Automated Driver Drowsiness Detection System

Princekumar Gupta[1], Dr. Archana Gupta[2]

[1]Department of Computer Engineering, K.J Somaiya College of Engineering, Mumbai, India.

[2]Professor, Department of Computer Engineering, K.J Somaiya College of Engineering, Mumbai, India. [1]princekumar.g@somaiya.edu , [2]archana.gupta@somaiya.edu

Abstract - Road accidents caused by distracted and drowsy driving continue to pose significant global safety challenges. This research presents an intelligent, deep learning-based system designed to detect and mitigate such behaviors in real-time. Utilizing dashboard camera footage and webcam input, the system performs dual tasks: distraction classification and drowsiness detection. Distraction detection leverages the State Farm Distracted Driver Detection dataset to categorize driver activities into ten classes, using both traditional Convolutional Neural Networks (CNNs) and transfer learning models like VGG16 and EfficientNet-B2. An innovative integration of an attention mechanism within EfficientNet-B2 enhances the model's focus on salient image regions, thereby improving classification performance. For drowsiness detection, facial and ocular cues are continuously analyzed using image processing techniques to compute the Percentage of Eye Closure (PERCLOS), with high blinking frequency triggering real-time alerts. The system further adapts to individual driver behavior and environmental variability through intelligent learning strategies. Designed for real-time execution with low latency, the proposed solution demonstrates strong accuracy and practical applicability. By integrating advanced deep learning methods with adaptive monitoring, this work contributes toward safer driving environments and lays foundational support for future autonomous vehicle technologies.

Keywords: Distracted Driving, Drowsiness Detection, CNN, VGG16, EfficientNet-B2, Face Detection, Eye Tracking, Deep Learning, Attention Mechanism.

1. INTRODUCTION

Road safety remains a major global concern, with over 1.3 million lives lost annually in traffic-related accidents. A significant portion of these incidents is due to human error, particularly from driver distraction and drowsiness. These factors impair a driver's awareness and reaction time, making real-time detection systems increasingly essential as vehicles become faster and more autonomous.

Distracted driving involves activities like mobile phone usage, interacting with infotainment systems, eating, or conversing with passengers each diverting attention from the road. As smart vehicle features grow, so do visual and cognitive distractions, increasing the risk of accidents, especially among young drivers. Proactive monitoring systems are needed to detect and respond to such behaviors instantly.

Drowsiness, while less obvious, is equally dangerous. Long drives, night shifts, or lack of rest often lead to fatigue that shows through subtle signs like eye closure, yawning, or head nodding. Traditional vehicle-based methods are reactive and may fail to detect early fatigue. This underscores the need for visual and physiological analysis through facial landmarks and eye tracking.

To address both problems, this study presents a unified deep learning-based system for real-time detection of drowsiness and distraction. It uses CNN models like VGG16 and EfficientNet-B2 with attention mechanisms for distraction classification, and PERCLOS-based eye closure tracking for drowsiness detection. The system issues real-time alerts, improving driver safety and supporting future applications in autonomous vehicles and smart transportation systems.

2. DESCRIPTION OF PROBLEM

Distracted and drowsy driving continues to be a major cause of road accidents worldwide, despite advancements in vehicle safety technologies. Traditional systems like lane departure warnings are often reactive and fail to detect early signs of driver inattention. Distractions from mobile phones, infotainment systems, or passengers reduce focus and delay response times, while drowsiness—especially during night driving—can lead to dangerous micro-sleeps without warning. Existing solutions struggle to detect these subtle indicators. To overcome this, the proposed research introduces a deep learning-based system that analyzes facial landmarks, eye movements, and driver behavior using CNN models with attention mechanisms. The system aims to detect both distraction and fatigue in real time and provide timely alerts, improving overall road safety.

3. MOTIVATION

With the increasing number of road accidents caused by distracted and drowsy driving, there is an urgent need for intelligent systems that can monitor driver behavior in real time. Traditional safety methods are often delayed or ineffective in detecting early signs of distraction or fatigue. This motivated the development of a deep learning-based solution that can proactively recognize unsafe driving patterns and help prevent accidents, ultimately contributing to safer roads and smarter transportation systems.

4. MAIN CONTRIBUTIONS

This research presents a unified, real-time driver monitoring system that effectively detects both distracted and drowsy driving behaviors. The system leverages deep learning models, specifically VGG16 and EfficientNet-B2, with transfer learning to classify ten distinct driver activities using the State Farm Distracted Driver Detection dataset. To enhance accuracy, an attention mechanism is integrated, allowing the model to focus on the most relevant regions within an image. For drowsiness detection, the system employs facial landmark detection to calculate the Percentage of Eye Closure (PERCLOS), enabling early identification of fatigue symptoms. The entire solution is deployed through a Flask-based web interface, supporting live webcam or video feed processing for real-time monitoring and alert generation. This comprehensive approach improves road safety by providing proactive, intelligent detection of unsafe driving behaviors.

