



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** X **Month of publication:** October 2026

DOI: <https://doi.org/10.22214/ijraset.2026.85176>

www.ijraset.com

Call: ☎ 08813907089

E-mail ID: ijraset@gmail.com

Smart Vision-Based Driver Safety Monitoring System

Mrs. Sowmya D¹, Sanjana M L², Shreya S³, Yashu A B⁴, Khushi N⁵

Department of Computer Science and Engineering, JNN College of Engineering, Shivamogga, Karnataka, India

Abstract: Many road accidents happen because a driver becomes tired or sleepy, and safety features such as seat belts and airbags only reduce the damage after a crash has already begun. This paper presents the design of a Smart Vision-Based Driver Safety Monitoring System that watches the driver through an ordinary dashboard or laptop camera and warns them when signs of drowsiness appear. The system looks at several clues together, namely how long the eyes stay closed, whether the driver is yawning, and how the head is positioned, instead of depending on a single clue. A deep learning model called a Convolutional Neural Network (CNN), improved with an attention mechanism known as CBAM, helps the system concentrate on the eye and mouth regions that matter most. To stay dependable when the view is poor, the system first checks the quality of each frame, adjusts how much weight each clue receives, smooths its decisions over a short period of time, and raises an alert only when two separate checks agree. A review of sixteen related studies from 2021 to 2026 shows the gaps that this design aims to close, and preliminary training results are reported to indicate the current stage of the implementation. The proposed solution is low-cost, needs no wearable sensors, and is intended to run on standard computing hardware.

Keywords: Driver Drowsiness Detection; Convolutional Neural Network; Attention Mechanism; CBAM; Adaptive Multi-Cue Monitoring; Computer Vision; Driver Safety.

I. INTRODUCTION

Road accidents are one of the leading causes of death and injury around the world, and a large share of them is linked to drivers who are tired or sleepy, especially during long journeys, monotonous highway stretches, and night-time travel. Drowsiness slows reaction time and weakens judgement, with effects comparable to those of a blood alcohol level of 0.05% or higher. Common safety features such as seat belts, airbags, and anti-lock brakes are reactive, because they reduce the harm of a crash rather than preventing it. This work therefore proposes a Smart Vision-Based Driver Safety Monitoring System that uses an ordinary camera, face analysis, and deep learning to notice early signs of drowsiness and warn the driver before a dangerous situation develops.

Watching the driver's face with a camera has become the most popular non-intrusive way to monitor fatigue, because it avoids the cost and discomfort of body-worn sensors such as EEG or EOG electrodes. Early camera-based methods rely on simple measurements such as the Eye Aspect Ratio (EAR), which compares how wide the eye is open with its width, together with fixed limits. These rules are cheap to compute but work poorly under changing light, sunglasses, partial occlusion, and head movement. Deep learning, and Convolutional Neural Networks (CNNs) in particular, learns useful patterns directly from images, and attention mechanisms further help the network focus on the face regions that matter most. The need is greatest for truck, bus, and taxi drivers, who work long hours with irregular sleep, and features such as lane departure warning react only after the vehicle has already drifted, whereas watching the face addresses the cause of the problem itself.

Existing systems, however, usually depend on a single clue such as eye closure, combine several clues in a fixed way whether or not each clue is actually visible, or need heavy computing that is difficult to deploy in a vehicle. This motivates the present work: a camera-only system that checks the quality of every frame, relies on whichever clues remain visible, and raises an alert only when its evidence is both strong and consistent over time. Because it needs only camera frames, the same system can serve private cars, commercial fleets, ride-sharing services, driver-training simulators, and low-cost embedded boards such as the Raspberry Pi or NVIDIA Jetson Nano.

The objectives of the proposed system are to detect the driver's face and facial landmarks from live camera frames; to check frame quality so that unreliable frames are flagged instead of being mis-classified; to combine eye, mouth, and head-pose evidence adaptively; to classify drowsiness using a CNN enhanced with an attention mechanism; to smooth decisions over time and confirm them with a probability and confidence check before alerting; and to deliver timely warnings in real time on standard, low-cost hardware without wearable sensors.

