A Multi-Cue Spatiotemporal Model for Real-Time Driver Drowsiness Detection

Sneha Barmaiya[1], Aman Kumar Patel[2], Megha Patidar[3] and Anand Singh Jalal*[4]

[1,2,3,4] School of Computer Science and Information Technology,
Devi Ahilya Vishwavidyalaya, Indore, India
[1] snehaabarmaiya@gmail.com, [2]amanp4631@gmail.com,
[3]meghapatidar13388@gmail.com, [4]anandsinghjalal@gmail.com*

Abstract. Driver drowsiness is a critical and persistent concern on modern roadways, and most of the accidents due to drowsiness can be avoided by detecting them in time. Previous works mostly rely on either a single-parameter cue, like thresholds on blink frequency or eye aspect ratio, or head-pose deviations. However, such isolated features often fail when these channels are subjected to real-world challenges like changes in illuminations, occlusions, and differences in individual behavior. This paper investigates early behavioral clues that indicate the onset of drowsiness in real-time to warn the driver well before attention decreases or control is lost. In this paper, we proposed a lightweight multi-cue detection model that overcomes single-feature dependency. The proposed method fuses MediaPipe Face Mesh with robust indicators, Temporal Landmark-based Eye Aspect Ratio for illumination-aware eye aspect analysis, Blink Morphology with-Closed-eye State Index and Head Pose and Spatial Dynamics. The fusion of this cue significantly enhances reliability under diverse driving scenarios, including low lighting and dynamic head movement.

Keywords: Driver drowsiness, Multi-cue, Driver fatigue.

1 Introduction

In the current era, drowsy driving is recognized as one of the major concerns for road accidents, primarily attributed to enhanced driving time, disorders related to sleep, bioeffects of medicines, and alcohol consumption. It has been evident through various research studies that there is a significant increase in the risk of accidents caused by drowsiness. It is evident that drowsy driving is responsible for an enhanced risk of accidents by 29-34%, establishing its widespread contribution to deaths on the roads [1].

The classical methods of correcting or self-assessing would not be effective, as fatigue takes time to develop, causing the drivers themselves to miss the initial signs of fatigue. Therefore, in recent times, the existing technology drifts predominantly towards the utilization of CNN-based systems [2], yawning detection modules [3], or analysis of eye closure. These systems would work effectively in a laboratory setting, but often not in real-world driving scenarios that encompass differences in blinking rates, head rotation, lighting, or intense computational power utilization for processing.

Machine learning-based approaches are used by extracting handcrafted features like HOG, LBP, or analysis of eye closure & blinks. Deep learning-based CNN framework

S. Sharma et al. (eds.), *Proceedings of the International Conference on Recent Advances in Intelligent and Sustainable Technologies (RAIST 2026)*, Atlantis Highlights in Intelligent Systems 17,
https://doi.org/10.2991/978-94-6239-707-1_25

is utilized that was further aided by temporal analysis of eyes for fatigue recognition [4]. Quiles-Cucarella et al. [5] utilized classic image processing techniques like EAR thresholding, Haar cascades, or SVM-based classifications. Kumar et al. [6] developed a drowsiness detection alert system using image processing techniques in association with a CNN-based framework. These systems tend to work entirely on parameters like EAR, blink times, or MAR, which would often be dynamic in real-world driving scenarios with respect to lighting or other factors, and would often not work consistently or would be sensitive to variations in such factors.

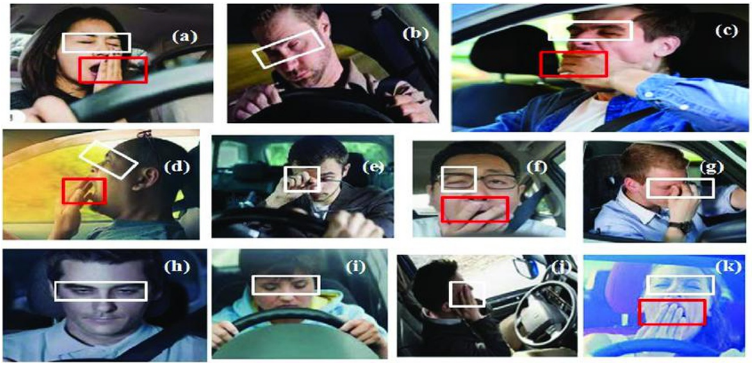

Fig. 1. Sample images of drowsiness [7]

Fig. 1 illustrates some of the visual cue indicating driver drowsiness. It portrays most common driver behaviors, including eye closure, drooping eyelids, yawning, and distracted facial movements associated with eating or talking. These cues are shared as common signals for fatigue and less alertness in drivers.

