Indian Journal of Computer Science and Technology
https://www.doi.org/10.59256/indjcst.20260501039
Volume 5, Issue1 (January-April 2026), PP: 273-279.
www.indjcst.com
ISSN No: 2583-5300

An Adam W-Optimized Vision Transformer Framework with Back propagation Training for Driver Drowsiness Detection for Smart Vehicular Safety

Mani Bakkiaraj P [1], Dr. K. Karuppasamy[2], Thenmalar R[3], Sinduja R[4]

[1] PG student, Department of Computer Science & Engineering, RVS College Of Engineering & Technology, Coimbatore, Tamilnadu, India.

[2] Head of Department, Professor, Department of Computer Science & Engineering, RVS College Of Engineering & Technology, Coimbatore, Tamilnadu, India.

[3] Project Guide, Assistant Professor, Department of Computer Science & Engineering, RVS College Of Engineering & Technology, Coimbatore, Tamilnadu, India.

[4]Project Coordinator, Assistant Professor, Department of Computer Science & Engineering, RVS College Of Engineering & Technology, Coimbatore, Tamilnadu, India.

To Cite this Article: *Mani Bakkiaraj P [1], Dr. K. Karuppasamy[2], Thenmalar R[3], Sinduja R[4], "An Adam W-Optimized Vision Transformer Framework with Back propagation Training for Driver Drowsiness Detection for Smart Vehicular Safety", Indian Journal of Computer Science and Technology, Volume 05, Issue 01 (January-April 2026), PP: 273-279.*

Abstract: *When drivers suffer from fatigue-induced cognitive impairment, the result is a tragic surge in worldwide traffic accidents that claim lives and drain vital economic resources. High-speed travel leaves no room for the slow reaction times that come with exhaustion, yet many operators don't realize they are impaired until it is too late. Because physical symptoms often lag behind actual cognitive decline, reactive safety systems simply aren't enough to prevent crashes. To truly move the needle on road safety, we must find ways to flag the very first physiological and behavioural warning signs. This paper focuses on catching fatigue in its earliest stages, allowing for the kind of proactive intervention that can stop an accident before it ever starts. Current methods, like analysing vehicle data often have reliability problems. Can be affected by environmental noise. Vehicle operators fatigue is a concern and fatigue detection can help prevent accidents. Road safety protocols need to be enhanced to reduce the number of accidents caused by fatigue Conventional detection methods have limitations and new approaches are needed to address this issue. To mitigate these deficiencies, this research proposes a real-time Driver Drowsiness Detection System leveraging optical sensing and deep representational learning architectures. The framework continuously acquires facial telemetry via a video stream, isolating ocular regions through facial landmark regression. An Adaptive Eye Characteristic Ratio (AECR) algorithm is employed to quantify prolonged ocular closure. A cardinal indicator of fatigue. Furthermore, a Vision Transformer (ViT) model analyses global spatial dependencies within the ocular features to categorize the operator's alertness state. Upon detection of somnolence, the system initiates immediate multi-modal alerts. A relational database backend logs temporal fatigue metrics for longitudinal performance analytics. Empirical validation under diverse illumination and pose conditions yielded a classification accuracy of approximately 92% with a false positive rate of 5% and sub-second inference latency. This cost-efficient, non-intrusive solution addresses the limitations of legacy systems and offers scalability for commercial fleets and private transport. Future iterations may integrate yawing analysis and infrared imaging for nocturnal efficacy.*

Key Words: *ROI Extraction, Optical Acquisition, AECR Algorithm, Vision Transformer (ViT), Optimization Engine, Alert Logic.*

I. INTRODUCTION

Overview-The mitigation of fatigue-related vehicular accidents is a critical objective in the domain of intelligent transportation systems. Somnolence severely degrades an operator's cognitive faculties, specifically affecting reaction latency and decision-making precision. Therefore, continuous vigilance monitoring is essential, particularly for long-haul logistics and nocturnal operations. The proposed architecture utilizes a computer vision-based methodology to analyse driver behaviour via live optical feeds. By quantifying ocular kinematics— specifically blink frequency and closure duration—the system accurately infers alertness levels. This non-contact approach eliminates the need for invasive bio-sensors, ensuring high user compliance. To maximize detection fidelity, the system integrates the **Adaptive Eye Characteristic Ratio (AECR)** for dynamic thresholding and a **Vision Transformer (ViT)** for extracting high-dimensional spatial features. Optimized via **AdamW** and trained using back propagation, the model ensures robust performance, triggering proactive alerts to avert potential collisions.

