



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





Multimodal Driver Drowsiness Detection Using CNN

Y. Sai Prasanna¹, R. Hemanth Kumar², P. Karthik³, P. Uma Venkata Mahesh⁴, R. Manasa⁵,
Dr. R. Satish⁶

Department of Electronics and Communication Engineering, Sir C. R. Reddy College of Engineering, Eluru, India¹⁻⁵

Professor, Department of Electronics and Communication Engineering, Sir C. R. Reddy College of Engineering,
Eluru, India⁶

ABSTRACT: Driver drowsiness is one of the major causes of road accidents worldwide, as fatigue significantly reduces a driver's attention, reaction time, and decision-making ability. Early detection of drowsiness can help prevent accidents and improve road safety. This paper presents a multimodal driver drowsiness detection system that identifies fatigue indicators using deep learning techniques. The proposed system utilizes Convolutional Neural Networks (CNN) to analyze visual cues associated with driver fatigue, specifically eye closure and yawning behavior.

Two separate CNN models are trained using eye and yawning image datasets to classify driver states. The eye detection model determines whether the driver's eyes are open or closed, while the yawning detection model identifies yawning or non-yawning conditions. The system processes input images through preprocessing steps including image resizing, normalization, and data augmentation before feeding them into the CNN models for classification.

The trained models are evaluated using performance metrics such as accuracy, confusion matrix, and classification report. Experimental results demonstrate that the proposed multimodal approach effectively identifies signs of driver fatigue and improves detection reliability by combining multiple visual indicators. The system can be extended for real-time driver monitoring applications, intelligent transportation systems, and vehicle safety technologies.

KEYWORDS: Driver Drowsiness Detection, Convolutional Neural Network (CNN), Deep Learning, Eye Closure Detection, Yawning Detection, Image Classification, Data Augmentation, Fatigue Monitoring, Road Safety, Computer Vision.

I. INTRODUCTION

Driver drowsiness has become a critical issue in road safety and is responsible for a large number of traffic accidents worldwide. Long driving hours, fatigue, and lack of sleep can significantly reduce a driver's attention, reaction time, and decision-making ability. When drivers become drowsy, their ability to safely control the vehicle decreases, increasing the risk of collisions and severe injuries. Therefore, developing intelligent systems capable of detecting driver fatigue at an early stage has become an important research area in intelligent transportation systems and computer vision.

Recent advancements in artificial intelligence and deep learning have enabled the development of automated monitoring systems that analyze visual cues from drivers to determine their level of alertness. Unlike traditional physiological monitoring systems that require wearable sensors such as electroencephalography (EEG) devices or heart-rate monitors, vision-based systems are non-intrusive and more convenient for practical applications. These systems utilize cameras and computer vision algorithms to analyze driver behavior and detect signs of fatigue.

The project presented in this work proposes a multimodal driver drowsiness detection system that analyzes two major indicators of fatigue: eye closure and yawning behavior. Eye closure is one of the most widely used indicators for detecting drowsiness because frequent blinking or prolonged eye closure often indicates reduced alertness. Similarly, yawning is a natural response to fatigue and can serve as another strong indicator of driver drowsiness. By analyzing



International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

both indicators simultaneously, the system can provide more reliable detection compared to systems that rely on a single feature.

The proposed system follows a structured workflow that begins with the collection of eye and yawning image datasets and progresses through stages including image preprocessing, feature learning using Convolutional Neural Networks (CNN), classification, and performance evaluation. In the initial stage, eye and yawning images are collected and organized into different classes such as open eye, closed eye, yawning, and non-yawning. These images form the primary input for training the deep learning models.

Following preprocessing, the system utilizes CNNs to automatically learn discriminative features from the images. CNNs are particularly effective for image classification tasks because they can capture spatial patterns, edges, textures, and hierarchical representations directly from the data. Separate CNN models are trained for eye state detection and yawning detection. Performance metrics such as accuracy, confusion matrix, precision, recall, and F1-score are used to analyze the effectiveness of the models.

II. RELATED WORK

Energy consumption forecasting and driver safety monitoring share similar challenges: both require robust pattern recognition from sequential or image data. Several studies have proposed diverse approaches to detect driver fatigue using computer vision and machine learning techniques [1].

Traditional approaches relied on physiological sensors such as EEG or ECG devices; however, these are intrusive and impractical for daily use. With advancements in deep learning, CNN-based frameworks have emerged as the preferred method for analyzing visual facial cues such as eye closure and yawning [5].

