

AN INTELLIGENT REAL-TIME DRIVER DROWSINESS DETECTION FRAMEWORK USING EYE CLOSURE, YAWNING, AND HEAD POSE ANALYSIS

Shaman Ali ^{*1}, Syed Muzammil Hussain², Muskan Laghari³, Sikandar Ali Bughio⁴,
Sajjad Ahmed⁵, Zoya Naz⁶

^{*1,2,3,4,5,6}Information Technology Centre, Sindh Agriculture University, Tandojam

DOI: <https://doi.org/10.5281/zenodo.23007504>

Keywords

driver drowsiness detection; eye aspect ratio; PERCLOS; yawning detection; head-pose estimation; temporal fusion; driver monitoring system; real-time computer vision

Article History

Received: 01 June 2026

Accepted: 23 August 2026

Published: 28 September 2026

Copyright Shaman Ali, Syed Muzammil Hussain

Corresponding Author: *
Sajjad Ahmed, Muskan Laghari, Zoya Naz, Sikandar Ali Bughio

Abstract

Drowsy driving remains a persistent cause of road crashes, and camera-based monitoring is attractive because it needs no contact with the driver. Many recent vision systems report high accuracy for single cues such as eye state, yet they often decide from individual frames, rely on fixed thresholds, or are evaluated on image-level splits in which the same person appears in training and test data. This study proposes a lightweight, temporally aware framework that combines eye closure, yawning and head pose into one interpretable drowsiness score and maps it to four progressive alert levels. Eye aspect ratio (EAR), inner-lip mouth aspect ratio (MAR) and perspective-n-point head pose are computed from a 478-point face mesh; each cue is converted into bounded evidence over time using blink-filtered PERCLOS and closure run-length, duration-constrained yawn events, and sustained pitch deviation with nod counting; the evidence streams are fused by a convex weighted sum (weights 0.55, 0.15 and 0.30), gated by landmark quality, and passed to a decision layer with persistence, hysteresis and a 2.0-s microsleep override. We implemented the framework in Python and evaluated it in three ways. On eight unseen subjects of the MRL Eye dataset (17,983 images, subject-independent split), the eye-state component reached an accuracy of 0.934, F1-score of 0.926, MCC of 0.871 and ROC-AUC of 0.987, with the largest losses for eyewear (F1 0.875) and strong reflections (F1 0.758). An automated simulation test confirmed that blinks and speech-like mouth openings did not trigger alerts while sustained drowsiness escalated to the critical level. In two logged live sessions (about 10 minutes), none of the 41 blink-length closures (shorter than 0.5 s) triggered the microsleep override; all 22 closures of 2.0 s or longer triggered the critical warning, and the complete pipeline ran at 54.7–86.2 frames per second on a laptop CPU. The results support the feasibility of a transparent, CPU-only multi-cue design and identify detection latency for gradual drowsiness as the main parameter to tune in driving-data validation.

1. INTRODUCTION

Driving is a sustained-attention task, and the ability to sustain attention degrades as sleep pressure builds. The degradation appears gradually: longer and slower blinks, periods of

partial eyelid closure, repeated yawning, a head that drifts forward and is jerked back, and eventually brief lapses in which the eyes stay closed for a second or more. All of these are

visible to an in-cabin camera, which is why camera-based monitoring has become the dominant approach to non-intrusive drowsiness detection (Kolus, 2024; Fonseca & Ferreira, 2025). Electrodermal signals such as EEG remain the laboratory reference for vigilance (Balam, 2024; Divvala & Mishra, 2024; Obaidan et al., 2024; Zayed et al., 2024; Feng et al., 2025; Lu et al., 2025; Zayed et al., 2025; Rahman et al., 2026; Xu et al., 2026), and cardiac, electrodermal, radar and thermal sensing have also been studied (Alguindigue et al., 2024; Fujiwara et al., 2024; Jiao et al., 2024; Siddiqui et al., 2024; Lopez et al., 2026; Nomura et al., 2026; Vaussenat et al., 2026). Ayas et al. (2026) combined driving performance with physiology to separate drowsiness from high cognitive load. These approaches, however, require electrodes or dedicated hardware (El Sahmarany et al., 2026), whereas many vehicles already have a single low-cost camera.

A camera observes behavior, not sleepiness, and every visual cue has innocent look-alikes. A blink resembles the start of a microsleep, talking and laughing open the mouth, and a mirror check moves the head. A detector that reacts to single-frame events produces frequent false alarms, which drivers soon learn to ignore; a detector tuned to avoid them may warn too late. The problem addressed here is therefore not whether eye, mouth and head cues can be measured, but how to turn noisy per-frame measurements into a decision that separates brief events from sustained drowsiness, remains interpretable, runs on ordinary hardware, and is validated on people the system has not seen.

The combination of eye, mouth and head cues is not new, and we do not claim it as a contribution. Recent studies combine eye and mouth features (Florez et al., 2024; Makhmudov et al., 2024; Essahraui et al., 2025; Ersoy et al., 2026) and, in at least one low-light system, head tilting, blinking and yawning with temporal factors (Saxena et al., 2026). This study contributes a lightweight temporal multi-cue decision framework whose every step is explicit and inspectable. Specifically, the paper makes four contributions:

- C1. Temporal evidence functions for each cue that convert frame-level EAR, MAR and head

angles into bounded evidence using duration constraints, so that blinks, speech, and brief head movements do not raise the score (Section 3.3).

- C2. A transparent fusion rule with non-negative weights summing to one, a landmark-quality gate, and a four-level alert policy with persistence, hysteresis and a microsleep override (Sections 3.4–3.5).

- C3. A subject-independent evaluation of the eye-state component on MRL Eye, including per-subject and per-condition analysis (Section 5.1).

- C4. Verification of the complete CPU-only pipeline through an automated simulation test and logged live sessions, reporting alert behavior and throughput (Sections 5.2–5.4).

The study addresses five research questions. RQ1: How accurately does the eye-state component detect eye closure on unseen subjects? RQ2: How robust is it to eyewear, reflections and lighting? RQ3: Does the temporal decision layer suppress isolated events while escalating sustained drowsiness? RQ4: Does the progressive alert mechanism behave as designed in live use? RQ5: Can the complete pipeline run in real time on a CPU without a GPU?

Section 2 reviews recent work and the gap it leaves. Section 3 presents the framework, Section 4 the materials and evaluation protocol, and Section 5 the results. Sections 6 to 8 discuss the findings, their limitations and ethical aspects, and Section 9 concludes.

2. Related Work

2.1 Vision-Based Detection and Single Cues

Two systematic reviews frame current practice: Kolus (2024) surveyed eye-activity measures, and Fonseca and Ferreira (2025) reviewed deep-learning models, together documenting a shift from handcrafted geometric measures to learned representations and a wide spread of datasets and protocols. Primary studies range from classical pipelines with Haar and HOG features (Quiles-Cucarella et al., 2024) and landmark-based machine learning (Essel et al., 2024; Essahraui et al., 2025; Delwar et al., 2025) to deep models designed for driver-assistance systems (Yang & Yi, 2024; Vidyathanathan Dixith et al., 2025), IoT-oriented designs (Das et al.,

2024), complete alerting systems (Bhanja et al., 2025; Alvarez Oviedo et al., 2025) and related tasks such as driver identification (Díaz-Santos et al., 2024), multi-method detection for automated vehicles (Beles et al., 2024), abnormal-expression detection (Yao et al., 2024) and supervised-learning frameworks over extracted features (Harb, 2025).

Eye closure is the most frequently used cue. The eye aspect ratio (Soukupová & Čech, 2016; foundational reference) and PERCLOS (Dinges & Grace, 1998; foundational reference) remain standard measures; recent work proposes adaptive eye ratios (Lamba et al., 2025), grades fatigue by the ratio of eye opening to closing (Zhao et al., 2025), and classifies eye crops with transfer-learning and custom CNNs (Madni et al., 2024; Salem & Waleed, 2024; Venkateswarlu & Ch, 2024; Jebrailey et al., 2024; Ramzan et al., 2024). Hassan et al. (2025) reported 99.15% eye-state accuracy with a Vision Transformer on MRL Eye using an 80/20 image split, with a real-time loop that raises an alarm after a fixed count of closed-eye frames. Yawning is slower and more ambiguous: MAR is cheap and has been combined with CNN eye classifiers on embedded hardware (Florez et al., 2024), but talking and singing, which YawDD deliberately includes (Abtahi et al., 2014; foundational reference; frame-level labels in Mujtaba et al., 2025), open the mouth repeatedly. Ersoy et al. (2026) calibrated driver-specific EAR and MAR thresholds and reported a 2–3% accuracy gain over fixed thresholds. Head pose appears mostly as an auxiliary or distraction cue (Hollósi et al., 2024; Ciesla & Ostermayer, 2024; Kazemi et al., 2025; Hollósi et al., 2025), although forward droop and nodding are classic signs of sleep onset; pose is usually estimated by solving a perspective-n-point problem (Lepetit et al., 2009; foundational reference).

2.2 Deep, Lightweight and Temporal Models

Transformers and attention models dominate recent accuracy gains (Jarndal et al., 2025; El-

Nabi et al., 2025; Hassan et al., 2026; Nurnoby & El-Alfy, 2026), alongside hybrid CNNs (Al-Gburi et al., 2025a, 2025b), semi-supervised learning from facial cues (Rahmani et al., 2026) and multimodal fusion with ear-EEG or uncertainty estimation (Mai et al., 2025; Zhang et al., 2025). Saxena et al. (2026) used attention-augmented InceptionV3 with low-light enhancement and explainability tools. A parallel line pursues efficiency: joint face and landmark localisation (Wu et al., 2024; Huang et al., 2026), detectors robust to difficult lighting (Liu et al., 2025), lightweight CNNs (Venkateswarlu & Ch, 2024), handcrafted features for edge devices (Owen & Surantha, 2025), compact attention networks with edge deployment (Hussein et al., 2026) and reconfigurable embedded systems (Belakhddhar, 2026). Temporal modelling is either learned, as in inception-attention time models (Lee et al., 2025), video models over key facial features (Yang et al., 2024), recurrent multimodal models (Huang et al., 2024) and the temporal deep belief network proposed with the NTHU-DDD benchmark (Weng et al., 2017; foundational reference), or rule-based, typically as consecutive-frame counters and PERCLOS windows. Multi-cue systems often fuse cues inside a classifier or ensemble (Xu et al., 2025), making it hard to see which cue produced a given alert.

2.3 Research Gap

Table 1 summarizes representative studies and Table 2 the gaps this study targets. Across the reviewed work, strong per-frame classifiers are often followed by simple counters; fixed thresholds are known to transfer poorly between individuals; image-level random splits can place the same person in training and test data; and runtime is reported on heterogeneous hardware. What remains insufficiently examined is a transparent, reproducible decision layer that combines the three cues over time, is evaluated on unseen subjects, and runs on a CPU.

Table 1. Representative 2024–2026 studies (reported figures are those of the original authors).

