
VIDEO-DRIVEN MONITORING OF DISTRACTED DRIVING USING DEEP LEARNING TECHNIQUES

S. Chanti*, Ch. Poojitha Ramya, Ch. Hanish, D. Siva Karthik, K. Leela Prasad

Dept. of Information Technology, GMR Institute of Technology, India.

Article Received: 11 March 2026

*Corresponding Author: S. Chanti

Article Revised: 31 March 2026

Dept. of Information Technology, GMR Institute of Technology, India.

Published on: 21 April 2026

DOI: <https://doi-doi.org/101555/ijarp.8115>

ABSTRACT:

Driver distraction is a major cause of road accidents and poses a serious threat to road safety. This project presents a deep learning-based system for detecting distracted driver behavior using both images and video inputs captured from a vehicle's dashboard camera. The system aims to automatically analyze driver actions and classify them as safe or distracted driving behaviors in real time. The proposed approach processes input data by resizing and normalizing images, while video clips are handled by extracting frames at regular intervals using OpenCV. Each frame is preprocessed to match the training conditions before being passed into the model. A Convolutional Neural Network (CNN) is implemented using transfer learning with a pre-trained VGG16 model to improve classification accuracy and reduce training complexity. The model classifies driver behavior into multiple categories such as safe driving, texting, phone usage, drinking, and other distractions. For video-based prediction, frame-wise classifications are aggregated using a majority voting technique to determine the final driver state. The system is integrated into a full-stack web application using Next.js for the frontend and Flask for the backend, enabling users to upload inputs and receive instant predictions. Experimental results demonstrate that the proposed system achieves reliable performance with efficient real-time detection, highlighting its potential for enhancing road safety through intelligent driver monitoring.

KEYWORDS: *Driver Distraction, Road Safety, Deep Learning, Convolutional Neural Network (CNN), Transfer Learning, VGG16, Video Processing, OpenCV, Full-Stack Web Application, Real-Time Monitoring.*

INTRODUCTION:

Road accidents caused by driver distraction have become a critical concern in modern transportation systems, leading to severe injuries, loss of life, and economic damage. With the increasing use of mobile phones, in-vehicle infotainment systems, and other distractions, ensuring driver attentiveness is essential for improving road safety. Traditional monitoring methods rely heavily on manual observation or sensor-based systems, which are often costly and inefficient.

Recent advancements in deep learning and computer vision have enabled automated, accurate, and real-time analysis of driver behavior using visual data. Video-based driver monitoring systems can capture facial expressions, hand movements, and posture to identify distracted driving activities. By leveraging Convolutional Neural Networks (CNNs) and transfer learning techniques, high accuracy can be achieved even with limited data. This project focuses on developing a deep learning-based driver distraction detection system using dashboard camera videos to classify driver behaviors and promote safer driving environments.

In this context, the proposed system utilizes dashboard camera videos to monitor driver behavior continuously. Frames are extracted and processed to detect patterns associated with distraction, and classification models are applied to categorize activities such as texting, phone usage, or safe driving. This approach not only improves detection efficiency but also supports real-time implementation. Ultimately, the project aims to contribute to intelligent transportation systems by providing a reliable and scalable solution for enhancing driver safety and reducing accident risks.

Literature Survey:

Mehdi Sefidgar Dilmaghani et al. (2024) focused on improving driver monitoring systems using event camera technology. Their study introduced a neuromorphic approach that captures high-speed facial movements such as eye blinks, gaze, and head pose with low latency. The main objective was to optimize bias settings in event cameras to enhance system performance. They proposed new evaluation metrics to assess features like face tracking and gaze estimation. Results showed that proper bias tuning significantly improves stability and accuracy. However, the system requires careful configuration of parameters for different environments. This research highlights the potential of event-based vision in real-time driver monitoring systems.

Xin Zuo et al. (2024) focused on detecting driver distraction caused by mobile phone usage using multi-modality physiological signals. Their study developed a framework based on multiscale entropy (MSaE) and BiLSTM to analyze EEG, ECG, and EMG signals. The main goal was to capture the complexity of physiological responses and improve distraction detection accuracy. Results showed that combining multiple physiological signals improves performance compared to single-source data. However, the system depends on sensor-based data collection, which may not always be practical. This research emphasized the effectiveness of deep learning and multimodal data in understanding driver mental states.