In this paper, a multi-cue based driver drowsiness detection model is proposed. The proposed model is experimented on real-time webcam video streams. The face alignment combined with tracking of key regions, such as ocular and oral ones, permitted the model to assess behavioral attributes including blinking frequency, stability of head movements. These subtle signs detected at an early stage allow the system to alert the driver quickly to reduce the chances of fatigue-induced incidents. Conclusively, the proposed model is expected to offer a dependable and efficient approach towards enhancing the safety of traffic and reducing accidents caused by drowsiness among drivers.

The remainder of this paper is structured as follows: Section 2 presents the related work, Section 3 discusses about datasets, Section 4 describes the proposed multi-cue driver drowsiness detection model, Section 5 demonstrates the experimental results, and Section 6 concludes the paper.

2 Related Work

The following section discusses the existing methods considered while developing driver drowsiness detection systems. Key recent advances and limitations of different methods are highlighted.

A real-time machine learning-based drowsy driver detection system using non-intrusive visual inputs taken from a camera mounted on the car dashboard has been proposed by Albadawi et al. [8]. Jahan et al. [9] developed a real-time system for driver drowsiness detection using a customized convolutional neural network to determine drowsiness based on eye state. Though the approach performs well, however, its efficiency for various subjects and illumination needs to be further confirmed to generalize the results to real-world situations.

A real-time framework for detecting a driver being drowsy has been proposed by Hassan et al. [10] using vision transformer and swin transformer models. The deep models have been strengthened using transfer learning techniques with pre-trained models like VGG19, DenseNet169, ResNet50V2, InceptionV3, and MobileNet. However, the computational complexities associated with the use of models like transformers might create certain limitations for real-time implementation.

Asif et al. [11] described a real-time system for driver drowsiness detection based on classical image processing and eye state classification. This method is based on real-time eye state identification (open or closed) based on certain visual characteristics of the face. Eye state identification is done by classical methods, and drowsiness is determined by a threshold on the time for which the eyes are closed. However, the system is tested in a controlled environment and not work efficiently in a more challenging environment, especially with low lighting.

Rajput et al. [12] presented a system for detecting drowsiness in drivers using transfer learning with a MobileNetV2 architecture. The method makes use of facial information obtained from images to build a learning framework for distinctive features that describe drowsiness and alertness classes. The MobileNetV2 weights are fine-tuned to enhance classification accuracy without sacrificing computational cost. This system shows that the developed low-cost model has a comparable level of accuracy, ensuring adaptability to real-time applications. Nevertheless, a possible deficiency of the system could arise when exposed to difficult challenges, such as changes in lighting and variability in drivers.

3 Datasets

Table 1 describes all the datasets used for training the proposed drowsiness detection model. These datasets together cover a wide variation in facial states, such as eye openness, eye closure, head orientation, and variation with/without glasses. The contribution of multiple datasets ensures the robustness of the model by offering diversity in training samples, hence improving detection performance under various lighting and facial conditions.

Table 1. Details of available datasets [13]

Dataset Name	Classes	Images	Size
Glasses Dataset	2	164	1.1 MB
MRL Eye Dataset	2	84,898	328.2 MB
NTHU-DDD	2	66,520	3.02 GB
Driver Drowsiness Dataset	2	41,790	2.76 GB

4 Proposed Model

The flow-diagram of the proposed model is shown in Fig. 2. The proposed model starts with robust preprocessing to enhance raw video frames, reduces noise, and normalizes the input in order to have reliable facial analysis in natural driving conditions. Face detection, cropping, and normalizing the frames with alignment were done to keep the consistency in landmark extraction. MediaPipe Face Mesh is used to get high-precision 3D facial landmarks and construct regions-of-interest in particular for eyes, mouth, and head pose. These landmarks are projected into a 2D and 3D geometrical space, computing indicators that reflect fatigue. Major metrics which describe eye closure behavior, such as Temporal Landmark–based Eye Aspect Ratio (TL-EAR), Blink Morphology and Closed-eye State Index (BMCSI), and Head Pose and Spatial Dynamics (HPSD), are derived. Lastly, these multi-cue indicators are combined to provide one unified fatigue score for the effective detection of drowsiness.

