

CHAUFFEUR BEHAVIOR RECOGNITION USING FACE RECOGNITION AND DEEP LEARNING

Abstract

High levels of focus are necessary for safe driving, but these behaviors are frequently overridden by distractions like tiredness, eating, drinking, talking, and phone calls. Sadly, these distractions play a significant role in the worrying increase in traffic accidents nowadays. The creation of software that can proactively inform drivers is essential to resolving this pressing problem. This study suggests a novel, lightweight architecture for convolutional neural networks that is intended to recognize different driving styles, enabling warning systems to provide accurate information and dramatically lowering traffic collisions. This network effectively recognizes and categorizes driver behaviors by fusing feature extraction and classifier modules. By combining these elements, in the neural network, feature extraction is intended to be more effective and efficient. The model may lower computing costs and enhance information flow through the network by utilizing depth-wise separable convolutions and adaptive connections. Furthermore, the Convolution Block Attention Module can assist the network in prioritizing crucial features, improving performance across a range of computer vision applications. Using a global average pooling layer and a soft max layer, the classifier module successfully calculates class probabilities. This carefully thought-out architecture makes sure that the network parameterization is optimized while yet maintaining good classification accuracy. On three benchmark datasets, the entire network is painstakingly trained and carefully assessed, confirming its dependability and robustness.

Keywords: Accident, CNN (Convolutional Neural Network), driver behaviors, face recognition.

Authors

Lohith C

Professor,
Engineering Department,
Garden City University
Bangalore, India

Amaresh A M

Assistant Professor,
Department of CSE, GSSSIETW,
Mysuru, India

Pundalik Chavan

Assistant Professor,
School of CSE,
REVA University,
Bangalore, India

Triveni N

Professor,
Engineering Department,
Garden City University
Bangalore, India

I. INTRODUCTION

The volume and complexity of the road traffic networks have significantly increased recently. Unfortunately, the frequency of accidents has gradually increased as a result of this expansion. Estimates from the World Health Organisation state that there are about 1.35 million annual fatalities and roughly 50 million collisions globally as a result of traffic accidents. Driver behaviour stands out as the main cause of this rise among the many other causes.

Furthermore, the assertion makes the case that if drivers could keep their concentration on the road, the accident rate may drop by a staggering four times. Alarming figures from the National Highway Transportation and Safety Administration (NHTSA) in the US show that 2,895 people died as a result of distracted driving.

It is possible to achieve great progress in lowering accidents brought on by driver distraction and inattention by combining efforts from several stakeholders, including car manufacturers, researchers, and legislators. This will make roads safer for everyone.

In order to combat driver distraction proactively, attempts are being undertaken to create warning and preventive mechanisms for drivers or increase the amount of vehicle automation. The use of automated driving capabilities is being investigated in order to lessen the dependency on drivers in urgent circumstances. Notably, some modern cars, especially those with high-tech features and equipment, already have rudimentary systems that can recognise some forms of driver inattention, such driver weariness, and issue the necessary alerts.

These programmes are frequently referred to as driver aid systems, which include a variety of automated driving functions like adaptive distance maintaining and lane keeping.

The extremely difficult task of operating a car necessitates constant focus and attention. Driver distraction and inattention must be addressed through a multifaceted strategy that includes education, technical developments, law, and behaviour modification. Increasing public awareness of the risks associated with distracted driving, enforcing stronger regulations and punishments, and incorporating cutting-edge safety systems into vehicles can all significantly lower the risks and make roads safer for everyone. NHTSA statistics show that distractions, fatigue, or absentmindedness all contribute to about 25% of crashes that are reported by the police. The probability of accidents is considerably increased by abnormal driving behaviour, such as drowsy, aggressive, inebriated, careless, or reckless driving. Only a handful of the available driver behaviour detection systems attempt to distinguish between several types; the majority concentrate on identifying one specific form of improper behaviour [5]. As a result, no technology exists for monitoring driver behaviour that can effectively discern between various abnormal driving behaviours.

We enumerate the numerous traits of the various driving styles in order to better comprehend them:

1. Rude or aggressive driving This includes tailgating (driving too closely behind another vehicle), dangerous lane changes, and abrupt changes in vehicle acceleration and speed.

