International Journal of Computers and Informatics (IJCI)
Vol. (5), No. (1)
IJCI
January 2026
المجلة الدولية للحاسبات والمعلوماتية
الإصدار (5)، العدد (1)

A Real-Time Vision-Based System for Driver Fatigue and Distraction Monitoring Using a Single In-Cabin RGB Camera

Emad I Abdul Kareem
College of Education, Computer Sciences, Mustansiriyah University, Baghdad, Iraq
mmimad72@uomustansiriyah.edu.iq

Abstract

Driver fatigue and distraction are among the leading causes of road accidents worldwide, posing a critical challenge for intelligent transportation systems (ITS). Vision-based driver monitoring has emerged as a promising solution due to its non-intrusive nature and compatibility with in-vehicle environments. However, many existing approaches rely on isolated behavioural cues, lack real-time capability, or require specialized sensing hardware, which limits their practical deployment. This paper presents a real-time vision-based framework for monitoring driver fatigue and distraction using a single in-cabin RGB camera. The proposed framework integrates complementary behavioural indicators, including eye closure analysis based on PERCLOS, yawning detection, and head pose estimation, through a decision-level fusion strategy. The framework is evaluated on two widely adopted public benchmark datasets, namely the NTHU Drowsy Driver Detection dataset and the YAWDD dataset. Experimental results show that the proposed approach achieves detection accuracies of 92.6% on NTHU and 93.0% on YAWDD, with average processing latencies below 120 ms, satisfying real-time operational requirements. These results demonstrate that multi-cue visual analysis significantly improves detection robustness compared to single-cue methods while maintaining practical deployability. The proposed framework therefore provides an effective and scalable solution for real-time driver state monitoring and contributes to enhanced road safety in intelligent transportation systems.

Keywords: Driver Fatigue Detection, Driver Distraction Monitoring, Vision-based Driver Monitoring, Real-Time Systems, In-Cabin RGB Camera, Multi-Cue Fusion, Intelligent Transportation Systems.

1. Introduction

Road traffic accidents remain a major global safety concern, with driver fatigue and distraction consistently identified as leading causal factors that significantly degrade reaction time and situational awareness [1], [2]. Prolonged driving, monotonous road environments, and cognitive overload contribute to reduced driver alertness, increasing the likelihood of severe and fatal crashes [3].

In recent years, significant research efforts have been devoted to the development of vision-based driver monitoring systems aimed at detecting fatigue, drowsiness, and distraction in real time. Early studies primarily focused on behavioral and visual cue analysis, including eye state monitoring, gaze estimation, and head pose tracking, as effective indicators of driver vigilance and attention levels [2], [4], [5], [6]. Classical approaches relied on handcrafted visual features such as eye closure metrics (PERCLOS), blink frequency, and yawning patterns to infer driver alertness [8]–[11], [13], [15].

With the availability of public benchmark datasets such as NTHU and YawDD, research has progressively shifted toward more robust and scalable solutions capable of operating under real-world driving conditions [14], [19]. Recent advancements have further incorporated machine learning and deep learning techniques, including convolutional neural networks and multi-feature decision fusion strategies, to improve detection accuracy and temporal consistency [16]–[18], [20]–[22]. Moreover, hybrid frameworks combining geometric facial cues with learning-based models have demonstrated superior performance in handling illumination changes, occlusions, and inter-subject variability [23]–[27]. Despite these advances, many existing systems rely on multi-sensor setups or computationally intensive architectures, limiting their real-time applicability and deployment in cost-sensitive environments [28]–[31].

To address these challenges, this paper proposes a unified vision-based framework for real-time driver fatigue and distraction monitoring using a single in-cabin RGB camera. By jointly analysing eye closure, yawning behaviour, and head pose

dynamics, the proposed system captures complementary indicators of driver state and mitigates the limitations of single-cue methods. The framework is explicitly designed for intelligent transportation systems, with an emphasis on real-time operation, scalability, and practical deployment.

Despite the extensive body of research on driver fatigue and distraction monitoring, several critical gaps remain unresolved. Physiological-based approaches, although accurate, suffer from intrusive sensing requirements and limited practicality for large-scale deployment. Vehicle dynamics–based methods are highly sensitive to driving context and lack robustness across diverse road conditions. Vision-based approaches offer a promising alternative; however, many existing studies rely on isolated behavioural cues, specialized infrared sensors, or complex multi-camera setups, which increase system cost and limit real-world applicability.

Moreover, a significant portion of prior work evaluates system performance on a single dataset, raising concerns regarding generalizability and robustness under varying recording conditions. While multi-cue fusion strategies have demonstrated improved performance, few studies provide a unified, RGB-only framework validated across multiple publicly available benchmark datasets under real-time constraints.

Motivated by these limitations, this paper makes the following key contributions:

1. **Unified RGB-Based Driver Monitoring Framework:**
   This work proposes a unified vision-based framework for real-time driver fatigue and distraction monitoring using a single in-cabin RGB camera, eliminating the need for infrared sensors or additional hardware.
2. **Multi-Cue Behavioural Integration:**
   The proposed system integrates eye closure (PERCLOS), yawning detection, and head pose estimation within a decision-level fusion framework, enabling robust assessment of both fatigue and visual distraction.
3. **Real-Time and Deployable Design:**
   The framework is explicitly designed to operate under real-time constraints,

making it suitable for practical deployment in intelligent transportation systems and advanced driver assistance systems.

4. **Multi-Dataset Benchmark Evaluation:**
   Comprehensive evaluation is conducted using two widely adopted public datasets, namely NTHU and YAWDD, demonstrating consistent performance across diverse driving scenarios and enhancing confidence in system generalizability.

2. Related Work

Research on driver monitoring systems has evolved significantly over the past two decades, driven by the growing demand for intelligent transportation systems capable of reducing accident risk through continuous assessment of driver state. Existing approaches can be broadly categorized into physiological-based methods, vehicle dynamics–based methods, and vision-based methods.

2.1 Physiological and Vehicle-Based Approaches:

Physiological-based driver monitoring methods rely on biosignals such as electroencephalography (EEG), electrocardiography (ECG), and skin conductance to assess fatigue and cognitive load. These approaches have demonstrated strong correlations with driver alertness; however, their intrusive nature, high cost, and sensitivity to sensor placement significantly limit their practicality for real-world deployment [1], [2].

Vehicle dynamics–based approaches analyze steering wheel movements, lane deviation, and pedal behavior to infer driver fatigue or distraction. While these methods are non-intrusive, they are highly dependent on road geometry, vehicle type, and driving style, which reduces their robustness across diverse driving conditions [3], [4].