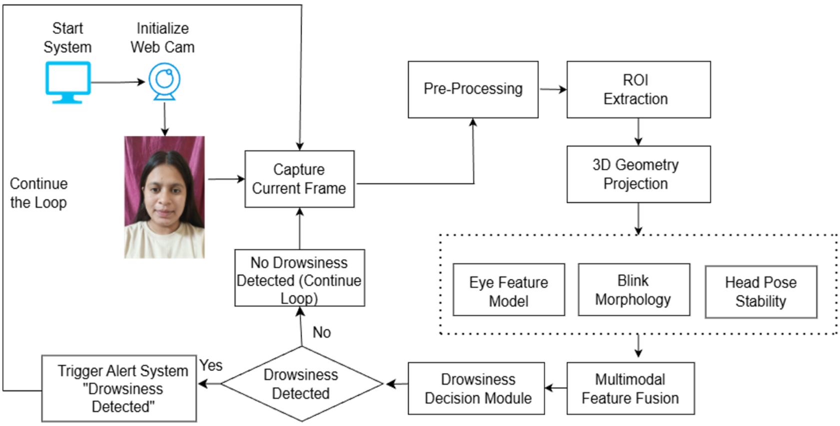

Fig. 2. Flow-Diagram of the proposed driver drowsiness detection model

4.1 Preprocessing

Pre-processing is an important aspect in the use of computer vision in detecting drowsiness since it ensures the images are of high quality and, in the process, any form of noise is removed and the inputs are normalized. In the use of an in-car camera, images collected are normalized and organized before the computer can process them for the detection of drowsiness. In this process, the facial Region of Interest is removed from the backgrounds that could cause interference.

4.2 Region of Interest (ROI)

Following augmentation and normalization, facial analysis is carried out by MediaPipe Face Mesh [14], yielding 468 3D facial landmarks at 30FPS in a single run. These landmarks are crucial for analyzing facial micro-movements that are integral to analyzing driver fatigue. Groups of these landmarks establish Regions of Interest, which are eye regions for TL-EAR & BMCSI, and regions surrounding the nose, chin, and cheeks for head pose stability.

4.3 3D Geometry Projection and Fatigue Indicator Computation

After extracting the ROI, the detected facial landmarks are transformed by the proposed model into meaningful geometric representations suitable for fatigue analysis. MediaPipe Face Mesh provides 468 normalized landmark coordinates

$$P_i = (x_i, y_i, z_i),\ i = 1,\ldots,468 \tag{1}$$

where x_i and y_i are normalized with respect to image dimensions and z_i represents the relative depth. To obtain robust indicators of fatigue, these normalized coordinates are projected both in 2D pixel space and in a 3D canonical face centred coordinate system. The detailed procedure of 3D geometry projection and computation of fatigue indicator has been carried out through following sequential steps:

2D and 3D Projection of Facial Landmarks [15]. *2D Pixel – Space Projection.* Each normalized coordinate is mapped to the actual image plane as

$$u_i = x_i W,\ v_i = y_i H \tag{2}$$

Where W and H denote the image width and height. These 2D points are u_i, v_i used for computing eye and mouth geometry.

3D Canonical Projection. Since MediaPipe's depth values are relative, the system stabilizes them by scaling with the detected facial bounding box width w_f:

$$X_i = (x_i - \bar{x}) \cdot w_f,\ \ Y_i = (y_i - \bar{y}) \cdot w_f,\ Z_i = z_i \cdot w_f \tag{3}$$

Where is $\bar{x}$, $\bar{y}$ the facial centroid.

This provides a canonical 3D representation that is invariant to camera distance and suitable for pose estimation. For higher precision, a subset of stable points (nose tip, eye corners, mouth corners, chin) are fitted to a 3D face template using Perspective-n-Point (PnP), yielding a rotation matrix R_t and Euler angles:

$$\theta_t = (\, yaw_t,\ pitch_t,\ roll_t \,) \tag{4}$$

4.4 Eye Feature Model

We used TL-EAR for more reliable eye-state detection under varying illumination and head movements.