Bergasa et al. [8] developed one of the earliest real-time driver monitoring systems combining eye blinking frequency, eye closure duration, head movements, and gaze direction. Their work demonstrated that combining multiple indicators significantly improves detection reliability. Singh et al. [3] introduced a method based on Eye Aspect Ratio (EAR) for detecting eye closure, which became widely adopted for its computational efficiency.

More recently, deep learning architectures including CNN and LSTM have demonstrated strong performance in both image classification and temporal pattern analysis [5, 6]. Hybrid models combining CNN spatial feature extraction with LSTM-based temporal analysis have been proposed to capture both single-frame and sequential patterns in video-based monitoring.

TABLE I: LITERATURE REVIEW

S.No	Author(s) & Year	Method Used	Key Findings	Limitations
1	Bergasa et al. (2006) [8]	Multi-feature visual monitoring	Combining multiple indicators improves reliability	Limited lighting robustness
2	Singh et al. (2018) [3]	Eye Aspect Ratio (EAR)	Computationally efficient eye closure detection	Fails under poor lighting
3	Abtahi et al. (2011) [6]	Yawning detection via CV	Validated yawning as fatigue indicator	Controlled conditions only
4	Weng et al. (2016) [9]	CNN-based classification	CNN outperforms SVM for eye state	Large dataset required
5	Baheti et al.	Deep learning	High accuracy in	Single indicator



International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

	(2018) [10]	eye closure	real-time	only
6	Patel et al. (2019) [11]	Mouth landmark analysis	Yawning complements eye detection	Occlusion sensitivity
7	Zhang et al. (2021) [12]	Multimodal deep learning	Multimodal outperforms single-feature	High computational cost
8	Wang et al. (2022) [13]	CNN under real-world conditions	Robust under illumination variations	Needs large diverse data
9	Proposed Work	Multimodal CNN (Eye + Yawn)	99% eye accuracy, 84% yawn accuracy	Reduced hardware dependency

III. PROPOSED SYSTEM

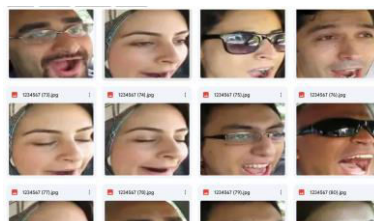
The proposed system is a simulation-based Driver Drowsiness Detection System developed using deep learning and image processing techniques. The system analyzes eye state and yawning behavior using image datasets to determine whether a driver is in an alert or drowsy condition. Unlike real-time monitoring systems, the proposed model works on pre-collected datasets and produces output in the form of classification results and performance evaluation metrics.

The system consists of five functional modules: Dataset Input Module, Image Pre-processing Module, CNN Training Module, Image Classification Module, and Performance Evaluation Module. By combining eye state detection and yawning detection, the system improves the reliability of drowsiness detection.

A. DATASET INFORMATION

TABLE II: DATASET INFORMATION

Attribute	Description
Data Source	Eye and Yawning Image Datasets
Image Classes	Open Eye, Closed Eye, Yawning, Non-Yawning
Eye Dataset Size	818 images (349 closed, 469 open)
Yawn Dataset Size	2942 images (1566 no yawning, 1376 yawning)
Data Type	Image-based (Static)
Target Variable	Driver Alertness State (Alert/Drowsy)





International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

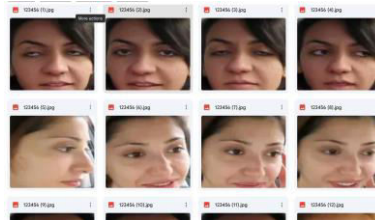


Fig. 3: Sample images from the yawning dataset (left) and the eye dataset (right).

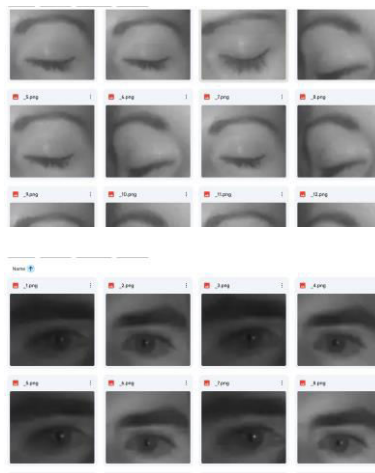


Fig. 4: Sample images from the closed-eye dataset (left) and open-eye dataset (right).

