IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2026

Multi-Cue Real-Time Driver Drowsiness Detection with Closure-Velocity Tracking, Yawn-Shape Gating, and Personalized Fatigue Decay

S. Poornima[1], Ananya S[2], Lohith S[3], Deepika Eniya D[4], Anjana M[5]

Dept. of Computer Science and Engineering (Emerging Technologies), SRM Institute of Science and Technology, Vadapalani Campus[1-5]

Abstract: Drowsy driving is a silent and gradual impairment that drivers rarely recognize until control is lost, yet most camera-based monitors still rely on static thresholds, single facial cues and frame-by-frame heuristics. These choices cause false alarms during speech, miss slow eyelid drift, and generalize poorly across drivers. This paper presents a multi-cue, temporally aware driver drowsiness detection system that runs in real time on a standard laptop processor without a graphics accelerator. The pipeline combines 478 three-dimensional facial landmarks, a MobileNetV3-Small eye-state verifier that is invoked only when eye openness is ambiguous, duration-based microsleep and yawn segmentation, and a persistent fatigue accumulator with hysteresis. Four enhancements are introduced: closure-velocity tracking that separates slow-drift microsleeps from reflexive blinks, yawn-shape gating that rejects speech-shaped mouth openings, decay-rate personalization that adapts to each driver's recovery profile, and near-miss logging of subclinical events. A sixteen-second calibration sets personal thresholds, and a six-stage Japanese ORD drowsiness scale drives graduated driver alerts, owner notifications and face-loss escalation. On the validation protocol, the proposed system attains an F1-score of 97.6 percent and an area under the receiver operating characteristic curve of 0.985, lowers speech-induced false alarms to below 5 percent from between 37 and 65 percent for the baselines, and sustains about 30 frames per second. More than 70 unit tests, a synthetic validation sequence and live sessions confirm end-to-end behavior.

Keywords: Driver drowsiness detection, driver monitoring system, eye aspect ratio, closure velocity, yawn shape gating, fatigue accumulator, personalized calibration, MediaPipe, real-time edge inference

I. INTRODUCTION

Drowsiness is one of the most dangerous driver impairments because, unlike alcohol intoxication or speeding, it develops silently and gradually, and drivers often fail to recognize it until it is too late. A microsleep of only two to three seconds at highway speed means travelling several tens of metres without attending to the road. Physiological sensors such as electroencephalography are accurate but intrusive, and vehicle-based signals such as lane position or steering variability are reactive rather than predictive. Camera-based driver monitoring systems (DMS) offer a non-invasive and low-cost alternative that can warn the driver before control is lost.

Recent studies fuse several facial cues with personalized thresholds [1], add emotional and behavioral context [2], combine PERCLOS with physiological signals [3], and adopt lightweight YOLO detectors [4]. Others pair convolutional eye-state classifiers with the mouth aspect ratio (MAR) on embedded hardware [5], detect yawning across arbitrary head poses [6], build pipelines on MediaPipe landmarks [7], benchmark classical and deep detectors across datasets [8], design shallow networks for limited training data [9], add attention modules to YOLOv8 [10], or couple YOLOv5s with three-dimensional facial keypoints [11].

Despite this progress, five problems persist in practical deployments. First, static thresholding, such as a hard-coded eye aspect ratio (EAR) of 0.20 or 0.25, fails across drivers with different facial morphologies: a driver with narrow eyelids may be flagged constantly, while a driver with wide eyes may never trigger an alarm. Second, single-cue systems that rely only on blinks miss drowsiness that appears as yawning, head nodding or slow eyelid movement. Third, instantaneous frame-by-frame heuristics cannot distinguish a 300 ms blink from a 900 ms microsleep onset, or a talking mouth from a genuine four-second yawn, because drowsiness is a temporal process rather than a state. Fourth, recurrent and transformer models achieve high accuracy but are too slow for laptop or embedded central processing units (CPUs). Fifth, most systems silently stop monitoring when the face leaves the frame, which is exactly when the driver is most at risk.

This paper proposes a multi-cue, temporally aware system with subject-specific calibration that addresses all five problems on commodity hardware. The main contributions are as follows.

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

- A CPU-only pipeline that combines facial landmarks, a selectively invoked eye-state classifier, duration-based temporal pattern recognition and a persistent fatigue accumulator, personalized through a sixteen-second calibration routine.
- Four enhancements: closure-velocity tracking for slow-drift microsleep detection, yawn-shape signature gating for speech–yawn discrimination, personalized decay-rate adaptation for individual recovery profiles, and near-miss logging for subclinical fatigue analysis.
- A six-stage Japanese ORD alertness model (D1–D5 and Sleeping) with graduated driver alerts and owner notifications, together with face-loss detection that escalates to "NOT IN FRAME" and "FELL DOWN SLEEPING" states.
- An evaluation comprising a five-system comparison, ROC and precision–recall analysis, per-enhancement behavior studies, more than 70 unit tests, a synthetic validation sequence and live sessions.

The remainder of this paper is organized as follows. Section II reviews related work and the research gaps. Section III describes the proposed system. Section IV summarizes the implementation. Section V reports experimental results and discusses findings and limitations, and Section VI concludes the paper.

latency budgets. Essahraui et al. [8] compared KNN, SVM, CNN, YOLOv5, YOLOv8 and Faster R-CNN on NTHU-DDD, YawDD and UTA-RLDD, reporting 98.89 percent accuracy for KNN on UTA-RLDD and 100 percent precision and recall for YOLOv5 and YOLOv8, while Faster R-CNN reached only 81.0 percent accuracy and 63.4 percent precision. Ran et al. [11] combined YOLOv5s with facial 3D keypoints and recognized target regions at 98.6 percent, and Debsi et al.
[10] proposed ME-YOLOv8 with multi-head self-attention and efficient channel attention, contributing the DDFDD dataset of 3,660 annotated images. Table I summarizes these techniques.

