Intelligent Systems with Applications 14 (2022) 200075

Contents lists available at ScienceDirect

Intelligent Systems with Applications

journal homepage: www.elsevier.com/locate/iswa

Automatic driver distraction detection using deep convolutional neural networks

Md. Uzzol Hossain[a], Md. Ataur Rahman[a], Md. Manowarul Islam[a,*], Arnisha Akhter[a], Md. Ashraf Uddin[b], Bikash Kumar Paul[c]

[a] Department of Computer Science and Engineering, Jagannath University, Dhaka, Bangladesh
[b] Internet Commerce Security Laboratory, Federation University, Australia
[c] Department of Information and Communication Technology, Mawlana Bhashani Science and Technology University, Bangladesh

ARTICLE INFO

Article history:
Received 17 April 2021
Revised 29 January 2022
Accepted 2 April 2022
Available online 4 April 2022

Keywords:
Driver distraction
Deep learning
Convolutional neural network
Transfer learning
Resnet50
MobileNetV2
Accuracy

ABSTRACT

Recently, the number of road accidents has been increased worldwide due to the distraction of the drivers. This rapid road crush often leads to injuries, loss of properties, even deaths of the people. Therefore, it is essential to monitor and analyze the driver's behavior during the driving time to detect the distraction and mitigate the number of road accident. To detect various kinds of behavior like- using cell phone, talking to others, eating, sleeping or lack of concentration during driving; machine learning/deep learning can play significant role. However, this process may need high computational capacity to train the model by huge number of training dataset. In this paper, we made an effort to develop CNN based method to detect distracted driver and identify the cause of distractions like talking, sleeping or eating by means of face and hand localization. Four architectures namely CNN, VGG-16, ResNet50 and MobileNetV2 have been adopted for transfer learning. To verify the effectiveness, the proposed model is trained with thousands of images from a publicly available dataset containing ten different postures or conditions of a distracted driver and analyzed the results using various performance metrics. The performance results showed that the pre-trained MobileNetV2 model has the best classification efficiency.

1. Introduction

Every day, thousands of people die due to road accidents, with the Middle and Lower-Middle-Class countries having the greatest mortality rates (Agrawal et al., 2019; Eraqi, Abouelnaga, Saad & Moustafa, 2019; Yan, Coenen & Zhang, 2016). Many car accidents occur worldwide, and 90% of them are caused by human or driver mistakes. According to WHO (World Health Organization, 2020), the number of deaths caused by road accidents is approximately 1.35 million each year and on average, 64 people die every day. Road accident is a major problem in many developing countries including Bangladesh. More than 5000 people were killed in road accidents across Bangladesh in 2019, an unprecedented increase from the previous year's projections. The number of deaths in road crashes increased to 5227, with 788 more people killed than in 2018. The annual report (Kamruzzaman, 2020) of Nirapad Sarak Chai (We Want Safe Roads), a non-profit organization dedicated to road safety stated that the key causes of road accidents are unqualified and unskilled drivers, vehicles with technological faults, poor traffic control, a lack of public awareness, reckless vehicle speeds on highways, faulty road constructions, and inadequate enforcement of traffic laws.

According to the statistics of Bangladesh Jatri Kalyan Samiti (The Daily Star, 2020), in 2016- the number of road accidents was 4312 and the number of deaths was 6055, in 2017- the number of road accidents was 4979 and the number of road death was 7397, in 2018- the number of road accidents was 5514 and number of deaths was 7221, in 2019- number of road accidents was 5516 and number of road death was 7855 shown in Fig. 1. Of all, 18.99% buses, 29.81% trucks, 21.4% motorcycle, 9.35% auto-rickshaw and others vehicle are involved in road accidents. Therefore, the graph showed that 90% of the road accident are occurred due to abnormal behavior of the driver such as driver's eating or drinking, operating radio, talking to others, texting or talking over cell phone etc. To reduce such non-professional activities, machine learning can play a significant role by detecting the distracted driver activ-

* Corresponding author.
E-mail address: manowar@cse.jnu.ac.bd (Md.M. Islam).

