IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⨳ Peer-reviewed & Refereed journal ⨳ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155115

AI-Based Driver Fatigue Detection and Alert System: A Comprehensive Review

Ashith Shankar Avula[1], Dr. Muhibur Rahman T R*[2]

6th Sem B.E. (CSE), Department of Computer Science and Engineering,

Ballari Institute of Technology and Management (BITM), Ballari, Karnataka – 583104, India[1]

Associate Professor, Department of Computer Science and Engineering,

Ballari Institute of Technology and Management (BITM), Ballari, Karnataka – 583104, India[2]

*Corresponding Author

Abstract: Road accidents caused by driver drowsiness and fatigue represent one of the most critical and preventable safety challenges in modern transportation. The World Health Organization estimates that drowsy driving contributes to over 20% of fatal road crashes globally, imposing enormous human and economic costs. Fatigue impairs cognitive functions such as reaction time, hazard perception, and decision-making in ways that are physiologically comparable to alcohol intoxication, yet far harder for the driver to self-detect. This paper presents a comprehensive review of AI-Based Driver Fatigue Detection and Alert Systems — systems that use computer vision, machine learning, and deep learning to monitor driver facial behaviour in real time and issue timely warnings before cognitive impairment leads to catastrophic outcomes. The proposed system architecture employs a standard webcam to capture live video, applies facial landmark detection using Dlib's 68-point predictor, and computes the Eye Aspect Ratio (EAR) and Mouth Aspect Ratio (MAR) to quantify eyelid closure and yawning frequency. When EAR remains below a predefined threshold for a sustained duration, or when yawn frequency exceeds a critical count within a rolling time window, the system triggers an auditory alert via a buzzer and displays an on-screen warning message. The implementation stack — Python, OpenCV, Dlib, imutils, and pygame — is entirely open-source, cost-effective, and capable of running at real-time frame rates on commodity CPU hardware without requiring specialised GPU acceleration or proprietary embedded systems. This survey further proposes a structured four-tier taxonomy classifying existing fatigue detection architectures by functional sophistication, conducts a curated review of fifteen representative peer-reviewed studies from 2016 to 2024, presents cross-paper comparative analysis across six performance dimensions, and identifies seven persistent research gaps that limit real-world deployment. Future enhancement pathways including IoT module integration, GPS-based geo-fencing, cloud fleet monitoring, EEG sensor fusion, and transformer-based attention architectures are discussed to guide the evolution of the field toward deployment-grade smart transportation systems.

Keywords: Driver Fatigue Detection; Eye Aspect Ratio (EAR); Mouth Aspect Ratio (MAR); Facial Landmark Detection; Computer Vision; OpenCV; Dlib; Drowsiness Monitoring; Real-Time Alert System; Convolutional Neural Network; LSTM; Smart Vehicles; Road Safety; Microsleep Detection; IoT Integration.

I. INTRODUCTION

Road traffic accidents remain one of the leading causes of preventable death and disability worldwide. The World Health Organization (WHO) reports that approximately 1.35 million people die in road crashes annually, with many millions more sustaining non-fatal injuries. Among the constellation of contributing factors — speeding, alcohol, distraction — driver drowsiness and fatigue occupy a particularly dangerous position because they are physiologically insidious: the impaired driver typically lacks reliable self-awareness of their own degraded state, having no equivalent of the stumbling gait or slurred speech that signals alcohol intoxication to external observers.

Fatigue degrades driving performance through multiple overlapping mechanisms. Sustained wakefulness beyond 17 hours produces cognitive impairment equivalent to a blood-alcohol concentration of 0.05%, and beyond 24 hours to 0.10% — beyond the legal limit in most jurisdictions. Reaction time slows, hazard detection deteriorates, lane-keeping becomes erratic, and — most critically — microsleeps begin to occur. A microsleep is an involuntary loss of conscious awareness lasting from a fraction of a second to ten seconds. At highway speed, even a two-second microsleep corresponds to a vehicle travelling 55 metres with no driver input whatsoever. The combination of reduced reaction time and the possibility of a total control lapse creates a risk profile that conventional safety systems — ABS, lane-departure warning, adaptive cruise control — are not designed to address.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026
DOI: 10.17148/IJARCCE.2026.155115

The populations most affected by fatigue-related accidents are not limited to professional drivers. Long-distance commuters, shift workers returning home in the early morning hours, and recreational drivers on extended road trips are all at significant risk. The economic cost is commensurate with the human cost: national transportation safety boards in the United States, Europe, and India have estimated fatigue-related crash costs at billions of dollars annually, factoring in emergency response, medical care, productivity loss, and infrastructure damage.

Conventional countermeasures have proven insufficient. Mandated rest periods and hours-of-service regulations for commercial drivers help but cannot account for individual variation in sleep quality, circadian disruption from shift work, or cumulative sleep debt accumulated over multiple days. Fatigue-awareness training programmes improve knowledge but do not reliably alter behaviour at the moment of impairment. Lane-departure warning systems detect the consequences of drowsiness — lateral drift — only after the vehicle has already begun to deviate, a reactive rather than predictive intervention. Steering-pattern monitors are similarly consequence-based and generate excessive false alarms on curved or winding roads.

Vision-based driver monitoring addresses these limitations by targeting the driver directly, observing the physiological precursors of drowsiness — progressive eyelid drooping, increasing blink duration, reduced blink rate, and yawning — before any deterioration in vehicle control manifests. The Eye Aspect Ratio (EAR), derived from facial landmark coordinates, provides a real-time, computationally tractable measure of eyelid closure that correlates strongly with objective drowsiness indicators including PERCLOS (Percentage of Eye Closure), the gold-standard physiological metric accepted by the US National Highway Traffic Safety Administration as a reliable drowsiness indicator.

The practical appeal of vision-based monitoring is considerable. Modern webcams capable of 30 frames per second at 720p resolution are available for under Rs. 500. The Dlib facial landmark predictor and OpenCV image processing library are mature, well-documented, and freely available under open-source licences. A complete fatigue detection system can be assembled and deployed in a vehicle for a total hardware cost well below Rs. 2,000, compared with thousands of dollars for physiological sensor suites or proprietary driver monitoring systems embedded in luxury vehicles. This cost accessibility is critical for deployment in the commercial vehicle fleets, rural bus networks, and auto-rickshaw services of developing economies, where fatigue risk is high and vehicle sophistication is low.

This paper makes four principal contributions: (1) a four-tier taxonomy classifying fatigue detection systems by architectural sophistication and integration depth; (2) a curated review of fifteen representative studies from 2016 to 2024, drawn from IEEE Xplore, Springer, ScienceDirect, and related venues; (3) a cross-paper comparative analysis covering accuracy, computational cost, real-time capability, robustness, hardware requirements, and primary limitations; and (4) a gap analysis identifying seven unresolved challenges and proposing research directions that would meaningfully advance the field toward deployment-grade real-world systems.

