each modality to be optimized separately and is easier to modify without retraining the entire model [58].

3 Materials and Methods

3.1 Data Acquisition

This study utilized publicly available data on general driver fatigue following the scarcity of emergency driving datasets. It employed labelled eye-tracking, yawning, and EEG datasets from Kaggle [60–62]; and ECG/EDA/RESP dataset from Zenodo [63] (See Table 1), enabling supervised learning.

Table 1. Summary of the datasets used to train base learners, including their sizes and sample descriptions

Dataset	Size	Sample Description
Eye-tracking videos [61]	144 videos (73 alert and 71 fatigued)	Recorded by 48 healthy adults using mobile and web cameras in mimicked driving scenarios
Eye-tracking images [61]	13,869 images (6,936 closed and 6,933 open eyes)	Captured from adults with and without glasses
Yawn images [60]	1,448 images (723 yawn images and 725 no yawn images)	Captured from healthy adults in driver positions
EEG signals [62]	Recordings from 12 participants	Continuous brain activity sampled at 1000 Hz in CNT file formats
ECG/EDA/RESP [63]	870 data points	Acquired in a single CSV file with 400 features derived from 78 participants, segmented into 10-second windows

3.2 Data Preprocessing

Data from all modalities underwent a rigorous preprocessing pipeline to ensure uniformity and enhance model performance.

Data Cleaning. Eye-tracking videos recorded in segments were merged accordingly. Eye-tracking images were resized to 64x64 pixels and converted to grayscale for computational efficiency. Yawn images were resized to 224x224 pixels, an appropriate size for its training model. The color of the yawn images was, however, retained as it was essential for accurate yawn detection. For the ECG/EDA/RESP dataset, features with more than 50% null values were dropped, leaving 342 features. The EEG dataset was cleaned using the MNE library with channel selection (based on the 10-20 system [64]), a 1–45 Hz band-pass filter to remove noise [65], and average re-referencing to improve the signal-to-noise ratio [66]. Data preprocessing also involved dropping highly correlated features to avoid multicollinearity [67].

Feature Extraction. For eye-tracking videos, OpenCV and MediaPipe Face Mesh were used for frame-by-frame video processing and facial landmarks identification. The Eye Aspect Ratio (EAR) was calculated for both eyes and then averaged to reduce noise and fluctuations [51]. This combined EAR was used to extract blink-related and PERCLOS features using an EAR threshold of 0.2, which is consistent with prior studies [68–70]. For EEG signals, a 12.0-second non-overlapping sliding window segmentation was applied to balance frequency resolution and signal stationarity. From these segments, time-domain features (mean, SD, max, min, peak-to-peak, RMS) and frequency-domain features (ratios of Delta, Theta, Alpha, Beta, and Gamma power bands) were extracted.

Normalization. For the image datasets, the default pixel range of a standard 8-bit image (0 to 255) was converted to a range of 0 to 1 by dividing the pixel values by 255.0. For the EEG dataset, z-score [71] and relative change normalization [72] were explored.

Data Splitting. Leave-One-Subject-Out (LOSO) was employed for the physiological datasets (EEG, ECG/EDA/RESP) to mitigate data leakage [73]. For the other datasets, holdout cross-validation was employed. The eye-tracking videos dataset was split into 80% training and 20% testing sets, given its small size. The image datasets for both yawn and eye-tracking were divided into 70% training, 15% validation, and 15% testing sets, given their relatively larger sizes, to improve model generalization during training with the validation set.

Data Augmentation. Data augmentation was conducted with TensorFlow on the training sets for the image datasets to diversify the training data and mitigate overfitting [74]. For eye-tracking images, augmentation included zooming, rotation, horizontal flipping, and pixel modification. Yawn image augmentation focused on preserving facial features, including horizontal flipping, small rotations and zooms, brightness and contrast adjustments, and center cropping.

3.3 Base Model Training

To address the first research question regarding the individual effectiveness of data modalities, separate base learners were trained for each data stream.

Eye-tracking Videos. RF, SVM, and LR classifiers were trained using 5-fold cross-validation to classify samples as fatigued or alert. RF was configured with 100 estimators; SVM used a radial basis function (RBF) kernel with probability set to True to enable probabilistic outputs for stacking; and LR employed L2 regularization to ensure convergence for the high-dimensional feature space. A fixed random seed was used for all base learners to ensure reproducibility.

Eye-tracking Images. A sequential CNN was developed for binary classification with 5-fold cross-validation in 50 epochs. Its architecture consisted of four convolutional blocks, followed by dense layers and a sigmoid output. The model was trained for 50 epochs with early stopping based on validation loss.

Yawn Images. Transfer learning with the pre-trained EfficientNetB0 architecture was used. The model was fine-tuned in two phases: first, by training a new classification head with the frozen EfficientNetB0 layers in 25 epochs, and then by unfreezing a portion of the base model for a final fine-tuning phase of 10 epochs with early stopping.

EEG Signals. RF, SVM, GB, LR, and NN models were trained using 12-fold LOSO cross-validation.

ECG/EDA/RESP Dataset. RF, SVM, and LR were trained using 78-fold LOSO cross-validation.

Base learners were primarily selected based on their strengths and the dataset requirements. LR was selected for interpretability [75], SVM to capture more complex relationships [76], NN for its superiority in regression problems [77], and ensemble methods such as RF and GB for their classification strengths [78]. Due to the smaller size of the yawn dataset, EfficientNetB0 was selected to improve model performance [79] and CNN was selected for the eye image dataset for its strengths in DL [80].

3.4 Late Fusion-Based Multimodal Model Development

To address the second research question, the prediction probabilities for the best-performing base learners for each data stream were stored. Data preprocessing for the metaclassifiers involved loading these prediction probabilities and developing a fusion dataset in two stages. First, by randomly combining the loaded prediction probabilities to address the difference in sample sizes and avoid data loss completely. Second, by infusing the samples with 30% of simulated missing modalities. The fusion dataset was split using a holdout cross-validation, where 80% of the data was allocated for training and 20% for testing.

