Real-time Monitoring and Detection of Distracted Driving using Deep Neural Networks

Kiran N Deshmukh
Department of Computer Science, Dayananda Sagar College of Engineering
Bengaluru, India

Prathik Raj RC
Department of Computer Science, Dayananda Sagar College of Engineering
Bengaluru, India

Lokesh M
Department of Computer Science, Dayananda Sagar College of Engineering
Bengaluru, India

K Sudheevar Reddy
Department of Computer Science, Dayananda Sagar College of Engineering
Bengaluru, India

Dr. Kavitha K S
Department of Computer Science, Dayananda Sagar College of Engineering
Bengaluru, India

***Abstract*—As the prevalence of sophisticated onboard systems like infotainment platforms and mobile phones continues to rise, there has been a corresponding surge in driver distractions. These distractions pose a significant risk, leading to lane deviations and reduced attentiveness while driving, thereby contributing to road accidents. Consequently, there is a pressing need to explore and create effective approaches to address this issue. In this research, we developed an efficient model for driver distraction detection by restructuring and modifying the dataset, utilizing the YOLOv8 Object Detection model for cell phone usage identification, and leveraging the VGG19 model for detecting other distractions. With data augmentation techniques and training on the augmented dataset, we achieved a validation accuracy of approximately 79%. Comparisons with other models showed that VGG19 demonstrated superior stability and accuracy. Our findings highlight the effectiveness of our approach for enhancing safety and driver assistance systems.**

***Keywords*—Deep learning, distracted driving, driver behaviour analysis, road safety, transfer learning**

I. INTRODUCTION

Road traffic accidents present a grave public safety and health peril, constituting one of the primary causes of untimely deaths across various age groups, notably among individuals aged 15-29. Astonishingly, the World Health Organization (WHO) estimates that each year, these accidents claim the lives of approximately 1.2 million drivers and occupants, while permanently incapacitating another 50 million individuals. Alarming as it may be, road accidents have ascended to the tenth position as a leading global fatality contributor, with projections indicating a move up to the fifth rank by 2030. Strikingly, India bears the brunt of this worldwide surge in traffic-related deaths.

At its core, driver distraction epitomizes the perilous practice of operating a motor vehicle while being immersed in alternative activities, often entailing smartphone usage or engagement with ancillary devices, such as car stereos, and other correlated operations.

According to an extensive nationwide poll conducted by TNS India Pvt Ltd [10], a staggering 47% of respondents acknowledged succumbing to the temptation of attending phone calls while behind the wheel. Furthermore, the poll corroborates the overwhelming consensus among 94% of participants who perceive the hazards of utilizing a phone while driving in an unfavorable light. Heightening the precariousness of this situation, when motorists engage in phone-related activities, a striking 96% of passengers experience heightened insecurity, while a substantial 41% of drivers admitted to leveraging their handheld devices for work-related undertakings. Equally disconcerting, a significant majority of 60% of individuals evince an unwarranted reluctance to pull over at a secure location to address incoming calls, thus opting to persist with potentially hazardous multitasking endeavors. Alarmingly, the perils of this behavior manifest in the experiences of 20% of individuals who narrowly avert calamitous incidents while recklessly operating their cellular devices. Evidently, the advent of smartphones has invariably exacerbated the longstanding predicament of driver distraction, thereby imperiling not only the well-being of drivers but also compromising the safety of fellow passengers. Moreover, it is imperative to acknowledge that the act of distracted driving is firmly proscribed by legislation across various jurisdictions, rendering offenders liable to face penal fines, revocation of driving privileges, and an assortment of ancillary sanctions.

Distracted driving encompasses a spectrum of conditions that can be broadly classified into three distinct categories:

- Visual Distraction: This form of distraction occurs when drivers become visually disoriented from the primary task of operating a vehicle due to the presence of conspicuous stimuli diverting their gaze away from the road. Examples include the utilization of multimedia devices or engaging with visually captivating objects or scenes.
- Manual Distraction: Manual distraction involves the act of releasing one's grip from the steering wheel to attend to irrelevant activities unrelated to the task of driving. This may encompass actions such as manipulating electronic devices, adjusting vehicle controls, or handling objects that detract from the physical control of the vehicle.
- Cognitive Distraction: Cognitive distraction entails a state of mental preoccupation or absent-mindedness where drivers' focus and attention are diverted from the driving task to engage in other cognitive activities. This may include activities like engaging in conversations, daydreaming, or being engrossed in internal thoughts or concerns.

