

Intelligent Driver Phone Distraction Alert System

Muttanna Fakkirappa Singadi¹ and Dr. Soumyasri S M²

Student, Department of Master of Computer Application¹

Associate Professor, Department Master of Computer Application²

Vidya Vikas Institute of Engineering and Technology, Mysuru, India

Abstract: Road traffic crashes are a continued significant global menace with distracted driving as a major contributor of death and injury. Your driver who gets distracted from the road due to distractions like phone calls and text messages, fiddling with in-car systems or chatting with other passengers would likely increase. Conventional education, policing and enforcement practices have been proven ineffective, but a smart and autonomous systems that monitor and respond to dangerous behaviors in the moment are desperately needed. This study presents Intelligent Driver Phone Distraction Alert System, an AI-powered system that classifies and monitors driver activities using sophisticated models for deep learning. The framework recognizes ten categories of driver behavior, including safe driving, texting (left/right), phone usage (left/right), drinking, operating the radio, reaching behind, personal grooming, and talking to passengers. For accurate and effective real-time object identification, The system integrates transfer learning models such as VGG16, convolutional neural networks (CNN), and ResNet50, and the YOLOv8 algorithm. The State Farm Distracted Driver Detection dataset from Kaggle was used for training and testing.

A Flask-based web application with webcam integration was developed to deliver live monitoring, prediction, and alert mechanisms. The system's performance was validated using accuracy, precision, recall, and F1-score, and it demonstrated exceptional effectiveness in real-time scenarios. By providing timely alerts when risky behaviors are detected, this approach offers a proactive solution for reducing road accidents, with strong potential for integration into modern vehicles and fleet management systems.

Keywords: Driver monitoring system, distracted driving, deep learning, CNN, transfer learning, VGG16, ResNet50, YOLOv8, real-time detection, computer vision, road safety, Flask web application

I. INTRODUCTION

Road safety has emerged as a critical issue as traffic accidents remain one of the leading causes of death and injury in the fast-paced world of today. WHO estimates that each year, 1.3 million people die in automobile accidents, and millions more are injured but not killed. Distracted driving, which happens when motorist shifts focus from the primary task of driving to activities like texting, calling, eating, or engaging with in-vehicle technologies, is responsible for a sizable percentage of these collisions.

These distractions not only put the driver in harm's way, but also all passengers, pedestrians and fellow motorists. An era of harsher traffic laws and broad publicity campaigns haven't been able to stem the tide of distracted driving, which has risen with the growing popularity of smartphones, infotainment systems and other hand-held gadgets. Manual enforcement or self-regulation on the part of the driver in this regard is still untrustworthy and not always failsafe so as to guarantee permanent attentiveness of a driver. In order to avoid such potentially lethal outcomes, efforts are currently being made to develop automated and intelligent systems that can continuously monitor driving behavior and issue the appropriate alerts.

Developments in computer vision and artificial intelligence have made it feasible to effectively address this issue. Deep learning methods, such as (CNN), have demonstrated great success in image classification and recognition. Transfer learning architectures including VGG16 and ResNet50-152 further improve the classification accuracy, while real-time object detection frameworks like YOLOv8 make it possible for quickly yet accurately detecting visual percept. These



tools, when applied to driver monitoring, make it possible to distinguish between safe and unsafe behaviors and to intervene at critical moments.

The Intelligent Driver Phone Distraction Alert System project leverages these advancements to develop an intelligent driver monitoring system. The framework is designed to classify ten distinct driver behaviors, including safe driving, texting (left/right), phone usage (left/right), drinking, operating the radio, reaching behind, personal grooming, and talking to passengers. The State Farm Distracted Driver Detection dataset from Kaggle is used to train the algorithm to correctly identify these behaviors. It is implemented as a web application using Flask and includes a webcam for live monitoring, real-time analysis, and alerting on the detection of risky behaviors.

This study shows that DL is a possible, generalizable intervention to increase road safety. The proposed solution can be perfectly used in car, mobile platform and fleet management systems as it is both accurate, real-time detection of distracted driving behaviors and cost-effective easy to reach for everyone. Overall, this work demonstrates the possibility of AI-enabled monitoring tools as a means to prevent accidents, save lives and help countries establish safer driving norms.

