INTERNATIONAL SCIENTIFIC JOURNAL OF ENGINEERING AND MANAGEMENT (ISJEM) ISSN: 2583-6129
CONFERENCE PROCEEDINGS: VOLUME: 05 ACME – 2026 DOI: 10.55041/ISJEM.ACME158
AN INTERNATIONAL SCHOLARLY || MULTIDISCIPLINARY || OPEN ACCESS || INDEXING IN ALL MAJOR DATABASE & METADATA

Vision Transformer-Based Real-Time Driver Drowsiness Monitoring with Enhanced Safety Performance

Jithendra Reddy Dandu Department of Electronics and Communication Engineering Annamacharya Institute of Technology and Science Tirupati,India
jithendrareddy.d@gmail.com

Poola Jyothika Reddy
Department of Electronics and communication Engieering
Annamacharya institute of technology and Sciences
Tirupati,India
Jyothikareddy345@gmail.com

Yadadhala Kalyan Reddy
Department of Electronics and communication Engineering
Annamacharya institute of technology and Sciences Tirupati,India
ykalyanreddy.8801@gmail.com

Chippagiri Mahesh Reddy
Department of Electronics and communication Engineering
Annamacharya institute of technology and Sciences
Tirupati,India
maheshreddyc072@gmail.com

Anduri Manisha Reddy
Department of Electronics and communication Engineering
Annamacharya institute of technology and Sciences
Tirupati,India
manireddy092200@gmail.com

ABSTRACT-Fatigue and drowsiness of drivers are one of the primary reasons for road accidents across the world. Driver monitoring systems are essential for road safety. This research proposes a real-time driver tiredness detection system using a fine-tuned Vision Transformer. The proposed system focuses on identifying eye states (open or closed) to determine the level of driver alertness and provide timely warnings to prevent accidents. A Vision Transformer architecture was fine-tuned by adding additional layers and training it on a large dataset consisting of 84,900 images representing open-eye and closed-eye conditions. The transformer-based model effectively captures spatial features and attention patterns in eye images, enabling accurate classification of driver alertness states. The system operates in real time using a camera that continuously monitors the driver's face and analyzes eye movements. When the model detects prolonged eye closure indicating possible drowsiness, an alarm mechanism is activated to alert the driver immediately. This real-time warning helps the driver regain attention and reduces the likelihood of fatigue-related accidents. The experimental results show that the proposed system has an accuracy of 98.8% along with high precision, recall, and F1-score for both open eye and closed eye detection. This demonstrates the results obtained by employing Vision Transformer-based models for improving the performance of the driver monitoring system significantly. This presented system for real-time monitoring along with the application of sophisticated deep learning (DL) models provides a favourable solution for improving the efficiency of the driver monitoring system. This work contributes significantly towards the development of an intelligent transportation system for improving road safety by reducing the number of accidents resulting from driver fatigue.

Keywords-Driver Drowsiness Detection, Vision Transformer (ViT), Computer Vision, Eye State Classification, Road Safety, Deep Learning, Real-Time Monitoring**.**

I. INTRODUCTION

Ensuring everyone's safety on the road depends critically on the alertness of drivers. The WHO estimates that 1.3 million people lose their lives in accidents every year. Twenty to fifty million more people suffer non-fatal injuries, many of which result in lifelong disability. The main reasons for traffic accidents are tiredness, drowsiness, and inattentive driving. As a result, a system that can alert drivers to fatigue and help them become more vigilant by sending out timely warnings may be able to reduce the number of accidents that occur, save money, and reduce personal suffering.

The research introduces an innovative approach to advance driver safety through the fine-tuning of the Vision Transformer (ViT) model on a dataset for the accurate analysis of driver drowsiness. By delving into advanced computer vision, the study seeks to develop a robust system for proactive detection of driver drowsiness, ultimately contributing to accident prevention and overall road safety improvement. It is worth emphasizing the preference for Vision Transformer (ViT) over traditional algorithms like CNN or KNN due to ViT's exceptional capability to capture intricate long-range dependencies in images. To guarantee the reliability of the training and testing, the model uses a large labeled dataset training approach, where the dataset is split into 80% for training, 10% for validation, and 10% for testing.