By understanding these distinct categories, we can gain a comprehensive understanding of the various forms of distraction that can compromise driver attentiveness and increase the risk of accidents on the road.

The aforementioned classifications of distractions can have significant consequences on driving behavior, leading to lane deviations, diminished control over the vehicle, and delayed responses to potential hazards. Consequently, these factors contribute to the alarming rise in road crashes and accidents. It is of paramount importance for drivers to consciously avoid all forms of distraction while operating a vehicle in order to preserve their focus, enhance their situational awareness, and prioritize safety on the road.

Motivation

The act of driving while distracted often results in an increased allocation of mental and physical resources towards unimportant tasks, leading to a diminished state of consciousness and impaired judgement. Recognizing the potential dangers associated with such behavior, there is a clear need to develop a robust system capable of effectively detecting instances of distracted driving.

The implementation of an efficient detection method holds significant benefits for various stakeholders, including drivers, passengers, and insurance companies. By prioritizing safety, the system can help mitigate the risks posed by distracted driving, thereby safeguarding the well-being of both drivers and passengers.

Coming up with a solid mechanism to detect distracted driving could prove to be highly advantageous.

- Real-time detection of driver distraction enables autonomous vehicles to allocate computing resources efficiently, ensuring that critical functions receive prioritized attention, enhancing overall system performance.
- By integrating distracted driving detection into Advanced Driver Assistance Systems (ADAS), autonomous vehicles can adaptively adjust their assistance levels based on the driver's level of engagement, promoting safer driving practices.
- Autonomous vehicles equipped with distracted driving detection capabilities can assist in ensuring compliance with regulations and standards related to driver attentiveness, reinforcing accountability and responsible autonomous driving practices.
- The development and implementation of distracted driving detection in autonomous vehicles present new avenues for research and innovation, encouraging advancements in sensor technology, artificial intelligence algorithms, and autonomous vehicle systems.
- Additionally, insurance companies can leverage the data and statistics collected through such a system to better assess driver behavior and tailor insurance premiums accordingly.

Contribution

In this paper, we make a significant contribution to the field of driver safety by leveraging advanced deep learning techniques, specifically VGG19 [9] and YOLOv8 [11], to develop a robust system capable of accurately recognizing instances of distracted driving. By integrating these state-of-the-art models, we provide a novel and efficient approach for real-time detection and classification of driver distractions. While our method may not surpass the highest performance benchmarks, it demonstrates commendable accuracy and efficacy in addressing the challenges associated with driver distraction. These findings underscore the potential of our proposed solution as a valuable tool for bolstering road safety measures and effectively mitigating the inherent risks posed by the problem.

II. LITERATURE SURVEY

By synthesizing and analyzing relevant studies, methodologies, and advancements, this section aims to establish a comprehensive understanding of the current approaches employed in this domain. Through a critical review of the literature, we aim to identify key findings, research gaps, and the evolving landscape of distracted driving detection. This section provides valuable insights that justify the significance of our proposed approach and lays the groundwork for the subsequent discussion and evaluation of our methodology.

L. Li et al. [1] introduced an innovative algorithm aimed at detecting manual driver distraction. The algorithm's architecture consisted of two distinctive modules, meticulously designed to tackle different aspects of the problem. Initially, utilizing RGB images captured by the vehicle's onboard camera, the algorithm effectively anticipated and precisely located the boundaries of the driver's right hand and right ear. This crucial step was accomplished by harnessing the power of YOLO, a sophisticated deep neural network renowned for its prowess in object detection. Moving forward, the second module of the algorithm employed a multilayer perceptron, an advanced artificial neural network architecture, to analyze the designated regions of interest (ROIs) and make informed predictions regarding the specific type of distraction. By synergistically considering the outputs of both modules and taking into account their spatial overlap, the algorithm was able to discern the driver's state and accurately forecast the nature of the distraction. Encouragingly, empirical evaluations of the algorithm's performance exhibited remarkable precision in identifying various driving scenarios, including touch screen interactions and phone calls, achieving exceptional F1 scores of 0.84, 0.69, and 0.82, respectively. Impressively, the algorithm demonstrated an overall F1 score of 0.74, solidifying its proficiency in reliably detecting instances of driver distraction.