Study	Year	Cues/signals	Method family	Deployment focus	Relevance to this study
Hassan et al.	2025	Eye state	ViT, Swin, CNNs (99.15% on MRL Eye)	Webcam loop	Strong classifier; counter-based decision
Saxena et al.	2026	Head tilt, blink, yawn	InceptionV3 + attention, low-light transformer	Low light	Closest multi-cue temporal system; heavier model
Ersoy et al. (preprint)	2026	EAR, MAR, PERCLOS, head	Personalized thresholds + CNNs	Real time	Threshold personalization (+2–3%)
Florez et al.	2024	Eyes, MAR	CNN + MAR	Embedded	Eye + mouth; no head cue
Makhmudov et al.	2024	Yawning, eye state	Machine learning	Real time	Two-cue design
Lamba et al.	2025	Adaptive eye ratio	Geometric	Road safety	Adaptive single cue
Xu et al.	2025	Hybrid facial features	Ensemble learning	—	Fusion inside classifier
Yang et al.	2024	Key facial features	Video deep model	—	Learned temporal model
Lee et al.	2025	Temporal sequence	Inception–attention time model	—	Learned temporal model
Wu et al.	2024	Face + landmarks	Joint localization	Efficiency	Efficient front-end
Jarndal et al.	2025	Facial images	Vision Transformer	Real time	Transformer cost on CPU
Hussein et al.	2026	Facial images	Compact dual attention	Edge, calibration	Edge reference
Owen & Surantha	2025	Handcrafted features	Classical CV	Edge devices	Supports geometric design
Kazemi et al.	2025	Eye tracking + head pose	Readiness estimation	Automated driving	Head pose in monitoring

Table 2. Research gaps and this study's response.

Gap	Evidence	Response here
Opaque or hand-tuned decision rules	Counters after strong classifiers (Hassan et al., 2025); fusion inside classifiers (Xu et al., 2025)	Explicit per-cue evidence, convex fusion, logged and inspectable scores
Isolated events trigger alarms	Blinks, speech and glances resemble drowsy behavior (Abtahi et al., 2014)	Blink filter, yawn duration rule, yaw gating, persistence and hysteresis

Gap	Evidence	Response here
Person leakage in evaluation	80/20 image splits are common (Hassan et al., 2025)	Subject-independent split on MRL Eye; per-subject reporting
Fixed thresholds generalize poorly	Lamba et al. (2025); Ersoy et al. (2026)	Calibration phase; per-subject and per-condition analysis
Heterogeneous runtime reporting	Owen & Surantha (2025); Hussein et al. (2026)	Full pipeline measured on a CPU-only laptop
Single on/off alarms	Bhanja et al. (2025); Hassan et al. (2025)	Four-level policy with microsleep override, verified in logs

3. Proposed Framework

3.1 Overview and Architecture

The framework rests on one principle: no single frame should be able to trigger a warning. A frame contributes measurements; only persistent patterns in those measurements contribute evidence; and only persistent evidence changes the alert level. Two exceptions are deliberate. A continuous eye closure longer than a critical duration escalates directly to the critical level, because a microsleep at speed is dangerous whatever the other cues say, and a drop in landmark quality holds the score instead of feeding it unreliable values.

Figure 1 shows the architecture. Frames from an in-cabin camera are resized to 640×480 pixels and passed to Media Pipe Face Land marker, which fits a 478-point face mesh based on the real-time facial geometry model of Kartynnik et al. (2019; foundational reference). Three cue modules turn landmark subsets into bounded evidence values $E(t)$, $Y(t)$ and $H(t)$ in $[0, 1]$; a fusion layer combines them into a drowsiness score $D(t)$; and a decision layer maps $D(t)$ to alert levels 0–3 and drives audio and voice output. A short calibration phase at start-up, during which the driver sits normally with eyes open and looks ahead, establishes the neutral head pose and per-driver baselines.

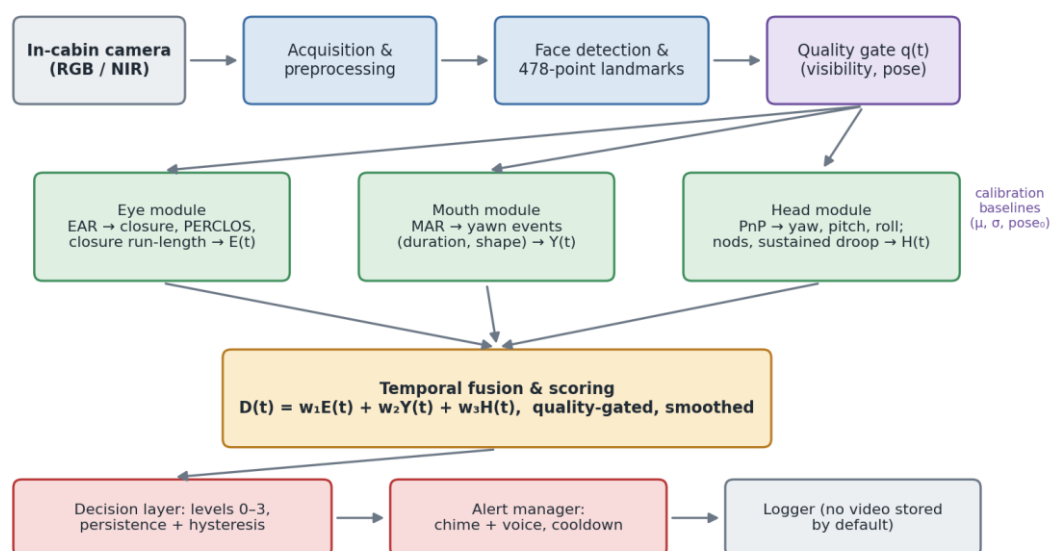


Figure 1. System architecture. Frame-level perception feeds three cue modules whose evidence streams $E(t)$, $Y(t)$ and $H(t)$ are fused into a quality-gated score $D(t)$; the decision layer maps $D(t)$ to alert levels 0–3. Calibration baselines are estimated at start-up.

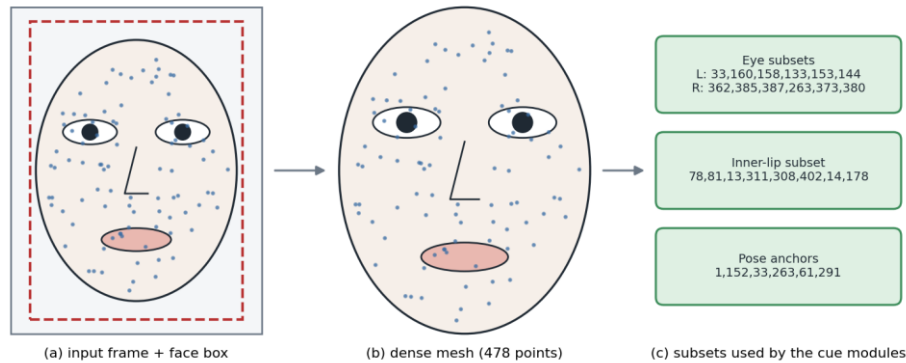


Figure 2. Face detection and landmark extraction: (a) face localization, (b) the 478-point mesh, (c) the landmark subsets used by the eye, mouth and head-pose modules.

3.2 Frame-Level Measurements

Eye closure. Six landmarks per eye (Figure 3) define the eye aspect ratio (Soukupová & Čech, 2016; foundational reference):

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

where p_1 and p_4 are the eye corners and p_2 , p_3 , p_6 , p_5 lie on the upper and lower lids (mesh indices 33, 160, 158, 133, 153, 144 and 362, 385, 387, 263, 373, 380). The numerator falls toward zero as the lids meet while the eye width changes little, so the ratio is approximately scale-

invariant. When the head turns, the far eye is foreshortened, so both eyes are combined with visibility weights v_L and v_R returned by the mesh model,

$$EAR_t = \frac{v_L EAR_{L,t} + v_R EAR_{R,t}}{v_L + v_R} \quad (2)$$

and the result is smoothed with an exponential moving average (EMA), where β controls how much of the previous value $\tilde{x}_{t-1}$ is retained:

$$\tilde{x}_t = \beta \tilde{x}_{t-1} + (1 - \beta) x_t, \quad 0 \leq \beta < 1 \quad (3)$$

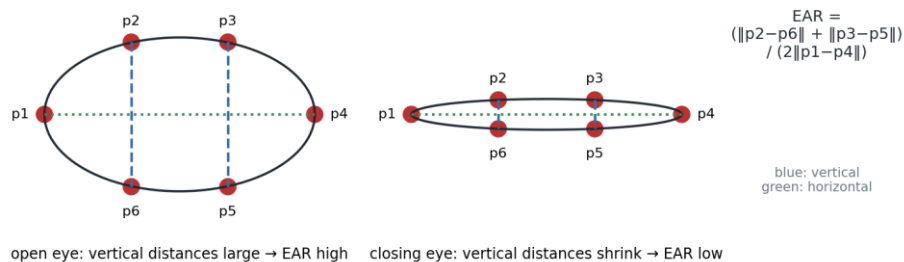


Figure 3. Eye aspect ratio. Vertical lid distances (blue) shrink during closure while the horizontal width (green) stays nearly constant.

Yawning. The mouth module uses eight inner-lip landmarks (78, 81, 13, 311, 308, 402, 14, 178; m_1 - m_8 in Figure 4):

$$MAR = \frac{\|m_2 - m_8\| + \|m_3 - m_7\| + \|m_4 - m_6\|}{3 \|m_1 - m_5\|} \quad (4)$$

where m_1 and m_5 are the inner mouth corners and the numerator terms are vertical openings at the left, center and right. Inner-lip points are preferred because the outer contour moves with smiling even when the mouth is closed.

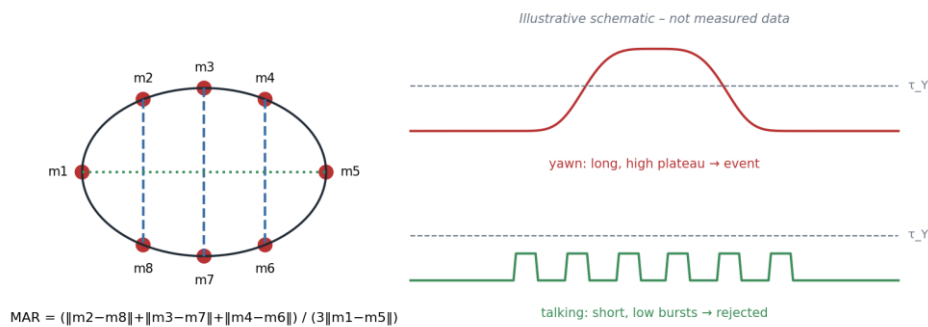


Figure 4. Mouth aspect ratio. Left: inner-lip landmarks. Right: illustrative schematic (not measured data) contrasting a yawn, a long high plateau accepted as an event, with talking, short bursts rejected by the duration rule.