II. LITERATURE SURVEY

Driver drowsiness is one of the primary causes of traffic accidents globally, making the development of intelligent monitoring systems an important research focus. Several researchers have explored Artificial Intelligence (AI) and computer vision techniques to detect fatigue and improve road safety. Ganapathy and Sankaradass proposed an AI-powered system that analyzes driver eye movements and facial features to detect signs of tiredness and trigger alerts [1]. Similarly, Thakur et al. developed a real-time fatigue detection system employing deep learning algorithms that monitor eye blinking patterns, facial expressions, and head movements to warn drivers when fatigue is detected [2]. Padmaja et al. emphasized the use of visual cues such as eye closure, yawning, and head posture to identify driver fatigue using AI models [3]. Kalla highlighted the importance of AI-based driver behavior analysis in Advanced Driver Assistance Systems (ADAS), where machine learning techniques help identify unsafe driving patterns and support accident prevention [4].

Recent studies have further improved these systems by integrating advanced technologies and hybrid models. Merakanapalli and Bodapati introduced a system that combines drowsiness detection with autonomous steering control, enabling the vehicle to navigate toward a safe zone when the driver becomes unresponsive [5]. Chikhale et al. proposed an AI-powered intelligent cognitive monitoring system designed to enhance transportation safety and fleet management efficiency [6]. Kamboj et al. provided a comprehensive review of machine learning, deep learning, and physiological approaches used in fatigue detection, highlighting the efficiency of DL approaches in improving accuracy [7]. Bend et al. analyzed the utilized of mobile technologies and eye-tracking metrics to study fatigue patterns among drivers [8]. Chari and Prashant proposed a hybrid DL and ML model for real-time driver drowsiness detection [9]. Additionally, Chukwunweike et al. emphasized the importance of privacy, security, and data integrity in AI-driven systems [10]. Overall, these studies demonstrate the potential of AI-based monitoring systems in

improving driver safety while also identifying areas for further enhancement in accuracy, reliability, and real-time performance.

III. METHODOLOGY

The methodology of the proposed system focuses on detecting driver drowsiness in real time using computer vision and DL approaches. The system starts by capturing video frames of the driver through a camera installed inside the vehicle. These frames are continuously processed to monitor the driver's facial behaviour during driving. The first stage of processing involves detecting the driver's face using computer vision algorithms. Once the face is identified, the system extracts the eye region because eye movements and blinking patterns are strong indicators of fatigue. The extracted eye images are then pre-processed to ensure consistent size and quality before being passed to the DL model. In this project, a fine-tuned Vision Transformer (VIT) model is used to analyse the eye images. To identify key visual elements, the Vision Transformer splits the image into tiny pieces and employs a self-attention process. This approach allows the model to capture relationships between different parts of the image more effectively than traditional convolutional neural networks. The trained model classifies each eye image into two categories: open eye or closed eye. The system continuously monitors these predictions over time to identify abnormal blinking patterns or prolonged eye closure. If the eyes remain closed for a predefined duration, the process determines that the driver may be drowsy. Once drowsiness is detected, the system activates an alert mechanism to warn the driver immediately. This methodology ensures that the driver's condition is monitored continuously and that early signs of fatigue are detected before they lead to dangerous situations

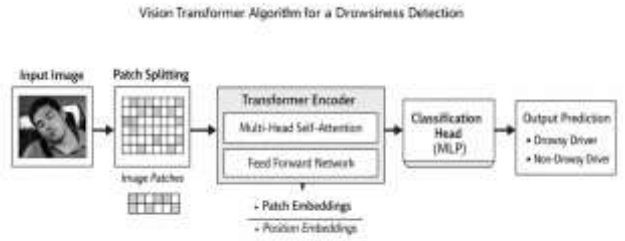

Fig-1 Block Diagram of vision transformer

When a camera captures an image or a video frame, the Vision Transformer (ViT) processes it through several stages to understand what is present in the image and then produce an output such as classification or detection. The working process can be explained in the following simple and clear steps.

Image Capture from Camera

The system first receives the input image from the camera feed. In real-time applications like a driver drowsiness detection system, a webcam continuously captures frames of the driver's face while the vehicle is in motion. Each captured frame is treated as an input image that will be analysed by the model.