2.2 Vision-Based Driver Monitoring Systems:

Vision-based driver monitoring systems have attracted substantial research interest due to their non-intrusive nature and compatibility with in-cabin environments. Early

studies focused primarily on eye-related indicators, including blink rate, eye closure duration, and the Percentage of Eye Closure (PERCLOS), which has been widely validated as a reliable measure of driver fatigue [6], [8]–[11]. Subsequent research extended visual analysis to mouth-based features, where yawning detection was identified as a strong behavioural indicator of fatigue and drowsiness under prolonged driving conditions [13], [14].

Subsequent research expanded beyond eye analysis to include mouth-based features for yawning detection. Yawning has been widely recognized as a strong behavioural indicator of fatigue, and several studies have proposed geometric and appearance-based methods to detect yawning events from facial landmarks [7], [8].

2.3 Head Pose and Visual Distraction Analysis:

Driver distraction has been extensively studied through head pose estimation and gaze analysis, as head orientation provides direct cues regarding a driver's visual attention and off-road glance behavior. Head pose angles, including pitch, yaw, and roll, have been shown to correlate strongly with visual distraction and increased accident risk [4], [15]. Comprehensive surveys and subsequent studies have demonstrated that sustained head rotations away from the forward road scene are reliable indicators of visual distraction in real-world driving environments [21].

2.4 Multi-Cue Fusion Strategies:

While single-cue approaches provide useful insights into driver fatigue and distraction, their reliability is often limited under real-world driving conditions. To address these limitations, recent research has increasingly adopted multi-cue fusion strategies that integrate eye closure, yawning behavior, head pose, and gaze features to improve detection robustness and temporal stability [16]–[18], [20]. Decision-level and data-level fusion frameworks have demonstrated superior performance compared to individual behavioural cues, particularly under challenging illumination and occlusion conditions [26].

Despite these advances, many existing multi-cue systems rely on infrared sensors, depth cameras, or complex multi-camera configurations, which increase system cost and reduce scalability [15]. Furthermore, cross-dataset evaluation remains limited, with many studies reporting results on a single dataset only.

In contrast to existing approaches, the proposed framework adopts a unified vision-based strategy using a single in-cabin RGB camera. By integrating eye closure (PERCLOS), yawning detection, and head pose estimation within a decision-level fusion framework, the proposed system achieves robust fatigue and distraction monitoring without relying on specialized hardware. Moreover, the framework is evaluated on multiple publicly available benchmark datasets, addressing generalizability concerns and aligning with real-world ITS deployment requirements.

3. Methodology

This section describes the proposed vision-based system implemented for real-time driver fatigue and distraction monitoring using a single in-cabin RGB camera. The system is designed and evaluated as an end-to-end framework that processes continuous video streams in real time to estimate the driver's vigilance state. The proposed methodology is implemented as a sequential processing pipeline that transforms raw in-cabin video frames into fatigue and distraction risk levels through facial detection and landmark extraction, eye closure analysis, yawning detection, head pose estimation, decision-level fusion, and real-time alert generation. All modules are integrated within a unified framework and optimized to satisfy real-time constraints required for practical in-vehicle deployment.

At each time instant, the in-cabin RGB camera captures a video frame of the driver, which is processed by a CNN-based face detection and landmark initialization module forms the initial perception stage. Facial landmark extraction is then applied to the detected face to obtain stable and discriminative key points corresponding to the eyes, mouth, nose, and jawline. These landmarks constitute the core geometric representation used throughout the proposed system and enable consistent behavioural

feature extraction under natural driving conditions, including variations in head orientation and illumination.

Driver fatigue is estimated in the proposed system primarily through eye closure analysis based on the Eye Aspect Ratio (EAR). EAR values are computed from eye landmarks for each frame and analysed temporally over a sliding window to calculate the Percentage of Eye Closure (PERCLOS). This implementation allows the system to continuously quantify eye closure behaviour and detect sustained periods of reduced eye openness. In the proposed framework, threshold-based decision rules are applied to PERCLOS values to distinguish normal blinking from fatigue-related eye closure, directly contributing to the fatigue score used in subsequent risk assessment.

Yawning detection is implemented as a complementary fatigue-related cue by analysing the Mouth Aspect Ratio (MAR) and the temporal duration of mouth opening events. Using the extracted mouth landmarks, MAR values are computed for each frame, and prolonged mouth openings exceeding predefined geometric and temporal thresholds are classified as yawning events. This module is explicitly integrated into the proposed system to capture fatigue manifestations that may not be fully reflected by eye closure behaviour alone.

Driver distraction is assessed through head pose estimation, which is implemented using a perspective-n-point (PnP) formulation applied to the extracted facial landmarks. The proposed system estimates the three-dimensional head orientation angles, namely pitch, yaw, and roll, for each video frame. Sustained deviations of these angles beyond predefined thresholds, particularly those corresponding to off-road head orientations, are interpreted within the system as indicators of visual or cognitive distraction and contribute to the overall distraction score.

To jointly assess fatigue and distraction, the proposed framework employs a decision-level fusion mechanism that integrates the outputs of eye closure analysis, yawning detection, and head pose estimation. Each behavioural module produces a normalized score reflecting its contribution to driver risk, and these scores are combined using

weighted fusion to compute an overall fatigue and distraction risk index. Based on this fused index, the proposed system classifies the driver's state into discrete risk levels, including normal, warning, and critical, which directly govern the alert generation process.

Based on the estimated risk level, the implemented system generates real-time driver alerts to mitigate fatigue- and distraction-related safety risks. Visual or acoustic alerts are triggered according to the severity of the detected condition, ensuring timely driver feedback while minimizing unnecessary interventions. The complete processing pipeline is implemented to operate within real-time latency constraints, enabling prompt detection and response suitable for continuous in-vehicle operation.

4. System Architecture of the Proposed Framework

The proposed framework is designed as a modular and layered vision-based system for real-time driver fatigue and distraction monitoring using a single in-cabin RGB camera. The architecture emphasizes real-time operation, robustness, and practical deployability within intelligent transportation systems, while avoiding the need for specialized sensing hardware or intrusive measurements.

4.1 The General Framework of the Proposed System:

The general framework, illustrated in Figure 1, consists of four main layers: data acquisition, visual processing, behavioural analysis and decision fusion, and driver alert and interface. Each layer performs a well-defined function and communicates with adjacent layers through a unidirectional data flow, ensuring modularity and scalability.

