# Chauffeur Behavior Recognition using Face Recognition and Deep Learning

Lohith C1,

1Assistant Professor, Department of CSE,

Gopalan College of Engineering and Management, Bangalore, India Email: [lohithchandrashekhar@gmail.com](mailto:lohithchandrashekhar@gmail.com)

# Amaresh A M2

2Assistant Professor, Department of CSE, GSSSIETW,

Mysuru, India

Email: [amarreva2021@gmail.com](mailto:amarreva2021@gmail.com)

Pundalik Chavan3

3Assistant Professor, School of CSE,

REVA University,

Bangalore, India

Email: [Pundalik.chavan29@gmail.com](mailto:Pundalik.chavan29@gmail.com)

# Triveni N4

4Assistant Professor, Department of ISE,

Cambridge Institute of Technology,

Bangalore, India

Email: [triveninagaraj33@gmail.com](mailto:triveninagaraj33@gmail.com)

# **ABSTRACT**

Driving is a set of behaviors that need high concentration. Sometimes these behaviors are dominated by other acts such as smoking, eating, drinking, talking, phone calls, adjusting the radio, or drowsiness. These are also the main causes of current traffic accidents. Therefore, developing applications to warn drivers in advance is essential. This research introduces a light-weight convolutional neural network architecture to recognize driver behaviors, helping the warning system to provide accurate information and to minimize traffic collisions. This network is a combination of feature extraction and classifier modules. By combining these elements, feature extraction in the neural network is intended to be more effective and efficient. The model may lower computing costs and enhance information flow through the network by utilising depth-wise separable convolutions and adaptive connections. Furthermore, the Convolution Block Attention Module can assist the network in prioritising crucial features, improving performance across a range of computer vision applications. The classifier module is comprised of a global average pooling and soft max layer to calculate the probability of each class. The overall design optimizes the network parameters and maintains classification accuracy. The entire network is trained and evaluated on three benchmark datasets.

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

## INTRODUCTION

Nowadays, road traffic systems have grown much in terms of quantity and complexity. Accordingly, the number of accidents also increased gradually. The World Health Organisation estimates that road traffic crashes cause over 1.35 million fatalities yearly and cause 50 million collisions overall. Driver behaviour is one of the primary factors contributing to a rise in accidents.

The statement above also mentioned that if the drivers really focused when driving, it could reduce the accident rate by four times. According to a statistic from the National Highway Transportation and Safety Administration (NHTSA) in the United States (US), about 2,895 people were killed in distracted driving accidents in 2019, accounting for 8.7% of all traffic accident deaths in that year.

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.

Thereby, they develop warning and prevention mechanisms for driver or increase the automation level of the vehicle to avoid dealing with driver distraction in the first place, e.g., with automated driving functionalities. Some modern vehicles above a certain vehicle class and equipment level may already have simple systems that can detect certain types of driver inattention, such as driver fatigue, and warn the driver accordingly.

Such systems are usually subsumed under the term driver assistance systems, which also include vehicle automation systems such as adaptive distance keeping or lane keeping.

Driving a vehicle is an extremely complex task. 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. According to NHTSA statistics, approximately 25% of police-reported crashes involve some form of driver inattention such as the driver is distracted, fatigued, or otherwise lost in thought. Abnormal driving behavior (e.g. drowsy, aggressive, drunk, careless, and reckless driving) is defined by the driver behavior, which increases the risk of the accidents. Most driver behavior detection systems only identify one type of abnormal driver behavior, whereas few papers attempt to distinguish among the different types of driver behavior [[5](https://ietresearch.onlinelibrary.wiley.com/doi/full/10.1049/ell2.12076#ell212076-bib-0005)]. Nevertheless, there is still no driver behavior monitoring system that can efficiently distinguish between different abnormal driver behaviors.

In order to clarify different styles of driver behaviour, we summarize the various characteristics of driving styles in the following.Aggressive driving behaviour includes unsafe lane change, quick change in car acceleration and speed, tailgating (drive too closely behind another vehicle).Distracted driving style is associated with inattention to necessary activities and task of driving toward other activities such as drinking, eating, and using smartphone or technologies in the car. Distracted driving usually followed by a quick driver reflex to rectify car situation.

Driver fatigue usually leads to the drowsy driving style that is accompanied by observable symptoms such as yawning, closing eyes, slower reactions and steering, rare use of brake, and lower revolutions per minute (RPM).