Image Pre-Processing

Before the image is given to the model, it goes through a preprocessing stage. In this step, the captured frame is resized to a fixed resolution such as 224 × 224 pixels. The pixel values are also normalized so that all images follow a similar scale. This step ensures that the model receives data in a consistent format, which helps improve the accuracy and stability of the system.

Image Patch Creation

Instead of analyzing the entire image at once, the Vision Transformer divides the image into smaller patches. For example, an image of 224 × 224 pixels can be divided into patches of 16 × 16 pixels. This results in 196 small patches. Each patch represents a small portion of the original image and will be processed individually by the model.

Patch Embedding

After the patches are created, each patch is converted into a numerical representation called an embedding. This process transforms visual information from the image patches into vectors that the transformer model can understand and process.

Positional Encoding

Since transformers do not automatically recognize the spatial location of image patches, positional information is added to each patch embedding. This step helps the model understand where each patch originally appeared in the image, allowing it to maintain the spatial structure of the image.

Transformer Encoder Processing

All patch embeddings are then passed into several Transformer Encoder layers. These layers are responsible for learning important patterns from the image. Each encoder layer contains:

- Multi-Head Self-Attention, which helps the model understand relationships between different patches in the image.
- Feed Forward Network (MLP), which further processes and refines the extracted features.
- Layer Normalization and Residual Connections, which help stabilize the learning process and improve the overall performance of the model.

Through multiple encoder layers, the model gradually learns meaningful visual patterns such as facial features, eye states, or closed eyelids, which are important for detecting driver drowsiness.

Classification Head

A special token called the classification token (CLS token) gathers information from all the image patches. This token is then passed through a fully connected layer that converts the learned features into a final prediction.

Output Generation

Finally, the system generates the output label based on the model's prediction. For example, in a driver monitoring system:

- Open Eyes → Driver is Alert
- Closed Eyes → Driver is Drowsy

If the process finds that the driver is drowsy, it can immediately trigger safety actions such as activating a buzzer, displaying a warning message on an LCD screen, or sending an alert notification. These actions help warn the driver and reduce the risk of accidents.

Hardware components:

The hardware setup of the proposed system is built using a combination of sensing, processing, and alert components arranged on a prototype board.

NODE MCU

This allows smooth communication between the vision-based monitoring setup and the hardware control unit.

Alcohol sensor:

An alcohol sensor is included to detect the presence of alcohol from the driver's breath. The sensor continuously monitors alcohol levels and sends the readings to the Arduino. If the detected value exceeds the safe limit, the controller triggers warning actions in the system.

Buzzer:

A buzzer is used as an alert device. Whenever the system detects abnormal conditions such as driver drowsiness or alcohol presence, the Arduino activates the buzzer to warn the driver immediately.

16x2 LCD Display:

It provides clear information such as system activation, driver condition, or warning alerts.

Motor driver:

A motor driver circuit connected to a DC motor is used to simulate the vehicle engine control. The motor driver allows the Arduino to control the motor safely by regulating the current supplied to it.

Power supply:

All these components are powered through a regulated power supply circuit, ensuring stable voltage for proper system operation.

IV. *RESULTS AND DISCUSSION*

The presented driver drowsiness detection system was successfully implemented using a fine-tuned Vision Transformer (ViT) model to classify the driver's eye state as open or closed in real time. The model was trained utilizing a dataset containing approximately 84,900 eye images, including both open-eye and closed-eye samples. The system was developed using Python with deep learning frameworks such as TensorFlow and PyTorch, and deployed through a web-based monitoring interface for real-time driver analysis.

During training, the model showed steady improvement in classification accuracy while the loss decreased, indicating successful learning and convergence. The trained ViT model was able to effectively identify visual patterns of eye states [11] as shown in fig-2.

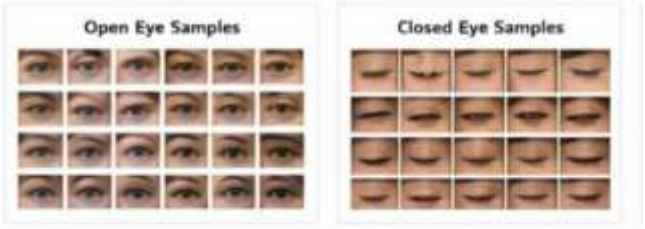

Fig-2: Dataset sample images (open-eye and closed-eye images).