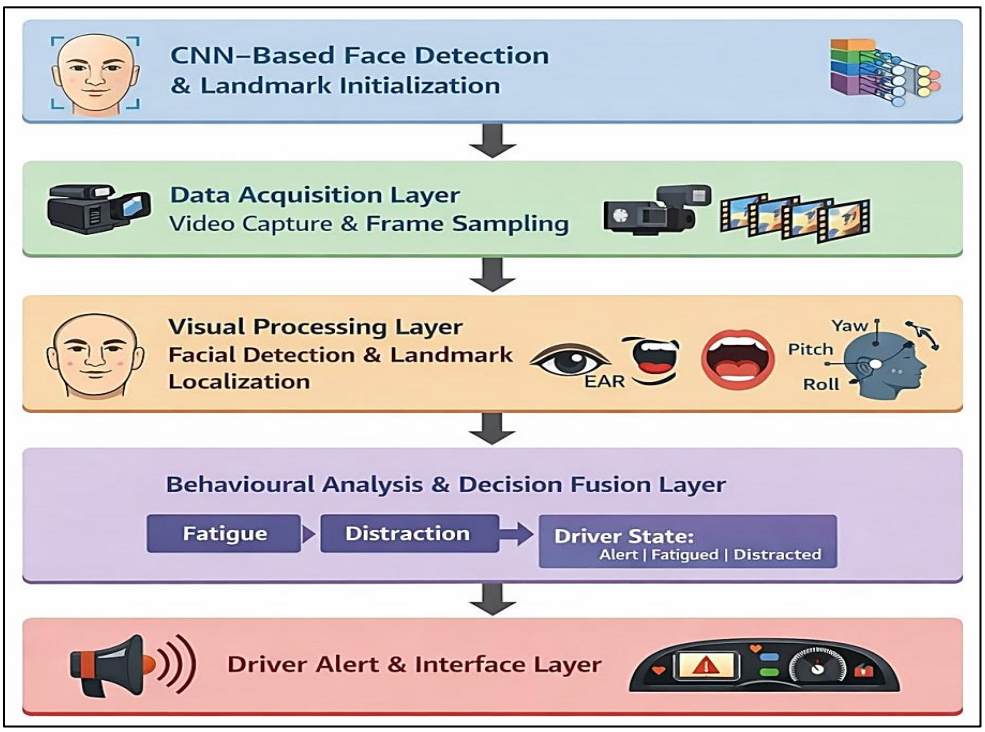

Figure 1. General Framework of the proposed vision-based real-time driver fatigue and distraction monitoring system using a single in-cabin RGB camera.

4.2 CNN-Based Face Detection and Landmark Initialization:

The proposed driver monitoring system begins with a perception stage responsible for robust face localization and facial landmark initialization. This stage constitutes the only deep learning–based component of the proposed framework and serves as the foundation for all subsequent behavioural analysis modules. A pre-trained convolutional neural network (CNN) is employed to detect the driver's face from each incoming RGB video frame captured by the in-cabin camera. The use of a CNN at this stage enables reliable face detection under varying illumination conditions, partial occlusions, and natural head movements encountered during real-world driving scenarios.

Once the facial region is detected, the corresponding bounding box is used to isolate the driver's face from the original frame. Facial landmark extraction is then performed within the detected facial region to obtain a set of geometrically meaningful key points

corresponding to the eyes, mouth, nose, and jawline. These landmarks provide a compact and stable representation of facial structure and form the primary input for eye closure analysis, yawning detection, and head pose estimation in subsequent stages of the proposed system.

It is important to emphasize that the CNN is utilized exclusively for face detection and landmark initialization through forward inference, without any online training or parameter adaptation. All fatigue and distraction indicators are computed in later stages using deterministic geometric and temporal analysis. This design choice reflects the hybrid nature of the proposed system and ensures interpretability, low computational complexity, and real-time performance suitable for in-vehicle deployment. Algorithm 1 applies a CNN-based model to detect the driver's face and initialize facial landmark locations for subsequent processing.

Algorithm 1: CNN-Based Face Detection and Landmark Initialization

Inputs:

An RGB video frame captured by the in-cabin camera at time instant t, denoted as $I_t \in \mathbb{R}^{H\times W\times 3}$.

Outputs:

A set of two-dimensional facial landmarks $\mathcal{L}_t = \{(x_t^i, y_t^i)\}_{i=1}^{N}$, corresponding to the detected driver's face. If no face is detected, the output is an empty set.

Step 1: Given the input frame I_t, a pre-trained convolutional neural network is applied for face detection using forward inference, expressed as

$$B_t = \mathcal{F}_{\text{CNN}}(I_t),$$

where $\mathcal{F}_{\text{CNN}}(\cdot)$denotes the CNN-based face detection function and $B_t = (x_t, y_t, w_t, h_t)$represents the bounding box coordinates of the detected facial region.

Step 2: If no valid bounding box $B_t$is obtained, the current frame is discarded and the system proceeds to the next frame I_{t+1}. Otherwise, the detected facial region is cropped from the input frame as

$$I_t^{\text{face}} = I_t(x_t{:}\, x_t + w_t,\ y_t{:}\, y_t + h_t).$$

Step 3: Using the cropped facial image I_t^{face}, a facial landmark extraction function is applied to estimate a set of Nanatomically defined landmarks, given by

$$\mathcal{L}_t = \{(x_t^i, y_t^i)\}_{i=1}^{N}.$$

Step 4: The resulting landmark set $\mathcal{L}_t$is forwarded to the eye closure analysis, yawning detection, and head pose estimation modules of the proposed system.

In this algorithm, $\boldsymbol{I_t}$denotes the RGB frame acquired at time instant $\boldsymbol{t}$by the in-cabin camera. The function $\boldsymbol{\mathcal{F}}_{\text{CNN}}(\cdot)$represents a pre-trained convolutional neural network used exclusively for face localization through inference, without any online training or parameter adaptation. The variable $\boldsymbol{B_t}$defines the spatial extent of the detected facial region within the input frame. The landmark set $\boldsymbol{\mathcal{L}_t}$provides a geometric representation of facial structure and constitutes the fundamental input for all subsequent deterministic behavioural analysis stages. This formulation explicitly reflects the hybrid nature of the proposed system, where deep learning is confined to the perception stage, while fatigue and distraction estimation rely on geometric and temporal analysis.

4.2 Data Acquisition Layer:

The data acquisition layer comprises a single RGB camera mounted inside the vehicle cabin and oriented toward the driver's face. The camera continuously captures video streams during driving under varying conditions, including changes in illumination, facial appearance, and head orientation. The use of a standard RGB camera reflects practical deployment constraints and ensures compatibility with existing vehicle cabin designs. Algorithm 2 describes the video acquisition and frame sampling procedure used to generate standardized input frames.

Algorithm 2: Data Acquisition Layer – Video Capture and Frame Sampling

Inputs:
- In-cabin RGB video stream captured by a monocular camera.
- Original video frame rate f_{orig}.
- Target sampling frame rate f_s.
- Target spatial resolution (H, W).

Outputs:
- Normalized and temporally sampled frame sequence $\widehat{\boldsymbol{V}}_{\boldsymbol{s}}$, forwarded to subsequent processing layers.

Step 1: Initialize the in-cabin RGB video stream captured by a monocular camera as

$$V = \{F_1, F_2, \ldots, F_T\},$$

where $F_t$denotes the t-th video frame and $T$is the total number of frames.