2. **Distracted Driving Methodology** This behaviour has to do with neglecting important driving duties while engaging in other activities, like eating, drinking, or using a smartphone or other technology. When a driver is distracted, they frequently react quickly to correct the problem in the car.

The development of an effective driver behaviour monitoring system continues to face considerable challenges, including the inability to distinguish between these various driving behaviour patterns.

The drowsy driving style, which is frequently the result of driver fatigue, is characterised by obvious signs like yawning, shutting eyes, slower reflexes and steering, sparing use of the brakes, and lower engine revolutions per minute (RPM).

On the other hand, drinking or using drugs has a major negative impact on a driver's mental capacity, which results in drunk driving, which has observable symptoms such inefficient braking, abrupt accelerations, and dangerous lane changes.

The traits that set the safe driving style apart from risky driving styles can also be used to identify it.

II. LITERATURE SURVEY

To carry out, it is very essential to understand the current scenario and the technology merits and demerits. To provide a solid basis for the project's goals, it is crucial to critically analyse and synthesise the data from various sources when doing a literature study. A well-done literature review aids in identifying knowledge gaps, shapes research topics, and directs the course of the project activity. Additionally, it indicates the researcher's knowledge of the body of prior study in the area.

Driving involves a variety of behaviours essential for maintaining road safety and calls for a high level of attention. Unfortunately, distractions like eating, drinking, talking on the phone, adjusting the radio, or being sleepy frequently obscure these behaviours. These interruptions have been identified as the main causes of the worrisome increase in road accidents.

The creation of programmes that can forewarn drivers is necessary to handle this urgent problem. This paper introduces a revolutionary lightweight convolutional neural network architecture for distinguishing various driving actions. The warning system can accurately deliver information and significantly lessen traffic collisions by effectively identifying certain behaviours. The network's successful protection of road users is ensured by the well-balanced combination of feature extraction and classifier modules. For distracted driving detection tasks using the provided benchmark datasets, the architecture and components described as a whole strive to optimise network parameters and maintain good classification accuracy [1]. With very high accuracy on two of the three datasets and a respectably high accuracy on the third, it appears that the model performs remarkably well.

Seek examination and advice from a healthcare provider with expertise in sleep medicine if you think you may have sleep apnea [2] or are exhibiting signs of sleep-

disordered breathing. Both sleep quality and overall health outcomes can be considerably enhanced by early detection and successful treatment. The SE-MSCNN (Single-Lead ECG SE-Multi-Scale Convolutional Neural Network), an innovative and effective technique for detecting supraventricular arrhythmia (SA) in single-lead ECG signals, is introduced in this research. Two essential parts make up the proposed SE-MSCNN: a channel-wise attention module and a multi-scaled convolutional neural network (CNN) module.

The authors use three sub-neural networks to extract variably scaled ECG information, with each network recording different-length neighbouring segments, so performance of SA identification is improved. This method enhances the model's ability to recognise SA patterns by allowing it to take into account various temporal contexts in the data.

The local concentration of feature fusion when utilising conventional concatenation approaches is one issue this study addresses. The authors use a squeeze-to-excitation block in a channel-wise attention module to get around this. Overall, the statement emphasises how well the suggested SE-MSCNN outperformed alternative SA detection techniques on the benchmark dataset for apnea-ECG. This model shows that it has potential for improving SA identification and has the potential to make a big impact on the fields of sleep medicine and healthcare. An evaluation of the model's performance and generalizability would thoroughly done, however, benefit from additional validation on more datasets and comparisons with other pertinent strategies. With the advantages of quick reaction and few parameters, the SE-MSCNN may be embeds into a wearable device to offer a SA detection service for people undergoing home sleep testing (HST).