5. RESEARCH LANDSCAPE OVERVIEW

The problem of driver drowsiness and distraction has gained significant research attention in recent years due to its direct correlation with road accidents and fatalities. Researchers have explored multiple modalities for monitoring driver alertness, including physiological signals, facial behavior analysis, and driving pattern monitoring. Among these, EEG and ECG-based methods are noted for their high accuracy in detecting fatigue. Fu et al. [1] proposed a federated learning approach using EEG signals, while Lee et al. [2] developed a wearable ECG/PPG system that detects drowsiness through unique recurrence plot patterns. Although effective, such physiological methods often require expensive and intrusive equipment, making them unsuitable for real-time, in-vehicle environments.

In contrast, vision-based approaches offer a non-intrusive alternative by analyzing facial features such as eye blink rate, head pose, and mouth movements. Deng and Wu [3] proposed a real-time detection system using facial features to identify signs of drowsiness, achieving reliable performance under standard conditions. Similarly, Safarov et al. [9] leveraged eye-blink detection and lip-boundary analysis using CNN-RNN models to detect fatigue with high accuracy. These methods are generally lightweight and

easier to deploy in real-time systems using only a standard webcam, but are still vulnerable to lighting changes and occlusions like sunglasses.

Behavioral and vehicular methods such as monitoring steering wheel movements, lane deviation, and speed patterns have also been explored [4], [5]. While these techniques provide valuable supplementary information, they are reactive rather than proactive, and their effectiveness is often reduced under varying road or vehicle conditions.

To overcome the limitations of individual methods, researchers have proposed hybrid systems that fuse multiple data sources. Ramzan et al. [2] and Nasri et al. [3] emphasized the importance of combining facial recognition, physiological signals, and vehicular telemetry to improve accuracy and robustness. Such systems can compensate for weaknesses in any one modality, offering a more comprehensive view of the driver's condition.

Deep learning has revolutionized the field of driver monitoring. CNN-based architectures, such as VGG16 and EfficientNet, have been widely adopted for image classification tasks, including detecting distracted driver behaviors [9], [12]. Qu et al. [6] highlighted how lightweight deep models such as MobileNet and EfficientNet are enabling real-time applications even on constrained hardware. Moreover, the inclusion of attention mechanisms, as in your proposed work, improves focus on critical facial regions and boosts classification accuracy.

Furthermore, embedded system implementations, such as those discussed by Zain et al. [10] and Titare et al. [11], demonstrate the practical feasibility of deploying drowsiness detection systems on platforms like Raspberry Pi using computer vision libraries such as OpenCV and Dlib. These systems compute real-time metrics like Eye Aspect Ratio (EAR) and Percentage of Eye Closure (PERCLOS), which are reliable indicators of drowsiness and yawning.

Distraction detection methods have also evolved. Kashevnik et al. [8] proposed a sensor-based framework for classifying manual, visual, and cognitive distractions. Their approach used YOLO and data fusion methods to detect behaviors like texting, phone use, and looking away from the road all of which are covered in the ten-class classification used in your project via the State Farm dataset.

Lastly, regulatory perspectives like those discussed in [7] and [15] highlight the growing demand for Advanced Driver Assistance Systems (ADAS) in the automotive industry, especially in commercial vehicles. As safety regulations evolve, systems like yours that combine fatigue and distraction detection will be essential components of intelligent transportation systems.

6. LIMITATION IN EXISTING SYSTEM VS OUR SYSTEM

Compared to existing driver monitoring systems, the proposed solution offers several key advancements in functionality, practicality, and performance. Most existing systems are limited to detecting either drowsiness or distraction individually, whereas our system is designed to identify both behaviors within a single integrated framework. Traditional systems often rely on additional hardware such as EEG sensors, IR cameras, or wearable devices, which can be expensive and intrusive. In contrast, our approach uses only a standard webcam, making it more accessible and non-intrusive for real-world deployment.

In terms of real-time capability, many existing models suffer from high latency or are designed for offline analysis. Our system, however, processes input frames in real time and provides instant feedback to the driver through audio alerts and optional email notifications. While traditional systems typically detect only a few distraction types such as texting or phone use, our model supports classification of ten distraction behaviors, including drinking, operating the radio, and interacting with passengers.

Drowsiness detection in many systems relies on reactive methods such as steering pattern analysis or lane deviation, which may not capture early signs of fatigue. Our solution leverages proactive, vision-based indicators like Eye Aspect Ratio (EAR), lip distance, and head posture to detect drowsiness before it leads to critical situations. Additionally, the use of EfficientNet-B2 with an integrated Convolutional Block Attention Module (CBAM) improves model accuracy by focusing on the most relevant features within the video frames.