https://doi.org/10.1016/j.iswa.2022.200075

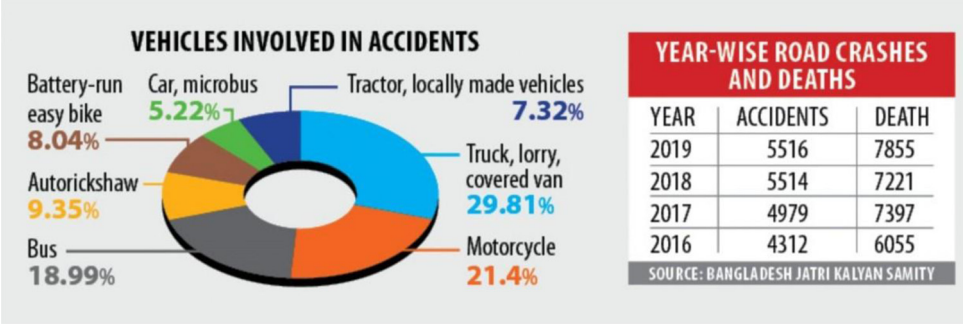


Fig. 1. Statistic of Road Accident in Bangladesh.

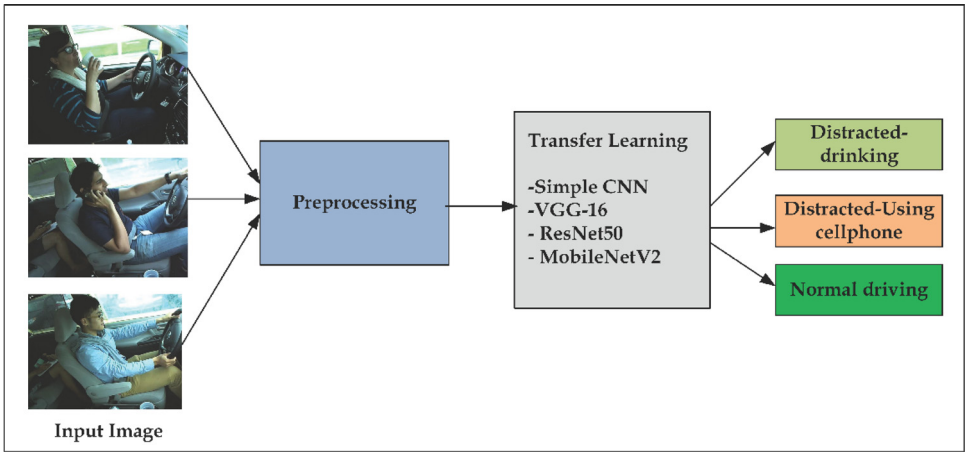


Fig. 2. Proposed Schematic Diagram.

ities and notify the driver to stop while these activities occur to prevent accidents.

Machine learning especially deep learning is concerned with the development and implementation of models or algorithms that enable a system to gain the necessary knowledge based on prior experience or data sets. Monitoring and analyzing the distracted driver behavior and hence addressing the problem using machine learning has been popular and attracted the researchers as an emerging solution nowadays. Using smart system like onboard camera or computer can be used to collect driving data and drivers' postures. This information can be used to train a variety of machine learning models such as- support vector machine, neural network and deep neural networks to learn driving behavior and identify abnormal activities like- sleeping, talking to others, or eating during the driving (Alvarez, Garcia, Naranjo, Anaya & Jimenez, 2014; Kouchak & Gaffar, 2019; Morton, Wheeler & Kochenderfer, 2016). In this concern, our primary goal is to design a convolutional neural network-based architecture that can easily detect and identify the distracted driver. Recently, researchers are exploring machine learning methodology to realize driving behaviors, building automated driver assistance facilities to mitigate road traffic incidents. Deep learning algorithms have demonstrated impressive performance by extracting various useful information from real-time image processing and computer vision due to the huge computational capabilities.

To detect various kinds of distracted behavior of a driver like using cell phone, talking to others, eating, sleeping or lack of concentration during driving, deep learning models can generate warnings for the distracted driver periodically. Fig. 2 depicts the schematic diagram of the current study.

We utilize various Convolutional Neural Network (CNN) architectures such as VGG-16, ResNet50, and MobleNetV2 and analyze the results using appropriate classification evaluation metrics such as average training loss, validation accuracy, training accuracy for various test experiments. We have found that the proposal can effectively detect various abnormal and distracted behavior such as drinking, talking and texting, etc. of a distracted driver. In this study, we focus on developing a CNN-based approach for detecting distracted drivers and determining the source of distraction. Thus, the outcome of this work may reduce the number of accidents accordingly if we can build an automated system in real time. The contribution of this paper is as follows:

- Propose a convolutional Neural network-based (CNN) based driver distraction detection model that uses various architectures like Resnet50, MobileNetV2 along with transfer learning.
- For performance comparison, several experiments are conducted using a different number of images collected from a publicly available dataset. This dataset comprises thousands of images with ten different gestures of a distracted driver including sleeping, talking to others, using cell phone etc.
- Finally, to verify the effectiveness, appropriate validation metrics such as average training time, accuracy, validation and training accuracy are considered and the results show that ResNet50 and MobileNetV2 architecture outperform the others models.