II. LITERATURE SURVEY

A review of recent research was carried out to identify effective techniques for real-time driver drowsiness detection. Sharma et al. [1] used deep learning-based face detection followed by a CNN that classifies eye state as open or closed and raises an alert when closure lasts too long, but the method relies on good lighting and clear eye visibility and ignores yawning and head pose. Islam et al. [2] trained a CNN directly on raw eye-state images, which removes manual feature design and generalises across drivers, although it needs a large, diverse dataset and loses accuracy under poor lighting or occlusion. Gupta et al. [3] likewise applied deep neural networks to real-time drowsiness detection. Singh et al. [4] tracked the Eye Aspect Ratio across frames to detect prolonged eye closure; the method is simple and cheap, but it is sensitive to lighting and glasses and ignores yawning and head nodding.

Verma et al. [5] proposed a hybrid CNN–LSTM in which the CNN extracts spatial features and the LSTM analyses their sequence, reducing false positives at a computational cost that suits embedded use poorly without GPU acceleration. Li et al. [6] combined a lightweight CNN with attention focused on the eyes and mouth, achieving high accuracy at low cost, though it remains sensitive to lighting and needs well-annotated data. Kumar et al. [7] used YOLOv8 for face detection followed by landmark-based EAR and MAR computation, improving robustness and speed but requiring a GPU for best performance. Singh et al. [8] deployed MobileNet on edge devices for low-latency in-vehicle detection, at the price of somewhat lower accuracy. Bhanja et al. [9] added PERCLOS (percentage of eye closure over time) to a CNN eye-state classifier to detect sustained fatigue rather than momentary blinks, but the system stays eye-centred and degrades under occlusion and poor lighting.

Albadawi et al. [10] combined EAR, MAR, and head pose from MediaPipe/Dlib landmarks for greater robustness, although the thresholds need per-person calibration. Nasir et al. [11] proposed SAFE-DRIVE-AI, a CNN–LSTM–attention framework with high accuracy that needs large datasets and GPU support. Hassan et al. [12] applied Vision Transformers to reach state-of-the-art accuracy at a very high computational cost. Chen et al. [13] fused facial, head-movement, and physiological signals to cut false detections, but required complex and costly sensor hardware. Rahman et al. [14] paired a CNN with a bidirectional LSTM to capture subtle changes in blinking and eye closure, with heavy memory and processing demands. Hassan et al. [15], the base paper for this work, enhanced a CNN with four attention mechanisms (SE, Spatial Attention, ECA, and CBAM) and benchmarked it against six transfer-learning models on the NTHU Driver Drowsiness Dataset, reporting over 99% accuracy with CBAM while remaining sensitive to lighting and occlusion. Finally, although addressing a different problem, Menda et al. [16] show that combining several complementary mechanisms, in their case multiple encryption keys for secure video streaming, improves robustness compared with relying on a single one.

Across these studies, lighter systems [1], [4], [8], [9] rely on a single cue and degrade under occlusion, poor lighting, or eyewear, whereas systems that add temporal modelling, transformers, or physiological sensors [5], [11]–[14] gain accuracy at a computational cost that limits low-cost real-time use. Multi-cue systems [6], [7], [10] combine cues statically rather than adapting to which cues are reliable in each frame, and few bring attention-enhanced features, adaptive multi-cue evidence, and quality-aware decision making together in a single camera-only system that runs on ordinary hardware. This gap motivates the system proposed in this paper.

III. PROPOSED METHODOLOGY

The proposed system is organized as a pipeline of six stages: video acquisition and face detection, frame-quality gating, adaptive cue selection, CNN+CBAM feature extraction, temporal smoothing with a two-gate decision, and alert generation, so that a standard camera feed can be turned into a reliable drowsiness warning with minimal delay and no wearable hardware.