Overall, the study shows how well the mobileVGG CNN architecture works to identify driving distractions. The method exhibits encouraging accuracy results while being computationally effective, which is essential for real-world applications like ADAS in automobiles. A World Health Organisation (WHO) study indicates that the number of fatal traffic collisions has been progressively increasing in recent years. However, the death rate relative to the world's population has stabilised in recent years. The National Highway Traffic Safety Administration (NHTSA) found that distracted driving is a significant cause of traffic accidents. Addressing driver distraction is necessary to increase traffic safety. The authors of the study outline a strategy for recognising and classifying driver distraction that is based on CNN. It is widely acknowledged that CNNs are effective at tasks involving pattern recognition and image processing. When creating safety features for ADAS, computational efficiency must be taken into account in addition to accuracy. The suggested mobileVGG architecture is designed to be accurate while still being efficient computationally. The performance of the recommended mobileVGG network is assessed by the authors using the Statefarm dataset from Kaggle and the inattentive driving detection dataset from American University in Cairo (AUC). The suggested mobileVGG outperforms earlier techniques with accuracy levels of 95.24% AUC dataset and 99.75% Statefarm dataset. The network also has a relatively low computational complexity and memory need with only 2.2 million parameters. The mobileVGG architecture outperforms earlier techniques with just 2.2 million parameters, achieving accuracy rates of 95% and above on the dataset known as AUC and 99% and above on the dataset known as Statefarm.

A study introducing the mobileVGG Convolutional Neural Network (CNN) architecture is described in the referenced paper. This architecture was developed to assess and categorise driver distraction precisely and fast. This study's major objective is to develop a CNN for ADAS, or advanced driver assistance systems, which increase traffic safety.

The importance of this study is highlighted by studies from the World Health Organisation (WHO), which highlights the disturbing trend of escalating fatal traffic accidents in recent years. However, it is encouraging to note that over this time, the global death rate has stabilised.

Driver distraction must be addressed since, according to the National Highway Traffic Safety Administration (NHTSA), it contributes significantly to traffic accidents.

If we wish to improve road safety, driver distraction must be effectively addressed. The convolutional neural network (CNN) designed specifically for this application, the mobileVGG architecture, is given in this research report. As a result of its outstanding performance in a range of visual identification tasks, CNNs are often used in image processing applications.

When designing safety features for Advanced Driver Assistance Systems (ADAS), the authors emphasise the significance of considering both computational correctness and efficiency. The basic objective is to create a CNN that can recognise driving distractions while using memory effectively.

The authors use two datasets—the inattentive driver detection dataset from the American University in Cairo (AUC) and the Statefarm dataset from Kaggle—to assess the effectiveness of their suggested mobileVGG network.

1.25 million people worldwide may away as a result of road accidents each year, and the number has been continuously climbing in recent years, according to the World Health Organisation (WHO). A fifth of these collisions are the result of distracted drivers. The present distracted driving detection study focuses on a small number of distractions, mostly cell phone use. Many improvised methods are unreliable. The authors of the study present the first dataset for driver distraction identification that is publicly available and contains more distracted postures than competitor datasets [5]. Furthermore, authors provide a solid, 90% accurate deep learning-based solution. We show that using a weighted ensemble of classifiers and a genetic approach, the classification confidence is increased. The system consists of a group of convolutional neural networks that have been genetically weighted. The authors also look into how different visual elements influence distraction detection through the localization of hands and faces as well as skin segmentation. Additionally, the authors offer a condensed version of our ensemble that operates in real-time and achieves a classification accuracy of 84.64%.

The study's authors [6] provide a fresh dataset for calculating "distracted driver" posture. For the assessment of driving position, we also propose a novel method with a classification accuracy of 95.78%. Higher classification confidence is made possible by the system, which is an ensemble of convolutional neural networks that has been genetically weighted. This research has a number of potential applications, including driver assistance

systems, vehicle safety, and automotive HMIs. The authors show that a weighted ensemble of classifiers using a genetic approach results in improved classification confidence. The authors also explore how different visual elements, such as hands and faces, influence distraction detection and categorization through the use of face and hand localizations. A pared-down ensemble that can operate in real-time and achieve a classification accuracy of 94.29% is also suggested by the authors.

The development of self-driving automobiles has undergone more research in the previous 20 years, with several businesses and academic institutions pushing the edge. The automatic identification of in-vehicle actions is crucial to the development of such vehicles. The authors describe a novel human-pose driven strategy for video-based monitoring of the driver's state/activity in response to the recent success of deep Convolutional Neural Networks (CNN) in visual identification tasks. The method infers the driver's status and activity from a single frame, enabling real-time operation. Additionally, the authors incorporate concepts from current research on visual recognition and transfer learning for human pose detection. These concepts are combined by the modified DenseNet [7], where one stream is concentrated on the latent body position and the other stream concentrates on appearance data. The suggested strategy is thoroughly tested on two difficult datasets made up of different ancillary non-driving activities. The experimental results show that the addition of latent body-pose into the current deep networks greatly improves the performance of driver activity recognition.