Using alcohol or drug affects the driver mental ability and leads to the drunk driving style that is accompanied by measurable symptoms such as lower use of brake, abrupt acceleration, and dangerous lane change. Finally, safe driving style could be identified by detecting the characteristics of risky driving styles.

## 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 is a set of behaviors that need high concentration. Sometimes these behaviors are dominated by other acts such as smoking, eating, drinking, talking, phone calls, adjusting the radio, or drowsiness. These are also the main causes of current traffic accidents. Therefore, developing applications to warn drivers in advance is essential. This research introduces a light-weight convolutional neural network architecture to recognize driver behaviors, helping the warning system to provide accurate information and to minimize traffic collisions. This network is a 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. This paper proposes a lightweight SA detection method of multi-scaled fusion network named SE-MSCNN based on single-lead ECG signals. The proposed SE-MSCNN mainly includes multi-scaled [convolutional neural network](https://www.sciencedirect.com/topics/engineering/convolutional-neural-network) (CNN) module and channel-wise attention module. In order to facilitate the SA detection performance, various scaled ECG information with different-length adjacent segments are extracted by three sub-neural networks. To overcome the problem of local concentration of [feature fusion](https://www.sciencedirect.com/topics/computer-science/feature-fusion) with concatenation, a channel-wise attention module with a squeeze-to-excitation block is employed to fuse the different scaled features adaptively. Furthermore, the ablation study and computational complexity analysis of the SE-MSCNN are conducted. Overall, the statement emphasises how well the suggested SE-MSCNN outperformed alternative SA detection techniques on the benchmark dataset for apnea-ECG. This shows that the model has potential for improving SA identification and has the potential to make a big impact on the fields of sleep medicine and healthcare. A thorough evaluation of the model's performance and generalizability would, however, benefit from additional validation on more datasets and comparisons with other pertinent strategies. The SE-MSCNN with the merits of quick response and lightweight parameters can be potentially embedded into a wearable device to provide an SA detection service for individuals in home sleep test (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. According to a World Health Organisation (WHO) research, the number of fatal traffic accidents has been rising steadily in recent years. However, in recent years, the death rate in relation to the global population has stabilised. Distracted driving is cited as a major contributing factor in traffic accidents, according to a survey conducted by the National Highway Traffic Safety Administration (NHTSA). To improve road safety, driver distraction must be addressed. The research paper's authors describe a CNN-based method for identifying and categorising driver distraction. The efficiency of CNNs in image processing and pattern recognition tasks is well recognised. Computational efficiency must be taken into account in addition to accuracy while developing safety features for ADAS. The proposed mobileVGG architecture is made to be accurate while yet being computationally effective. The inattentive driver detection dataset from American University in Cairo (AUC) and Statefarm's dataset from Kaggle are used by the authors to assess the performance of the proposed mobileVGG network. With 95.24% accuracy on the AUC dataset and 99.75% accuracy on the Statefarm dataset, the proposed mobileVGG beats earlier methods. With only 2.2 million parameters, the network also has a relatively low computational complexity and memory need. With only 2.2 million parameters, the mobileVGG architecture beats earlier methods, reaching accuracy rates of 95.24% on the AUC dataset and 99.75% on the Statefarm dataset.

The article that is supplied covers a study that introduces the mobileVGG Convolutional Neural Network (CNN) architecture for identifying and categorising driver distraction. The authors want to create a CNN for safety features in Advanced Driver Assistance Systems (ADAS) that is accurate and computationally effective. According to a World Health Organisation (WHO) research, the number of fatal traffic accidents has been rising steadily in recent years. However, in recent years, the death rate in relation to the global population has stabilised. Distracted driving has been identified as a major contributing factor in traffic accidents, according to a survey conducted by the National Highway Traffic Safety Administration (NHTSA). To improve road safety, driver distraction must be addressed. The research study introduces the convolutional neural network-based mobileVGG architecture. CNNs have demonstrated excellent performance in a number of visual recognition tasks and are frequently utilised in image processing jobs. Computational efficiency and accuracy must both be taken into account when designing safety features for Advanced Driver Assistance Systems. The authors want to create a CNN that is quick and efficient in terms of memory utilisation, in addition to being accurate when classifying driving distraction. The inattentive driver detection dataset from the American University in Cairo (AUC) and Statefarm's dataset from Kaggle are used by the authors to assess the performance of their suggested mobileVGG network.