Lastly, while many existing systems lack a complete user interface and are confined to lab setups, our model is implemented through a Flask-based web application that supports both video uploads and live webcam input. This makes the system lightweight, user-friendly, and scalable for broader adoption in real driving environments.

7. CONTEXT OF THE STUDY

Driving is a complex activity that demands continuous attention, quick decision-making, and physical alertness. However, in today's fast-paced world, drivers often face fatigue, mental stress, and various distractions while on the road. Long driving hours, night shifts, and monotonous highways can easily lead to drowsiness, while distractions like mobile phones, infotainment systems, and passenger interaction contribute to lapses in focus.

Traditional vehicle safety systems such as airbags, ABS, and lane departure warnings are reactive they come into play after a risky situation has already occurred. While some vehicles now include basic driver monitoring systems, many are still limited to single-parameter detection like steering behavior or lane keeping. These approaches fail to capture subtle physiological and behavioral signs of inattention.

With advancements in computer vision and deep learning, it has become feasible to analyze a driver's face, eyes, and behavior using simple camera inputs. These technologies enable real-time detection of risky states like drowsiness and distraction before an accident occurs. Moreover, such systems can be implemented with minimal hardware, making them cost-effective for wide adoption.

This study leverages the power of CNN-based deep learning models and facial landmark detection to develop a dual-purpose driver monitoring system. The goal is to identify both drowsy and distracted behaviors with high accuracy and low latency, while providing instant alerts to prevent accidents. The system is designed to be lightweight, practical, and adaptable for use in real driving conditions, contributing to the broader vision of intelligent and safer transportation.

8. RESEARCH METHODOLOGY

The proposed research adopts a vision-based, deep learning approach to detect driver drowsiness and distraction in real time. The methodology begins with data acquisition, followed by preprocessing, model training, facial landmark analysis, and finally, system integration for deployment.

For distraction detection, the State Farm Distracted Driver Detection dataset was used, containing images categorized into ten classes such as texting, talking on the phone, drinking, and reaching behind. Each image was resized to 128×128 pixels and normalized to ensure uniform input for the model. For drowsiness detection, facial landmarks were extracted using a standard webcam and the Dlib library. Key regions like eyes and lips were tracked to identify signs of fatigue through Eye Aspect Ratio (EAR) and lip distance measurements.

Two deep learning models were used: VGG16 and EfficientNet-B2, both pre-trained on ImageNet and fine-tuned on our dataset using transfer learning. To enhance performance, an attention mechanism was integrated into EfficientNet-B2, allowing the model to focus on the most relevant parts of the input image. The models were trained using the Adam optimizer with categorical cross-entropy loss, and SMOTE was applied to balance the class distribution in the dataset.

For drowsiness detection, a threshold-based approach was applied: if the EAR fell below 0.20 for multiple consecutive frames, drowsiness was flagged. Yawning was detected by measuring the vertical gap between the upper and lower lips, and head posture was analyzed using a slope-based method to detect backward glances or inattention.

The system was implemented using Flask, allowing users to upload videos or use live webcam feeds. Each video frame

is processed in real time for both distraction and drowsiness detection. On detecting risky behavior, the system generates a sound alert and optionally sends an email notification with the driver's snapshot.

Finally, the system was evaluated using metrics such as accuracy, precision, recall, F1-score, and a confusion matrix. The distraction model achieved an accuracy of 94.65%, demonstrating its effectiveness in real-world scenarios.

9. PRE-TRAINED MODEL TRANSFER LEARNING

CNN :

One type of deep learning model is Convolutional Neural Networks (CNNs). It is a significant and positive advancement in deep learning, particularly for problems involving image processing and classification. Naturally, they are made to automatically extract significant and valuable information from unprocessed image data, taking inspiration from the intricate workings of the human visual cortex. In essence, a CNN is composed of several layers that work together to gradually transform input images into predictions. The fundamental components of a CNN architecture are convolutional, pooling, and fully linked layers.

- **Convolutional Layers:** The fundamental building blocks of CNNs, these layers identify patterns and characteristics in the spatial content of the input images. Convolutional operations use tiny filters or kernels to swipe across the input image, computing dot products at each location. This is an excellent technique for highlighting and differentiating various features, which can include edges, textures, and forms. When we filter it collectively, the various filters in this layer are employed to find more sophisticated types of features.

- **Pooling Layers:** After the convolutional layers, pooling layers play a crucial role in resampling the feature maps created by the earlier layers. Naturally, two well-liked pooling techniques that minimize the spatial dimensions of feature maps without losing any crucial information are max pooling and average pooling. This step reduces processing complexity and helps prevent overfitting.

