Indonesian Journal of Electrical Engineering and Computer Science
Vol. 39, No. 1, July 2025, pp. 592~602
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v39.i1.pp592-602

Real-time driver drowsiness detection based on integrative approach of deep learning and machine learning model

Gowrishankar Shiva Shankara Chari, Jyothi Arcot Prashant
Department of Computer Science and Engineering, M.S. Ramaiah University of Applied Sciences, Bengaluru, India

Article Info

Article history:

Received Jan 15, 2025
Revised Mar 12, 2025
Accepted Mar 26, 2025

Keywords:

Deep learning
Driver drowsiness detection
Facial landmark prediction
Multi-stage decision fusion
Road accident

ABSTRACT

Driver drowsiness is a major factor that contributing to road accidents. Several researches are ongoing to detect driver drowsiness, but they suffer from the complexity and cost of the models. This paper introduces a hybrid artificial intelligence (AI)-driven framework integrating deep learning (DL) and machine learning (ML) models for real-time drowsiness detection. The system utilizes a robust DL model to classify driver states based on facial images and support vector machine (SVM) model is trained to develop a cost-efficient yet robust facial landmark detector to extract key features such as eye aspect ratio (EAR) and mouth aspect ratio (MAR). We also introduce a multi-stage decision fusion mechanism that combines convolutional neural network (CNN) probability scores with EAR/MAR thresholds to enhance detection reliability and reduce false positives. Experimental results demonstrate that the proposed model achieves 98% accuracy and F1-score, significantly outperforming traditional DL approaches. Additionally, the SVM-based landmark predictor shows improved efficiency with lower mean squared error (MSE) without having higher computational requirements.

Corresponding Author:

Gowrishankar Shiva Shankara Chari
Department of Computer Science and Engineering, M.S. Ramaiah University of Applied Sciences
Bengaluru, India
Email: shankar.iq@gmail.com

1. INTRODUCTION

The road safety remains a significant concern worldwide due to increasing traffic congestion and the growing number of vehicles on the roads [1], [2]. One of the major causes of road accidents is driver fatigue and drowsiness, which reduces alertness, slows down reaction time, and impairs decision-making ability [3], [4]. Several methods have been developed to detect driver drowsiness using physiological signals, behavioral patterns, and facial expressions also based on focused on monitoring steering patterns, lane deviation, and head movements [5], [6]. However, these approaches often produce false positives because external factors like road conditions, weather changes, and vehicle types can influence driving behavior. Common approach of physiological monitoring, which includes electroencephalography (EEG), heart rate variability (HRV), and eye-tracking sensors are highly accurate in controlled environments but are expensive, intrusive, and impractical for real-world applications because they require specialized hardware [7]. The work carried out by Vicente *et al.* [8] and Sahayadhas *et al.* [9] explored HRV and combined physiological and behavioral measures for drowsiness detection, achieving improved accuracy in controlled settings. However, real-time applicability is limited due to sensor placement issues and physiological variability. Shahbakhti *et al.* [10] enhanced EEG-based detection by integrating eye blink features with band power analysis, but EEG signals remain highly susceptible to noise and movement artifacts. Similarly, Fujiwara *et al.* [11] developed a self-attention autoencoder using electrocardiography (ECG) R-R intervals, demonstrating high sensitivity, but real-world monitoring remains challenging due to motion-induced distortions. Advancements in deep learning (DL) and machine learning (ML) have significantly improved facial-based drowsiness detection. Bai *et al.* [12] proposed a spatial-temporal graph convolutional neural network (CNN) to address illumination changes and occlusions, but its high computational cost limits real-time feasibility. Florez *et al.* [13] evaluated CNN architectures such as InceptionV3, VGG16, and ResNet50V2, achieving 99.71% accuracy on the NITYMED dataset, yet its focus on nighttime driving restricts generalization. Maior *et al.* [14] utilized eye aspect ratio (EAR) with ML classifiers, performing well on the DROZY dataset, but lacked real-time adaptability. Other approaches, such as fuzzy logic-based feature detection by AlKishri *et al.* [15] and support vector machine (SVM)-based drowsiness detection by Shukla *et al.* [16], demonstrated improvements but were outperformed by CNN-based models on larger datasets. To further enhance efficiency, Biswal *et al.* [17] developed an IoT-based real-time facial landmark detection system, though deployment challenges remain. Lamaazi *et al.* [18] introduced edge-based drowsiness detection, ensuring privacy and reduced latency, but edge computing limitations constrain DL model complexity. Zhang *et al.* [19] tackled this by implementing federated transfer learning, enabling distributed training while preserving privacy. Ahmed *et al.* [20] improved accuracy using a dual InceptionV3 model with facial subsampling, yet high computational demands introduced inference delays, limiting real-time usability. Jebraeily *et al.* [21] further optimized CNN models using genetic algorithms, enhancing performance but increasing computational overhead. Efforts to optimize hardware efficiency have also been explored. Nguyen *et al.* [22] developed a miniaturized EEG-based system with tiny neural networks, reducing processing delays but still limited by EEG signal reliability in real-world driving. Mousavikia *et al.* [23] accelerated DL inference on FPGA, achieving faster computation and lower power consumption, but FPGA-based implementations require costly hardware modifications. Lastly, Madni *et al.* [24] leveraged transfer learning with eye movement behavior analysis, combining visual geometry group (VGG)-16 with a light gradient-boosting classifier, yet its real-world effectiveness requires further validation.

Despite advancements in physiological, behavioral, AI-based drowsiness detection, existing approaches face several limitations that reduce their effectiveness for real-time driver monitoring in real-world conditions. One major challenge is that many DL-based models are trained on limited datasets, that do not adequately represent variations in lighting, facial orientations, or ethnic diversity, or environmental factors. As a result, these models struggle with occlusions (e.g., glasses, masks, poor lighting) and head movements. Additionally, many studies use either DL or feature-based approaches independently, without an integrated fusion mechanism that combines both methodologies for enhanced accuracy. Lastly, the existing models often prioritize accuracy over computational efficiency, which are not suitable for real-time deployment in automotive applications where resource-efficient processing is essential.