The remainder of the paper is arranged as follows. Section 2 provides a brief review of previous studies regarding the identification of distracted driver behavior using machine learning algorithms. In Section 3 and 4, we describe the proposed methodology along with the CNN architectures adopted in the proposal. Section 5 presents the findings of the experiments as well as a discussion. Finally, section 6 concludes this study with future works.

2. Literature survey

Machine learning techniques are now being used in several applications. For example, cyberbullying detection (Islam et al., 2020), Human biometric recognition, medical record diagnosis (Ahamed et al., 2021; Ahmed et al., 2021), and human activity are some of the applications mentioned here. This section describes some related literature related to our work that is relevant to machine learning and deep learning approaches.

2.1. Machine learning approaches

Nowadays, machine learning algorithm to be trained using real-time images to detect distracted drivers is becoming more popular. For instance, Feng and Yue (2019) built a more inclusive distracted driving dataset using ten categorized images. They used a variety of machine learning approaches, including linear SVM, softmax, naive bayes, decision tree, and neural network, where the SVM classifier had the highest classification accuracy of 72.80%. Fernández, Usamentiaga, Carús and Casado (2016) explored the use of an SVM-based detector to determine if a driver is using a cellphone while driving. Images of drivers' faces were used to conduct the investigation. The hidden Markov model (HMM) is designed to detect driver behavior (Iversen, Møller, Morales & Madsen, 2016). AdaBoost and Hidden Markov models are explored by Craye and Karray (2015) to detect distracted behaviors of a driver but this approach is based on data generated in an indoor environment.

Using machine learning and fuzzy set theory, authors proposed a novel approach for testing driver distraction while performing a secondary task like chatting on a mobile phone, which is investigated as the secondary activity (Aksjonov, Nedoma, Vodovozov, Petlenkov & Herrmann, 2017, 2018a). There are several different methods, such as artificial neural networks or multi-layer perceptron classifiers, that have been studied to model the behavior of any distracted driver (Bejani & Ghatee, 2018; Jabbar et al., 2018). For example, the driver models using machine learning algorithms such as ANN and ANFIS are presented (Aksjonov, Nedoma, Vodovozov, Petlenkov & Herrmann, 2018b).

Nowadays, researchers can easily acquire eye and vehicle dynamics data, enabling them to investigate relations between these data and a driver's stress (Boril, Omid Sadjadi, Kleinschmidt & Hansen, 2010; Lanat'a et al., 2014). Traditional machine learning approaches were mainly utilized in the literature to manually extract features from data and then integrate the features to construct stress detection models (Nakisa, Rastgoo, Tjondronegoro & Chandran, 2018). Even though the handmade features technique produces outstanding results, appropriately extracting meaningful and representative qualities is always a challenge. Furthermore, the method of handcrafted features needs specialized knowledge and is more prone to noise and data volatility. As a result, several academics have developed multimodal data-based stress level detection algorithms. The authors developed a driving simulator that combines multimodal data, such as video data (facial activity) and physiological data, to assess stress levels (Benoit et al., 2009). To measure the driver's stress levels, the model used facial movements such as blinking, yawning, and head tilting, as well as ECG and electrical skin reactions. The authors presented a methodology to identify and predict driver stress levels and driving performance. Physiological signals, video recordings (eye data, head movement), and environmental data are multimodal data. With an accuracy of 86%, the model employed a support vector machine (SVM) classifier to discriminate between two stress levels (no stress and stress) (Rigas, Goletsis & Fotiadis, 2011).

Aksjonov, Nedoma, Vodovozov, Petlenkov and Herrmann (2018a) proposed a hybrid technique by combining a modified machine learning algorithm and fuzzy logic to detect and evaluate driver distraction while conducting secondary tasks. Their model includes a subsystem that measures the deviation of drivers from their normal activities. The driver distraction is calculated using the Fuzzy logic. Torres, Ohashi and Pessin (2019) offered a non-intrusive method for automatically distinguishing between drivers and passengers when reading a message in a vehicle. They collected data from smartphone sensors and fed the data into their machine learning model. In various settings, they simulated and assessed seven cutting-edge machine-learning techniques. The authors' model includes Convolutional Neural Network with Gradient Boosting.

