International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 08 | Aug - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Using Face Recognition and Deep Learning Identifying the Driver Behaviors

Sinchana K.P,
Don Bosco Institute of Technology,
Bengaluru, India
Email: sinchkp.1999@gmail.com

Shravya G Gowda
Don Bosco Institute of Technology,
Bengaluru, India
Email: shravyagowda97@gmail.com

ABSTRACT

Driving entails a series of actions demanding intense focus. Sometimes, these actions get overshadowed by other behaviors like smoking, eating, drinking, conversing, making phone calls, adjusting the radio, or even drowsiness. These behaviors stand as the primary culprits behind today's traffic accidents. Thus, it's imperative to create applications that can forewarn drivers. This study proposes a streamlined convolutional neural network structure for identifying driver behaviors, thereby enhancing the accuracy of the alert system and minimizing traffic collisions. This network combines modules for extracting features and making classifications. By merging these components, the feature extraction within the neural network is designed to be more potent and efficient. The model aims to decrease computational costs and amplify data flow within the network by employing depth-wise separable convolutions and adaptive connections. Additionally, the Convolution Block Attention Module aids the network in prioritizing vital features, resulting in improved performance for various computer vision tasks. The classifier module includes global average pooling and a softmax layer to compute class probabilities. The overall architecture optimizes network parameters while upholding classification precision. The entire network is trained and evaluated using three benchmark datasets.

Keywords— Incident, Convolutional Neural Network (CNN), driver actions, facial identification.

I. INTRODUCTION

In the present era, road traffic systems have experienced significant growth both in terms of quantity and complexity. Consequently, the incidence of accidents has also gradually risen. The World Health Organization's estimations reveal that road traffic accidents result in more than 1.35 million fatalities annually and contribute to a total of 50 million collisions. A key factor contributing to this surge in accidents is driver behavior.

The aforementioned statement also highlights that if drivers were more attentive while operating vehicles, the accident rate could be reduced by a factor of four. According to data from the National Highway Traffic Safety Administration (NHTSA) in the United States, distracted driving was responsible for 2,895 deaths in 2019, accounting for 8.7% of all traffic-related fatalities that year.

Significant progress in curbing accidents caused by driver distraction and lack of attention can be achieved through collaboration among various stakeholders, including car manufacturers, researchers, and lawmakers. This collaborative effort can lead to the development of warning and prevention mechanisms for drivers or the advancement of vehicle automation to mitigate driver distraction from the outset, such as through automated driving features. Some modern vehicles, particularly those of a certain class and equipped with specific features, may already possess rudimentary systems capable of detecting certain forms of driver inattention, like fatigue, and subsequently notifying the driver. These systems generally fall under the category of driver assistance systems, which encompass features like adaptive distance maintenance and lane-keeping.

Driving a vehicle is a highly intricate task, necessitating a multifaceted approach to tackle driver distraction and inattention. This strategy should encompass education, technological advancements, legal measures, and behavior modification. By increasing public awareness about the perils associated with distracted driving, implementing stricter regulations and penalties, and integrating state-of-the-art safety systems into vehicles, the risks can be substantially diminished, fostering safer roads for everyone. NHTSA statistics indicate that around 25% of police-reported crashes involve some form of driver inattention, whether due to distraction, fatigue, or absentmindedness. Aberrant driving behaviors, such as drowsiness, aggression, intoxication, carelessness, and recklessness, contribute to increased accident risks. While most driver behavior detection systems typically focus on identifying a single type of abnormal driver behavior, a limited number of studies attempt to differentiate among various forms of such behavior.

However, a comprehensive driver behavior monitoring system capable of efficiently distinguishing between diverse abnormal driver behaviors has yet to be developed. To differentiate between distinct driver behavior styles, we can outline various characteristics associated with these styles. Aggressive driving behavior encompasses hazardous lane changes, sudden shifts in acceleration and speed, and tailgating. Distracted driving style entails diverting attention from essential driving tasks to secondary activities like eating, drinking, using smartphones, or in-car technologies. This often leads to quick driver reflexes to correct the vehicle's situation. Driver fatigue manifests as the drowsy driving style, marked by observable signs such as yawning, closed eyes, sluggish reactions and steering, infrequent brake usage, and reduced revolutions per minute (RPM).

The consumption of alcohol or drugs impairs the driver's mental faculties and gives rise to the drunk driving style, characterized by noticeable indicators like reduced brake application, abrupt accelerations, and perilous lane changes. Finally, the safe driving style can be identified by recognizing the traits associated with risk-free driving behaviors.

II. LITERATURE SURVEY

Executing the project effectively necessitates a thorough grasp of the current situation and the advantages and disadvantages of the technology involved. Establishing a strong foundation for the project's objectives requires a meticulous analysis and synthesis of data from diverse sources during the literature review process. A well-executed literature review serves the purpose of identifying gaps in knowledge, shaping research themes, and guiding the trajectory of project endeavors. Moreover, it serves as an indicator of the researcher's familiarity with the existing body of work in the field.