Training was performed using four tree-based ML models: RF, HGB, XGB, and LGB, due to their strengths in handling heterogeneous features and non-linear relationships. HGB, XGB, and LGB were trained on the fusion dataset due to their inherent ability to handle missing values (NaN). In contrast, RF was trained on a separate version of the dataset, which was created by imputing the missing data with 0, as it cannot inherently deal with missing data modalities. RF was included as a benchmark model due to its broad adoption in the literature [11–13].

Ensemble Model Training. Ensemble models for the individual metaclassifiers were trained using the following methods: simple average in which equal weights were assigned, weighted average which assigned model-specific weights according to the overall accuracy, dynamic weighted average which dynamically allocated more weights to models with high confidence, majority voting which selected the most voted class, and best confidence selection which chose the metaclassifier with the highest confidence. Confidence was defined as the absolute distance of the predicted probability from 0.5, representing a natural decision boundary. Best confidence defaults to the weighted average when all models exhibited low confidence (threshold < 0.25).

3.5 Performance Evaluation

The performance of all base learners and metaclassifiers was evaluated on test data using accuracy, AUC-ROC, confusion matrices, and a per-class report on precision, recall, and F1-score using sklearn's classification_report(). Additionally, a missing modality impact analysis was performed to assess the metaclassifiers' robustness under four test scenarios (complete data, one missing modality, two missing modalities, and three missing modalities), with a maximum of three missing modalities per data point to avoid complete system failure. In this analysis, False Negative Rate (FNR), False Positive Rate (FPR), and Standard Deviation (SD) were employed.

3.6 Experimentation Environment

The training, evaluation, and inference of base learners and the metaclassifiers were conducted on Google Colab, which provided sufficient computational resources. The training time for the base learners and metaclassifiers varied depending on model complexity and dataset size, with the longest being the CNN classifier for the yawn dataset, which required 58 minutes. Inference for single samples was sufficiently fast to support near real-time deployment. All experiments were run using standard ML and DL libraries in Python, ensuring reproducibility and portability across different platforms.

4 Results

4.1 Base Model Performance Analysis

Behavioral Modalities. The behavioral data modalities demonstrated a strong classification accuracy. For the eye-tracking video dataset, the accuracy of RF (82.76%) outperformed that of SVM (75.86%) and LR (79.31%), demonstrating RF's effectiveness in capturing complex relationships in the heterogeneous video features. RF consistently outperformed SVM and LR in a per-class analysis in other metrics (see Table 2).

Table 2. Classification report for eye-tracking video fatigue detection ML model

Model	Class	Precision	Recall	F1-score	Support
RF	Alert	78%	93%	85%	15
	Fatigued	91%	71%	80%	14
SVM	Alert	70%	93%	80%	15
	Fatigued	89%	57%	70%	14
LR	Alert	74%	93%	82%	15
	Fatigued	90%	64%	75%	14

For the eye-tracking image dataset, the CNN model achieved a high performance with a test accuracy of 96.68%, highlighting its ability to learn fine-grained visual features of eye states. The per-class analysis in Table 3 illustrates the model's consistent performance across all metrics.

Table 3. Classification report for the eye-tracking image classification model

Class	Precision	Recall	F1-score	Support
Closed Eyes	97%	97%	97%	1041
Open Eyes	97%	97%	97%	1040

Similarly, for the yawn image dataset, the EfficientNetB0 architecture achieved high performance with a test accuracy of 86.19%. When compared to the performance of the raw CNN (54.53%) on the yawn dataset, the ability of transfer learning to address data scarcity by leveraging pre-trained models is demonstrated. The per-class analysis in Table 4 emphasizes the impact that transfer learning has had on our small yawn dataset.

Table 4. Classification report for the yawn image classification model

Class	Precision	Recall	F1-score	Support
No Yawn	83%	89%	86%	99
Yawn	89%	84%	87%	111

Physiological Modalities. The physiological data modalities proved to be powerful indicators of internal physiological states related to fatigue, despite their small sample sizes. For the EEG dataset, the SVM model achieved the highest LOSO accuracy of 71.56% (see Table 5) with the relative change normalization strategy.

Table 5. EEG classification model performance evaluation

Model	LOSO Accuracy	LOSO F1-score	LOSO Precision	LOSO Recall
RF	68.56%	64.67%	73.88%	57.50%
GB	66.31%	60.85%	72.69%	52.55%
SVM	71.56%	68.16%	77.49%	60.83%
LR	68.64%	64.53%	74.35%	57.00%
NN	69.72%	66.67%	74.23%	60.50%
NN	69.72%	66.67%	74.23%	60.50%

The ECG/EDA/RESP dataset performed exceptionally well, with RF achieving a LOSO accuracy of 95.96% and a near-perfect AUC-ROC score of 99.49% (see Table 6), emphasizing RF's effectiveness in capturing complex, non-linear relationships within the diverse and heterogeneous features in the ECG/EDA/RESP dataset.

Table 6. ECG/EDA/RESP ML performance evaluation report

Model	LOSO Accuracy	LOSO F1-score	LOSO AUC-ROC
RF	95.96%	93.59%	99.49%
SVM	94.34%	91.08%	98.58%
LR	94.55%	91.84%	98.05%

4.2 Metaclassifier Results and Missing Modalities Impact Analysis

Fig. 1 and Fig. 2 illustrate a comparative analysis of the individual and ensemble metaclassifiers. The weights for the weighted average ensemble were: *RF=0.251, HGB=0.249, XGB=0.252, LGB=0.249*, and the selection frequencies for the best confidence selection model were: *XGB; 480 (80.0%), HGB; 57 (9.5%), weighted average*

32 (5.3%), RF; 17 (2.8%), and LGB; 14 (2.3%). From Fig. 1, dynamic weighted average demonstrates the highest accuracy, while XGB demonstrates the lowest FNR and SD across all missing modality scenarios (also see Table 7), positioning XGB as the best-performing metaclassifier. These results highlight XGB's classification capacity in complex, diverse datasets in the presence of missing modalities. RF, on the other hand, struggled the most, likely due to its limited capability with missing modalities.