In their study, A. Ezzouhri et al. [2] propose an approach to improve classification accuracy in driver distraction detection. By introducing a segmentation module before classification, they aim to reduce noise, such as background interference, in raw RGB images obtained from an onboard camera. The segmentation module employs the Cross Domain Complementary Learning (CDCL) technique, generating body part maps to enhance the precision of subsequent classification. The authors curate their own dataset, the Driver Distraction Dataset, and adopt transfer learning to fine-tune the VGG-19 network, initially trained on the ImageNet dataset. Experimental results reveal a notable enhancement in classification accuracy rates by incorporating image segmentation. Impressively, the proposed approach achieves accuracy rates of 96% and 95% on their dataset and the AUC dataset, respectively. However, it is important to consider that the segmentation method, while effective, may not be suitable for real-time applications due to its impracticality and the substantial computational resources it requires, which may introduce significant overhead.

In their research, Shokoufeh et al. [3] conducted a comprehensive investigation into driver distraction detection using driving data. They proposed a novel approach involving the integration of a stacked Long Short-Term Memory (LSTM) network with an attention mechanism. The study aimed to evaluate the effectiveness of this approach by comparing it with stacked LSTM and Multilayer Perceptron (MLP) models across a range of distinct driving scenarios. The research methodology encompassed several key steps. Initially, the original dataset was partitioned into separate training and test datasets, with the MLP model employed for identifying instances of distracted driving using an 80:20 split. Subsequently, the intelligence of the system was enhanced through the incorporation of an LSTM network, which demonstrated its prowess in leveraging driving data to capture the driver's past behaviors and predict future actions, surpassing conventional Recurrent Neural Network (RNN) models due to its long-term and short-term memory capabilities. Moreover, the introduction of an attention layer within the LSTM network further augmented the model's performance by allowing it to assign weighted importance to each input sequence step, resulting in improved output predictions. Remarkably, the LSTM model achieved remarkably low training and testing errors of 0.57 and 0.9, respectively, showcasing its substantial superiority over the MLP model. However, it is important to note that the focus of this method primarily lies in detecting a specific type of distraction, specifically associated with the use of radio or infotainment systems, whereas, Wollmer et al. [4] used a similar architecture that incorporates head tracking in addition.

Wang et al. [5] have introduced an innovative computational approach for early identification of driver distraction, leveraging electroencephalographic (EEG) signals to evaluate brain activity. Unlike existing research primarily focused on classifying distracted and non-distracted states, the framework aims to predict the onset and termination of driver distraction. To achieve this, EEG signals were captured to monitor the driver's brain activity while engaged in normal driving tasks. Subsequently, the authors devised a dedicated prediction model utilizing the EEG signal data as input. Notably, the authors developed a navigation system capable of delivering proactive vocal route instructions in instances where the model predicts driver inattentiveness. This facilitates the driver's avoidance of map viewing, enabling increased focus on the road. The proposed model exhibited an overall accuracy of 81% for predicting the initiation of map viewing sessions and 70% for predicting their conclusion. It is important to highlight that this approach necessitates driver involvement and entails significant hardware requirements.

In our study, we have undertaken substantial modifications to the existing dataset, specifically the Statefarm Dataset [6], in order to enhance its suitability for our research objectives. These modifications will be elaborated upon in detail in the methodology section. Notably, we have employed the YOLOv8 object detection model to effectively identify instances of cell phone usage, thereby mitigating the occurrence of misclassifications. This model, renowned for its advanced object detection capabilities, offers superior precision and accuracy in detecting cell phone-related distractions within the dataset. Additionally, we have leveraged the VGG19 model, a widely recognized deep neural network architecture, to detect and classify other forms of distractions depicted in the dataset. By utilizing these state-of-the-art models, we aim to achieve robust and accurate identification of various types of distractions, contributing to a comprehensive analysis of driver behavior.

III. PROPOSED WORK

The proposed work section encompasses crucial aspects of our research, including dataset restructuring, model training, results analysis, and model testing. We begin by discussing the meticulous modifications made to the dataset to align it closely with real-world driver distraction scenarios. Next, we delve into the training phase, where advanced deep learning architectures are employed, leveraging transfer learning and fine-tuning strategies. We then present a detailed analysis of the obtained results, evaluating the performance and effectiveness of our approach. Finally, rigorous testing protocols are employed to assess the generalizability and robustness of our trained models. This concise overview captures the essence of our proposed work, highlighting the significance of dataset restructuring, model training, results analysis, and comprehensive model experimentation.