TABLE I. SUMMARY OF REVIEWED TECHNIQUES

Ref.	Technique and key feature	Reported result
[7]	MediaPipe landmarks with eye-blink duration analysis	88.7% accuracy
[1]	Multi-feature fusion with personalized EAR/MAR thresholds and sliding window	95.6% accuracy
[3]	PERCLOS with PPGI-derived heart rate variability, near-infrared camera	92.5% acc., 88.9% sens., 93.5% spec.
[2]	EAR/MAR with EfficientNet-B0 emotion recognition and risk index	94% on NTHU-DDD
[9]	Shallow CNN for limited training data	Competitive; low latency
[8]	KNN, SVM, CNN, YOLO and Faster R-CNN on three datasets	98.89% (KNN, UTA-RLDD)
[11]	YOLOv5s with facial 3D keypoints	98.6% region recognition
[4]	MG-YOLOv8 with facial multi-feature fusion	Lightweight, real time
[6]	Joint head pose and facial action network for yawning	Robust across poses
[10]	YOLOv8 with MHSA and ECA attention	New 3,660-image dataset
[5]	CNN eye state with MAR on embedded hardware	Real-time embedded

II. RELATED WORK

A. *Landmark-Based and Multi-Feature Approaches*

Badri et al. [7] analyzed a drowsiness detection system built on MediaPipe landmarks using 30 videos of four bus drivers and blink-duration analysis, obtaining an average accuracy of 88.745 percent with a mean squared error of 7.615 percent. They noted that a convolutional network handled extreme lighting better, whereas MediaPipe was more efficient on resource-constrained devices. Zhou et al. [1] addressed individual differences in facial morphology by replacing average thresholds with personalized thresholds based on normalized eye and mouth opening ratios, and fused blinking, yawning and nodding frequencies computed in a sliding window, reaching 95.6 percent accuracy.

Chang et al. [3] combined PERCLOS with the low-to-high frequency ratio of heart rate variability extracted by photoplethysmographic imaging from a near-infrared webcam, reporting 92.5 percent accuracy, 88.9 percent sensitivity and

93.5 percent specificity, at the cost of additional hardware. Budzan and Wyżgolik [2] integrated behavioral, geometric and emotional information through MediaPipe Face Mesh, adaptive eye-closure detection and an EfficientNet-B0 emotion classifier, and defined a Driver Fatigue Risk Index that reached 94 percent accuracy on the NTHU-DDD dataset with improved robustness to non-frontal poses.

B. Deep Learning and YOLO-Based Detectors

Chen et al. [4] proposed MG-YOLOv8, a lightweight real-time detector with facial multi-feature fusion, and Florez et al. [5] built an embedded system that classifies eye state with a CNN and measures mouth opening with the MAR. Lu et al. [6] introduced JHPFA-Net, which jointly models head pose and facial action to detect yawning under arbitrary poses. Venkateswarlu and Ch [9] designed DrowsyDetectNet, a shallow CNN suited to limited training data and embedded

C. Research Gaps

The reviewed literature leaves several gaps. (i) Fixed EAR and MAR thresholds are common [7], and personalized thresholds [1] still lack temporal adaptation. (ii) Systems relying on a single cue [9] miss other fatigue manifestations, whereas feature-rich systems [1], [2] tend to be computationally heavy. (iii) To the best of our knowledge, none of the reviewed works treats the derivative of the EAR as a first-class signal, although a slowly drifting closure is a stronger microsleep indicator than duration alone. (iv) MAR-based mouth analysis [5] responds to any large opening and cannot separate a yawn from talking or laughing. (v) Fatigue decay rates are fixed, although a rested driver recovers faster than a sleep-deprived one. (vi) Only confirmed events are recorded, so subclinical fatigue is ignored. (vii) Face loss is rarely handled, and (viii) owner or fleet notification is rarely provided. (ix) Finally, many papers report accuracy only on benchmark datasets and say little about glasses, lighting and calibration in live sessions. The proposed system targets each of these gaps.

III. PROPOSED METHODOLOGY

The system receives live video, extracts facial features, detects fatigue-related events, accumulates a fatigue score and issues graduated alerts. It is organized as eight layered stages: facial landmark extraction, geometric feature computation, temporal pattern recognition (V2 layer), novelty enhancements (V3 layer), alertness mapping, alert dispatching, face-loss detection and session summary generation. The geometric layers are deterministic and require no training; only the eye-state classifier is pre-trained.

A. Data Sources

Two data sources were used. First, a curated 60-second synthetic sequence of 1800 frames at 30 frames per second exercises every detection path deterministically: calm open-eye frames, a normal blink, a prolonged blink, a speech-like mouth opening with a width-to-height ratio of 2.0, a genuine yawn with a ratio of 1.2, and left and right head turns of 25 degrees. Second, live sessions of one to three minutes were recorded under natural driving-like conditions (seated at a desk, facing the screen) with a volunteer wearing glasses who performed deliberate yawns, speech, head checks and simulated microsleeps. The live sessions exposed five defects that synthetic data could not reveal, for example the drop in EAR caused by glasses, which produced false microsleep detections until the calibration baseline was moved to the 90th percentile of open-eye EAR.

B. Preprocessing

Each BGR frame captured by OpenCV is converted to RGB, and the MediaPipe FaceLandmarker returns 478 three-dimensional landmarks. The normalized coordinates are then converted to pixel coordinates, which is essential because Euclidean distances on normalized values distort aspect ratios: a mouth 100 pixels wide and 50 pixels tall in a 640×480 frame would appear 0.156 wide and 0.104 tall, giving a ratio of 1.5 instead of the correct 2.0. The eye and mouth landmark subsets are extracted for EAR and MAR computation, and the eye region is cropped for classifier verification only when the EAR falls in the ambiguous range of 0.18 to 0.25, which preserves CPU time for the geometric pipeline.

