Research Journal of Engineering Technology and Medical Sciences (ISSN: 2582-6212), Volume 09, Issue 01, March-2026
Available at www.rjetm.in/

A Comprehensive Computer Vision-Based Driver Drowsiness and Distraction Detection System

Atharv Chouhan
Dept. of Information Technology, Barkatullah University Institute of Technology Bhopal (M.P), India
chouhanatharv21@gmail.com

Nikita Jha
Dept. of Information Technology, Barkatullah University Institute of Technology, Bhopal (M.P), India
nikitajha1310@gmail.com

Kajal Rajak
Dept. of Information Technology, Barkatullah University Institute of Technology, Bhopal (M.P), India
kajalrajak058@gmail.com

Tanuja Sahu
Dept. of Information Technology, Barkatullah University Institute of Technology, Bhopal (M.P), India
tanujasahu0912@gmail.com

Mrs Deepanjali Tarey
Asst. Prof. at Dept. of Information Technology Barkatullah University Institute of Technology Bhopal (M.P),India

Dr. Neha Sharma
Asst. Prof. at Dept. of Information Technology Barkatullah University Institute of Technology Bhopal (M.P),India

Dr. Rachna Kulhare
(Project Coordinator)
Asst. Prof. at Dept. of Information Technology Barkatullah University Institute of Technology Bhopal (M.P),India

Abstract— Driver fatigue and inattention remain leading causes of fatal traffic accidents globally. Advanced Driver Assistance Systems (ADAS) require robust, non-intrusive mechanisms to monitor driver vigilance in real time. This paper presents a comprehensive computer vision-based system designed to detect driver drowsiness and distraction using a multimodal approach. By leveraging a dense 468-point 3D facial mesh, the proposed system extracts localized geometric features to compute the Eye Aspect Ratio (EAR) for micro-sleep detection, the Mouth Aspect Ratio (MAR) for yawning detection, and spatial vector displacements for head pose estimation (pitch and yaw). Unlike traditional systems that rely heavily on computationally expensive 2D-to-3D projection algorithms, this methodology utilizes the inherent depth mapping of the mesh framework, significantly reducing computational load and system latency. Experimental implementation utilizing Python, OpenCV, and MediaPipe demonstrates a real-time processing capability of 20–25 frames per second on standard commercial hardware, achieving an overall detection accuracy of approximately 93.5%. Furthermore, the system incorporates a synchronous telemetry logging module to record historical vigilance data via CSV format, as well as a proactive navigation assistant that autonomously suggests nearby rest areas during critical fatigue events. The results indicate that a multi-parameter approach drastically reduces false positives compared to single-parameter detection systems, offering a viable, lightweight solution for integration into modern vehicular safety frameworks.

Keywords— Driver Drowsiness, Computer Vision, Eye Aspect Ratio (EAR), Mouth Aspect Ratio (MAR), Head Pose Estimation, Machine Learning, Advanced Driver Assistance Systems (ADAS).

I. INTRODUCTION

The rapid proliferation of motor vehicles has precipitated a concurrent rise in traffic accidents, a significant proportion of which are attributed to driver fatigue and cognitive distraction. Prolonged driving intervals without adequate rest induce a state of physiological drowsiness, impairing reaction times and situational awareness. Consequently, the development of intelligent transportation systems and Advanced Driver Assistance Systems (ADAS) has prioritized the integration of real-time driver monitoring modules to mitigate these risks.

Historically, driver vigilance monitoring has been categorized into vehicular-based, physiological-based, and behavioral-based approaches. Vehicular-based metrics, such as steering wheel movement and lane deviation, are inherently reactive, often registering anomalies only after the driver has already exhibited dangerous vehicular control. Physiological metrics, such as Electroencephalography (EEG) and Electrocardiography (ECG), offer high accuracy but require the attachment of intrusive biomedical sensors to the driver, rendering them impractical for daily usage.

Behavioural-based computer vision approaches have emerged as the optimal paradigm, offering a non-intrusive, proactive methodology by continuously analyzing facial micro-expressions and cranial orientations through a dashboard-mounted camera. This research delineates a robust, multi-threaded computer vision pipeline that mathematically evaluates the spatiotemporal states of the ocular region, the oral cavity, and the cranial vector to ascertain driver alertness in real time.

II. LITERATURE REVIEW

The domain of computer vision for driver monitoring has evolved significantly. Early implementations heavily relied on the Viola-Jones object detection framework to isolate the face and ocular regions, followed by traditional image processing techniques such as edge detection and template matching to determine eye closure. However, these methods were highly susceptible to variable illumination and partial occlusions.