Problem Statement Driver fatigue represents a pervasive global safety hazard. Extended operational durations, circadian rhythm disruptions, and cumulative cognitive load significantly impair driver alertness. This physiological state precipitates delayed motor responses and loss of vehicular control, often culminating in catastrophic accidents. The insidious nature of fatigue, characterized by gradual onset and episodes of micro sleep, renders self-diagnosis unreliable. While legacy detection systems exist, they possess inherent flaws: vehicle-based heuristics (e.g., lane deviation) are reactive and environmentally dependent, while physiological sensors (e.g., EEG) are physically intrusive and impractical for widespread deployment. Existing vision-based solutions often rely on static thresholds that fail to account for inter-subject anatomical variability and utilize architectures that struggle with complex spatial dependencies. This project addresses these lacunae by engineering an adaptive, transformer-based system capable of real-time, personalized fatigue detection.

Objective The primary objective is to architect and deploy a real-time, vision-based safety system capable of monitoring driver vigilance to enhance road safety. Specific technical goals include:

- To implement continuous video acquisition for real-time ocular behavioural analysis.
- To quantify fatigue metrics utilizing blink frequency and closure duration analysis.
- To develop an **Adaptive Eye Characteristic Ratio (AECR)** algorithm that dynamically calibrates to individual user baselines.
- To deploy a **Vision Transformer (ViT)** for superior feature extraction and state classification.
- To optimize the deep learning model utilizing back propagation and the **AdamW** optimizer.
- To integrate a low-latency alert mechanism for immediate hazard mitigation.

Scope

- **System Architecture:** Development of an AI-driven, real-time monitoring framework utilizing standard optical hardware.
- **Biometric Analysis:** Implementation of AECR to analyse ocular temporal patterns, including micro sleeps and blink rate variability.
- **Deep Learning Integration:** Utilization of Transformer architectures (ViT) to supersede traditional CNNs in feature spatialization.
- **Alert Mechanism:** Design of a multi-modal feedback loop (auditory/visual) for immediate driver re-engagement.
- **Adaptive Heuristics:** Personalization of detection thresholds to accommodate diverse facial anatomies.
- **Optimization:** Application of advanced training protocols (AdamW) to ensure model generalization across varying environmental conditions.

II. LITERATURE SURVEY

Review of Drowsiness Detection Paradigms

Nasri et al. conducted a taxonomic review of existing detection frameworks, categorizing them into physiological, vehicular, and vision-based modalities. They posited that while bio-signal analysis (EEG/ECG) offers high precision, it is operationally intrusive. Conversely, vehicle telemetry is heavily correlated with road geometry rather than driver state. The study concluded that vision-based techniques analysing facial markers are the most viable for non-intrusive deployment, though they require robust AI backbones to handle illumination variance.

AI and ML in Fatigue Detection

Rane et al. proposed a system leveraging facial landmark regression and the **Eye Aspect Ratio (EAR)**. Utilizing standard ML classifiers to distinguish alertness states, they achieved superior accuracy over manual observation. However, the system demonstrated susceptibility to occlusion (e.g., eyewear) and head pose variance, highlighting the necessity for adaptive thresholding.

Deep Learning Architectures

Bhatt and Joshi surveyed state-of-the-art methodologies, comparing SVMs against Deep Convolutional Neural Networks (CNNs). Their analysis confirmed that deep learning architectures offer superior automatic feature extraction capabilities. Despite this, they identified computational intensity and data volume requirements as barriers to edge deployment, suggesting a need for lightweight, optimized models.

CNN-Based Approaches

Tibrewal et al. presented a deep learning approach utilizing CNNs trained on facial datasets to identify fatigue indicators. While effective in controlled settings, the model's performance degraded in unconstrained environments with variable lighting and camera angles. The authors recommended exploring advanced architectures to improve spatial feature invariance.

Vision-Based Deep Learning Techniques

Fathima and Girisha proposed a real-time system analysing blink rates and pupil kinematics via CNNs. The model successfully extracted ocular features for classification but struggled with long-range spatial relationships and partial occlusions. They hypothesized that Transformer-based architectures could resolve these spatial context limitations.