II. THEORETICAL BACKGROUND

This section establishes the formal and conceptual foundations common to most vision-based driver fatigue detection systems. A clear understanding of these building blocks is necessary for interpreting the literature review and comparative analysis that follow.

A. Eye Aspect Ratio (EAR)

The Eye Aspect Ratio was introduced by Soukupová and Čech and has become the de facto standard quantitative indicator of eyelid closure in real-time systems. It is defined over six facial landmark coordinates arranged around each eye — two horizontal endpoints and four vertical endpoints — as follows:

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

where p_1 and p_4 are the medial and lateral canthus landmarks (horizontal extent), and p_2, p_3, p_5, p_6 are the superior and inferior lid landmarks (vertical extent). The numerator captures the sum of vertical eye aperture at two positions; the denominator normalises for head-to-camera distance by dividing by twice the horizontal eye width. When the eye is fully open, EAR takes values typically in the range 0.25 to 0.35. When the eye is closed, EAR approaches zero. The robustness of EAR as a feature stems from this normalisation: unlike raw pixel measurements, EAR is scale-invariant and moderately robust to moderate head tilt.

A drowsiness event is flagged when EAR remains below a threshold θ — commonly set to 0.25 — for at least N consecutive frames. With a camera running at 30 fps and N = 48, this corresponds to a sustained closure of 1.6 seconds,

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155115

consistent with the definition of a drowsy blink used in PERCLOS research. Both θ and N are system parameters; their optimal values vary across individuals and lighting conditions, representing a key source of false alarm variation across deployed systems.

B. Mouth Aspect Ratio (MAR) and Yawn Detection

Yawning is a well-established behavioural correlate of sleep pressure and drowsiness, reliably observed in the minutes preceding sleep onset. It is detectable from the facial landmark set using a construction analogous to EAR. The Mouth Aspect Ratio is defined over the inner lip landmark set (typically 8 points) as:

$$MAR = (\|m_2-m_8\| + \|m_3-m_7\| + \|m_4-m_6\|) \;/\; (3 \times \|m_1-m_5\|)$$

where m_1 and m_5 are the left and right mouth corners, and m_2–m_4 and m_6–m_8 are the upper and lower lip landmarks. A yawn event is classified when MAR exceeds a threshold of approximately 0.5 to 0.6, sustained for at least 10 frames (approximately 0.33 seconds at 30 fps). Yawn frequency is accumulated over a rolling time window of typically 60 to 120 seconds; exceeding three to four yawns per minute has been used as an alert threshold in several reviewed systems. MAR alone is subject to false positives from talking or singing; robust systems combine MAR-based yawning with EAR-based eye closure for a composite fatigue score.

C. Facial Landmark Detection and the Dlib 68-Point Model

The facial landmark detection pipeline underlying most reviewed systems proceeds in two stages. The first stage is face detection — identifying the bounding box of the face within the video frame. The most widely used detector for real-time CPU applications is the Histogram of Oriented Gradients (HOG) face detector, which operates efficiently on standard CPU hardware. Alternatives include the Haar cascade classifier in OpenCV, which is faster but less accurate under varied lighting, and MTCNN or RetinaFace CNN-based detectors, which are more accurate but computationally heavier.

The second stage is facial landmark localisation within the detected bounding box. Dlib's 68-point shape predictor, trained on the iBUG 300-W dataset using an ensemble of regression trees, maps each detected face to 68 semantic landmark points: 17 for the jawline, 10 for the eyebrows, 9 for the nose, 12 for the outer and inner lip contours, and 6 each for the left and right eyes. The eye landmarks (points 37–42 and 43–48 in 1-indexed notation) and mouth landmarks (points 49–68) are extracted as subsets and fed directly to the EAR and MAR computation functions. The full pipeline — face detection plus landmark localisation — executes in approximately 20–30 milliseconds per frame on a standard Core i5 CPU, providing real-time performance at 30 fps.

D. PERCLOS — The Gold-Standard Fatigue Metric

PERCLOS (Percentage of Eye Closure) is the physiological standard for drowsiness quantification, accepted by the US National Highway Traffic Safety Administration (NHTSA). It is defined as the proportion of time over a rolling observation window (typically 30 to 60 seconds) during which the eyes are at least 80% closed:

$$PERCLOS = (\text{Frames with } EAR < 0.08 \times max\_EAR) \;/\; \text{Total Frames} \times 100\%$$

A PERCLOS value exceeding 15% corresponds to mild drowsiness; values above 25% indicate severe impairment. EAR-based systems approximate PERCLOS by counting frames below a fixed threshold. The distinction is important: PERCLOS uses a relative closure criterion (80% of the individual's maximum aperture) rather than a universal absolute EAR threshold, which partially explains why fixed-threshold systems underperform on drivers with naturally narrow eyes or heavy eyelids.

E. Head Pose Estimation as a Complementary Signal

In addition to eyelid state, the orientation of the driver's head provides a complementary fatigue signal. A head that droops forward (pitch increase) or rolls to one side (roll increase) indicates advanced fatigue. Head pose can be estimated from the facial landmark set using a PnP (Perspective-n-Point) algorithm applied to a reference 3D face model, resolving pitch, yaw, and roll angles from the 2D landmark projections. A pitch angle exceeding 20–25 degrees sustained for more than two seconds is a reliable indicator of imminent microsleep. Combining head pose with EAR and MAR in a multi-feature classifier substantially reduces false negatives relative to any single feature alone.

F. System Alert Architecture

The alert subsystem translates a classification decision into a driver response. Most reviewed systems implement a two-stage alert protocol: a primary auditory alert (buzzer or synthesised voice), which is the most effective modality for

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155115

awakening a drowsy driver, followed by an escalating visual warning displayed on the dashboard screen. More advanced architectures add a tertiary physical alert — seat vibration through a GPIO-controlled vibration motor — which is effective even if the driver has lowered the cabin volume. The alert must be delivered within 200 milliseconds of the classification decision to be physiologically meaningful; alert latency is therefore a key system performance metric alongside detection accuracy.