The training accuracy and loss graphs further confirmed the learning behaviour of the model across multiple epochs as shown in fig-3.

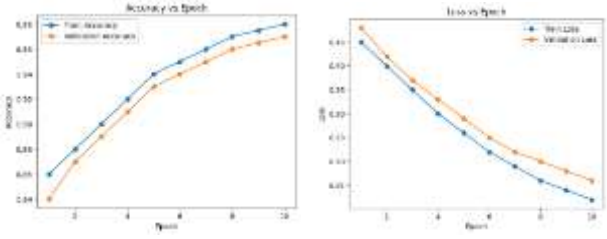

Fig-3: Training accuracy and loss graph.

The confusion matrix results are shown in fig-4 that most eye images were classified correctly with very few misclassifications, demonstrating high detection performance.

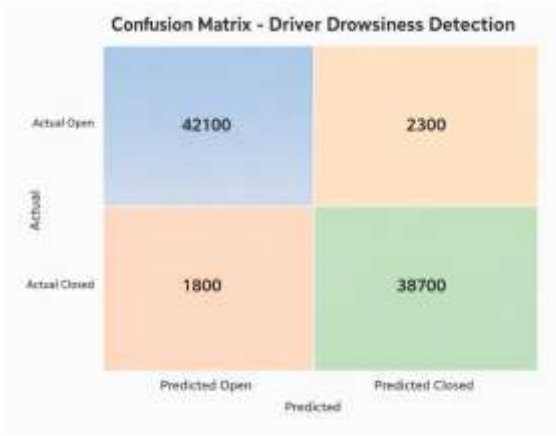

Fig-4: Confusion matrix of classification results.

For real-time testing, the system captured video frames from a webcam and analysed them using the trained model as shown in fig-5 and fig-6. When the driver's eyes remained closed beyond a predefined duration, the system detected drowsiness and generated an alert warning.

Fig-5: Real-time detection output showing "Driver Awake".

Fig-6: Real-time detection output showing "Drowsiness Detected".

To improve usability, a web monitoring interface was developed where users can access the detection system.

The web system consists of:

- Login Page – Provides secure access to the monitoring system.
- Dashboard Interface – Displays the live camera feed and detection results.
- Detection Status Display – Shows driver state such as Driver Awake or Drowsiness Detected.
- Alert Notification – Generates a warning message when drowsiness is detected.

Fig-7: Web login page.

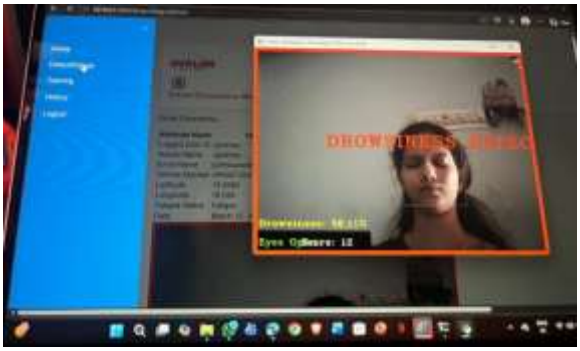

Fig-8: Monitoring dashboard showing live detection.

The experimental results demonstrate that the Vision Transformer model effectively detects driver drowsiness by capturing global image features and contextual relationships between pixels. This capability enables improved eye-state recognition compared to conventional CNN-based approaches.

However, certain challenges were observed during testing. Detection accuracy may slightly decrease under poor lighting conditions, different camera angles, or when the driver wears glasses. Future improvements may include larger datasets, infrared cameras, and additional sensors to further improve system robustness and reliability.

The hardware kit was designed to demonstrate the real-time operation of the system. It consists of a camera module, processing unit (laptop or embedded system), buzzer alert system, and power supply. The camera continuously captures the driver's face, and the processed output from the model determines whether an alert should be activated.

Fig-9: Hardware kit showing output of driver state

When drowsiness is detected, the buzzer generates an audible alert to wake the driver and prevent potential accidents. This hardware integration ensures that the system can function as a practical driver safety device in real vehicles.

Overall, the developed system offers a reliable real-time driver monitoring solution that can help mitigate road accidents caused by driver fatigue.