A. Video Acquisition and Face Detection

Frames are captured from a dashboard-mounted or laptop camera using OpenCV [19]. Each frame is resized and passed to a HOG+SVM face detector, and a 68-point facial-landmark predictor from the dlib library [20] then locates the eyes, mouth, and face outline. This lighter detector is chosen over GPU-dependent alternatives so that the face-detection stage stays deployable on ordinary hardware.

B. Frame-Quality Gating

Before any decision is made, each frame passes a quality check that confirms the face is visible, the lighting is adequate, and the eyes and mouth are not hidden by sunglasses, shadows, or heavy occlusion. A frame that fails the check is flagged as unreliable instead of being forced into a prediction, so that the system never gives a confident-looking answer from poor evidence.

C. Adaptive Multi-Cue Selection

Three groups of cues are computed from the landmarks: the Eye Aspect Ratio (EAR) for eye closure, the Mouth Aspect Ratio (MAR) for yawning, and head pose for nodding or tilting. Rather than always combining them equally, the system dynamically re-weights these cues according to which of them the current frame can actually support, so that mouth and head-pose evidence gain importance when sunglasses make the eyes unreliable. The EAR and MAR thresholds can be personalised to each driver through a short calibration phase at start-up.

D. NN+CBAM Feature Extraction

The eye and mouth regions are passed to a custom CNN backbone enhanced with the Convolutional Block Attention Module (CBAM), which applies channel attention followed by spatial attention so that the network learns both what to look at and where to look. The base work [15] examined four attention mechanisms (SE, Spatial Attention, ECA, and CBAM) and benchmarked them against six transfer-learning architectures (VGG19, ResNet50V2, DenseNet169, InceptionV3, MobileNet, and InceptionResNetV2) on the NTHU Driver Drowsiness Dataset [18]; CBAM is adopted here because it combined high reported accuracy with a light computational load. The model is built with Keras [17].

E. Temporal Smoothing and Two-Gate Decision

A single frame is not enough to judge drowsiness, because a normal blink or a brief glance away can look like a warning sign. The system therefore keeps a short sliding window of recent frames and computes the proportion of time for which each cue indicates fatigue, in the same spirit as PERCLOS but across eye, mouth, and head-pose evidence together. A two-gate check is then applied: the smoothed drowsiness probability must exceed a probability threshold, and the model's confidence in that decision must also exceed a confidence threshold, before an alert is allowed. A light sliding window is used instead of a full recurrent network, trading some temporal modelling capacity for real-time speed on modest hardware.

F. Alert Generation and Monitoring Interface

When both gates are passed, the system raises a drowsiness alert to warn the driver. A monitoring interface shows the live driver status, the cues currently in use, and any alerts, and it reports clearly when the face is not detected or the frame quality is too low to trust. Because the pipeline needs only camera frames, the design is intended to be low-cost and deployable on standard laptops and embedded boards, with future scope for fleet-monitoring dashboards and richer behaviour analysis.

G. Preliminary Training and Evaluation

Figs. 1 and 2 show the results of the current training and evaluation run of the classifier. Training and validation accuracy both settle at about 64.3% within the first two epochs and remain unchanged for the rest of the 15 epochs shown (Fig. 1). On the 9,078 test samples (3,239 not drowsy and 5,839 drowsy), every sample was predicted as drowsy (Fig. 2): the drowsy class has a recall of 1.00 and an F1-score of 0.78, while the not-drowsy class has a precision, recall, and F1-score of 0.00, giving a macro-averaged F1-score of 0.39. The overall accuracy of 0.64 equals the share of drowsy samples in the test set (5,839 of 9,078), which shows that the model has so far learned only the class distribution and not features that separate the two classes. These results are therefore treated as a baseline for the current stage of the work, and improving the training so that the model separates the two classes is the immediate next step.