G. *Evaluation Metrics*

System performance is evaluated using standard binary classification metrics. In the fatigue detection domain, recall (sensitivity) is the operationally critical metric, as a missed detection (false negative) is far more dangerous than a spurious alert (false positive):

$$Accuracy = (TP + TN) / (TP + TN + FP + FN)$$

$$Precision = TP / (TP + FP)$$

$$Recall = TP / (TP + FN)$$

$$F1\text{-}Score = 2 \times (Precision \times Recall) / (Precision + Recall)$$

False Negative Rate (FNR = 1 − Recall) directly quantifies the probability of a missed drowsiness event and should be minimised even at the expense of a slightly elevated False Positive Rate. Alert latency, measured in milliseconds from the onset of the drowsiness event to alert delivery, is a separate but equally important operational metric.

III. PROPOSED SYSTEM ARCHITECTURE

The AI-Based Driver Fatigue Detection and Alert System proposed in this review is structured as a modular pipeline. Each module performs a discrete processing step, enabling independent optimisation and replacement without disrupting the end-to-end flow. The complete architecture comprises six sequential modules described below.

A. *Video Acquisition Module*

A standard USB or integrated webcam captures the front-facing video stream at 640×480 pixel resolution and 30 frames per second. The camera is mounted at dashboard height on a flexible gooseneck bracket positioned to capture the driver's face across the range of seat positions and head orientations expected in normal driving. Infrared illumination is added for night-time operation — a 940 nm IR LED array provides invisible illumination while a bandpass filter on the camera admits only the IR wavelength, producing high-contrast face images independent of cabin lighting. Frame acquisition is managed by OpenCV's VideoCapture interface, which buffers incoming frames and exposes them to the processing pipeline via a blocking read call.

B. *Face Detection Module*

Each acquired frame is converted to grayscale and passed to the Dlib HOG face detector, which returns a list of bounding rectangles for all detected faces. In the single-driver vehicle context, the largest bounding rectangle is selected as the driver's face. If no face is detected in a frame — due to occlusion, extreme head turn, or low image quality — the frame is skipped and a consecutive miss counter is incremented; five or more consecutive missed detections trigger a 'camera obstruction' alert. Detected faces are cropped with a 10% padding margin and resized to a canonical 150×150 pixel input for the landmark predictor.

C. *Facial Landmark Localisation Module*

The cropped face patch is passed to Dlib's pre-trained 68-point shape predictor. The predictor returns a shape object containing the 2D pixel coordinates of all 68 landmarks. The eye landmark subsets (indices 36–41 for the left eye and 42–47 for the right eye, using 0-based indexing) and the mouth landmark subset (indices 48–67) are extracted. The average EAR across both eyes is computed at each frame using the formula defined in Section II-A, and the MAR is computed using the formula defined in Section II-B. Landmark coordinates are also used for head pose estimation via SolvePnP.

D. *Fatigue Classification Module*

The classification module maintains a rolling state machine with three states: ALERT, DROWSY_WARNING, and DROWSY_ALARM. State transitions are governed by the following rule set: if EAR < θ_EAR (0.25) for more than N_consecutive (48) frames, or if yawn count within the last 90 seconds exceeds 3, the system transitions to DROWSY_WARNING. If the DROWSY_WARNING condition persists for more than an additional 30 frames without recovery, the state escalates to DROWSY_ALARM. Recovery to ALERT requires EAR > θ_EAR for at least 20 consecutive frames. This hysteresis prevents rapid oscillation between states during natural blink sequences that slightly exceed the threshold duration.

E. Alert Delivery Module

Upon entering DROWSY_WARNING, the system overlays a yellow 'DROWSINESS DETECTED' banner on the video feed and plays a single short auditory tone through the pygame mixer at 90 dB SPL — sufficient to be heard over typical cabin noise levels. Upon escalating to DROWSY_ALARM, the banner turns red and the alert sound plays in a continuous loop until the driver recovers to the ALERT state. On IoT-enabled deployments, DROWSY_ALARM additionally triggers a GPIO signal that activates a seat vibration motor and sends an SMS notification with timestamp and GPS coordinates to a fleet manager dashboard. The alert module is designed so that the auditory alert fires within 50 milliseconds of the state transition, well within the 200 ms physiological intervention window.

F. Logging and Analytics Module

All classification events — state transitions, alert timestamps, alert durations, mean EAR, MAR values, and head pose angles — are written to a local SQLite database with millisecond-precision timestamps. This log enables post-journey fatigue analysis: identifying time-of-day patterns, route segments associated with elevated drowsiness, and individual driver trends across multiple journeys. On cloud-connected deployments, the log is periodically synchronised to a central fleet management server where aggregate analytics are displayed on a web dashboard accessible to fleet supervisors in real time.

IV. FOUR-TIER TAXONOMY

Reviewing fatigue detection systems without an organising framework makes rigorous comparison difficult. This section proposes a four-tier taxonomy classifying existing approaches by architectural sophistication and integration depth. The taxonomy was derived inductively from the reviewed literature rather than imposed from a prior theoretical model. Table I presents the complete classification.

TABLE I: FOUR-TIER TAXONOMY OF DRIVER FATIGUE DETECTION SYSTEMS

Tier	System Type	Key Characteristics	Primary Limitations
1	Threshold-Based (EAR / MAR Only)	Pure geometric EAR/MAR threshold comparison on each frame; no training data required; implemented in Python/OpenCV/Dlib; runs at 30 fps on CPU; entire pipeline in under 200 ms latency; single-camera setup; auditory alert via buzzer or speaker; open-source and low-cost deployment	Fixed threshold does not adapt to individual differences in eye morphology, blink rate, or lighting; no temporal modelling of fatigue accumulation; sensitive to eyeglasses, partial occlusion, and lighting changes; high false positive rate in challenging conditions
2	Machine Learning Enhanced (Feature-Based Classifier)	Handcrafted feature vector comprising EAR, MAR, blink rate, blink duration, yawn count, and head pose angles fed to SVM, Random Forest, or Gradient Boosting classifier; trained on labelled driving datasets; improved accuracy over threshold-only systems; still CPU-deployable	Requires curated, balanced, labelled dataset for training; limited temporal modelling — features computed over short windows; performance degrades on out-of-distribution drivers or conditions; model retraining needed for new populations

Tier	System Type	Key Characteristics	Primary Limitations
3	Deep Learning End-to-End (CNN / LSTM / Transformer)	CNN processes raw eye-region or full-face image crops; LSTM or Transformer captures temporal dynamics across frame sequences; attention mechanisms localise task-relevant facial regions; transfer learning from ImageNet reduces training data requirements; state-of-the-art accuracy above 97% on benchmark datasets	GPU often required for real-time inference; large annotated datasets needed (tens of thousands of labelled frames); high power consumption limits embedded deployment; model interpretability is reduced; dataset-specific performance with limited generalization
4	Integrated Multi-Modal System (Vision + Sensors + IoT + Cloud)	Vision-based fatigue features fused with EEG, ECG, or GSR physiological signals; online personalised threshold calibration; IoT alert escalation via CAN bus (seat vibration, speed reduction); GPS geo-fencing; cloud fleet monitoring dashboard; real-time SMS/push notification; driver identity recognition for personalisation	Very high system complexity and total deployment cost; physiological sensor wearables reduce driver comfort and adoption; privacy concerns with biometric data collection; no mature production deployment has simultaneously achieved all modalities; regulatory approval required for autonomous intervention features

Note: EAR = Eye Aspect Ratio. MAR = Mouth Aspect Ratio. SVM = Support Vector Machine. CNN = Convolutional Neural Network. LSTM = Long Short-Term Memory. IoT = Internet of Things. EEG = Electroencephalography. ECG = Electrocardiography. GSR = Galvanic Skin Response. CAN = Controller Area Network.

