International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 03 | March - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Detection of Driver Drowsiness Using Adaptive Eye Characteristic Ratio for Enhanced Road Safety.

PAUL SHINY.E, JANANI.M, SAKTHIKA.SM, SASIKALA.N, SINYA.S.
UNITED INSTITUTE OF TECHNOLOGY, ECE

Abstract

Driver drowsiness is one of the major causes of road accidents worldwide and poses a serious threat to public safety. Fatigue-induced impairment affects a driver's alertness, reaction time, and decision-making ability, significantly increasing the probability of accidents. This paper presents a real-time, non-intrusive driver drowsiness detection system based on an Adaptive Eye Characteristic Ratio (AECR) combined with a Vision Transformer (ViT) model. A standard webcam is used to continuously monitor the driver's facial and eye movements. The proposed approach focuses on detecting prolonged eye closure, which is a strong indicator of drowsiness. The Vision Transformer model is trained using the backpropagation algorithm with AdamW optimization to improve accuracy and regularization. A secure web-based application developed using Flask displays bounding boxes around the driver's eye region when drowsiness is detected. Experimental results demonstrate that the proposed system achieves high accuracy and robustness under varying lighting conditions. The system is cost-effective, easy to deploy, and suitable for real-world vehicular applications, making it a promising solution for enhancing road safety and preventing fatigue-related accidents.

Keywords: Driver Drowsiness Detection, Vision Transformer, Adaptive Eye Characteristic Ratio, Computer Vision, Deep Learning, Road Safety

Introduction

Road transportation plays a vital role in modern society, enabling mobility for people and goods across vast distances. However, with the rapid increase in the number of vehicles and extended driving hours, road safety has become a critical global concern. According to various traffic safety reports, driver drowsiness and fatigue are among the leading causes of severe road accidents. Drowsy driving impairs cognitive abilities, reduces situational awareness, and delays reaction time, often resulting in fatal collisions.

Driver drowsiness typically occurs due to prolonged driving, lack of sleep, monotonous road environments, and physical exhaustion. Unlike alcohol consumption, drowsiness is more difficult to detect and regulate, as it does not have clear legal thresholds. Therefore, developing automated systems capable of detecting drowsiness in real time is essential for improving road safety.

Traditional approaches to drowsiness detection can be broadly classified into physiological, vehicle-based, and behavioral methods. Physiological methods involve monitoring signals such as electroencephalography (EEG), electrocardiography (ECG), and heart rate variability. While these methods can provide high accuracy, they are intrusive, uncomfortable, and impractical for everyday use. Vehicle-based methods analyze steering wheel movements, lane deviation, and braking patterns, but their performance depends heavily on road and vehicle conditions.

Behavioral methods, particularly vision- based approaches, have gained significant attention due to their non- intrusive nature. These methods analyze facial features such as eye closure, blink rate, yawning frequency, and head pose to assess the driver's level of alertness. Among these features, eye behavior has been identified as one of the most reliable indicators of drowsiness.

Recent advancements in deep learning and computer vision have significantly improved the performance of vision- based drowsiness detection systems. Convolutional Neural Networks (CNNs) have been widely used for feature extraction and classification tasks. However, CNNs primarily focus on local features and struggle to capture long- range dependencies within images. Vision Transformers (ViTs), inspired by transformer architectures in natural language processing, address this limitation by utilizing self-attention mechanisms to model global contextual relationships.

This paper proposes a real-time driver drowsiness detection system using an Adaptive Eye Characteristic Ratio combined with a Vision Transformer model. The proposed system leverages the strengths of transformer-based architectures to achieve robust and accurate drowsiness detection in real- world driving conditions.