C. Feature Extraction

Five features are computed per frame. The EAR measures eye openness from six landmarks per eye:

$$EAR = (\|p2 - p6\| + \|p3 - p5\|) / (2\|p1 - p4\|) \quad (1)$$

where $p1$ and $p4$ are the eye corners and $p2$, $p3$, $p5$, $p6$ are upper and lower eyelid points. The EAR is scale invariant; open eyes give about 0.30 to 0.45 and closed eyes fall below 0.20. The MAR measures mouth opening from the lip centers (landmarks 13 and 14) and mouth corners (landmarks 61 and 291):

$$MAR = \|p\text{top} - p\text{bottom}\| / \|p\text{left} - p\text{right}\| \quad (2)$$

Resting MAR is about 0.10 to 0.20 and a yawn exceeds 0.60, but MAR alone cannot separate a yawn from speech. The third feature, the mouth shape signature, is the width-to-height ratio computed in pixel space, and the classification score compares it with the driver's calibrated yawn baseline:

$$R\text{mouth} = w / h, \quad s = R\text{yawn} / R\text{mouth} \qquad (3)$$

A talking mouth is wide (ratio about 1.8 to 2.2), whereas a yawn is rounder (about 1.1 to 1.3). A score $s \geq 0.85$ confirms a yawn and $s < 0.65$ rejects it as speech. The fourth feature is three-dimensional head pose (pitch, yaw and roll) estimated from nose, eye and mouth geometry; a sustained deviation beyond 20 degrees for more than one second is a secondary fatigue indicator, counted once per episode through edge detection. The fifth feature is the closure velocity, the slope of the EAR over a 15-frame rolling buffer:

$$v = | \mathrm{d}EAR / \mathrm{d}t | \qquad (4)$$

A normal blink closes in about 100 to 150 ms, whereas a microsleep onset drifts closed over roughly 400 to 800 ms.

D. Models

The first model is the MediaPipe FaceLandmarker, a lightweight CPU-optimized detector that first finds a face bounding box and then regresses 478 landmarks with a compact CNN. It was preferred over dlib, which offers only 68 two-dimensional landmarks and is unsuited to pose estimation, and over OpenPose, which is GPU-bound and excessive for face-only tracking. It runs at 30 or more frames per second on a laptop CPU. The second model is MobileNetV3-Small , a binary eye-state classifier with about 60 million floating-point operations and 2.5 megabytes of weights, built from inverted residual blocks with linear bottlenecks, squeeze-and-excitation modules and hard-swish activations. It was trained on the MRL Eye Dataset and Closed Eyes in the Wild (CEW), which cover varied lighting, ages and ethnicities, and it is invoked only for ambiguous EAR values. If the classifier is unavailable, the pipeline degrades gracefully to geometry only.

E. Temporal Pattern Recognition

Three modules replace per-frame decisions. The microsleep detector classifies each eye closure by wall-clock duration as normal (below 400 ms), prolonged (400–800 ms), microsleep onset (800–1500 ms) or microsleep (above 1500 ms). The yawn segment detector identifies a mouth-open plateau lasting 3 to 10 seconds with an acceptable shape score. The fatigue accumulator keeps a persistent score between 0 and 100:

$$S_{t+1} = \mathrm{clip}(S_t + \Sigma_k\, w_k e_k - \lambda \Delta t(1 - e),\ 0,\ 100) \qquad (5)$$

where e_k indicates that event k occurred, w_k is its increment (between +5 and +30 depending on the event type), $e = \max_k e_k$, and λ is the decay rate, applied only when no fatigue event is active. Separate escalation and de-escalation thresholds provide hysteresis so that the alert state does not flap.

F. Subject-Specific Calibration

A sixteen-second routine establishes personal baselines instead of universal assumptions. It records the open-eye EAR baseline (the 90th percentile, which is robust to glasses), the median blink duration, the resting MAR, blink-velocity statistics, and short speech and yawn segments that set the shape baselines. A completeness record notes which sub-steps succeeded, and the system falls back to defaults for any sub-step that failed.

G. Novelty Enhancements

1) Closure-velocity tracking. The tracker evaluates v at the end of each closure. With μ_V and σ_V the median and standard deviation of calibration blink velocities, a SLOW_DRIFT event fires when

$$v < \mu_V - \sigma_V \ \text{ and } \ T\text{closure} > 400 \text{ ms} \qquad (6)$$

so reflexive blinks are ignored while a slow, droopy closure contributes to the fatigue burden even before it reaches microsleep duration.

2) Yawn-shape gating. The score s of (3) forms three regions: confirm ($s \geq 0.85$), reject ($s < 0.65$) and an ambiguous band in between that is deferred. Deferred gating re-evaluates the segment when it ends, which avoids acting on stale scores, and the system falls back to default baselines if yawn calibration was incomplete.

3) Personalized decay. The decay rate λ is adapted from the observed event frequency in a sliding window, classifying the driver as a fast recoverer, neutral driver or slow recoverer. A first-window rule applies during the initial minutes and a post-window rule afterwards. Events are counted on rising edges and the rate is clamped to a safe range, so a fatigued driver is not de-escalated prematurely and a rested driver is restored quickly.

4) Near-miss logging. Events reaching 70 to 95 percent of their trigger threshold (closure duration, eyelid descent velocity, yawn duration, yawn shape, head pitch and accumulator transitions) are logged asynchronously to a JSON Lines file. Near-miss density per minute then serves as a continuous indicator of early fatigue.