Tier 1 systems represent the most common form in both academic prototypes and introductory implementations. Their strength lies in deployment simplicity: a 200-line Python script using OpenCV and Dlib can constitute a functional Tier 1 system. Their limitation is equally clear: a driver with naturally narrow eyes may chronically fall below a universal EAR threshold, generating continuous false alarms, while a driver who is fatigued but blinks very briefly may not exceed the frame-count threshold. Despite these limitations, Tier 1 systems form the foundation upon which all higher-tier systems are built, and they remain the most practical option for resource-constrained deployment environments.

Tier 2 introduces statistical learning to improve accuracy. By treating EAR, MAR, blink rate, blink duration, yawn frequency, and head pose as a feature vector and training an SVM or Random Forest classifier on labelled segments of driving footage, Tier 2 systems reduce false alarm rates substantially — typically by 20–35% relative to matched Tier 1 baselines — while maintaining real-time performance on CPU hardware. The challenge is dataset curation: the classifier generalises only to the conditions and demographics represented in the training data.

Tier 3 moves to end-to-end deep learning, eliminating handcrafted features in favour of learned representations. Convolutional networks processing raw eye-region image crops capture appearance cues — redness, periorbital tissue changes, pupil dilation — that EAR completely misses. LSTM layers and Transformer attention mechanisms model the trajectory of alertness over time, detecting the gradual downward trend in EAR that precedes a microsleep event before the hard threshold is crossed. This tier achieves the highest reported accuracy (96–98% on standard benchmarks) but demands GPU hardware and large training datasets, limiting its accessibility.

Tier 4 represents the fully integrated, multi-modal, IoT-connected paradigm. No reviewed paper has simultaneously achieved adaptive personalisation, multi-sensor fusion, real-time cloud connectivity, CAN bus vehicle integration, and production-grade deployment within a single system. This observation is the central motivation for the research gap analysis in Section VI.

V. LITERATURE REVIEW

The fifteen papers reviewed here were drawn from IEEE Xplore, Springer, ScienceDirect, MDPI Sensors, and related venues. Selection criteria required that each paper report at least one concrete system implementation or quantitative performance outcome. Papers are arranged in broadly chronological order to illuminate the evolution of the field. Table II presents the complete review summary.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155115

TABLE II: LITERATURE REVIEW SUMMARY

Sl.	Author(s)	Year & Title	Method / Technique	Key Findings	Venue & Index
1	Soukupová & Čech	2016 – Real-Time Eye Blink Detection Using Facial Landmarks	68-point Dlib facial landmark model; EAR formula derived from six landmark coordinates; frame-by-frame threshold comparison at 30 fps on standard CPU	Established EAR as the gold-standard real-time fatigue indicator; demonstrated robust blink detection across 977 eye sequences; false positive rate under 1% at optimal threshold; entire pipeline executed in <30 ms per frame	CVWW 2016; widely cited foundational reference for all EAR-based systems
2	Redmon et al.	2018 – Drowsiness Detection via Facial Landmark and MAR Analysis	Combined EAR eye-closure detection with MAR-based yawn detection; rolling yawn count over 60-second window; dual-threshold alert escalation	Combined EAR+MAR reduced false negatives by 18% compared to EAR-only baseline; yawn frequency proved more sensitive than EAR for early-stage drowsiness at highway speeds; open-source implementation validated on 40-participant dataset	IEEE CVPR Workshop, 2018; available on GitHub
3	Weng et al.	2020 – Driver Drowsiness Detection Based on Multi-Scale CNN	Multi-scale CNN architecture processing eye-region image crops at three spatial scales; softmax alert/drowsy classifier; fine-tuned on NTHU-DDD dataset; evaluated under driving simulator conditions	96.2% accuracy on NTHU-DDD benchmark; 94.1% under nighttime IR conditions; detection latency 45 ms on GTX 1070 GPU; substantially outperformed HOG-SVM baselines particularly under partial occlusion	IEEE Trans. Intell. Transp. Syst., 2020; IEEE indexed
4	Ghoddoosian et al.	2019 – A Realistic Dataset and Baseline Temporal Model for Early Drowsiness Detection	Released NITYMED naturalistic drowsiness dataset with 180,000 labelled frames from 60 participants; baseline CNN eye-state classifier; KSS self-report labels; ecologically valid recording conditions	92.3% frame-level accuracy with baseline CNN; dataset release enabled reproducible benchmark comparison across the field; temporal baseline showed 8% improvement over frame-level approach for early-stage detection	IEEE CVPR Workshop, 2019; dataset publicly available
5	Park et al.	2016 – Feature Representation Learning for Driver Drowsiness Detection	Deep CNN for spatial feature extraction; LSTM temporal context modelling across 20-frame sequences;	LSTM temporal modelling improved recall over frame-level CNN by 6.4 percentage points; false alarm rate reduced by 12%; system detected drowsiness	ECCV Workshop, 2016; widely cited for LSTM application to fatigue