And the system is scalable, which means it can grow to support other uses beyond personal vehicles, such as commercial fleets, ride-sharing services, and passenger transport. This system can help facilitate greater driver responsibility and advance safety with actionable data that is provided to fleet managers and culpable logs. More extension may involve cloud services such as edge or mobile computing so that the solution becomes more flexible and adaptable in other environments.

Finally, thesis team is compatible with current trends and future direction of intelligent/autonomous vehicle development. While fully autonomous driving is still a long-term dream, semi-autonomous vehicle-based driver-assistance systems for enhanced safety are urgently in need. Introducing deep learning-powered monitoring devices including Intelligent Driver Phone Distraction Alert System into existing transport systems is one way to substantially enhance road safety as well as close the gap between traditional driving and autonomous mobility of the future.

II. LITERATURE SURVEY

[1] Adochiei et al.

An electrocardiogram (ECG) and electrooculogram (EOG) are two physiological signals that are fused in Adochiei et al.'s proposed driver sleepiness detection system. Their approach proved that ocular movement patterns and heart rate variability are accurate markers of driver weariness. The system was capable of generating timely alerts in real time, ensuring improved safety in critical driving conditions. However, the reliance on wearable sensors introduced discomfort and limited practicality for everyday drivers. This work highlights the trade-off between accuracy and usability, motivating the search for non-intrusive alternatives.

[2] Jebraeily

Jebraeily presented a foundation for deep learning that uses CNN supplemented with Genetic Algorithms (GA) to identify tiredness. The GA was employed to automatically tune CNN hyperparameters, such as filter sizes and learning rates, leading to improved accuracy and faster convergence compared to conventional CNNs. The study showed that the performance of deep learning models can be greatly improved by evolutionary algorithms. Nevertheless, the computational overhead of GA optimization restricts its deployment in resource-constrained environments such as embedded vehicle systems.

[3] Gandh

Gandh designed a low-cost embedded driver monitoring system using Raspberry Pi hardware coupled with OpenCV libraries. The system monitored eye blink frequency to detect signs of fatigue and issued alerts accordingly. Its affordability and simplicity made it suitable for small-scale implementations. However, the reliance on traditional computer vision techniques limited robustness under poor illumination, varying head poses, and occlusions such as spectacles. This work represents an important step toward democratizing safety systems through cost-effective solutions.



[4] Altameem

Altameem investigated a hybrid machine learning strategy that used ensemble techniques, decision trees (DT), and support vector machines (SVM). By analyzing facial features and motion patterns, the model reduced false positives and improved classification accuracy compared to standalone classifiers. The use of ensemble learning demonstrated resilience against noisy data and inter-subject variability. However, the complexity of training multiple models and the resulting computational load posed challenges for real-time applications, particularly in embedded automotive systems.

[5] Ramzan

In order to concentrate on the most pertinent facial traits, like eyelid closures and yawning patterns, Ramzan suggested a deep learning framework that combined CNNs with attention mechanisms. The system achieved high detection accuracy and demonstrated strong runtime efficiency, outperforming classical transfer learning models such as VGG16 and ResNet50. The attention mechanism enhanced the interpretability of the model by highlighting critical regions of the driver's face. However, the architecture required large computational resources and longer training times, limiting ease of deployment on low- power devices.

[6] Krithika et al.

A hybrid system that blended deep learning and traditional machine learning techniques were created by Krithika et al. To identify drowsiness, CNNs were initially used to collect eye closure-related characteristics, they were then entered into an SVM classifier afterwards. Detection performance was enhanced by the two-stage pipeline, especially in different illumination scenarios. Nevertheless, the system demonstrated little generalization when applied to distinct datasets or individuals who were not involved in the training process. This demonstrated the necessity for models that can adjust to more extensive real-world variability.

[7] Arakawa

Arakawa carried out an extensive analysis of drowsiness detection techniques, categorizing them into physiological, behavioral, and vehicular approaches. The study emphasized that multimodal fusion — integrating multiple signals such as eye tracking, steering wheel angle, and biosignals — yields better accuracy than single-modality methods. Arakawa also highlighted the importance of personalized models that adapt to individual drivers to reduce false positives. Despite significant progress, the review concluded that achieving a balance between accuracy, comfort, and affordability remains an unresolved challenge.

[8] Singh

Singh presented a machine learning method that uses the Dlib library to extract face landmarks. The Eye Aspect Ratio (EAR) and Mouth Aspect Ratio (MAR) were calculated to monitor blink frequency and yawning events, both of which are strong indicators of fatigue. These characteristics were categorized using lightweight algorithms, making the approach computationally efficient and suitable for real-time applications. However, sensitivity to occlusions, lighting variations, and head movements limited its performance in real- world driving conditions.