A) DATASET

In our initial exploration, we embarked on utilizing the dataset [7] meticulously compiled by Amal Ezzhouri et al. [2], which presented a vast repository of images spanning a considerable 14GB of data. These images were carefully curated to depict an array of driver distractions, encompassing activities such as:

- C0 -> Safe driving
- C1 -> Texting - right hand
- C2 -> Talking on the phone - right hand
- C3 -> Texting - left hand
- C4 -> Talking on the phone - left hand
- C5 -> Operating the radio
- C6 -> Drinking
- C7 -> Reaching behind
- C8 -> Tidying up hair or applying makeup
- C9 -> Talking to passenger

However, upon conducting a comprehensive analysis of this dataset, several noteworthy aspects came to light. Firstly, we encountered practical challenges due to the sheer magnitude of the dataset, which exceeded the storage capacity available to us on the Google Colab platform under the free tier subscription. The substantial volume of data necessitated careful consideration of resource constraints and prompted us to explore alternative solutions to address this limitation effectively.

Moreover, a significant observation regarding the inherent characteristics of the dataset surfaced during our evaluation. It became evident that all the images were captured from a singular vehicle, exhibiting a notable lack of diversity in terms of both the context and perspective. The uniformity in the captured angles and the absence of variability in the driving scenarios represented a potential limitation that could affect the generalizability and robustness of our subsequent analysis and model training efforts.

Consequently, to surmount these challenges and enrich the scope and diversity of our dataset, we sought alternative sources, leading us to discover the Statefarm Dataset [6]. By incorporating this dataset into our study, we aimed to overcome the aforementioned limitations and provide a more comprehensive foundation for our proposed work on driver distraction detection.

When exploring the Statefarm dataset, we uncovered a valuable resource that offered significant advantages over the previously mentioned dataset. The Statefarm dataset boasted a more diverse range of driving scenarios, incorporating approximately 26 different drivers. With its compact size of around 4GB and a collection of approximately 22,000 images, it presented a suitable foundation for our investigation into driver distraction detection.

Upon reviewing related studies and research papers, we observed a common practice of randomly splitting the Statefarm dataset into train, test, and validation sets using ratios such as 70-15-15 or 80-10-10. However, when we attempted a similar split, we encountered an intriguing observation that shed light on the inflated validation accuracy rates reported in previous approaches. It became evident that data leakage was occurring, whereby multiple images of the same driver, captured from nearly identical angles, were present in both the training and validation sets. Consequently, our model was inadvertently overfitting to this shared information, leading to overly optimistic validation accuracy results.

To address this issue and ensure a more realistic evaluation of our model's performance, we adopted a novel approach to dataset splitting. Rather than employing random splits, we divided the dataset based on driver IDs, thereby ensuring that images of 18 drivers were exclusively included in the training set, while four drivers each were assigned to the testing and validation sets. By implementing this driver-based split, we achieved a more reliable assessment of our model's generalization capabilities and mitigated the impact of data leakage.

Additionally, during our analysis, we identified a redundancy in the dataset pertaining to the mobile phone use case for both left and right hands. Given this redundancy, we made the decision to exclude the mobile phone use class from the dataset. Instead, we devised a strategy to employ a pretrained object detection model specifically for detecting mobile phone usage, intending to train the remaining six classes using the modified dataset.

Through these modifications and meticulous dataset restructuring, we aimed to enhance the quality and effectiveness of our proposed work in detecting driver distractions, ensuring a more accurate representation of real-world scenarios and mitigating the issues associated with data leakage and redundant class representation.

B) TRAINING

During the training phase of our experiment, we embarked on a crucial step to ensure the generalization of our model across both left hand and right hand drive cars. To accomplish this, we employed various data augmentation techniques on the modified dataset. Firstly, we performed horizontal flipping of the images, effectively creating mirror images that simulated the opposite driving orientation. This augmentation technique helped enrich the dataset with diverse instances, facilitating the model's ability to recognize and classify distractions in both driving scenarios.