Head pose. Six mesh points (nose tip 1, chin 152, eye corners 33 and 263, mouth corners 61 and 291) are fitted to a generic 3-D face model with a perspective-n-point solver (Lepetit et al., 2009; foundational reference):

$$s[u_i \ v_i \ 1]^T = K[R|t][X_i \ Y_i \ Z_i \ 1]^T, \quad i = 1, \dots, 6 \quad (5)$$

where (u_i, v_i) are image coordinates, (X_i, Y_i, Z_i) the model coordinates, K the camera matrix (focal length approximated by image width when calibration is unavailable), R and t the rotation

and translation, and s a scale factor. Pitch θ , yaw ψ and roll ϕ follow from the elements r_{ij} of R ,

$$\theta = \text{atan2}(r_{32}, r_{33}), \quad \psi = \text{atan2}(-r_{31}, \sqrt{r_{32}^2 + r_{33}^2}), \quad \phi = \text{atan2}(r_{21}, r_{11}) \quad (6)$$

and are expressed relative to the neutral pose $(\theta_0, \psi_0, \phi_0)$ estimated during calibration (Figure 5):

$$\Delta\theta_t = \tilde{\theta}_t - \theta_0, \quad \Delta\psi_t = \tilde{\psi}_t - \psi_0, \quad \Delta\phi_t = \tilde{\phi}_t - \phi_0 \quad (7)$$

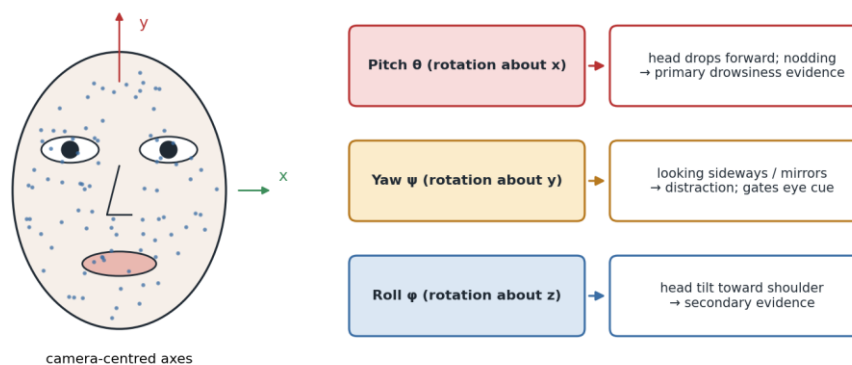


Figure 5. Head-pose estimation. Pitch captures forward droop and nodding (primary drowsiness evidence); yaw captures sideways glances (used to gate the other cues); roll captures lateral tilt (secondary evidence).

3.3 Temporal Evidence

Temporal analysis separates everyday behavior from drowsiness. Blinks last a few hundred milliseconds, a spoken syllable opens the mouth for a fraction of a second, and a mirror check turns the head for a second or two, whereas

drowsiness accumulates as longer closures, repeated yawns and persistent droop (Figure 6). Eye evidence. Closure and its current duration are

$$c_t = 1[\widehat{\text{EAR}}_t < \tau_E], \quad L_t = (L_{t-1} + 1) c_t \quad (8)$$

where $c_t = 1$ when the smoothed EAR is below the threshold τ_E and L_t counts consecutive closed frames (converted to seconds). PERCLOS is computed over a window of N_W frames counting only closures longer than a blink cut-off B_{blink} , a deliberate modification of classical PERCLOS (Dinges & Grace, 1998):

$$\text{PERCLOS}_t = \frac{1}{N_W} \sum_{k=t-N_W+1}^t c_k 1[L_k^* \geq L_{\text{blink}}] \quad (9)$$

where L_k^* is the length of the closure run containing frame k . Eye evidence combines slow accumulation with the current event:

$$E_t = \lambda \min\left(1, \frac{\text{PERCLOS}_t}{P_{\max}}\right) + (1 - \lambda) \min\left(1, \frac{L_t}{L_{\max}}\right) \quad (10)$$

where $P_{\max}$ and $L_{\max}$ are the PERCLOS level and closure duration, treated as full evidence and λ balances the two terms.

Yawn evidence. A yawn event j over $[a_j, b_j]$ is accepted only when the smoothed MAR stays above τ_Y throughout, the interval lasts at least d_Y seconds, and the peak exceeds $\kappa \tau_Y$ with $\kappa > 1$:

$$\text{yawn}_j \Leftrightarrow \widehat{\text{MAR}}_t > \tau_Y \quad \forall t \in [a_j, b_j], \quad b_j - a_j \geq d_Y, \quad \max_{t \in [a_j, b_j]} \widehat{\text{MAR}}_t > \kappa \tau_Y \quad (11)$$

Speech produces short, shallow openings that fail the duration conditions. Because yawning is a slow sign, its evidence decays over minutes, with decay constant T_Y and N_Y recent yawns treated as full evidence:

$$Y_t = \min\left(1, \frac{1}{N_Y} \sum_{j: b_j \leq t} e^{-(t-b_j)/T_Y}\right) \quad (12)$$

Head evidence. A frame is flagged as a drowsy posture when the head has dropped forward without a large sideways turn, or is strongly tilted:

$$h_t = 1[\Delta\theta_t > \tau_P \wedge |\Delta\psi_t| < \tau_\psi] \vee 1[|\Delta\phi_t| > \tau_R] \quad (13)$$

where τ_P , τ_ψ and τ_R are pitch, yaw and roll thresholds. Sustained droop (flagged runs R_k of at least d_H seconds) and nod events (a fast pitch drop followed by recovery) form the head evidence, with $n_{\text{nod}}(t)$ nods in the window, N_{nod} nods treated as full evidence and α balancing the terms:

$$H_t = \alpha \frac{1}{N_W} \sum_{k=t-N_W+1}^t h_k 1[R_k \geq d_H] + (1 - \alpha) \min\left(1, \frac{n_{\text{nod}}(t)}{N_{\text{nod}}}\right) \quad (14)$$

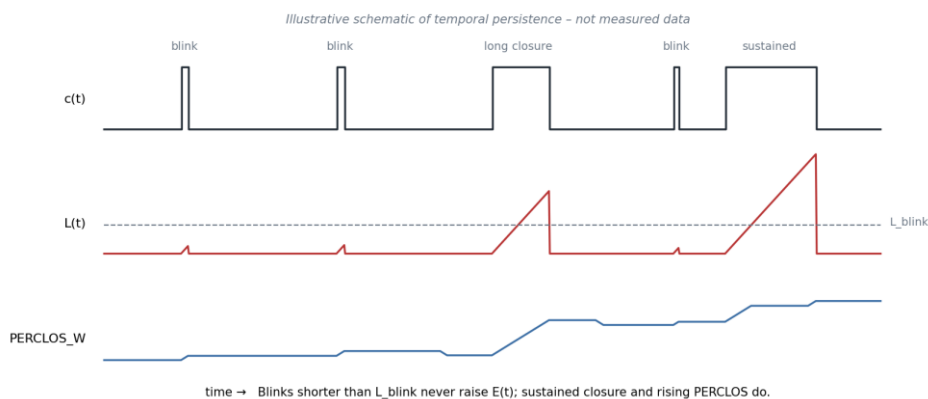


Figure 6. Temporal processing (illustrative schematic, not measured data). Closure indicator $c(t)$, closure run-length $L(t)$ and blink-filtered PERCLOS: blinks shorter than L_{blink} do not contribute; long and sustained closures raise both components of the eye evidence.

3.4 Multi-Cue Fusion and Quality Gating

The three evidence streams are smoothed, clipped to $[0, 1]$ and combined by a convex weighted sum:

$$D_t^{\text{raw}} = w_1 E_t + w_2 Y_t + w_3 H_t, \quad w_i \geq 0, \quad w_1 + w_2 + w_3 = 1 \quad (15)$$

The convex form keeps D on the same scale as its inputs, so thresholds have a stable meaning, and it lets anyone read how much each cue contributed to an alert. In the prototype, the weights were set to $w = (0.55, 0.15, 0.30)$, giving the eye cue the largest share because eyelid behavior is the most direct visual correlate of reduced vigilance (Kolus, 2024), and the yawn cue the smallest because yawns are rare and ambiguous. Landmark quality q_t , computed from mesh visibility and reduced when the head is turned far to the side, gates the update so that the score holds its previous value when measurements are unreliable:

$$D_t = q_t D_t^{\text{raw}} + (1 - q_t) D_{t-1}, \quad q_t \in [0, 1] \quad (16)$$

3.5 Decision and Progressive Alerts

The decision layer maps D_t to a level $\ell_t \in \{0, 1, 2, 3\}$ using entry thresholds $\theta_1 < \theta_2 < \theta_3$, a

persistence time T_{on} for escalation, and a hysteresis margin δ with release time T_{off} for de-escalation:

$$\ell_t \leftarrow k \quad \text{if } \ell_{t-1} < k \quad \text{and} \quad D_\tau \geq \theta_k \quad \forall \tau \in [t - T_{\text{on}}, t] \quad (17)$$

$$\ell_t \leftarrow k - 1 \quad \text{if } \ell_{t-1} = k \quad \text{and} \quad D_\tau < \theta_k - \delta \quad \forall \tau \in [t - T_{\text{off}}, t] \quad (18)$$

Escalation requires the score to stay above a threshold, so one noisy spike cannot cause an alert; de-escalation requires the score to fall clearly below it, so the level does not oscillate near a boundary. In addition, a continuous eye closure of $L_{\text{crit}} = 2.0$ s or longer sets $\ell_t = 3$ immediately (microsleep override). Each level has its own tone and spoken message (Figure 7), written so that the advice escalates without ever implying that continuing to drive while sleepy is acceptable: Level 1, "Warning: signs of drowsiness detected. Please remain alert."; Level 2, "You appear to be drowsy. Please consider stopping at a safe location." Level 3, "Please stop driving and park the vehicle safely before resting." Voice alerts are not new in themselves; the contribution is their coupling to a graded, persistence-controlled score.

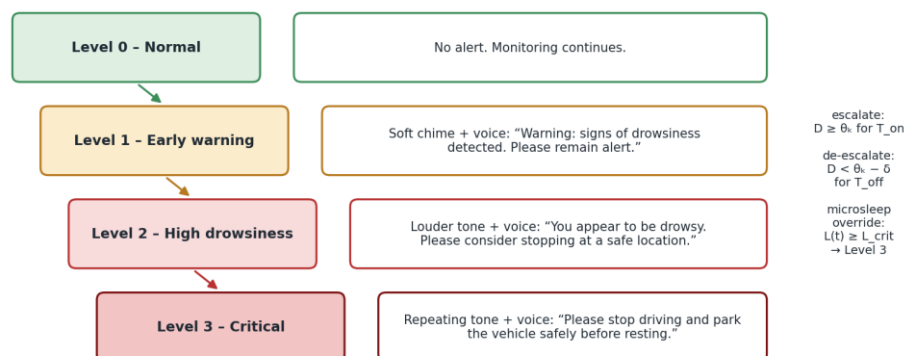


Figure 7. Progressive alert mechanism. Escalation requires persistence, de-escalation requires hysteresis, and a closure of 2.0 s or longer escalates directly to Level 3.

Table 3. Parameters of the implemented prototype.