Driving encompasses a range of behaviors demanding heightened concentration. At times, these behaviors are overshadowed by other actions such as smoking, eating, drinking, conversing, phone usage, adjusting the radio, or drowsiness. These actions are also the primary contributors to current traffic accidents. Consequently, the development of applications to proactively alert drivers becomes indispensable. This study introduces a streamlined convolutional neural network architecture designed for identifying driver behaviors, enhancing the warning system's ability to deliver precise information and mitigate traffic collisions. This architecture combines feature extraction and classifier components. For the task of detecting distracted driving using established benchmark datasets, the architecture and its constituent elements collectively aim to fine-tune network parameters and sustain a high level of classification accuracy [1]. Attaining remarkably high accuracy on two out of three datasets and a notably strong accuracy on the third dataset underscores the model's impressive performance.

If you suspect that you might have sleep apnea [2] or are displaying symptoms of disrupted breathing during sleep, it is recommended to seek evaluation and guidance from a healthcare professional specializing in sleep medicine. Identifying and addressing sleep apnea early on can significantly improve both the quality of your sleep and your overall health outcomes.

In this research article, a novel and lightweight method for detecting sleep apnea (SA) is proposed, known as SE-MSCNN, which stands for Single-Lead ECG-based Multi-Scaled Fusion Network. This approach primarily comprises two key components: a multi-scaled convolutional neural network (CNN) module and a channel-wise attention module. To enhance the accuracy of SA detection, various scaled ECG data with different-length adjacent segments are extracted through three sub-neural networks. To address the challenge of integrating features from different scales effectively, a channel-wise attention module incorporating a squeeze-to-excitation block is utilized to intelligently combine the diverse scaled features. The study also includes an ablation analysis and a computational complexity assessment of the SE-MSCNN.

The overall findings underscore the remarkable performance of the proposed SE-MSCNN compared to other existing SA detection methods, particularly on the apnea-ECG benchmark dataset. This highlights the potential of the model to significantly enhance the identification of sleep apnea and contribute to advancements in sleep medicine and healthcare. However, for a comprehensive evaluation of the model's effectiveness and applicability, further validation on additional datasets and comparisons with other relevant strategies are recommended. Leveraging its advantages of rapid responsiveness and efficient resource utilization, the SE-MSCNN could potentially be integrated into a wearable device, offering a sleep apnea detection service for individuals undergoing home sleep tests (HST).

In general, the study demonstrates the effectiveness of the mobileVGG CNN architecture in identifying instances of distracted driving. The approach yields promising accuracy outcomes while maintaining computational efficiency, a crucial aspect for practical implementations like Advanced Driver Assistance Systems (ADAS) in vehicles. Recent research from the World Health Organization (WHO) indicates a steady increase in fatal traffic accidents, yet the global death rate has stabilized in recent times. Distracted driving is identified as a significant contributor to these accidents, as revealed by a survey conducted by the National Highway Traffic Safety Administration (NHTSA). Addressing driver distraction is imperative for enhancing road safety.

The authors of the research paper introduce a CNN-based approach for the detection and categorization of driver distraction. CNNs are renowned for their efficacy in tasks involving image processing and pattern recognition. While accuracy is essential, the development of safety features for ADAS must also consider computational efficiency. The proposed mobileVGG architecture strikes a balance between accuracy and computational efficiency. To assess its performance, the authors utilize datasets from the American University in Cairo (AUC) and Statefarm obtained from Kaggle for detecting inattentive drivers.

Remarkably, the proposed mobileVGG outperforms previous methodologies, achieving an accuracy of 95.24% on the AUC dataset and 99.75% accuracy on the Statefarm dataset. This accomplishment is particularly noteworthy given that the network utilizes only 2.2 million parameters, resulting in relatively low computational complexity and memory requirements. These results underscore the superiority of the mobileVGG architecture compared to earlier approaches, achieving accuracy levels of 95.24% on the AUC dataset and 99.75% on the Statefarm dataset while utilizing only 2.2 million parameters.

The provided article discusses a research study that introduces the mobileVGG Convolutional Neural Network (CNN) architecture for the identification and categorization of driver distraction. The authors aim to develop a CNN specifically designed for safety enhancements in Advanced Driver Assistance Systems (ADAS), emphasizing both accuracy and computational efficiency. Recent research by the World Health Organization (WHO) indicates a consistent increase in fatal traffic accidents; however, the death rate relative to the global population has stabilized in recent times. Notably, distracted driving has been recognized as a significant contributing factor in traffic accidents, as evidenced by a survey conducted by the National Highway Traffic Safety Administration (NHTSA). To enhance road safety, addressing driver distraction is imperative.

The research study introduces the mobileVGG architecture, a convolutional neural network approach, as a solution. CNNs have showcased exceptional performance across various visual recognition tasks and are commonly employed in image processing applications. Designing safety features for Advanced Driver Assistance Systems necessitates a careful balance between computational efficiency and accuracy. The authors' goal is to create a CNN that offers both swift and efficient memory utilization, alongside accurate classification of driving distractions. To evaluate the performance of the proposed mobileVGG network, the authors utilize datasets from the American University in Cairo (AUC) and Statefarm, specifically designed to detect instances of inattentive driving.