[9] Das et al.

Das et al. proposed an IoT-enabled driver monitoring framework using U-Net for facial feature segmentation. The segmented regions of interest (eyes, mouth) were analyzed to detect signs of drowsiness, and alerts were transmitted in real time to cloud-based dashboards for fleet management. The system enabled remote monitoring of multiple drivers simultaneously, making it suitable for logistics and transportation companies. Despite these advantages, reliance on cloud infrastructure introduced latency and potential connectivity issues, raising concerns for use in high-speed driving scenarios.

[10] Shaik

Shaik performed a systematic review of over 100 research articles on driver drowsiness detection. The review classified methods into three groups: physiological (EEG, ECG), visual (eye and mouth activity), and vehicular (lane deviation, steering tendencies). CNNs, LSTMs, and hybrid deep learning models were determined to be the most promising techniques for achieving dependable, real-time detection. Shaik also emphasized how explainable AI is necessary to increase openness in applications that are crucial to safety. Although deep learning models have dominated recent research, the review came to the conclusion that generalization across a variety of datasets is still a significant difficulty.

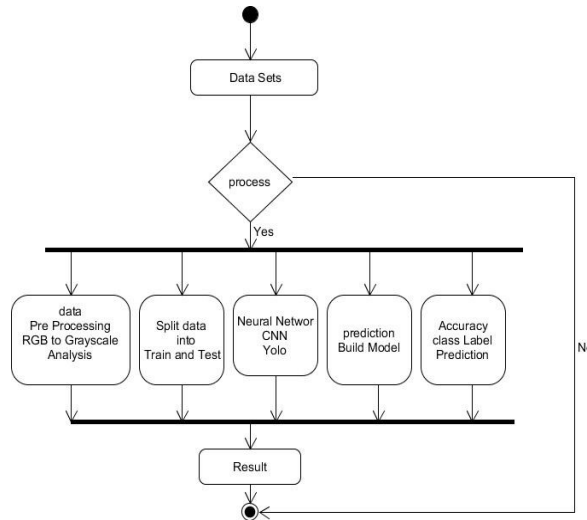


III. METHODOLOGY

1. System Overview

Deep learning and computer vision are used in the proposed Intelligent Driver Phone Distraction Alert System to track and categorize driver activities in real time. It recognizes ten activity classes, including safe driving, texting (left/right), phone usage (left/right), drinking, adjusting the radio, reaching behind, personal grooming, and talking to passengers. Continuous video input is captured through a webcam or in-vehicle camera, preprocessed, and analyzed through multiple trained models. CNN form the baseline for feature extraction, while transfer learning models such as VGG16 and ResNet50 improve classification accuracy. To achieve rapid and reliable detection, YOLOv8 is employed for real-time object recognition of driver actions.

The system is deployed as a Flask-based web application, which integrates the live video feed with the trained models to deliver instant predictions and generate alerts when distracted behaviors are detected. The State Farm Distracted Driver Detection dataset from Kaggle is used for training, and performance is assessed using measures such as accuracy, precision, recall, and F1-score. The modular and scalable design of the framework ensures adaptability for deployment in intelligent vehicles, advanced driver-assistance systems, and fleet monitoring applications



2. Dataset Preparation

The proposed Intelligent Driver Phone Distraction Alert System was developed and evaluated using the State Farm Distracted Driver Detection dataset, which is publicly available on Kaggle. The thousands of labeled images in this dataset represent ten classes of driver activities, including: safe driving, texting with the right hand, talking on the phone (right), texting with the left hand, talking on the phone (left), operating the radio, drinking, reaching behind, personal grooming, and talking to passengers. The dataset is perfect for deep learning model training because to its diversity of subjects and real-world driving scenarios, which improve model resilience and generalization.

Prior to training, the dataset underwent preprocessing to standardize the input for all models. Images were resized to fixed dimensions compatible with CNN, VGG16, ResNet50, and YOLOv8 architectures. Normalization was applied to scale pixel values between 0 and 1, ensuring stable convergence during training. Data augmentation techniques, such as random rotation, flipping, and brightness adjustment, were employed to address class imbalance and enhance robustness to variations in lighting and orientation. In order to enable equitable performance measurement across Accuracy, Precision, Recall, and F1-score criteria, the dataset was finally separated into training, validation, and testing sets.