2.2. Deep learning approaches

The researchers have recently shown that Deep learning approaches have outperformed conventional machine learning strategies for automatically detecting distracted postures. For example, deep learning is used to classify driver stress levels (Rastgoo, Nakisa, Maire, Rakotonirainy & Chandran, 2019) while tracing the drivers' movements and recognize their behavioral patterns (Galarza, Egas, Silva, Velasco & Galarza, 2018). The authors developed a CNN-based framework that recognizes and reveals the source of driver distraction. The efficiency of their enhanced VGG16 architecture is increased by 96.31% Baheti, Gajre and Talbar (2018). Yan et al. (2016) examined CNNs to detect distracted behavior using four different distraction postures.

Using a Deep Learning algorithm, Abouelnaga, Eraqi and Moustafa (2017) developed real-time distracted driver posture classification. Deep learning models such as AlexNet, InceptionV3, Majority Voting Ensemble, and GA-Weighted Ensemble were used to evaluate the proposal. For driver behavior prediction via sensory-fusion architecture, the authors used the Recurrent Neural Networks (RNNs) algorithms (Jain, Singh, Koppula, Soh & Saxena, 2016). They have tracked the driver's face and head position and combined the driver's features from both inside and outside the vehicle using GPS, road camera, and vehicle dynamics. Kim, Choi, Jang and Lim (2017) suggested a system for detecting driver distraction using RestNet50 and MobileNet CNN models. However, their analysis only looked at two types of distractions: looking in front and not looking in front postures. Mase et al. (2020) proposed a hybrid deep learning model that combines pretrained InceptionV3 and stacked Bidirectional LSTMs for detecting the distracted behavior with an accuracy of around 92.70%. Ou and Karray (2019) proposed a generative adversarial network-based driver distraction detection system (GANs). The author used the Internet to gather a diverse data set of drivers in various driving conditions and behavior patterns and train generative models for a variety of driving scenario.

Shahverdy, Fathy, Berangi and Sabokrou (2020) suggested a 2D Convolutional Neural Network (CNN) for detecting driver distractions and identified five types of driving styles using driving signals such as acceleration, gravity, throttle, speed, and Revolutions Per Minute (RPM), including normal, aggressive, distracted, drowsy, and drunk. They presented a three-phase strategy based on a lightweight 1D-CNN to detect driving behavior from vehicle signals (Shahverdy, Fathy, Berangi & Sabokrou, 2021).

Majdi, Ram, Gill and Rodríguez (2018) recommended CNNs to detect driver distraction postures and applied the U-Net CNN architecture for collecting context around the objects. The Distracted Driver dataset from the American University in Cairo (AUC) was used to train their model. When compared to Support Vector Classifiers and other CNN designs, their findings indicate a significant improvement concerning accuracy level. Similarly, Eraqi et al. (2019) suggested for utilizing four distinct CNN architectures to create a weighted ensemble of CNNs. Next, the CNNs are trained using the AUC distracted driver dataset's five separate

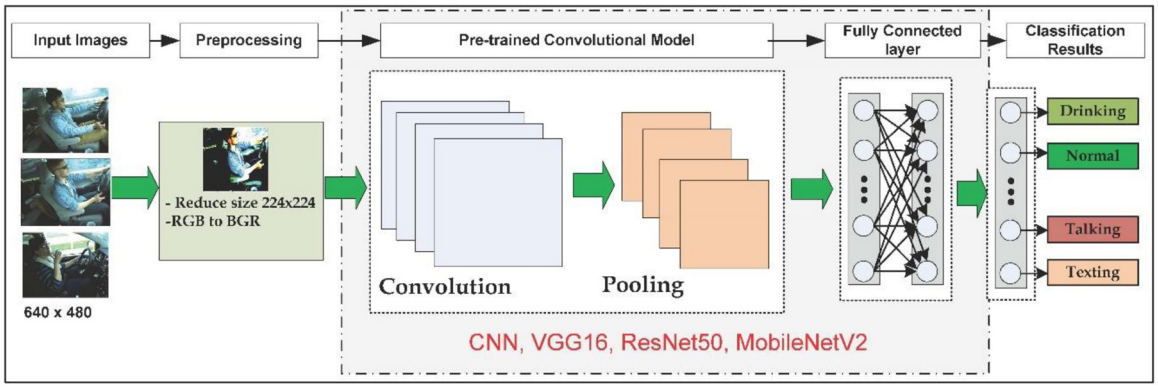


Fig. 3. Work-flow Diagram of the Proposal.