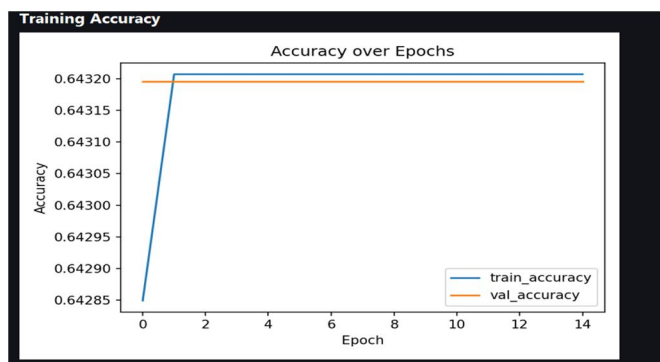


Fig. 1. Training and validation accuracy over epochs.

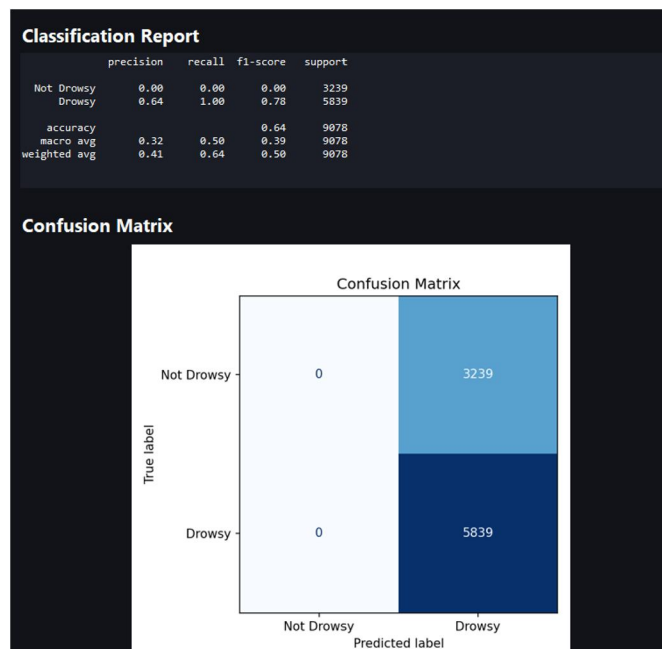


Fig. 2. Classification report and confusion matrix on the test set.

IV. CONCLUSION

This paper has presented the design of a Smart Vision-Based Driver Safety Monitoring System that combines computer vision and deep learning to detect driver drowsiness from an ordinary camera. By combining frame-quality gating, adaptive eye, mouth, and head-pose cues, a CNN+CBAM feature extractor, sliding-window smoothing, and a two-gate probability-confidence check, the proposed approach aims to remain reliable under sunglasses, poor lighting, and partial occlusion while staying light enough for real-time use on standard hardware. The CBAM-enhanced CNN of the base work [15] reports over 99% accuracy, but the preliminary results of the present implementation reflect only the class distribution and show that further refinement of the training is needed. Future work will complete this refinement and evaluate the full adaptive framework on real driving footage, validating its robustness, latency, and false-alarm rate.