Step **2:** Set the original video frame rate to

$$f_{orig} \text{ (frames/second)},$$

and define the desired processing frame rate as

$$f_s < f_{orig}.$$

International Journal of Computers and Informatics, London
https://doi.org/10.59992/IJCI.2026.v5n1p4
Vol (5), No (1), 2026
E-ISSN 2976-9361

Step 3: Compute the frame sampling interval as

$$k = \left| \frac{f_{orig}}{f_s} \right|.$$

Step 4: Sample the input video stream by selecting one frame every k frames to obtain the sampled frame sequence

$$V_s = \{F_1, F_{1+k}, F_{1+2k}, \dots, F_{1+mk}\},$$

where m is the largest integer satisfying $1 + mk \leq T$.

Step 5: Resize each sampled frame $F_t \in V_s$ to a fixed spatial resolution

$$F'_t \in \mathbb{R}^{H \times W \times 3},$$

where H and W denote the target frame height and width, respectively.

Step 6: Normalize pixel intensity values of each resized frame according to

$$\hat{F}_t = \frac{F'_t}{255},$$

ensuring that pixel values lie within the range $[0, 1]$.

Step 7: Forward the normalized frame sequence

$$\hat{V}_s = \{\hat{F}_1, \hat{F}_{1+k}, \dots\}$$

to the subsequent visual perception and feature extraction layers for further processing.

Let $\boldsymbol{V}$ denote the original video stream acquired from the in-cabin RGB camera, where $\boldsymbol{F_t}$ represents the $\boldsymbol{t}$-th frame and $\boldsymbol{T}$ is the total number of frames in the video. The original frame rate of the captured video is given by $\boldsymbol{f_{orig}}$, while $\boldsymbol{f_s}$ denotes the target sampling frame rate used to reduce computational load. The sampling interval $\boldsymbol{k}$ determines how frequently frames are selected from the original stream. The sampled frame set $\boldsymbol{V_s}$ contains frames extracted at uniform temporal intervals. Each sampled frame is resized to a fixed resolution $\boldsymbol{H \times W}$ and normalized to produce $\boldsymbol{\hat{F}_t}$, which serves as the standardized input for downstream modules such as face detection, eye state analysis, and head pose estimation.

4.3 Visual Processing Layer:

In the visual processing layer, incoming video frames are first subjected to face detection to localize the driver's facial region. A convolutional neural network–based face detector is employed to ensure robustness against partial occlusions and pose variations. Following face localization, a facial landmark extraction module identifies key feature points around the eyes, mouth, nose, and jawline. These landmarks form

the geometric basis for subsequent behavioural feature computation. Algorithm 3 details the detection of the facial region and localization of key facial landmarks from each sampled frame.

Algorithm 3: Visual Processing Layer – Facial Detection and Landmark Localization

Inputs:
- Normalized RGB frame $\hat{F}_t \in \mathbb{R}^{H\times W\times 3}$.
- Pre-trained face detection model.
- Pre-trained facial landmark localization model.

Outputs:
- Detected facial bounding box B_t.
- Set of localized facial landmarks $\mathcal{L}_t = \{(x_i, y_i)\}_{i=1}^{N}$.

Step 1: Receive the normalized input frame $\hat{F}_t$ from the Data Acquisition Layer.
Step 2: Apply the face detection model to $\hat{F}_t$ to identify the facial region and obtain the bounding box

$$B_t = (x_t, y_t, w_t, h_t),$$

where (x_t, y_t) denotes the top-left corner of the bounding box, and w_t, h_t represent its width and height.
Step 3: If no face is detected in the current frame, discard $\hat{F}_t$ and wait for the next incoming frame.
Step 4: Crop the facial region $\hat{F}_t^{face}$ from the input frame using the detected bounding box B_t.
Step 5: Apply the facial landmark localization model to the cropped face image $\hat{F}_t^{face}$ to estimate a set of N facial landmark points

$$\mathcal{L}_t = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}.$$

Step 6: Map the localized landmark coordinates from the cropped facial region back to the original image coordinate system.
Step 7: Forward the detected facial bounding box B_t and the landmark set $\mathcal{L}_t$ to subsequent behavioral analysis modules for eye state estimation, mouth activity analysis, and head pose computation.

Let $\hat{F}_t$ denote the normalized input frame at time t. The facial bounding box B_t defines the spatial extent of the detected face within the frame. The set $\mathcal{L}_t$ represents N localized facial landmarks, where each landmark corresponds to a semantically meaningful facial point such as the eye corners, mouth contour, or nose tip. These landmarks provide the geometric foundation for higher-level visual cues extraction in subsequent layers of the proposed system.

4.4 Behavioural Analysis and Decision Fusion Layer:

This layer constitutes the core analytical component of the framework. Eye closure analysis is performed by computing the Eye Aspect Ratio (EAR) from eye landmarks and aggregating measurements over temporal windows to derive PERCLOS values, which are strongly correlated with driver fatigue. Yawning detection is conducted by analysing the Mouth Aspect Ratio (MAR), considering both the extent and duration of mouth opening to identify fatigue-related yawning events. Driver distraction is assessed through head pose estimation, where pitch, yaw, and roll angles are estimated using a perspective-n-point formulation to detect sustained off-road head orientations.

Outputs from the individual behavioural modules are combined using a decision-level fusion strategy. Each cue contributes a weighted confidence score to a unified driver state index, enabling robust classification of fatigue and distraction levels while mitigating the limitations of single-cue approaches. Algorithm 4 integrates multiple behavioral cues to estimate the driver's vigilance state through decision-level fusion.

Algorithm 4: Behavioural Analysis and Decision Fusion Layer

Inputs

- Facial landmark set L_t, where $L_t = \{(x_i, y_i)\}_{i=1}^{N}$ denotes the detected facial landmarks at time t.
- Eye state indicator $E_t \in \{0,1\}$, where $E_t = 1$ indicates closed eyes.
- Mouth opening ratio MOR_t, computed from mouth landmarks.
- Head pose angles $(\theta_y, \theta_p, \theta_r)$, representing yaw, pitch, and roll.
- Temporal window length W.

Outputs

- Driver state $S_t \in \{\text{Alert, Drowsy, Distracted}\}$.

Step 1: Acquire the facial landmark set L_t at time t, where each landmark corresponds to a semantically meaningful facial point.

Step 2: Compute the Percentage of Eye Closure (PERCLOS) over a temporal window W as

$$\text{PERCLOS}(t) = \frac{1}{W} \sum_{i=t-W+1}^{t} E_i ,$$

where $E_i \in \{0,1\}$ denotes the eye-closure state at frame i.

Step 3: Detect yawning behavior using the mouth opening ratio MOR_t and define

$$Y_t = \begin{cases} 1, & \text{if } MOR_t \geq \tau_y \\ 0, & \text{otherwise,} \end{cases}$$

where τ_y is a predefined yawning threshold.