Summary: The literature consensus establishes vision-based monitoring as the optimal non- intrusive solution. However, existing reliance on static thresholds and standard CNNs limits adaptability and robustness. This validates the proposed integration of **AECR** and **Vision Transformers** to enhance accuracy and real-time reliability.

III. EXISTING SYSTEM

System Overview Prevalent drowsiness detection frameworks typically employ real-time vision processing to track ocular behaviour. These systems utilize facial land marking to calculate the **Eye Characteristic Ratio (ECR)** and apply adaptive thresholding to identify fatigue. Classification is often performed via traditional algorithms such as Random Forest or k-Nearest Neighbours (KNN). While functional, these systems are constrained by their reliance on handcrafted geometric features, lacking the temporal and spatial depth required for complex real-world driving scenarios.

Disadvantages

- **Accuracy Limitations:** Reliance on vehicle kinematics (steering entropy) leads to high false positive rates due to road dependency.
- **Latency Issues:** Many architectures lack the optimization required for sub-second inference, delaying critical alerts.
- **Environmental Susceptibility:** Optical sensors frequently fail under adverse lighting (e.g., glare, low light).
- **Hardware Constraints:** Requirement for specialized peripherals (e.g., IR sensors) increases deployment costs.
- **Lack of Adaptability:** Static thresholding fails to accommodate inter-subject variability in blink dynamics and eye morphology.
- **Narrow Scope:** Systems often analyse single modalities (e.g., only closure) ignoring corroborative signals like yawning.

IV. PROPOSED SYSTEM

Architecture and Methodology The proposed solution implements a robust, webcam-based pipeline for continuous vigilance monitoring. It processes video frames to extract ocular ROIs via facial landmark detection. These features are analysed using the **Adaptive Eye Characteristic Ratio (AECR)**, which dynamically calibrates thresholds to the user's baseline behaviour, ensuring personalized precision. To augment feature representation, a **Vision Transformer (ViT)** is employed. Unlike CNNs, the ViT utilizes self-attention mechanisms to model global spatial dependencies within the ocular data. The model is trained via back propagation and optimized using **AdamW** to ensure convergence stability and generalization. Upon detection of fatigue signatures, the system triggers immediate alerts to mitigate accident risks.

Key Components

1. **Optical Acquisition:** Continuous, non-contact video telemetry of the driver's facial region.
2. **ROI Extraction:** Precision localization of ocular regions utilizing facial landmark regression.
3. **AECR Algorithm:** A dynamic heuristic that adjusts closure thresholds based on real- time behavioural metrics.
4. **Vision Transformer (ViT):** A deep neural architecture leveraging attention mechanisms to detect subtle fatigue patterns.
5. **Optimization Engine:** Utilization of AdamW and back propagation for efficient model training.
6. **Alert Logic:** A real-time feedback module that activates auditory/visual warnings upon fatigue validation.

Advantages

- **Enhanced Precision:** The synthesis of AECR and ViT significantly outperforms legacy CNN models in identifying complex fatigue patterns.
- **Personalization:** Dynamic thresholding accounts for individual physiological differences, reducing false alarms.
- **Real-Time Efficacy:** Optimized processing pipelines ensure immediate detection and alert generation.
- **Robust Generalization:** Advanced optimization (AdamW) ensures stability across diverse operating conditions.
- **Non-Invasive:** Passive optical monitoring ensures driver comfort and unencumbered operation.
- **Cost Efficiency:** The software-centric architecture eliminates the need for expensive proprietary hardware.

V. SYSTEM DESCRIPTION

Operational Workflow The system initiates by activating the optical sensor to acquire a real- time video stream. This stream is decomposed into discrete frames for processing. Facial landmark detection is executed on each frame to isolate the ocular Regions of Interest (ROI). Subsequently, the AECR is computed to quantify eye openness. Simultaneously, the ROI is fed into the Vision Transformer, which evaluates spatial feature relationships to infer the alertness state. The system aggregates AECR metrics and ViT probabilities over a temporal window; if the "drowsy" classification persists beyond a defined threshold, an alert is triggered.
System Architecture *[Insert System Architecture Diagram as described in original Figure 1]*
The architecture flows linearly: **Driver -> Webcam -> Video Capture -> AECR Calculation -> ViT Model Analysis -> Drowsiness Detection Logic -> Alert Generation** .