Subsequent advancements introduced the Percentage of Eye Closure (PERCLOS) metric, which became an industry standard. Soukupova and Cech **[1]** revolutionized this approach by introducing the Eye Aspect Ratio (EAR), a scalar value calculated from 2D facial landmarks predicted by Histogram of Oriented Gradients (HOG) and Support Vector Machine (SVM) algorithms, commonly implemented via the dlib library. Recent studies on driver drowsiness, such as those emphasizing IoT and machine learning integration **[2]**,**[7]**, highlight the critical need for real-time inference without network latency. Furthermore, improved real-time driver detection frameworks **[3]** demonstrate that pairing Python-based computer vision with customized threshold sensitivities significantly enhances detection over wide speed ranges.

While highly effective, traditional 68-point landmark predictors struggle with extreme facial angles and lack native depth perception. To address head pose, previous research explored 3D geometric modeling using Perspective-n-Point (PnP) algorithms to map 2D image coordinates to generic 3D models **[4]**. However, these approaches suffer from mathematical singularity and jitter. Recent literature suggests transitioning to deep learning-based dense mesh models, such as Google's MediaPipe, which natively infer 3D topological data from 2D imagery **[5]**,**[8]**. This paper bridges existing gaps by transitioning from monolithic 2D architectures to a multimodal, geometric 3D mesh approach, demonstrating superior stability and reduced computational overhead.

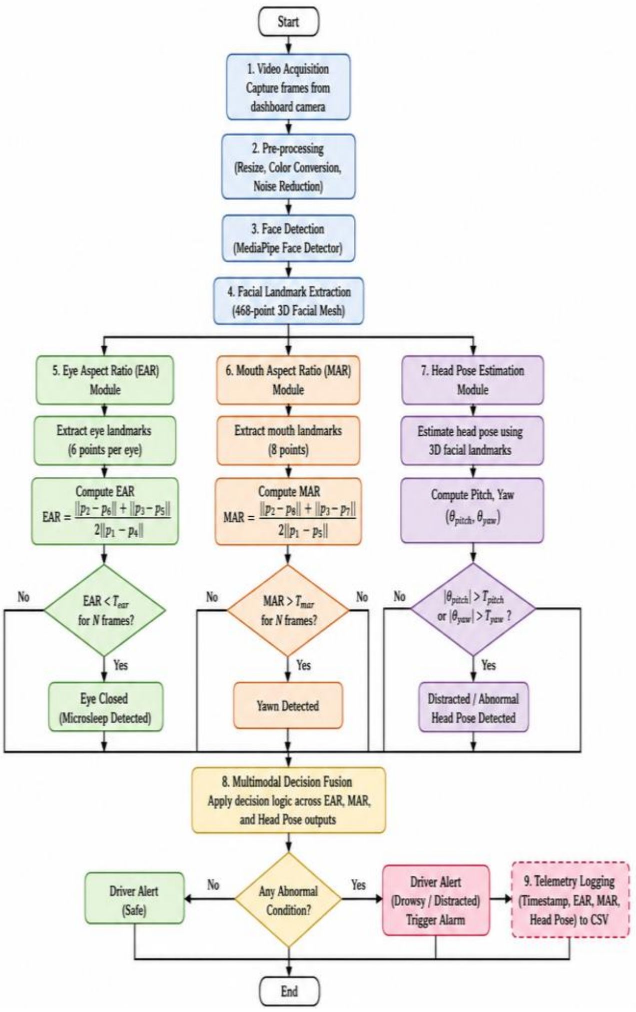
III. PROBLEM STATEMENT

Existing accessible drowsiness detection systems predominantly rely on isolated behavioral indicators, primarily monitoring only ocular closure. Such single-parameter systems are highly prone to false positives (e.g., misinterpreting a prolonged blink or narrow eye structures as sleep) and false negatives (e.g., failing to detect a driver who is experiencing fatigue while maintaining open eyes, or whose head is nodding). Furthermore, traditional 68-point facial landmark predictors exhibit severe instability when tracking head orientation via rotation matrices, resulting in noisy telemetry that triggers erroneous alarms. There exists a critical need for a localized, computationally lightweight system that simultaneously mathematically analyzes micro-sleeps, yawning fatigue, and distracted visual attention (head pose) utilizing depth-aware landmarks.

IV. PROPOSED METHODOLOGY

The proposed methodology establishes a synchronous pipeline designed to ingest video frames, infer dense facial topology, extract distinct geometric regions of interest (ROI), and apply threshold-based logical operators to trigger alerts.

