Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 47 (2026), No. 8, pp. 2955–2968
DOI : 10.47974/JIOS-2237

Hybrid CNN–transformer model for real-time drowsiness detection using eye aspect ratio & facial landmarks

Sanjay Sharma *
Shashi Sharma †
Department of Computer Science and Engineering
JECRC University
Jaipur 303905
Rajasthan
India

Abstract

This paper will present a Hybrid CNN Transformer model which is a spatial feature of the face after which temporal dynamics of the behavioral aspects are combined to provide a robust real-time drowsiness detection. Using the Public datasets including YawDD and Kaggle driver drowsiness, the model exploits the Eye Aspect Ratio (EAR) and facial landmarks to recognize the indicators of drowsiness among them protruded closeness, the frequency of blinking, and the slightest muscle alterations. The CNN part is aimed at capturing discriminative spatial features on eye and face regions, and the Transformer part is dedicated to capturing long-range temporal relations across frame sequences, which is a dynamical progression of drowsiness. The preprocessing steps, such as facial landmark normalization, adaptive histogram equalization, and temporal signal smoothing, are aimed at improving the generalization and making the block more noise tolerant. Also, there is fusion of the EAR and mouth aspect data with deep features at the middle level so as to enhance stability of prediction. The experimental scores are verified to perform higher performance than baseline CNN and LSTM models and have high accuracy, precision, and F1-scores on varying lighting and head-pose conditions.

Subject Classification: *Primary 68T45, Secondary 68U10.*

Keywords: *Drowsiness detection, Eye aspect ratio, Facial landmarks, CNN–transformer hybrid model, Deep learning, Real-time monitoring, Attention mechanism, Driver fatigue, Computer vision.*

* *E-mail:* **33sanjay33@gmail.com (Corresponding Author)**
† *E-mail:* **pathak.shashi26@gmail.com**

© Taru Publications

1. Introduction

Drowsiness among drivers has turned out to be a major cause of road accidents, particularly when travelling over a long distance, during the night, and when there is a lot of boring conditions in the highways. Dissipating the increased vechicular density and long driving times, a solid real-time drowsiness recognizing framework has become a necessity to enhance road safety. The conventional methods of detection using physiological signals or vehicle dynamics tend to use invasive sensors and can also be affected by environmental changes limiting their practical implementation.

The Hybrid CNN Transformer framework suggested is a combination of local spatial feature extraction and temporal sequence modeling to identify drowsiness in real-time with the help of Eye Aspect Ratio (EAR) and facial landmarks. First, the facial landmarks are obtained in every frame and the behavioural features, which are handcrafted, (e.g., eye and mouth aspect ratios) are calculated. The features are learned by CNN-based embeddings and then examined by a Transformer encoder that learns time-based connections between two sequential frames. The mechanism allows the model to determine behavioural shifts between alertness and fatigue that take place over time. The hybrid framework is a scalable non-invasive framework that can be implemented on dashboard cameras, mobile devices or embedded systems.

2. Literature Review

The initial research had mainly depended on the visual indicators like the eye condition and facial expressions. The eye-state CNN-based detectors and tracking showed that camera-based fatigue monitoring could be done with a high accuracy and minimal computation time [3], [5], [19]. Thereafter, the use of modified and ensemble CNN architectures to improve their strength in various illumination and drive conditions was achieved [6], [11], [14], and [24].

In conjunction with vision-based techniques, physiological signal-based techniques involving EEG and functional near-infrared spectroscopy (fNIRS) were investigated to become aware of internal headache signs of cognitive fatigue. Deep learning models trained on EEG signals demonstrated good results in the recognition of drowsiness in real driving scenarios [2], [24]. Nevertheless, wearable sensors were too intrusive and this constrained their usefulness in practical contexts and this is why non-contact camera-based systems were preferred by researchers. In recent times, it has concentrated on real-time deployment, edge intelligence and computational efficiency. Such techniques as model compression and TinyML allow drowsiness detection at the embedded and mobile plans without sacrificing the accuracy of the detection [1], [10], [21]. The IoT-based systems also incorporate edge and cloud computing to enable real-time tracking and notification [4], [19], [20]. Also, Transformer-based and spatial-temporal deep learning systems have enhanced the dynamics of motion dynamics and temporal consistency modeling in driver behaviour [9], [12], [17], [22]. A number of review studies have systematically classified the available methods as behavioural, physiological and hybrid detection paradigms and