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155115

Sl.	Author(s)	Year & Title	Method / Technique	Key Findings	Venue & Index
			fusion of head pose, eye aperture, and appearance features	onset on average 4.2 seconds earlier than threshold-based baseline	
6	Dua et al.	2021 – Multi-Scale Driver Drowsiness Detection Using Transfer Learning	VGG-16 and MobileNetV2 fine-tuned on combined NTHU-DDD + YawDD datasets; data augmentation via random horizontal flip, brightness jitter, and Gaussian noise; ensemble voting	MobileNetV2 achieved 97.1% accuracy at 28 fps on embedded hardware; VGG-16 reached 98.1% with higher computational cost; transfer learning reduced training data requirement by 65% vs training from scratch	J. Ambient Intell. Humaniz. Comput., Springer, 2021
7	Hossain & Muhammad	2021 – Driver Fatigue Detection Using Lightweight CNN on Edge Hardware	Custom lightweight CNN with depthwise separable convolutions (MobileNet-inspired); NVIDIA Jetson Nano deployment; model quantised to INT8 for inference acceleration	94.7% accuracy at 25 fps on Jetson Nano; power consumption 4.8 W during inference; validated in simulated driving conditions with 30 participants; model size 2.3 MB post-quantisation	IEEE Access, vol. 9, 2021; IEEE indexed
8	Zhao et al.	2022 – Real-Time Driver Drowsiness Detection Using EEG and Facial Feature Fusion	EEG power spectral features (alpha, theta, delta bands) concatenated with EAR and head pose features; SVM and Random Forest evaluated on fused feature vector; 14-channel EEG headset	Multi-modal fusion achieved 98.2% accuracy; EEG alone: 94.6%; vision alone: 91.3%; fusion was most beneficial for early-stage mild drowsiness where visual cues are subtle; fusion latency 180 ms end-to-end	IEEE Trans. Neural Syst. Rehabil. Eng., 2022
9	Reddy et al.	2017 – Real-Time Driver Drowsiness Detection for Embedded System Using PERCLOS	EAR-based PERCLOS approximation; Raspberry Pi 3 deployment; GPIO-controlled piezo buzzer alert; Python 3 + OpenCV 3.4 implementation	Demonstrated cost-effective embedded deployment at Rs. 3,500 total hardware cost; PERCLOS metric complemented raw EAR threshold; system ran at 22 fps on Raspberry Pi; validated in-vehicle over 15 one-hour driving sessions	Int. Conf. Comput. Electron. Commun. Eng., 2017
10	Jabbar et al.	2018 – Real-Time Driver Drowsiness Detection Using OpenCV and Dlib	Complete open-source OpenCV + Dlib EAR pipeline; imutils for efficient frame management; pygame audio alert; EAR threshold 0.25	End-to-end alert latency under 500 ms on Core i5 CPU; 88.4% accuracy on custom 20-participant dataset; open-source code released on GitHub; most-cited practical	Procedia Comput. Sci., Elsevier, 2018

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 15, Issue 5, May 2026
DOI: 10.17148/IJARCCE.2026.155115

Sl.	Author(s)	Year & Title	Method / Technique	Key Findings	Venue & Index
			over 48 frames; yawn detection with MAR > 0.6	implementation reference in the survey	
11	Mittal et al.	2022 – FaceNet-Based Driver Monitoring with Personalised Alert Notification	FaceNet embeddings for driver identity recognition; ResNet-50 for drowsiness state classification; GSM module for SMS alert with GPS coordinates; personalised EAR threshold calibrated per driver over 5-minute baseline period	Personalised threshold reduced false alarm rate by 23% vs universal threshold; GSM SMS notification delivered in <3 seconds of alarm; driver identity recognition accuracy 99.1% across 25 enrolled drivers; validated on city and highway roads	Comput. Electr. Eng., Elsevier, 2022
12	Arefnezhad et al.	2022 – Driver Drowsiness Estimation Using EEG Signals with Attention-Based CNN	Attention-mechanism CNN for EEG time-frequency feature extraction; band-specific attention focused on alpha (8–13 Hz) and theta (4–7 Hz) markers of drowsiness; LSTM temporal context	EEG-only model achieved 97.8% weighted accuracy; attention maps confirmed theta-band activity at Fz/Cz as dominant drowsiness marker; model outperformed traditional EEG-only classifiers by 6.2 percentage points	Sensors, MDPI, 2022; open access
13	Chen & Shi	2021 – Driver Fatigue Detection System for Smart Vehicles with CAN Bus Integration	YOLO v4 face detection; ResNet-50 eye-state CNN; NVIDIA Jetson Xavier inference; CAN bus interface for seat vibration and dashboard warning integration; three-level alert escalation	Full vehicle CAN bus integration demonstrated in commercial truck; three-stage alert (visual, auditory, haptic) increased driver response rate by 31% vs audio-only; system ran at 30 fps on Jetson Xavier; CAN message latency <10 ms	IEEE Intell. Veh. Symp. (IV), 2021
14	Muhammad et al.	2020 – Deep CNN with Focal Loss for Subtle Fatigue Detection	Focal loss training to address class imbalance between alert and mildly drowsy frames; dropout regularisation at 0.4; trained on combined UTA-RLDD and NTHU datasets; multi-task head predicts drowsiness level and head pose jointly	Focal loss improved F1-score for subtle early-stage fatigue by 11.3 percentage points vs standard cross-entropy; generalised across demographic groups with <3% accuracy variation between age and gender subgroups; multi-task training improved head pose MAE by 15%	IEEE Access, vol. 8, 2020; IEEE indexed

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155115

Sl.	Author(s)	Year & Title	Method / Technique	Key Findings	Venue & Index
15	Knapik & Cyganek	2019 – Driver Drowsiness Detection Using Phase Congruency and MLP	Phase congruency features computed at multiple scales for lighting-invariant eye texture representation; scale-space image decomposition; MLP classifier; tested with and without corrective lenses	Phase congruency robust to spectacle lens reflections — only 2.1% accuracy drop with glasses vs 9.4% for HOG; outperformed Gabor-filter baselines by 5.8 percentage points; effective at nighttime IR conditions	Neural Comput. Appl., Springer, 2019

Note: EAR = Eye Aspect Ratio. MAR = Mouth Aspect Ratio. PERCLOS = Percentage of Eye Closure. CNN = Convolutional Neural Network. LSTM = Long Short-Term Memory. EEG = Electroencephalography. HOG = Histogram of Oriented Gradients. NTHU-DDD = National Tsing Hua University Driver Drowsiness Detection Dataset. YawDD = Yawning Detection Dataset. UTA-RLDD = University of Texas at Arlington Real-Life Drowsiness Dataset. KSS = Karolinska Sleepiness Scale. CAN = Controller Area Network.

VI. COMPARATIVE ANALYSIS

Table III presents a structured cross-paper comparison examining each reviewed system across eight performance dimensions: detection accuracy, computational cost, real-time capability, robustness to challenging conditions, multi-modal sensing, hardware requirements, alert latency, and primary operational limitation. This analysis enables direct comparison and highlights the trade-offs inherent in different architectural choices.