Yuheng Sun et al. (2025) focused on noncontact multimodal sensing for driver distraction detection. Their study proposed a system that integrates behavioral features and physiological signals such as heart rate variability using remote photoplethysmography (rPPG). The main objective was to distinguish between different types of distractions, including fatigue and external interference. Results demonstrated high accuracy in real-time detection using RGB-IR camera systems. However, the system complexity increases due to multimodal integration. This research highlights the importance of combining visual and physiological data for robust and accurate driver monitoring systems..

Sara A. Alameen et al. (2023) focused on detecting driver drowsiness using deep learning-based spatiotemporal models. Their study developed a hybrid model combining 3D-CNN and LSTM to extract both spatial and temporal features from video sequences. The main goal was to improve the detection of drowsiness indicators such as eye closure, yawning, and head movements. Results showed high accuracy across multiple datasets, demonstrating the effectiveness of combining CNN and LSTM. However, performance may vary under challenging conditions like poor lighting or occlusion. This research emphasized the importance of temporal feature learning in real-time driver monitoring systems.

Zizheng Guo et al. (2026) focused on improving cross-domain driver distraction detection using deep learning. Their study proposed a Dynamic Adaptation Framework (DAF) that aligns both global and local features to handle variations between different datasets and environments. The main objective was to enhance model generalization across diverse real-world conditions. Results showed that the proposed method significantly outperforms existing approaches in accuracy, precision, and recall across multiple datasets (as seen in Table II on page 10). However, the model involves higher computational complexity due to multi-level alignment. This research highlights the importance of domain adaptation techniques in building robust and scalable driver monitoring systems.

METHODOLOGY:

The proposed system detects distracted driver behavior using a video-based deep learning approach. Video clips captured from a dashboard camera are processed by extracting frames at regular intervals using OpenCV. Each frame is resized, normalized, and zero-mean scaled to match training conditions. A Convolutional Neural Network (CNN) with transfer learning is applied, utilizing a pre-trained VGG16 model for efficient feature extraction. Additional fully connected layers classify driver behavior into ten categories, including safe driving, texting, phone usage, and passenger interaction. This method reduces training complexity and improves accuracy. Frame-level predictions are combined using a majority voting technique to generate a reliable video-level driver behavior classification.

The proposed system processes dashboard camera video by extracting frames and applying preprocessing techniques such as resizing, normalization, and augmentation. Transfer learning is utilized with pre-trained VGG16 model to extract deep features and classify driver behaviors. Frame-level predictions from both models are combined using a majority voting ensemble method to improve accuracy. Finally, aggregated frame predictions generate the overall video-level classification as safe or distracted driving.

Data Collection and Preprocessing

Video data is collected using a dashboard camera capturing real-time driver behavior under various conditions. The recorded video is segmented into frames at regular intervals using OpenCV for detailed analysis. Each frame is resized to a fixed input size (e.g., 224×224) to match deep learning model requirements. Pixel values are normalized and zero-mean scaled to ensure consistency and faster convergence during training. Noise reduction techniques are applied to improve image quality, while data augmentation methods such as rotation, flipping, and brightness adjustment are optionally used to increase dataset diversity and enhance model generalization.

Training the model

The training phase utilizes transfer learning with pre-trained deep learning models such as VGG16 and MobileNet to efficiently extract high-level features from the input images. The top layers of these models are modified and fine-tuned using the prepared dataset to adapt them to driver behavior classification tasks. The models are trained on labeled data representing different driving activities, optimizing parameters through backpropagation and minimizing loss using suitable optimizers like Adam. Batch processing and multiple epochs

are employed to improve learning and model accuracy. Regularization techniques such as dropout are applied to prevent overfitting, ensuring better generalization on unseen data.

Architecture Design:

The architecture design of the proposed driver distraction detection system is based on a deep learning framework that integrates transfer learning and ensemble techniques for improved performance. Pre-trained convolutional neural networks such as VGG16 and MobileNet are used as the backbone models due to their proven efficiency in image classification tasks. The input video is first converted into frames, and each frame is preprocessed and fed into both models. The top layers of these pre-trained networks are replaced with custom fully connected layers to adapt them to the specific task of driver behavior classification. Transfer learning enables the system to leverage previously learned features, reducing training time and improving accuracy even with limited datasets. Each model independently extracts high-level features and generates predictions for different driving activities.