- **Fully Connected Layers:** Following the convolutional and pooling layers, fully connected layers will naturally incorporate the learned features before performing the final classification. The model is able to comprehend intricate relationships between various feature types since each neuron in these layers is obviously related to every other neuron in the layer before it.
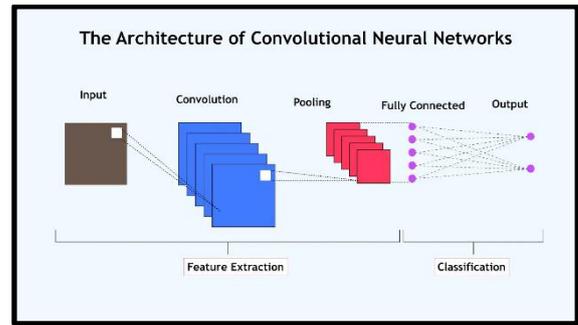

Figure 9.1: - CNN Architecture Diagram.

VGG16:

The Visual Geometry Group (VGG) at the University of Oxford created the deep Convolutional Neural Network architecture known as VGG16. It was first presented in Karen Simonyan and Andrew Zisserman's paper "Very Deep Convolutional Networks for Large-Scale Image Recognition" and placed first in the 2014 ImageNet Large Scale Visual Recognition Challenge (ILSVRC). VGG16 is already familiar with general features like edges, textures, and forms because of its ImageNet pretraining.
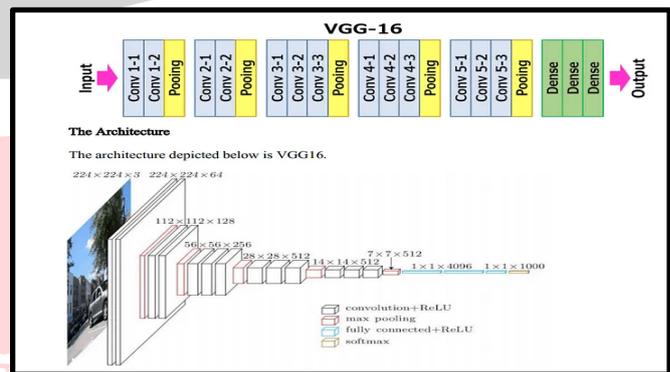

Figure 9.2: - VGG16 Architecture Diagram.

Efficient-Net B2

EfficientNetB2 is a member of the Efficient Net family of convolutional neural network (CNN) designs, which Google unveiled and showcased in 2019. On a range of picture classification applications, it demonstrates a scalable and efficient model architecture that genuinely preserves computational efficiency while attaining good and creative performance. EfficientNetB2 is an excellent option for real-world applications and scenarios with restricted resources since it strikes a compromise between model size and accuracy. With fewer parameters, EfficientNetB2 can execute traditional convolutional networks. The performance of EfficientNetB2 has been promising.
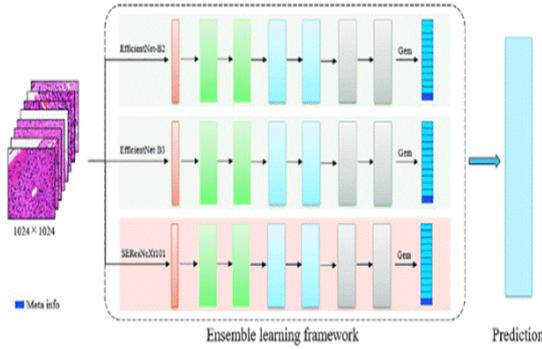
Figure 9.3: - EfficientNetB2 Architecture Diagram.

Efficient-Net B2 with Attention Mechanism

To enhance the model's focus and improve interpretability, a Convolutional Block Attention Module (CBAM) was integrated into the EfficientNet-B2 architecture. This attention mechanism applies both channel and spatial attention, enabling the model to concentrate more effectively on critical regions within the image, such as the driver's hands, face orientation, and presence of mobile phones. By guiding the network toward these relevant features, the modified model demonstrates improved performance across all evaluation metrics, including accuracy, precision, recall, and F1-score. Notably, it also reduces misclassification in challenging categories like "reaching behind" and "talking to passengers," where contextual cues are often subtle.

Figure 9.4: - EfficientNetB2 with Attention Mechanism.