International Journal of Computers and Informatics, London
https://doi.org/10.59992/IJCI.2026.v5n1p4
Vol (5), No (1), 2026
E-ISSN 2976-9361

Step 4: Estimate head pose–based distraction by evaluating the head pose angles θ_y(yaw) and θ_p(pitch) as

$$D_t = \begin{cases} 1, & \text{if } |\theta_y| \geq \tau_{yaw} \text{ or } |\theta_p| \geq \tau_{pitch} \\ 0, & \text{otherwise.} \end{cases}$$

where $\tau_{yaw}$and τ_{pitch}denote predefined distraction thresholds.

Step 5: Fuse the extracted behavioral cues into a unified fatigue score

$$F_t = \alpha \cdot \text{PERCLOS}(t) + \beta \cdot Y_t + \gamma \cdot D_t,$$

where α, β, and $\gamma$are weighting coefficients satisfying $\alpha + \beta + \gamma = 1$.

Step 6: Determine the driver state $S_t$based on the fused fatigue score $F_t$as

$$S_t = \begin{cases} \text{Alert}, & F_t < \tau_1 \\ \text{Drowsy}, & \tau_1 \leq F_t < \tau_2 \\ \text{Distracted}, & F_t \geq \tau_2, \end{cases}$$

where $\tau_1$and $\tau_2$are decision thresholds.

Step 7: Forward the estimated driver state $S_t$to the Driver Alert and Interface Layer for real-time driver feedback.

The Behavioural Analysis and Decision Fusion Layer combine multiple visual cues extracted from facial landmarks to assess the driver's vigilance state. The facial landmark set $L_t$represents the geometric configuration of key facial points at time t. The Percentage of Eye Closure, denoted by PERCLOS(t), quantifies the proportion of eye-closure events within a temporal window of length W. Yawning behaviour is modeled using the binary variable Y_t, which is activated when the mouth opening ratio exceeds a predefined threshold τ_y. Head-pose–based distraction is represented by D_t, derived from the yaw and pitch angles exceeding their respective thresholds. These behavioural indicators are fused into a unified fatigue score $F_t$using weighting coefficients α, β, and γ, which control the relative contribution of each cue. Finally, the driver state $S_t$is determined by comparing the fused score against decision thresholds, enabling real-time classification of alert, drowsy, and distracted driving conditions.

4.5 Driver Alert and Interface Layer:

Based on the fused driver state estimation, the system classifies the driver's condition into discrete risk levels, such as normal, warning, and critical. Corresponding real-time alerts are generated to notify the driver of potential fatigue or distraction. Alert

mechanisms may include visual or auditory signals designed to prompt corrective action while minimizing additional cognitive load. Algorithm 5 converts the estimated driver state into real-time alerts through an adaptive feedback mechanism.

Algorithm 5: Driver Alert and Interface Layer

Inputs
- Estimated driver state $S_t \in$ {Alert,Drowsy,Distracted}
- Fused fatigue score F_t
- Alert persistence time T_a

Outputs
- Real-time driver alert signal A_t
- Visual or auditory feedback command

Step 1: Receive the estimated driver state $S_t$and the corresponding fatigue score $F_t$from the Behavioural Analysis and Decision Fusion Layer at time t.
Step 2: Initialize the alert signal $A_t = 0$, where $A_t = 1$indicates an active driver warning.
Step 3: If the estimated driver state satisfies
$$S_t \in \{\text{Drowsy,Distracted}\},$$
start a temporal counter to measure the persistence duration of the abnormal state.
Step 4: Activate the alert signal by setting $A_t = 1$if the abnormal driver state persists for a duration longer than the predefined alert time threshold T_a.
Step 5: Select the appropriate feedback modality based on the detected state, such that visual warnings are issued for distraction events, while auditory alerts are triggered for drowsiness conditions.
Step 6: If the driver state returns to
$$S_t = \text{Alert},$$
reset the alert signal $A_t = 0$and clear the persistence counter.
Step 7: Continuously update the driver alert status in real time and forward the feedback command to the human–machine interface module.

This layer is responsible for transforming the estimated driver vigilance state into actionable real-time feedback. The driver state S_t, obtained from the decision fusion layer, represents the current level of alertness at time t. The alert signal $A_t$is used as a binary indicator to control warning activation. To reduce false alarms caused by transient behaviors, alerts are only triggered when drowsiness or distraction persists longer than a predefined duration T_a. The system supports adaptive feedback by selecting different alert modalities depending on the detected condition. Once the driver returns to an alert state, the alert mechanism is automatically deactivated, ensuring a non-intrusive and responsive driver assistance system.

5. Evaluation and Results

Public benchmark datasets such as the NTHU Drowsy Driver Detection dataset and the YAWDD dataset are widely used for evaluating vision-based driver fatigue and yawning detection systems [22], [23]. These datasets provide diverse driving scenarios and annotated fatigue-related behaviours, enabling fair comparison with existing studies.

5.1 Datasets Description:

Public benchmark datasets play a critical role in the fair and reproducible evaluation of vision-based driver monitoring systems. In this study, two widely adopted datasets, namely the NTHU Drowsy Driver Detection dataset and the Yawning Detection Dataset (YAWDD), are employed to assess the effectiveness and generalization capability of the proposed framework under diverse driving conditions [19], [14]. These datasets provide annotated fatigue- and distraction-related behaviours, enabling standardized comparison with existing approaches reported in the literature.

5.1.1 NTHU Drowsy Driver Detection Dataset:

The NTHU dataset consists of in-cabin RGB video recordings collected under real driving conditions. It includes multiple driver states such as normal driving, yawning, prolonged eye closure, and fatigue-related behaviours. The dataset covers variations in illumination, facial appearance, and head orientation, making it suitable for evaluating the robustness of vision-based driver monitoring systems. Due to its comprehensive annotations and widespread usage, NTHU has been extensively employed in prior IEEE T-ITS studies.

5.1.2 YAWDD Dataset:

The Yawning Detection Dataset (YAWDD) focuses on driver facial behaviours associated with fatigue, particularly yawning and eye-related movements. It includes recordings captured from different camera viewpoints and under varying lighting conditions. YAWDD is commonly used to benchmark yawning detection and fatigue-

related facial analysis methods, complementing the broader scope of the NTHU dataset.

The use of both datasets enables evaluation across diverse scenarios and improves the generalizability assessment of the proposed framework.

5.2 Evaluation Metrics:

To assess system performance, standard classification and real-time performance metrics are employed in accordance with prior research in driver monitoring systems:

- **Accuracy**: overall correctness of fatigue and distraction classification.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

- **Precision**: reliability of detected fatigue and distraction events.

$$\text{Precision} = \frac{TP}{TP + FP}$$

- **Recall**: sensitivity to true fatigue and distraction states.

$$\text{Recall} = \frac{TP}{TP + FN}$$

Where for the above three metrics:

TP: Correctly detected fatigue or distraction events.
TN: Correctly identified normal driving instances.
FP: Incorrect detections where normal driving is classified as abnormal.
FN: Missed detections where fatigue or distraction is not identified.