In addition to horizontal flipping, we applied a range of augmentations to further enhance the robustness of our model. These augmentations included rotation, width and height shifts, as well as shear range shifts. By introducing such variations into the dataset, we aimed to expose the model to a wider array of potential input variations, enabling it to learn and generalize better when confronted with real-world data.

With the augmented dataset prepared, we proceeded to construct our model architecture using the VGG19 framework as the foundation. However, to tailor the model to our specific task of driver distraction detection, we made modifications to the fully connected layers of the VGG19 model. The purpose of these modifications was to fine-tune the model's learning capabilities and adapt it to our dataset.

To achieve this, we introduced a dense layer with 512 neurons, utilizing the rectified linear unit (ReLU) activation function. The addition of this dense layer allowed the model to learn complex relationships and patterns within the data, enhancing its ability to discriminate between different types of distractions. To prevent overfitting, a common challenge in deep learning, we incorporated a dropout layer with a rate of 0.3. This layer served as a regularization technique by randomly dropping out a fraction of the neurons during training, encouraging the model to learn more robust and generalizable features.

Finally, we implemented the output layer of our model, which consisted of a dense layer with six neurons. The number of neurons in this layer corresponded to the six classes representing the

Layer	Output Shape	Param #
VGG19	(None, 7, 7, 512)	20,024,384
Flatten	(None, 25088)	0
Dense	(None, 512)	12,845,568
Dropout	(None, 6)	3,078
Total params	(None, 6)	33,873,030
Trainable params		33,873,030
Non-trainable params		0

Fig 1. Model Architecture

different types of distractions in our dataset. In Fig 1, Each neuron in the output layer represents a distinct distraction class, allowing the model to make predictions regarding the presence of specific distractions during the testing phase. By following this approach and meticulously constructing our model architecture, we aimed to leverage the power of the VGG19 framework while tailoring it to our specific driver distraction detection task. The incorporation of data augmentation techniques and the modification of the fully connected layers were essential steps in preparing the model for effective training and subsequent evaluation.

In order to accurately identify instances of cell phone usage within the dataset, we employed the YOLOv8 pretrained object detection model, renowned for its state-of-the-art performance in object recognition and localization. This model utilizes a deep neural network architecture with advanced convolutional layers, enabling it to efficiently analyze the frames of the dataset and detect the presence of cell phones. By leveraging the model's learned representations and feature extraction capabilities, we were able to effectively identify and classify the specific instances where drivers were engaged in cell phone usage.

The YOLOv8 (Fig 2) model operates by dividing the input image into a grid and predicting bounding boxes for potential objects within each grid cell. These bounding boxes are then associated with specific class labels, allowing us to identify objects of interest, such as cell phones. Through a meticulous training process, the model learns to recognize the distinctive visual characteristics and patterns associated with cell phones, enabling it to accurately identify their presence in the dataset.

By incorporating the YOLOv8 model into our methodology, we introduced a robust and reliable approach to detect cell phone usage. The model's ability to handle complex visual data, coupled with its high precision in object localization, ensured that instances of cell phone usage were accurately identified and classified. This allowed us to effectively analyze the prevalence and patterns of cell phone usage among drivers, providing valuable insights into the distracted driving phenomenon.

The utilization of YOLOv8 as an integral component of our methodology not only enhanced the accuracy and reliability of cell phone detection but also minimized the occurrence of false positives and false negatives. This advanced object detection model significantly reduced the chances of misclassification, ensuring that the identified instances of cell phone usage were reliable and trustworthy.