To enhance the robustness of the system, an ensemble approach using majority voting is implemented. The predictions from VGG16 and MobileNet at the frame level are combined, and the final class is determined based on the majority decision. This reduces the chances of misclassification and improves overall reliability. Furthermore, frame-level predictions are aggregated over time to generate a video-level classification, ensuring consistency in the final output. The architecture is designed to handle real-time inputs efficiently while maintaining high accuracy. By combining the strengths of multiple models and leveraging transfer learning, the system achieves a balance between computational efficiency and performance, making it suitable for real-world driver monitoring applications.

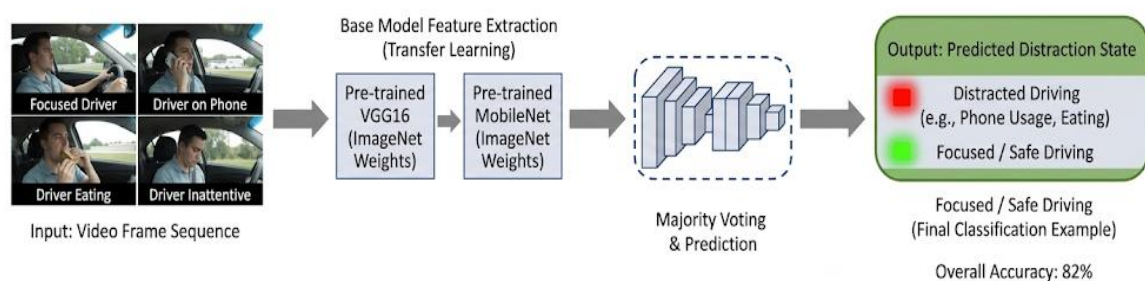


Figure 2: Distracted Driver Detection.

Evaluation and Testing

The evaluation and testing of the proposed driver distraction detection system are conducted to assess its accuracy and reliability. The dataset is divided into training and testing sets to evaluate performance on unseen data. Metrics such as accuracy, precision, recall, and F1-score are used to measure classification effectiveness across different driver behavior categories. The performance of individual models, VGG16 and MobileNet, is compared with the ensemble majority voting approach to verify improvements.

Additionally, confusion matrices are analyzed to identify misclassification patterns and improve model performance. The system is also tested using real-time video inputs to ensure its practical applicability. These evaluations demonstrate the robustness and effectiveness of the proposed system in real-world driving scenarios.

Applications in Real-Time

The proposed system can be deployed in real-time driver monitoring systems to detect distracted driving and enhance road safety. It can be integrated into smart vehicles to provide instant alerts and warnings to drivers. The system supports intelligent transportation systems by enabling continuous monitoring. It also helps in preventing accidents by identifying risky driver behaviors in real time.

Real-Time Driver Monitoring Systems

The system continuously monitors driver behavior and detects distractions in real time. This enables quick intervention to improve safety and reduce accidents.