Distracted Driving Behaviors

Dataset used: Roboflow
https://universe.roboflow.com/newaccount/project-part2/dataset/2

Size: 8688 Images

Classification (10 Classes)

Classes:

- Safe Driving
- Texting - Right

- Talking on Phone - Right
- Texting - Left
- Talking on Phone - Left
- Operating Radio
- Drinking
- Reaching Behind
- Hair/Makeup
- Talking to Passenger

Figure 9.5: - Dataset Images

Per class: 800 approx.
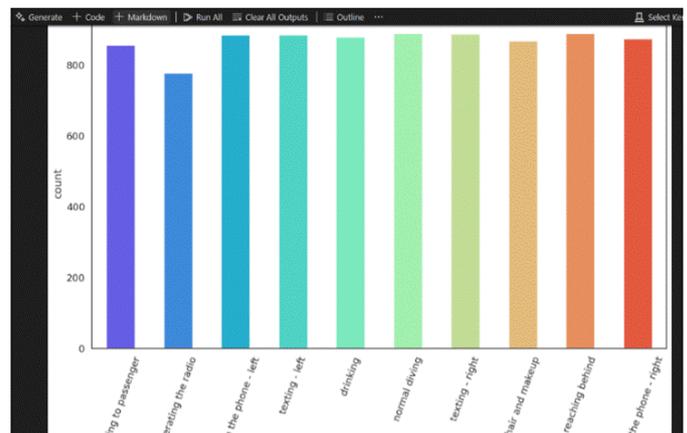
Figure 9.6: - Classification of dataset

10. OPERATIONAL WORKFLOW
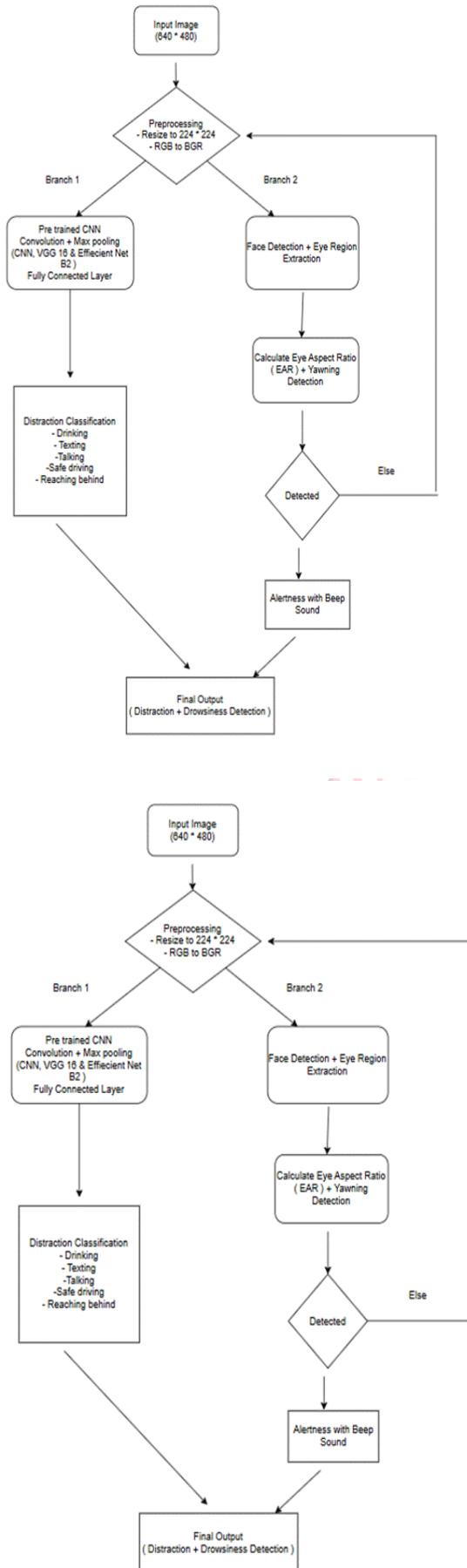

The operational workflow of the proposed driver monitoring system is designed to function in real-time and involves multiple coordinated stages. It begins with the acquisition of video input, which can either be a live feed from a webcam or a pre-recorded video uploaded through a user-friendly web interface developed using Flask. Once the video input is received, individual frames are extracted and passed through a preprocessing pipeline. In this step, each frame is resized to 128×128 pixels and normalized to ensure consistency with the model's input requirements. This preprocessing helps in reducing computational load and enhancing model performance.

For distracted driving detection, the preprocessed frames are fed into a fine-tuned EfficientNet-B2 deep learning model, which has been augmented with a Convolutional Block Attention Module (CBAM). This module helps the network to give more importance to relevant areas in the image, such as the driver's hands, face orientation, and the presence of distracting objects like mobile phones. The model classifies the driver's activity into one of ten predefined distraction categories such as texting, drinking, or talking to passengers.

Simultaneously, for drowsiness detection, the system utilizes Dlib's 68-point facial landmark detection to track the eyes, mouth, and head position. Using these landmarks, the system computes critical indicators such as the Eye Aspect Ratio (EAR) to monitor eye closure duration, lip distance to detect yawning, and head posture slope to determine if the driver is looking away or backward. If the EAR drops below a predefined threshold or the lip distance exceeds a certain limit over several frames, the driver is considered drowsy.