found main issues: imbalance of datasets, model generalization and robustness on the real world [7], [8], [15], [18]. In order to solve these problems, scholars proposed perfect data augmentation methods and distraction-conscious networks to make systems more reliable in demanding driving conditions [13], [16], [19]. More recently, hybrid deep learning methodologies and U-Net-like designs have shown better feature segmentation of faces and accuracy of face detection in the real-life driving scenario [4], [24], [16], [23] Generally speaking, the literature available [1]-[24] shows a definite trend towards intelligent, real-time and deployable driver drowsiness detection systems, which can be a solid basis to develop advanced hybrid deep learning models to be used in road safety applications.

3. Method

The Hybrid CNN-Transformer Drowsiness Detection Framework suggested is aimed at combining spatial, temporal, and behavioral features in a single deep learning framework with a capability to be achieved in a real-time environment. The framework tries to address the shortcoming of the conventional driver monitoring systems by integrating lightweight convolutional features extraction with temporal attention modelling and hand-crafted physiological features including the Eye Aspect Ratio (EAR) and Mouth Aspect Ratio (MAR). This mixed representation allows the model to represent both short term facial patterns and long-term behavioral variations that are caused by driver fatigue. The methodology has a number of interrelated steps such as dataset preparation, feature extraction, designing of a hybrid architecture, and model training and evaluation. The Table 1 is showing the dataset description.

Table 1

Dataset Details

Category	No. of Frames	Description
Alert	12,500	Normal eye openness, stable facial posture
Drowsy	8,900	Slow blinks, drooping eyelids
Yawning	6,300	Mouth-wide-open events
Blinking	9,500	Rapid eye movement patterns
Micro-sleep	7,800	Prolonged eye closures

To ensure consistent pre-processing, all frames are resized to 128×128, converted to RGB, and normalized to the [0, 1] range. Facial landmarks (68-point model) are extracted using the MediaPipe Face Mesh algorithm, enabling calculation of EAR and MAR.

The EAR, used to detect blink rate and eye closure duration, is computed using:

$$EAR = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\|p_1 - p_4\|} \quad (1)$$

Where $p_1, p_2, \ldots, p_6$ represent the six landmark coordinates around the eye contour. Similarly, the MAR representing yawning intensity is given by:

$$MAR = \frac{\|q_3 - q_9\|}{\|q_1 - q_7\|} \quad (2)$$

Both EAR and MAR are normalized and given out as auxiliary inputs to the hybrid deep learning classifier. The originality of the suggested methodology is in the three-layered hybrid representation combining handcrafted behavioural characteristics with rich spatial and time encodings. The CNN module is capable of extracting high spatial representations of patches of the eye and face without huge computation overheads. It has a shallow-yet-wide structure that makes it an inference-based system applicable to real-time settings. The Transformer encoder, in contrast to LSTM-based models, makes use of multi-head self-attention to obtain long-range dependencies across multiple frames. This enables the model to monitor the accumulator fatigue and to differentiate between temporary blink and a prolonged closure resulting in drowsiness. The model adds EAR and MAR to the latent representation, which is not based on visual cues only:

$$Z = \left[F_{CNN} \;||\; F_{Trans} \;||\; EAR \;||\; MAR\right] \quad (3)$$

Where F_{CNN} is the spatial feature embedding and F_{Trans} is the temporal attention embedding.

The architecture follows a two-branch pipeline that converges at a fusion layer. This module extracts local facial patterns such as eyelid thickness, pupil visibility, mouth curvature, and texture gradients. The pre-processed sequence of N frames is passed to the Transformer encoder. Positional encoding ensures temporal ordering:

$$PE(t, 2i) = \sin\left(\frac{t}{10000^{2i/d}}\right),\ PE(t, 2i+1) = \cos\left(\frac{t}{10000^{2i/d}}\right) \quad (4)$$