B. SYSTEM WORKFLOW

The proposed system processes data through ten sequential stages. The process begins with Data Collection, where eye and yawning image datasets are loaded from pre-organized folders. In Data Preprocessing, images are converted to uniform format and any inconsistencies are resolved. Feature Engineering involves extracting relevant visual patterns from image data such as eye shape and mouth opening patterns.

Data Transformation resamples and structures images into input arrays and class labels for model training. Dataset Splitting divides the data using an 80:20 training-testing ratio, ensuring that model evaluation is performed on unseen samples. Model Training with CNN builds convolutional layers that extract features, pooling layers that reduce spatial dimensions, fully connected layers that learn class-discriminative patterns, and a Softmax output layer.

Model Evaluation measures performance using accuracy, precision, recall, F1-score, and confusion matrix. Prediction Generation applies the trained model to new input images. Result Visualization uses Matplotlib and Seaborn to compare actual versus predicted classifications through graphs and confusion matrices.



International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

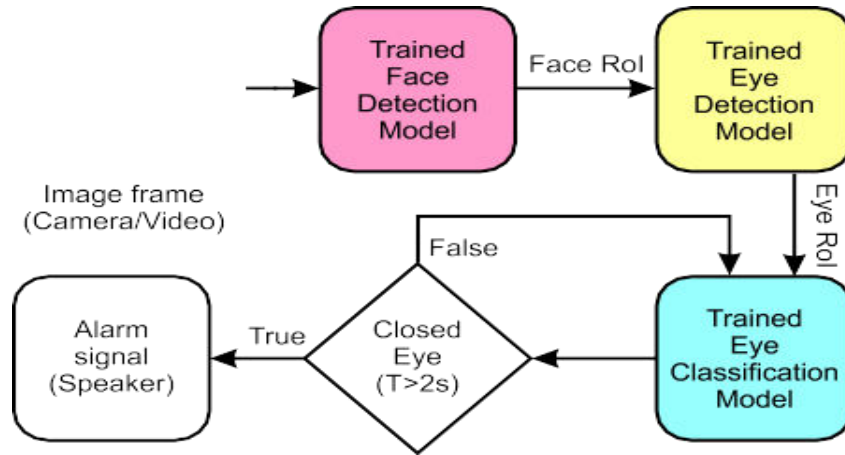


Fig. 1: Block diagram of the iris recognition pipeline motivating the proposed drowsiness detection architecture.

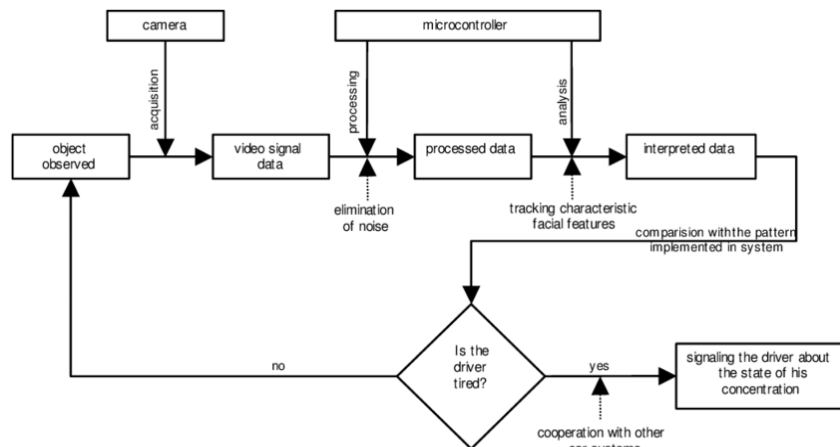


Fig. 2: General workflow of a vision-based driver monitoring system.

C. CNN MODEL ARCHITECTURE

In this implementation, the CNN architecture consists of convolution layers for feature extraction, pooling layers for spatial reduction, a flatten layer, fully connected dense layers, and a Softmax output layer. A dropout layer follows the convolution stack to reduce overfitting by randomly disabling a fraction of neurons during training, improving the model’s generalization to unseen data.

The model is compiled using categorical cross-entropy loss and the Adam optimizer. Training continues for multiple epochs until the model converges. The architecture is kept lightweight to allow deployment on standard hardware without requiring GPU acceleration for inference.