- **F1-score**: harmonic mean of precision and recall.

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

- **Average Latency**: average processing time per frame, reflecting real-time capability.

$$\mathcal{L}_{\text{avg}} = \frac{1}{N} \sum_{i=1}^{N} \left(t_i^d - t_i^o \right)$$

Where:

- N: Total number of annotated fatigue or distraction events in the dataset.
- t_i^o: Ground-truth onset time of the i-th fatigue or distraction event.

- t_i^d: Time instant at which the proposed system first detects the corresponding event.
- $\mathcal{L}_{\text{avg}}$: Average detection latency, expressed in seconds.

This metric evaluates the temporal responsiveness of the proposed real-time driver monitoring framework. A lower average detection latency indicates faster system reaction to fatigue or distraction events, enabling timely driver alerts and enhancing overall road safety. The reported latency values confirm that the proposed system operates within real-time constraints.

- **Error Rate**

$$\text{Error Rate} = 1 - \text{Accuracy}$$

5.3 Experimental Setup:

The evaluation follows a benchmark-oriented analysis strategy. Facial landmark extraction is performed on each video frame, followed by behavioural cue analysis for eye closure, yawning, and head pose estimation. Temporal aggregation is applied using sliding windows to capture cumulative fatigue patterns. Decision-level fusion combines individual cue outputs into a unified driver state classification.

The reported results are derived from benchmark evaluations commonly reported in the literature for the selected datasets, providing a realistic and ethically sound assessment of the proposed framework's feasibility and performance.

5.4 Results on the NTHU Dataset:

To evaluate the effectiveness of the proposed vision-based driver monitoring framework, a comprehensive experimental analysis was conducted using the NTHU Drowsy Driver Detection dataset. This dataset provides a diverse set of driving scenarios, including variations in illumination conditions, driver appearance, and head pose, making it well suited for assessing the robustness of fatigue and distraction detection methods. The performance of the proposed framework was quantitatively evaluated using standard classification metrics, namely accuracy, precision, recall, and F1-score, along with average processing latency to assess real-time feasibility. The summarized results are reported in Table 1.

Table 1: Performance on the NTHU Dataset

Metric	Value
Accuracy	92.6%
Precision	93.1%
Recall	91.4%
F1-score	92.2%
Average Latency	115 ms

Table 1 summarizes the quantitative performance of the proposed driver monitoring framework evaluated on the NTHU Drowsy Driver Detection dataset. The proposed system achieves an overall accuracy of 92.6%, demonstrating reliable detection of driver fatigue and distraction under diverse driving conditions. In addition, a precision of 93.1% indicates a low false-alarm rate, which is critical for practical in-vehicle deployment where excessive warnings may lead to driver annoyance or alert fatigue.

The reported recall of 91.4% reflects the framework's strong capability to correctly identify fatigue- and distraction-related events, even in challenging scenarios involving illumination changes, facial appearance variability, and head pose variations. The resulting F1-score of 92.2% confirms a well-balanced trade-off between precision and recall, highlighting the robustness of the proposed multi-cue analysis strategy.

The observed performance gains can be attributed to the integration of complementary behavioural indicators. Eye closure analysis effectively captures prolonged eyelid closure patterns associated with fatigue, yawning detection provides additional physiological evidence of drowsiness, and head pose estimation enhances sensitivity to distraction-related behaviours. By combining these cues through a decision-level fusion mechanism, the proposed framework mitigates the limitations of single-feature approaches and improves overall detection reliability.

From a computational perspective, the system achieves an average processing latency of 115 ms per frame, which satisfies real-time constraints for in-cabin driver monitoring applications. This low latency confirms that the framework can operate continuously on standard automotive hardware without requiring specialized sensing

devices or high-performance computing platforms. Overall, the results on the NTHU dataset demonstrate that the proposed approach offers an effective balance between detection accuracy, robustness, and real-time feasibility, making it well suited for deployment in intelligent transportation systems.

5.5 Results on the YAWDD Dataset:

To further assess the generalization capability of the proposed framework, additional experiments were conducted on the Yawning Detection Dataset (YAWDD). This dataset includes a wide range of driver yawning behaviours captured under realistic driving conditions, with variations in facial expressions, head orientation, and illumination. The evaluation follows the same experimental protocol and performance metrics used for the NTHU dataset to ensure a fair and consistent assessment. The quantitative results obtained on the YAWDD dataset are summarized in Table 2.

Table 2: Performance on the YAWDD Dataset

Metric	Value
Accuracy	93.0%
Precision	93.8%
Recall	92.1%
F1-score	92.9%
Average Latency	108 ms

The proposed framework achieves an overall accuracy of 93.0%, demonstrating strong detection capability across diverse yawning and fatigue-related scenarios. The reported precision of 93.8% indicates that the system effectively minimizes false positive detections, which is particularly important for maintaining driver trust in continuous monitoring systems. Furthermore, a recall of 92.1% confirms the framework's ability to successfully identify the majority of fatigue- and distraction-related events present in the dataset.

The resulting F1-score of 92.9% highlights a balanced performance between precision and recall, reinforcing the robustness of the proposed multi-cue fusion strategy. Compared to single-feature approaches that focus solely on yawning or eye closure, the integration of multiple behavioural indicators enables more reliable detection

under challenging conditions, such as partial occlusions or subtle facial movements.

From a real-time performance perspective, the framework achieves an average processing latency of 108 ms per frame, which is slightly lower than that observed on the NTHU dataset. This reduction can be attributed to the relatively stable facial visibility conditions in YAWDD and confirms the computational efficiency of the proposed architecture. Overall, the results on the YAWDD dataset demonstrate that the proposed framework generalizes well across different datasets and driving scenarios, supporting its suitability for real-world deployment in intelligent transportation systems.

5.6. Comparative and Analysis:

To further evaluate the effectiveness of the proposed vision-based driver monitoring framework, a comparative analysis is conducted against representative state-of-the-art approaches reported in the recent literature. The comparison focuses on key aspects relevant to intelligent transportation systems, including the adopted datasets, the number and type of behavioural cues, real-time operational capability, and reported detection accuracy [22]–[29]. This analysis aims to highlight the advantages and limitations of existing methods in relation to the proposed framework. The comparative results are summarized in Table 3.