This paper introduces a hybrid AI-driven framework that integrates DL-based image classification with a lightweight yet robust SVM-based facial landmark detector to enhance accuracy while maintaining computational efficiency for real-time driver monitoring. The articulation of research gap is as follows: The detection of driver drowsiness has drawn a lot of interest lately, but many of the current techniques have drawbacks, including large false positive rates, insensitivity to occlusions, and poor accuracy in practical settings. Eye closure (EAR) and yawning (MAR), two independent facial metrics that are sensitive to fleeting facial alterations and ambient conditions (such as lighting and occlusions), are commonly used in simple threshold-based drowsiness detection systems. Furthermore, a lot of DL-based systems don't integrate multi-modal data sources or incorporate different face signals to make better decisions.

In order to overcome these difficulties, this study suggests a hybrid AI-driven architecture that incorporates a multi-stage decision fusion mechanism, ML-driven facial landmark tracking, and DL-based facial state prediction. The main contributions are:

- Multi-stage fusion approach: to reduce false alarms brought on by fleeting facial changes, we incorporate EAR, MAR, and the CNN-based probability score into a counter-based decision fusion system. This method guarantees that an alert will only be triggered by persistent signs of drowsiness, such as extended eye closing, frequent yawning, or poor confidence in the CNN forecast.
- Adaptability to environmental variability: by combining several indications that can identify tiredness through various aspects of facial behavior, the model successfully manages occlusions (such as masks or sunglasses) and changing illumination conditions.
- Real-time performance: the system is appropriate for practical implementation in embedded systems for driver monitoring since it delivers real-time accuracy (98%) and robustness with low computational cost.
- Increased accuracy and robustness: the suggested fusion methodology guarantees sustained detection, in contrast to other techniques that depended on a single metric or threshold, making the system for real-time sleepiness detection more dependable and accurate. The novelty of the proposed system is the integration of a light-weight facial landmark detection module with a robust DL-based prediction model, which enables accurate drowsiness detection under complex real-world conditions using facial features.

2. METHOD

This section presents the system design and implementation procedures for the proposed hybrid AI-driven framework inspired from the findings of our prior work [25]. The system integrates DL-based image classification with a lightweight SVM-based facial landmark detector for accurate and real-time driver drowsiness detection. In order to ensure real-time detection, the framework considers both DL predictions and facial landmark tracking operate simultaneously. The workflow of the proposed integrative system is shown in Figure 1.

As shown in Figure 1, the input to the proposed framewok is the real-time video feeds aptured from a driver-facing camera. Each video frame is extracted sequentially, with every third frame sampled to reduce computational load while maintaining detection efficiency. The extracted frames are resized to match the input size of the DL model and then processed in parallel threads i.e., one for DL-based drowsiness classification and the other for SVM-based facial landmark detection. The DL module implements efficientNet CNN model classifies each frame as drowsy or alert by analyzing facial features. Simultaneously, the SVM-based landmark detector facial landmarks for a given frame. The system then computes EAR and MAR feature to measure eye closure and yawning frequency, which are critical indicators of drowsiness. In order to ensure detection reliability, the system employs a multi-stage decision fusion mechanism, which combines DL-based probability scores with EAR and MAR thresholds. If drowsiness indicators persist for consecutive frames, an alert is generated, and promp driver to take action. Additionally, the alert data is stored in a cloud-based system, which enables post-drive review via a web application. This approach ensures that drivers can analyze their drowsiness history and make necessary adjustments to their driving behavior. The proposed framework is designed to be computationally efficient, resource-friendly, and scalable, thereby making it suitable for real-world deployment scenario with very minimal false positives and negatives.

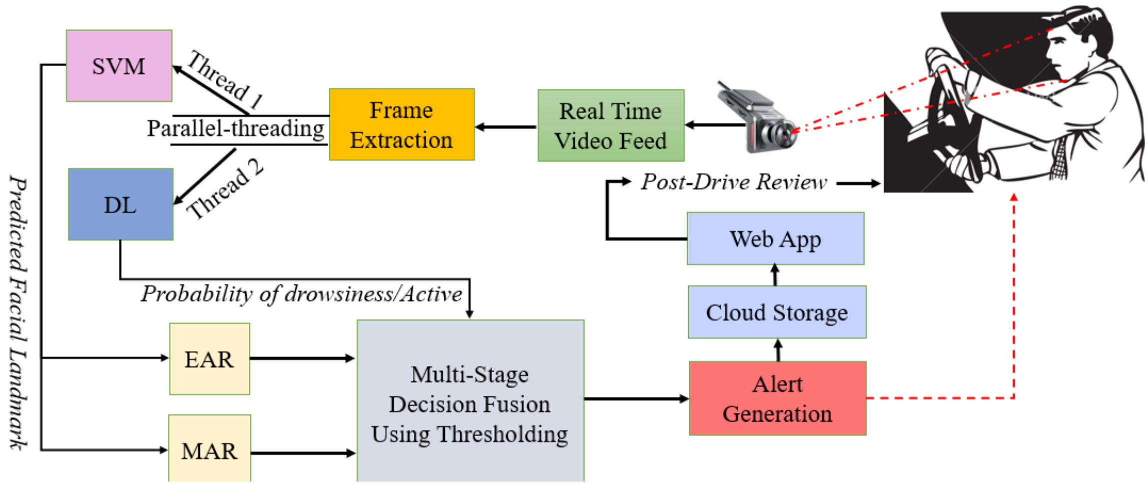


Figure 1. Illustrates workflow of proposed integrative system

2.1. Dataset description

This study created a custom dataset by scraping 6,000 images from sources like Google and Bing, and adding 12,000 images from Kaggle. The images were manually reviewed, and 10,444 relevant images were selected, excluding those with yawning to focus on facial features like eye closure and head position. The images were resized to 224×224 pixels and split into training, validation, and testing sets in a 80:10:10 ratio, with random shuffling to avoid order bias. The dataset consist of training set of 4,284 and drowsy set consists of 4,077 images whereas there are 531 images for active state and 509 images for drowsy state for validation dataset. Finally, testing images consists of 532 images for active state and 511 images for drowsy state.