Parameter	Value	Note
Input resolution / landmark model	640 × 480 / MediaPipe Face Landmarker (478 points)	Package version 1.0.1 reported by the self-test

Parameter	Value	Note
EAR closure threshold τ_E	0.15 (0.17 in one session)	Displayed on the live overlay
MAR opening threshold τ_Y	0.40	Displayed on the live overlay
Fusion weights w_1, w_2, w_3	0.55, 0.15, 0.30	Confirmed from logged scores (Section 5.3)
Microsleep override L_{crit}	2.0 s	Confirmed from logged override onsets
Minimum Level-3 duration	≈ 3 s	Observed in logs (3.1–3.2 s minimum)
Calibration	Neutral posture, eyes open, looking ahead	At start-up
Audio	Tone per level + offline text-to-speech (pyttsx3)	Non-blocking thread

3.6 Implementation and Complexity

The system is written in Python with OpenCV for capture, image processing and the PnP solver, MediaPipe for the face mesh, NumPy for geometry and temporal state, and pyttsx3 for offline speech. A capture thread writes the newest frame into a one-slot buffer so that stale frames are dropped rather than queued; an inference thread runs perception, cue modules, fusion and decision; and an alert thread plays tones and speech without blocking inference.

Every processed frame can be logged to CSV with all measurements, evidence values, the quality term, the score, the level and the microsleep flag. Algorithm 1 gives the per-frame procedure. Per frame, cost is dominated by the landmark model; EAR, MAR and fusion need a fixed number of arithmetic operations, PnP with six points is a small fixed-size problem, and the temporal state is updated incrementally with running sums, so each update is $O(1)$ in time and $O(N_W)$ in memory.

Algorithm 1: Proposed Real-Time Driver Drowsiness Detection FrameworkInput: video stream V ; parameters Π (Table 3)Output: alert level ℓ_t and score D_t for each frame; alert events

```

1 load face-mesh model; set PnP model points and camera matrix  $K$ 
2  $E, Y, H, D \leftarrow 0$ ;  $\ell \leftarrow 0$ ;  $L \leftarrow 0$ ; yawn_list  $\leftarrow \emptyset$ ; ring buffers for  $c, h$ 
3 CALIBRATE: record neutral pose ( $\theta_0, \psi_0, \varphi_0$ ) and baseline EAR/MAR
4 for each frame  $f_t$  with timestamp  $t$ :
5    $f_t \leftarrow$  resize and colour-convert  $f_t$ 
6   (found,  $P, vis$ )  $\leftarrow$  FACE_MESH( $f_t$ )
7   if not found:  $q_t \leftarrow 0$ ; go to 14
8   EARt  $\leftarrow$  eq. 1-3; MARt  $\leftarrow$  eq. 4 (+EMA)
9   ( $\Delta\theta, \Delta\psi, \Delta\varphi$ )  $\leftarrow$  PnP pose, eq. 5-7
10   $q_t \leftarrow$  QUALITY( $vis, \Delta\psi$ )
11   $E_t \leftarrow$  eq. 8-10;  $Y_t \leftarrow$  eq. 11-12;  $H_t \leftarrow$  eq. 13-14
12   $D_{raw} \leftarrow w_1 \cdot E_t + w_2 \cdot Y_t + w_3 \cdot H_t$   $\triangleright$  eq. 15
13   $D_t \leftarrow q_t \cdot D_{raw} + (1 - q_t) \cdot D_{t-1}$   $\triangleright$  eq. 16
14   $\ell_t \leftarrow$  DECIDE( $D_t, \ell_{t-1}$ , history)  $\triangleright$  eq. 17-18
15  if  $L_t \geq L_{crit}$ :  $\ell_t \leftarrow 3$   $\triangleright$  microsleep override
16  if  $\ell_t$  changed or cooldown expired: ALERT( $\ell_t$ )  $\triangleright$  non-blocking
17  LOG( $t$ , measurements,  $E_t, Y_t, H_t, q_t, D_t, \ell_t$ )

```

4. Materials and Methods**4.1 Evaluation Design**

We evaluated the framework at three levels, each answering different research questions. The eye-state component was evaluated quantitatively on a public dataset with a subject-independent split (RQ1, RQ2). The temporal decision logic was verified with an automated simulation test in which the true drowsy period is known by construction (RQ3). The complete pipeline was run live with a webcam and logged frame by frame to verify alert behavior and throughput (RQ3-RQ5).

4.2 MRL Eye Dataset and Subject-Independent Protocol

The MRL Eye dataset (Fusek, 2018; foundational dataset reference) contains 84,898 infrared eye crops (as reported by Hassan et al., 2025) annotated with eye state and image conditions, including eyewear, reflections, and lighting. Because images of one person are highly correlated, a random image-level split would let a model learn individual appearance and would overstate performance on new drivers. The data were therefore partitioned by subject. Six subjects (s0001, s0002, s0012,

s0023, s0031, s0037) formed the validation set, eight different subjects (s0003, s0004, s0005, s0019, s0021, s0025, s0035, s0036) formed the test set, and the remaining subjects were used for training; no test subject contributed any image to training or validation. We trained a multilayer perceptron (MLP) eye-state classifier on the training subjects; we selected its decision threshold (0.48 on the output score) on the validation subjects, then applied it once, unchanged, to the 17,983 test images. Closed eyes are the positive class. Results are reported overall with 95% bootstrap confidence intervals, per subject, and per image condition using the dataset's own annotations.

4.3 Automated Simulation Test

A self-test script checks the implementation before each run. It verifies the EAR and MAR formulas on synthetic landmark sets, the PnP solution on projected model points with known rotation, and the temporal and alert logic on a synthetic 120-s drive. In the synthetic drive, the first 50 s represent an alert driver, with normal blinks and short mouth openings that imitate speech, and the remaining 70 s represent a drowsy driver, with longer eye closures and two

yawns. The script plots every signal together with the resulting alert level, which allows direct inspection of whether isolated events are suppressed and sustained drowsiness is escalated.

4.4 Live Sessions

The complete application was run on a Windows laptop with its built-in webcam. MediaPipe's Python runtime on Windows executes on the CPU, so no GPU acceleration was involved. Two adult participants took part, sitting in front of the camera after the calibration phase and alternating between normal posture, simulated drowsiness (long eye closures, forward head drops, yawns), and sideways head turns. Two sessions were logged to CSV, one row per logged frame, with all measurements, evidence values, the quality term, the raw and gated scores, the alert level and the microsleep flag. We also captured sixteen screenshots of the on-screen overlay. These sessions have no frame-level ground-truth labels, so they are used to verify behavior and throughput, not to estimate detection accuracy.

4.5 Metrics

Classification performance uses the standard confusion-matrix counts, where TP, TN, FP and FN are true positives, true negatives, false positives and false negatives:

$$\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN}, \quad \text{Prec} = \frac{TP}{TP + FP} \quad (19)$$

$$\text{Spec} = \frac{TN}{TN + FP}, \quad \text{F1} = \frac{2 \text{Prec} \cdot \text{Rec}}{\text{Prec} + \text{Rec}}, \quad \text{FPR} \quad (20)$$

$$\text{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

(21)

F1-score and the Matthews correlation coefficient (MCC) are the primary measures because they remain informative when classes are imbalanced; specificity and the false-positive rate describe how often open eyes are mistaken for closed ones, and ROC-AUC describes ranking quality independently of the threshold. Throughput is reported as frames per second,

$$\ell_{\text{frame}} = \sum_{s \in S} \ell_s, \quad \text{FPS} = \frac{N_{\text{frames}}}{t_{\text{end}} - t_{\text{start}}} \quad (22)$$

where the frame latency ℓ_{frame} is the sum of the latencies ℓ_s of the pipeline stages $s \in S$, and N_{frames} frames are processed between t_{start} and t_{end} .

5. Results

5.1 Eye-State Classification on Unseen Subjects (RQ1, RQ2)

Table 4 and Figures 8–9 report the eye-state component on the eight MRL Eye test subjects. The classifier detected closed eyes with high sensitivity (recalled 0.972) and moderate specificity (0.907), for an F1-score of 0.926, MCC of 0.871 and ROC-AUC of 0.987. Its errors were asymmetric: 963 open-eye images (9.3%) were labelled closed, against 215 closed-eye images (2.8%) labelled open. For drowsiness detection, this is the less dangerous direction, because a missed closure delays a warning, whereas a spurious single-frame closure is exactly the kind of isolated event that the blink filter and run-length requirement of Section 3.3 are designed to absorb.

Table 4. Eye-state classification on MRL Eye, 8 unseen test subjects (17,983 images; closed eye = positive class; threshold 0.48 chosen on validation subjects).

Metric	Value	95% CI
TP / TN / FP / FN	7,417 / 9,388 / 963 / 215	—
Accuracy	0.934	0.895–0.976
Precision	0.885	—
Recall (sensitivity)	0.972	—
Specificity	0.907	—

Metric	Value	95% CI
F1-score	0.926	0.849–0.978
MCC	0.871	0.781–0.951
ROC-AUC	0.987	—
FPR / FNR	0.093 / 0.028	—

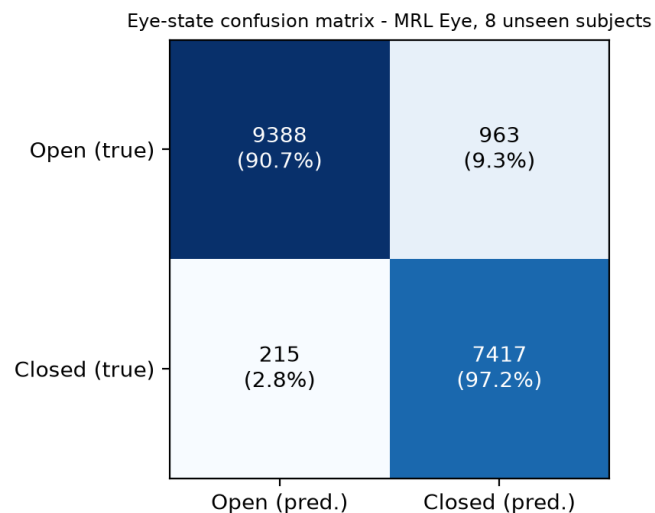


Figure 8. Confusion matrix of the eye-state classifier on the 8 unseen MRL Eye test subjects (row percentages in brackets).

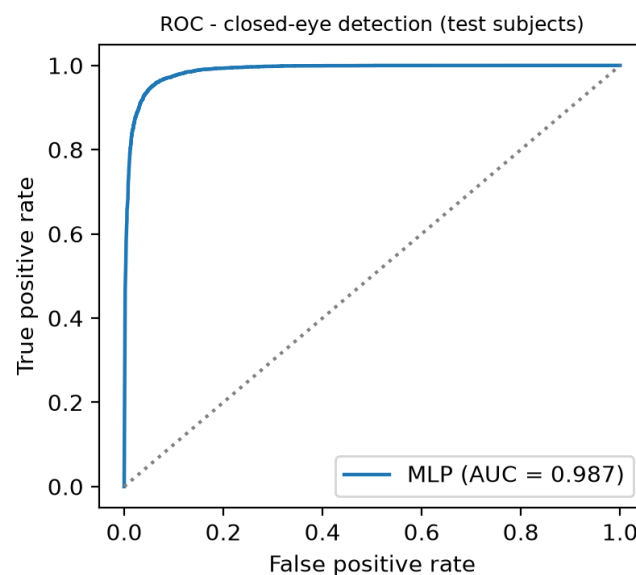


Figure 9. ROC curve of closed-eye detection on the 8 unseen MRL Eye test subjects (AUC = 0.987).