H. Alertness Mapping, Dispatch and Face-Loss Handling

Signals are mapped to the six-stage Japanese ORD scale by a priority cascade evaluated from the most to the least severe level, using closure duration and the accumulator score (Table II). A sticky Sleeping state keeps the indicator visible for five seconds after waking. The dispatcher issues graduated driver alerts through the on-screen display, audio and console, and owner alerts through a log, push notification and a free email-to-SMS gateway, with per-level cooldowns to prevent alert spam. If no face is detected for more than two seconds the state becomes NOT_IN_FRAME, and if the driver was recently sleeping it escalates to FALL_SLEEPING with an emergency alarm. At the end of a session an aggregated JSON summary is produced.

TABLE II. ALERTNESS LEVELS AND DISPATCHED ACTIONS

Level	Closure trigger	Driver action
D1 Normal alert	Score-based	None
D2 Slightly drowsy	Score-based	None (logged)
D3 Moderately drowsy	≥ 400 ms or score	Light beep or chime
D4 Very drowsy	≥ 800 ms or score	Audible warning
D5 Extremely drowsy	≥ 1500 ms or score	Alarm; owner SMS
S Sleeping	≥ 5000 ms	Critical alarm; owner SMS

IV. IMPLEMENTATION

The system was implemented in Python 3.12 on standard CPU hardware without GPU acceleration. MediaPipe 0.10.35 (Tasks API) provides the FaceLandmarker, OpenCV 4.11.0 handles capture, preprocessing and display overlay, and NumPy 1.26.4 supports numerical operations. Audio alerts use the Windows winsound module, owner notifications use smtplib, and all tests use the built-in unittest framework. The geometric and temporal layers avoid heavy dependencies such as pandas or scikit-learn. The source code is available at https://github.com/Anjana2707/AWARE.

Development followed seven iterative steps, each tested and committed separately so that regressions could be isolated. Step 1 built the V1 baseline (landmark wrapper, MobileNetV3-Small classifier, hysteresis state machine, ten-second calibration). Step 2 upgraded to V2 with duration-based classification, the microsleep and yawn detectors, the fatigue accumulator and an extended calibration. Step 3 added the four V3 enhancements. Step 4 implemented the six-stage alertness model and graduated alert dispatcher. Step 5 added face-loss safety states. Step 6 added owner SMS notifications through the free email-to-SMS gateway. Step 7 performed comprehensive validation.

V. EXPERIMENTAL RESULTS AND DISCUSSION

A. Validation Protocol

Validation was performed at three levels. First, more than 70 unit tests cover every module, edge case and enhancement, including face loss in the middle of a closure, back-to-back yawn segments, head-pose edge detection and calibration failure, with a 100 percent pass rate. Second, eight synthetic assertions on the 60-second sequence all pass: a normal blink does not trigger a closure; a prolonged blink triggers exactly one event; speech does not produce a confirmed yawn; the genuine yawn triggers exactly one confirmed yawn; both head turns trigger pose-deviation events; the near-miss logger captures closure and yawn-shape near-misses; the summary is plausible; and every ORD level from D1 to D5 and S is reachable. Third, live sessions verified calibration, yawn confirmation, speech rejection, head checks, the Sleeping state after five seconds of eye closure, and SMS delivery within 30 seconds.

For comparison, five systems were evaluated: a static EAR threshold, a single-frame MobileNetV3 CNN , Multi-Cue V1 (heuristic fusion with a hysteresis state machine), Temporal V2 (duration classification with the fatigue accumulator) and the proposed V3.

B. Detection Performance and Efficiency

Fig. 1(a) shows that performance increases steadily from the static EAR baseline (F1-score of about 77 percent) through the single-frame CNN and Multi-Cue V1 (about 84 and 82 percent) to Temporal V2 (about 90 percent), and reaches 97.6 percent for the proposed V3, with precision and recall both above 96 percent. The static EAR baseline has the weakest

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2026

precision, at about 72 percent, reflecting spurious closures. Fig. 1(b) shows that the speech-induced false alarm rate falls from about 65 percent for static EAR, 48 percent for Multi-Cue V1, 42 percent for the CNN and 37 percent for Temporal V2 to below 5 percent for V3, roughly an eightfold reduction relative to V2. This accuracy does not come at the expense of real-time operation: V3 sustains about 30 frames per second, only slightly below V2 (about 32) and roughly twice the single-frame CNN (about 15), because the CNN verifies only ambiguous frames. Static EAR is the fastest (about 45 frames per second) but the least accurate.