Illumination-Corrected Eye Aspect Ratio [15]. The Eye Aspect Ratio (EAR) is computed from 2D landmark distances

$$EAR_t = (|p_2\text{-}p_6| + |p_3\text{-}p_5|) \,/\, (2|p_1\text{-}p_4|) \tag{5}$$

Here, p_i means the i-th landmark point of the eye.
EAR is sensitive to lighting and image noise therefore, the system applies photometric correction.

Illumination Correction. The mean grayscale intensity of the eye region at time t is denoted as I_t. The correction coefficient is defined as:

$$\alpha_t = I_{ref} \,/\, (I_t+\varepsilon) \tag{6}$$

Where I_{ref} is the running illumination reference and ε prevents division by zero.
The illumination-normalized EAR becomes:

$$EAR_t^L = EAR_t \cdot \alpha_t \tag{7}$$

TL-EAR is preferred over EAR because it records the dynamic process of eye blinking, which is crucial in processes related to fatigue identification. In fact, EAR can neither record blink rates nor blink durations.

To suppress noise and ensure temporal stability, an exponential filter is applied:

$$TL\text{-}EAR_t = \lambda \cdot EAR_t^L + (1\text{-}\lambda) \cdot TL\text{-}EAR_{t\text{-}1} \tag{8}$$

Where $0 < \lambda < 1$controls responsiveness.

Outcome: TL-EAR provides a smooth and lighting-invariant measure of eyelid closure, enabling reliable blink detection even under fluctuating illumination.

4.5 Blink Morphology and Closed-eye State Index (BMCSI) [16]

While TL-EAR indicates blink presence, it does not differentiate normal (fast, symmetric) blinks from fatigue-related (slow, asymmetric) blinks. BMCSI models the blink waveform shape using curvature and temporal morphology.

Let:

$$B = \{TL - EAR_{t1} \,, .. ,TL - EAR_{\,tn} \} \tag{9}$$

represent the TL - EAR sequence during a blink.

Here:

B denotes the blink signal sequence.

TL - EAR_{ti} represents the Temporal Local Eye Aspect Ratio value at time instant t_i.

t_1 to t_n correspond to consecutive video frames covering the complete blink duration, from eye opening to closure and reopening.

n is the total number of frames involved in the blink.

Curvature Analysis (Ki). Curvature at discrete point i is computed using the second finite difference:

$$K_i = | B_{i+1} - 2B_i + B_{i-1}| \quad (10)$$

Higher curvature corresponds to sharper blinks, fatigue blinks typically show shallow, low-curvature profiles.

Asymmetry Measure. Let A_L and A_R denote the eye-specific areas under the blink curves. Blink asymmetry is quantified as:

$$A = (|A_L - A_R|) / ((A_L + A_R)/2 + \varepsilon) \quad (11)$$

Duration Normalization. Blink duration is:

$$D = (t_n - t_1) / T_{ref} \quad (12)$$

where T_{ref} is typical blink duration under alert conditions.

BMCSI Formulation. The composite fatigue-oriented shape index is defined as:

$$BMCSI = w_1 \max(k_i) + w_2 A + w_3 D \quad (13)$$

Where w_1, w_2, w_3 are empirically chosen weights.

Interpretation. Low curvature (κ_i) indicates slow eyelid motion, commonly associated with fatigue. High asymmetry (A) reflects uneven eye-closing and opening behavior, a characteristic of drowsy blinks. Prolonged blink duration (D) signifies delayed eye reopening, which is a strong indicator of driver fatigue.

4.6 Head Pose and Spatial Dynamics (HPSD) [17]

Using the 3D head orientation ($\theta_t = yaw_t, pitch_t, roll_t$), the system tracks posture instability.

Fatigue often causes slow lateral tilting or forward nodding. To quantify this, the variance of pitch and roll is computed over a sliding temporal window W:

$$HPSD_t = Var\,(pitch_{\,t-W:t}) + Var\,(roll_{\,t-W:t}) \quad (14)$$

High variance corresponds to loss of postural control, a strong physiological signature of drowsiness.