The World Health Organization (WHO) reported 1.25 million deaths yearly due to road traffic accidents worldwide and the number has been continuously increasing over the last few years. Nearly fifth of these accidents are caused by distracted drivers. Existing work of distracted driver detection is concerned with a small set of distractions (mostly, cell phone usage). Unreliable*ad hoc* methods are often used. In the paper, [5] Authors present the first publicly available dataset for driver distraction identification with more distraction postures than existing alternatives. In addition, authors propose a reliable deep learning-based solution that achieves a 90% accuracy. The system consists of a genetically weighted ensemble of convolutional neural networks; we show that a weighted ensemble of classifiers using a genetic algorithm yields a better classification confidence. Authors also study the effect of different visual elements in distraction detection by means of face and hand localizations, and skin segmentation. Finally, authors present a thinned version of our ensemble that could achieve 84.64% classification accuracy and operate in a real-time environment.

In the paper [6] , authors present a new dataset for "distracted driver" posture estimation. In addition, we propose a novel system that achieves 95.98% driving posture estimation classification accuracy. A genetically weighted ensemble of convolutional neural networks, which allows for increased classification confidence, is what makes the system special. Applications of this research are possible in a number of fields, including driver assistance systems, automobile safety, and human-machine interfaces in cars. Authors show that a weighted ensemble of classifiers using a genetic algorithm yields in better classification confidence. Authors also study the effect of different visual elements (i.e. hands and face) in distraction detection and classification by means of face and hand localizations. Finally, Authors present a thinned version of our ensemble that could achieve a 94.29% classification accuracy and operate in a realtime environment.

Over the past two decades, there has been increasing research in developing self-driving vehicles, with many industries pushing the bounds alongside academia. Automatic recognition of in-vehicle activities plays a key role in developing such vehicles. In this work, authors propose a novel human-pose driven approach for video-based monitoring of driver’s state/activity and is inspired by the recent success of deep Convolutional Neural Network (CNN) in visual recognition tasks. The approach infers the driver’s state/activity from a single frame and thus, could operate in real-time. Authors also bring together ideas from recent works on human pose detection and transfer learning for visual recognition. The adapted DenseNet [7] integrates these ideas under one framework, where one stream is focused on the latent body pose and the other stream is on appearance information. The proposed method is extensively evaluated on two challenging datasets consisting various secondary non-driving activities. The experimental results demonstrate that the driver activity recognition performance improves significantly when the latent body-pose is integrated into the existing deep networks.

By providing real-time monitoring and alerts to drivers who may be engaging in distracting behaviours, driver distraction detection systems have the potential to greatly improve traffic safety. These devices can assist drivers become more aware of their degree of attention and encourage safer driving habits. Further study and the development of advanced driver assistance systems (ADAS) that actively intervene to reduce distractions and enhance traffic safety could both benefit from the data acquired by the system. An assisted driving testbed is developed for the purpose of creating realistic driving experiences and validating the distraction detection algorithms. The authors collected a dataset which consists of images of the drivers in both normal and distracted driving postures. Four deep convolutional neural networks[8] including VGG-16, AlexNet, GoogleNet, and residual network are implemented and evaluated on an embedded graphic processing unit platform. In addition, they developed a conversational warning system that alerts the driver in real-time when he/she does not focus on the driving task. Experimental results show that the proposed approach outperforms the baseline one which has only 256 neurons in the fully-connected layers. Furthermore, the results indicate that the GoogleNet is the best model out of the four for distraction detection in the driving simulator testbed.

## PROPOSED METHODOLOGY



**Figure 1: 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, ImageDataGenerator is used to produce new images. The rotation range is set to 360 as the images are spiral and can be rotated any number of degrees without changing the image’s meaning. You can try other image transformations available under ImageDataGenerator class. But be careful while applying any augmentation as certain transformations may lower the CNN model’s lower accuracy.

The data distribution after augmentation —The images are also reshaped to a common size (128, 128, 1). Images are further normalised before fitting the dataset to the model.

1. **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:

**1. Haar Cascade Training:** Haar cascades are trained using a large dataset of positive and negative images. The positive images contain examples of faces, while the negative images do not. The training process involves extracting Haar-like features 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.

**2. 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.

**3. 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.

**4. 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.

**5. 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.

**6. 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.

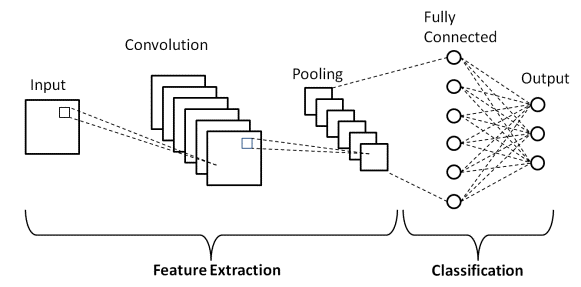
**7. Face Recognition:** Once the faces have been detected, further steps can be applied for face recognition. This may involve extracting facial features, such as landmarks or descriptors, from the detected faces and comparing them against a database of known faces using techniques like eigenfaces, 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.