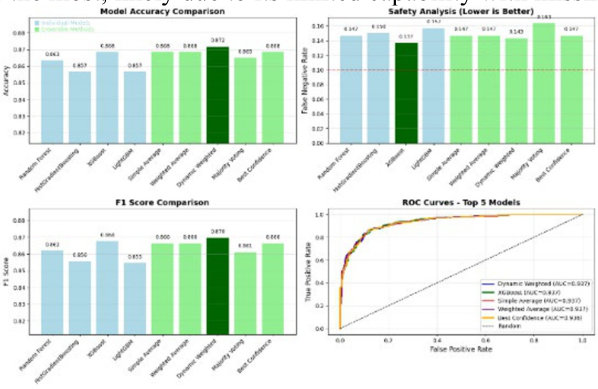

Fig. 1. Performance comparison of the top five metaclassifiers based on accuracy, FNR, F1-score, and ROC

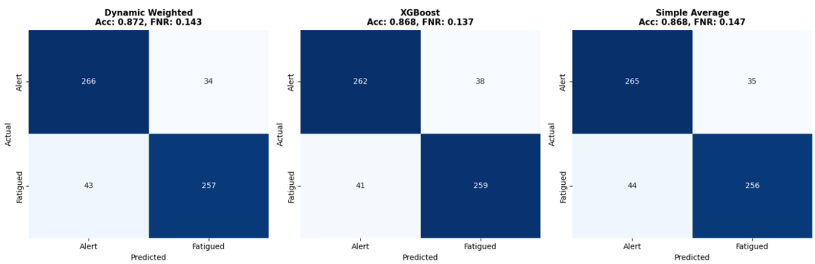

Fig. 2. Confusion matrices for the top three performing models by accuracy. Note that Dynamic Weighted shows superior identification of the Alert class, whereas XGB demonstrates higher precision in identifying Fatigued samples

Table 7. Robustness analysis across missing modality scenarios, reporting average performance variation and the minimum (worst-case) accuracy observed

Model	Average Accuracy	SD	Worst Accuracy
HGB	79.2%	0.041	72.0%
LGB	79.2%	0.043	71.6%
Simple Average	79.0%	0.040	71.8%
Dynamic Weighted	78.9%	0.041	71.6%
XGB	78.7%	0.039	72.4%
RF	77.5%	0.050	70.0%

5 Discussion

This study extends existing research in ML-based driver fatigue detection [11, 12, 74, 81, 82] by proposing a late fusion-based multimodal approach tailored for the dynamic environment of emergency response driving. Although this study utilizes publicly available data that do not exclusively represent emergency response driving conditions, the selected datasets possess fatigue indicators (features) that are consistent across different contexts and tasks. This is mainly because fatigue manifestation at behavioral and physiological levels has shown to be task-independent [83], and are therefore transferable across domains. Nevertheless, we acknowledge that the datasets collected may not fully represent emergency response driving; thus, the results of this study are framed as a feasibility study rather than a deployment-ready solution.

Our study's findings confirm that while individual data modalities can effectively capture symptoms of fatigue, a fused system is essential for achieving the robustness required for real-world applications. Our base model analysis revealed that modalities possess distinct strengths for fatigue detection. The eye-tracking image dataset, for example, achieved a high accuracy of 96.68% with CNN, consistent with prior research [84, 85], highlighting the reliability of periorbital data as a strong indicator of drowsiness. Similarly, the ECG/EDA/RESP dataset performed exceptionally well, with a 95.96% accuracy with RF, underscoring the power of physiological signals as direct indicators of internal fatigue states.

In contrast, the EEG dataset achieved an accuracy of 71.56% with SVM, a relatively lower accuracy compared to other studies [86, 87]. This performance was likely due to the limited EEG sample size or our cross-validation choice of LOSO. LOSO is highly recommended for physiological data [73, 88–90] to prevent data leakage over holdout cross-validation, employed by both [86, 87]. However, LOSO has a particular trade-off in accuracy. A study by [91] employed both cross-validation techniques on EEG data for driver fatigue detection and achieved completely different test results with LOSO (70.45%) as compared to the holdout method (90.73%).

Nevertheless, each modality has its own limitations in detecting driver fatigue. EEG-based fatigue detection, for instance, presents an inherent generalization challenge due to high inter- and intra-subject variability caused by the unique differences in brain activity across individuals [92–95]. Similarly, ECG, EDA, and RESP data are non-specific to fatigue, as their variations can also be influenced by factors such as stress, gender, ethnicity, lifestyle, emotion, and external factors like noise and heat [96–98]. Furthermore, studies by [85, 99, 100] highlight the challenges associated with behavioral data in driver fatigue detection, including issues with lighting and occlusion. These factors emphasize the importance of multimodal driver fatigue detection approaches for better performance and reliability.

Our multimodal model overcomes the challenges of unimodal approaches by combining behavioral and physiological data modalities. Additionally, it addresses a common challenge in multimodal driver fatigue detection studies, where early and intermediate fusion strategies fail to account for missing modalities, by exploring the use of stacking late fusion. With this approach, XGB demonstrated good performance with an accuracy of 86.83%, a low FNR of 0.137, and a minimal variation of 0.039 SD across all the simulated missing modality scenarios.