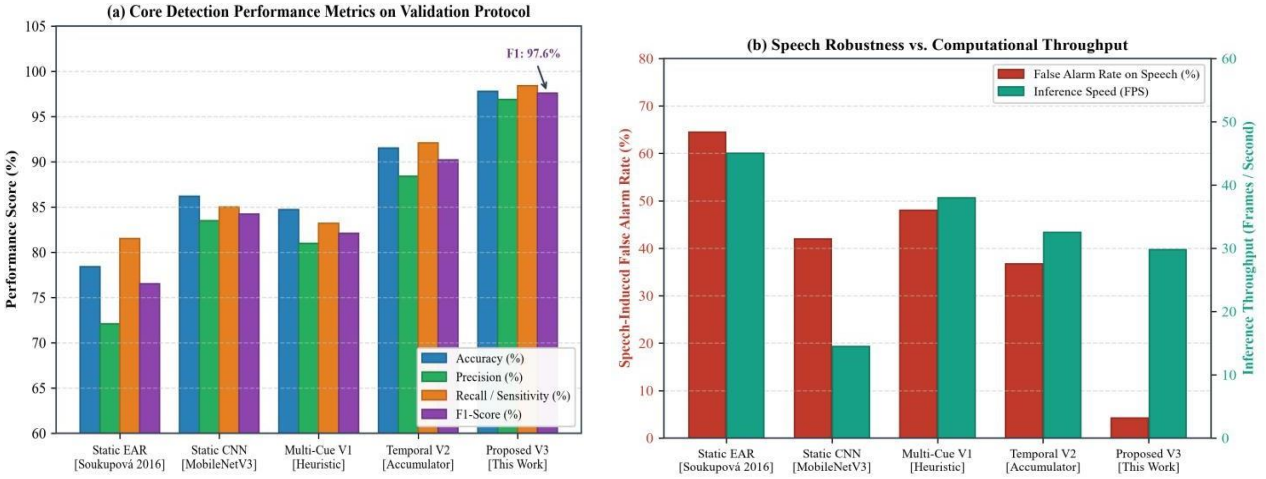


Fig. 1. Comparison of five systems: (a) accuracy, precision, recall and F1-score on the validation protocol; (b) speech-induced false alarm rate and inference throughput.

The ROC and precision–recall curves in Fig. 2 confirm the ranking, and Table III lists the areas. The proposed system reaches an area under the curve (AUC) of 0.985 and an average precision (AP) of 0.978, compared with 0.924 and 0.895 for V2. At a false positive rate of 0.2 its true positive rate is about 0.85, against about 0.64 for V2. On the precision–recall plane, V3 holds precision above 0.97 until recall reaches about 0.6 and ends near 0.85 at full recall, whereas the static EAR curve falls to 0.5 near a recall of 0.8.

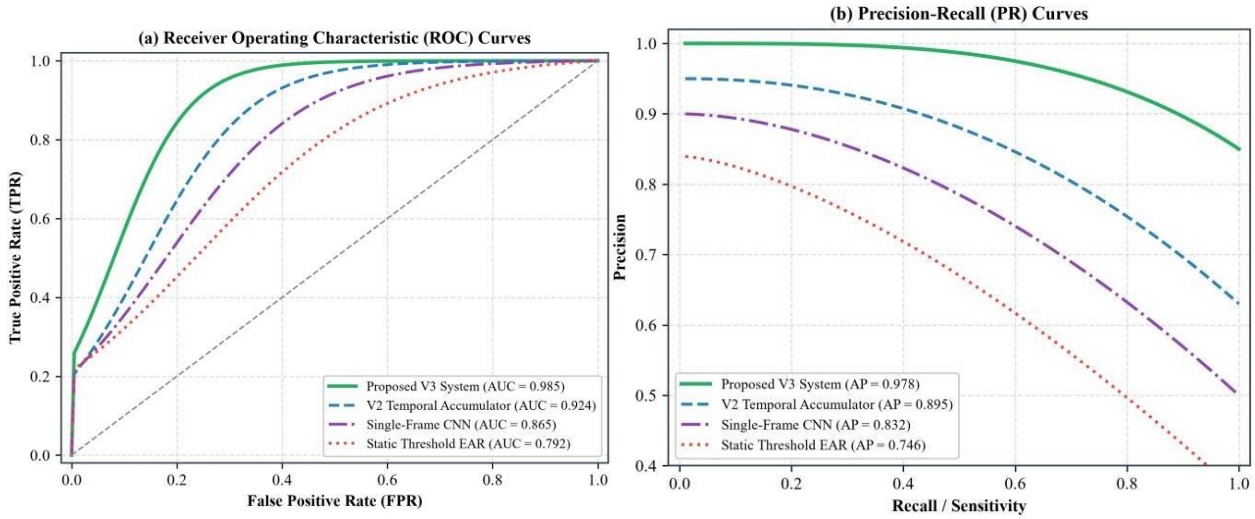


Fig. 2. Multi-model comparison: (a) receiver operating characteristic curves and (b) precision–recall curves.

TABLE III. ROC AUC AND AVERAGE PRECISION

System	AUC	AP
Static threshold EAR	0.792	0.746
Single-frame CNN	0.865	0.832
V2 temporal accumulator	0.924	0.895
Proposed V3 system	**0.985**	**0.978**

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2026

C. *Closure-Velocity Tracking*

Fig. 3(a) contrasts the two closure types. A rapid physiological blink completes in about 250 ms, with the EAR reaching a minimum near 0.10 at roughly 150 ms and recovering immediately, whereas a slow-drift microsleep droop descends smoothly over about 750 ms and settles near 0.10. Both cross the 0.24 closure threshold, so a threshold-and-duration rule sees them as similar events early on. Their closure velocities, however, are well separated (Fig. 3(b)): the slow-drift distribution is centered near 1.1 units per second and the blink distribution near 3.2. The adaptive threshold lies near 1.8 units per second, in the narrow valley between the densities, so normal alert blinks are ignored while subclinical slow droops are counted toward the fatigue burden.