International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

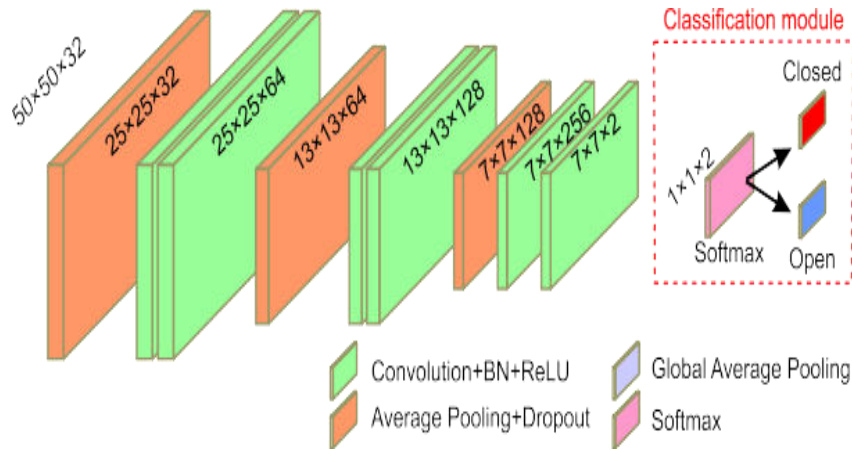


Fig. 5: CNN architecture showing convolutional feature extraction blocks and the softmax classification module.

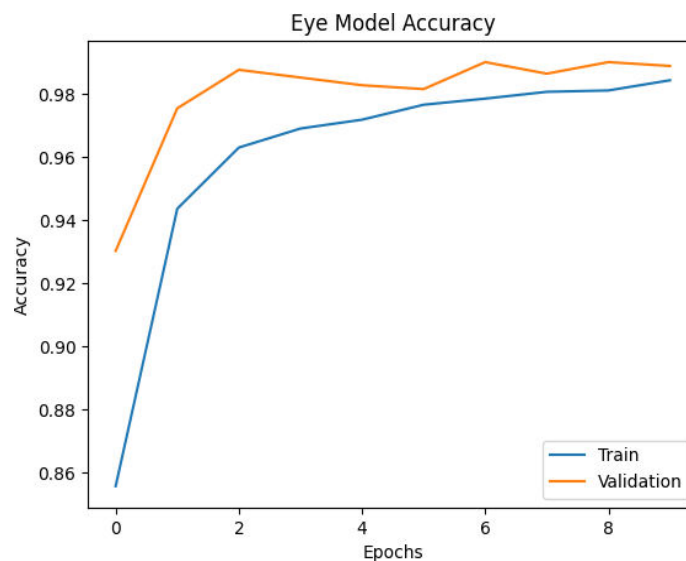
IV. RESULTS AND EVALUATION

The evaluation of the proposed system is divided into two components: eye state classification and yawning detection classification. Both models were trained on labeled datasets and tested on held-out samples using standard performance metrics.

A. EYE DETECTION MODEL RESULTS

TABLE III: EYE CLASSIFICATION REPORT

Class	Precision	Recall	F1-Score	Support
Closed Eye	0.99	0.99	0.99	349
Open Eye	0.99	0.99	0.99	469
Overall Accuracy	—	—	99%	818





International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

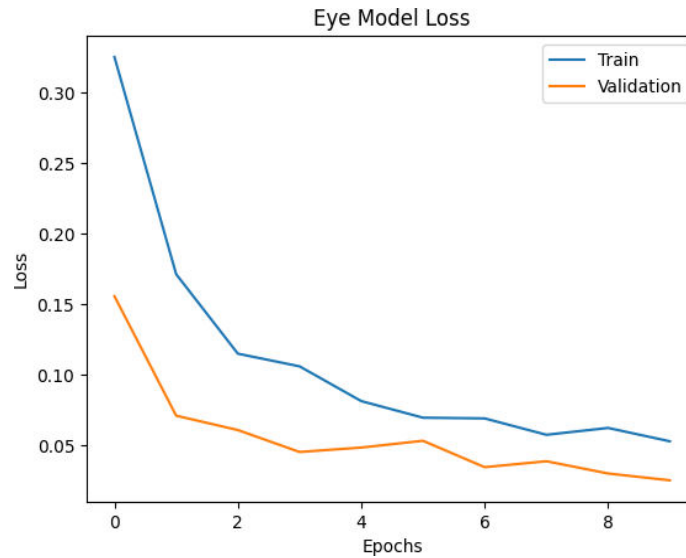


Fig. 6: Eye-state CNN model training and validation accuracy (left) and loss (right) across 10 epochs.