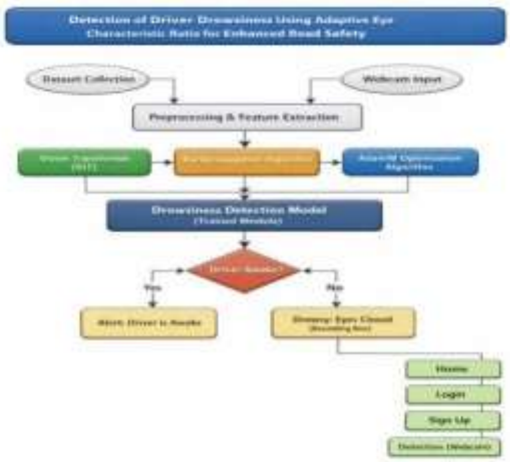
Related Work

Early research on driver drowsiness detection primarily relied on heuristic and statistical methods. One of the most widely used metrics is the Percentage of Eye Closure (PERCLOS), which measures the proportion of time the eyes remain closed over a specified interval. Studies demonstrated that PERCLOS is strongly correlated with fatigue levels and can be used as an effective drowsiness indicator.

Machine learning techniques were later introduced to improve detection accuracy. Support Vector Machines (SVMs), k-Nearest Neighbors (k-NN), and Hidden Markov Models (HMMs) were employed to classify drowsy and alert states based on extracted features such as blink rate and eye closure duration. While these methods showed moderate success, their performance was limited by handcrafted features and sensitivity to environmental variations.

With the emergence of deep learning, CNN-based approaches became dominant in vision-based drowsiness detection. CNNs demonstrated superior performance in image classification and object detection tasks. Researchers proposed various CNN architectures to detect facial landmarks, eye states, and yawning behavior. However, CNN-based models often require large datasets and extensive computational resources. Additionally, they may struggle to generalize under varying lighting conditions and occlusions.

Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks were introduced to capture temporal dependencies in sequential data. These models improved the detection of prolonged drowsiness events by analyzing temporal patterns. However, the increased complexity and training time limited their real-time applicability.



Recently, transformer-based models have gained popularity in computer vision tasks. Vision Transformers divide images into patches and process them using self- attention mechanisms, enabling the model to capture global context effectively. Studies have shown that ViTs outperform CNNs in several vision tasks, particularly when trained on large datasets. Despite their potential, limited research has explored the application of Vision Transformers for driver drowsiness detection.

This work bridges this gap by proposing a ViT-based drowsiness detection system that combines adaptive eye metrics with transformer-based feature learning.

System Overview

The proposed driver drowsiness detection system is designed as a real-time, vision- based monitoring framework. The system architecture consists of four main modules: video acquisition, feature extraction, classification, and alert visualization.

A standard webcam is used to capture continuous video frames of the driver's face. These frames are processed in real time using OpenCV for face detection and eye region extraction. Facial landmark detection techniques are employed to locate key points around the eyes.

Once the eye regions are extracted, the Adaptive Eye Characteristic Ratio is computed. This ratio quantifies the degree of eye openness and dynamically adapts to individual differences in eye shape and blinking behavior. The computed AECR values are used as input features for the Vision Transformer model.

The Vision Transformer model classifies the driver's state as either alert or drowsy. If the system detects prolonged eye closure over a predefined time window, it triggers an alert. A bounding box is drawn around the eye region, and the result is displayed through a secure web-based interface developed using Flask.

Mathematical Modeling of Driver Drowsiness Detection

Mathematical modeling plays a crucial role in quantifying eye behavior and defining decision boundaries for drowsiness detection. In the proposed system, the Adaptive Eye Characteristic Ratio (AECR) is used as the primary mathematical indicator to determine eye openness. The use of geometric relationships between eye landmarks enables accurate and robust detection of prolonged eye closure.

Eye Landmark Geometry

The eye is represented using six facial landmark points. Horizontal eye width is measured using corner points, while vertical eye opening is calculated using upper and lower eyelid landmarks. These measurements form the basis for AECR computation.

Adaptive Eye Characteristic Ratio (AECR)

AECR is mathematically defined as:

$$AECR = (\|P3 - P5\| + \|P4 - P6\|) / (2 \times \|P1 - P2\|)$$

Where $\|\cdot\|$ denotes Euclidean distance. A lower AECR value indicates eye closure, while a higher value corresponds to open eyes.