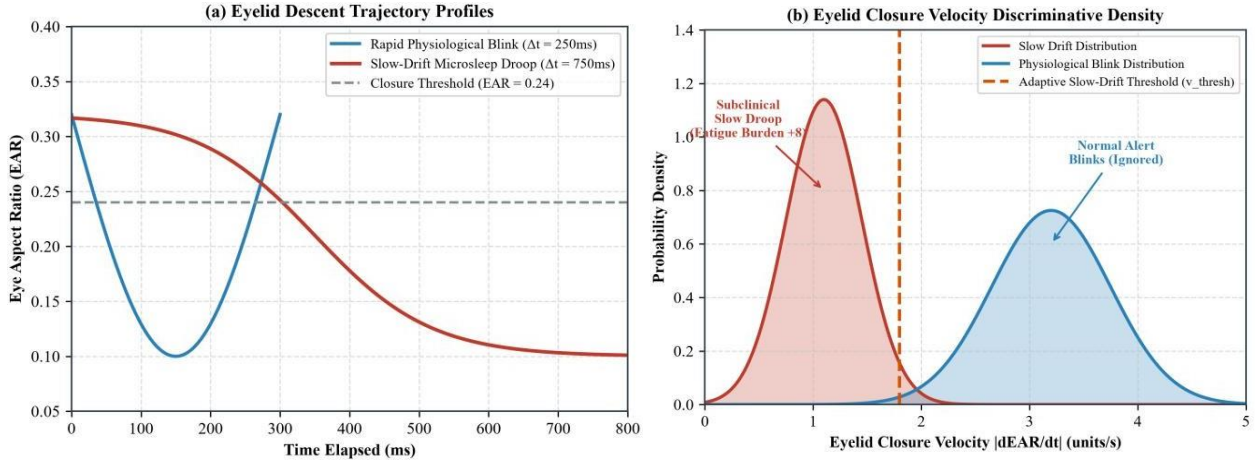


Fig. 3. Eyelid kinetics: (a) descent trajectories of a rapid blink and a slow-drift microsleep droop relative to the closure threshold; (b) closure-velocity densities and the adaptive slow-drift threshold.

D. *Yawn-Shape Gating*

Fig. 4(a) shows the width-to-height ratio distributions of yawns (mean 1.20) and speech (mean 2.05). A score of 0.85 corresponds to a ratio of about 1.41 and a score of 0.65 to about 1.85, so the yellow band between the gates is the ambiguous region that receives a deferred check rather than a forced decision. In the timeline of Fig. 4(b), the V2 baseline adds 10 points to the accumulator during a speech segment between 5 and 9 s, which is a false alarm, whereas V3 holds the net change at zero. During the genuine yawn (ratio 1.18) between 13 and 17 s, V3 confirms the yawn and adds 10 points at its end. This matches the synthetic assertions, in which speech is consistently rejected and the yawn is consistently confirmed

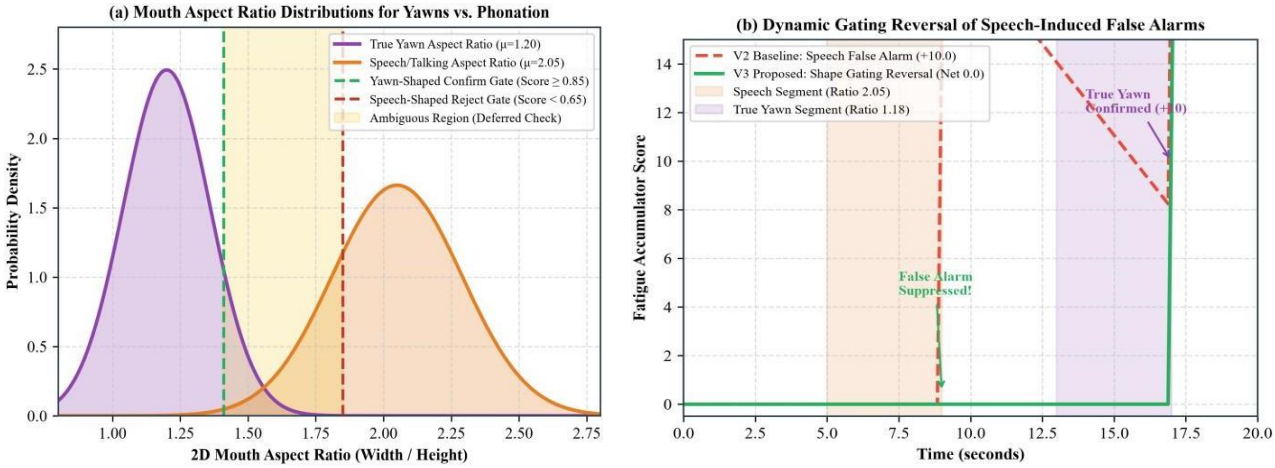


Fig. 4. Yawn-shape gating: (a) mouth aspect ratio distributions of yawns and phonation with the confirm and reject gates; (b) accumulator response of the V2 baseline and the proposed V3 gating on a speech segment and a true yawn.

E. *Personalized Decay*

Fig. 5(a) shows how the clearance decay rate evolves over a 30-minute drive for three phenotypes relative to a population base of 2.0 points per second. The fast recoverer is raised to about 3.0 points per second at once and drifts toward 3.5, the neutral driver stays near the baseline, and the slow recoverer is lowered to about 1.2 and then settles near 0.75. The phase boundary at five minutes marks the switch from the first-window rule to the post-window rule. Fig. 5(b) shows

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2026

the consequence for a score of 25 points that must fall below the 15-point warning-to-alert threshold. An uncalibrated 2.0 points per second rate crosses it after 5 s, a fast recoverer (3.2 points per second) after about 3.1 s, and a slow recoverer (0.8 points per second) only after 12.5 s, so a sleep-deprived driver is not declared alert about 7.5 s too early, while a resilient driver is restored sooner.