The World Health Organization (WHO) issued a report stating that global road traffic accidents result in 1.25 million annual deaths, with a consistent upward trend in recent years. Nearly one-fifth of these accidents can be attributed to drivers who are distracted. Current efforts in detecting distracted drivers have been limited to a narrow range of distractions, primarily focusing on cell phone usage. These approaches often lack reliability and resort to ad hoc methods. In [5], the authors introduce a groundbreaking initiative by providing the first publicly accessible dataset designed for identifying driver distraction. This dataset encompasses a broader array of distraction postures compared to existing alternatives. Furthermore, the authors propose a dependable solution grounded in deep learning principles, achieving an impressive 90% accuracy rate.

Their system entails a genetically weighted ensemble of convolutional neural networks. Notably, the authors demonstrate that incorporating a weighted ensemble of classifiers via a genetic algorithm significantly enhances classification confidence. The study also delves into exploring the impact of various visual cues, such as facial and hand localizations, along with skin segmentation, in distraction detection. Additionally, the authors present a streamlined version of their ensemble that maintains an 84.64% classification accuracy and functions effectively in real-time scenarios.

In a separate paper [6], the authors introduce a fresh dataset intended for estimating "distracted driver" postures. Alongside this, they propose an innovative system attaining an impressive 95.98% accuracy in classifying driving postures. This achievement is attributed to a genetically weighted ensemble of convolutional neural networks, which significantly bolsters classification confidence. The potential applications of this research extend across several domains, including driver assistance systems, automotive safety, and human-machine interfaces within vehicles. The authors underline that leveraging a weighted ensemble of classifiers through a genetic algorithm leads to heightened classification confidence. Additionally, they scrutinize how different visual components, like hands and face, impact distraction detection and classification using methods such as face and hand localization. Ultimately, the authors present a more streamlined version of their ensemble, which sustains a commendable 94.29% classification accuracy while operating in real-time environments.

Over the preceding two decades, there has been a growing body of research dedicated to the advancement of autonomous vehicles, with numerous industries collaborating closely with academia to push the boundaries of knowledge. One pivotal aspect of developing such vehicles involves the automatic recognition of activities taking place within the vehicle itself. In this study, the authors propose an innovative approach centered on human pose analysis, designed to monitor a driver's state and activity using video-based methods. This approach draws inspiration from the recent successes of deep Convolutional Neural Networks (CNNs) in visual recognition tasks. By analyzing a single frame, the method is capable of swiftly inferring the driver's current state or activity, thereby enabling real-time operation.

The authors also amalgamate insights from recent investigations into human pose detection and transfer learning for visual recognition. These concepts are seamlessly integrated within the adapted DenseNet framework [7], where one stream is dedicated to capturing the underlying body pose, while the other stream focuses on capturing appearancerelated information. The proposed technique is subjected to rigorous assessment using two demanding datasets that encompass a range of secondary non-driving activities. The experimental findings unequivocally illustrate that the integration of the latent body pose into existing deep networks significantly enhances the performance of driver activity recognition.

The introduction of real-time monitoring and alerts through driver distraction detection systems holds tremendous promise for enhancing traffic safety. These systems have the potential to elevate drivers' awareness of their attention levels and promote safer driving behaviors. Furthermore, deeper exploration and the development of advanced driver assistance systems (ADAS) that actively intervene to mitigate distractions and bolster traffic safety stand to benefit from the wealth of data generated by such systems. To this end, a testbed for assisted driving has been created to simulate authentic driving scenarios and validate the efficacy of the distraction detection algorithms.

In the pursuit of this endeavor, the authors compiled a comprehensive dataset encompassing images of drivers assuming both typical and distracted driving postures. Employing an embedded graphics processing unit platform, the authors implemented and evaluated four deep convolutional neural networks [8]—VGG-16, AlexNet, GoogleNet, and residual network. Moreover, they devised a real-time conversational warning system designed to alert drivers when their focus deviates from the task of driving. Empirical results underscore the superiority of the proposed approach over a baseline model with only 256 neurons in fully-connected layers. Importantly, the outcomes also point to GoogleNet as the most effective model among the four for detecting distractions in the context of a driving simulator testbed.

III. PROPOSED METHODOLOGY

IV.