When either drowsiness or distraction is detected, the system instantly activates an audio alert (beep sound) to capture the driver's in drowsiness detection. If the unsafe behavior persists, an email notification is optionally sent to a pre-registered emergency contact. This email contains a snapshot image of the driver during the alert-triggering event, providing visual context for external monitoring.

All these operations are executed in real-time, ensuring immediate response and maximum road safety. The results, including predicted class labels and detection status, are overlaid on the video frames and displayed to the user via the web interface. This seamless integration of computer vision, deep learning, and web technologies creates an efficient, automated pipeline capable of improving driver safety through continuous, intelligent monitoring.

11. RESULTS

- **Model Accuracy attained:**

  The model achieved an overall accuracy of 94.65%, which demonstrates its effectiveness in detecting alertness and drowsiness in various

Figure 10.1: - System Architecture

lighting and angle conditions.

- **Loss:**

  The loss function is used to measure how far the model's predictions are from the actual labels. In the case of EfficientNet-B2, which is commonly used for multi-class classification, the model uses Categorical Cross-Entropy Loss.

  This loss function compares the predicted class probabilities with the true labels and penalizes incorrect predictions. The goal during training is to minimize this loss, so the model becomes more accurate over time.
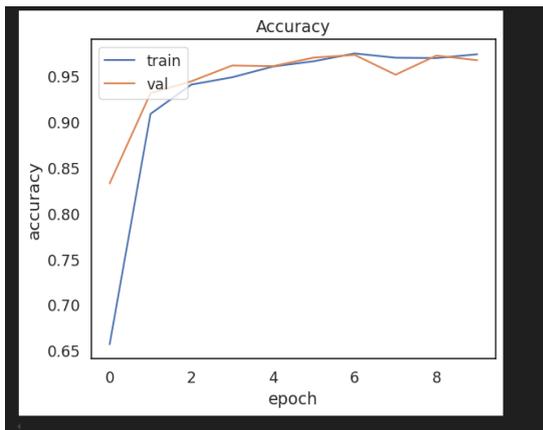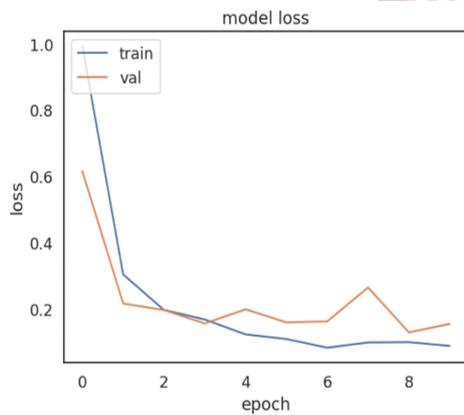

Figure 11.1: - Accuracy

Figure 11.2: - Model loss

Confusion matrix:

The confusion matrix is used to evaluate the performance of the classification model by comparing the predicted labels with the actual ground truth labels. It is a square matrix where rows represent the actual classes and columns represent the predicted classes.

Each cell in the matrix indicates how many times a particular class was predicted correctly or incorrectly. This allows for a detailed understanding of model performance, beyond just overall accuracy.
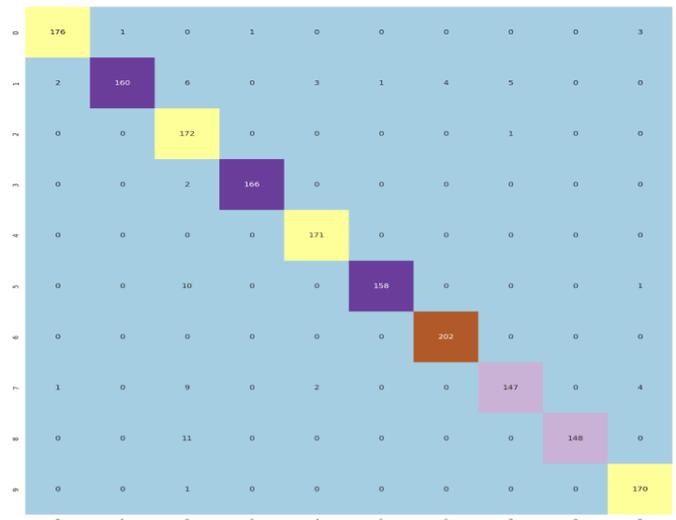