The inherent demands of emergency response driving justify our design choices. Unlike monotonous highway driving, the "unstable" nature of emergency driving renders vehicle kinematics data less reliable. By excluding such data, our model correctly identifies fatigue without misclassifying necessary maneuvers as signs of impairment. Moreover, the model excludes fatigue from cognitive underload, as it is rare in emergency response driving scenarios.

This study acknowledges some limitations. First, small sample sizes for the acquired data impacted feature selection and model performance. The study could benefit from additional data collection to improve model performance. Second, the simulation of missing modalities may not capture every failure in real-world emergency driving. Future studies could conduct in-depth research on driver fatigue, particularly in emergency response driving, to incorporate more solutions to the various other problems that exist in this area. Lastly, some sensors, such as EEG headsets and ECG wristbands, can be intrusive for emergency response operations [43], a challenge that will be further explored in our future work. Nonetheless, our study demonstrates how independently trained modality-specific models can be integrated into a single multimodal fatigue detection model that caters to the demands of emergency response driving.

6 Conclusion

This paper presents a multimodal ML model for driver fatigue detection in emergency response driving scenarios, leveraging stacking late fusion to address missing modalities. Among the trained metaclassifiers, the dynamic weighted average ensemble achieved the highest accuracy of 87.17%, followed closely by XGB with an accuracy of 86.83%, which also has the lowest FNR of 0.137. Additionally, robustness analysis with missing modalities showed XGB to be the most stable metaclassifier with the least SD of 0.039.

This study's findings demonstrate the effectiveness of the late fusion stacking approach for integrating behavioral and physiological data, accounting for missing modalities, a common challenge in the real world yet often overlooked, and render a model tailored for emergency response driving needs.

Future research will extend this feasibility study to deployment through domain-specific data collection, exploring the approaches to embed the proposed models into vehicular systems, considering factors such as real-time data acquisition, on-board processing, alert mechanisms, and human factors such as sensor placement and comfort, system usability, response time, and user perceptions on the system.

Acknowledgement. The authors would like to express their sincere appreciation to the Baden-Württemberg Stipendium for sponsoring this research and to the host university, Duale Hochschule Baden-Württemberg (DHBW), where this research was conducted.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. WHO: Road traffic injuries, https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries, last accessed 2025/09/07.
2. Kashani, A.T., Moghadam, M.R., Amirifar, S.: Factors affecting driver injury severity in fatigue and drowsiness accidents: a data mining framework. J Inj Violence Res. 14, 75, (2022). https://doi.org/10.5249/JIVR.V14I1.1679.
3. Dawson, D., Reynolds, A.C., Van Dongen, H.P.A., Thomas, M.J.W.: Determining the likelihood that fatigue was present in a road accident: A theoretical review and suggested accident taxonomy. Sleep Med. Rev. 42, 202–210 (2018). https://doi.org/10.1016/j.smrv.2018.08.006.
4. Moradi, A., Nazari, S.S.H., Rahmani, K.: Sleepiness and the risk of road traffic accidents: A systematic review and meta-analysis of previous studies. Transp Res Part F Traffic Psychol Behav. 65, 620-629, (2019). https://doi.org/10.1016/J.TRF.2018.09.013.
5. Tefft, B.C.: Acute Sleep Deprivation and Risk of Motor Vehicle Crash Involvement. AAA Found. Traffic Saf. 28, (2016). https://aaafoundation.org/acute-sleep-deprivation-risk-motor-vehicle-crash-involvement.
6. National Highway Traffic Safety Administration (NHTSA). Ground ambulance crashes. Accessed Aug. 28, (2025). https://www.usfa.fema.gov/blog/nhtsa-publishes-data-on-ground-ambulance-crashes.
7. Sammadile, Z., Luwemba, G., Mgawe, B., Nyambo, D.: Augmented Reality-Based Outdoor Navigation System for Urban Firefighting: A Case of Dar es Salaam, Tanzania. Indian Journal of Science and Technology. 18, 2491-2499, (2025). https://doi.org/10.17485/IJST/v18i31.793.
8. Hossain, M.M., Zhou, H., Das, S.: Data mining approach to explore emergency vehicle crash patterns: A comparative study of crash severity in emergency and non-emergency response modes. Accid Anal Prev. 191, 107217, (2023). https://doi.org/10.1016/J.AAP.2023.107217.
9. Schagen, I.: European Road Safety Observatory Road Safety Thematic Report-Fatigue. (2021).
10. Guo, F., Zhou, Y., Wang, X., Li, W., Cai, J.: Literature review of driving fatigue research based on bibliometric analysis. J. Traffic Transp. Eng. Engl. Ed. 11, 1401–1419 (2024). https://doi.org/10.1016/j.jtte.2024.03.005.
11. Misra, A., Samuel, S., Cao, S., Shariatmadari, K.: Detection of Driver Cognitive Distraction Using Machine Learning Methods. IEEE Access. 11, 18000-18012, (2023). https://doi.org/10.1109/ACCESS.2023.3245122.
12. He, D., Wang, Z., Khalil, E.B., Donmez, B., Qiao, G., Kumar, S.: Classification of Driver Cognitive Load: Exploring the Benefits of Fusing Eye-Tracking and Physiological Measures. Transp Res Rec. 2676, 670-681, (2022). https://doi.org/10.1177/03611981221090937.
13. Rahman, H., Ahmed, M.U., Barua, S., Funk, P., Begum, S.: Vision-based driver's cognitive load classification considering eye movement using machine learning and deep learning. Sensors. 21, (2021). https://doi.org/10.3390/S21238019.