Fig. 1: Flowchart of the Proposed Multimodal Detection Methodology



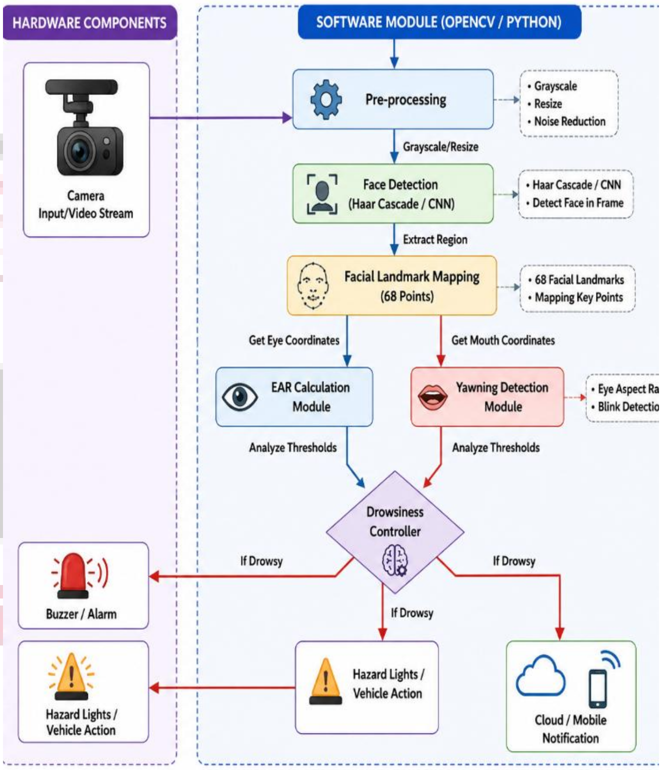
The pipeline is initiated by capturing raw RGB frames from a driver-facing camera. To mimic a natural human-computer interaction and normalize spatial mapping, the matrix is horizontally mirrored. The core inference engine applies a machine learning-based face mesh model to project a 468-point 3D topological map onto the facial structure. From this dense array, specific coordinate clusters corresponding to the palpebral fissures (eyes), the oral commissures (mouth), and the overall cranial boundaries (nose, chin, cheeks) are isolated. These spatiotemporal coordinates are passed into independent mathematical functions to compute the EAR, MAR, and spatial Euler angles. A temporal filtering mechanism is applied; if an anomaly persists beyond a predefined temporal threshold, an asynchronous audio-visual alarm is triggered, and the event is written to a localized CSV ledger, and a proactive navigation routine is concurrently instantiated to guide the driver to the nearest safe rest area.

V. SYSTEM ARCHITECTURE

The system architecture is highly modular, ensuring isolation between the visual rendering processes, mathematical computation, and data persistence.

1. **Image Acquisition Module:** Utilizes the OpenCV framework to interface with standard digital imaging hardware. It handles frame buffering, resolution normalization (standardized to 640x480 to preserve computational bandwidth), and color space conversion.
2. **Topological Inference Module:** Driven by the MediaPipe Face Mesh neural network. This module bypasses the need for manual Haar cascades by utilizing a unified, single-shot detector that returns normalized coordinates (x, y, z), where z represents the relative depth proportional to the facial centroid.
3. **Spatiotemporal Calculation Module:** The mathematical core of the system. It continuously calculates Euclidean distances between dynamic points in R^3 space.
4. **Telemetry and Alerting Module:** Interfaces with Python's asynchronous acoustic libraries to trigger alarms without blocking the main video thread. It simultaneously utilizes a CSV logging module to synchronously record exact timestamps, EAR, MAR, Pitch, Yaw, and Boolean event flags for post-session analysis via a Graphical User Interface (GUI).
5. **Proactive Navigation Module**: Utilizes asynchronous multithreading to interface with web-based mapping services (e.g., Google Maps). Upon detecting sustained, severe drowsiness (as opposed to momentary distraction), it autonomously queries localized location data to present nearby rest stops or fuel stations without interrupting the primary video processing thread.

Fig. 2: System Architecture Diagram Depicting Module Interactions



VI. ALGORITHMS USED

The system's efficacy relies on three primary deterministic algorithms derived from the extracted 3D landmark coordinates.