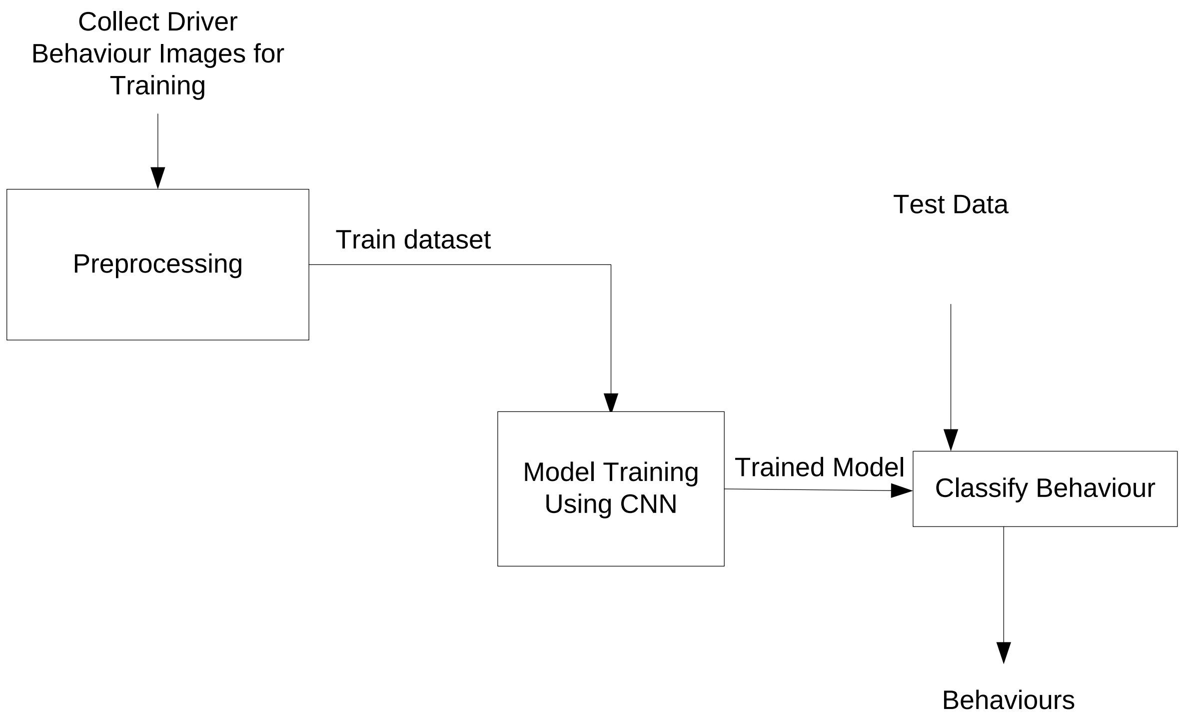


Figure 1: Block Diagram

The above block diagram mentions the modules as described below:

A. Data Augmentation

The utilization of data augmentation presents a potent strategy for artificially enlarging dataset sizes. This technique proves particularly advantageous when dealing with limited training data, effectively bolstering the efficacy and robustness of deep learning models. In this context, the ImageDataGenerator is harnessed to generate novel images. To accommodate the spiral nature of the images, a rotation range of 360 degrees is designated, as this allows for various rotations without altering the inherent meaning of the images. Exploring alternative image transformations within the purview of the ImageDataGenerator class is a possibility; nevertheless, caution is advised when implementing augmentations, as certain transformations could potentially undermine the lower accuracy of the CNN model.

Following augmentation, a uniform data distribution is achieved, and the images are uniformly resized to dimensions of (128, 128, 1). Furthermore, prior to being integrated into the model, the images undergo normalization.

B. Face Recognition

Haar cascades have gained substantial popularity in the realm of object detection tasks, particularly in face detection, owing to their blend of speed and precision. Although Haar cascades served as a pioneering breakthrough upon their inception, more sophisticated object detection techniques, such as the Single Shot Multibox Detector (SSD) and the You Only Look Once (YOLO) algorithm, have since surpassed them in terms of both swiftness and accuracy. Nevertheless, Haar cascades retain their value and efficiency for specific applications and environments. The key steps entailed in face recognition through Haar cascades include:

1. Haar Cascade Training: Training Haar cascades involves a substantial dataset containing affirmative and negative images. Positive images encompass facial examples, while negative images lack them. The training process encompasses the extraction of Haar-like features from these images, utilizing machine learning algorithms like AdaBoost to create a robust classifier capable of discerning between faces and non-faces based on these features.

2. Haar Cascade XML File: Upon completing the training, the resulting classifier is saved as an XML file, recognized as the Haar cascade XML file. This file encapsulates information about the acquired features and the classifier thresholds.

3. Face Detection: Employing Haar cascades to detect faces within an input image involves a sliding window approach. At each stage, a diverse set of rectangular windows moves across the image, computing Haar-like features for each window.

4. Integral Image: Computing Haar-like features for every window can be resource-intensive. An integral image is introduced to optimize this process, enabling swift calculation of Haar-like features within any image's rectangular region.

5. Classifier Application: Haar-like features computed for each window are contrasted with the acquired features stored within the Haar cascade XML file. Based on classifier thresholds, the algorithm ascertains whether each window encompasses a face. Windows identified as faces are designated as potential facial regions.

6. False Positive Mitigation: The face detection process might yield false positive identifications, where non-facial

zones are wrongly classified as faces. To mitigate this, supplementary methods like non-maximum suppression or overlapping region elimination can be employed to eradicate redundant or overlapping identifications and enhance the final array of detected faces.