4.7 Multimodal Fatigue Score Fusion [18] [19]

The fusion weights w_E, w_B, and w_H are associated with the contribution of eye-based, blink-based and head-based features, respectively, towards the decision-making process. The weights are selected based on an empirical evaluation of the system on a validation set. Three normalized indicators are combined:

S_E: TL-EAR-based eye closure severity

S_B: BMCSI-based blink morphology deviation

S_H: HPSD-based posture instability

The fused fatigue score is computed as:

$$S_t = w_E\, S_E + w_B\, S_B + w_H\, S_H \tag{15}$$

where $w_E + w_B + w_H = 1$.

Finally, drowsiness is declared when:

$$S_t \geq T \tag{16}$$

where T is determined through ROC-based threshold optimization on the validation dataset.

5 Result and Discussion

The proposed model is evaluated on driver drowsiness dataset (DDD), glasses dataset, MRL Eye dataset, NTHU-DDD dataset and compared with other state-of-the-art methods. The results show that the proposed approach outperforms existing drowsiness-detection methods across all studies as compared in Table 2. Traditional machine learning-based methods with HOG, LBP, and blink features only provide a mediocre performance; deep CNN and IoT-based models demonstrated better performance however, limited in terms of their consistency and computationally expensive. The classical thresholding methods on EAR estimates coupled with Haar cascades provide low reliability due to sensitivity to illumination conditions and pose angles. In contrast, the proposed model, which combines MediaPipe Face Mesh with TL-EAR, BMCSI, and HPSD, demonstrated higher accuracy and proved to possess much stronger stability of facial conditions, this confirms the effectiveness and robustness of the proposed hybrid model in real-time detection of driver drowsiness.

An ablation study is conducted to analyze the contribution of Eye, Blink and Head posture cue in the fusion model. When the model uses all the three cue, it gains highest accuracy of 95.2%, whereas not considering eye as a cue resulted in an accuracy of 87.6%, when not considering blink rate an accuracy of 91.3% is achieved and not considering head posture resulted in an accuracy of 92.1%. These results indicate that each cue contributes effectively in the detection of drowsiness and removing any cue can affect the accuracy of the model.

Table 2. Accuracy Comparison of Detection Models on dataset

Reference	Features	Approach	Accuracy (%)
[20]	Eye and Mouth	Logistic regression	92
[21]	Eye, Mouth, and Head	3D convolutional networks	76.2
[22]	Eye State, Yawn	VGG16	91
Proposed Model	Eye, Blink, Head	Multi-Cue Spatiotemporal Model	95

To ensure generalizability, the model performance is also validated on different benchmark datasets as illustrated in Table 3. It is evident that the model performance is consistent throughout different datasets.

Table 3. Result analysis of the proposed model

Dataset	Precision(%)	Recall(%)	Accuracy(%)
Driver Drowsiness Dataset (DDD)	96.1	95.4	95.8
Glasses Dataset	91.3	89.6	90.8
MRL Eye Dataset	97.4	96.9	97.1
NTHU-DDD	95.2	94.1	94.7

6 Conclusion

In this work, we present a driver drowsiness detection model that integrates multiple parameter such as MediaPipe Face Mesh, TL-EAR, BMCSI, and HPSD. MediaPipe Face Mesh is used to get high-precision 3D facial landmarks and construct regions-of-interest. TL-EAR is used for more reliable eye-state detection under varying illumination and head movements. BMCSI is utilized to calculate the frequency of blinks. HPSD uses 3D head orientation to track head posture and instability. Considering multiple cues resulted in a more robust and accurate model as compared to existing models. In future other cues can be explored to further improve performance and make system more reliable.

References

[1] A. Moradi, S. S. H. Nazari, and K. Rahmani, "Sleepiness and the risk of road traffic accidents: A systematic review and meta-analysis of previous studies," Transportation Research Part F: Traffic Psychology and Behaviour, vol. 65, pp. 620–629, Aug. 2019, doi: 10.1016/j.trf.2018.09.013.

[2] N. Datta, T. Mahmud, M. Begum, M. T. Aziz, D. Islam, M. F. B. Aziz, K. Kochkarov, T. Eshchanov, V. S. O. Uglu, S. Parmanov, and M. S. Hossain, "Convolutional neural network-based real-time drowsy driver detection for accident prevention," TELKOMNIKA (Telecommunication Computing Electronics and Control), vol. 23, no. 3, pp. 1-??, 2025, doi:10.12928/telkomnika.v23i3.26059.