V. ACKNOWLEDGMENT

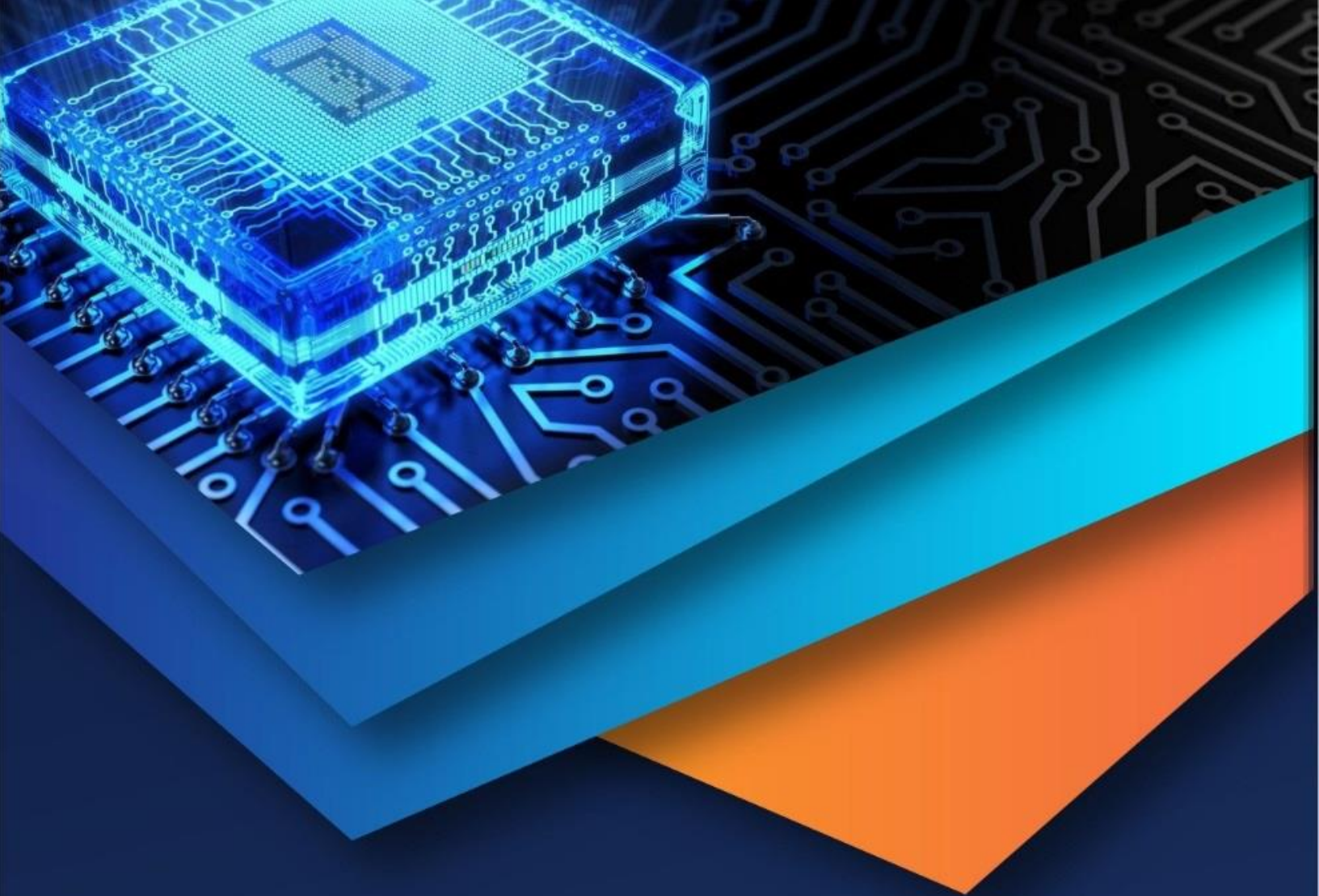
The authors would like to express their sincere and heartfelt gratitude to all those who have supported this work. We thank our respected guide, Mrs. Sowmya D, Assistant Professor, Department of Computer Science and Engineering, for her constant guidance, insightful suggestions, and continuous support throughout the development of this project. We are equally grateful to Dr. Jalesh Kumar, Professor and Head of the Department of Computer Science and Engineering, and Dr. Y. Vijaya Kumar, Principal, JNN College of Engineering, Shivamogga, for their unwavering support and inspiration. We also thank our families and friends for their constant encouragement, and we gratefully acknowledge JNN College of Engineering and its Department of Computer Science and Engineering for providing the resources, technical knowledge, and motivation to carry out this project.