7. Face Recognition: Following face detection, subsequent stages can be executed for face recognition. This could involve the extraction of facial attributes, such as landmarks or descriptors, from the detected faces, followed by comparison against a database of known faces using approaches like eigenfaces, Fisherfaces, or deep learning-based methods.

Haar cascades provide a relatively expedited approach to face detection and have been extensively employed across a spectrum of applications, encompassing face recognition, analysis of facial expressions, and identification of facial attributes.

C. Deep Learning Method for Training and Classification

Although Haar cascades were revolutionary upon their inception, modern and more sophisticated techniques, such as convolutional neural networks (CNNs) based on deep learning or pre-trained models like MTCNN and OpenFace, have demonstrated enhanced efficacy in face detection and recognition endeavors.

CNN Model Architecture

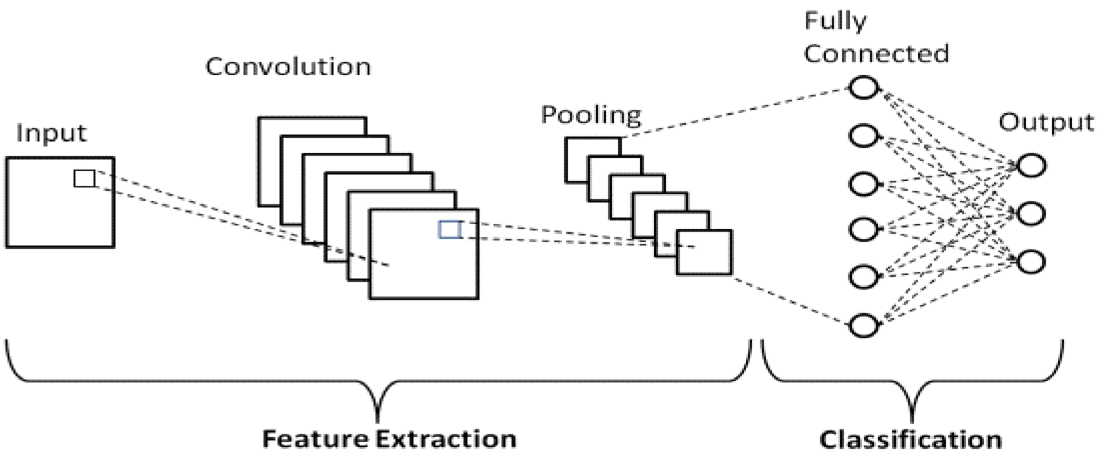
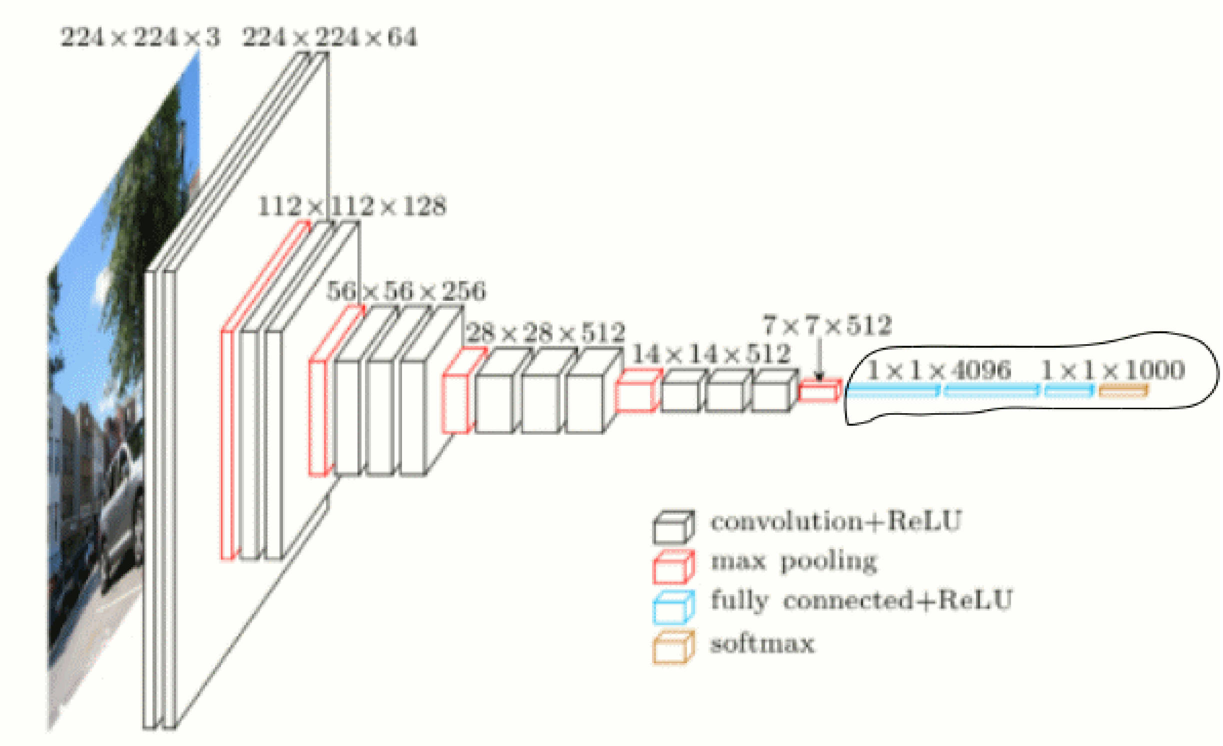