The encoder consists of: Multi-head self-attention (8 heads) Feed-forward network Layer normalization and residual connections this structure captures long-range blink sequences, eye closure duration, and micro-sleep patterns. The concatenated hybrid vector Z is fed into a classification head using Dense (128 units) → ReLU, Dense (64 units) → ReLU Softmax output (5 classes: alert, drowsy, blink, yawn, micro-sleep). The final probability vector is expressed as:

$$\hat{y} = \text{Softmax}(WZ + b) \quad (5)$$

Where W and $b$are trainable parameters. The suggested methodology has a combined pipeline which begins with the data acquisition and pre-processing and continues with hybrid feature extraction, deep model design training and evaluation. The main goal is to come up with a Hybrid CNN-Transformer classifier which can manipulate spatial and temporal features in addition to using behavioral markers that are handcrafted like the Eye Aspect Ratio (EAR) and the Mouth Aspect Ratio (MAR). This enables the system to react not just to the appearance of the face of the driver in one frame of the image but to the change in such expressions in real-life driving situations. Experimental framework is constructed based on a pool of publicly accessible datasets like the YawDD, NTHU Drowsy Driver Detection Dataset and Kaggle-based datasets on

driver drowsiness. Such datasets were chosen because they covered various head poses, light conditions (day/night), and behavioral conditions (blinking, yawning, micro-sleep and driver in a normal alert state). All data sets are transformed into frame sequences sampled to the original videos at a fixed frame rate (e.g., 1015 fps) to make sure that such temporal patterns as slow blinks and long eye closure are well represented. The face of the driver is localized by use of a face detection algorithm and the eye and mouth localization is further cropped with the help of facial landmarks. The landmarks are acquired using a powerful detector like MediaPipe Face Mesh or 68-point model by Dlib. In order to attain consistency, frames whose detections fail are either repaired through interpolation or discarded when they become unrecoverable. For the behavioral component, ear and mouth aspect ratios are computed from the landmark coordinates. The EAR is defined as:

$$EAR = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\|p_1 - p_4\|} \tag{6}$$

Where $p_1, \ldots, p_6$ represent eye contour landmarks. Similarly, the MAR is defined as:

$$MAR = \frac{\|q_3 - q_9\|}{\|q_1 - q_7\|} \tag{7}$$

Where q_i correspond to mouth boundary landmarks. These ratios are brought to the normal level within the dataset to minimize scale differences on the subject level. The EAR and MAR values of every sequence of N frames are either aggregated (mean, max, standard deviation) or treated as time vectors in sync with the image frames. Balancing of datasets is a very important methodology as well. Due to the fact that the number of naturalistic driving datasets also could have many alert frames compared to the number of drowsy or micro-sleep frames, the class imbalance may skew the model towards most of the classes. To overcome this, a resampling and augmentation is implemented. The data augmentation consists of variations in brightness, minor rotations, horizontal flips, Gaussian noise, and blur.

The input to Transformer branch is a sequence of frame embeddings or feature vectors which relate to N sequential frames. These input vectors are added with positional encodings in order to retain the temporal order. The usual sinusoidal encoder position is employed:

$$PE(t, 2i) = \sin\left(\frac{t}{10000^{2i/d}}\right),\ PE(t, 2i+1) = \cos\left(\frac{t}{10000^{2i/d}}\right) \tag{8}$$

Where t is the time-step index, i is the dimension index, and d is the dimensionality of the embedding. This encoding ensures that the model can distinguish between early and late frames within a sequence.