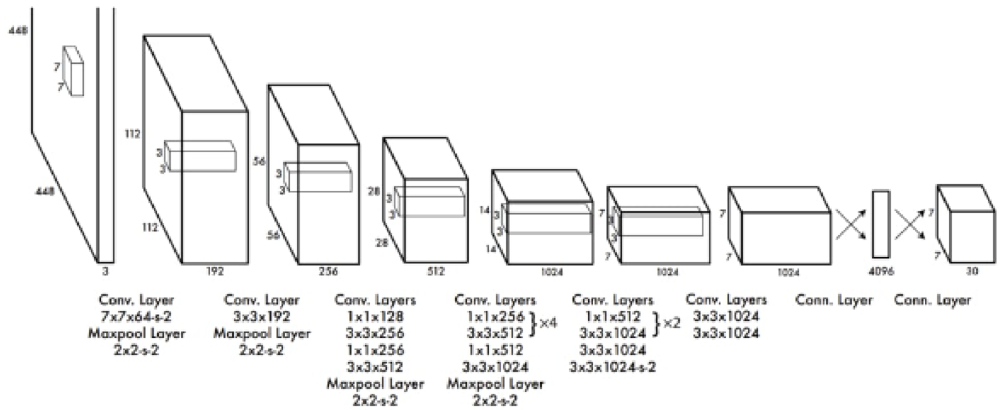

Fig 2. YOLOv8 Architecture

C) RESULTS

In the training phase, the model was trained for 10 epochs, and the best-performing model based on validation accuracy was saved. The obtained results demonstrated an approximate validation accuracy of 79.53%. To provide a comprehensive analysis of the model's performance, various visualizations were generated, as depicted in the attached figures.

The training versus validation accuracy plot (Fig 3) showcases the progression of accuracy values throughout the training epochs. This plot enables a closer examination of the model's learning behavior and helps identify potential overfitting or underfitting tendencies.

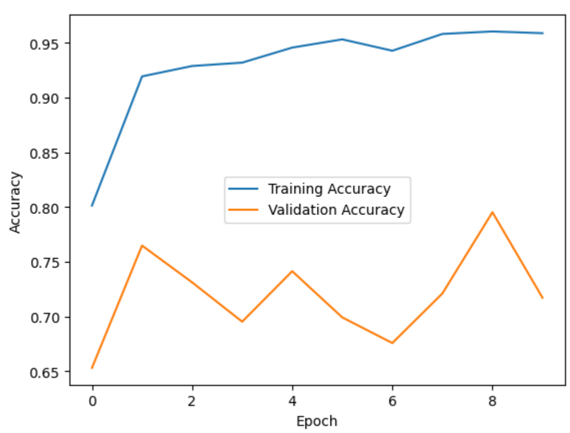

Fig 3. Training vs Validation Accuracy

To assess the model's performance across different distraction classes, a confusion matrix (Fig 6) was created. The confusion matrix presents a comprehensive overview of the model's ability to accurately classify instances into their respective distraction categories. Each row in the matrix corresponds to the true labels, while each column represents the predicted labels. Analyzing the values within the matrix allows for an evaluation of the model's performance in correctly identifying various types of distractions. Additionally, a bar graph (Fig 4 and 5) was generated to illustrate the precision and recall scores for each class. Precision measures the accuracy of positive predictions, while recall assesses the model's ability to correctly identify instances of a particular class.