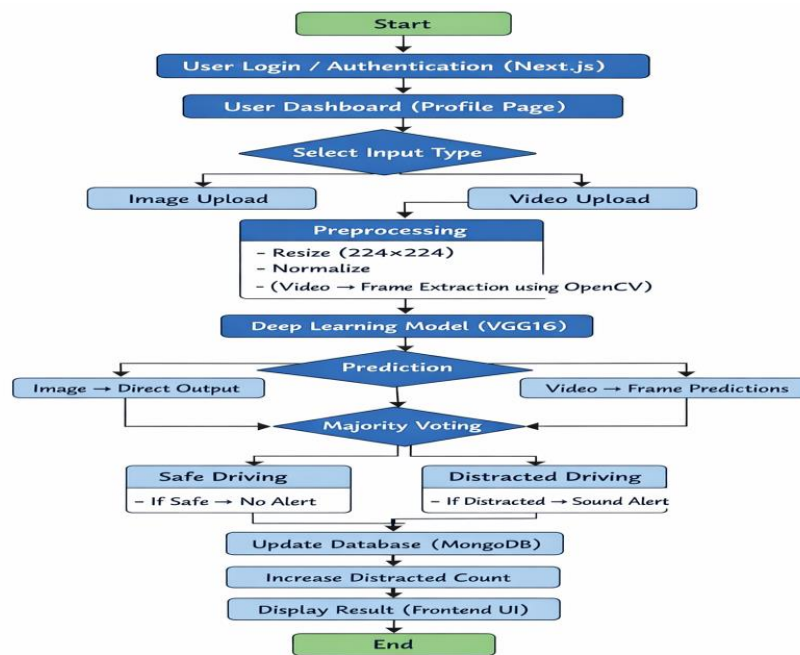
Smart Vehicle Integration and Alerts

The model can be integrated into vehicles to provide instant alerts for unsafe behavior. It helps drivers regain attention and maintain safe driving conditions.

Intelligent Transportation and Safety Systems

The system supports large-scale monitoring of driver behavior across vehicles. This contributes to improved traffic safety and reduced accident rates.

Flow Diagram



RESULTS:

The experimental results demonstrate that the proposed distracted driver detection system effectively identifies driver behavior using deep learning techniques. The use of the VGG16 model significantly improves classification accuracy by extracting robust visual features from input images and video frames. Furthermore, the implementation of a majority voting algorithm for video analysis enhances prediction reliability by minimizing errors from individual frame predictions. The integration of the model with a full-stack web application ensures smooth real-time interaction, allowing users to upload inputs and receive immediate feedback. Additionally, the system successfully stores and tracks distracted driving instances in the database. Overall, the system proves to be efficient, scalable, and suitable for real-time applications, with potential improvements achievable by incorporating larger datasets and more advanced architectures.

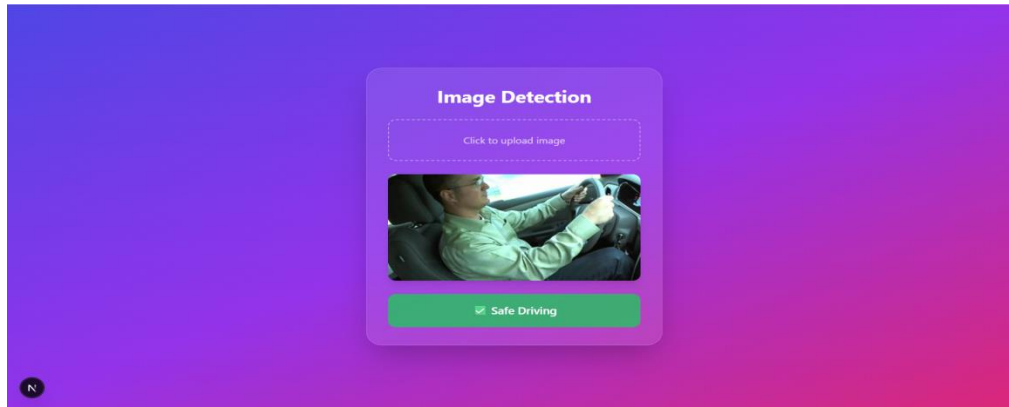


Figure 3: Image Detection (Safe Driving)

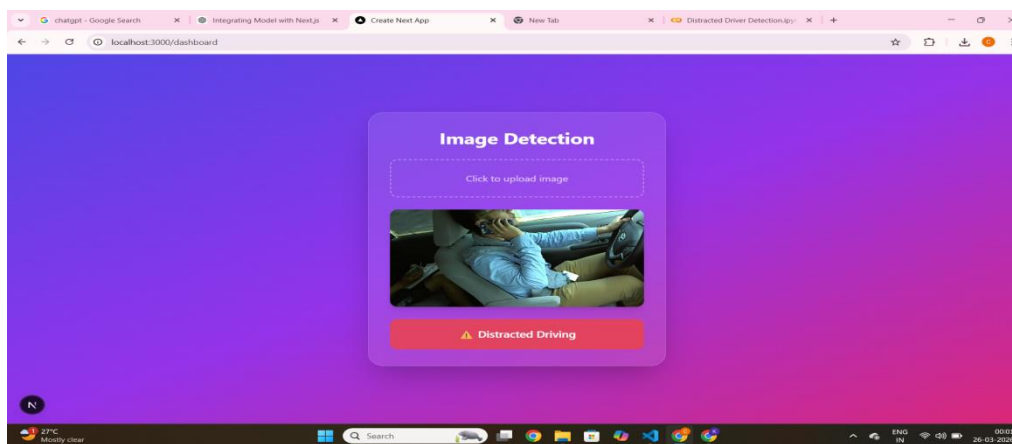


Figure 4: Video Detection (Distracted Driving)