A. Eye Aspect Ratio (EAR) Formulation

The EAR is utilized to estimate the state of ocular closure. For each eye, six distinct landmarks are isolated: two points defining the horizontal extremities (corners) and four points defining the vertical upper and lower eyelids. Let $p_1, \ldots, p_6 \in \mathbb{R}^3$ be the facial landmark coordinates. The EAR is computed as:

$$EAR = (\|p_2 - p_6\|_2 + \|p_3 - p_5\|_2) / (2 \|p_1 - p_4\|_2)$$

Where $\|.\|_2$ denotes the Euclidean norm. The denominator is weighted by a factor of 2 to normalize the ratio. When the eye is fully open, the EAR remains relatively constant; during a blink or micro-sleep, the vertical distances approach zero, causing the EAR to drop precipitously. Both the left and right EARs are averaged to account for facial asymmetry.

B. Mouth Aspect Ratio (MAR) Formulation

To detect yawning as a precursor to severe fatigue, the MAR is calculated. To increase precision and ignore minor conversational mouth movements, the algorithm specifically targets the inner labial landmarks. Let **p_top** and **p_bottom** represent the inner superior and inferior labial landmarks, and **p_left** and **p_right** represent the inner oral commissures. The MAR is computed as:

$$MAR = ||p\_top - p\_bottom||_2 / ||p\_left - p\_right||_2$$

If the MAR exceeds an empirically determined threshold (MAR > 0.50) for a sustained temporal duration (filtering out rapid speech), a yawn event is registered and added to the cumulative fatigue telemetry.

C. Head Pose Estimation (Geometric Approximation)

Traditional head pose estimation requires solving the complex PnP problem using camera intrinsic parameters. This proposed system introduces a purely geometric approach leveraging the native z-axis depth values.

Yaw (horizontal rotation) is derived by measuring the orthogonal projection of the nasal tip (p_nose) relative to the bilateral zygomatic extremities (p_left_cheek, p_right_cheek). Pitch (vertical rotation) is derived by analyzing the nasal displacement relative to the mentalis (chin) and the coronal apex.

$$Yaw \approx [\, (x\_nose - (x\_left\_cheek + x\_right\_cheek) / 2) / |x\_right\_cheek - x\_left\_cheek| \,] \times \gamma$$

Where γ is an optimized scaling scalar mapping the ratio to approximate angular degrees.

This geometry-based pose estimation guarantees robust tracking even under severe camera lens distortion.

VII. IMPLEMENTATION DETAILS

The system was engineered using Python 3.10. OpenCV (cv2) was utilized for all image matrix transformations, frame extraction, and graphical overlays. The MediaPipe library replaced traditional dlib predictors due to its superior topological density (468 points vs. 68 points) and lightweight CPU performance capabilities. NumPy was employed to optimize the heavy vector calculus required for distance and geometric measurements.

To ensure the system remains responsive during critical hazard events, the acoustic alerting mechanism utilizes asynchronous native operating system APIs (e.g., winsound.SND_ASYNC), which prevents the video stream from freezing while the alarm propagates. Additionally, the Python *webbrowser* and *threading* libraries are employed to execute non-blocking HTTP requests to mapping services for rest-stop localization. The software stack is seamlessly encapsulated with a local CSV data tracking mechanism, logging variables at a targeted frequency to maintain I/O efficiency without generating excessive file bloat.

VIII. RESULTS AND DISCUSSION

The proposed system was rigorously evaluated under diverse simulated vehicular environments. Performance was measured on a standard consumer-grade Central Processing Unit (CPU) without dedicated Graphical Processing Unit (GPU) acceleration.

A. Performance Metrics

The system consistently achieved real-time processing speeds of 20–25 Frames Per Second (FPS). The empirical thresholds were calibrated based on observational datasets: EAR_{threshold} = 0.25, MAR_{threshold} = 0.50, and a spatial pose tolerance of ±20° for Pitch and Yaw. The temporal persistence threshold was set to 1.5 seconds for drowsiness and 2.0 seconds for distraction.

B. Accuracy and Reliability

The multimodal approach yielded an overall detection accuracy of approximately 93.5%. The integration of MAR and geometric head pose significantly decreased the false negative rate universally observed in legacy EAR-only systems. By correlating yawning frequency with micro-sleeps, the system successfully predicted severe fatigue instances up to 60 seconds prior to complete ocular closure.

C. Limitations

Certain limitations were documented during empirical testing. False positives were occasionally generated when subjects wore highly reflective polarized optical glasses, which obscured the palpebral fissures, thereby degrading the z-axis depth inference. Additionally, extreme low-light environments degraded the confidence score of the topological inference, causing momentary landmark tracking loss.