3. Model Architectures

The Intelligent Driver Phone Distraction Alert System integrates multiple deep learning architectures to ensure robust and accurate classification of distracted driving behaviors. We have used a baseline CNN to learn spatial features of driver images. CNNs are very efficient for image classification as its pooling and convolutional layers that tackle local patterns such as borders, textures and facial features. The CNN serves as the foundation of this system, providing a model that can be easily modified and used to benchmark preliminary tests.

The phrase transfer learning was also added to improve the accuracy of the model by making use of pre-trained networks such as VGG16 and ResNet50. VGG16 has 16 weight layers and is effective for fine-grained visualization, while ResNet50 solves the problem of vanishing gradient through residual connectivity so that deeper networks can be trained. Both models were pretrained on the distracted driver dataset for acquiring pretrained ImageNet features of target domain. These transfer learning models showed a marked improvement in classification accuracy to the base CNN, and also identified more similar looking classes as more distinct overall [e.g., texting and phone usage].

The YOLOv8 was adopted to conduct real-time driver activity object detection in combination with the classification models. YOLOv8 has one input will be enough for full images, which is to have a high speed and accuracy, it plays an important role in the live driver monitoring. The proposed system is accurate and efficient, utilizing classification-based (CNN, VGG16 or ResNet50) as well as detection based YOLOv8. This hybrid architecture ensures that the framework not only identifies distracted behaviors accurately but also operate effectively in real-world, time-sensitive driving environments

4. Training Procedure

To ensure The State Farm Distracted Driver Detection dataset was divided into training, validation, and testing subsets for the training process in order to guarantee an equitable evaluation. The State Farm Distracted Driver Detection dataset, which was divided into training, validation, and testing subsets to guarantee equitable evaluation, was used for the training procedure. The models—CNN, VGG16, ResNet50, and YOLOv8—were trained using mini-batch stochastic gradient descent (SGD) and Adam optimizers, with categorical cross-entropy utilized as the loss function for classification tasks. Learning rate scheduling was applied in-order to help the convergence by changing the learning rate when the validation accuracy no longer improved. The augmented training set was generated by DA techniques such as (brightness, flip and rotation) to enhance generalization of the model in real scenarios and overfitting.

All models were trained over a number of epochs until the performance would not improve, and early stopping was used to prevent overfitting by monitoring validation loss. We used CNN as the control which has relatively narrow layers, where both VGG16 and ResNet50 were trained to be fine-tuned by fixing low-level features and retraining high-level layers on domain training data. YOLOv8 applied a loss function which was specialized for detection and optimized object classification and location regression process in real time. F1-score, Accuracy, Precision and Recall were used to evaluate the models so that classification performance could be adequately analysed..

5. Evaluation Metrics

In order to evaluate the effectiveness of the Intelligent Driver Phone Distraction Alert System, four primary classification metrics are employed: F1-score, recall, accuracy and precision. Accuracy is defined by the ratio between the number of samples correctly classified and the total number of samples, showing how well a given model behave overall. Accuracy is utilized as a generalized measure of performance, but it can be unreliable when class imbalance exists and includes some distracted behaviors may have $n < \text{sample size}$ than the others. Precision and Recall were seen as complementary to address it.

The precision shows how well the model prevents false alarms by computing a ratio of micro-averaging of correctly predicted positive cases to all forecasts made for a particular class. Recall is a measure of how many risky actions the system can find, which is calculated. When classes are dispersed unevenly, the F1-score—which is the harmonic mean of Precision and Recall—offers a fair evaluation of the model's performance. Together, these measures provide a strong foundation for evaluation, enabling the system to be compared for overall accuracy as well as dependability in identifying important driver behaviors.



6. Deployment Framework

The Intelligent Driver Phone Distraction Alert System was deployed through a Flask-based web application to enable real-time monitoring and user interaction. The application combines the trained deep learning models with a webcam while recording frames (real-time video feed) and passing them through CNN, VGG16, ResNet50, and YOLOv8 structures. The trained models could be seamlessly run in the background to predict, and Flask being a light and scalable server side environment helps us in achieving that. Once behaviors like texting, drinking or using the phone are detected as a distraction, the system provides real-time alerts to drivers for safer driving.