Figure 11.3: - Confusion matrix
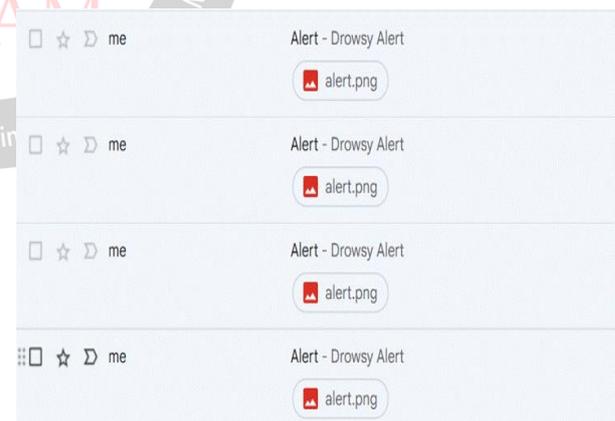

Figure 11.4: - Drowsiness detection

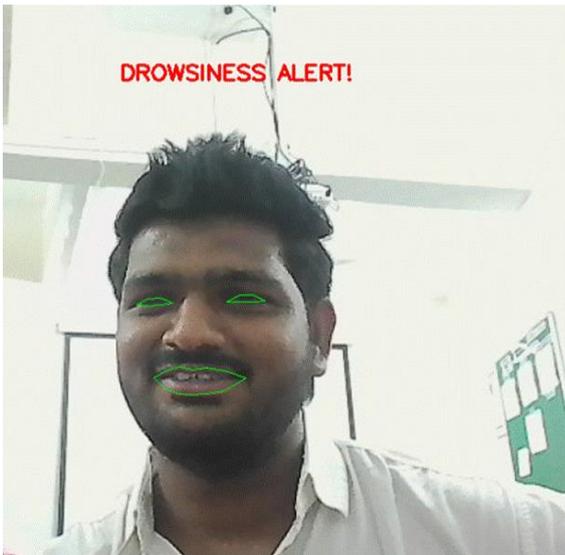Figure 11.5: - Snapshot of drowsiness alert via email

Figure 11.6: - Alerting on drowsiness detected

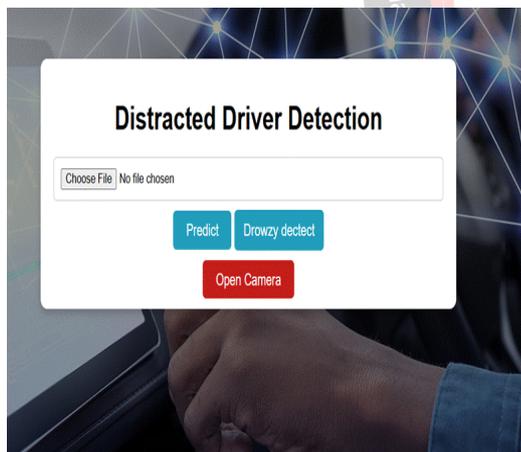Figure 11.7: - Alerting the driver while texting

Figure 11.8: - User Interface

12. SYSTEM TESTINGS AND VALIDATION

The proposed system was rigorously evaluated using a combination of real-time webcam inputs and pre-recorded video-based test scenarios to assess its accuracy in detecting both distracted and drowsy driving behaviors. The evaluation encompassed a variety of realistic driving situations, including normal driving, using a mobile phone, drinking beverages, yawning, prolonged eye closure, and reaching behind each simulating common causes of distraction or fatigue. The EfficientNet-B2 model integrated with an attention mechanism demonstrated high accuracy in classifying distraction categories, effectively distinguishing between different types of non-driving activities. Meanwhile, drowsiness detection was successfully achieved through continuous monitoring of facial landmarks. Key metrics such as Eye Aspect Ratio (EAR) and lip distance were computed in real time to identify early signs of fatigue, including eye closure and yawning. The system's performance in live testing scenarios validated its robustness and reliability, confirming its potential for real-world deployment in intelligent driver assistance systems.

Alerts were tested via audio beeps and optional email notifications on drowsiness detection with driver snapshots. The system responded within 1–2 seconds of detecting unsafe behavior. Functional testing also confirmed smooth operation of both real-time webcam mode and video upload through the Flask web interface. The system remained stable even under edge cases such as no face detected or fast head movements.

13. CONCLUSION

This research presents an intelligent, real-time driver monitoring system capable of detecting both distracted and drowsy driving behaviors. By integrating deep learning models such as VGG16 and EfficientNet-B2 with an attention mechanism, the system accurately classifies ten types of distracted driver activities. For drowsiness detection, facial landmarks and PERCLOS-based eye closure analysis enable the identification of early signs of fatigue. The entire solution operates using a standard webcam and is implemented through a user-friendly Flask-based web interface, making it both practical and scalable. Real-time alerts through sound and email notifications further enhance its effectiveness. The proposed system demonstrates high accuracy and responsiveness, contributing significantly to road safety and paving the way for integration into advanced driver assistance systems and intelligent transportation frameworks.

REFERENCES

[1] X. Qin, Y. Niu, H. Zhou, X. Li, W. Jia, and Y. Zheng, "Driver drowsiness EEG detection based on tree federated learning and interpretable network," Int. J. Neural Syst., vol. 33, no. 3, pp. 2350009:1– 2350009:17, 2023.