Driver distraction detection systems have the potential to significantly increase road safety by offering real-time monitoring and notifications to drivers who may be participating in distracting behaviour. These tools can help motorists increase their awareness of their level of focus and promote safer driving practises. The system's data collection might be used to advance research and develop cutting-edge driver assistance technologies (ADAS) that actively intervene to lessen distractions and improve traffic safety. To create realistic driving scenarios and validate the distraction detection algorithms, an assisted driving testbed is being constructed. Images of the drivers in both normal and distracted driving postures were gathered for the authors' dataset. On a platform with embedded graphics processing units, VGG-16, AlexNet, GoogleNet, and residual network are four deep convolutional neural networks [8] that are implemented and assessed. They also created a conversational warning system that notifies the driver in real-time when he or she is not paying attention to the task of driving. The proposed methodology outperforms the baseline method, which uses just 256 neurons in the fully linked layers, according to experimental results. Additionally, the outcomes show that the GoogleNet model outperforms the other three in the driving simulator testbed for distraction detection.

III. PROPOSED METHODOLOGY

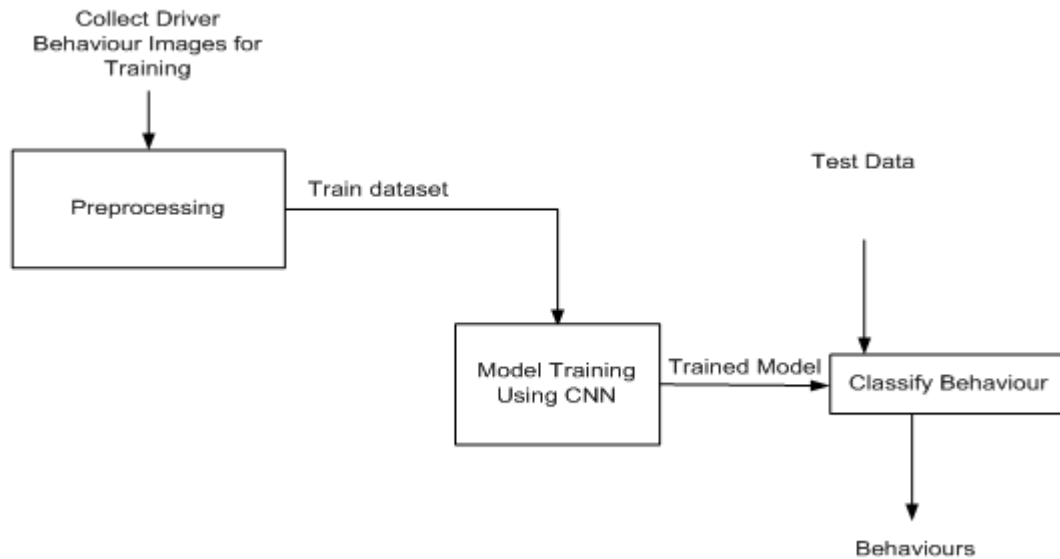


Figure1: Block Diagram

The above block diagram mentions the modules as described below:

- 1. Data Augmentation:** The data augmentation approach is powerful and may be used to artificially expand the size of a dataset. It is especially helpful when working with sparse training data and enhances the performance and durability of deep learning models. Here, fresh images are created using ImageDataGenerator. Since the images are spirals, they may be rotated in any direction without altering their meaning, hence the rotation range is set at 360 degrees. You can experiment with the additional picture transformations offered by the ImageDataGenerator class. However, exercise caution when using any augmentation because some changes may reduce the CNN model's poorer accuracy.

The distribution of the data following augmentation Additionally, the photos are resized to a standard size (128, 128, 1). Before fitting the dataset to the model, the images are further normalised.