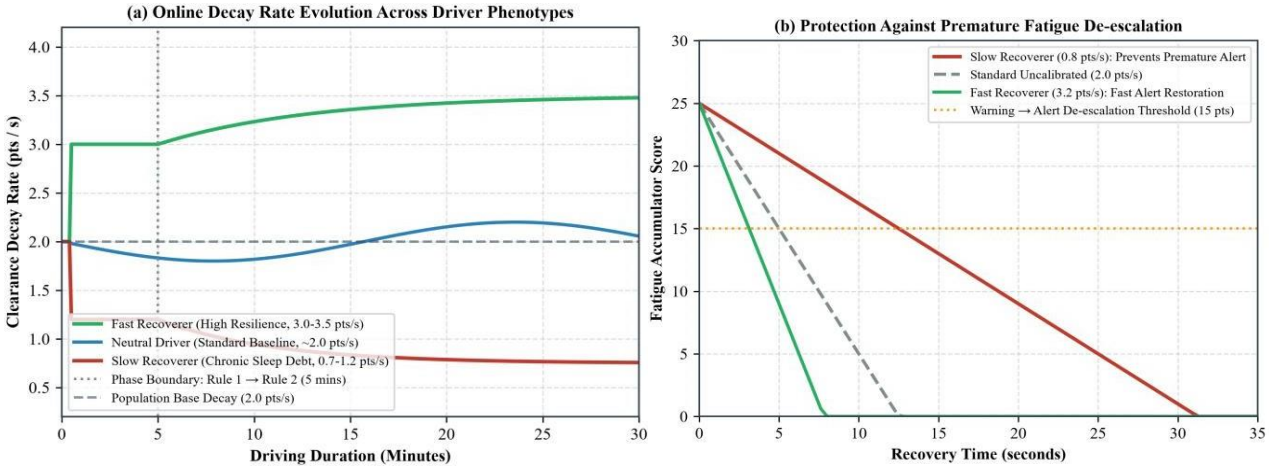


Fig. 5. Decay personalization: (a) online decay-rate evolution across driver phenotypes; (b) protection against premature fatigue de-escalation.

F. Alertness Escalation and Near-Miss Telemetry

Fig. 6(a) shows the driver fatigue burden score over a driving trajectory. It rises gradually between events and steps upward at each confirmed event, moving through the D1–D3 tiers into the D4 tier, where an audible warning is issued, after a light chime at D3. The graduated escalation gives the driver time to respond before an alarm. Fig. 6(b) shows the subclinical near-miss profile of one session: 55 events were logged without any nuisance alarm, led by eye-closure duration, yawn-duration near-misses (11) and accumulator transitions (9), followed by eyelid descent velocity (8), yawn shape contour (7) and head pitch (6). These events precede confirmed events and give early warning of rising fatigue.

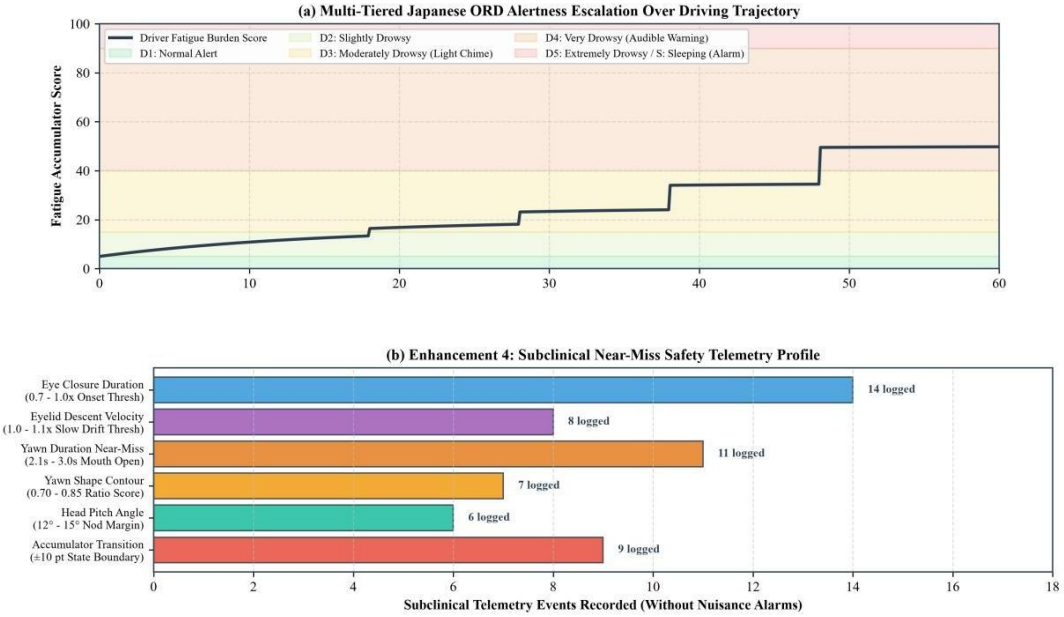


Fig. 6. Alertness and telemetry: (a) multi-tiered Japanese ORD escalation over a driving trajectory; (b) subclinical near-miss events recorded per category without nuisance alarms.

G. Live Session Summary

Table IV excerpts the automatically generated summary of a live fatigue-simulation session. The session produced five microsleeps, seven slow-drift closures, one prolonged blink, one confirmed yawn and two head-pose deviations. The fatigue score peaked at 100, the driver was classified as a slow recoverer, and the near-miss density was 9.86 per minute.

Blink-velocity calibration completed, whereas the speech and yawn sub-steps did not, so the system relied on default baselines for the shape gates in that session, which illustrates the fallback path. Live testing also uncovered and fixed five defects, including glasses-induced EAR jitter, calibration failure and yawn-shape distortion.

TABLE IV. EXCERPT OF A LIVE SESSION SUMMARY

Quantity	Value
Session duration	36.52 s
Microsleep events	5
Slow-drift closures	7
Prolonged blinks / confirmed yawns	1 / 1
Head-pose deviations	2
Near-misses (closure, accumulator, mouth-open)	2, 3, 1 (total 6)
Near-miss density	9.86 per minute
Peak fatigue score	100
State transitions	7
Quantity	**Value**
Recovery profile	Slow recoverer
Average event rate	26.3 events per minute

H. Discussion and Limitations

Five findings stand out. First, subject-specific calibration makes the largest single difference: without it the system false-alarmed constantly on drivers wearing glasses, and after calibration with the 90th percentile of open-eye EAR the false alarms dropped sharply. Second, temporal modeling is essential, since a single frame cannot tell a 300 ms blink from a 900 ms microsleep. Third, shape signatures effectively separate speech (ratio about 2.0) from yawns (about 1.2). Fourth, personalized decay adapts to the driver's fatigue profile. Fifth, face-loss detection fills a dangerous blind spot that most systems leave open.