Per-subject results (Table 5) varied considerably, with accuracy from 0.837 to 0.986 and F1 from

0.573 to 0.984. The two weakest subjects, s0021 and s0025, kept accuracy above 0.83 but had F1-scores near 0.6, indicating many false closed-eye predictions relative to the number of genuinely closed eyes. A pooled figure hides this spread;

this is why we use subject-independent evaluation and include a calibration phase in the framework.

Table 5. Per-subject results on the MRL Eye test subjects.

Subject	Images	Accuracy	Recall	Specificity	F1
s0003	679	0.951	0.968	0.875	0.970
s0004	1,069	0.953	0.953	0.000†	0.976
s0005	736	0.969	0.973	0.857	0.984
s0019	6,175	0.983	0.997	0.970	0.982
s0021	987	0.837	0.896	0.828	0.601
s0025	1,502	0.883	0.861	0.885	0.573
s0035	642	0.986	0.857	0.990	0.800
s0036	6,193	0.900	0.960	0.869	0.866

† For s0004, accuracy equals recall, indicating that almost all this subject's test images are closed-eye; its specificity rests on very few open-eye images and should not be interpreted.

Image conditions mattered (Table 6). With glasses, F1 fell from 0.938 to 0.875, and with large reflections on the eye region it fell to 0.758 (accuracy 0.830); the lighting annotation made

little difference (F1 0.928 for poor and 0.919 for good lighting). The small-reflection subgroup contains only 305 images, so its estimate is imprecise. These results support the design choice to weight eye evidence by landmark quality and to let the head and yawn cues carry more of the decision when the eye cue is degraded.

Table 6. Eye-state results by image condition (MRL Eye annotations).

Condition	Value	Images	Accuracy	F1
Glasses	No	14,220	0.944	0.938
Glasses	Yes	3,763	0.900	0.875
Reflections	None	16,211	0.945	0.940
Reflections	Small	305	0.902	0.722
Reflections	Big	1,467	0.830	0.758
Lighting	Bad	15,220	0.935	0.928
Lighting	Good	2,763	0.931	0.919

5.2 Temporal Logic on a Simulated Drive (RQ3)

All self-test checks passed: EAR and MAR formulas, PnP head pose, temporal logic and alert levels, MediaPipe installation and model file. Figure 10 shows the simulated drive. No alert was raised during the 50-s alert phase, so the blinks and speech-like mouth openings did not accumulate evidence. The two yawns in the drowsy phase raised Y and the fused score but,

on their own, did not cross the first alert threshold. As longer closures accumulated, the system escalated to Level 1 at approximately 100 s, Level 2 at approximately 103 s and Level 3 at approximately 116 s (times read from the plot). The first alert therefore came about 50 s after the simulated onset of drowsiness. This conservative response prevents false alarms, but such a delay would be too long if it reflected real drowsiness

onset, so detection latency is the main quantity to tune.

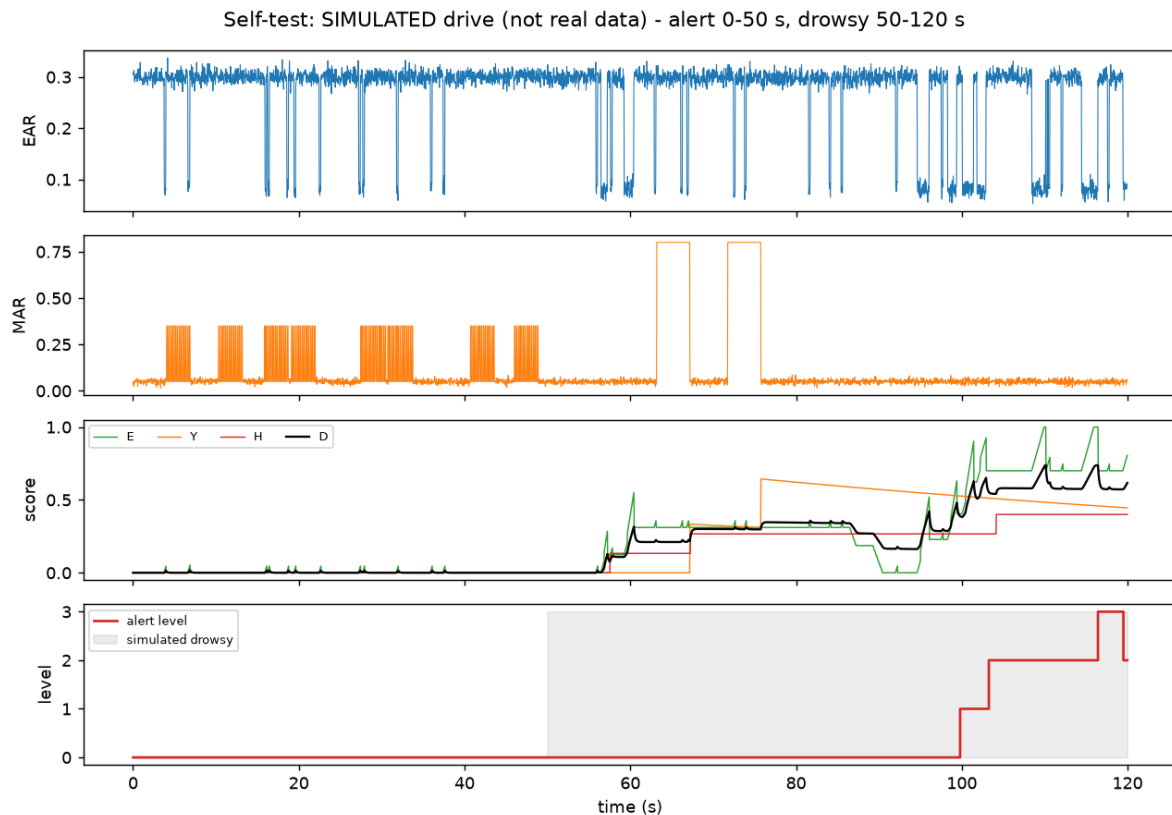


Figure 10. Automated self-test on a simulated 120-s drive (synthetic signals, not real data). From top: EAR, MAR, cue evidence E, Y, H with fused score D, and alert level against the simulated drowsy period (grey).

5.3 Live Sessions and Alert Behaviour (RQ3, RQ4)

Figure 11 shows representative frames from the live application, including the calibration phase,

all four alert levels and two microsleep events, and Table 7 lists values read from the overlay. Figure 12 plots the first logged session in full and Table 8 summarises both logged sessions.

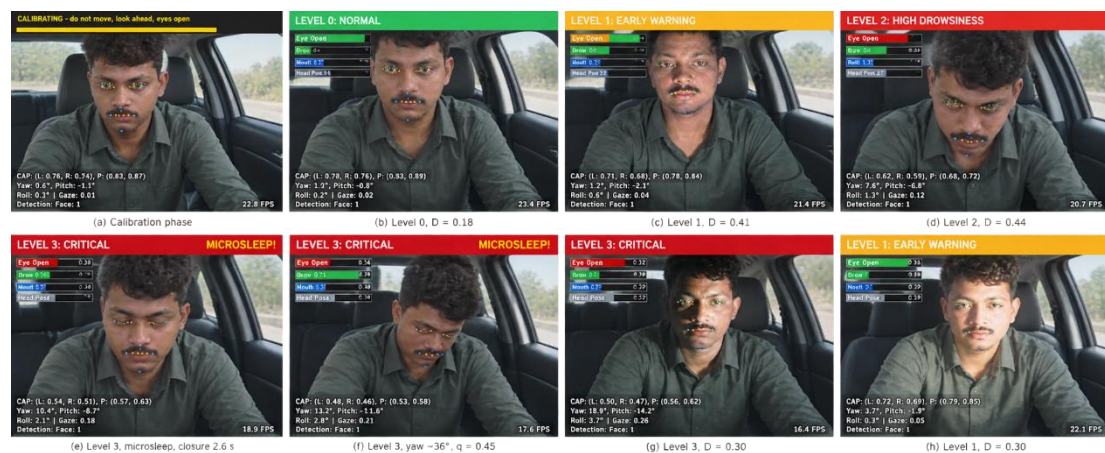


Figure 11. Live prototype (screenshots, participants' faces shown with consent). (a) Calibration; (b)–(d) Levels 0–2; (e) Level 3 from the microsleep override after a 2.6-s closure; (f)

Level 3 with the head at yaw -36° , where quality drops to about 0.45; (g)–(h) the same score $D = 0.30$ shown with Level 3 and Level 1, illustrating that the level depends on its history.

Table 7. Selected values displayed by the live prototype (read from screenshots; MS = microsleep override active).

Level	D	E	Y	H	EAR (τ_E)	PERCLOS	Closure (s)	Pitch / yaw / roll ($^\circ$)
0	0.18	0.09	0.07	0.40	0.32 (0.15)	0.02	0.0	+11 / +7 / -9
1	0.41	0.48	0.09	0.40	0.30 (0.15)	0.10	0.0	+2 / +13 / -9
1	0.30	0.33	0.02	0.40	0.22 (0.15)	0.07	0.0	+7 / +7 / -7
2	0.44	0.52	0.06	0.50	0.36 (0.15)	0.11	0.0	+17 / +6 / -5
2	0.55	0.70	0.04	0.53	0.33 (0.15)	0.21	0.0	+10 / +13 / -16
3	0.30	0.33	0.02	0.40	0.24 (0.15)	0.07	0.0	+7 / +8 / -7
3 (MS)	0.68	1.00	0.05	0.40	0.05 (0.15)	0.24	2.6	+11 / +6 / -8
3 (MS)	0.58	1.00	0.22	0.00	0.12 (0.17)	0.37	5.6	+6 / -36 / +5

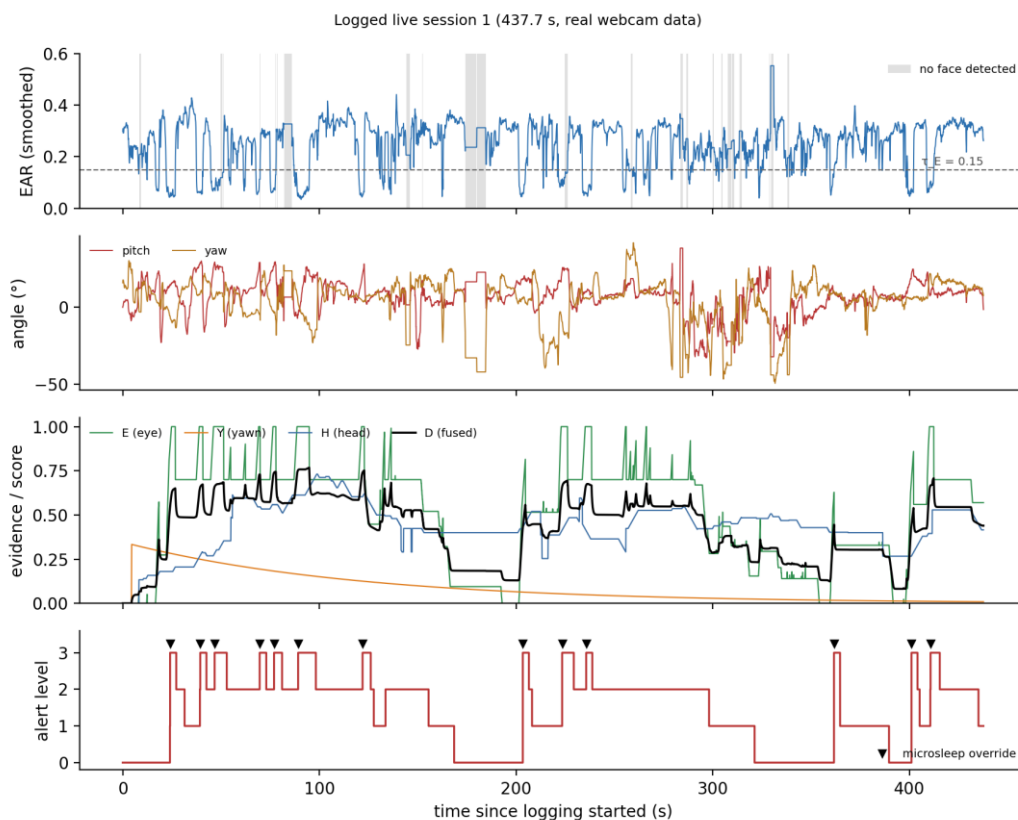


Figure 12. Logged live session 1 (real webcam data, 437.7 s). From top: smoothed EAR with the closure threshold (grey: no face detected); pitch and yaw; evidence E, Y, H and fused score D; alert level with microsleep-override events (triangles).