- 2. Face Recognition:** Haar cascades have been widely used for object detection tasks, especially for face detection, due to their speed and accuracy. While Haar cascades were groundbreaking when introduced, more advanced object detection methods like the Single Shot Multibox Detector (SSD) and the You Only Look Once (YOLO) algorithm have surpassed them in terms of speed and accuracy. Nonetheless, Haar cascades remain a valuable and efficient option for certain applications and environments.

Here are the main steps involved in face recognition using Haar cascades:

- Haar Cascade Training:** Haar cascades are trained using a big collection of photos, both good and negative. While faces are absent from the negative photos, they are there in the positive ones. Extraction of Haar-like characteristics is a step in the

training process from these images and using machine learning algorithms, such as AdaBoost, to create a strong classifier that can distinguish between faces and non-faces based on these features.

- **Haar Cascade XML File:** Once the training process is complete, the resulting classifier is saved as an XML file, known as the Haar cascade XML file. This XML file contains the information about the learned features and the classifier thresholds.
- **Face Detection:** To detect faces in an input image using Haar cascades, the image is passed through a sliding window technique. At each step, a rectangular window of different sizes is moved across the image, and Haar-like features are calculated for each window.
- **Integral Image:** Calculating Haar-like features for each window can be computationally expensive. To optimize the process, an integral image is used. The integral image allows for the quick calculation of the Haar-like features within any rectangular region of the image.
- **Applying the Classifier:** The Haar-like features calculated for each window are compared with the learned features stored in the Haar cascade XML file. Based on the classifier thresholds, the algorithm determines whether each window contains a face or not. If a window is classified as a face, it is marked as a potential face region.
- **False Positive Removal:** The face detection process may generate some false positive detections, where non-face regions are classified as faces. To mitigate this, additional techniques such as non-maximum suppression or overlapping region removal can be applied to eliminate redundant or overlapping detections and refine the final set of detected faces.
- **Face Recognition:** Once the faces have been detected, further steps can be applied for face recognition. This may involve extracting facial features, such as landmarks or descriptors, from the detected faces and comparing them against a database of known faces using techniques like eigenfaces, Fisherfaces, or deep learning-based approaches.

Haar cascades provide a relatively fast method for face detection, and they have been widely used in various applications, including face recognition, facial expression analysis, and facial attribute detection.

3. **Deep Learning Method for Training and Classification:** See while Haar cascades were groundbreaking at the time of their introduction, newer and more advanced methods, such as deep learning-based approaches using convolutional neural networks (CNNs) or pre-trained models like MTCNN or OpenFace, have shown superior performance in face detection and recognition tasks.