picture sources: raw images, skin-segmented images, face images, hands images, and face and hands images. According to the results, individual CNNs exhibited the best accuracy when trained on raw images. Following that, the predictions from the various CNNs are merged using a weighted Genetic Algorithm (GA), and the findings demonstrate that the fusion is more accurate than standalone CNNs and majority voting fusion.

3. Methodology

The process of detecting abnormal behavior of a distracted driver is presented in Fig. 3. In the proposal, firstly, the image is preprocessed for training the model with different postures of the driver including, eating, cell-phone texting, talking to others. Secondly, the pre-trained Convolutional model consists of different CNN-based deep learning architecture including ResNet50, MobileNetV2 has been adopted to identify the distracted behavior of the driver. Finally, the model is evaluated using the testing data images and we analyze the results of the proposal.

The steps that are carried out in the proposal are explained in this section:

1. Input Dataset: To train any learning model, we need to input images that represent the status of any distracted driver like gossiping, sleeping, or eating during the driving. This raw data can be collected using on-board computer system or external camera. We use an ideal dataset that contains thousands of images for our implementation.
2. Preprocessing: Preprocessing the images before fitting them into the learning module is essential to obtain accurate results. For example, we have to rotate and resize the system's input images.
3. Convolutional Neural Network: The CNNs are trained on raw images and can automatically extract high-level features from raw input features that are used for the detection, classification, and segmentation. Different architectures like ResNet50, VGG16, and MobileNetV2 are incorporated in our model. These architectures apply a filter to an input image to generate a feature map that summarizes the presence of detected features. The pooling layer is responsible for reducing the size of the extracted vectors which minimize the required computational energy to process the data by dimensionality reduction. In addition, it can identify and extract the dominant feature that are necessary for effective classification or detection process. Finally, the fully connected layers are similar to the traditional multi-layer perceptron neural network that connects every neuron layer-wise. Here, backpropagation is applied to every iteration of training make the system learn.
4. Classification: In this phase, the CNN identify each of the distracted behavior based on the learned experience of the previous phase.

4. Pre-trained model transfer learning

Transfer learning which is one of the studies in deep learning is implemented in multi-tasking and idea drift. Transfer learning is popular in deep learning, given the massive resources necessary to train deep learning models or the vast and hard datasets on which deep learning models are taught. In deep learning, transfer learning only works if the model features learned in the first task are general. In transfer learning, we train a base network on a base dataset and task, then re-purpose or transfer the acquired features to a second target network to be trained on a target dataset and task in transfer learning. This approach is more likely to succeed if the characteristics are generic, i.e., applicable to both the base and target tasks, rather than being particular to the base job. Transfer learning can be used in someone's predictive modeling problems. Two mainly used transfer learning approaches are the develop model approach and pre-trained model approach.

For the classification of image data, a large number of high performing models have been developed and deployed on ILSVRC (ImageNet Large Scale Visual Recognition Challenge). ImageNet, which is the source of the picture used in the competition, has led to a number of advancements in convolutional neural network construction and training. Furthermore, many of the models that were used in the competitions were published under a permissive license. These models can be utilized in computer vision applications as the foundation for transfer learning. The learning features are helpful and the models learn to detect the generic features from the images. The models achieve state-of-the-art performance for specific image identification tasks and remain effective for the tasks for which the models were actually developed. Fig. 4 shows the basic of transfer learning.

Pre-trained model weights are very easily accessible because of having freely downloadable features. The model weights and expedient APIs are downloaded and used in the same model architecture using several deep learning libraries, including Keras.

To derive the precise and succinct features set from the training data, transfer learning can be used and we adopted four CNN-based architectures for effective deep features extraction: simple CNN, VGG-16, Resnet50, and MobileNetV2. Since these architectures were pre-trained on a large image dataset, they managed to learn a strong representation of low-level features such as edges, rotation, lighting, and patterns, which can be used to extract deep features from new distracting images. Thus, these models can be useful for extracting specific features using transfer learning concepts. Below, we describe each of the model precisely.