Table 8. Summary of the two logged live sessions (computed from the CSV logs; descriptive, no ground-truth labels).

Quantity	Session 1	Session 2
Logged duration (s) / logged frames	437.7 / 5,590	161.3 / 1,347
Frames with a face detected (%)	92.8	97.9
Mean quality q / frames with $q < 0.5$ (%)	0.878 / 9.2	0.919 / 4.7
Time at Level 0 / 1 / 2 / 3 (%)	25.3 / 22.4 / 39.5 / 12.8	6.1 / 67.3 / 6.7 / 20.0
Eye closures: total / < 0.5 s / ≥ 1 s / ≥ 2 s	58 / 28 / 21 / 13	28 / 13 / 14 / 9
Longest continuous closure (s)	7.7	5.7
Microsleep overrides (closure at trigger, s)	13 (2.00–2.07)	9 (2.00–2.13)
Level-3 episodes, median duration (range) (s)	13, 3.4 (3.1–9.0)	9, 3.2 (3.2–6.4)
Maximum fused score D	0.768	0.592
Median $ D_{\text{raw}} - (0.55E + 0.15Y + 0.30H) $	0.0008	0.0013

The logs support five observations. First, the fused score follows Equation 15 with $w = (0.55, 0.15, 0.30)$: the median absolute difference between the logged raw score and this weighted sum is 0.0008–0.0013, with larger differences only during rapid changes, where evidence smoothing lags. Second, the blink filter and microsleep override behaved as designed: none of the 41 closures shorter than 0.5 s triggered the override, while all 22 closures of 2.0 s or longer did, at closure lengths of 2.00–2.13 s. Third, every Level-3 episode was entered through the override, and the fused score never exceeded 0.77, so in these sessions the critical level was reached by long closures and not by accumulated evidence alone. Fourth, levels depended on history, as intended: identical evidence ($D = 0.30$) appeared with Level 3 and with Level 1, and Level 3 lasted at least about 3 s after each override before de-escalating. Fifth, the quality term fell to about 0.45 when the head turned to yaw -36° , and head evidence stayed near zero in session 2 (mean $H = 0.008$), where the head was mostly turned sideways (median yaw -12°) rather than dropped forward, which is consistent with the yaw-gating rule of Equation 13.

The eye evidence also decays more slowly than the eyes reopen. In the Level 2 frame with EAR 0.33, E remained 0.70 because PERCLOS (0.21) still reflected the preceding closures. This is intended, because drowsiness does not vanish

when the eyes open, so warnings persist for a while after recovery.

5.4 Real-Time Performance (RQ5)

The frame rate displayed by the running application ranged from 54.7 to 86.2 FPS across the fifteen screenshots in which it was legible (median 77.1 FPS). These readings cover the complete pipeline, including landmark extraction, the three cue modules, fusion, decision and overlay rendering, on a laptop CPU. They are well above 15 FPS, the rate at which the temporal resolution of closure measurement (about 67 ms per frame) remains adequate for blink-length events. The readings are instantaneous values rather than a controlled benchmark; repeated-replay measurements of per-stage latency, CPU utilization, and memory on several devices remain part of future work.

6. Discussion

Eye-state component (RQ1, RQ2). On subjects it had never seen, the eye-state classifier reached an F1-score of 0.926 and ROC-AUC of 0.987. These values are lower than the 99% accuracies often reported for MRL Eye (e.g., Hassan et al., 2025), but the protocols differ: an image-level 80/20 split lets images of the same person appear on both sides, whereas the subject-wise split used here measures generalization to new people, which is what a deployed system faces. The per-subject spread (F1 0.573–0.984) and the

losses with glasses and reflections are the more informative findings. They show that a single fixed model can serve most drivers well and a few poorly, which supports per-driver calibration of the kind proposed by Lamba et al. (2025) and Ersoy et al. (2026), and they confirm that eyewear and specular reflections remain the main visual obstacle, consistent with the low-light and imaging-robustness focus of recent work (Liu et al., 2025; El-Nabi et al., 2025; Saxena et al., 2026).

Temporal decision layer (RQ3). The simulation and the live logs agree on the central design claim: short events did not generate alerts, and sustained closure did. None of the 41 blink-length closures in the live sessions triggered the override, and no alert occurred during the simulated alert phase despite blinks and speech-like mouth movement. The same evidence also exposes the main weakness of the current settings. In the simulation, the first alert came about 50 s after the drowsy phase began, and in the live sessions the fused score never reached the critical threshold on its own; all critical alerts came from the 2.0-s override. The score-based path is therefore conservative, and most protection against sudden sleep currently comes from the override. This is a safe failure mode with respect to false alarms, but it may delay warnings for gradual drowsiness and lowering the entry thresholds or shortening the PERCLOS window are the first adjustments to test.

Alert policy (RQ4). The logs show that the four-level policy behaves as specified: levels depend on history rather than on the instantaneous score, Level 3 is held for a consistent minimum

of about 3 s, and quality gating reacts to large head turns. Because eye evidence decays slowly, warnings persist for some time after the eyes reopen. Whether this is desirable is a human-factors question; a driver who has just had a microsleep arguably should keep hearing advice to stop, but persistent warnings after genuine recovery may reduce acceptance.

Computational cost (RQ5). The complete pipeline ran at a median of 77.1 FPS on a laptop CPU. Almost all computations are in one pre-trained landmark model, and the cue, temporal and fusion stages add only a fixed number of arithmetic operations per frame, so the feasibility of the framework on a given device is essentially the feasibility of the landmark model on that device. Transformer classifiers applied to every frame (Jarndal et al., 2025; Hassan et al., 2025) may offer higher per-frame accuracy but at a much higher cost on CPU-only hardware; the present design instead places complexity in the temporal model.

7. Limitations

Table 9 lists the main limitations. The most important is that the full drowsiness decision has not yet been evaluated against labelled drowsiness data. The eye-state results are quantitative and subject-independent, but they measure one component; the simulation measures the logic on synthetic signals; and the live sessions verify behavior without ground-truth labels. The cue-level ablation, the comparison with baselines, and detection latency on real drowsiness therefore remain to be established, as described in Section 9.

Table 9. Limitations and their expected effect.

Limitation	Expected effect	Planned response
No labelled drowsiness evaluation of the full framework	Accuracy, latency and false alarms of the fused decision are unknown	Subject-independent evaluation on NTHU-DDD and YawDD
Eye-state results from one dataset and one classifier	Generalization to other cameras uncertain	Cross-dataset tests; learned vs. geometric eye cue
Conservative score thresholds	Late warnings for gradual drowsiness	Validation-based threshold and weight selection

Limitation	Expected effect	Planned response
Hand-set fusion weights	Possibly sub-optimal cue balance	Weight search with sensitivity analysis
Two participants, acted behavior, indoor	Limited diversity; exaggerated behavior	Larger, diverse, real-vehicle study
Eyewear and reflections	Degraded eye evidence (F1 down to 0.758)	Near-infrared camera; rely more on head and yawn cues
Visual cues only	Drowsiness with eyes open is missed	Add a non-visual channel
Instantaneous FPS readings	No per-stage latency or resource profile	Repeated-replay benchmark on several devices

8. Safety and Ethical Considerations

The framework is a driver-assistance tool. It can warn a driver and advise stopping; it cannot establish that a driver is fit to drive, and the absence of an alert must never be read as evidence of alertness. For this reason, the messages never suggest that continuing is acceptable; Level 2 already recommends stopping at a safe location, and Level 3 instructs the driver to park and rest. Facial video is biometric data. In deployment mode, frames are processed in memory and discarded; only derived signals are logged, logging can be disabled, and no identity recognition is performed. The live recordings shown in this paper were made with the participants' informed consent. Integrating this into a production vehicle would require validation on diverse real drivers, assessment under applicable functional-safety and driver-monitoring requirements, fairness analysis across demographic groups (Nurnoby & El-Alfy, 2026), and fail-safe behavior when the camera is blocked or no face is found.

9. Conclusion and Future Work

This paper presented a lightweight, interpretable framework for real-time driver drowsiness detection that combines eye closure, yawning and head pose through explicit temporal evidence functions, a convex fusion rule, landmark-quality gating and a four-level alert policy with a microsleep override. The eye-state component reached an F1-score of 0.926, MCC of 0.871 and ROC-AUC of 0.987 on eight unseen MRL Eye subjects, with the largest degradation for eyewear and strong reflections.

A simulation test and two logged live sessions showed that blinks and speech-like mouth movements did not trigger alerts, that every eye closure of 2.0 s or longer raised a critical warning, that alert levels followed the intended persistence and hysteresis, and that the complete pipeline ran at a median of 77.1 FPS on a laptop CPU. They also showed that, with the current settings, the score-based path responds conservatively to gradual drowsiness.

Future work follows directly. The first priority is a subject-independent evaluation of the complete decision on labelled video from NTHU-DDD (Weng et al., 2017) and YawDD (Abtahi et al., 2014; Mujtaba et al., 2025), including a cue-level ablation (single cues, pairs, all three with and without temporal processing), baselines ranging from EAR and MAR rules to a MobileNetV3 eye classifier (Howard et al., 2019) and a transformer face classifier, and event-level measures such as detection latency and false alerts per hour. The second is validation-based selection of fusion weights and alert thresholds, comparing fixed and personalised thresholds. The third is a real-vehicle study with a larger, more diverse group of drivers, followed by the addition of a non-visual channel to detect drowsiness not visible on the face.

Declarations

Data availability: MRL Eye is publicly available from its providers. Code, configuration files and the anonymized session logs are available from the corresponding author on reasonable request. Live video is not shared to protect participant privacy.

Ethics: The live sessions involved adult volunteers who gave informed consent, including consent to publish the images in Figure 11.