Figure 2: Basic CNN Architecture

Figure 3: CNN Model with four Convolutional Layers



The CNN model structure employed in the implementation comprises the following features:

- The model incorporates four Convolutional Layers, each with 128, 64, 32, and 32 filters, respectively.
- Within the convolutional layers, filters of different sizes are utilized.
- Subsequent to each convolutional layer, a MaxPool2D layer is applied.
- Following the convolutional block, there are two Fully Connected layers.

V. ALGORITHMS

Steps involved in the operation of a Convolutional Neural Network (CNN) algorithm, which is widely employed for various computer vision tasks like image classification, object detection, and image segmentation, are outlined as follows:

1. Data Preprocessing: Initiate by assembling and preprocessing your training data. This encompasses gathering a dataset of labeled images and executing essential preprocessing actions such as resizing, normalizing pixel values, and augmenting the data (e.g., rotation, scaling, flipping) to enrich the variety of training samples.

2. Architecture Design: Determine the architecture of your CNN, including the number of convolutional layers, pooling layers, fully connected layers, and output layers. Consider task complexity, image dimensions, and the desired number of classification classes.

3. Model Initialization: Initialize your CNN model and set the initial layer weights. Common methods for initialization include random assignment or adopting pre-trained weights from models suited to analogous tasks (transfer learning).

4. Forward Propagation: Conduct forward propagation through the CNN layers. This encompasses feeding the input image through convolutional layers, applying activation functions (e.g., ReLU) for non-linearity, and pooling layers (e.g., MaxPooling) for feature map downsampling.

5. Flattening: Following convolutional and pooling layers, transform the output feature maps into a one-dimensional vector. This prepares the data for interaction with fully connected layers.

6. Fully Connected Layers: Establish connections between the flattened output and one or more fully connected layers, which operate as a conventional neural network. Activation functions (e.g., ReLU) are applied here as well.

7. Output Layer: Introduce an output layer with the appropriate number of neurons. For classification, the softmax activation function is often chosen to yield class probabilities. Linear activation might be used for regression tasks.

8. Loss Function: Select an appropriate loss function contingent on your problem's nature. Cross-entropy loss is common for classification, while mean squared error (MSE) is suitable for regression.

9. Backpropagation: Execute backpropagation to compute gradients of the loss function concerning weights and biases. This entails retrogressing errors through the network, adjusting weights using optimization algorithms like stochastic gradient descent (SGD), Adam, or RMSprop.

10. Training: Train the CNN iteratively by feeding training data through the network, computing loss, and adjusting weights through backpropagation. This continues for a set number of epochs or convergence.

11. Validation: After each training epoch, assess CNN performance on a distinct validation dataset. This aids in monitoring model generalization, enabling early stopping or hyperparameter adjustments.

12. Testing: Ultimately, evaluate the trained CNN using a separate testing dataset to gauge its performance on unseen data. Calculate metrics such as accuracy, precision, recall, and F1 score to assess model effectiveness.

VI. RESULTS AND DISCUSSION

With the implementation of algorithms mentioned, achived the lossand accuracy with more efficient and graphs shown below represents the loss and accuracy for the model with the epochs

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 08 | Aug - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