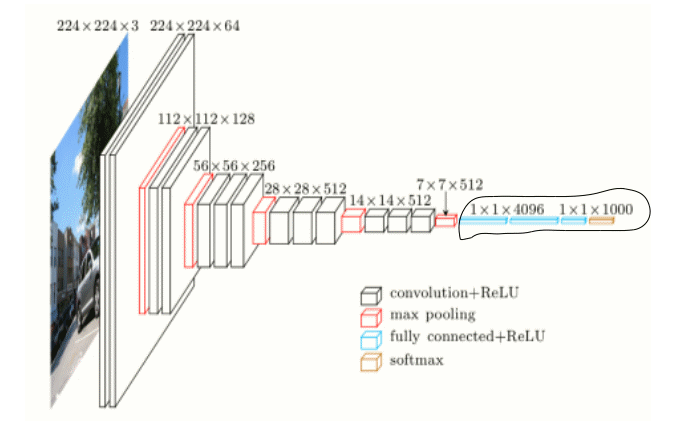
1. **Deep Learning Method for Training and Classification**

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: Basic CNN Architecture**

**Figure 3: CNN Model with four Convolutional Layers**

The implementation uses a CNN model architecture with the following characteristics —

* The model contains four Convolutional Layers with 128, 64, 32, and 32 filters, respectively.
* The convolutional layers contain filters with varying filter sizes.
* A MaxPool2D layer follows each convolutional layer.

Two Fully Connected layers follow the convolutional block

## 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:** Start by gathering and preprocessing your training data. This involves collecting a labeled dataset of images and performing necessary preprocessing steps such as resizing, normalizing pixel values, and augmenting the data (e.g., rotation, scaling, flipping) to increase the diversity of your training samples.

**2. Architecture Design:** Decide on the architecture of your CNN. This involves determining the number of convolutional layers, pooling layers, fully connected layers, and output layers. Consider the complexity of the task, the size of the input images, and the number of classes you want to classify.

**3. Model Initialization:** Initialize your CNN model and set the initial weights for the layers. Common initialization methods include random initialization or using pre-trained weights from models trained on similar tasks (transfer learning).

**4. Forward Propagation:** Perform forward propagation through the layers of the CNN. This involves passing the input image through the convolutional layers, applying activation functions (e.g., ReLU) to introduce non-linearity, and pooling layers (e.g., MaxPooling) to downsample the feature maps.

**5. Flattening:** After the convolutional and pooling layers, flatten the output feature maps into a 1D vector. This prepares the data for the fully connected layers.

**6. Fully Connected Layers:** Connect the flattened output to one or more fully connected layers. These layers act as a traditional neural network, where each neuron is connected to every neuron in the previous layer. Apply activation functions (e.g., ReLU) to these layers as well.

**7. Output Layer:** Add an output layer with the appropriate number of neurons. For classification tasks, a common choice is the softmax activation function, which produces class probabilities. For regression tasks, a linear activation function may be used.

**8. Loss Function:** Select an appropriate loss function based on the nature of your problem. For classification tasks, cross-entropy loss is commonly used, while mean squared error (MSE) is often used for regression tasks.

**9. Backpropagation:** Perform backpropagation to calculate the gradients of the loss function with respect to the weights and biases. This step involves propagating the errors backward through the network, updating the weights using optimization algorithms such as stochastic gradient descent (SGD), Adam, or RMSprop.

**10. Training:** Train your CNN by iteratively feeding the training data through the network, calculating the loss, and updating the weights using backpropagation. This process continues for a fixed number of epochs or until convergence.

**11. Validation:** After each training epoch, evaluate the performance of your CNN on a separate validation dataset. This helps monitor the model's generalization and allows for early stopping or hyperparameter tuning.

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

## 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

## 

## 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

## CONCLUSION

This architecture exploits the advantages of standard convolution, depthwise separable convolution operation, and proposed adaptive connections to extract feature maps. Finally, the classifier is applied to recognize ten driver behaviors. This work applied several techniques for reducing the number of network parameters and increasing the accuracy. On the other hand, it was also tested on different resolution videos with good processing speeds.

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-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.