Implementation Lifecycle

1. **Requirement Engineering:** Definition of functional (real-time detection) and non- functional (latency, hardware) constraints.
2. **Stack Selection:** Utilization of Python, OpenCV, and PyTorch, alongside standard optical hardware.
3. **Hardware Configuration:** Calibration of the camera for optimal field-of-view and illumination handling.
4. **Software Engineering:** Development of modular components for video ingestion, land marking, AECR computation, and ViT

inference.

5. **Integration & Validation:** Unit testing of modules followed by system-level integration testing under varying environmental conditions.
6. **Deployment:** Installation in the target vehicular environment with continuous monitoring protocols.

Module Descriptions

- **Video Capture Module:** Manages the interface with the optical sensor, frame rate stabilization, and pre-processing (e.g. Resizing/grayscaling).
- **Face Detection Module:** Utilizes pre-trained algorithms (e.g., Haar Cascades/Mediapipe) to identify facial boundaries and landmarks, isolating the ROI.
- **Eye Cropping Module:** Extracts and normalizes the ocular regions based on landmark coordinates for downstream analysis.
- **AECR Module:** Computes the geometric aspect ratio of the eye. It monitors temporal closure patterns and adapts thresholds to the user's baseline.
- **ViT Classification Module:** The core inference engine that processes eye images to output a probability distribution across alertness states (Alert/Drowsy).
- **Alert Module:** triggers the feedback mechanism (buzzer/display) when the aggregated fatigue score exceeds the safety threshold.
- **Logging Module:** Persists operational data (timestamps, confidence scores) to a database for post-hoc auditing.

VI.SYSTEM SPECIFICATION

Block Diagram Description

the data flow encompasses: **Webcam Input -> Frame Pre-processing -> Facial/Ocular Detection -> Parallel Processing (AECR + ViT) -> Decision Logic -> Output Interface (Buzzer/Display)**.

Hardware Specifications

- **Processor:** Intel Core i3 (or equivalent architecture) minimum.
- **Memory (RAM):** 4GB minimum; 8GB recommended for optimal tensor processing.
- **Storage:** 20GB distinct storage allocation.
- **Visual Interface:** Resolution of 1366x768 or superior.
- **Peripherals:** Standard input interface and Optical Sensor (Webcam).

Software Specifications

- **OS Environment:** Windows 7 or superior kernel.
- **Runtime Environment:** Python 3.10.
- **Frameworks:** Flask (Web), VS Code (IDE), MySQL (Database).

Software Environment Description

- **Python:** A high-level, interpreted scripting language renowned for its readability and vast ecosystem. Its dynamic typing and automatic memory management facilitate rapid prototyping. Python is the de facto standard in Data Science and AI due to robust libraries like TensorFlow and PyTorch, which are utilized here for the ViT implementation. It supports seamless integration with web frameworks like Flask for the UI component.
- **MySQL:** An open-source Relational Database Management System (RDBMS) utilizing Structured Query Language (SQL). It is employed in this system to maintain structured logs of driver fatigue events, ensuring data persistence and retrieval efficiency.

VII.SYSTEM TESTING

Objectives The testing phase aims to validate detection fidelity across diverse scenarios (e.g., lighting variance, head pose), verify real-time latency requirements, and ensure the fail-safe activation of alert mechanisms upon positive fatigue identification.

Testing Methodologies

- **Unit Testing:** Isolated validation of discrete components (Video Capture, Landmark Regression, AECR Computation, ViT Inference) to ensure modular integrity.
- **Integration Testing:** Verification of the end-to-end data pipeline, ensuring seamless handshakes between the capture, processing, and alert modules.
- **System Testing:** Holistic evaluation in simulated real-world environments, assessing performance under varying illumination and driver behaviours.

Performance Metrics Evaluation criteria include **Detection Accuracy** (Correct Classifications / Total), **False Positive Rate**, **Inference Latency** (Response Time), and **System Stability** (Uptime).