2.2. DL based driver drowsiness prediction

The proposed system employs EfficientNetV2B0, an optimized DL architecture designed for efficient feature extraction and classification. EfficientNetV2B0 is a lightweight CNN that applies mobile bottleneck convolutions (MBConv) and inverted residual blocks to achieve optimal accuracy with minimal computational cost. In the proposed work, this learning model is enhanced with additional custom layers to improve drowsiness detection accuracy and fine-tuned for binary classification (drowsy vs. active) using a two-phase training strategy to balance model generalization and task-specific adaptation. Table 1 summarizes the custom DL model adopted for driver drowsiness prediction using facial image.

In the proposed system, driver drowsiness prediction is formulated as a binary classification problem, where the model determines whether a driver is drowsy ($Y=1$) or alert ($Y=0$) based on facial images. For a given image X, the model computes the probability of drowsiness using the sigmoid activation function, such that $P(Y=1|X)=\sigma(W^T f(X)+b)$, where, W and b are the trainable parameters called weights and bias, respectively, $f(X)$ represents the feature vector extracted from the final dense layer, where $\sigma(z)$ ensures that the output is a probability score in the range [0,1]. The model is trained using the binary cross-entropy loss function, which is defined as:

$$[\mathcal{L}=-\frac{1}{N}\sum_{i=1}^{N}[y_i\log(\hat{y}_i)+(1-y_i)\log(1-\hat{y}_i)]] \tag{1}$$

Where, y_i represents the ground truth label (1 for drowsy, 0 for alert), $\hat{y}_i$ is the predicted probability and N is the total number of the training samples. The training strategy incorporates data augmentation and a two-phase learning process to enhance model generalization and performance. In the initial training phase (feature extraction phase), the base layers of the DL model are frozen to retain the pre-trained features from ImageNet, while only the custom layers (dense, dropout, and output layer) are trained. The model undergoes 20 epochs of training with an initial learning rate of 0.001. In the fine-tuning phase (task-specific adaptation phase), the last 20 layers of DL model (EfficientNetV2B0) are unfrozen to allow task-specific adjustments while leveraging pre-trained knowledge. In order to optimize model performance, the following hyperparameters were fine-tuned as illustrated in Table 2.

Table 1. The architectural details of the custom DL

Layer type	Input shape	Filter size	Number of layers	Output shape
Input layer	(224, 224, 3)	-	1	(224, 224, 3)
Stem Conv2D (Base)	(224, 224, 3)	3x3	1	(112, 112, 32)
MBConv1_3x3 (Base)	(112, 112, 32)	3x3	1	(112, 112, 16)
MBConv6_3x3 (Base)	(112, 112, 16)	3x3	2	(56, 56, 24)
MBConv6_5x5 (Base)	(56, 56, 24)	5x5	2	(28, 28, 40)
MBConv6_3x3 (Base)	(28, 28, 40)	3x3	3	(14, 14, 80)
MBConv6_5x5 (Base)	(14, 14, 80)	5x5	3	(7, 7, 112)
MBConv6_3x3 (Base)	(7, 7, 112)	3x3	4	(7, 7, 192)
Conv2D (Custom)	(7, 7, 192)	3x3	1	(7, 7, 128)
GlobalAveragePooling2D	(7, 7, 128)	-	1	(128)
Dense layer	(128)	-	1	(1024)
Dropout	(1024)	-	1	(1024)
Dense layer	(1024)	-	1	(128)
Output layer (Dense)	(128)	-	1	(1)

Table 2. The details of hyperparameter optimization

Hyperparameter	Value	Purpose
Batch size	32	Ensures stable training and prevents memory overload
Optimizer	Adam	Adaptive learning rate for efficient weight updates
Initial learning rate	0.001	Faster convergence in the initial training phase
fine-tuning learning rate	0.0001	Prevents sudden weight changes during fine-tuning
Dropout rate	0.5	Reduces overfitting in dense layers
Total epochs	50	20 for initial training + 30 for fine-tuning