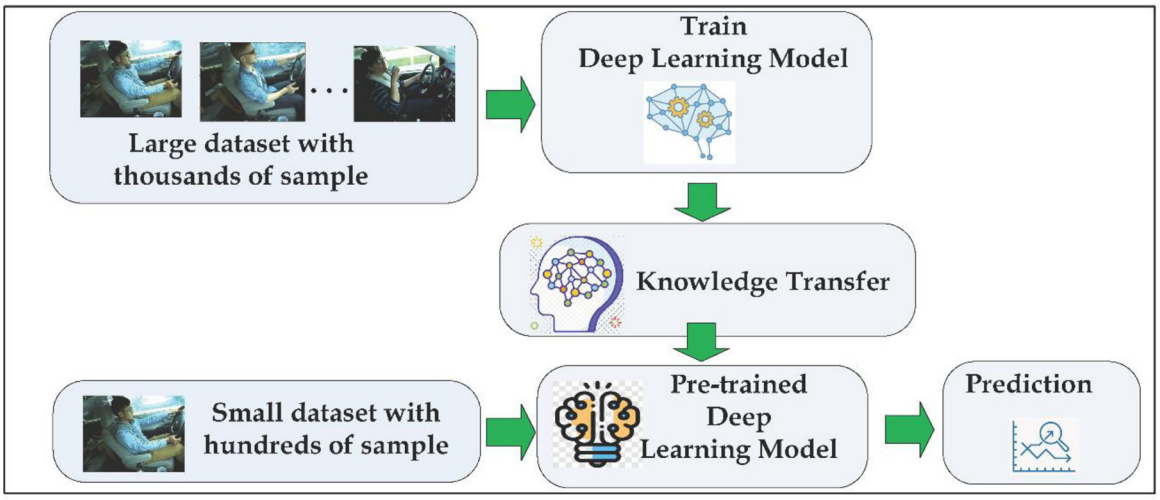


Fig. 4. Transfer Learning.

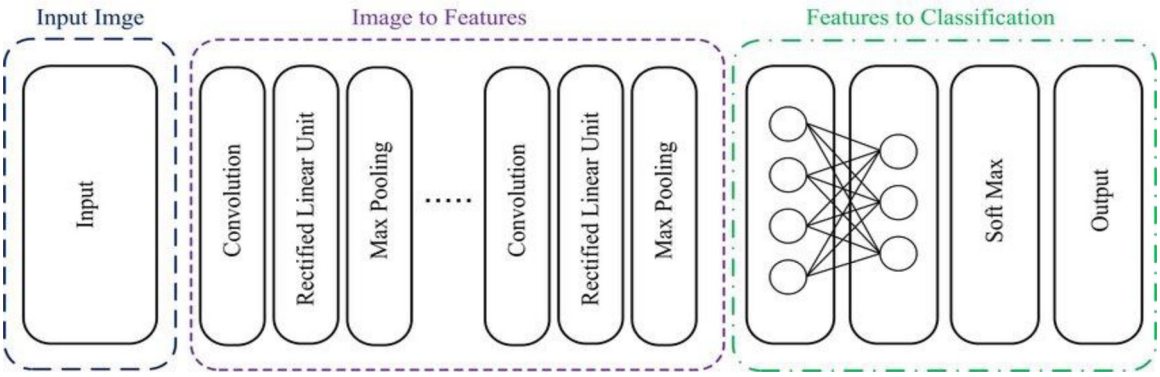


Fig. 5. Architecture of Convolutional Neural Network.

4.1. Convolutional neural network

A convolutional neural network (CNN) is a deep neural network used to classify images and detect object. A CNN has three layers: a convolutional layer, a pooling layer, and one or more fully connected layers as shown in Fig. 5.

4.1.1. Convolutional layer

ConvNet is motivated by the arrangement of the Visual Cortex and is akin to the connectivity pattern of Neurons in the Human Brain. This layer extracts features from the input image and generates a feature map that characterizes the presence of those features. Convolution retains the spatial relationship between pixels by learning the deep features of tiny small squares of an image (Yamashita, Nishio, Do & Togashi, 2018).

4.1.2. Pooling layer

These layers reduce the spatial dimensions of a large input image. Spatial pooling, also known as down sampling, is a technique for reducing the dimensionality of each map while preserving essential data. Max pooling, which produces the maximum value in the input features and average, is one of the most popular pooling techniques.