14. Prabhakar, G., Mukhopadhyay, A., Murthy, L., Modiksha, M., Sachin, D., Biswas, P.: Cognitive load estimation using ocular parameters in automotive. Transp. Eng. 2, 100008, (2020). https://doi.org/10.1016/J.TRENG.2020.100008.
15. Yan, X., Abas, A.: Advancements and Perspectives in Fatigue Driving Detection: A Comprehensive Review. ICCK Trans. Intell. Unmanned Syst. 1, 4-15, (2024). https://doi.org/10.62762/TIUS.2024.767724.
16. Zhang, M.: Research on Driver's Fatigue Detection Based on Information Fusion. Comput. Mater. Contin. 79, 1039-1061, (2024). https://doi.org/10.32604/CMC.2024.048643.
17. Li, S., Du, C., Zhao, Y., Huang, Y., Zhao, H.: What Makes for Robust Multi-Modal Models in the Face of Missing Modalities?, http://arxiv.org/abs/2310.06383, (2023). https://doi.org/10.48550/arXiv.2310.06383.
18. Wang, M., Fan, S., Li, Y., Xie, Z., Chen, H.: Missing-modality enabled Multi-modal Fusion Architecture for medical data. 164. 104796. (2025). https://doi.org/10.1016/j.jbi.2025.104796.
19. Boulahia, S.Y., Amamra, A., Madi, M.R., Daikh, S.: Early, intermediate, and late fusion strategies for robust deep learning-based multimodal action recognition. Mach Vis Appl. 32, 1-18, (2021). https://doi.org/10.1007/S00138-021-01249-8.
20. Pereira, L.M., Salazar, A., Vergara, L.: A comparative analysis of early and late fusion for the multimodal two-class problem. IEEE Access. 11, 84283–84300 (2023).
21. Yang, Q., Zhao, Y., Cheng, H.: MMLF: Multi-modal Multi-class Late Fusion for Object Detection with Uncertainty Estimation, http://arxiv.org/abs/2410.08739, (2024). https://doi.org/10.48550/arXiv.2410.08739.
22. Rwigema, J., Mfitumukiza, J., Tae-Yong, K.: A hybrid approach of neural networks for age and gender classification through decision fusion. Biomed. Signal Process. Control. 66, 102459 (2021). https://doi.org/10.1016/j.bspc.2021.102459.
23. Chandola, Y., Virmani, J., Bhadauria, H.S., Kumar, P.: End-to-end pretrained CNN-based computer-aided classification system design for chest radiographs. Primers in Biomedical Imaging Devices and Systems, (2021).
24. Kalagotla, S.K., Gangashetty, S.V., Giridhar, K.: A novel stacking technique for prediction of diabetes. Comput Biol Med. 135, 104554, (2021). https://doi.org/10.1016/J.COMPBIOMED.2021.104554.
25. Zeller, R., Williamson, A., Friswell, R.: The effect of sleep-need and time-on-task on driver fatigue. Transp. Res. Part F Traffic Psychol. Behav. 74, 15–29 (2020). https://doi.org/10.1016/j.trf.2020.08.001.
26. National Highway Traffic Safety Administration (NHTSA): Drowsy Driving | NHTSA, https://www.nhtsa.gov/risky-driving/drowsy-driving, last accessed 2025/08/13.
27. Tefft, B.C.: Prevalence of motor vehicle crashes involving drowsy drivers, United States, 2009-2013. AAA Foundation for Traffic Safety Washington, DC (2014). https://aaafoundation.org/prevalence-motor-vehicle-crashes-involving-drowsy-drivers-united-states-2009-2013.

28. Jaiswal, A., Zadeh, M.Z., Hebri, A., Makedon, F.: Assessing Fatigue with Multimodal Wearable Sensors and Machine Learning, http://arxiv.org/abs/2205.00287, (2022). https://doi.org/10.48550/arXiv.2205.00287.
29. Jones, A.W.: Alcohol, its analysis in blood and breath for forensic purposes, impairment effects, and acute toxicity. WIREs Forensic Sci. 1, e1353 (2019). https://doi.org/10.1002/wfs2.1353.
30. Lidestam, B., Thorslund, B., Selander, H., Näsman, D., Dahlman, J.: In-Car Warnings of Emergency Vehicles Approaching: Effects on Car Drivers' Propensity to Give Way. Front. Sustain. Cities. 2, (2020). https://doi.org/10.3389/frsc.2020.00019.
31. Hsiao, H., Chang, J., Simeonov, P.: Preventing Emergency Vehicle Crashes: Status and Challenges of Human Factors Issues. Hum. Factors. 60, 1048–1072 (2018). https://doi.org/10.1177/0018720818786132.
32. Li, Z., Li, S.E., Li, R., Cheng, B., Shi, J.: Online Detection of Driver Fatigue Using Steering Wheel Angles for Real Driving Conditions. Sensors. 17, 495 (2017). https://doi.org/10.3390/s17030495.
33. Gormley, M., Walsh, T., Fuller, R.: Risks in the driving of emergency service vehicles. Ir. J. Psychol. 29, 7–18 (2008). https://doi.org/10.1080/03033910.2008.10446270.
34. May, J.F., Baldwin, C.L.: Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies. Transp. Res. Part F Traffic Psychol. Behav. 12, 218–224 (2009). https://doi.org/10.1016/j.trf.2008.11.005.
35. Patterson, P.D., Higgins, J.S., Van Dongen, H.P.A., Buysse, D.J., Thackery, R.W., Kupas, D.F., Becker, D.S., Dean, B.E., Lindbeck, G.H., Guyette, F.X., Penner, J.H., Violanti, J.M., Lang, E.S., Martin-Gill, C.: Evidence-Based Guidelines for Fatigue Risk Management in Emergency Medical Services. Prehosp. Emerg. Care. 22, 89–101 (2018). https://doi.org/10.1080/10903127.2017.1376137.
36. Patterson, P.D., Robinson, K., National Association of State EMS Officials: Fatigue in Emergency Medical Services Systems. (2019). https://doi.org/10.21949/1525960.
37. Albadawi, Y., Takruri, M., Awad, M.: A Review of Recent Developments in Driver Drowsiness Detection Systems. Sensors. 22, 2069 (2022). https://doi.org/10.3390/s22052069.
38. Qing, W., BingXi, S., Bin, X., Junjie, Z.: A PERCLOS-Based Driver Fatigue Recognition Application for Smart Vehicle Space. In: 2010 Third International Symposium on Information Processing. pp. 437–441 (2010). https://doi.org/10.1109/ISIP.2010.116.
39. Li, Q.: Advancements in driver fatigue detection: A comprehensive analysis of eye movement and facial feature approaches. Appl. Comput. Eng. 65, 75–80 (2024). https://doi.org/10.54254/2755-2721/65/20240471.
40. Chang, R.C.-H., Wang, C.-Y., Chen, W.-T., Chiu, C.-D.: Drowsiness Detection System Based on PERCLOS and Facial Physiological Signal. Sensors. 22, 5380 (2022). https://doi.org/10.3390/s22145380.