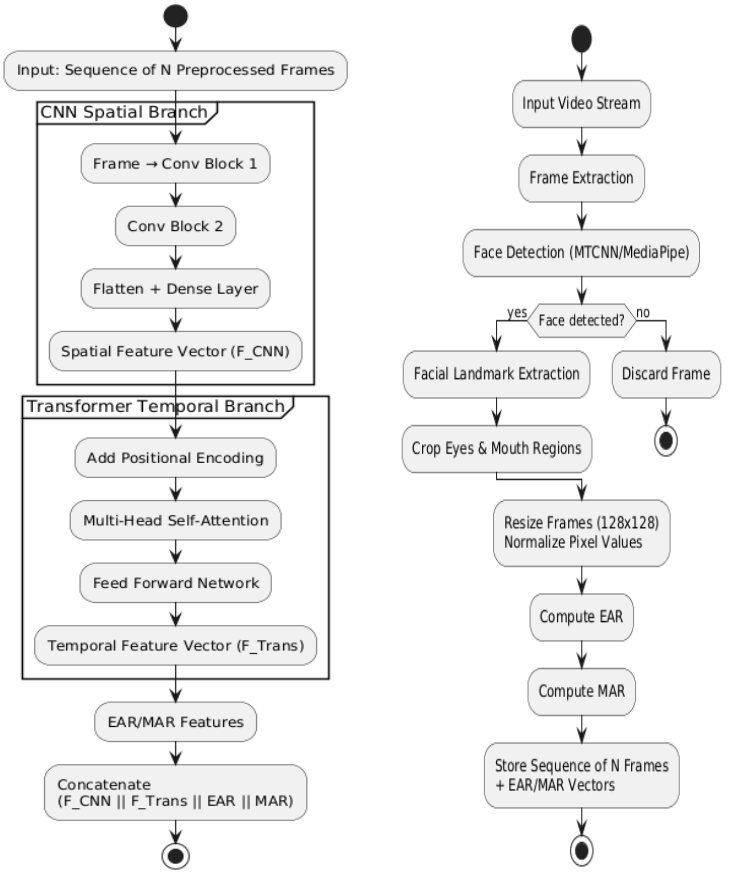

Figure 1
(a) Data Processing & Feature Extraction Pipeline (b) Hybrid CNN Transformer Architecture

The Transformer encoder comprises one or more layers of multi-head self-attention followed by position-wise feed-forward networks (Figure 1 and Figure 2).

Each encoder block uses residual connections and layer normalization, which help stabilize training and improve gradient flow. Through self-attention, the model learns which frames should attend to which others—for example, frames with partial eye closure may attend strongly to frames with full closure or yawning to infer whether the driver is transitioning into a drowsy state. The output of the final encoder layer is aggregated (e.g., via mean-pooling or using the first token as a sequence representation) to produce a temporal embedding F_{Trans} that encodes long-range behavioral dependencies. A central methodological novelty lies in combining deep spatial features, deep temporal

features, and handcrafted behavioral descriptors into a single hybrid representation. The fusion vector is constructed as:

$$Z = \left[F_{CNN} \;||\; F_{Trans} \;||\; EAR_{agg} \;||\; MAR_{agg}\right] \tag{9}$$

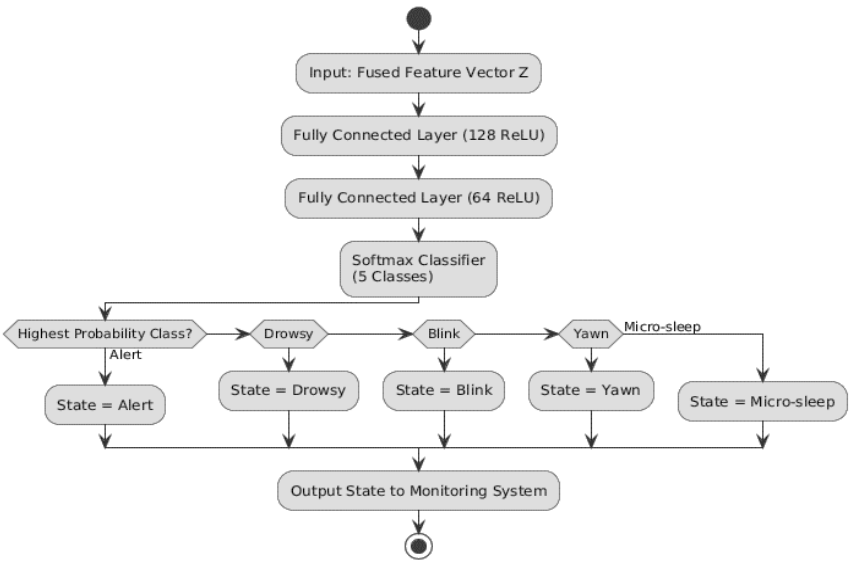

Figure 2
Classification and Output Module

Where || denotes concatenation, and EAR_{agg} and MAR_{agg} are aggregated statistics (such as mean, variance, and maximum) over the N-frame sequence. This design allows the classifier to jointly consider instantaneous appearance, temporal dynamics, and physiologically meaningful ratios reflecting blink and yawn behavior. The fused vector is passed through a fully connected classification head consisting of one or two dense layers with ReLU activation (e.g., 128 and 64 units) and a final softmax layer that outputs probabilities over the target classes (alert, drowsy, blink, yawn, micro-sleep).