2.3. ML based facial landmark prediction model

Facial landmark detection plays a crucial role in identifying key facial features such as the eyes and mouth, which serve as indicators of driver vigilance. By accurately detecting these landmarks, the system can compute essential metrics such as the EAR and MAR, which are fundamental in assessing drowsiness and yawning frequency. Although various landmark detection methods exist, but dlib-68 [26], [27], a pre-trained ML model, has been widely adopted due to its lightweight nature and reasonable accuracy across diverse face orientations. However, dlib-68 is highly sensitive to occlusions and lacks adaptability to dynamic facial changes, which affects its robustness in real-world driving scenarios. Therefore, the proposed study develops a lightweight yet robust facial landmark detector utilizing a pre-trained CNN for feature extraction and SVM for landmark prediction. The proposed approach integrates publicly available iBug 300-W face landmark dataset [26], and custom dataset from the previous module used for drowsiness prediction. Since publicly available datasets often lack variations in dynamic conditions, the custom dataset is enriched with diverse facial expressions, occlusions (e.g., masks, sunglasses), and different lighting conditions. Here, Dlib is employed to generate ground-truth facial landmarks for each image in the our custom dataset, and then both datasets (iBug 300-W and our custom image dataset) are combined into a single comprehensive dataset, consisting two classes namely images and labels (containing facial coordinates). Then for training SVM model, the dataset is split into training (80%), validation (10%) and testing (10%) subsets. Afterwards, the study then performs feature extraction operation using a lightweight ResNet-18 model with pre-trained ImageNet weights. Since ResNet-18 is not readily available in TensorFlow, we customize ResNet-50 by reducing the number of layers and removing unnecessary bottleneck layers by replacing three convolutional layers with two to maintain computational efficiency while preserving feature extraction capability. The extracted deep features from ResNet-18 further serve as optimized input representations for training the SVM model to predict landmark coordinates. Facial landmark detection is formulated as a regression problem, where the goal is to predict the x and y coordinates of facial key points based on extracted features. In this regard, for a given input image I the extracted feature vector $\mathrm{F} \in \mathbb{R}^{\mathrm{d}}$ from the CNN model serves as the input to the SVM model that learns a function $f\colon \mathbb{R}^{d} \to \mathbb{R}^{2}$ that maps each input feature vector to a predicted landmark coordinate such that: $\left(\hat{Y} = f(\mathrm{F}) = WF + b\right)$, where, $\hat{Y} = (\hat{x}, \hat{y})$ represents the predicted facial landmark coordinates and W and b are the weights and bias parameters. Since facial landmark prediction involves learning a non-linear mapping, the training of SVM is carried out considering RBF kernel, which helps to transform the feature space into a higher-dimensional representation, such that: $K\left(F_i, F_j\right) = \exp\left(-\gamma \,||F_i - F_j||^2\right)$, where, $K\left(F_i, F_j\right)$ represents the similarity between two feature vectors, γ is a tunable hyperparameter that controls the width of the Gaussian kernel and $||F_i - F_j||^2$ is the squared Euclidean distance between feature vectors. Another hyperparameter considered is the regularization parameter (c) that controls the trade-off between achieving a low error on the training set and minimizing model complexity to prevent overfitting. On the other hand, a tolerance factor ε is considered to specifies the margin of tolerance for prediction errors in regression problems of facial coordinates prediction. Therefore, the objective function for SVM based landmark prediction model is formulated is as follows:

$$[\min_{W,b} \frac{1}{2} ||W||^2 + C \sum_{i=1}^{N} \max\left(0, \left|Y_i - \widehat{Y}_i\right| - \varepsilon\right)] \tag{2}$$

Where, Y_i and $\widehat{Y}_i$ are the ground-truth and predicted landmark coordinates, N is the number of training samples and the term $\max\left(0, \left|Y_i - \widehat{Y}_i\right| - \varepsilon\right)$, enforces the tolerance margin, thereby ensuring that predictions within ε are not penalized. Once SVM model is trained then its is evaluated using the mean squared error (MSE) on test dataset. The predicted landmarks are then used to compute EAR and MAR as follows:

$$EAR = \frac{||P_2 - P_6|| + ||P_3 - P_5||}{2 \times ||P_1 - P_4||} \tag{3}$$

Where, P_1, P_4 denotes horizontal eye landmarks, P_2, P_6 and P_3, P_5 are the vertical eye landmark, $\| P_a - P_b \|$ is the Euclidean distance between points P_a and P_b. Since the EAR remains nearly constant when eyes are open and rapidly decreases during a blink, it serves as an effective indicator for eye closure detection.

$$\mathrm{MAR} = \frac{||P_2 - P_8|| + ||P_3 - P_7|| + ||P_4 - P_6||}{2 \times ||P_1 - P_5||} \tag{4}$$

Where, P1, P5 → horizontal mouth corner points, P2, P8 → vertical upper and lower lip points (outer), P3, P7 → vertical upper and lower lip points (middle), P4, P6 → vertical upper and lower lip points (inner), and ∣Pa−Pb∣ represents the Euclidean distance between two points.

2.4. Multi-stage decision fusion mechanism

To ensure accurate and reliable drowsiness detection, the proposed system implements a multi-stage decision fusion mechanism that combines DL-based predictions with geometric thresholding techniques. This approach minimizes false positives and false negatives by integrating the CNN probability score with EAR and MAR. The system continuously monitors EAR, MAR, and the CNN probability score. If any of these metrics exceed their predefined thresholds for a specified number of consecutive frames, an alert is triggered. The predefined threshold values and conditions for decision fusion are summarized in Table 3. Here, EAR is used to determine prolonged eye closure, MAR is used to detect yawning, and the CNN probability score assesses the overall likelihood of drowsiness based on facial feature analysis. These combined indicators ensure that temporary distractions or brief blinks do not mistakenly trigger alerts, thereby improving the system's robustness.

In addition, a counter-based approach is implemented to track the number of consecutive frames where EAR, MAR, or CNN probability exceed their respective thresholds. Each counter is updated per frame, ensuring that only sustained indicators of drowsiness lead to an alarm. This prevents false alarms caused by momentary changes in facial expression, lighting conditions, or brief head movements. If any of these counters surpass their defined limit, the system generates an alert and logs the event for further analysis.

Table 3. Multi-stage decision fusion thresholds

Metric	Purpose	Threshold	Consecutive frames	Evaluation criteria
EAR	Detects prolonged eye closure	0.26	15	Trigger alert if EAR <0.26 for 15 consecutive frames
MAR	Identifies yawning frequency	0.05	20	Record yawn if MAR >0.05 for 20 consecutive frames
CNN probability	Measures overall likelihood of drowsiness	0.35	15	Trigger alert if CNN probability <0.3 for 15 consecutive frames

3. RESULTS AND DISCUSSION

The design and development of the system was carried out using the Python programming language, and the execution of the model was done on the Anaconda distribution on a Windows 10 machine.