Funding: This research received no external funding. Conflicts of interest: The authors declare no conflict of interest.

REFERENCES

References marked [Foundational] predate 2024 and are cited only for a fundamental measure, algorithm or dataset definition. References marked [Preprint] have not completed peer review at the time of writing.

- Abtahi, S., Omidyeganeh, M., Shirmohammadi, S., & Hariri, B. (2014). YawDD: A yawning detection dataset. In Proceedings of the 5th ACM Multimedia Systems Conference (pp. 24–28). ACM. [Foundational]
- Al-Gburi, S. H., Al-Sammak, K. A., Marghescu, I., Oprea, C. C., Drăgulescu, A.-M. C., Alduais, N. A. M., Alheeti, K. M. A., & Al-Sammak, N. A. H. (2025a). EffRes-DrowsyNet: A novel hybrid deep learning model combining EfficientNetB0 and ResNet50 for driver drowsiness detection. *Sensors*, 25(12), 3711. <https://doi.org/10.3390/s25123711>
- Al-Gburi, S. H., Al-Sammak, K. A., Marghescu, I., Oprea, C. C., Drăgulescu, A.-M. C., Suci, G., Alheeti, K. M. A., Alduais, N. A. M., & Al-Sammak, N. A. H. (2025b). Introducing a novel fast neighbourhood component analysis–deep neural network model for enhanced driver drowsiness detection. *Big Data and Cognitive Computing*, 9(5), 126. <https://doi.org/10.3390/bdcc9050126>
- Alguindigue, J., Singh, A., Narayan, A., & Samuel, S. (2024). Biosignals monitoring for driver drowsiness detection using deep neural networks. *IEEE Access*, 12, 93075–93086. <https://doi.org/10.1109/ACCESS.2024.3423723>
- Alvarez Oviedo, A., Mamani Villanueva, J. F., Echaiz Espinoza, G. A., Villanueva, J. M. M., Salazar, A. O., & Villarreal, E. R. L. (2025). Design of a system for driver drowsiness detection and seat belt monitoring using Raspberry Pi 4 and Arduino Nano. *Designs*, 9(1), 11. <https://doi.org/10.3390/designs9010011>
- Ayas, S., He, D., & Donmez, B. (2026). Driver state classification: Identifying high cognitive load and drowsiness through driver performance and physiology. *IEEE Transactions on Intelligent Transportation Systems*, 27(1), 1105–1115. <https://doi.org/10.1109/TITS.2025.3633499>
- Balam, V. P. (2024). Systematic review of single-channel EEG-based drowsiness detection methods. *IEEE Transactions on Intelligent Transportation Systems*, 25(11), 15210–15228. <https://doi.org/10.1109/TITS.2024.3442249>
- Belakhddhar, I. (2026). Connected embedded system for drowsiness detection based on a reconfigurable set of features. *Sensors*, 26(4), 1195. <https://doi.org/10.3390/s26041195>
- Beles, H., Vesselenyi, T., Rus, A., Mitran, T., Scurt, F. B., & Tolea, B. A. (2024). Driver drowsiness multi-method detection for vehicles with autonomous driving functions. *Sensors*, 24(5), 1541. <https://doi.org/10.3390/s24051541>
- Bhanja, A., Parhi, D., Gajendra, D., Sinha, K., & Sahoo, A. K. (2025). Driver drowsiness shield (DDSH): A real-time driver drowsiness detection system. *ROBOMECH Journal*, 12(1), 18. <https://doi.org/10.1186/s40648-025-00307-4>
- Ciesla, M., & Ostermayer, G. (2024). A multimodal recurrent model for driver distraction detection. *Applied Sciences*, 14(19), 8935. <https://doi.org/10.3390/app14198935>

- Das, S., Pratihari, S., Pradhan, B., Jhaveri, R. H., & Benedetto, F. (2024). IoT-assisted automatic driver drowsiness detection through facial movement analysis using deep learning and a U-Net-based architecture. *Information*, 15(1), 30. <https://doi.org/10.3390/info15010030>
- Delwar, T. S., Singh, M., Mukhopadhyay, S., Kumar, A., Parashar, D., Lee, Y., Rahman, M. H., Sejan, M. A. S., & Ryu, J. Y. (2025). AI- and deep learning-powered driver drowsiness detection method using facial analysis. *Applied Sciences*, 15(3), 1102. <https://doi.org/10.3390/app15031102>
- Díaz-Santos, S., Cigala-Álvarez, Ó., González-Sosa, E., Caballero-Gil, P., & Caballero-Gil, C. (2024). Driver identification and drowsiness detection while driving. *Applied Sciences*, 14(6), 2603. <https://doi.org/10.3390/app14062603>
- Dinges, D. F., & Grace, R. (1998). PERCLOS: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance (Tech Brief FHWA-MCRT-98-006). Federal Highway Administration. [Foundational]
- Divvala, C., & Mishra, M. (2024). Deep learning-based attention mechanism for automatic drowsiness detection using EEG signal. *IEEE Sensors Letters*, 8(3), 1–4. <https://doi.org/10.1109/LSENS.2024.3363735>
- El-Nabi, S. A., Ibrahim, A. F., El-Rabaie, E.-S. M., Hassan, O. F., Soliman, N. F., Ramadan, K. F., & El-Shafai, W. (2025). Driver drowsiness detection using Swin transformers and diffusion models for robust image denoising. *IEEE Access*, 13, 71880–71907. <https://doi.org/10.1109/ACCESS.2025.3561717>
- El Sahmarany, L., Alkhaldi, M., & Alzahrani, S. I. (2026). Sensing technologies and physiological parameters for real-time driver drowsiness detection: A comprehensive review. *Sensors*, 26(11), 3333. <https://doi.org/10.3390/s26113333>
- Ersoy, G., Tatar, M. A., Tonbul, E., & Kırbız, S. (2026). Improving driver drowsiness detection via personalized EAR/MAR thresholds and CNN-based classification [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2604.22479>
- Essahraoui, S., Lamaakal, I., El Hamly, I., Maleh, Y., Ouahbi, I., El Makkaoui, K., Filali Bouami, M., Pławiak, P., Alfarraj, O., & Abd El-Latif, A. A. (2025). Real-time driver drowsiness detection using facial analysis and machine learning techniques. *Sensors*, 25(3), 812. <https://doi.org/10.3390/s25030812>
- Essel, E., Lacy, F., Albalooshi, F., Elmedany, W., & Ismail, Y. (2024). Drowsiness detection in drivers using facial feature analysis. *Applied Sciences*, 15(1), 20. <https://doi.org/10.3390/app15010020>
- Feng, W., Wang, X., Xie, J., Liu, W., Qiao, Y., & Liu, G. (2025). Real-time EEG-based driver drowsiness detection based on convolutional neural network with Gumbel-softmax trick. *IEEE Sensors Journal*, 25(1), 1860–1871. <https://doi.org/10.1109/JSEN.2024.3492176>
- Florez, R., Palomino-Quispe, F., Alvarez, A. B., Coaquira-Castillo, R. J., & Herrera-Levano, J. C. (2024). A real-time embedded system for driver drowsiness detection based on visual analysis of the eyes and mouth using convolutional neural network and mouth aspect ratio. *Sensors*, 24(19), 6261. <https://doi.org/10.3390/s24196261>
- Fonseca, T., & Ferreira, S. (2025). Drowsiness detection in drivers: A systematic review of deep learning-based models. *Applied Sciences*, 15(16), 9018. <https://doi.org/10.3390/app15169018>
- Fujiwara, K., Iwamoto, H., Hori, K., & Kano, M. (2024). Driver drowsiness detection using R-R interval of electrocardiogram and self-attention autoencoder. *IEEE Transactions on Intelligent Vehicles*, 9(1), 2956–2965. <https://doi.org/10.1109/TIV.2023.3308575>

- Fusek, R. (2018). Pupil localization using geodesic distance. In *Advances in Visual Computing: ISVC 2018 (Lecture Notes in Computer Science, Vol. 11241, pp. 433–444)*. Springer. [Foundational; MRL Eye dataset]
- Harb, H. (2025). An efficient drowsiness detection framework for improving driver safety through supervised learning models. *World Electric Vehicle Journal*, 16(11), 620. <https://doi.org/10.3390/wevj16110620>
- Hassan, O. F., Ibrahim, A. F., Gomaa, A., Makhoul, M. A., & Hafiz, B. (2025). Real-time driver drowsiness detection using transformer architectures: A novel deep learning approach. *Scientific Reports*, 15, 17493. <https://doi.org/10.1038/s41598-025-02111-x>
- Hassan, O. F., Ibrahim, A. F., Gomaa, A., Alshathri, S., El-Shafai, W., Makhoul, M. A., & Hafiz, B. (2026). Comprehensive benchmarking of attention enhanced CNNs and transfer learning models for real time driver drowsiness detection. *IEEE Access*, 14, 17667–17701. <https://doi.org/10.1109/ACCESS.2026.3658594>
- Hollósi, J., Ballagi, Á., Kovács, G., Fischer, S., & Nagy, V. (2024). Bus driver head position detection using capsule networks under dynamic driving conditions. *Computers*, 13(3), 66. <https://doi.org/10.3390/computers13030066>
- Hollósi, J., Kovács, G., Sysyn, M., Kurhan, D., Fischer, S., & Nagy, V. (2025). Driver distraction detection in extreme conditions using Kolmogorov–Arnold networks. *Computers*, 14(5), 184. <https://doi.org/10.3390/computers14050184>
- Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q. V., & Adam, H. (2019). Searching for MobileNetV3. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 1314–1324)*. [Foundational]
- Huang, Z., Liu, S., Huang, J., & Huang, J. (2026). Nighttime driver fatigue detection based on real-time joint face and facial landmarks detection. *Modelling*, 7(2), 60. <https://doi.org/10.3390/modelling7020060>
- Huang, Z., Tang, W., Tian, Q., Huang, T., & Li, J. (2024). Air traffic controller fatigue detection based on facial and vocal features using long short-term memory. *IEEE Access*, 12, 56663–56682. <https://doi.org/10.1109/ACCESS.2024.3391427>
- Hussein, Y. M., Hassan, R. F., & Chisab, R. F. (2026). Guided dual-attention networks for compact driver drowsiness detection: Performance, calibration, and edge deployment. *Technologies*, 14(9), 544. <https://doi.org/10.3390/technologies14090544>
- Jarndal, A., Tawfik, H., Siam, A. I., Alsyoud, I., & Cheaitou, A. (2025). A real-time vision transformers-based system for enhanced driver drowsiness detection and vehicle safety. *IEEE Access*, 13, 1790–1803. <https://doi.org/10.1109/ACCESS.2024.3522111>
- Jebraeily, Y., Sharafi, Y., & Teshnehlal, M. (2024). Driver drowsiness detection based on convolutional neural network architecture optimization using genetic algorithm. *IEEE Access*, 12, 45709–45726. <https://doi.org/10.1109/ACCESS.2024.3381999>