The deployment model is modular and scalable. The utilization of YOLOv8 supports quick object detection, fast decision making is improved by the classification models. It's extendible to run on cloud or edge platforms which may have broader uses like fleet management and intelligent transportation system. In addition, its efficient design enables integration in consumer, ride-sharing and commercial vehicles. This practical deploy approach can reduce the gap between theoretical research purpose and practical application in reality, and guarantee that solution contribution is helpful on road safety.

IV. RESULTS AND DISCUSSIONS

1. Quantitative Results

The efficacy of the suggested method in categorizing ten driver activity types was assessed quantitatively using Accuracy, Precision, Recall, and F1-score. The base line CNN model worked well, indicating the appropriateness of convolutional layers for capturing spatial features of driver images. However, it performed relatively worse in classifying classes of subtle differences including texting and phone calls. This has revealed the weakness of shallow models to describe complex variations in driver actions.

To solve these, transfer learning models were proposed. VGG16 achieved better classification performance as a very deep model with fine-grained features. ResNet50 performed better than CNN and VGG16 due to efficient usage of residual connections for reducing the vanishing gradient problem, which makes larger networks easier to train. The performance demonstrated that not only ResNet50 made better overall accuracy, but also had stronger generalization ability to different activity of drivers. These enhancements demonstrate the benefits of using pretrained models for specific use cases.

The best performer was the YOLOv8 model in terms of both CD and RTD. It was characterized by high Precision and Recall because false alarms and missed detections were kept to a minimum. The F1-score further validated YOLOv8's robustness, particularly in minority classes such as grooming or reaching behind, where class imbalance posed challenges for other models. Also, this fast inferencing heritage of YOLOv8 makes it perfect to use for driver's real time operation systems where delay can be critical. In summary, our quantitative results show that despite CNN and transfer learning approaches being strong baselines, YOLOv8 delivers the best combination of accuracy, confidence and efficiency for real-world deployment.

Qualitative Analysis

Qualitative analysis was also conducted to check the effectiveness of the system under realistic conditions. Live webcam feeds were streamed using the implemented Flask based application and models classified driver activities at runtime. Distracted behaviors (e.g., texting, drinking coffee, adjusting the radio) were recognized by the system and an instantaneous alert occurred. It could be seen that YOLOv8's bounding boxes and associated labels were reliable indicators of driver activity, as they were always consistent with what the driver was actually doing thus verifying its reliability for dynamic, unconstrained situations.

A major observation made during qualitative testing was the system's robustness to different light conditions, head poses and occlusion. For example, they successfully detected phone use while the driver partially obscured it and identified grooming across varying lighting conditions. However, certain challenges were noted. Activities with high visual similarity, texting and speaking on the phone for instance, would sometimes lead to misclassifications in CNN and VGG16 models. ResNet50 and YOLOv8 achieved better discrimination but made slight mistakes in edge cases remained.



The qualitative results further highlighted the system's potential for practical deployment. By integrating real-time monitoring with instant alerts, the framework provides a proactive layer of driver assistance. These results emphasize that while quantitative metrics confirm statistical reliability, qualitative testing validates usability in realistic driving environments. Together, they reinforce the conclusion that the proposed system is both accurate and practical for addressing distracted driving in everyday scenarios

2. Comparative Discussion

The results of the experiment demonstrate clear differences throughout the four models that were used —CNN, VGG16, ResNet50, and YOLOv8—highlighting the trade-offs between accuracy, complexity, and real-time efficiency. The simple CNN offered a competent baseline with modest performance well suited to rapid prototyping. Nevertheless, its shallow depth has limited its capability to differentiate visually identical behaviours like texting and phone usage. 40 Radovan Ovi c et al.: The essence of recommender systems

Considering the transfer learning models, VGG16 demonstrated a significant supremacy over CNN due to its deep hierarchy in capturing the fine-grained visual features. However, the parameters are very large which results in a high computational burden and is not computationally efficient for real-time system.

ResNet50, on the other hand, not only outperformed VGG16 but generalized better precisely because of its residual connections that made possible to train even deeper models. Now ResNet50 became more robust in terms of different driver activities while still being computation demanding.