CONCLUSION:

The proposed system for video-driven monitoring of distracted driving effectively demonstrates the use of deep learning to enhance road safety. By analyzing dashboard camera video inputs, the system accurately classifies driver behaviors such as safe driving and various distractions. The integration of computer vision enables automated, real-time monitoring, reducing reliance on manual observation. The use of MobileNet ensures fast and efficient predictions suitable for real-time applications, while VGG16 improves feature extraction and classification accuracy. Additionally, the majority voting algorithm combines predictions from both models, enhancing overall performance and minimizing errors. Experimental results show that the ensemble approach outperforms individual models in terms of accuracy. Although the system has minor limitations, such as sensitivity to lighting conditions and computational requirements, it remains a scalable and effective solution. With future improvements like advanced models and fatigue detection, the system can further contribute to safer driving environments.

REFERENCES:

1. Mehdi Sefidgar Dilmaghani, Waseem Shariff, Muhammad Ali Farooq, Joe Lemley, and Peter Corcoran, “*Optimization of Event Camera Bias Settings for a Neuromorphic Driver Monitoring System*,” *IEEE Access*, vol. 12, pp. 32959–32970, 2024.
2. Xin Zuo, Chi Zhang, Fengyu Cong, Jian Zhao, and Timo Hämäläinen, “Mobile Phone Use Driver Distraction Detection Based on MSaE of Multi-Modality Physiological Signals,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 11, pp. 17650–17663, Nov. 2024.
3. Yuheng Sun, Mingyue Zhou, Fengyi Cai, Xiaoqian Yang, Changping Li, and Ying Yi, “*Noncontact Multimodal Sensing for Driver Distraction Detection via Behavioral-Physiological Fusion*,” *IEEE Sensors Journal*, vol. 25, no. 20, pp. 38576–38584, 2025.
4. Sara A. Alameen and Areej M. Alhothali, “A Lightweight Driver Drowsiness Detection System Using 3D-CNN With LSTM,” *Computer Systems Science & Engineering*, vol. 44, no. 1, pp. 895–910, 2023.
5. Zizheng Guo, Qing Liu, Guofa Li, Xiyuan Luo, Guanglei Wang, and Cristina Olaverri-Monreal, “*A Dynamic Multi-Level Feature Alignment Method for Domain Adaptive Driver Distraction Detection*,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 27, no. 1, pp. 1279–1290, 2026.
6. F. Sajid, A. R. Javed, A. Basharat, N. Kryvinska, A. Afzal, and M. Rizwan (2021), “An Efficient Deep Learning Framework for Distracted Driver Detection,” *IEEE Access*, vol. 9, pp. 169270–169280.
7. Y.-S. Poon, C.-Y. Kao, Y.-K. Wang, C.-C. Hsiao, M.-Y. Hung, Y.-C. Wang, and C.-P. Fan (2021), “Driver Distracted Behavior Detection Technology with YOLO-Based Deep Learning Networks,” 2021 IEEE .
8. O. N. Neamah, T. A. Almohamad, and R. Bayir (2024), “Enhancing Road Safety: Real-Time Distracted Driver Detection Using Nvidia Jetson Nano and YOLOv8,” 2024 .
9. P. Wang, H. Li, B. Zhang, and M. Xia (2024), “Metric-based few-shot learning method for driver distracted behaviors detection,” *IEEE, Conference Publication*.
10. K. Anusha, B. Rikshith, M. Chandra Kanth, and M. Maheshwar (2024), “Real-Time Distracted Driver Detection and Monitoring using Computer Vision,” *IEEE, Conference Publication*.