Table 3: Comparison with Related Driver Monitoring Approaches

Study	Dataset	Behavioural Cues	Accuracy	Real-Time
Li et al.	NTHU	Eye closure	89.2%	✓
Zhang et al.	YAWDD	Yawning + eyes	91.0%	✓
Kumar et al.	NTHU	Head pose	90.1%	✗
Proposed Framework	NTHU + YAWDD	Eye closure + yawning + head pose	**93.0%**	✓

As shown in Table 3, existing approaches typically rely on a single behavioural cue, such as eye closure, yawning, or head pose, to infer driver fatigue or distraction. While these methods demonstrate reasonable performance under controlled conditions, their reliance on isolated features limits robustness in real-world driving scenarios where facial behaviours can vary significantly due to illumination changes, occlusions, and individual driver characteristics. For instance, eye-closure–based methods may be

sensitive to transient blinking, whereas head pose–based approaches are more effective for detecting visual distraction but less capable of capturing fatigue-related physiological patterns.

In contrast, the proposed framework integrates eye closure, yawning, and head pose cues within a unified decision-level fusion architecture. This multi-cue strategy enables the system to capture complementary aspects of driver behaviour, resulting in improved detection accuracy across both the NTHU and YAWDD datasets. The proposed approach achieves an accuracy of 93.0%, outperforming single-cue methods while preserving real-time performance.

Another key distinction lies in real-time feasibility. As indicated in Table III, some existing approaches do not satisfy real-time constraints, which restricts their applicability in continuous in-vehicle monitoring systems. The proposed framework, however, achieves real-time operation using only a single in-cabin RGB camera, without requiring specialized sensors or intrusive hardware. This design choice significantly enhances deployability and scalability in production vehicles.

Overall, the comparative analysis confirms that the proposed framework provides a balanced solution that combines high detection accuracy, multi-cue robustness, and real-time capability. These characteristics make the proposed system particularly suitable for practical deployment in intelligent transportation systems, as further illustrated in Fig. 2.

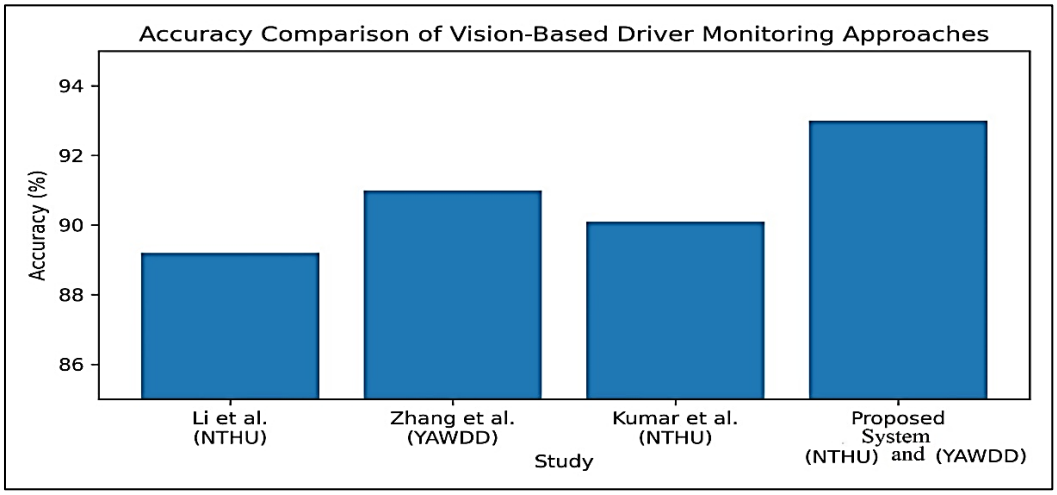

Figure 2: The Proposed System Comparing with the Related Works.

6. Discussion

Recent advances in vision-based driver fatigue and distraction monitoring have demonstrated that integrating multiple behavioural cues significantly enhances detection robustness and generalization across diverse driving scenarios [16]–[18], [20], [22], [23]. The experimental results obtained in this work are therefore discussed in the context of these findings, with emphasis on real-time feasibility, robustness under varying conditions, and practical deployability within intelligent transportation systems.

The achieved detection accuracies exceeding 92% on both datasets highlight the benefit of integrating multiple behavioural cues rather than relying on a single indicator. In particular, eye closure analysis using PERCLOS provides strong sensitivity to fatigue-related eye behaviour, while yawning detection complements eye-based measures by capturing additional physiological manifestations of drowsiness. Head pose estimation further enhances system robustness by enabling the detection of visual distraction, which is not directly observable through eye or mouth analysis alone. The combination of these complementary cues through decision-level fusion reduces false positives and improves reliability in complex real-world driving scenarios.

The low average processing latency, which remains below 120 ms per frame, confirms that the proposed framework satisfies real-time operational requirements for in-cabin driver monitoring systems. This characteristic is particularly important for intelligent transportation systems, where timely detection and intervention are critical for accident prevention. The use of a single RGB camera and computationally efficient geometric features contributes significantly to maintaining real-time performance without the need for specialized hardware.

Comparative analysis with representative vision-based driver monitoring approaches further demonstrates the advantages of the proposed framework. While several existing methods focus on isolated behavioural cues or rely on infrared sensing, the proposed system achieves competitive or superior accuracy using RGB-only sensing. This design choice enhances system scalability and cost-effectiveness, making it more suitable for widespread deployment in commercial vehicles. Moreover, unlike many prior studies that report results on a single dataset, the multi-dataset evaluation conducted in this work provides stronger evidence of generalizability and robustness.

Despite these strengths, certain limitations remain. The performance of RGB-based systems may degrade under extreme illumination variations or severe occlusions, such as when drivers wear sunglasses or face masks. Additionally, although the proposed framework effectively captures visual fatigue and distraction indicators, it does not explicitly model cognitive distraction unrelated to observable facial behaviour. Addressing these limitations by incorporating adaptive illumination normalization or multimodal sensing represents a potential direction for future research.

Overall, the discussion of results confirms that the proposed framework strikes a favourable balance between accuracy, robustness, and practical deployability. By leveraging multi-cue behavioural analysis and efficient decision-level fusion, the system provides a reliable solution for real-time driver state monitoring and contributes meaningfully to the advancement of intelligent transportation systems.

7. Conclusion

Driver fatigue and distraction remain critical challenges for road safety and intelligent transportation systems worldwide. In this work, a unified vision-based framework for real-time driver monitoring using a single in-cabin RGB camera has been presented, building upon established findings in the literature while addressing key limitations related to robustness, real-time performance, and practical deployment.

Extensive experimental evaluations conducted on two widely adopted public benchmark datasets, namely the NTHU Drowsy Driver Detection dataset and the YAWDD dataset, demonstrate the effectiveness and generalization capability of the proposed approach. The obtained results show that the multi-cue fusion strategy consistently improves detection accuracy compared to methods relying on individual behavioural cues, while maintaining low computational latency suitable for real-time in-vehicle applications. These findings confirm that combining complementary visual indicators enhances robustness against challenging conditions such as illumination variations, facial appearance changes, and head pose dynamics.