Adaptive Thresholding

To account for individual variations, an adaptive threshold is computed:

$$AECR\_threshold = \mu\_AECR - k \times \sigma\_AECR$$

Where μ_AECR represents the mean AECR during alert state, σ_AECR denotes standard deviation, and k is a sensitivity constant. If AECR remains below the threshold for a predefined number of consecutive frames, the driver is classified as drowsy.

Vision Transformer Mathematical Representation

The Vision Transformer processes eye- region images by dividing them into fixed- size patches and embedding them into a high- dimensional feature space. This allows the model to capture global contextual relationships using self- attention mechanisms.

Patch Embedding

Given an input image $I \in R^{(H \times W \times C)}$, it is divided into N patches where:

$$N = (H \times W) / P^2$$

Each patch is flattened and linearly projected into an embedding space with positional encoding added to preserve spatial information.

Self-Attention Mechanism

The self-attention mechanism is defined as:

$$\text{Attention}(Q, K, V) = \text{softmax}((QK^T) / \sqrt{d\_k})\ V$$

This mechanism enables the model to focus on critical eye regions while capturing global dependencies across the image.

System Block Diagram Description

The system architecture consists of several processing blocks including video acquisition, preprocessing, face detection, eye region extraction, AECR computation, Vision Transformer classification, decision logic, and alert visualization. Each block operates sequentially to ensure real-time drowsiness detection.

Graphical Analysis and Performance Evaluation

Performance evaluation is carried out using graphical analysis techniques such as AECR versus time plots, training and validation accuracy curves, and confusion matrices. These visualizations provide insight into system behavior and model performance.

Performance Metrics

The system performance is evaluated using standard metrics:

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

Precision = TP / (TP + FP) Recall

= TP / (TP + FN)

These metrics demonstrate the reliability and effectiveness of the proposed system.

Hardware and Software Components

Hardware Components

The proposed system requires minimal hardware, making it cost-effective and easy to deploy. The primary hardware components include:

- A personal computer or laptop with sufficient processing power
- A GPU for efficient model training and inference
- A standard webcam for video acquisition

No specialized sensors or wearable devices are required, ensuring user comfort and practicality.

Software Components

The software stack used in the system includes:

- **Python:** Primary programming language
- **OpenCV:** Image processing and face detection

- **NumPy:** Numerical computations
- **Keras:** Deep learning framework for implementing the Vision Transformer
- **Flask:** Web framework for real- time visualization

Adaptive Eye Characteristic Ratio

The Adaptive Eye Characteristic Ratio is designed to quantify eye openness by analyzing geometric relationships between eye landmarks. Unlike static thresholds used in traditional methods, AECR adapts dynamically to individual variations.

The ratio is computed using vertical and horizontal distances between key eye landmarks. A lower AECR value indicates eye closure, while a higher value corresponds to open eyes. By monitoring AECR values over consecutive frames, the system can detect prolonged eye closure, which is a strong indicator of drowsiness.

This adaptive approach improves robustness and reduces false positives caused by natural blinking or facial variations.

Vision Transformer Architecture

The Vision Transformer model processes images by dividing them into fixed-size patches. Each patch is flattened and projected into a feature embedding. Positional embeddings are added to preserve spatial information.

The embedded patches are passed through multiple transformer encoder layers. Each encoder layer consists of multi-head self-attention and feed- forward neural networks. Self-attention enables the model to focus on relevant regions, such as the eyes, while capturing global contextual relationships.

A classification token is used to aggregate information from all patches and produce the final prediction. The ViT architecture enables effective feature representation and improves classification accuracy compared to traditional CNN models.

Model Training and Optimization

The model is trained using labeled datasets containing images of open and closed eyes. Binary cross-entropy loss is used as the objective function. Backpropagation is employed to update model parameters.