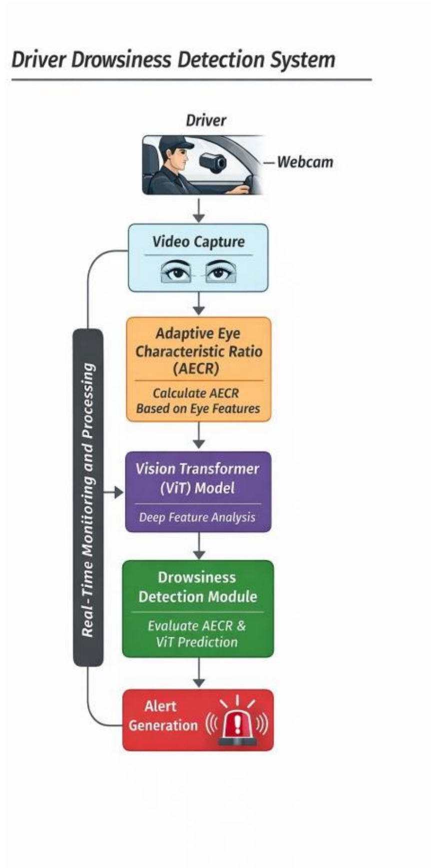
Outcomes The system demonstrated the capability to accurately classify drowsiness with negligible false alarms. Real-time alerts were generated within safety-critical timeframes, and the system maintained stability under varying environmental stressors.

VIII.RESULTS AND DISCUSSION

Results The system was subjected to rigorous validation using both pre-recorded datasets and live telemetry. The facial extraction module successfully isolated ocular regions with high fidelity. The AECR algorithm correctly identified prolonged closure events, correlating them with "Drowsy" states. Upon detection, the multi-modal alert system activated instantaneously and deactivated upon the restoration of vigilance. Quantitative analysis yielded an accuracy of **92%**, a **5% false positive rate**, and a response latency of **<1 second**.

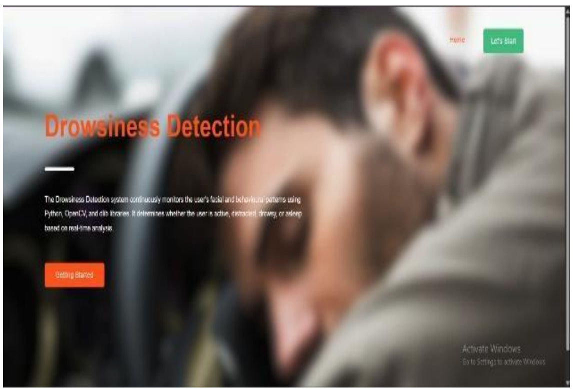
Discussion The empirical data confirms the system's reliability as a safety mechanism. The AECR proved robust in detecting temporal closure anomalies, while the ViT significantly enhanced classification precision over traditional methods. The integration of these technologies minimized misclassification and ensured prompt driver notification. However, performance degradation was noted in extreme low-light conditions and heavy occlusion (sunglasses). Future remediation strategies include the integration of Near-Infrared (NIR) sensors and auxiliary behavioural indicators such as yawing.