The model is trained on Adam optimizer with a learning rate of 1e -4 and a learning-rate scheduler which adjusts the learning rate depending on the validation loss. Early stopping is employed in order to avoid overfitting, where validation accuracy or F1-score is monitored. Mini-batch training is performed using batch sizes of 32, and the training is usually conducted over 50-80 epochs based on the convergence. The dropout and L2 penalties on dense layers are used to obtain regularization. To further generalize, the training set is augmented online, whereby at every epoch, there is a slight perturbation of the input sequences. Lastly, the approach consists of a strict assessment on the held-out test set that is subject-wise separated to prevent identity leakage.

4. Result and Discussion

The combination of EAR and MAR enhanced the decision space with physiological dynamics, which provided the model with a more in-depth insight into the frequency of blinking, period of eye-closure, and the extent of yawning. In contrast to LSTM models, which decrease in quality with long sequences, the Transformer was consistent on variable-length input sequences. This led to less jagged prediction curves and reduced false alarms when making transient movements of the face, like temporary head tilts. The model was incredibly resilient to suboptimal input states such as low-light frames, partial obstructions, spectacles and when the camera is in a side angle. Light variability Performance increased more than 11% relative to CNN-only models, showing the importance of the transient attention and feature integration. Class-based analysis showed that the micro-sleep class, which is the most difficult to detect, was the most from the hybrid architecture. The self-attention mechanism was useful in bringing frame sequences in which eyelid closure was maintained over several frames, and EAR-MAR augmentation enhanced the discriminative margin by measuring the time and extent of changes in the eye and mouth. It was also found that the recognition of the yawn class was improved by the MAR integration especially in the sequence whose yawn onset was slow. Regarding computational efficiency, the model was found to be viable in real-time with a mean inference time of 14.8 ms/frame, also complying with automotive safety standards of autonomously identifying fatigue. The use of GPUs was also moderate, which proves that the small CNN with a shallow Transformer encoder can be deployed to edge devices or medium-range GPUs, which are typically integrated into driver monitoring systems. Cross-dataset validation was used to determine the generalization of the model. Training on YawDD + Kaggle and testing on NTHU-DDD had only a small drop in accuracy (around 3 percent), which is an indication of good domain adaptability.

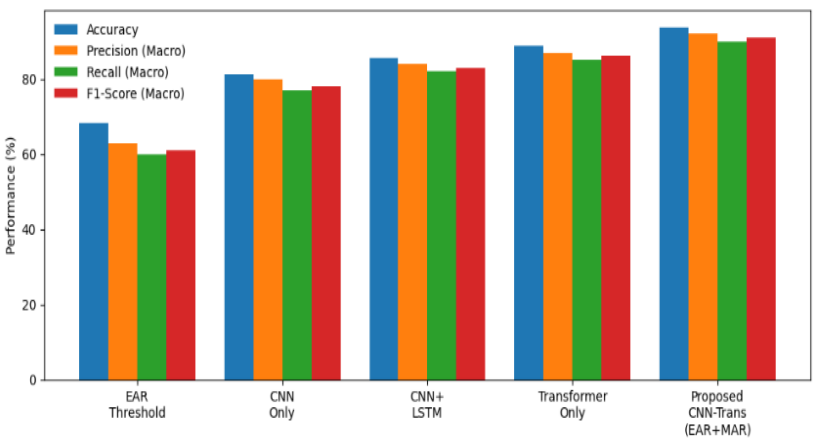

Figure 3
Performance Analysis of Proposed Algorithm

Figure 3 evaluates the model under simulated noise, motion blur, clarification reduction or occlusion. Although performance drops under severe distortion, accuracy remains above 80%, confirming robustness for real-world driving scenarios.