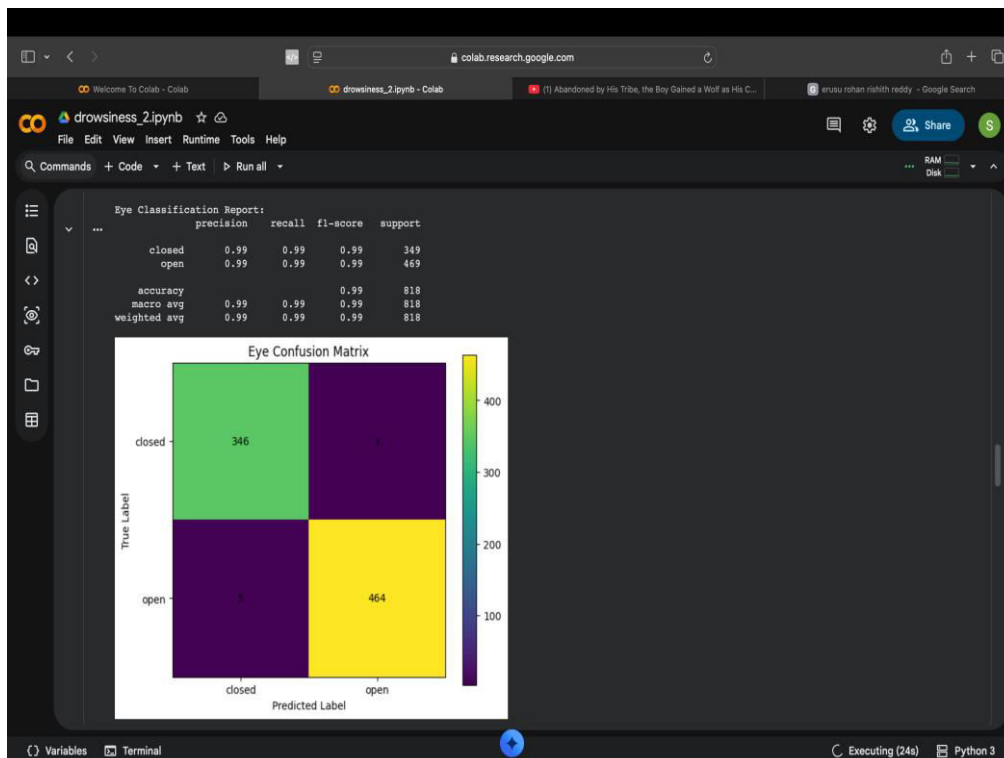


Fig. 7: Confusion matrix for eye-state detection model.

The eye detection model achieved exceptional performance with 99% accuracy. Only 8 out of 818 test images were misclassified (3 closed eyes misclassified as open, and 5 open eyes misclassified as closed). This high accuracy validates that the CNN model effectively learns the visual characteristics that distinguish open from closed eye states. The strong precision and recall values indicate that the model is both reliable and sensitive to eye closure—the primary indicator of driver drowsiness.



International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

B. YAWNING DETECTION MODEL RESULTS

TABLE IV: YAWNING CLASSIFICATION REPORT

Class	Precision	Recall	F1-Score	Support
No Yawning	0.81	0.92	0.86	1566
Yawning	0.89	0.75	0.81	1376
Overall Accuracy	—	—	84%	2942