[3] S. Bandewar, A. A. Labhsetwar, A. K. Lad, M. J. Lahamage, A. A. Laddha, P. D. Kurkute and G. S. Kurandale, "Drowsiness and yawning detection using facial landmark tracking,"

IJRaset Journal for Research in Applied Science and Engineering Technology, 2023, doi:10.22214/ijraset.2023.57034.

[4] A. Benmohamed and H. Zarzour, "A deep learning-based system for driver fatigue detection," International Journal of Intelligent Systems and Applications (ISI), vol. 29, no. 05, pp. 1–13, Oct. 2024, doi: 10.18280/isi.290511

[5] E. Quiles-Cucarella, J. Cano-Bernet, L. Santos-Fernández, C. Roldán-Blay, and C. Roldán-Porta, "Multi-Index Driver Drowsiness Detection Method Based on Driver's Facial Recognition Using Haar Features and Histograms of Oriented Gradients," Sensors, vol. 24, no. 17, Art. no. 5683, 2024, doi: 10.3390/s24175683.

[6] A. Kumar et al., "Driver Alert System Using Convolutional Neural Network," International Journal of Latest Technology in Engineering, Management & Applied Science (IJLTEMAS), vol. 14, no. 4, pp. 581–587, May 2025, doi: 10.51583/IJLTEMAS.2025.140400063.

[7] M. Itani, M. Jike, N. Watanabe, and Y. Kaneita, "Sleep deficiency and motor vehicle crash risk in the general population: A prospective cohort study," BMC Medicine, 2018.

[8] Y. Albadawi, A. AlRedhaei, and M. Takruri, "Real-time machine learning-based driver drowsiness detection using visual features," Journal of Imaging, vol. 9, p. 91, 2023.

[9] I. Jahan, K. M. A. Uddin, and S. A. Murad, "4D: A real-time driver drowsiness detector using deep learning," Electronics, vol. 12, no. 1, p. 235, 2023.

[10] 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, no. 1, 2025.

[11] H. Asif and G. Mustafa, "Driver drowsiness detection system by real-time eye state identification," Spectrum of Engineering Sciences, vol. 3, no. 8, pp. 167–177, 2025.

[12] I. S. Rajput, S. Tyagi, V. Pandey, A. Upreti, and J. K. Kukreja, "Enhancing driver safety with MobileNetV2-based transfer learning for drowsiness detection," Procedia Computer Science, vol. 259, pp. 1239–1248, 2025.

[13] 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.

[14] Google, "MediaPipe Face Mesh," 2020. [Online].
Available: https://developers.google.com/mediapipe/solutions/vision/face_mesh

[15] J. Čech and T. Soukupová, "Real-time eye blink detection using facial landmarks," Center for Machine Perception, Czech Technical University, Prague, Tech. Rep., pp. 1–8, 2016.

[16] X. Zhang, J. Liu, and M. Chen, "Head pose estimation for driver monitoring using PnP-based 3D face models," Pattern Recognition Letters, vol. 128, pp. 15–22, 2020.

[17] S. Lee, S. Kim, and H. Park, "Multimodal driver fatigue detection using eye, head pose, and yawning features," IEEE Access, vol. 9, pp. 123456–123468, 2021.

[18] J. Gwak, A. Hirao, and M. Shino, "An investigation of early detection of driver drowsiness using ensemble machine learning based on hybrid sensing," Applied Sciences, vol. 10, no. 8, p. 2890, 2020.

[19] A. Kumar and R. Patra, "Driver drowsiness monitoring system using visual behaviour and machine learning," in Proc. IEEE Symp. Comput. Appl. Ind. Electron. (ISCAIE), Penang, Malaysia, Apr. 2018, pp. 339–344.

[20] J. Yu, S. Park, S. Lee, and M. Jeon, "Driver drowsiness detection using condition-adaptive representation learning framework," IEEE Trans. Intell. Transp. Syst., vol. 20, pp. 4206–4218, 2019.

[21] M. Dua, R. Singla, S. Raj, A. Jangra, and Shakshi, "Deep CNN models-based ensemble approach to driver drowsiness detection," Neural Computing and Applications, vol. 33, pp. 3155–3168, 2021.

[22] A. Aytekin and V. Mençik, "Detection of driver dynamics with VGG16 model," *Applied Computer Systems*, vol. 27, no. 1, 2022.