- Threshold selection and justification: both statistical analysis and empirical experimentation were used to determine the criteria for CNN probability, EAR, and MAR. Several threshold values were examined in the early phases of model building to assess their effects on false positives and false negatives. Based on an optimization procedure meant to reduce detection mistakes, these thresholds were chosen. For each of the indicators (EAR, MAR, and CNN probability), we performed a receiver operating characteristic (ROC) curve analysis in order to statistically support the values. Through these evaluations, we were able to choose thresholds that offered the best possible balance between true positive rate (TPR) and false positive rate (FPR). In our test dataset, the best criteria for attaining low false-positive rates and good accuracy were found to be EAR=0.26, MAR=0.05, and CNN probability=0.35. In order to make sure our model's performance was in line with the most advanced techniques in this field, we also examined empirical research that had employed comparable thresholds for facial landmarks and drowsiness detection and modified those results.
- Threshold evaluation: a cross-validation procedure was used to further refine the assessment of these thresholds. We confirmed that the chosen thresholds produced the best possible trade-off between sensitivity and specificity by examining precision-recall curves. The validation phase's empirical data verified that the thresholds selected were appropriate for real-world situations where prompt and precise identification of driver fatigue is essential.

The performance evaluation of the suggested drowsiness prediction model is shown in Figure 2. One of the main justifications for choosing EfficientNetV2B0 is its compound scaling approach, which judiciously maximizes the model's depth, width, and resolution. This allows for high accuracy scaling of the model without a significant increase in computational complexity. Since driver drowsiness detection is a real-time process where accuracy and speed are crucial, EfficientNetV2B0 offers a great balance between the two. The comparison with other models are as follows:

- ResNet: because of its deep architecture and residual connections, ResNet tends to demand greater processing resources even if it is popular and has shown success in DL applications. ResNet might not be as effective as EfficientNetV2B0 in real-time applications such as sleep detection in terms of compute cost and inference time, which is crucial for guaranteeing low-latency answers.
- MobileNet: compared to more sophisticated models, MobileNet is faster and better suited for devices with less resources, but it also loses some accuracy. For this work, EfficientNetV2B0 was chosen over MobileNet due to the requirement for high accuracy in identifying small indicators of drowsiness, such as slight eye closes or changes in head position.
- Vision transformers (ViT): particularly when working with big datasets, vision transformers have demonstrated exceptional performance in picture classification tasks. To function at their best, they are said to need a lot of data and a lot of processing power. EfficientNetV2B0 offers a better option than vision transformers because the study's dataset is relatively small and its objective is to obtain effective, real-time performance.

The confusion matrix, displayed in Figure 2(a), demonstrates the model's accuracy in identifying the driver's status as either active or sleepy. The model has a high true positive and true negative rate, with few false positives and false negatives, according to the confusion matrix. The accuracy of the suggested improved EfficientNet is contrasted with that of other popular CNN architectures, such as DenseNet121, VGG-16, and EfficientNet, in Figure 2(b). As demonstrated, the suggested model outperforms competing architectures by a significant margin and attains the maximum accuracy. The proposed model retains a high F1-score, guaranteeing that both precision and recall are optimized and verifying its robustness in detecting drowsiness, as shown in the comparative study of the F1-score in Figure 2(c). Because of its effectiveness and performance balance, EfficientNetV2B0 was selected as the foundational model for driver drowsiness detection in this investigation. EfficientNetV2B0 is a lightweight CNN architecture that uses inverted residual blocks and mobile bottleneck convolutions (MBConv) to achieve good performance at low computational cost. For real-time applications, this paradigm is perfect, especially in resource-constrained settings like embedded systems or mobile devices, which are common applications for driver monitoring systems. EfficientNetV2B0 is a good option for this application since it provides the best possible balance of speed, accuracy, and efficiency for detecting driver sleepiness. The analysis in Figure 2 demonstrates that the proposed DL model outperforms other DL models, achieving 98% accuracy and F1-score. The confusion matrix confirms that the model effectively distinguishes between drowsy and active states, with minimal misclassifications. When comparing with DenseNet121, VGG-16, and EfficientNet, the fine-tuned EfficientNet exhibits superior performance with reduced false positives, ensuring robust drowsiness detection across diverse facial variations and environmental conditions. Figure 3 demonstrates performance analysis of the proposed SVM based facial landmark detection.