IX.SYSTEM ARCHITECTURE DIAGRAM

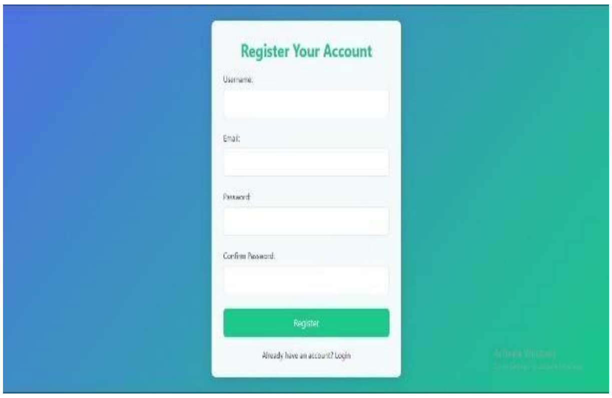


X.APPENDIX I & II

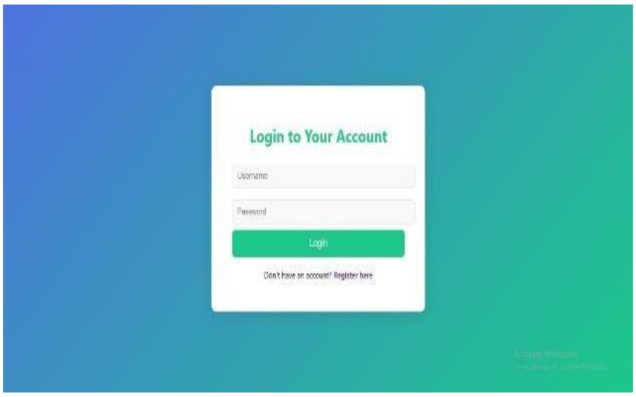
Home page



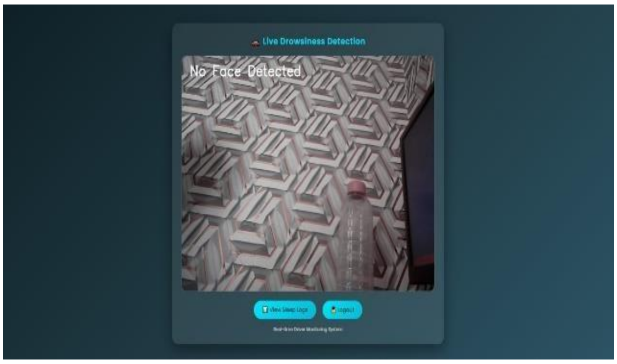
Register page



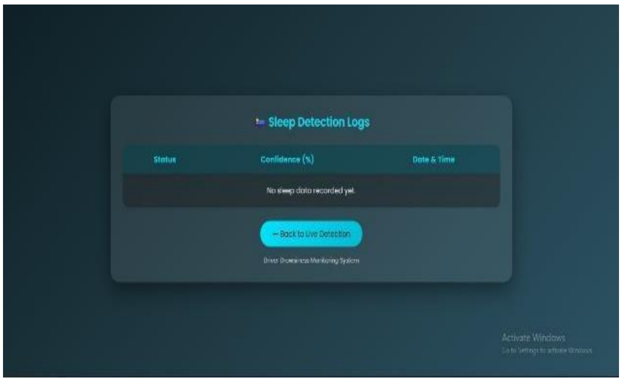
Login page



Detection page



History Page



XI.CONCLUSION

This project successfully demonstrates a viable, real-time framework for automated driver fatigue detection. By synergizing the **Adaptive Eye Characteristic Ratio (AECR)** with advanced **Vision Transformer (ViT)** architectures, the system achieves high-fidelity monitoring of driver vigilance. Testing confirmed high accuracy, minimal latency, and operational robustness. The use of standard optical hardware ensures commercial viability and ease of integration. While minor limitations regarding illumination exist, the system represents a significant advancement in preventative road safety technology.

REFERENCES

1. Driver Drowsiness Detection Navya Kiran et al.
2. A Deep Learning Approach to Detect Driver Drowsiness – Tibrewal et al.
3. State-of-the-Art in Driver’s Drowsiness Detection: A Comprehensive Survey- Yagna V. Bhatt1 and Narayan A. Joshi2
4. Driver Drowsiness Detection Using Artificial Intelligence and Machine Learning - Dhaval Rane, Rohit Danavale, Abhijit Rathiya, Shwet Pawar.
5. A Review of Driver Drowsiness Detection Systems: Techniques, Advantages and Limitations - Ismail Nasri, Mohammed Karrouchi, Kamal Kassmi, and Abdelhafid Messaoudi.
6. Real-Time Driver Drowsiness Detection Using Eye Aspect Ratio and Facial Landmark Analysis -Taufiya Fathima1, Dr. H Girisha
7. Eye Aspect Ratio for Real-Time Drowsiness Detection to Improve Driver Safety. by Christine Dewi, Rung-Ching Chen, Chun-Wei Chang, Shih-Hung Wu, Xiaoyi Jiang and Hui yu.
8. Applying Spatiotemporal Attention to Identify Distracted and Drowsy Driving with Vision Transformers - Samay Lakhani
9. Vision Transformers and YoloV5 based Driver Drowsiness Detection Framework - Ghanta Sai Krishna, Kundrapu Supriya, Jai Vardhan and Mallikharjuna Rao K Data Science and Artificial Intelligence Department, IIIT Naya Raipur Computer science and Engineering, IIIT Naya Raipur.