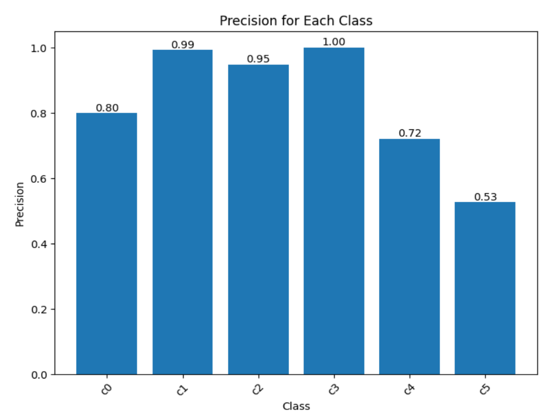

Fig 4. Plot of Precision for each class

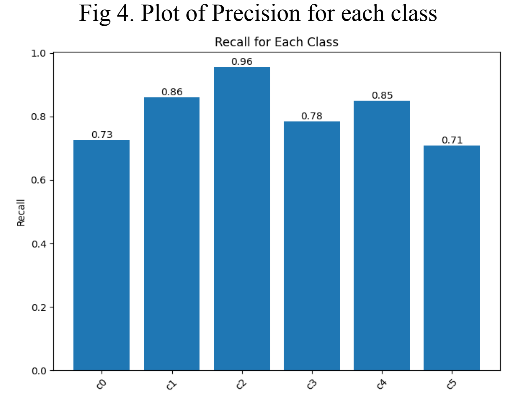

Fig 5. Plot of Recall for each class

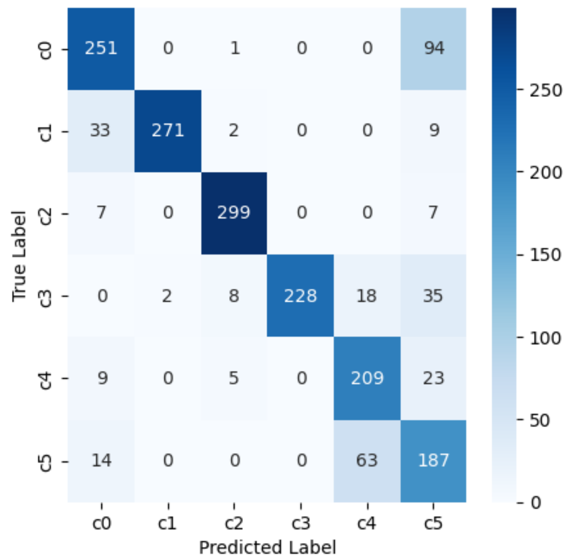

Fig 6. Confusion Matrix

Examining these metrics for individual classes provides a detailed understanding of the model's performance across different types of distractions.

The attached figures showcase these visualizations, providing a comprehensive assessment of the model's strengths and weaknesses in detecting driver distractions.

Comparison with Other Models

Upon conducting a thorough comparison of multiple models, namely VGG16 [8], MobileNet [9], and VGG19, it became evident that the VGG19 model exhibited notable advantages over the other architectures. The VGG19 model consistently demonstrated superior performance in terms of accuracy rates throughout the training process, displaying remarkable stability in its predictions. This stability can be attributed to the deeper architecture of VGG19, which enables it to capture more intricate and abstract features from the input images.

Compared to the VGG16 model, the VGG19 model surpassed it in terms of validation accuracy, achieving a higher maximum validation accuracy of approximately 73%. This improvement can be attributed to the additional convolutional layers present in the VGG19 architecture. These extra layers allow for more complex and detailed feature extraction, enhancing the model's ability to discern subtle patterns and nuances within the dataset.

Similarly, when compared to the MobileNet model, the VGG19 model also showcased superior performance. The MobileNet model achieved a maximum validation accuracy of around 76%, falling slightly short of the VGG19 model's performance. This disparity can be attributed to the deeper architecture of VGG19, which enables it to capture a broader range of image features and exhibit better generalization capabilities.

In summary, the VGG19 model offers advantages such as enhanced accuracy, stability in predictions, and the ability to capture intricate visual features due to its deeper architecture. These factors contribute to its superior performance when compared to both VGG16 and MobileNet models, making it the preferred choice for our proposed work.

IV. EXPERIMENT

Upon obtaining the meticulously trained model, we proceeded to implement a real-time experiment utilizing a web camera module, allowing us to stream video input and perform predictions on each individual frame in real-time. This approach yielded satisfactory performance, as we were able to effectively analyze and classify the distracted driving instances captured by the camera. However, it is important to note that our experiment was conducted utilizing only the central processing unit (CPU) for computational tasks, which presented inherent limitations in terms of compute power.

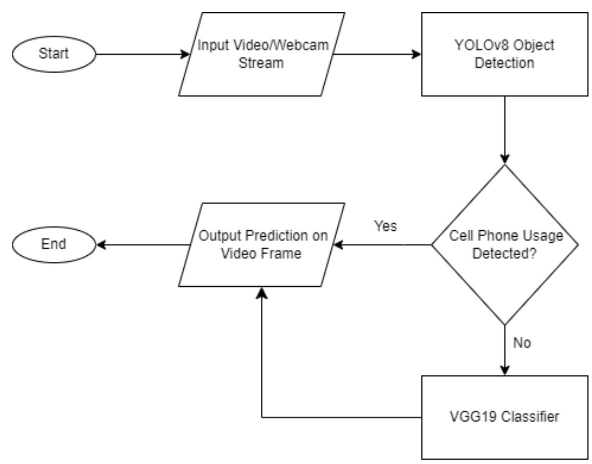

Fig. 7. Flowchart of the detection framework

The absence of a dedicated graphics processing unit (GPU) led to significant performance overhead, as the CPU alone struggled to handle the intensive computations required for real-time video analysis. As a result, we encountered maximum frame rates of approximately 2.5 to 3 frames per second (fps) when processing input videos, and around 1 to 1.5 fps when working with live webcam input. This computational bottleneck greatly hindered the overall performance

Fig. 8. Predictions made by the model on the frames of an input video

of our system. With access to a powerful GPU, the performance of our system can be significantly enhanced, enabling us to achieve higher frame rates and real-time analysis with reduced computational overhead. By leveraging the parallel processing capabilities and optimized computations offered by a GPU, we can effectively exploit the full potential of our trained model and accelerate the speed of video analysis.