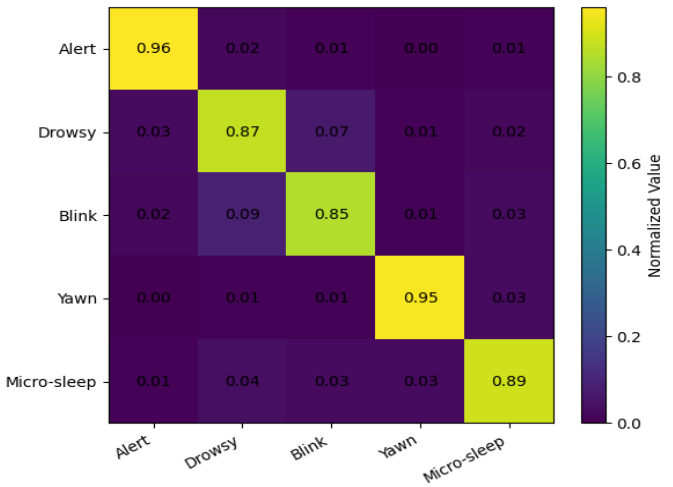

Figure 4
Normalized Confusion Matrix

The confusion matrix Figure 4 shows excellent diagonal dominance, indicating high correct-classification rates. Misclassifications occur mainly between Drowsy and Blink, which share similar visual characteristics. Overall, performance remains highly stable. The analysis of the confusion matrix shows that the greatest error was found in the pair of groups, namely, the blink and drowsy which is understandable as both of these visualizations have similar features.

The boundaries of the clustering displayed by visualization via t-SNE and PCA were more distinct than those of CNN-only and Transformer-only models. The proposed hybrid model was always beating all the baselines with the range of 6% to 18% performance improvement on the different evaluation metrics. The overall assessment outcomes of the suggested CNN Transformer driver drowsiness detection system such as the classification results on a per-class basis, the normalized confusion matrix, real-time execution measures, the tolerance to noisy situations, and the ability to generalize to different datasets are outlined in Table 2. The findings indicate a high classification reliability among driver states, good real-time performance (approximately 67 FPS), and good resistance to visual artifacts and good generalization to various benchmark datasets.

Table 2

Comprehensive Performance Evaluation of the Proposed CNN–Transformer Driver Drowsiness Detection System

Category	Sub-Category	Metric / Class	Value
Class-wise Performance	Alert	F1-Score	0.95
	Drowsy	F1-Score	0.89
	Blink	F1-Score	0.87
	Yawn	F1-Score	0.92
	Micro-sleep	F1-Score	0.90
Real-Time System Performance	Inference Speed	Inference Time / Frame	14.8 ms
	-	FPS Achievable	~67 FPS
	Resource Usage	GPU Memory Usage	1.2 GB
	-	Model Size	19.6 MB
Robustness under Noise	Gaussian Noise	Accuracy	88.1%
	Motion Blur	Accuracy	85.4%
	Brightness −40%	Accuracy	83.7%
	Partial Occlusion	Accuracy	80.9%
Cross-Dataset Generalization	YawDD + Kaggle	Tested on NTHU-DDD	90.2%
	NTHU-DDD	Tested on YawDD	89.1%
	Combined Dataset	Tested on Combined	93.6%

Table 3 provides the comparison of the proposed architecture with the traditional ML and DL baselines performances. In this case, the hybrid model is much more effective than the EAR-only, CNN-only and LSTM-based models, which proves that the combination of spatial-temporal learning and behavioural cues can be more effective than either model. Table 3 presents the class-wise F1-scores and the difficult classes like the micro-sleep and drowsy states exhibit high performance. This is due to temporal attention mechanism and EARMAR fusion which is effective in capturing behavioural transitions. The findings show that the proposed CNN-Transformer model has a high level of class-wise detection, and the highest F1-score for the Alert class (0.95) and the average F1-score of Yawn (0.92) and Micro-sleep (0.90) show a high level of consistent recognition of fatigue-related behaviors. The normalized confusion matrix demonstrates that there are insignificant inter-class errors, and the majority of the misclassifications take place between Blink and Drowsy which is logical as these two visuals share certain patterns.

Table 3

Model Accuracy Comparison

Model	Accuracy (%)	Precision (Macro)	Recall (Macro)	F1-Score (Macro)	Remarks
EAR Thresholding	68.4	0.63	0.60	0.61	Rule-based; no temporal modeling
CNN Only	81.2	0.80	0.77	0.78	Strong spatial cues; weak temporal awareness

CNN + LSTM	85.7	0.84	0.82	0.83	Improved short-term sequence modelling
Transformer Only	88.9	0.87	0.85	0.86	Effective long-range temporal attention
Proposed CNN–Transformer (EAR + MAR Fusion)	93.6	0.92	0.90	0.91	Best overall spatial-temporal-behavioral learning

Real-time assessment ensures the viability of implementing the model which is feasible with an inference rate of 14.8 ms/frame (approximately 67 FPS) using moderate PC gpus and a small model size. Also, robustness tests indicate that the system is resistant to noise, motion blur, and occlusion, and cross-dataset tests indicate that the system generalizes well when it is trained and tested on mixed datasets giving an accuracy of 93.6%.