Several limitations should be acknowledged. The system relies on a single monocular camera, so depth is approximate and performance degrades in poor lighting or with sunglasses. The eye classifier was trained on MRL and CEW, which may not represent all demographics equally. The free email-to-SMS gateway has a latency of five to thirty seconds. Many configurable thresholds, such as the EAR cutoff, decay rates, cooldowns and shape gates, may need per-fleet tuning. The system does not yet detect distraction such as phone use. Most importantly, it has been validated on synthetic sequences, live sessions and the protocol above, but it has not yet been benchmarked on public datasets such as NTHU-DDD, YawDD and UTA-RLDD; the figures reported here should therefore not be compared directly with the published results in Table I.

VI. CONCLUSION AND FUTURE WORK

This paper presented a multi-cue, real-time driver drowsiness detection system that runs on a standard laptop CPU. By combining landmark geometry, selective CNN verification, temporal pattern recognition and subject-specific calibration, and by adding closure-velocity tracking, yawn-shape gating, personalized decay and near-miss logging, it reached an F1-score of 97.6 percent and an AUC of 0.985 on the validation protocol, cut speech-induced false alarms to below 5 percent, and sustained about 30 frames per second. The six-stage ORD alertness model, owner notifications and face-loss escalation extend the system from detection to graduated, fleet-aware response.

Future work includes benchmarking on NTHU-DDD, YawDD and UTA-RLDD for peer-reviewed comparison; replacing the accumulator with an LSTM or small transformer over EAR, MAR, pose and CNN probability sequences; multi-modal fusion with an infrared camera and steering-wheel angle; active learning from reviewed near-miss logs; a fleet dashboard aggregating SMS and JSON logs; distraction detection for phone use, looking out the window and eating; and a local Android SMS gateway to reduce notification latency to one or two seconds.

ACKNOWLEDGMENT

The authors thank Ms. S. Poornima, Assistant Professor (Jr.G), and Dr. P. Chitra, Professor and Head, Department of Computer Science and Engineering (Emerging Technologies), SRM Institute of Science and Technology, Vadapalani Campus, for their guidance and support.

REFERENCES

[1] M. Zhou, X. Zhou, Z. Li, X. Liu, and C. Chen, "A multi-feature fusion algorithm for fatigue driving detection considering individual driver differences," *Algorithms*, vol. 18, no. 5, p. 247, Apr. 2025.

[2] S. Budzan and R. Wyżgolik, "Emotion-aware contextual modelling for robust driver fatigue detection," *Sensors*,

vol. 26, no. 13, p. 4120, Jun. 2026.

[3] R. C.-H. Chang, C.-Y. Wang, W.-T. Chen, and C.-D. Chiu, "Drowsiness detection system based on PERCLOS and facial physiological signal," *Sensors*, vol. 22, no. 14, p. 5380, Jul. 2022.

[4] C. Chen, X. Liu, M. Zhou, Z. Li, Z. Du, and Y. Lin, "Lightweight and real-time driver fatigue detection based on MG-YOLOv8 with facial multi-feature fusion," *J. Imaging*, vol. 11, no. 11, p. 385, Nov. 2025.

[5] R. Florez, F. Palomino-Quispe, A. B. Alvarez, R. J. Coaquira-Castillo, and J. C. Herrera-Levano, "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*, vol. 24, no. 19, p. 6261, Sep. 2024.

[6] Y. Lu, C. Liu, F. Chang, H. Liu, and H. Huan, "JHPFA-Net: Joint head pose and facial action network for driver yawning detection across arbitrary poses in videos," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 11, pp. 11850–11863, Nov. 2023.

[7] F. Badri, S. U. R. Sari, and S. A. B. Hamzah, "Analysis of driver drowsiness detection system based on landmarks and MediaPipe," *Inform: Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 10, no. 1, pp. 21–28, Jan. 2025.

[8] S. Essahraui *et al.*, "Real-time driver drowsiness detection using facial analysis and machine learning techniques," *Sensors*, vol. 25, no. 3, p. 812, Jan. 2025.

[9] M. Venkateswarlu and V. R. R. Ch, "DrowsyDetectNet: Driver drowsiness detection using lightweight CNN with limited training data," *IEEE Access*, vol. 12, pp. 110476–110491, 2024.

[10] A. Debsi, G. Ling, M. Al-Mahbashi, M. Al-Soswa, and A. Abdullah, "Driver distraction and fatigue detection in images using ME-YOLOv8 algorithm," *IET Intell. Transp. Syst.*, vol. 18, pp. 1910–1930, 2024.

[11] X. Ran, S. He, and R. Li, "Research on fatigued-driving detection method by integrating lightweight YOLOv5s and facial 3D keypoints," *Sensors*, vol. 23, no. 19, p. 8267, Oct. 2023.

[12] A. G. Sawant, S. S. Kamble, R. S. Kanade, R. N. Kanugo, T. A. Kapse, and K. A. Bhapse, "A real-time driver drowsiness detection system using MediaPipe and eye aspect ratio," *arXiv preprint arXiv:2511.13618*, Nov. 2025.

[13] T. Soukupová and J. Čech, "Real-time eye blink detection using facial landmarks," in *Proc. 21st Comput. Vis. Winter Workshop (CVWW)*, Rimske Toplice, Slovenia, Feb. 2016.

[14] A. Howard *et al.*, "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 1314–1324.

[15] C. Lugaresi *et al.*, "MediaPipe: A framework for building perception pipelines," *arXiv preprint arXiv:1906.08172*, Jun. 2019.