- Jiao, Y., Zhang, C., Chen, X., Fu, L., Jiang, C., & Wen, C. (2024). Driver fatigue detection using measures of heart rate variability and electrodermal activity. *IEEE Transactions on Intelligent Transportation Systems*, 25(6), 5510–5524. <https://doi.org/10.1109/TITS.2023.3333252>
- Kartynnik, Y., Ablavatski, A., Grishchenko, I., & Grundmann, M. (2019). Real-time facial surface geometry from monocular video on mobile GPUs [Preprint]. *arXiv*. <https://doi.org/10.48550/arXiv.1907.06724> [Foundational]
- Kazemi, M., Rezaei, M., & Azarmi, M. (2025). Evaluating driver readiness in conditionally automated vehicles from eye-tracking data and head pose. *IET Intelligent Transport Systems*, 19(1), e70006. <https://doi.org/10.1049/itr2.70006>
- Kolus, A. (2024). A systematic review on driver drowsiness detection using eye activity measures. *IEEE Access*, 12, 97969–97993. <https://doi.org/10.1109/ACCESS.2024.3424654>
- Lamba, P. S., Jain, R., Khan, S., Alanazi, S. M., Jain, A., Zamani, A. T., Panwar, A., & Abdallah, A. M. A. (2025). Detection of driver drowsiness using adaptive eye characteristic ratio for enhanced road safety. *IEEE Access*, 13, 84313–84328. <https://doi.org/10.1109/ACCESS.2025.3566981>
- Lee, M., Cha, M., & Woo, J. (2025). In-depth inception-attention time model: An application for driver drowsiness detection. *Electronics*, 14(6), 1069. <https://doi.org/10.3390/electronics14061069>
- Lepetit, V., Moreno-Noguer, F., & Fua, P. (2009). EPnP: An accurate O(n) solution to the PnP problem. *International Journal of Computer Vision*, 81(2), 155–166. <https://doi.org/10.1007/s11263-008-0152-6> [Foundational]
- Liu, G., Wu, K., Lan, W., & Wu, Y. (2025). YOLO-FDCL: Improved YOLOv8 for driver fatigue detection in complex lighting conditions. *Sensors*, 25(15), 4832. <https://doi.org/10.3390/s25154832>
- Lopez, M.-J., Palacios-Arias, C., Ordonez, B., & Quingatuna, N. (2026). Real-time radar sensor performance enhancement via edge machine learning for driver drowsiness detection. *IEEE Sensors Letters*, 10(9), 6008504. <https://doi.org/10.1109/LSSENS.2026.3715426>
- Lu, D., Huang, H., Yi, Z., & Pan, J. (2025). Confidence-aware multimodal network for driver drowsiness detection: Leveraging EEG-EOG signals enhanced by knowledge distillation. *IEEE Transactions on Instrumentation and Measurement*, 74, 1–13. <https://doi.org/10.1109/TIM.2025.3591857>
- Madni, H. A., Raza, A., Sehar, R., Thalji, N., & Abualigah, L. (2024). Novel transfer learning approach for driver drowsiness detection using eye movement behavior. *IEEE Access*, 12, 64765–64778. <https://doi.org/10.1109/ACCESS.2024.3392640>
- Mai, N.-D., Nguyen, H.-T., & Chung, W.-Y. (2025). Multimodal driver drowsiness detection using facial expressions and ear-EEGs with a lightweight auto-denoising network. *IEEE Transactions on Intelligent Transportation Systems*, 26(6), 7819–7832. <https://doi.org/10.1109/TITS.2025.3559098>
- Makhmudov, F., Turimov, D., Xamidov, M., Nazarov, F., & Cho, Y.-I. (2024). Real-time fatigue detection algorithms using machine learning for yawning and eye state. *Sensors*, 24(23), 7810. <https://doi.org/10.3390/s24237810>

- Mujtaba, A., Radchenko, G., Masana, M., & Prodan, R. (2025). YawDD+: Frame-level annotations for accurate yawn prediction [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2512.11446>
- Nomura, A., Nagumo, K., Oiwa, K., & Nozawa, A. (2026). Improving generalization of drowsiness classification from facial skin temperature distribution using data preprocessing in CNN. *IEEE Sensors Journal*, 26(5), 7132–7141. <https://doi.org/10.1109/JSEN.2025.3650001>
- Nurnoby, M. F., & El-Alfy, E.-S. M. (2026). Gender-aware driver drowsiness detection using multi-stream shifted-window-based hierarchical vision transformers. *Applied Sciences*, 16(7), 3353. <https://doi.org/10.3390/app16073353>
- Obaidan, H. B., Hussain, M., & AlMajed, R. (2024). EEG_DMNet: A deep multi-scale convolutional neural network for electroencephalography-based driver drowsiness detection. *Electronics*, 13(11), 2084. <https://doi.org/10.3390/electronics13112084>
- Owen, V., & Surantha, N. (2025). Computer vision-based drowsiness detection using handcrafted feature extraction for edge computing devices. *Applied Sciences*, 15(2), 638. <https://doi.org/10.3390/app15020638>
- Quiles-Cucarella, E., Cano-Bernet, J., Santos-Fernández, L., Roldán-Blay, C., & Roldán-Porta, C. (2024). Multi-index driver drowsiness detection method based on driver's facial recognition using Haar features and histograms of oriented gradients. *Sensors*, 24(17), 5683. <https://doi.org/10.3390/s24175683>
- Rahman, S., Farhan, S., Sharma, R. R., & Faisal, K. N. (2026). HatSeqNet: A sequence-based deep learning model for cross-subject driver drowsiness detection using single-channel EEG signal. *IEEE Sensors Journal*, 26(9), 13703–13710. <https://doi.org/10.1109/JSEN.2026.3677604>
- Rahmani, C., Benlamoudi, A., Bounab, Y., Bekhouche, S. E., Samai, D., Dornaika, F., Taleb, A., & Belhaouari, S. B. (2026). A semi-supervised neural framework for real-time drowsiness detection using facial cues. *IEEE Access*, 14, 12816–12836. <https://doi.org/10.1109/ACCESS.2026.3656468>
- Ramzan, M., Abid, A., Fayyaz, M., Alahmadi, T. J., Nobanee, H., & Rehman, A. (2024). A novel hybrid approach for driver drowsiness detection using a custom deep learning model. *IEEE Access*, 12, 126866–126884. <https://doi.org/10.1109/ACCESS.2024.3438617>
- Salem, D., & Waleed, M. (2024). Drowsiness detection in real-time via convolutional neural networks and transfer learning. *Journal of Engineering and Applied Science*, 71(1), 122. <https://doi.org/10.1186/s44147-024-00457-z>
- Saxena, S., Angel, Khurana, M., Tiwari, S., & Shakya, H. K. (2026). Low-light driver drowsiness detection for real-time safety assistance using dual attention mechanisms in deep learning model. *Scientific Reports*, 16, 18234. <https://doi.org/10.1038/s41598-026-44442-3>
- Siddiqui, H. U. R., Akmal, A., Iqbal, M., Saleem, A. A., Raza, M. A., Zafar, K., Zaib, A., Dudley, S., Arambarri, J., Castilla, Á. K., & Rustam, F. (2024). Ultra-wide band radar empowered driver drowsiness detection with convolutional spatial feature engineering and artificial intelligence. *Sensors*, 24(12), 3754. <https://doi.org/10.3390/s24123754>
- Soukupová, T., & Čech, J. (2016). Real-time eye blink detection using facial landmarks. In *Proceedings of the 21st Computer Vision Winter Workshop*. [Foundational]

- Vaussean, F., Bhattacharya, A., Payette, J., Saidi, A., Bellemin, V., Renaud-Dumoulin, G.-G., Cloutier, S. G., & Gagnon, G. (2026). Early drowsiness detection via second-order derivative analysis of heart rate variability: A non-contact ECG approach with machine learning. *Sensors*, 26(4), 1348. <https://doi.org/10.3390/s26041348>
- Venkateswarlu, M., & Ch, V. R. R. (2024). DrowsyDetectNet: Driver drowsiness detection using lightweight CNN with limited training data. *IEEE Access*, 12, 110476–110491. <https://doi.org/10.1109/ACCESS.2024.3440585>
- Vidyanathan Dixith, S., Jadhav, S., Kim, Y., & Jayakumar, N. (2025). Deep learning-based drowsiness detection system for driver's safety. *IEEE Access*, 13, 154080–154102. <https://doi.org/10.1109/ACCESS.2025.3604728>
- Weng, C.-H., Lai, Y.-H., & Lai, S.-H. (2017). Driver drowsiness detection via a hierarchical temporal deep belief network. In *Computer Vision – ACCV 2016 Workshops (Lecture Notes in Computer Science, Vol. 10118, pp. 117–133)*. Springer. [Foundational; NTHU-DDD dataset]
- Wu, Q., Li, N., Zhang, L., & Yu, F. R. (2024). Driver drowsiness detection based on joint human face and facial landmark localization with cheap operations. *IEEE Transactions on Intelligent Transportation Systems*, 25(12), 19633–19645. <https://doi.org/10.1109/TITS.2024.3443832>
- Xu, C., Huang, W., Liu, J., & Li, L. (2025). Detecting driver drowsiness using hybrid facial features and ensemble learning. *Information*, 16(4), 294. <https://doi.org/10.3390/info16040294>
- Xu, T., Guo, H., Zhang, Y., & Li, Q. (2026). MTSF-Net: A multiscale temporal-spatial fusion model for EEG-based driver fatigue detection. *IEEE Sensors Journal*, 26(6), 8423–8432. <https://doi.org/10.1109/JSEN.2026.3654563>
- Yang, E., & Yi, O. (2024). Enhancing road safety: Deep learning-based intelligent driver drowsiness detection for advanced driver-assistance systems. *Electronics*, 13(4), 708. <https://doi.org/10.3390/electronics13040708>
- Yang, L., Yang, H., Wei, H., Hu, Z., & Lv, C. (2024). Video-based driver drowsiness detection with optimised utilization of key facial features. *IEEE Transactions on Intelligent Transportation Systems*, 25(7), 6938–6950. <https://doi.org/10.1109/TITS.2023.3346054>
- Yao, K., Wang, Z., Guo, F., & Li, F. (2024). Driver abnormal expression detection method based on improved lightweight YOLOv5. *Electronics*, 13(6), 1138. <https://doi.org/10.3390/electronics13061138>
- Zayed, A., Belhadj, N., Ben Khalifa, K., Bedoui, M. H., & Valderrama, C. (2024). Efficient generalized electroencephalography-based drowsiness detection approach with minimal electrodes. *Sensors*, 24(13), 4256. <https://doi.org/10.3390/s24134256>
- Zayed, A., Trabes, E., Tarrillo, J., Ben Khalifa, K., & Valderrama, C. (2025). Efficient embedded system for drowsiness detection based on EEG signals: Features extraction and hardware acceleration. *Electronics*, 14(3), 404. <https://doi.org/10.3390/electronics14030404>
- Zhang, Y., Xu, X., Du, Y., & Zhang, N. (2025). TMU-Net: A transformer-based multimodal framework with uncertainty quantification for driver fatigue detection. *Sensors*, 25(17), 5364. <https://doi.org/10.3390/s25175364>

Zhao, Q., Wang, X., & Lei, Z. (2025). Driver fatigue degree detection based on eyes opening and closing ratio. *IEEE Instrumentation & Measurement Magazine*, 28(3), 30-37. <https://doi.org/10.1109/MIM.2025.10982115>