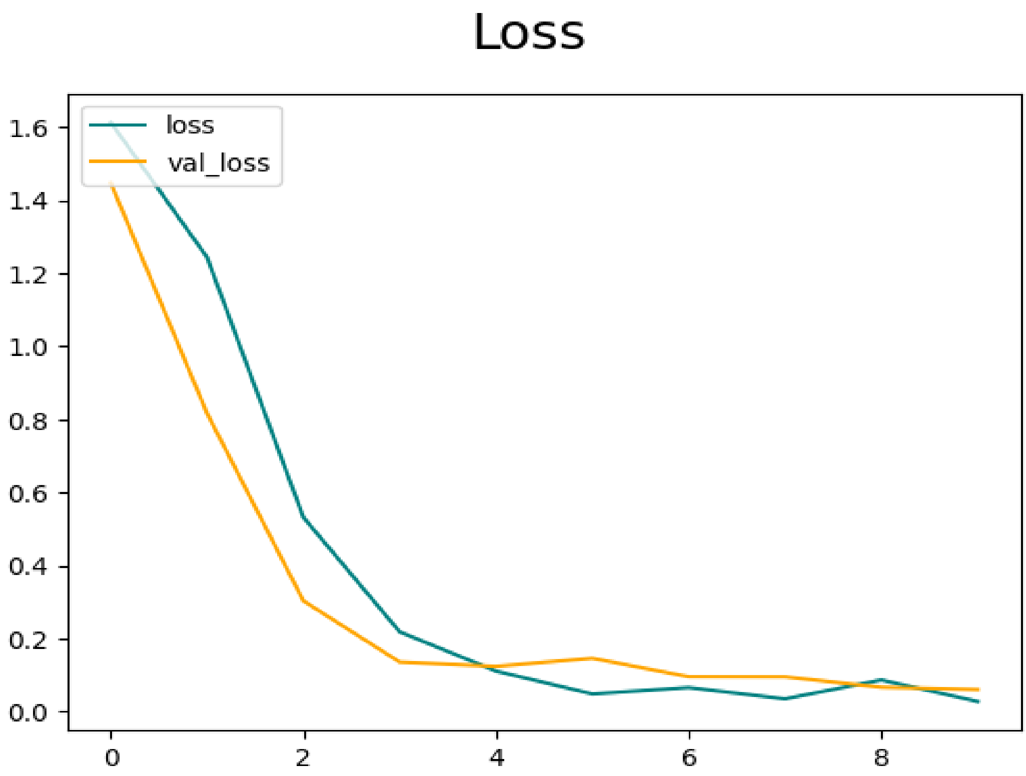


Figure 4: Loss Graph for the Model with the epochs considered as 10

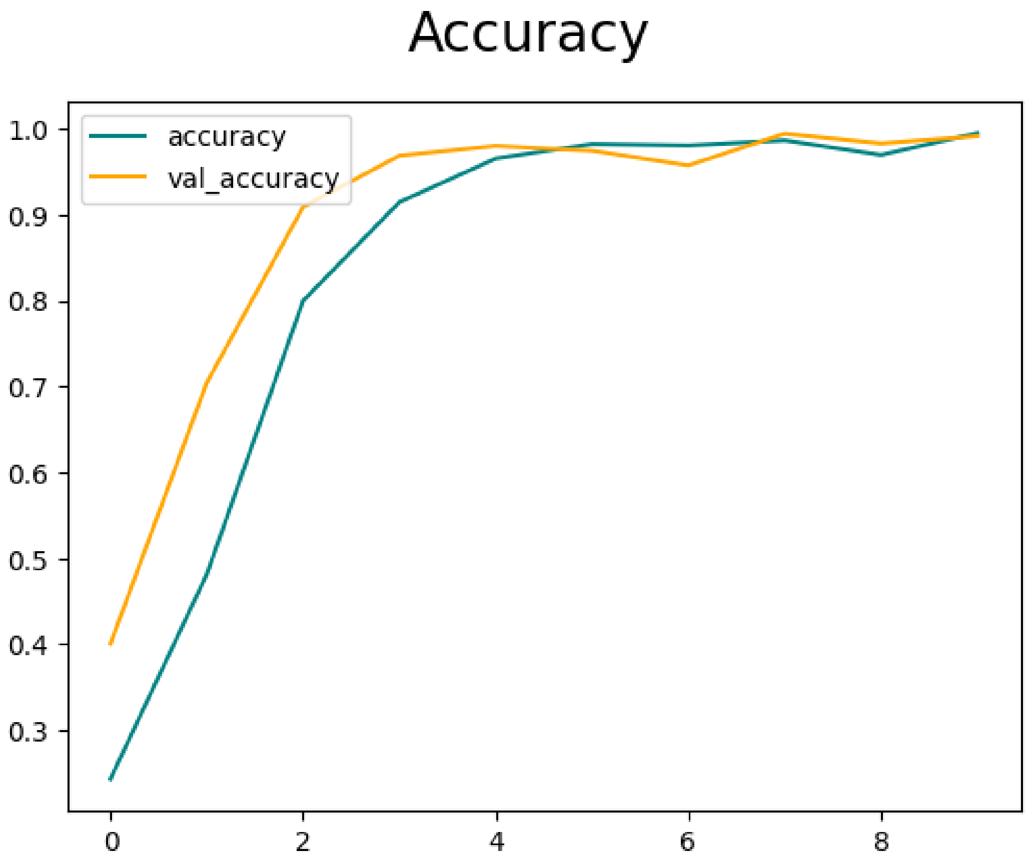


Figure 5: Accuracy Graph for the Model with the epochs considered as 10

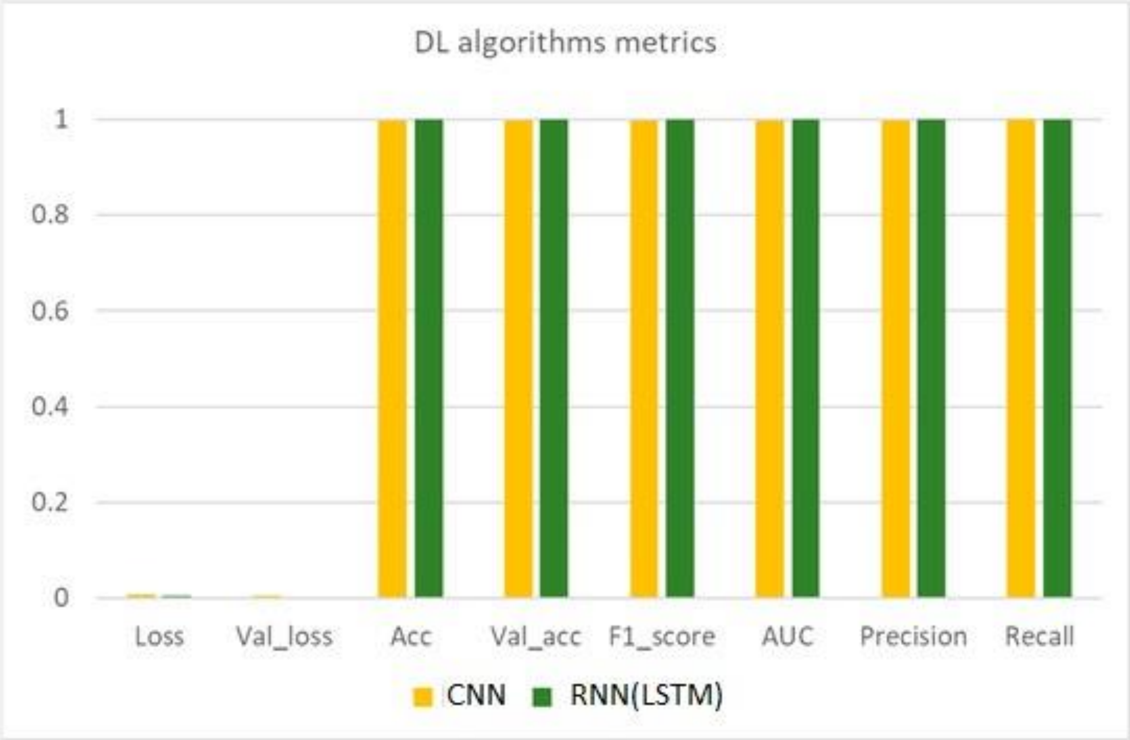


Figure 6: Algorithms comparison

VII.CONCLUSION

This design leverages the benefits of standard convolution, depthwise separable convolution operations, and introduces novel adaptive connections to capture feature maps. Subsequently, the classifier is employed to identify ten distinct driver behaviors. This study incorporated various strategies to decrease network parameters and enhance accuracy. Additionally, the system's performance was evaluated on videos of different resolutions, demonstrating efficient processing speeds.

Numerous potential avenues for future advancements and enhancements exist within the realm of CNN-based Driver Behavior Detection. Here are a few instances: Enhanced Precision: Aiming to elevate the precision of the CNN model remains a central objective for future endeavors. Nuanced Behavior Detection: Expanding the system's capabilities to detect more intricate behaviors beyond the fundamental categories. Real-Time Capability: Critical optimization of both the CNN model and the overall system for real-time performance is essential. This could encompass techniques like model compression (e.g., quantization, pruning), harnessing hardware accelerators (e.g., GPUs, TPUs), or investigating lightweight network architectures designed for resource-restricted environments such as embedded systems or edge devices.

REFERENCES