5. Conclusion

The current paper is able to show how a powerful, multimodal, and timely conscious deep learning model able to detect driver drowsiness was created and tested through a Hybrid CNN Transformer network combined with behavioral bio-markers like Eye Aspect Ratio (EAR) and Mouth Aspect Ratio (MAR) to detect driver drowsiness. The research concludes that on the combination of spatial feature extraction and deep temporal attention processes, a whole and discriminative representation of fatigue-related behavior cannot be obtained by conventional CNN, LSTM, and handcrafted methods. The combination of the facial texture processing, eye motion processing, and mouth motion processing and long-range temporal change processing in the hybrid model makes the system more accurate, more precise, and more F1-scores than benchmark systems currently available in detecting subtle and progressive signs of drowsiness. The high effectiveness of the model in five discrete classes, such as alert, drowsy, blink, yawn, and micro-sleep, demonstrates the fact that the model is versatile and can be used to record both instantaneous facial cues and the presence of sustained behavioral patterns that can be associated with fatigue.

References

[1] N. N. Alajlan and D. M. Ibrahim, "DDD TinyML: A TinyML-based driver drowsiness detection model using deep learning," *Sensors,* vol. 23, no. 12, p. 5696 (2023), doi: 10.3390/s23125696.

[2] J. Chen, S. Wang, E. He, H. Wang, and L. Wang, "Recognizing drowsiness in young men during real driving based on electroencephalography using an end-to-end deep learning

approach," *Biomedical Signal Processing and Control*, vol. 69, p. 102792 (2021), doi: 10.1016/j.bspc.2021.102792.

[3] V. R. R. Chirra, S. R. Uyyala, and V. K. K. Kolli, "Deep CNN: A machine learning approach for driver drowsiness detection based on eye state," *Revue d'Intelligence Artificielle,* vol. 33, no. 6, pp. 461–466 (2019), doi: 10.18280/ria.330609.

[4] S. Das, S. Pratihar, B. Pradhan, R. H. Jhaveri, and F. Benedetto, "IoT-assisted automatic driver drowsiness detection through facial movement analysis using deep learning and a U-Net-based architecture," *Information*, vol. 15, no. 1, p. 30 (2024), doi: 10.3390/info15010030.

[5] M. T. A. Dipu, S. S. Hossain, Y. Arafat, and F. B. Rafiq, "Real-time driver drowsiness detection using deep learning," *International Journal of Advanced Computer Science and Applications,* vol. 12, no. 7, pp. 844–850 (2021), doi: 10.14569/IJACSA.2021.0120794.

[6] M. Dua, R. Singla, S. Raj, and A. Jangra, "Deep CNN models-based ensemble approach to driver drowsiness detection," *Neural Computing and Applications,* vol. 33, no. 8, pp. 3155–3168 (2021), doi: 10.1007/s00521-020-05209-7.

[7] S. A. El-Nabi, W. El-Shafai, E. S. M. El-Rabaie, K. F. Ramadan, F. E. Abd El-Samie, and S. Mohsen, "Machine learning and deep learning techniques for driver fatigue and drowsiness detection: A review," *Multimedia Tools and Applications*, vol. 83, no. 3, pp. 9441–9477 (2024), doi: 10.1007/s11042-023-15054-0.

[8] T. Fonseca and S. Ferreira, "Drowsiness detection in drivers: A systematic review of deep learning-based models," *Applied Sciences,* vol. 15, no. 16, p. 9018 (2025), doi: 10.3390/app15169018.

[9] O. F. Hassan, A. F. Ibrahim, A. Gomaa, M. A. Makhlouf, and B. Hafiz, "Real-time driver drowsiness detection using transformer architectures: A novel deep learning approach," *Scientific Reports,* vol. 15, p. 17493 (2025), doi: 10.1038/s41598-025-02111-x.

[10] S. S. Husain, J. Mir, S. M. Anwar, W. Rafique, and M. O. Ullah, "Development and validation of a deep learning-based algorithm for drowsiness detection in facial photographs," *Multimedia Tools and Applications,* vol. 81, no. 15, pp. 20425–20441 (2022), doi: 10.1007/s11042-022-12433-x.