- **CNN Model Architecture**

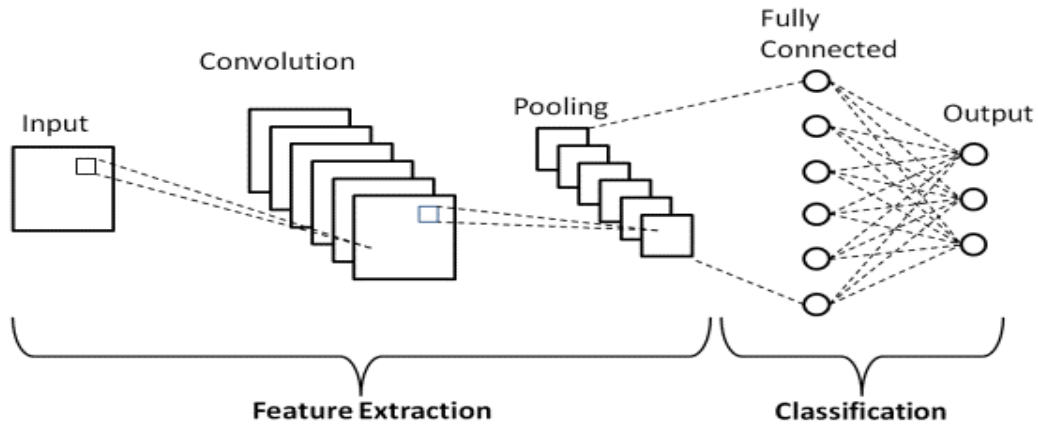


Figure 2: BasicArchitecture of CNN

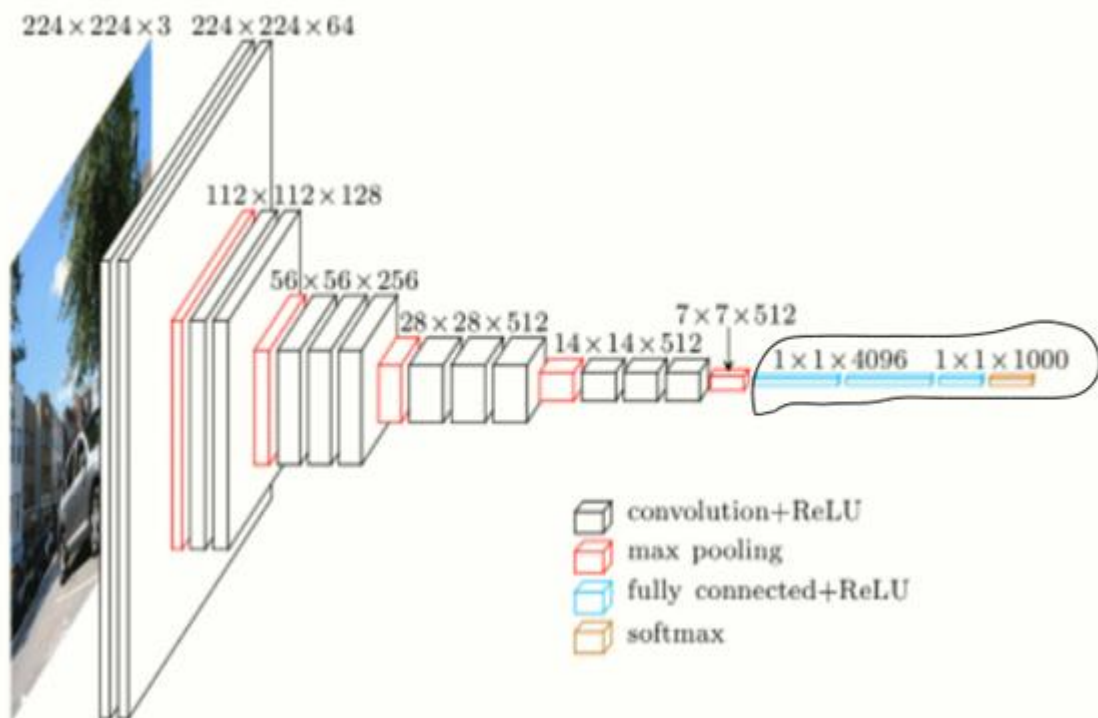


Figure 3: Model with four Convolutional Layers in CNN

The application makes use of a CNN model architecture that has the following features:

- Four convolutional layers, each containing 128 64 32 and 32 filters, make up the model.
- Filters of various filter sizes are present in the convolutional layers.
- Each convolutional layer is followed by a MaxPool2D layer.
- The convolutional block is followed by two fully connected layers.

IV. ALGORITHMS

CNN Algorithm Working steps

The steps involved in training a Convolutional Neural Network (CNN) algorithm. CNNs are widely used for various computer vision tasks, including image classification, object detection, and image segmentation.

Here's a general outline of the steps:

- 1. Data Preprocessing:** Compile and prepare your training data first. In order to improve the diversity of your training samples, you must first gather a labelled collection of images and take the necessary preprocessing processes, such as resizing, normalising pixel values, and augmenting the data (e.g., rotation, scaling, and flipping).
- 2. Architecture Design:** Select the CNN's architectural style. Counting convolutional layers, pooling layers, fully connected layers, and output layers is part of this process. Think on how difficult the work is, how big the input photos are, and how many classes you wish to identify.
- 3. Model Initialization:** Set the initial weights for the layers of your CNN model. Initialization techniques that are frequently used include random initialization and transfer learning with pre-learned weights from models trained on related tasks.
- 4. Forward Propagation:** Move information forward through the CNN's layers. This entails running the input image through the convolutional layers, adding non-linearity with activation functions like ReLU, and downsampling the feature maps with pooling layers like MaxPooling.
- 5. Flattening:** Create a 1D vector from the output feature maps after applying the convolutional and pooling layers. This gets the data ready for the layers that are fully connected.
- 6. Connect the output that has been flattened to one or more completely linked layers in step six.** Each neuron in these layers is connected to every other neuron in the layer below, acting as in a conventional neural network. Apply activation techniques to these layers as well, such as ReLU.
- 7. Add an output layer with the necessary quantity of neurons.** The softmax activation function, which yields class probabilities, is a popular choice for classification applications. A linear activation function may be utilised for regression tasks.
- 8. Loss Function:** Based on the specifics of your issue, choose an appropriate loss function. Cross-entropy loss is frequently utilised for classification jobs while mean square error (MSE) is frequently employed for regression activities.
- 9. Using backpropagation,** determine the gradients of the loss function with respect to the weights and biases. This stage entails changing the weights using optimisation techniques such stochastic gradient descent (SGD), Adam, or RMSprop, and propagating the mistakes backward through the network.