[2] H. Lee, J. Lee, and M. Shin, "Using wearable ECG/PPG sensors for driver drowsiness detection based on distinguishable pattern of recurrence plots," Electronics, vol. 8, no. 2, 2019.

[3] W. Deng and R. Wu, "Real-time driver-drowsiness detection system using facial features," IEEE Access, vol. 7, pp. 118727–118738, 2019.

[4] A. G. Wheaton, R. A. Shults, D. P. Chapman, E. S. Ford, and J. B. Croft, "Drowsy Driving and Risk Behaviors 10

States and Puerto Rico, 2011–2012," Morb. Mortal. Wkly. Rep., vol. 63, no. 26, pp. 557–562, Jul. 2014.

[5] B. C. Tefft, "Prevalence of motor vehicle crashes involving drowsy drivers, United States, 1999– 2008," Accid. Anal. Prev., vol. 45, pp. 180–186, Mar. 2012, doi: 10.1016/j.aap.2011.05.028.

[6] P. Philip et al., "Fatigue, sleep restriction and driving performance," Accid. Anal. Prev., vol. 37, no. 3, pp. 473–478, May 2005

[7] Union, E. Regulation (eu) 2019/2144 of the European parliament and of the council of ministers (2019).

[8] Macrae, C. Learning from the failure of autonomous and intelligent systems: Accidents, safety, and sociotechnical sources of risk. Risk Anal. 42, 1999–2025 (2022).

[9] Kaplan, S., Guvensan, M. A., Yavuz, A. G. & Karalurt, Y. Driver behavior analysis for safe driving: A survey. IEEE Trans. Intell. Transp. Syst. 16, 3017–3032 (2015).

[10] Nascimento JC, Figueiredo MA, Marques JS. Trajectory classification using switched dynamical hidden markov models. IEEE Trans Image Process. 2009.

[11] Calderara S, Prati A, Cucchiara R. Markerless body part tracking for action recognition. Int J Multimedia Intel Secur. 2010.

[12] Ohn-Bar E, Tawari A, Martin S, Trivedi MM. On surveillance for safety critical events: in-vehicle video networks for predictive driver assistance systems. Computer Vision Image Understand. 2015.

[13] R. L. Olson, R. J. Hanowski, J. S. Hickman, and J. Bocanegra, ''Driver distraction in commercial vehicle operations,'' U.S. Dept. Transp., Federal Motor Carrier Saf., Washington, DC, USA, Tech. Rep. FMCSA-RRR-09-042, 2009.

[14] K. Young, M. Regan, and M. Hammer, ''Driver distraction: A review of the literature,'' Monash Univ. Accident Res. Centre, Tech. Rep. 206, 2003.

[15] Revised General Safety Regulation, Eur. Commission, Belgium, Luxembourg, 2019.

[16] Castignani, G., Derrmann, T., Frank, R., & Engel, T. (2015). Driver behavior profiling using smartphones: A low-cost platform for driver monitoring. Castignani et al. discussed driver behavior profiling using smartphones as a cost-effective solution for driver monitoring.

[17] Zhang, X., Zheng, N., Wang, F., & He, Y. (2011). Visual recognition of driver hand-held cell phone use based on hidden CRF. Zhang et al. focused on visual recognition of driver hand-held cell phone use using hidden conditional random fields.

[18] Le, T. H. N., Zheng, Y., Zhu, C., Luu, K., & Savvides, M. (2016). Multiple scale Faster-RCNN approach to driver's cell-Phone usage and hands on steering wheel detection. Le et al. proposed a multiple scale Faster R-CNN approach for detecting driver's cell phone usage and hands on the steering wheel.

[19] Seshadri, K., Juefei-Xu, F., Pal, D. K., Savvides, M., & Thor, C. P. (2015). Driver cell phone usage detection on Strategic Highway Research Program (SHRP2) face view videos. Seshadri et al. worked on driver cell phone usage detection using face view videos from the Strategic Highway Research Program (SHRP2).

[20] Koesdwiady, A., Bedawi, S. M., Ou, C., & Karray, F. (2017). End-to- End Deep Learning for Driver Distraction Recognition. Koesdwiady et al. explored end-to-end deep learning for driver distraction recognition.

[21] Aksjonov, A., Nedoma, P., Vodovozov, V., Petlenkov, E., & Herrmann, M. (2019). Detection and evaluation of driver distraction using machine learning and fuzzy logic. Aksjonov et al. utilized machine learning (ML) and fuzzy logic techniques to detect and evaluate driver distraction effectively.

[22] Apostoloff, N., & Zelinsky, A. (2004). Vision in and out of vehicles: Integrated driver and road scene monitoring. Apostoloff and Zelinsky discussed an integrated system for monitoring both drivers and the road scene, utilizing vision technologies.