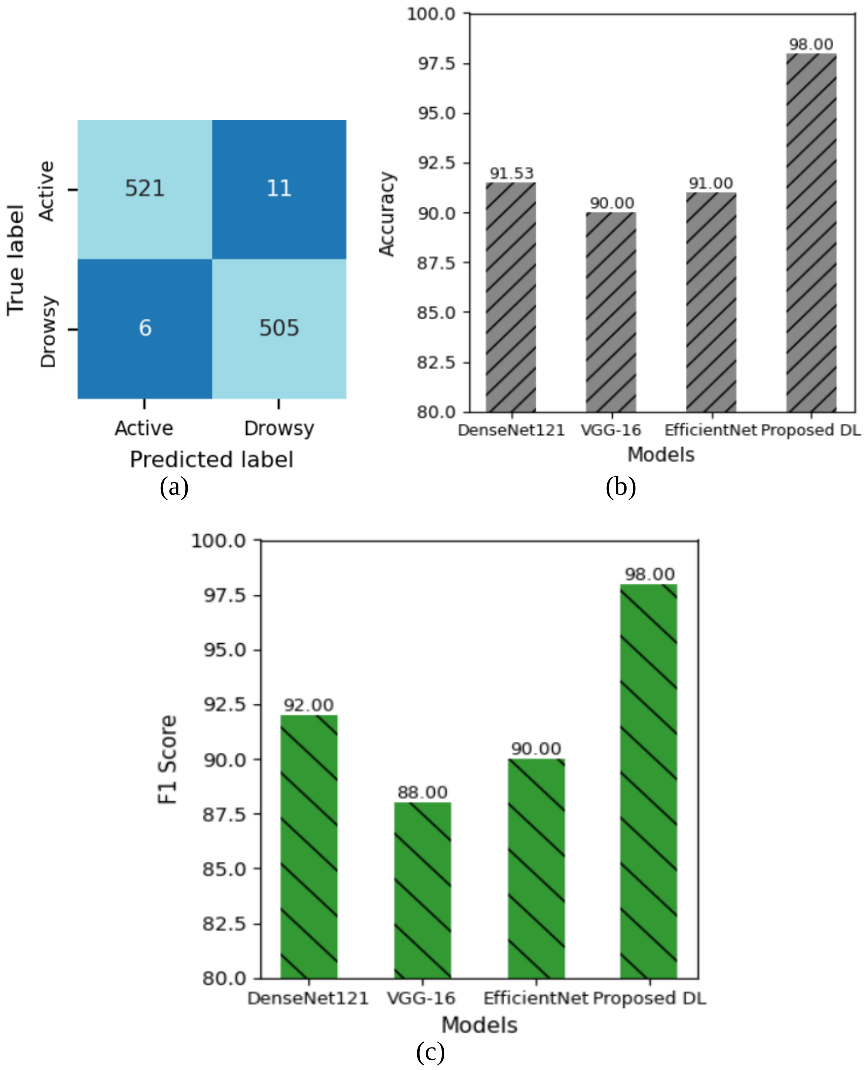


Figure 2. Performance evaluation of the proposed drowsiness prediction model, including: (a) confusion matrix, (b) comparative analysis of accuracy, and (c) comparative analysis of F1-score

The effectiveness of the SVM-based facial landmark prediction model is assessed in Figure 3. The model's ability to correctly anticipate important facial features including the eyes, nose, and mouth is seen in Figure 3(a), where the predicted landmarks are displayed on a sample image. The model's ability to handle variations in face expressions is demonstrated by the good alignment between the predicted and actual facial landmarks. The MSE analysis for the test and validation datasets is shown in Figure 3(b). High accuracy in face landmark prediction is indicated by the low MSE values for both datasets, which are 0.00398 for validation and 0.00397 for testing. Last but not least, Figure 3(c) contrasts the processing times of the suggested SVM model and Dlib. It demonstrates that, despite a minor computational time increase (0.320s vs. 0.280s) for the SVM-based model, this small delay is compensated for by the enhanced accuracy and robustness under various facial conditions.

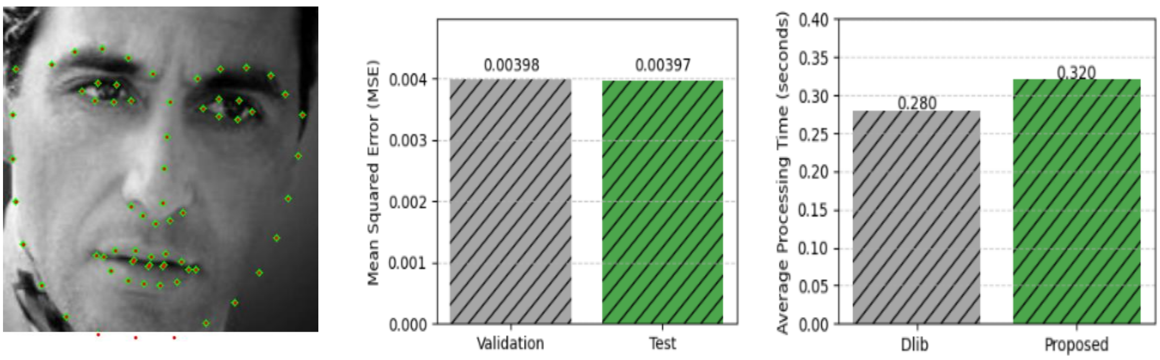
The next Figure 4 shows analysis with respect to real-time testing of the proposed drowsiness detection model on a subject under different conditions. The first image indicates an alert state, where the model correctly identifies no signs of drowsiness. The second image signaling a yawning event and system triggers an alert without wake up. On the other hand, the third image classifying the subject as awake.

We expand the performance analysis to include sensitivity, specificity, and AUC-ROC in addition to precision and recall in order to give a thorough assessment of the suggested approach. These measures provide a more impartial evaluation of the model's performance, especially when it comes to binary classification tasks like detecting driver drowsiness. The major performance indicators for the suggested model are compiled in the accompanying table (Table 4), which also contrasts them with those of other cutting-edge DL models currently in use.

The suggested model successfully balances precision (0.98), recall (0.98), sensitivity (0.98), and specificity (0.98), as demonstrated in Table 4, demonstrating its resilience in accurately identifying both alert and drowsy states. The model's outstanding ability to discriminate between the two classes is further evidenced by its AUC-ROC score of 0.99. The suggested model performs better than other industry-leading techniques like DenseNet121, VGG-16, and efficientNet in every important performance indicator, demonstrating its exceptional efficacy in detecting driver drowsiness.

Key findings of this study are as follows: a hybrid AI-driven framework for real-time driver drowsiness detection is presented in this paper, with an astonishing 98% accuracy rate in differentiating between active and drowsy states. A strong and dependable solution is provided by the multi-stage decision fusion technique, which combines facial landmark detection with DL-based facial analysis. The approach lowers false alarms and guarantees that only persistent sleepiness indicators cause alerts by combining CNN probability ratings, EAR, and mouth aspect ratio (MAR). The suggested approach is now at the forefront of driver safety technology thanks to these findings.