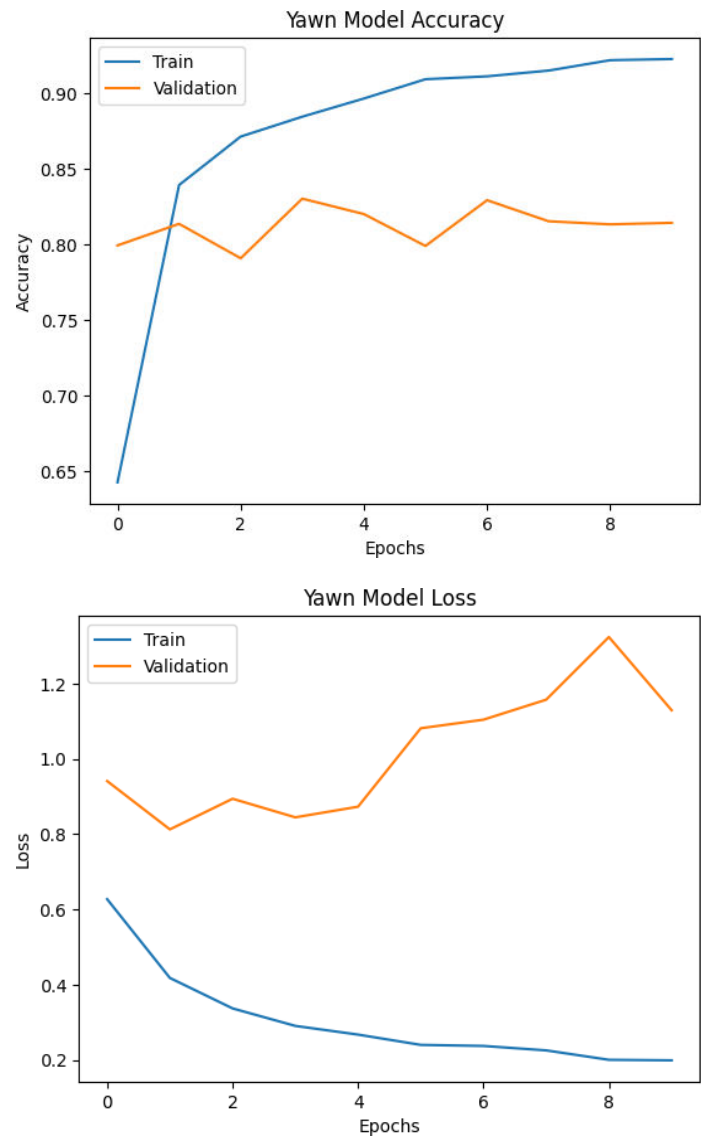


Fig. 8: Yawning CNN model training and validation accuracy (left) and loss (right) across 10 epochs.



International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

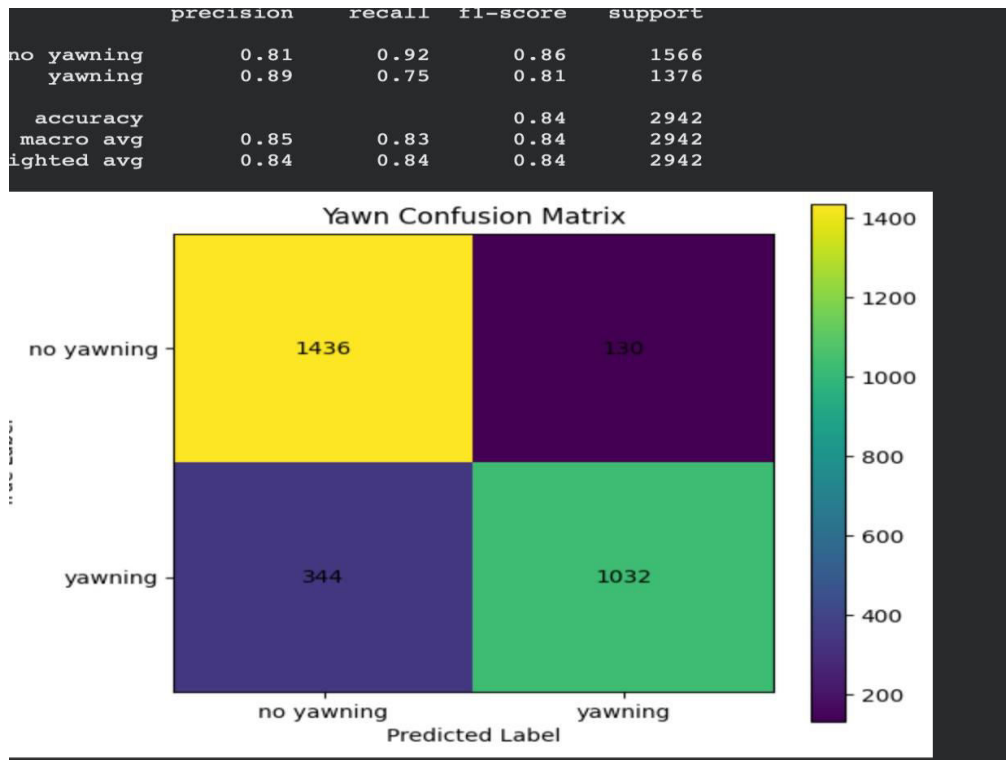


Fig. 9: Confusion matrix for yawning detection model.