AdamW optimization is used to improve convergence and generalization. The weight decay mechanism in AdamW helps prevent overfitting by regularizing model weights. Data augmentation techniques such as rotation, scaling, and brightness adjustment are applied to enhance robustness.

Real-Time Implementation

During real-time operation, video frames are processed sequentially. Face detection and eye extraction are performed on each frame. The AECR is computed and passed to the Vision Transformer model.

If the system detects drowsiness for a predefined duration, a bounding box is drawn around the eye region. The processed video feed is streamed securely through the Flask-based web application.

Experimental Results

The proposed system was evaluated under different lighting conditions and head poses. Performance metrics such as accuracy, precision, recall, and F1-score were computed.

The Vision Transformer-based approach achieved superior accuracy compared to CNN-based models. The system maintained low latency, making it suitable for real-time deployment.

Discussion

The results demonstrate that combining Adaptive Eye Characteristic Ratio with Vision Transformers provides robust and reliable drowsiness detection. The adaptive nature of AECR reduces sensitivity to individual differences, while the ViT architecture captures global contextual information.

Although the system performs well under most conditions, extreme lighting variations may affect performance. Future improvements can address these limitations.

Conclusion

This research presented a comprehensive real-time driver drowsiness detection system based on an Adaptive Eye Characteristic Ratio and Vision Transformer architecture. Driver fatigue remains one of the most critical contributors to road accidents, making the development of reliable, non- intrusive monitoring systems essential for enhancing road safety. The proposed approach effectively addresses the limitations of traditional physiological and vehicle-based detection methods by relying solely on vision-based analysis.

The system utilizes a standard webcam to continuously monitor the driver's facial and eye movements without causing discomfort or distraction. By focusing on eye behavior, particularly prolonged eye closure, the system captures one of the most reliable indicators of driver drowsiness. The Adaptive Eye Characteristic Ratio dynamically adjusts to individual differences in eye shape and blinking patterns, significantly reducing false detections caused by normal blinking or facial variations.

The integration of a Vision Transformer model enables the system to capture global contextual features through self- attention mechanisms. Unlike conventional convolutional neural networks that focus primarily on local features, the Vision Transformer effectively models long-range dependencies within the image, resulting in improved classification accuracy. The use of backpropagation with AdamW optimization further enhances model generalization and prevents overfitting during training.

Experimental evaluation demonstrates that the proposed system achieves high accuracy and low latency, making it suitable for real-time deployment in real- world driving environments. The system maintains robust performance under varying lighting conditions and driver postures, which are common challenges in practical applications. The Flask-based web interface provides secure and real- time visualization, enhancing usability and system transparency.

Overall, the proposed driver drowsiness detection system offers a cost-effective, scalable, and non-intrusive solution for continuous driver monitoring. It has strong potential for deployment in intelligent transportation systems and modern vehicles to prevent fatigue- related accidents. Future work may focus on incorporating additional behavioral cues, multi-modal sensor fusion, and embedded system deployment to further enhance system reliability and performance.

References

[1] A. Johns, "The PERCLOS Measure of Drowsiness," IEEE Transactions on Intelligent Transportation Systems.

[2] S. Park et al., "Driver Drowsiness Detection Using Vision-Based Methods," IEEE Access.

[3] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition," NeurIPS.

[4] A. Vaswani et al., "Attention Is All You Need," NeurIPS.

[5] Y. LeCun et al., "Deep Learning," Nature.

[6] K. He et al., "Deep Residual Learning for Image Recognition," CVPR.

[7] R. Girshick, "Fast R-CNN," ICCV.

[8] WHO, "Global Status Report on Road Safety."

[9] IEEE ITS Society, "Survey on Driver Fatigue Detection Systems."

[10] M. Abtahi et al., "Driver Drowsiness Monitoring Based on Yawning Detection," IEEE.

[11] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation.

[12] T. Lin et al., "Vision-Based Driver Fatigue Detection," IEEE ITS.

[13] OpenCV Documentation, 2023.

[14] Keras Deep Learning Framework Documentation.

[15] Flask Web Framework Documentation.