[1] Duy-linh nguyen , muhamad dwisnanto putro, and kang-hyun jo “Driver Behaviors Recognizer Based on Light-Weight Convolutional Neural Network Architecture and Attention Mechanism”, IEEE, Digital Object Identifier 10.1109/ACCESS.2022.3187185, July 2022

[2] Xianhui Chen, Ying Chen, Wenjun Ma, Xiaomao Fan, Ye Li, “Toward sleep apnea detection with lightweight multi-scaled fusion network”, Elsevier, 8 April 2022. https://doi.org/10.1016/j.knosys.2022.108783

[3] B. Baheti, S. Talbar, and S. Gajre, ‘‘Towards computationally efficient and realtime distracted driver detection with MobileVGG network,’’ *IEEE Trans. Intell. Vehicles*, vol. 5, no. 4, pp. 565–574, Dec. 2020.

[4] B. Baheti, S. Gajre, and S. Talbar, ‘‘Detection of distracted driver using convolutional neural network,’’ in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 1145–11456.

[5] H. M. Eraqi, Y. Abouelnaga, M. H. Saad, and M. N. Moustafa, ‘‘Driver distraction identification with an ensemble of convolutional neural networks,’’ *J. Adv. Transp.*, vol. 2019, 2019.

[6] Y. Abouelnaga, H. M. Eraqi, and M. N. Moustafa, ‘‘Real-time distracted driver posture classification,’’ *CoRR*, vol. abs/1706.09498, Dec. 2018, doi: 10.48550/arXiv.1706.09498.

[7] A. Behera and A. H. Keidel, ‘‘Latent body-pose guided DenseNet for recognizing driver’s fine-grained secondary activities,’’ in *Proc. 15th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Nov. 2018, pp. 1–6.

[8] D. Tran, H. Manh Do, W. Sheng, H. Bai, and G. Chowdhary, ‘‘Real-time detection of distracted driving based on deep learning,’’ *IET Intell. Transp. Syst.*, vol. 12, no. 10, pp. 1210–1219, Dec. 2018.