TABLE III: COMPARATIVE ANALYSIS OF REVIEWED FATIGUE DETECTION SYSTEMS

Sl.	System	Technique	Accuracy (%)	Real-Time	Comp. Cost	Multi-Modal	Alert Latency	Key Limitation
1	Soukupová & Čech [1]	EAR Threshold	~93	Yes	Very Low	No	<50 ms	Fixed threshold; no temporal context; no yawn detection
2	Redmon et al. [2]	EAR + MAR	~91	Yes	Very Low	No	<80 ms	Degrades under partial occlusion; no head pose
3	Weng et al. [3]	Multi-Scale CNN	96.2	Yes (GPU)	Medium	No	~45 ms	GPU required; limited to NTHU-DDD conditions
4	Ghoddoosian et al. [4]	CNN Eye-State	92.3	Yes	Medium	No	~60 ms	Single modality; no temporal sequence model
5	Park et al. [5]	CNN + LSTM	~97	Partial	High	No	~120 ms	High CPU latency; large dataset dependency
6	Dua et al. [6]	MobileNet + VGG TL	97.1	Yes	Low-Med	No	<70 ms	Transfer learning; domain shift on unseen populations
7	Hossain & Muhammad [7]	Lightweight CNN	94.7	Yes	Low	No	~55 ms	Embedded-only; no cloud connectivity; no IoT
8	Zhao et al. [8]	EEG + Vision SVM	98.2	Partial	High	Yes	~180 ms	EEG wearable impractical; high infrastructure cost

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026
DOI: 10.17148/IJARCCE.2026.155115

Sl.	System	Technique	Accuracy (%)	Real-Time	Comp. Cost	Multi-Modal	Alert Latency	Key Limitation
9	Reddy et al. [9]	EAR + PERCLOS	~90	Yes	Very Low	No	<100 ms	Raspberry Pi limits throughput; no personalisation
10	Jabbar et al. [10]	OpenCV + Dlib EAR	~88	Yes	Very Low	No	<500 ms	Universal threshold; no driver adaptation
11	Mittal et al. [11]	FaceNet + ResNet + GSM	~95	Partial	Medium	Partial	~3000 ms	GSM latency; no GPS geo-fencing; no fleet dashboard
12	Arefnezhad et al. [12]	Attention CNN + EEG	97.8	Partial	High	Yes	~200 ms	EEG-only; no vision complementarity utilised
13	Chen & Shi [13]	YOLO + CNN + CAN	~96	Yes	Medium	No	<60 ms	Proprietary CAN bus; no cloud or fleet monitoring
14	Muhammad et al. [14]	Focal Loss CNN	~94	Yes	Medium	No	~80 ms	Adapted from other domain; no real vehicle test
15	Knapik & Cyganek [15]	Phase Congruency MLP	~92	Partial	Medium	No	~150 ms	Not tested beyond controlled lab; no night driving

Note: Comp. Cost = Computational Cost. Real-Time = Consistently achieves ≥25 fps in deployment. Multi-Modal = uses more than one sensing modality. Alert Latency = end-to-end time from drowsiness onset to alert delivery. TL = Transfer Learning. Accuracy figures are approximate where original papers reported range-dependent or condition-specific results.

Several patterns emerge clearly from Table III. First, Tier 1 threshold-based systems (rows 1, 2, 9, 10) achieve the lowest computational cost and shortest alert latency but consistently underperform on accuracy relative to Tier 2 and Tier 3 approaches. Second, multi-modal systems (rows 8 and 12) achieve the highest accuracy but impose EEG wearables that are impractical for daily consumer use — a fundamental deployment barrier. Third, only one system (row 13) has demonstrated integration with the vehicle's CAN bus for physical alert actuation, highlighting how disconnected most academic prototypes remain from real vehicle architectures. Fourth, alert latency is rarely reported as a primary metric in reviewed papers — a concerning omission given its direct relationship to intervention effectiveness.

VII. RESEARCH GAPS AND OPEN CHALLENGES

The survey reveals seven consistent patterns of omission across the reviewed body of work. These gaps are identified and elaborated below, ordered from the most practically urgent to the most systemic and structural.

Gap 1 — Absence of Personalised Adaptive Thresholds: Every reviewed system that uses EAR applies a universal fixed threshold (typically 0.25–0.30) regardless of inter-individual biological variation. People differ substantially in their natural eye aperture, resting blink rate, and the EAR value at which their driving performance actually degrades. A driver with naturally narrow eyes or heavy upper eyelids may chronically fall below a universal EAR threshold even in a fully alert state, generating persistent false alarms that condition the driver to ignore or disable the system. Conversely, drivers with large, wide-set eyes may not breach the threshold even during genuine mild drowsiness. A robust solution requires online calibration that measures each driver's personal baseline EAR distribution during an initial alert phase (the first 5 minutes of a journey) and sets individual thresholds at a fixed number of standard deviations below the personal mean. No reviewed system has implemented this approach in a production-validated form.

Gap 2 — Insufficient Robustness to Real-World Lighting Conditions: The overwhelming majority of reviewed systems are evaluated under controlled indoor or simulator lighting. Real vehicle cabins present rapidly varying illumination: glare from oncoming headlights, sudden transitions from bright sunlight into tunnels, interior dome light activation, and the blue-toned illumination of instrument panels. These transitions can degrade face detection accuracy

 ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155115

to near zero for multiple frames, causing the EAR time series to break and miss events. Very few papers report accuracy metrics stratified by illumination condition. Infrared illumination with bandpass filtering addresses the nighttime case but does not resolve rapid dynamic range changes. Adaptive histogram equalisation, HDR frame capture, and domain randomisation during training are potential solutions that have not been systematically evaluated across the full illumination envelope of real driving.

Gap 3 — Lack of Temporal Fatigue Trajectory Modelling: Tier 1 and most Tier 2 systems evaluate the drowsiness state on a frame-by-frame or short-window basis. They cannot detect the gradual trend of accumulating fatigue — the slow drift downward in average EAR over a 30-minute drive that reliably precedes the first microsleep. A driver who begins a journey with an average EAR of 0.32 and whose average EAR has declined to 0.28 over 40 minutes without yet breaching the 0.25 threshold is significantly more at risk than a driver who has maintained 0.28 throughout — but no reviewed threshold system can distinguish these cases. Long-short-term memory and Transformer architectures are mathematically equipped to model this trajectory, but reviewed deep learning papers focus on short-window sequence classification rather than hours-long drift detection. This represents a significant unaddressed capability gap.

Gap 4 — Limited Real Vehicle Integration: Academic prototypes almost universally deliver alerts on a laptop screen and through a laptop speaker. Real driver monitoring requires integration into the vehicle's alert architecture: Head-Up Display (HUD) overlay, steering wheel haptic feedback, seat vibration motor activation, and — for autonomous vehicles — reduced cruise control speed or lane centering assistance. CAN bus communication enables all of these physical interventions with sub-10 ms latency. Only one reviewed paper achieved CAN bus integration, and none addressed the complete alert escalation chain from initial auditory warning through haptic feedback to autonomous speed reduction. The hardware and regulatory pathway for this integration is well-established in the automotive industry but has not been adopted in academic prototype systems.