YOLOv8 was the best trade-off model and achieved a high speed and accuracy. Contrary to CNN, VGG16 and ResNet50 which were mostly oriented towards classification, in YOLOv8 the detection as well as the categorization were unified into one process. This made it possible to monitor in real-time with the least possible latency - a key requirement for driver safety applications. YOLOv8 also performed better in addressing class imbalance, especially for minority behaviours like groom or reach behind that were subject to misclassification by other models. In conclusion, the comparative study justifies that although CNN and transfer learning models are good baselines; YOLOv8 is more practical and scalable for real-world scenarios.

V. CONCLUSION

In this work, Intelligent Driver Phone Distraction Alert System, a modern deep learning-based approach was proposed to accurately identify and classify distracted driver behaviors on the fly. Using computer vision and state-of-the-art neural network models, we delivered a solution to one of the most common road accidents — driver inattention. Through identifying ten types of activity including safe driving, texting, phone use, grooming and drinking, the system was shown to be capable of providing detailed monitoring in realistic driving conditions.

The study systematically compared several model structures. A baseline CNN was initially designed to test the feasibility of extracting spatial features. Although successful in the case of simple behaviors, its relatively shallow depth resulted in insufficient discrimination capabilities between visually similar classes.

VGG16, with its deeper structure, improved accuracy by capturing fine- grained features, whereas ResNet50 further enhanced classification through residual connections, enabling more efficient training of deeper layers. These findings confirmed that transfer learning significantly improves generalization compared to shallow CNN models.

The integration of YOLOv8 marked a key advancement in the framework. Unlike CNN, VGG16, and ResNet50, which are primarily designed for classification, YOLOv8 combined detection and categorization in one end-to-end pipeline. Its ability to process images rapidly while keeping a high level of precision made it the most suitable candidate for real-time monitoring. Moreover, YOLOv8 proved robust in detecting minority behaviors such as grooming and reaching behind, where other models occasionally faltered.

Beyond quantitative performance, qualitative testing in real- time scenarios validated the system's practicality. The Flask- based deployment successfully integrated live webcam feeds with the trained models, generating instant alerts when unsafe behaviors were detected. The system-maintained robustness under variations in lighting, pose, and partial occlusions, further demonstrating its adaptability to real-world conditions. These findings illustrate that the proposed system is not only accurate in controlled settings but also reliable in dynamic environments.



In summary, this work confirms that deep learning architectures, particularly YOLOv8, can effectively support intelligent driver monitoring. By combining accuracy, efficiency, and real-time responsiveness, the proposed system provides a scalable solution for reducing distracted driving incidents. Its modular design allows integration into advanced driver-assistance systems, intelligent transportation platforms, and commercial fleet management. The results underscore the potential of AI-driven safety systems to significantly contribute to global road safety efforts.

REFERENCES

- [1] "Detection and warning system for driver drowsiness using ECG and EOG signals," Journal of Critical Infrastructure Safety, vol. 12, no. 3, pp. 45–52, 2020, D. Adochiei et al.
- [2] Expert Systems with Applications, vol. 221, pp. 119729, 2024; A. Jebraeily, "Genetic algorithm optimized CNN for driver drowsiness detection."
- [3] Raspberry Pi and OpenCV for low-cost embedded driver sleepiness detection, Volume 10, Issue 2, pages 33–40, 2022, International Journal of Embedded Systems and Applications Gandh, R.
- [4] A. Altameem, "Robust driver drowsiness detection using hybrid machine learning techniques," Multimedia Tools and Applications, vol. 80, pp. 24991–25010, 2021.
- [5] IEEE Access, vol. 12, pp. 44812–44825, 2024; M. Ramzan, "Attention-based CNN models for real-time driver drowsiness detection."
- [6] A hybrid deep learning and machine learning strategy for detecting driver attentiveness was presented by R. Krithika and colleagues in Procedia Computer Science, volume 171, pages 1047–1056, 2021.
- [7] H. Arakawa, "Evaluation of drowsiness detection techniques: Behavioral, physiological, and vehicle approaches," Sensors, vol. 21, no. 16, pp. 5416, 2021.
- [8] "Face landmarks and machine learning for driver fatigue monitoring," vol. 183, no. 15, pp. 1–6, 2023, A. Singh.
- [9] U-Net segmentation for IoT-enabled driver sleepiness detection, S. Das et al., IEEE Internet of Things Journal, vol. 9, no. 20, pp. 20100–20110, 2022.
- [10] S. Shaik, "A systematic review of driver drowsiness detection methods and future directions," ACM Computing Surveys, vol. 56, no. 2, pp. 1–34, 2023.