4.1.3. Fully connected layer

This layer is similar to a multi-layer neural network in terms of functionality. This layer predicts how closely each value corresponds to each class in a classification task after the features have been extracted by the convolution layers and down sampled by the pooling layers.

4.2. VGG-16

The VGG-16, as shown in Fig. 6, is one of the most well-known CNN architectures, with 16 convolutional layers and a highly standardized architecture. VGG-16 is a simpler architecture model that uses fewer hyperparameters and is a popular choice for extracting features from images using transfer learning (Dash, 2019). The convolution layer, for example, employs 3×3 filters with a stride of one and 2×2 max-pooling with a stride of two.

4.3. ResNet50

ResNet50 is a 50-layer deep CNN with 48 Convolution layers, one MaxPool layer, and one Average Pool layer. ResNet links the nth layer input directly to the $(n+x)$th layer, allowing for the stacking of additional layers and can be used for image recognition shown in Fig. 7. In comparison to standard classification architectures such as VGG-16, ResNet-50 has demonstrated faster performance and lower computational costs (S Jahromi et al., 2019).

4.4. MobileNetV2

The MobileNet-v2 is a deep convolutional neural network with 53 layers. It's a powerful feature extractor that can detect and segment objects. It is built on an inverted residual structure, with

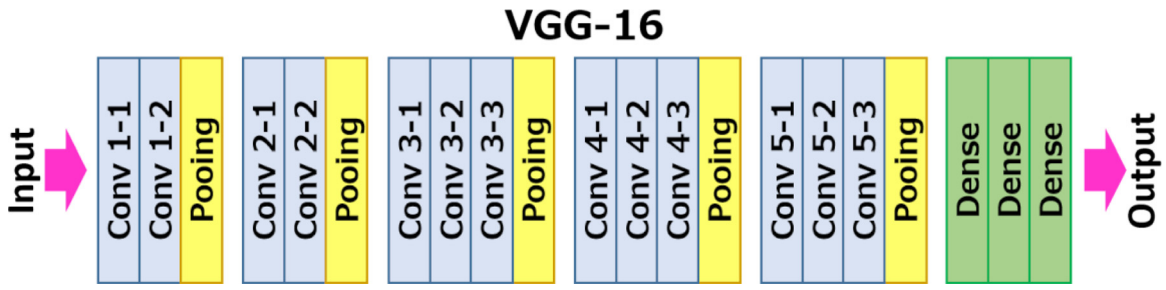


Fig. 6. Architecture of VGG16.

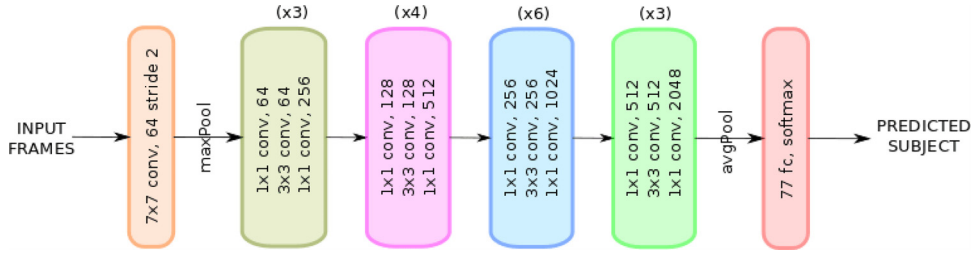


Fig. 7. Architecture of ResNet50.

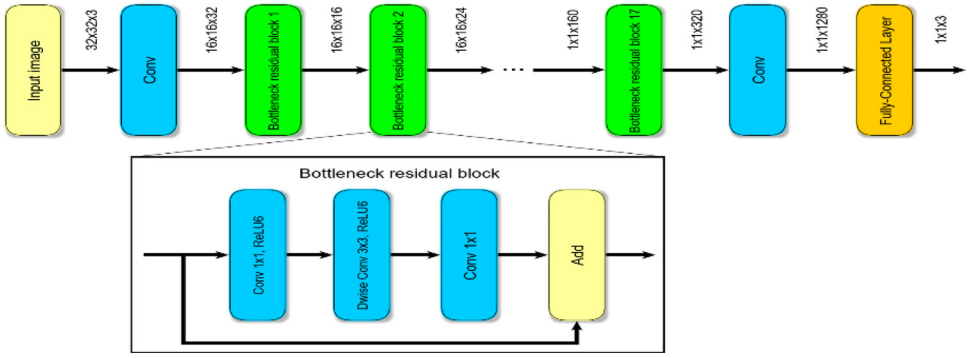


Fig. 8. Architecture of MobileNetV2.

residual connections between bottleneck layers shown in Fig. 8 (Seidaliyeva, Akhmetov, Ilipbayeva & Matson, 2020).