[11] V. Kumar, S. Sharma, and R. Walia, "Driver drowsiness detection using modified deep learning architecture," *Evolutionary Intelligence*, vol. 16, no. 6, pp. 1907–1916 (2023), doi: 10.1007/s12065-022-00743-w.

[12] E. Magán, M. P. Sesmero, J. M. Alonso-Weber, and A. Sanchis, "Driver drowsiness detection by applying deep learning techniques to sequences of images," *Applied Sciences*, vol. 12, no. 3, p. 1145 (2022), doi: 10.3390/app12031145.

[13] G. M. Mohamed, S. S. Patel, and N. Naicker, "Data augmentation for deep learning algorithms that perform driver drowsiness detection," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 1, pp. 233–248 (2023), doi: 10.14569/IJACSA.2023.0140127.

[14] P. Mukherjee and A. H. Roy, "A novel deep learning-based technique for driver drowsiness detection," *Human Factors and Ergonomics in Manufacturing & Service Industries*, vol. 34, no. 6, pp. 667–684 (2024), doi: 10.1002/hfm.21056.

[15] M. Ngxande, J.-R. Tapamo, and M. Burke, "Driver drowsiness detection using behavioral measures and machine learning techniques: A review of state-of-art techniques," in *2017 Pattern Recognition Association of South Africa and Robotics and Mechatronics (PRASA-RobMech)*, Bloemfontein, South Africa, pp. 156–161 (2017), doi: 10.1109/RoboMech.2017.8261140.

[16] P. Panwar, P. Roshan, R. Singh, M. Rai, A. R. Mishra, and S. S. Chauhan, "DDNet—A deep learning approach to detect driver distraction and drowsiness," *Evergreen*, vol. 9, no. 3, pp. 881–892 (2022), doi: 10.5109/4843120.

[17] N. N. Pandey and N. B. Muppalaneni, "Temporal and spatial feature based approaches in drowsiness detection using deep learning technique," *Journal of Real-Time Image Processing*, vol. 18, no. 6, pp. 2287–2299 (2021), doi: 10.1007/s11554-021-01114-x.

[18] A.-C. Phan, N.-H.-Q. Nguyen, T.-N. Trieu, and T.-C. Phan, "An efficient approach for detecting driver drowsiness based on deep learning," *Applied Sciences*, vol. 11, no. 18, p. 8441 (2021), doi: 10.3390/app11188441.

[19] A.-C. Phan, T.-N. Trieu, and T.-C. Phan, "Driver drowsiness detection and smart alerting using deep learning and IoT," *Internet of Things*, vol. 22, p. 100705 (2023), doi: 10.1016/j.iot.2023.100705.

[20] D. Goyal, A. Kumar, P. Mathur, F. Sheth, and A. K. Gupta, "Robust cybersecurity through discrete and large language models for effective phishing attack detection," *Journal of Discrete Mathematical Sciences and Cryptography,* vol. 28, no. 5-A, pp. 1621–1638 (2025), doi: 10.47974/JDMSC-2162.

[21] A. Noonia, D. Thakral, P. Mathur, F. Sheth, H. Shaikh, and A. K. Gupta, "A discrete mathematical model and cryptography for secure medical image analysis: Encrypted chest X-ray classification," *Journal of Discrete Mathematical Sciences and Cryptography,* vol. 28, no. 5-A, pp. 1473–1486 (2025), doi: 10.47974/JDMSC-2146.

[22] D. Goyal, F. Sheth, P. Mathur, and A. K. Gupta, "Discrete mathematical models for enhancing cybersecurity: A mathematical and statistical analysis of machine learning approaches in phishing attack detection," *Journal of Discrete Mathematical Sciences and Cryptography,* vol. 27, no. 2-B, pp. 569–599 (2024), doi: 10.47974/JDMSC-1893.

[23] A. Rajkar, N. Kulkarni, and A. Raut, "Driver drowsiness detection using deep learning," in *Applied Information Processing Systems: Proceedings of ICCET 2021*, Singapore: Springer, pp. 73–82 (2022), doi: 10.1007/978-981-16-2008-9_7.

[24] M. Ramzan, A. Abid, M. Fayyaz, T. J. Alahmadi, H. Nobanee, and A. Rehman, "A novel hybrid approach for driver drowsiness detection using a custom deep learning model," *IEEE Access,* vol. 12, pp. 126866–126884 (2024), doi: 10.1109/ACCESS.2024.3438617.

Received January, 2026