The yawning detection model achieved an accuracy of 84%, which is lower than the eye model but considered strong for behavioral detection tasks. Detecting yawning presents greater challenges because mouth shapes vary significantly across individuals and are more sensitive to lighting conditions and head orientation. The confusion matrix reveals 1436 correctly classified non-yawning cases and 1032 correctly classified yawning cases, with 344 yawning instances missed (false negatives). The higher false negative rate for yawning aligns with known difficulties in behavioral cue detection.

V. DISCUSSION

The results confirm that both CNN models are capable of learning meaningful patterns from image datasets associated with driver drowsiness. The eye detection model demonstrates near-perfect performance because eye closure is a visually distinctive and consistent feature across individuals and lighting conditions. The model’s 99% accuracy underscores its suitability as the primary fatigue indicator.

The yawning detection model’s 84% accuracy reflects the inherent difficulty of detecting behavioral patterns from static images. Yawning involves dynamic mouth movements that are better captured from video sequences. Nevertheless, the model’s ability to identify yawning in 75% of actual yawning instances is a meaningful contribution to improving overall system reliability when combined with the eye model.

The multimodal approach—combining both detection streams—is consistent with findings in related literature [12, 13], which show that multi-indicator systems significantly outperform single-feature detection. The use of separate models for each indicator allows the system to be modular, enabling independent updates to each component without affecting the overall pipeline. This architecture also supports future integration of additional fatigue indicators such as head pose estimation or gaze tracking.



International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

VI. CONCLUSION

This work presented a simulation-based multimodal driver drowsiness detection system using Convolutional Neural Networks trained on eye and yawning image datasets. The system was designed to analyze visual fatigue indicators and generate classification outputs with performance metrics for evaluation.

The eye detection model achieved 99% accuracy, demonstrating exceptional reliability in distinguishing open and closed eye states. The yawning detection model achieved 84% accuracy, providing a complementary fatigue indicator that enhances overall system robustness. Together, these models form a lightweight, non-intrusive detection pipeline that does not require physiological sensors or specialized hardware.

The proposed system demonstrates that integrating multiple visual fatigue indicators using deep learning is a practical and effective approach to driver drowsiness detection. Future work may incorporate real-time video processing, LSTM-based temporal analysis of sequential frames, deployment on embedded systems, and integration with vehicle alert systems to further enhance road safety.

REFERENCES

- [1] M. Eriksson and N. Papanikolopoulos, "Eye-tracking for detection of driver fatigue," IEEE Intelligent Transportation Systems Conference, 1997.
- [2] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, 1997.
- [3] S. Singh and N. Papanikolopoulos, "Monitoring driver fatigue using facial analysis techniques," IEEE Intelligent Transportation Systems Conference, 1999.
- [4] G. Bradski, "The OpenCV Library," Dr. Dobb's Journal of Software Tools, 2000.
- [5] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," Advances in Neural Information Processing Systems, 2012.
- [6] S. Abtahi, B. Hariri, and S. Shirmohammadi, "Driver drowsiness monitoring based on yawning detection," IEEE International Instrumentation and Measurement Technology Conference, 2011.
- [7] F. Chollet, Deep Learning with Python, 2nd ed.; Manning Publications, 2021.
- [8] L. M. Bergasa et al., "Real-time monitoring of driver vigilance," IEEE Transactions on Intelligent Transportation Systems, 2006.
- [9] C. H. Weng et al., "CNN-based driver drowsiness detection," IEEE Transactions on Human-Machine Systems, 2016.
- [10] B. Baheti et al., "Detection of distracted driver using convolutional neural network," CVPR Workshops, 2018.
- [11] M. Patel et al., "Yawning detection for driver fatigue monitoring," International Journal of Engineering Research, 2019.
- [12] Y. Zhang et al., "Multimodal driver fatigue detection using deep learning," IEEE Access, 2021.
- [13] X. Wang et al., "Vision-based driver drowsiness detection using deep learning," Sensors, 2022.
- [14] World Health Organization, Global Status Report on Road Safety, 2023.
- [15] Google, MediaPipe: A Framework for Building Multimodal Applied ML Pipelines, 2020.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  ijircce@gmail.com



www.ijircce.com

Scan to save the contact details