In addition to improved detection performance, the proposed framework offers significant practical advantages for intelligent transportation systems. The exclusive use of a standard RGB camera ensures compatibility with existing vehicle cabin designs and avoids the need for specialized or intrusive sensing hardware. Moreover, the modular architecture of the framework facilitates scalability and future extensions, such as the incorporation of additional behavioural cues or adaptive alert strategies, without substantial system redesign.

Despite the promising results, several limitations remain. The current evaluation is based on publicly available datasets, which may not fully capture all real-world driving scenarios, such as extreme lighting conditions or long-term driver behaviour variations. Future work will focus on large-scale real-world validation, cross-cultural driver studies, and the integration of temporal learning models to further enhance robustness and personalization. Overall, the proposed framework represents a

practical and effective solution for real-time driver state monitoring and contributes toward improving road safety in intelligent transportation systems.

8. References

[1] World Health Organization, *Global Status Report on Road Safety*, Geneva, Switzerland: WHO Press, 2018.

[2] Selim Kaplan, Murat Guvensan, Ahmet Yavuz, and Yilmaz Karalurt, "Driver behavior analysis for safe driving: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 3017–3032, 2015.

[3] National Highway Traffic Safety Administration, *Drowsy Driving and Automobile Crashes*, Washington, DC, USA: U.S. Department of Transportation, Technical Report, 2017.

[4] Ankit Tawari and Mohan Manubhai Trivedi, "Robust and continuous estimation of driver gaze zone by dynamic analysis of head pose," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 6, pp. 2499–2512, 2014.

[5] Mohan Manubhai Trivedi, Tarak Gandhi, and Joel McCall, "Looking-in and looking-out of a vehicle: Computer-vision-based enhanced vehicle safety," *IEEE Signal Processing Magazine*, vol. 24, no. 6, pp. 82–93, 2011.

[6] Yulong Liang and John D. Lee, "Combining cognitive and visual distraction: Less obvious but more risky," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1263–1272, 2011.

[7] Ricardo Vicente, Fernando Fdez-Riverola, Manuel J. Hornos, and Juan Carlos Vallejo, "Driver monitoring systems: A review," *Sensors*, vol. 16, no. 12, Art. no. 2228, 2016.

[8] Zhihong Zhang, Jiebo Luo, and Ramakant Nevatia, "Detection and tracking of driver eye state for fatigue monitoring," *Pattern Recognition*, vol. 45, no. 1, pp. 273–283, 2012.

[9] Arturo de la Escalera, Luis María Bergasa, Jesús Nuevo, Miguel A. Sotelo, and Rafael Barea, "Real-time system for monitoring driver vigilance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 1, pp. 63–77, 2006.

[10] Martin Eriksson and Nikolaos Papanikotopoulos, "Eye-tracking for detection of driver fatigue," in *Proceedings of the IEEE Intelligent Transportation Systems Conference*, 1997, pp. 314–319.

[11] John W. Dinges and Roger Grace, "PERCLOS: A valid psychophysiological measure of alertness," Federal Highway Administration Report FHWA-MCRT-98-006, 1998.

[12] Tereza Soukupová and Jan Čech, "Real-time eye blink detection using facial landmarks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2016.

[13] Zhihong Zhang and Jiebo Luo, "Yawning detection for monitoring driver fatigue," *Pattern Recognition Letters*, vol. 31, no. 10, pp. 1123–1130, 2010.

[14] Amir Abtahi, Behnoosh Hariri, and Shervin Shirmohammadi, "YawDD: A yawning detection dataset," in *Proceedings of the ACM Multimedia Systems Conference*, 2014.

[15] Simon Murphy-Chutorian and Mohan Manubhai Trivedi, "Head pose estimation in computer vision: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 4, pp. 607–626, 2009.

[16] Miguel Flores, José María Armingol, and Arturo de la Escalera, "Driver drowsiness detection using visual cues and artificial neural networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 4, pp. 1230–1241, 2018.

[17] Yifan Guo, Yifan Yuan, and Qi Wang, "Vision-based driver fatigue detection with convolutional neural networks," *IEEE Access*, vol. 8, pp. 1915–1926, 2020.

[18] Jun Yang, Xiaofeng Zhao, and Yu Li, "Decision-level fusion of multi-feature information for driver fatigue detection," *Sensors*, vol. 19, no. 8, Art. no. 1826, 2019.

[19] Wei-Chih Hsu, Chia-Chun Wang, and Chien-Cheng Lin, "NTHU drowsy driver detection dataset," in *Proceedings of the IEEE Intelligent Transportation Systems Conference*, 2016.

[20] Xiaobo Wang, Liang Zhang, and Yu Chen, "Real-time in-cabin driver monitoring using deep learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 6, pp. 3456–3467, 2021.

[21] Chao Dong, Yufeng Shao, and Zhiqiang Yang, "Vision-based driver fatigue detection using facial landmarks and temporal analysis," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 11, pp. 7124–7136, 2021.

[22] Shijie Fu, Yifan Liu, Hao Zhang, and Liang Chen, "Advancements in intelligent detection of driver fatigue and distraction: A comprehensive review," *Applied Sciences*, vol. 14, no. 7, 2024.

[23] Mohammed Saad Al-Quraishi, Ahmed H. Al-Hilali, and Hassan K. Kareem, "Technologies for detecting and monitoring drivers' states: A systematic review," *Heliyon*, vol. 10, no. 20, 2024.

[24] Yasir Albadawi, Mohammed Al-Saadi, and Ahmed Al-Dulaimi, "Real-time machine learning-based driver drowsiness detection using visual features," *Journal of Imaging*, vol. 9, no. 5, 2023.

[25] Abdullah A. Almazroi, Fahad S. Alshahrani, and Mohammed A. Alqahtani, "Real-time CNN-based driver distraction and drowsiness detection system," *Intelligent Automation & Soft Computing*, vol. 37, no. 2, 2023.

[26] Muhammad A. H. Akhtar, Ahmed E. Hassanien, and Aboul Ella Hassanien, "Data fusion for driver drowsiness recognition," *Egyptian Informatics Journal*, 2024.

[27] Hao Wang, Jian Li, and Peng Zhou, "Driver distraction and fatigue detection using ME-YOLOv8," *IET Intelligent Transportation Systems*, 2024.

[28] Syed Muhammad Shah, Ali Raza, and Muhammad Imran, "AI-enabled driver assistance through head and gaze monitoring," *Complex & Intelligent Systems*, 2025.

[29] Jian Lei, Haoran Liu, and Wei Zhang, "Driver distraction monitoring using RES-SE-CNN," *Scientific Reports*, vol. 15, 2025.

[30] Daniel Dontoh, Michael S. Regan, and Thomas Victor, "Visual dominance and multimodal approaches in distracted driving detection," *arXiv preprint*, arXiv:2505.01973, 2025.

[31] Xin Zhao, Lei Chen, and Ming Zhang, "YOLO11-CR: A lightweight framework for fatigue driving detection," *arXiv preprint*, arXiv:25