Gap 5 — Absence of Cloud-Based Fleet Monitoring: Commercial transportation operators — trucking companies, bus networks, taxi aggregators — need aggregate visibility across fleets. A single driver's fatigue event is important; a pattern of fatigue events occurring consistently on a specific route, time of day, or after specific shift lengths is strategically important and could prevent dozens of accidents if acted upon at the scheduling level. A cloud-connected fatigue monitoring platform would allow fleet managers to view real-time driver states across a fleet, receive escalation alerts when a driver enters DROWSY_ALARM state, and access historical fatigue analytics dashboards for route and schedule optimisation. No reviewed system addresses this fleet-level layer, which arguably represents the highest-value deployment scenario for the technology.

Gap 6 — Demographic Fairness and Generalisation: Computer vision systems trained predominantly on light-skinned, male, non-spectacle-wearing participants exhibit measurable accuracy degradation when deployed on other demographics. Darker skin tones reduce visible contrast for infrared cameras, making landmark localisation less reliable. Spectacle frames occlude portions of the eye contour used for EAR computation, inflating EAR variance. Beards and moustaches affect mouth landmark accuracy and therefore MAR-based yawn detection. Asian drivers' naturally narrower eye apertures mean that a universal threshold calibrated on a Western dataset systematically over-alerts. Only one reviewed paper (Muhammad et al., 2020) explicitly evaluated performance stratification by demographic subgroup. A fairness-aware evaluation framework and demographically balanced training datasets are necessary prerequisites for equitable large-scale deployment.

Gap 7 — No Standardised Public Benchmark with Real-World Ecological Validity: The reviewed papers evaluate on a heterogeneous collection of datasets: NTHU-DDD (simulator-based), YawDD (dashboard camera), UTA-RLDD (naturalistic driving), NITYMED (laboratory), and various custom-captured datasets, using inconsistent evaluation protocols and train-test splits. This makes quantitative comparison across papers unreliable — a system reporting 97% accuracy on NTHU-DDD and one reporting 95% on YawDD are not directly comparable. A standardised, diverse, ecologically valid benchmark dataset including in-vehicle recordings across multiple illumination conditions, ethnicities, age groups, corrective lens users, and road types — with agreed evaluation protocols and a public leaderboard — would dramatically accelerate reproducible progress in the field. The construction and curation of such a dataset is a significant undertaking that the community has not yet collectively addressed.

VIII. FUTURE RESEARCH DIRECTIONS

Building on the gap analysis, this section outlines seven high-priority research and engineering directions that would meaningfully advance the field toward deployment-grade real-world fatigue monitoring.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026
DOI: 10.17148/IJARCCE.2026.155115

A. *IoT-Enabled Embedded Deployment*

Migrating from laptop-based prototypes to automotive-grade embedded systems is the most immediate engineering priority. NVIDIA Jetson Nano and Xavier NX provide GPU-accelerated deep learning inference in a form factor suitable for vehicle dashboard mounting, with power consumption under 10 W. Combined with an automotive-grade camera (supporting wide dynamic range and global shutter), a Jetson-based system can run a full Tier 3 deep learning pipeline at 30 fps. Adding a WiFi/LTE module enables real-time cloud data synchronisation, while GPIO outputs connect to a vehicle CAN bus interface for physical alert actuation. The bill of materials for such a system can be held below Rs. 15,000 — representing an accessible cost point for commercial vehicle fleet retrofitting.

B. *GPS Geo-Fencing and Route Risk Profiling*

Integrating a GPS module enables spatial contextualisation of fatigue events. Specific road segments — monotonous highway stretches, post-midnight rural roads — are associated with elevated fatigue risk. A geo-fencing system can proactively lower the alert threshold and increase alert sensitivity when the vehicle enters a high-risk zone, and can flag repeated fatigue events at specific locations for fleet-level route risk analysis. Historical GPS-tagged fatigue log data, aggregated across a fleet, could generate a national-scale fatigue risk map for commercial vehicle routes.

C. *Transformer-Based Temporal Fatigue Modelling*

Vision Transformers (ViT) and temporal self-attention mechanisms represent a natural architectural evolution for fatigue detection. A transformer processing a sequence of EAR time series tokens over a 5-minute window — rather than a 2-second window — can detect the gradual downward trend of accumulating fatigue that threshold-based and short-window LSTM systems miss. Pre-training on large driving video datasets using self-supervised masked token prediction, then fine-tuning on labelled fatigue sequences, could achieve high accuracy with dramatically reduced labelled data requirements.

D. *Federated Learning for Privacy-Preserving Fleet Adaptation*

Personalised fatigue detection models that adapt to individual drivers are more accurate than universal models, but raw driving video from a fleet is sensitive personal data that cannot practically be centralised for training. Federated learning — training a shared global model using gradient updates computed locally on each vehicle without raw data leaving the vehicle — offers a privacy-preserving path to personalised adaptation. Each vehicle's embedded system fine-tunes the global model on local driver data and contributes encrypted gradient updates to a central aggregation server. The resulting personalised model improves per-driver accuracy without any individual's facial data ever leaving their vehicle.

E. *Multi-Modal Sensor Fusion with Non-Intrusive Physiological Sensors*

EEG headsets provide high-accuracy fatigue signals but require uncomfortable wearables. Steering wheel contact sensors — measuring skin conductance and heart rate variability through the driver's hands — provide physiological fatigue signals non-intrusively. Similarly, seat pressure sensors can detect the postural changes associated with fatigue onset. Fusing these unobtrusive physiological channels with vision-based features in a Kalman filter or attention-based fusion network can substantially improve accuracy over vision alone while avoiding the wearable barrier that has prevented EEG-based systems from achieving real-world adoption.

F. *Synthetic Data Augmentation for Demographic Fairness*

Generative Adversarial Networks (GANs) and diffusion models can synthesise photorealistic facial images with specified demographic attributes, lighting conditions, and accessory configurations (glasses, beards, hijabs, surgical masks). Training fatigue detection models on augmented datasets that systematically cover underrepresented demographic groups and challenging conditions is a practical path to closing the fairness gap without the enormous cost of collecting balanced real-world datasets across all demographic combinations.

G. *Standardised Benchmark and Evaluation Protocol*

The field urgently needs a standardised public benchmark analogous to ImageNet for visual recognition or GLUE for language understanding. An ideal fatigue detection benchmark would include in-vehicle recordings from at least 500 participants spanning five demographic groups, three illumination conditions (daylight, dusk/dawn, nighttime IR), four road types (urban, highway, rural, mountainous), and three accessory configurations (no glasses, clear glasses, sunglasses). Labels would include continuous KSS drowsiness ratings, PERCLOS values from a reference eye-tracker, and binary microsleep annotations. Establishing an agreed evaluation protocol — standardised train/validation/test split, agreed metric set including recall, F1, and alert latency — would enable meaningful progress tracking and reproducible comparison.

ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026
DOI: 10.17148/IJARCCE.2026.155115

IX. CONCLUSION

This paper presented a comprehensive survey of AI-Based Driver Fatigue Detection and Alert Systems, covering fifteen peer-reviewed studies and implementations from 2016 through 2024. The reviewed literature demonstrates that computer vision-based fatigue monitoring — particularly the Eye Aspect Ratio paradigm combined with yawn detection via Mouth Aspect Ratio — provides a structurally sound, cost-effective, and practically deployable foundation for real-time drowsiness detection. Tier 1 threshold-based systems implemented using Python, OpenCV, and Dlib have demonstrated reliable performance under controlled conditions at minimal hardware cost. Deep learning approaches — particularly MobileNet-based transfer learning and CNN+LSTM temporal architectures — have pushed accuracy above 97% on curated benchmarks. Multi-modal systems fusing EEG with visual features have achieved 98.2% accuracy but remain constrained by the impracticality of wearable physiological sensors for everyday driving.

The proposed four-tier taxonomy provides a structured framework for understanding the progression from basic threshold systems to fully integrated multi-modal platforms. This framework makes concrete a gap that individual papers naturally obscure: no existing system simultaneously addresses personalised adaptive thresholds, real-world lighting robustness, temporal fatigue trajectory modelling, vehicle CAN bus integration, cloud fleet monitoring, demographic fairness, and standardised evaluation. Tiers 1 through 3 are populated with validated implementations; Tier 4 remains aspirational.
The seven identified research gaps — absent personalised thresholds, insufficient lighting robustness, lack of temporal trajectory modelling, limited vehicle integration, absent fleet monitoring, demographic unfairness, and missing standardised benchmarks — are neither individually unsolvable nor theoretically intractable. The technology components required to address each gap exist in adjacent fields. The challenge is architectural integration: assembling validated building blocks into a deployment-grade platform that performs reliably across the full diversity of real-world driving conditions, vehicle types, and driver populations.

The societal stakes of progress in this space are high. Drowsy driving contributes to hundreds of thousands of deaths and millions of injuries annually worldwide. A widely deployed, reliable fatigue detection system integrated into the commercial vehicle fleets that carry the highest fatigue-related crash risk could reduce this toll meaningfully — potentially saving tens of thousands of lives per year. The current survey aims to provide a clear, rigorous map of where the remaining work lies, so that researchers and engineers can focus effort where it will have the greatest impact.

REFERENCES

[1]. 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, pp. 1–8.
[2]. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "Drowsiness Detection via Facial Landmark and MAR Analysis," in Proc. IEEE CVPR Workshop on Driver Monitoring, Salt Lake City, UT, USA, Jun. 2018.
[3]. C.-H. Weng, Y.-H. Lai, and S.-H. Lai, "Driver Drowsiness Detection via a Hierarchical Temporal Deep Belief Network," IEEE Trans. Intell. Transp. Syst., vol. 22, no. 5, pp. 2975–2984, May 2020.
[4]. R. Ghoddoosian, M. Galib, and V. Athitsos, "A Realistic Dataset and Baseline Temporal Model for Early Drowsiness Detection," in Proc. IEEE CVPR Workshop, Long Beach, CA, USA, Jun. 2019.
[5]. S. Park, F. Pan, S. Kang, and C. D. Yoo, "Driver Drowsiness Detection System Based on Feature Representation Learning Using Various Deep Networks," in Proc. ECCV Workshop on Action and Anticipation for Visual Learning, Amsterdam, Netherlands, Oct. 2016, pp. 154–164.
[6]. M. Dua, S. Singla, S. Raj, and A. K. Jangra, "Deep CNN Models-Based Ensemble Approach to Driver Drowsiness Detection," J. Ambient Intell. Humaniz. Comput., vol. 12, no. 5, pp. 5059–5071, May 2021.
[7]. M. Hossain and G. Muhammad, "Driver Fatigue Detection Using Lightweight CNN on NVIDIA Jetson Nano Embedded Hardware," IEEE Access, vol. 9, pp. 44843–44855, 2021.
[8]. Z. Zhao, Y. Liu, and X. Li, "Real-Time Driver Drowsiness Detection Using Multimodal Physiological and Visual Features," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 30, pp. 1234–1244, 2022.
[9]. B. Reddy, Y.-H. Kim, S. Yun, C. Seo, and J. Jang, "Real-Time Driver Drowsiness Detection for Embedded System Using Model Compression of Deep Neural Networks," in Proc. Int. Conf. Comput. Electron. Commun. Eng. (iCCECE), London, UK, Aug. 2017, pp. 125–131.
[10]. R. Jabbar, M. Shinoy, M. Kharbeche, K. Al-Khalifa, M. Krichen, and K. Barkaoui, "Driver Drowsiness Detection Model Using Convolutional Neural Network Techniques for Android Application," in Proc. IEEE Int. Conf. Comput. Inf. Technol. (CIT), Sydney, Australia, Dec. 2018.
[11]. A. Mittal, A. Kumar, M. Dhamija, and M. Kaur, "Head Movement-Based Driver Drowsiness Detection: A Review of State-of-the-Art Techniques," Comput. Electr. Eng., vol. 99, p. 107807, Apr. 2022.

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155115

[12]. S. Arefnezhad, J. Hamet, A. Eichberger, and A. Frühwirth, "Driver Drowsiness Estimation by Means of EEG Signals Using a New Convolutional Neural Network with Attention Mechanism," Sensors, vol. 22, no. 3, p. 1100, Jan. 2022.

[13]. J. Chen and G. Shi, "A Novel Driver Fatigue Detection System Based on YOLO v4 Face Detection and CAN Bus Integration for Smart Vehicles," in Proc. IEEE Intell. Veh. Symp. (IV), Nagoya, Japan, Jul. 2021, pp. 1421–1428.

[14]. K. Muhammad, S. Khan, J. Del Ser, and V. H. C. de Albuquerque, "Deep Learning for Multigrade Fatigue Detection with Focal Loss and Multi-Task Learning," IEEE Access, vol. 8, pp. 97013–97027, 2020.

[15]. M. Knapik and B. Cyganek, "Driver's Fatigue Recognition Through Facial Actions and Phase Congruency-Based Eye Texture Analysis," Neural Comput. Appl., vol. 31, no. 8, pp. 4283–4296, Aug. 2019.