When compared to earlier research, our method overcomes the shortcomings of earlier techniques by recognizing environmental changes and facial occlusions, which frequently make it more difficult to detect drowsiness in real-world situations. Our model uses a more dynamic fusion approach, including both facial geometry information and DL predictions, in contrast to conventional systems that just use basic thresholds or certain facial traits. Under a variety of driving circumstances, this hybrid method improves accuracy and resilience. In summary, this study shows how AI-powered systems might enhance driver safety by precisely identifying tiredness in real time. As technology develops, it may be able to significantly reduce driver fatigue-related collisions, ultimately making roadways safer everywhere.



(a) (b) (c)

Figure 3. Performance evaluation of the proposed SVM based landmark prediction, including: (a) predicted facial landmark, (b) MSE analysis, and (c) processing time analysis

Figure 4. Demonstrates real-time testing of the proposed drowsiness detection model

Table 4. Comparison of performance metrics for different models

Model	Precision	Recall	Sensitivity	Specificity	F1-Score	AUC-ROC
Proposed model	0.98	0.98	0.98	0.98	0.98	0.99
DenseNet121	0.94	0.96	0.95	0.94	0.95	0.97
VGG-16	0.92	0.94	0.93	0.92	0.93	0.96
EfficientNet	0.95	0.97	0.96	0.95	0.96	0.98

4. CONCLUSION

This study suggested a hybrid AI-driven architecture that combines a multi-stage decision fusion mechanism with DL-based predictions for real-time driver sleepiness detection. Through the integration of facial landmark monitoring with a ML-based facial landmark prediction model, the suggested system achieved a high detection accuracy of 98%, providing dependable categorization between drowsy and active states. Additionally, the counter-based method of tracking successive frames guarantees that an alert is only triggered by persistent signs of drowsiness, minimizing false alarms brought on by fleeting head movements or changes in facial expression. The device notifies the driver in real time when an alert is set off, and all instances of drowsiness are recorded for study after the drive.

Future research must address a few shortcomings, even if the suggested approach has great accuracy and resilience in controlled settings. First, harsh environmental conditions like intense glare, poor lighting, or severe weather might alter the model's performance by affecting the accuracy of landmark prediction and the visibility of facial features. Furthermore, the system is primarily dependent on facial landmark detection, which may have trouble detecting occlusions (such as long hair, sunglasses, or face masks). To improve the model's resilience in these difficult circumstances, more investigation is required. Future directions for the system could include adding more sensors, including eye-tracking devices or in-car cameras, to better identify indicators of tiredness when face occlusions are present. In order to facilitate real-time deployment in automotive systems, the model can also be tuned for quicker processing. The creation of adaptive rule-based systems that dynamically modify detection thresholds in response to current environmental conditions, such lighting or the driver's position, may be the subject of future study. Furthermore, broadening the dataset to encompass more varied and authentic driving scenarios, including nighttime or extended rides, may improve the model's capacity for generalization.

FUNDING INFORMATION

The authors state that no funding was involved in this research.

AUTHOR CONTRIBUTIONS STATEMENT

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Gowrishankar Shiva Shankara Chari	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Jyothi Arcot Prashant		✓		✓		✓		✓	✓	✓	✓	✓		

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

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The datasets used and analyzed during this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] D. Albalate and X. Fageda, "Congestion, road safety, and the effectiveness of public policies in urban areas," *Sustainability*, vol. 11, no. 18, p. 5092, 2019, doi: 10.3390/su11185092.

[2] E. Basili *et al.*, "The impact of road accidents on hospital admissions and the potential of ADAS in containing health expenditure: Evidence from Piedmont data," *Transp. Res. Interdiscip. Perspect.*, vol. 25, 2024, doi: 10.1016/j.trip.2024.101125.

[3] A. Chopra, N. Kumar, and R. K. Kaushal, "A comprehensive analysis of driver drowsiness detection techniques," in *Applied Data Science and Smart Systems*, London: CRC Press, 2024, pp. 134–139. doi: 10.1201/9781003471059-19.

[4] D. Léger, E. Pepin, and G. Caetano, "The economic burden of sleepy driving," *Sleep Med. Clin.*, vol. 14, no. 4, pp. 423–429, 2019, doi: 10.1016/j.jsmc.2019.07.004.

[5] M. Kamboj *et. al.*, "Advanced detection techniques for driver drowsiness: a comprehensive review of machine learning, deep learning, and physiological approaches," *Multimed. Tools Appl.*, 2024, doi: 10.1007/s11042-024-19738-z.

[6] G. Sikander and S. Anwar, "Driver fatigue detection systems: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 6, pp. 2339–2352, 2019, doi: 10.1109/tits.2018.2868499.

[7] A. A. Saleem *et. al.*, "A systematic review of physiological signals-based driver drowsiness detection systems," *Cogn. Neurodyn.*, vol. 17, pp. 1229–1259, 2023, doi: 10.1007/s11571-022-09898-9.

[8] J. Vicente, P. Laguna, A. Bartra, and R. Bailón, "Drowsiness detection using heart rate variability," *Med. Biol. Eng. Comput.*, vol. 54, no. 6, pp. 927–937, 2016, doi: 10.1007/s11517-015-1448-7.

[9] A. Sahayadhas, K. Sundaraj, and M. Murugappan, "Drowsiness detection during different times of day using multiple features," *Australas. Phys. Eng. Sci. Med.*, vol. 36, no. 2, pp. 243–250, 2013, doi: 10.1007/s13246-013-0200-6.

[10] M. Shahbakhti *et al.*, "Simultaneous eye blink characterization and elimination from low-channel prefrontal EEG signals enhances driver drowsiness detection," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 3, pp. 1001–1012, 2022, doi: 10.1109/jbhi.2021.3096984.

[11] K. Fujiwara, H. Iwamoto, K. Hori, and M. Kano, "Driver drowsiness detection using R-R interval of electrocardiogram and self-attention autoencoder," *IEEE Trans. Intell. Veh.*, vol. 9, no. 1, pp. 2956–2965, 2024, doi: 10.1109/tiv.2023.3308575.