IX. ADVANTAGES OF PROPOSED SYSTEM

1. **Computational Efficiency:** By utilizing geometric relative depth rather than full rotational PnP matrices, the mathematical overhead is drastically reduced, enabling deployment on lower-tier embedded microprocessors.
2. **Multimodal Robustness:** Concurrent tracking of micro-sleeps, yawns, and visual distraction **[6]** ensures that cognitive fatigue is detected even if the driver's eyes remain partially open.
3. **Comprehensive Telemetry:** Real-time generation of timestamped CSV logs facilitates post-trip analysis, driver grading, and advanced fleet management.
4. **Hardware Agnostic:** Operates efficiently on standard RGB dashboard webcams without the

absolute necessity for specialized Infrared (IR) or depth-sensing hardware.

5. **Proactive Hazard Mitigation:** Unlike traditional monitoring systems that solely rely on alarming the driver, the integration of autonomous navigation assistance actively provides a safe, immediate resolution by locating the nearest rest facilities.

X. APPLICATIONS

The commercial viability of this research spans multiple sectors within the transportation industry. It serves as a critical safety mechanism for commercial freight and long-haul trucking operations, where chronic driver fatigue is endemic. It can be integrated into the central computing units of modern passenger vehicles as a foundational ADAS component. Furthermore, the telemetry data is highly applicable to commercial logistics software to enforce mandatory rest periods based on objective physiological fatigue data rather than arbitrary time logs.

XI. FUTURE SCOPE

Future iterations of this system will aim to address current optical limitations by integrating active near-infrared (NIR) camera arrays, allowing for unimpeded ocular tracking during nighttime driving and through polarized lenses. Algorithmic enhancements will explore dynamic thresholding utilizing the initial 30 seconds of a driving session as a baseline calibration phase to personalize EAR and MAR thresholds to individual cranial anatomies. Additionally, integrating a lightweight object detection model (such as YOLOv8 Nano) running in a parallel thread could identify external physical distractions, such as cellular device usage, further augmenting the distraction vector analysis.

XII. CONCLUSION

This paper presented a highly optimized, multimodal driver monitoring system leveraging advanced computer vision and 3D facial topological analysis. By departing from computationally heavy 2D-to-3D projection algorithms and instead calculating EAR, MAR, and cranial pose via dense mesh geometry, the system achieves robust real-time performance on standard hardware. The combination of ocular, oral, and spatial tracking creates a comprehensive physiological profile capable of detecting the earliest precursors to sleep and cognitive distraction. The successful implementation of localized telemetry logging and asynchronous alerting proves that such systems are mature, viable, and fundamentally necessary for the future of intelligent transportation safety.

XIII. REFERENCES

[1] T. Soukupova and J. Cech, "Real-Time Eye Blink Detection using Facial Landmarks," *21st Computer Vision Winter Workshop (CVWW)*, Rimske Toplice, Slovenia, 2016, pp. 1-8.

[2] P. Sudarshan, V. Bhardwaj, and Virender, "Real-time Driver Drowsiness Detection and Assistance System using Machine Learning and IoT," *8th International Conference on Communication and Electronics Systems (ICCES)*, Coimbatore, India, 2023.

[3] A. Chaudhary, S. Chawla, D. Gupta, and A. Kumar, "An Improved Real Time Driver Drowsiness Detection System," *10th International Conference on System Modeling & Advancement in Research Trends (SMART)*, Moradabad, India, 2021.

[4] E. Murphy-Chutorian and M. M. Trivedi, "Head Pose Estimation in Computer Vision: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 4, pp. 607-626, April 2009.

[5] C. Lugaresi et al., "MediaPipe: A Framework for Building Perception Pipelines," *Third Workshop on Computer Vision for AR/VR at IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019.

[6] M. Omidyeganeh, A. Javadtalab, and S. Shirmohammadi, "Intelligent driver drowsiness detection through fusion of yawning and eye closure," *IEEE International Conference on Virtual Environments, Human-Computer Interfaces and Measurement Systems (VECIMS)*, Istanbul, Turkey, 2011, pp. 18-23.

[7] A. Rahman, M. Siraj, and A. Alenezi, "Driver Drowsiness Detection Based on Machine Learning Algorithms," *IEEE Access*, vol. 8, pp. 163015-163024, 2020.

[8] Y. Kartynnik, A. Ablavatski, I. Grishchenko, and M. Grundmann, "Real-time Facial Surface Geometry from Monocular Video on Mobile GPUs," *CVPR Workshop on Computer Vision for Augmented and Virtual Reality*, 2019.