41. Makhmudov, F., Turimov, D., Xamidov, M., Nazarov, F., Cho, Y.-I.: Real-Time Fatigue Detection Algorithms Using Machine Learning for Yawning and Eye State. Sensors. 24, 7810 (2024). https://doi.org/10.3390/s24237810.
42. Vijayalakshmi, S., Elavendhan, K.P., Suriya, E., Harrisasthaa, G.: Drowsiness Detection Using CNN-LSTM. Int. J. Res. Appl. Sci. Eng. Technol. 13, 79–87 (2025). https://doi.org/10.22214/ijraset.2025.66717.
43. Halomoan, J., Ramli, K., Sudiana, D., Gunawan, T.S., Salman, M.: A New ECG Data Processing Approach to Developing an Accurate Driving Fatigue Detection Framework with Heart Rate Variability Analysis and Ensemble Learning. Information. 14, 210 (2023). https://doi.org/10.3390/info14040210.
44. Nasri, I., Karrouchi, M., Kassmi, K., Messaoudi, A.: A Review of Driver Drowsiness Detection Systems: Techniques, Advantages and Limitations, http://arxiv.org/abs/2206.07489, (2022). https://doi.org/10.48550/arXiv.2206.07489.
45. Ramzan, M., Khan, H.U., Awan, S.M., Ismail, A., Ilyas, M., Mahmood, A.: A Survey on State-of-the-Art Drowsiness Detection Techniques. IEEE Access. 7, 61904–61919 (2019). https://doi.org/10.1109/ACCESS.2019.2914373.
46. Bamidele, A.A., Kamardin, K., Abd Aziz, N.S.N., Sam, S.M., Ahmed, I.S., Azizan, A., Bani, N.A., Kaidi, H.M.: Non-intrusive driver drowsiness detection based on face and eye tracking. Int. J. Adv. Comput. Sci. Appl. 10, (2019).
47. Jiao, Y., Zhang, C., Chen, X., Fu, L., Jiang, C., Wen, C.: Driver Fatigue Detection Using Measures of Heart Rate Variability and Electrodermal Activity. IEEE Trans. Intell. Transp. Syst. 25, 5510–5524 (2024). https://doi.org/10.1109/TITS.2023.3333252.
48. Du, G., Zhang, L., Su, K., Wang, X., Teng, S., Liu, P.X.: A Multimodal Fusion Fatigue Driving Detection Method Based on Heart Rate and PERCLOS. IEEE Trans. Intell. Transp. Syst. 23, 21810–21820 (2022). https://doi.org/10.1109/TITS.2022.3176973.
49. Zhu, S.: An Evtol Driver Fatigue Detection Model Based on Multimodal Fusion of Facial Features and Physiological Signals. J. Int. Nat. Sci. Stud. Vol. 2, (2025).
50. Hu, S., Gao, Q., Xie, K., Wen, C., Zhang, W., He, J.: Efficient detection of driver fatigue state based on all-weather illumination scenarios. Sci. Rep. 14, 17075 (2024). https://doi.org/10.1038/s41598-024-67131-5.
51. Cao, S., Feng, P., Kang, W., Chen, Z., Wang, B.: Optimized driver fatigue detection method using multimodal neural networks. Sci. Rep. 15, 12240 (2025). https://doi.org/10.1038/s41598-025-86709-1.
52. Samiee, S., Azadi, S., Kazemi, R., Nahvi, A., Eichberger, A.: Data fusion to develop a driver drowsiness detection system with robustness to signal loss. Sensors. 14, 17832–17847 (2014). https://doi.org/10.3390/s140917832.
53. Sun, W., Zhang, X., Peeta, S., He, X., Li, Y.: A Real-Time Fatigue Driving Recognition Method Incorporating Contextual Features and Two Fusion Levels. IEEE Trans. Intell. Transp. Syst. 18, 3408–3420 (2017). https://doi.org/10.1109/TITS.2017.2690914.