[12] J. Bai *et al.*, "Two-stream spatial–temporal graph convolutional networks for driver drowsiness detection," *IEEE Trans. Cybern.*, vol. 52, no. 12, pp. 13821–13833, 2022, doi: 10.1109/tcyb.2021.3110813.

[13] R. Florez *et. al.*, "A CNN-based approach for driver drowsiness detection by real-time eye state identification," *Appl. Sci. (Basel)*, vol. 13, no. 13, p. 7849, 2023, doi: 10.3390/app13137849.

[14] C. B. S. Maior, M. J. das C. Moura, J. M. M. Santana, and I. D. Lins, "Real-time classification for autonomous drowsiness detection using eye aspect ratio," *Expert Syst. Appl.*, vol. 158, no. 113505, p. 113505, 2020, doi: 10.1016/j.eswa.2020.113505.

[15] W. AlKishri, A. Abualkishik, and M. Al-Bahri, "Enhanced image processing and fuzzy logic approach for optimizing driver drowsiness detection," *Appl. Comput. Intell. Soft Comput.*, vol. 2022, pp. 1–14, 2022, doi: 10.1155/2022/9551203.

[16] R. K. Shukla, A. K. Tiwari, and A. K. Jha, "An efficient approach of face detection and prediction of drowsiness using SVM," *Math. Probl. Eng.*, vol. 2023, no. 1, 2023, doi: 10.1155/2023/2168361.

[17] A. K. Biswal, D. Singh, B. K. Pattanayak, D. Samanta, and M.-H. Yang, "IoT-based smart alert system for drowsy driver detection," *Wirel. Commun. Mob. Comput.*, vol. 2021, no. 1, pp. 1–13, 2021, doi: 10.1155/2021/6627217.

[18] H. Lamaazi, A. Alqassab, R. A. Fadul, and R. Mizouni, "Smart edge-based driver drowsiness detection in mobile crowdsourcing," *IEEE Access*, vol. 11, pp. 21863–21872, 2023, doi: 10.1109/access.2023.3250834.

[19] L. Zhang, H. Saito, L. Yang, and J. Wu, "Privacy-preserving federated transfer learning for driver drowsiness detection," *IEEE Access*, vol. 10, pp. 80565–80574, 2022, doi: 10.1109/access.2022.3192454.

[20] M. Ahmed, S. Masood, M. Ahmad, and A. A. Abd El-Latif, "Intelligent driver drowsiness detection for traffic safety based on multi-CNN deep model and facial subsampling," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 19743–19752, 2022, doi: 10.1109/tits.2021.3134222.

[21] Y. Jebraeily, Y. Sharafi, and M. Teshnehlab, "Driver drowsiness detection based on convolutional neural network architecture optimization using genetic algorithm," *IEEE Access*, vol. 12, pp. 45709–45726, 2024, doi: 10.1109/access.2024.3381999.

[22] H.-T. Nguyen, N.-D. Mai, B. G. Lee, and W.-Y. Chung, "Behind-the-ear EEG-based wearable driver drowsiness detection system using embedded tiny neural networks," *IEEE Sens. J.*, vol. 23, no. 19, pp. 23875–23892, 2023, doi: 10.1109/jsen.2023.3307766.

[23] S. K. Mousavikia, E. Gholizadehazari, M. Mousazadeh, and S. B. O. Yalcin, "Instruction set extension of a RiscV based SoC for driver drowsiness detection," *IEEE Access*, vol. 10, pp. 58151–58162, 2022, doi: 10.1109/access.2022.3177743.

[24] H. A. Madni, A. Raza, R. Sehar, N. Thalji, and L. Abualigah, "Novel transfer learning approach for driver drowsiness detection using eye movement behavior," *IEEE Access*, vol. 12, pp. 64765–64778, 2024, doi: 10.1109/access.2024.3392640.

[25] G. S. Shankara Chari and J. A. Prashant, "Scaling effectivity in manifold methodologies to detect driver's fatigueness and drowsiness state," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 13, no. 2, p. 1227, Jun. 2024, doi: 10.11591/ijai.v13.i2.pp1227-1240.

[26] IBug300W Accessed: Feb. 15, 2025. [Online]. Available: https://www.kaggle.com/code/gomidor/facial-landmark-detection

[27] B. Johnston and P. de Chazal, "A review of image-based automatic facial landmark identification techniques," *EURASIP J. Image Video Process.*, vol. 2018, no. 1, 2018, doi: 10.1186/s13640-018-0324-4.

BIOGRAPHIES OF AUTHORS

Gowrishankar Shiva Shankara Chari completed his Master degree in 2021 from BITS Pilani and Bachelor degree in 2006 from Visvesvaraya Technological University, India. He is pursuing his doctoral degree in artificial intelligence at the Department of Computer Science and Engineering, M.S. Ramaiah University of Applied Sciences, Ramaiah Technology Campus, Bengaluru, Karnataka, India. He has 18 years of Industry experience and working as solution architect. His research interest is in the field of machine learning, deep learning, AI, and cloud computing. He can be contacted at email: shankar.iq@gmail.com.

Jyothi Arcot Prashant completed her Ph.D. in 2020, Master degree in 2009, Bachelor degree in 2002 from Visvesvaraya Technological University, India. She works as Faculty of Engineering and Technology, Department of Computer Science and Engineering, Ramaiah University of Applied Sciences, M.S. Ramaiah Technology Campus, Bengaluru, Karnataka, India. She has 17 years of experience in teaching and has published many research papers in journals indexed in SCI/SCIE, WOS, Scopus and presented papers in several National and International conferences. Her research interest is in wireless sensor network, IoT, embedded systems, AI, ML, and deep learning. She can be contacted at email: jyothi.cs.et@msruas.ac.in.