Table-1: Existing Models with Proposed Model

Author & year	Technology /Model	Accuracy (%)	Limitations
Nalavade et al., 2024	Deep Neural Network (DNN)	92%	Sensitive to lighting variations and driver head movements.
Yang & Yi,2024	Deep Learning with ADAS Integration	94%	Requires high computational resources and advanced hardware.
Ahmed et al., 2021	Multi-CNN Deep Learning Model	93%	Requires large datasets and high training time.
Proposed System	Vision Transformer (ViT) Based Drowsiness Detection	98.%	Requires GPU resources for efficient training and deployment.

V. CONCLUSION

The research has successfully developed a real-time Driver Drowsiness Monitoring system for drivers, employing a combination of image processing techniques and a Vision Transformer (ViT) with additional layers to discern the drowsiness of drivers. With an impressive accuracy rate of 98.8%, the model, trained on an 80-10-10 data split, showcases its robustness in identifying drowsiness. Furthermore, the integration of an alarming sound feature enhances the system's practicality by promptly capturing the driver's attention. With its proactive monitoring of driver attentiveness, this innovative solution represents a major advancement in road safety.

VI. FEATURE SCOPE

The system can be optimized for real-world conditions such as poor lighting, weather changes, and varying driver postures. Using infrared or thermal cameras can improve nighttime detection. Deploying the model on edge devices like Raspberry Pi or NVIDIA Jetson enables real-time, low-cost operation. Integration with ADAS/IoT and continuous learning can enhance alerts, personalization, and reliability.

REFERENCES

1.Ganapathy, Manimozhi, and Veeramalai Sankaradass. "AI-Powered Driver Drowsiness Detection System for Augmented Road Safety." *2025 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI)*. IEEE, 2025.

2.Thakur, Samrat Rajkumar Singh, Laxman Thakre, and Manthan Ghosh. "Combatting Driver Fatigue with Real-Time AI-Powered Detection and Prevention Technologies." *2025 12th International Conference on Emerging Trends in Engineering & Technology-Signal and Information Processing (ICETET-SIP)*. IEEE, 2025.

3.Padmaja, C., et al. "AI-Powered Road Safety: Detecting Driver Fatigue through Visual Cues." *International Journal of Information* 12.3 (2023).

4.Kalla, Dinesh. "AI-powered driver behavior analysis and accident prevention systems for advanced driver assistance." *International Journal of Scientific Research and Modern Technology (IJSRMT) Volume* 1 (2022).

5.Merakanapalli, Saibabu, and Sai Jagadish Bodapati. "Autonomous Steering Control using AI-Based Driver Drowsiness Detection and Safe-Zone Navigation." *International Journal of Research Publications in Engineering, Technology and Management (IJRPETM)* 8.5 (2025): 12773-12784.

6.Chikhale, Shraddha, et al. "AI-Powered Intelligent Cognitive Lens for Driver Monitoring for Safe Transportation and Increased Fleet Efficiency Abstract." *Symposium on International Automotive Technology (2026)*. SAE Technical Paper, 2026.

7.Kamboj, Muskan, et al. "Advanced detection techniques for driver drowsiness: a comprehensive review of machine learning, deep learning, and physiological approaches." *Multimedia Tools and Applications* 83.42 (2024): 90619-90682.

8.Bend, Julia, Markus Gödker, and Thomas Franke. "AI and Mobile Technologies for Driver Fatigue Detection: Sex Differences Revealed by Eye-Tracking Metrics." *International Journal of Interactive Mobile Technologies* 19.11 (2025): 143-158.

9.Chari, Gowrishankar Shiva Shankara, and Jyothi Arcot Prashant. "Real-time driver drowsiness detection based on integrative approach of deep learning and machine learning model." *Indonesian Journal of Electrical Engineering and Computer Science* 39.1 (2025): 592-602.

10. Chukwunweike, Joseph Nnaemeka, et al. "The role of deep learning in ensuring privacy integrity and security: Applications in AI-driven cybersecurity solutions." *World Journal of Advanced Research and Reviews* 23.2 (2024): 2550.

11. https://www.kaggle.com/datasets/sehriyarmemmedli/open-closed-eyes-dataset