54. Barnum, G., Talukder, S., Yue, Y.: On the Benefits of Early Fusion in Multimodal Representation Learning, http://arxiv.org/abs/2011.07191, (2020). https://doi.org/10.48550/arXiv.2011.07191.
55. Huang, K., Shi, B., Li, X., Li, X., Huang, S., Li, Y.: Multi-modal Sensor Fusion for Auto Driving Perception: A Survey, http://arxiv.org/abs/2202.02703, (2024). https://doi.org/10.48550/arXiv.2202.02703.
56. Fayyad, J., Jaradat, M.A., Gruyer, D., Najjaran, H.: Deep Learning Sensor Fusion for Autonomous Vehicle Perception and Localization: A Review. Sensors. 20, 4220 (2020). https://doi.org/10.3390/s20154220.
57. Bird, J.J., Faria, D.R., Premebida, C., Ekárt, A., Vogiatzis, G.: Look and Listen: A Multi-modality Late Fusion Approach to Scene Classification for Autonomous Machines. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 10380–10385 (2020). https://doi.org/10.1109/IROS45743.2020.9341557.
58. Kuhn, C.B., Hofbauer, M., Petrovic, G., Steinbach, E.: Introspective Failure Prediction for Autonomous Driving Using Late Fusion of State and Camera Information. IEEE Trans. Intell. Transp. Syst. 23, 4445–4459 (2022). https://doi.org/10.1109/TITS.2020.3044813.
59. Ma, M., Ren, J., Zhao, L., Testuggine, D., Peng, X.: Are Multimodal Transformers Robust to Missing Modality? Presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2022).
60. Raju, S.: yawn_eye_dataset_new, https://www.kaggle.com/datasets/serena-raju/yawn-eye-dataset-new, (2020).
61. Fahmy, H.: Drowsiness Detection Dataset, https://www.kaggle.com/datasets/hazemfahmy/openned-closed-eyes, (2023).
62. Xu, J.: EEG driver fatigue detection, https://www.kaggle.com/datasets/jcxuitsme/eeg-driver-fatigue-detection, (2023). https://doi.org/10.6084/m9.figshare.5202739.v1.
63. Meteier, Q.: A dataset on the physiological state and behavior of drivers in conditionally automated driving. Data Brief. 47, 109027, (2023). https://doi.org/10.1016/J.DIB.2023.109027.
64. Munday, J.A.: Instrumentation and electrode placement. Respir. Care Clin. N. Am. 11, 605–615, viii (2005). https://doi.org/10.1016/j.rcc.2005.08.004.
65. Kanske, P., Kotz, S.A.: Effortful control, depression, and anxiety correlate with the influence of emotion on executive attentional control. Biol. Psychol. 91, 88–95 (2012). https://doi.org/10.1016/j.biopsycho.2012.04.007.
66. Nunez, P.L., Srinivasan, R.: Electric Fields of the Brain: The Neurophysics of EEG. Oxford University Press (2006).
67. Tsagris, M., Pandis, N.: Multicollinearity. Am. J. Orthod. Dentofacial Orthop. 159, 695–696 (2021). https://doi.org/10.1016/j.ajodo.2021.02.005.
68. Karim, R.U., Mahdi, S., Samin, A., Zereen, A.N., Abdullah-Al-Wadud, M., Uddin, J.: Optimizing Stroke Recognition With MediaPipe and Machine Learning: An Explainable AI Approach for Facial Landmark Analysis. IEEE Access. 13, 32636-32660, (2025). https://doi.org/10.1109/ACCESS.2025.3550577.

69. Kuwahara, A., Nishikawa, K., Hirakawa, R., Kawano, H., Nakatoh, Y.: Eye fatigue estimation using blink detection based on Eye Aspect Ratio Mapping(EARM. Cogn. Robot. 2, 50-59, (2022). https://doi.org/10.1016/J.COGR.2022.01.003.
70. Mehta, S., Dadhich, S., Gumber, S., Bhatt, A.J.: Real-Time Driver Drowsiness Detection System Using Eye Aspect Ratio and Eye Closure Ratio. SSRN Electron. J. (2019). https://doi.org/10.2139/SSRN.3356401.
71. Lal, S.K.L., Craig, A.: A critical review of the psychophysiology of driver fatigue. Biol. Psychol. 55, 173–194 (2001). https://doi.org/10.1016/S0301-0511(00)00085-5.
72. Gevins, A., Smith, M.E.: Neurophysiological measures of cognitive workload during human-computer interaction. Theor. Issues Ergon. Sci. (2003). https://doi.org/10.1080/14639220210159717.
73. Mekruksavanich, S., Hnoohom, N., Phaphan, W., Jitpattanakul, A.: A Comparative Study of Validation Methods for Sensor-Based Human Activity Recognition Using Deep Learning Models. In: 10th International Conference on Digital Arts, Media and Technology, DAMT 2025 and 8th ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering. pp. 686-691,. , NCON (2025). https://doi.org/10.1109/ECTIDAMTNCON64748.2025.10962074.
74. Strle, G., Kosir, A., Sodnik, J., Pececnik, K.S.: Physiological Signals as Predictors of Cognitive Load Induced by the Type of Automotive Head-Up Display. IEEE Access. 11, 87835-87848, (2023). https://doi.org/10.1109/ACCESS.2023.3305383.
75. Salih, A.M., Wang, Y.: Are Linear Regression Models White Box and Interpretable? (2024). https://doi.org/10.48550/ARXIV.2407.12177.
76. Liu, X.: Comparison of different machine learning models: Linear model, forest, and SVM. Appl. Comput. Eng. 51, 225–230 (2024). https://doi.org/10.54254/2755-2721/51/20241467.
77. Nadh, B.S.S., Sha, A., Addanki, V., Yerramreddy, D.R., S, N.K.: Performance Analysis of Machine Learning-Based Scheduling Policies for Operating Systems. 2023 7th Int. Conf. Electron. Commun. Aerosp. Technol. ICECA. 771–776 (2023). https://doi.org/10.1109/ICECA58529.2023.10395782.
78. Bony, M.N.V.A., Das, P., Pervin, T., Akter, S., Anjum, N., Alam, M., Erie, PA.: COMPARATIVE PERFORMANCE ANALYSIS OF MACHINE LEARNING ALGORITHMS FOR BUSINESS INTELLIGENCE: A STUDY ON CLASSIFICATION AND REGRESSION MODELS. Frontline Mark. Manag. Econ. J. 04, 72–92 (2024). https://doi.org/10.37547/marketing-fmmej-04-11-06.
79. Sahu, A., Das, P.K., Meher, S., Panda, R., Abraham, A.: An Efficient Deep Learning-Based Breast Cancer Detection Scheme with Small Datasets. Presented at the Cham (2023). https://doi.org/10.1007/978-3-031-35510-3_5.
80. Saleem, M., Senan, N., Wahid, F., Aamir M., Samad, A., Khan, M.: Comparative Analysis of Recent Architectures of Convolutional Neural Network. Math. Probl. Eng. 2022, (2022). https://doi.org/10.1155/2022/7313612.
81. Xin, H.J., Shamsuddin, S., Kamat, S.R., Ibrahim, M.S., Setiawan, R.: Investigation on Factors Affecting Cognitive Skills in Detection of Driving Fatigue. J. Adv. Res. Appl. Sci. Eng. Technol. 51, 270-280, (2025). https://doi.org/10.37934/ARASET.51.2.270280.