REFERENCES

- [1] S. K. Sharma, A. Choudhary, and R. Kumar, "Driver drowsiness detection using deep learning and computer vision," IEEE Access, vol. 9, pp. 92187–92199, 2021.
- [2] M. R. Islam, M. S. Hossain, and A. Alamri, "Vision-based driver drowsiness detection using CNN," IEEE Access, vol. 9, pp. 34519–34530, 2021.
- [3] N. Gupta, R. Sharma, and P. Arora, "Real-time driver drowsiness detection using deep neural networks," in Proc. 12th Int. Conf. on Computing, Communication and Networking Technologies (ICCCNT), 2021, pp. 1–6.
- [4] P. Singh, S. Jain, and A. K. Singh, "Real-time driver drowsiness detection using facial features and eye aspect ratio," in Proc. Int. Conf. on Electronics and Sustainable Communication Systems (ICESC), 2022, pp. 1–6.
- [5] K. Verma, A. Mishra, and S. Srivastava, "Driver drowsiness detection using hybrid deep learning model," in Proc. Int. Conf. on Intelligent Systems and Smart Infrastructure (ICISSI), 2023, pp. 1–5.
- [6] J. Li, X. Wang, et al., "Vision-based driver drowsiness detection using lightweight CNN and attention mechanism," IEEE Trans. Intell. Transp. Syst., vol. 25, no. 2, pp. 1234–1245, 2024.



- [7] S. Kumar, R. Patel, et al., “Real-time driver drowsiness detection using YOLOv8 and facial landmark analysis,” *Multimedia Tools Appl.*, Springer, vol. 83, no. 5, pp. 14321–14338, 2024.
- [8] A. Singh, P. Verma, et al., “Edge-based driver drowsiness detection using MobileNet and IoT,” *IEEE Internet Things J.*, vol. 11, no. 4, pp. 6780–6791, 2024.
- [9] A. Bhanja, S. Paul, et al., “A real-time driver drowsiness detection system using deep learning,” *Int. J. Mach. Learn. Cybern.*, Springer, vol. 16, no. 1, pp. 123–136, 2025.
- [10] M. A. Albadawi, H. Alshammari, et al., “Real-time driver drowsiness detection using facial analysis,” *MDPI Sensors*, vol. 25, no. 3, p. 812, 2025.
- [11] O. Nasir, M. A. Khan, et al., “SAFE-DRIVE-AI: A CNN–LSTM–attention framework for driver drowsiness detection,” *Eng., Technol. Appl. Sci. Res. (ETASR)*, vol. 15, no. 1, pp. 1–8, 2025.
- [12] O. F. Hassan, A. Mahmood, et al., “Real-time driver drowsiness detection using transformer architectures: A novel deep learning approach,” *Sci. Rep.*, Nature Publishing Group, vol. 15, no. 1, p. 9821, 2025.
- [13] H. Chen, Y. Zhao, et al., “Driver drowsiness detection using multimodal deep learning,” *Neurocomputing*, Elsevier, vol. 580, p. 127450, 2025.
- [14] M. Rahman, T. Islam, et al., “Real-time driver drowsiness detection using hybrid CNN-BiLSTM model,” *Neural Comput. Appl.*, Springer, vol. 37, no. 3, pp. 2345–2360, 2025.
- [15] O. F. Hassan, A. F. Ibrahim, A. Gomaa, S. Alshathri, and W. El-Shafai, “Comprehensive benchmarking of attention-enhanced CNNs and transfer learning models for real-time driver drowsiness detection,” *IEEE Access*, vol. 14, pp. 12310–12330, 2026.
- [16] J. Menda, B. N. Marrapu, R. T. Bheri, and C. S. N. Chebodula, “Efficient and secure video streaming using multi-key cryptography,” *International Research Journal of Modernization in Engineering Technology and Science*, vol. 6, no. 11, Nov. 2024.
- [17] F. Chollet, *Keras: The Python Deep Learning Library*, 2017. [Online]. Available: <https://keras.io>.
- [18] NTHU Driver Drowsiness Detection Dataset, 2023. [Online]. Available: <http://cv.cs.nthu.edu.tw/php/callforpaper/datasets/DDD/>.
- [19] OpenCV, 2024. [Online]. Available: <https://opencv.org>.
- [20] dlib C++ Library, 2024. [Online]. Available: <http://dlib.net>.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)