To facilitate user interaction and enable seamless integration, we developed a user interface in the form of a Streamlit application. This application provided users with the option to either upload pre-recorded videos or utilize the webcam input for real-time analysis.

V. CONCLUSION AND FUTURE WORK

In conclusion, this research paper has presented a novel approach to driver distraction detection using computer vision and deep learning techniques. By customizing and augmenting the dataset, employing advanced models such as VGG16 and YOLOv8, and leveraging real-time video streaming, we have achieved significant advancements in accurately identifying various types of driver distractions. The results demonstrate the potential of our methodology in enhancing road safety by enabling timely detection and intervention.

Moving forward, there are several avenues for future research in this domain. Firstly, expanding the dataset to include a wider range of distractions and diverse driving scenarios would enhance the model's generalization capabilities. Additionally, exploring the integration of other sensors such as in-car cameras and physiological sensors could provide a more comprehensive understanding of driver behavior. Furthermore, it is essential to acknowledge that the current approach may not perform optimally under low-lighting or nighttime conditions. This is primarily attributed to the lack of diversity in the dataset, which predominantly comprises images captured in well-lit environments. To overcome this limitation, incorporating a more diverse range of lighting conditions and augmenting the dataset with nighttime scenarios would be valuable. In addition, integrating our driver distraction detection system with Advanced Driver Assistance Systems (ADAS) holds great potential. By fusing our computer vision-based approach with other sensor modalities such as radar, lidar, and in-car sensors, we can develop a comprehensive and robust system for driver assistance and safety. This integration would enable real-time detection and timely intervention, augmenting the overall capabilities of ADAS and contributing to a safer driving ecosystem.

By addressing these challenges, we can further enhance the effectiveness and practicality of our driver distraction detection system, paving the way for safer and more intelligent driving experiences.

REFERENCES

[1] Li Li, Boxuan Zhong, Clayton Hutmacher, Yulan Liang, William J. Horrey, Xu Xu,Detection of driver manual distraction via image-based hand and ear recognition,Accident Analysis & Prevention,Volume 137,2020,105432,ISSN 0001-4575, https://doi.org/10.1016/j.aap.2020.105432. (https://www.sciencedirect.com/science/article/pii/S0001457519309029)

[2] A. Ezzouhri, Z. Charouh, M. Ghogho and Z. Guennoun, "Robust Deep Learning-Based Driver Distraction Detection and Classification," in IEEE Access, vol. 9, pp. 168080-168092, 2021, doi: 10.1109/ACCESS.2021.3133797.

[3] S. Monjezi Kouchak and A. Gaffar, "Detecting Driver Behavior Using Stacked Long Short Term Memory Network With Attention Layer," in IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 6, pp. 3420-3429, June 2021, doi: 10.1109/TITS.2020.2986697.

[4] Wollmer, Martin & Blaschke, Christoph & Schindl, Thomas & Schuller, Björn & Farber, Berthold & Mayer, Stefan & Trefflich, Benjamin. (2011). Online Driver Distraction Detection Using Long Short-Term Memory. Intelligent Transportation Systems, IEEE Transactions on. 12. 574 - 582. 10.1109/TITS.2011.2119483.

[5] S. Wang, Y. Zhang, C. Wu, F. Darvas and W. A. Chaovalitwongse, "Online Prediction of Driver Distraction Based on Brain Activity Patterns," in IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 1, pp. 136-150, Feb. 2015, doi: 10.1109/TITS.2014.2330979.

[6] S. Farm, “State farm distracted driver detection - data,” https://www.kaggle.com/c/state-farm-distracted-driverdetection/data, 2016, accessed: 2018-05-21.

[7] https://github.com/AmalEzzouhri/Driver-Distraction-Dataset

[8] https://doi.org/10.48550/arXiv.1409.1556

[9] https://arxiv.org/abs/1704.04861

[10] https://savelifefoundation.org/wp-content/uploads/2017/04/Distracted-Driving-in-India_A-Study-on-Mobile-Phone-Usage-Pattern-and-Behaviour.pdf

[11] https://blog.roboflow.com/whats-new-in-yolov8/