82. Ansari, S., Du, H., Naghdy, F., Stirling, D.: Automatic driver cognitive fatigue detection based on upper body posture variations. Expert Syst Appl. 203, 117568, (2022). https://doi.org/10.1016/J.ESWA.2022.117568.
83. Miao, X., Xue, C., Li, X., Yang, L., Miao, X., Xue, C., Li, X., Yang, L.: A Real-Time Fatigue Sensing and Enhanced Feedback System. Information. 13, (2022). https://doi.org/10.3390/info13050230.
84. Gupta, B.B., Gaurav, A., Chui, K.T., Arya, V.: Deep Learning Model for Driver Behavior Detection in Cyber-Physical System-Based Intelligent Transport Systems. IEEE Access. 12, 62268-62278, (2024). https://doi.org/10.1109/ACCESS.2024.3393909.
85. Dziuda, Ł.: Evaluation of a Fatigue Detector Using Eye Closure-Associated Indicators Acquired from Truck Drivers in a Simulator Study. Sensors. 21, 6449, (2021). https://doi.org/10.3390/S21196449.
86. Wang, L., Johnson, D., Lin, Y.: Using EEG to detect driving fatigue based on common spatial pattern and support vector machine. Turk. Jour-Nal Electr. Eng. Comput. Sci. 29, 1429-1444, (2021). https://doi.org/10.3906/elk-2008-83.
87. Hasan, M.M., Islam, M.N., Khandaker, S., Sulaiman, N., Islam, A., Hossain, M.M.: Ensemble-based machine learning models for vehicle drivers' fatigue state detection utilizing EEG signals. Facta Univ. - Ser. Electron. Energ. 37, 671-686, (2024). https://doi.org/10.2298/FUEE2404671H.
88. Rehman, S.U., Ali, A., Khan, A.M., Okpala, C.: Human Activity Recognition: A Comparative Study of Validation Methods and Impact of Feature Extraction in Wearable Sensors. Algorithms. 17, 556, (2024). https://doi.org/10.3390/A17120556.
89. Lee, G., Lee, S.: Importance of Testing with Independent Subjects and Contexts for Machine-Learning Models to Monitor Construction Workers' Psychophysiological Responses. J Constr Eng Manag. 148, 04022082, (2022). https://doi.org/10.1061/(ASCE)CO.1943-7862.0002341.
90. Lu, Y., Fridlyand, J., Tang, T., Qi, T., Simon, N., Leng, N.: The Future will be Different than Today: Model Evaluation Considerations when Developing Translational Clinical Biomarker. (2021).
91. Kucuk, K., Yurteri, E., Semiz, B.: Electroencephalography Analysis Frameworks for the Driver Fatigue Problem: A Benchmarking Study. (2025).
92. Wei, C.S., Lin, Y.P., Wang, Y.T., Lin, C.T., Jung, T.P.: A subject-transfer framework for obviating inter- and intra-subject variability in EEG-based drowsiness detection. Neuroimage. 174, 407-419, (2018). https://doi.org/10.1016/j.neuroimage.2018.03.032.
93. Saha, S., Baumert, M.: Intra- and Inter-subject Variability in EEG-Based Sensorimotor Brain Computer Interface: A Review. In: Front Comput Neuro-sci (2020).
94. Yuan, D., Yue, J., Xiong, X., Jiang, Y., Zan, P., Li, C.: A regression method for EEG-based cross-dataset fatigue detection. Front Physiol. 14, (2023). https://doi.org/10.3389/FPHYS.2023.1196919.
95. Feng, X., Guo, Z., Kwong, S.: ID3RSNet: cross-subject driver drowsiness detection from raw single-channel EEG with an interpretable residual shrinkage network.

Front Neurosci. 18, 1508747, (2025). https://doi.org/10.3389/FNINS.2024.1508747.
96. Doberenz, S., Roth, W.T., Wollburg, E., Maslowski, N.I., Kim, S.: Methodological considerations in ambulatory skin conductance monitoring. Int. J. Psychophysiol. 80, 87–95 (2011). https://doi.org/10.1016/j.ijpsycho.2011.02.002.
97. Sammito, S., Thielmann, B., Böckelmann, I.: Update: factors influencing heart rate variability–a narrative review. Front. Physiol. 15, (2024). https://doi.org/10.3389/fphys.2024.1430458.
98. Srinivasan, A.G., Smith, S.S., Pattinson, C.L., Mann, D., Sullivan, K., Salmon, P., Soleimanloo, S.S.: Heart rate variability as an indicator of fatigue: A structural equation model approach. Transp. Res. Part F Traffic Psychol. Behav. 103, 420–429 (2024). https://doi.org/10.1016/j.trf.2024.04.015.
99. Tarafder, S., Badruddin, N., Yahya, N., Nasution, A.H.: Drowsiness Detection Using Ocular Indices from EEG Signal. Sensors. 22, 4764, (2022). https://doi.org/10.3390/S22134764.
100. Bhanja, A., Parhi, D., Gajendra, D., Sinha, K., Sahoo, A.K.: Driver drowsiness shield (DDSH): a real-time driver drowsiness detection system. ROBOMECH J. 12, 1-11, (2025). https://doi.org/10.1186/S40648-025-00307-4/FIGURES/9.