- 10. Training:** To train your CNN, feed the training data into the network iteratively, compute the loss, and use backpropagation to update the weights. This process continues until convergence or a predetermined number of epochs.
- 11. Validation:** After each training iteration, assess how well your CNN performed using a different validation dataset. This enables early stopping or hyperparameter adjustment while also assisting in monitoring the generalisation of the model.
- 12. Testing:** To evaluate the trained CNN's performance on unseen data, examine it on a different testing dataset. To assess the model's efficacy, compute measures like accuracy, precision, recall, and F1 score.

V. RESULTS AND DISCUSSION

The loss and accuracy for the model with the epochs were achieved with the use of the aforementioned techniques in a more efficient manner, as demonstrated in the graphs below.

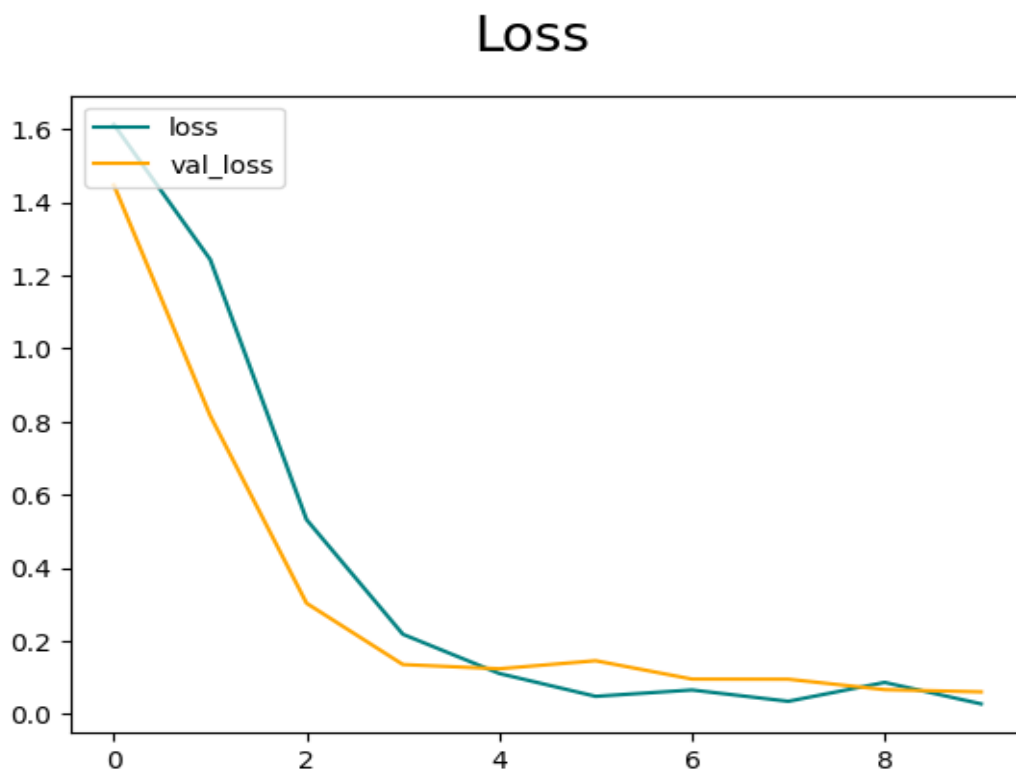


Figure 4: Loss Graph for the Model with the epochs considered as 10

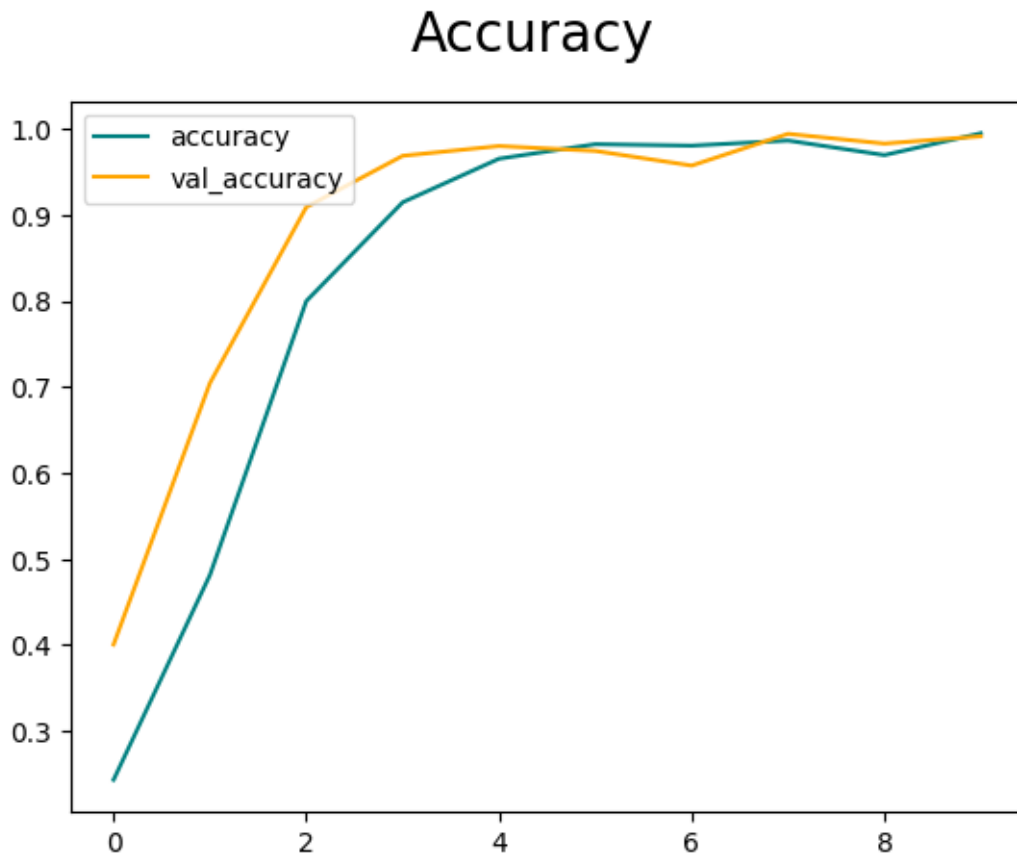


Figure 5: Accuracy Graph for the Model with the epochs considered as 10

VI. CONCLUSION

This architecture employs proposed adaptive connections, depthwise separable convolution operation, and standard convolution to extract feature maps. The next step is to identify ten driving behaviour using the classifier. Numerous techniques were utilised in this study to reduce the amount of network parameters and increase accuracy. However, it was also tested on videos of various resolutions and processing rates.

There are several potential areas for future work and improvements in the field of Driver Behavior Detection using CNN. Here are a few examples: **Improved Accuracy:** Enhancing the accuracy of the CNN model is always a focus for future work. **Fine-Grained Behavior Detection:** Expanding the system to detect more fine-grained behaviors beyond the basic classes. **Real-Time Performance:** Optimizing the CNN model and the overall system for real-time performance is crucial. This may involve model compression techniques (e.g., quantization, pruning), leveraging hardware accelerators (e.g., GPUs, TPUs), or exploring lightweight network architectures tailored for resource-constrained environments, such as embedded systems or edge devices

REFERENCES

- [1] Duy-linhnguyen ,muhamaddwisnantoputro, 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.