4.5. Architectural comparison

VGG-16, ResNet50 and MobileNetV2 are three basic transfer learning architecture of CNN. The performance accuracy of these models depends on various factors including the nature of dataset, the number of layers, the number of epochs, and batch size during the simulation. Table 1 shows the architectural comparison of the pre-trained transfer learning networks. Table 1 demonstrates that MobileNetV2 has a lower number of weights and the model is faster than the other models. Our experimental results also shows that MobileNetV2 provides a higher level of accuracy.

5. Experimental results analysis

The proposed approach is evaluated and validated on an Intel Core i7 processor with a 4 GB graphics card, a 64-bit windows operating system. Firstly, we describe the dataset for the experiments. Then, we evaluate our model using several experiments and show the results in detail.

5.1. Dataset

We have collected a large number of images from the State Farm Distracted Driver Detection competition hosted by State Farm (Kaggle, 2016). This dataset contains a large number of images that comprise ten different class postures of the distracted driver as shown in Table 2 and some example of driver images are depicted in Fig 9.

5.2. Experimental setup

For our experiments, we take 4000 and 6000 images from the dataset for two different experiments where each of the set contains almost equal number of images of each 10 classes. We choose 80% images for training and 20% for validation or testing for each experiment.

5.3. Data preprocessing

Before applying the model, the images are needed to preprocess such as rotation, resizing, and so on. Firstly, the input images are resized from 640×480 to 224×224. Next, the image is converted from RGB to BGR by subtracting the mean value (103.939, 116.779, 123.68) depicted in Fig. 10. Each pixel value was subtracted from the RGB mean value over all pixels. Mean subtraction is used since training our algorithm requires multiplying weights and applying biases in order to trigger the activation during the back-propagation.

Table 1
Architectural Comparison of VGG-16, ResNet50 and MobileNetV2.

SL	Properties	VGG-16	ResNet50	MobileNetV2
1.	image	224×224×3	224×224×3	224×224×3
2.	weight	Imagenet	Imagenet	imagenet
3.	size	528	98	14
4.	Total layers	16	50	53
5.	Convolution layer	13	48	53
6.	Max pool	5	1	1
7.	Activation function	Softmax	Softmax	Softmax
8.	Total parameters	138.3 million	25.6million	3.5 million
Advantages/Limitations				
VGG16	• It is painfully slow to train. • Weights are quite large.			
ResNet50	• Lower complexity than VGG-16. • Deeper than VGG-16.			
MobileNetV2	• Faster than others. • Low-power models parameterized to meet the resource constraints. • Preferable in vision-based mobile and embedded applications.			

Fig. 9. Example Images of Drivers.

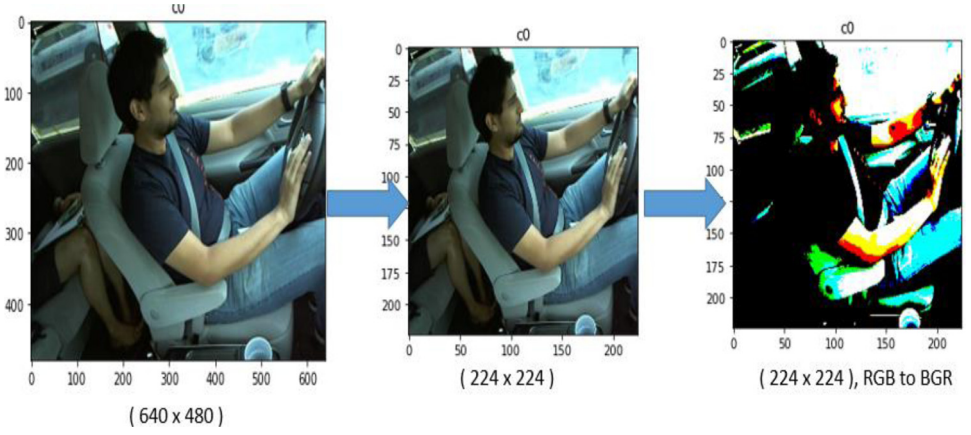


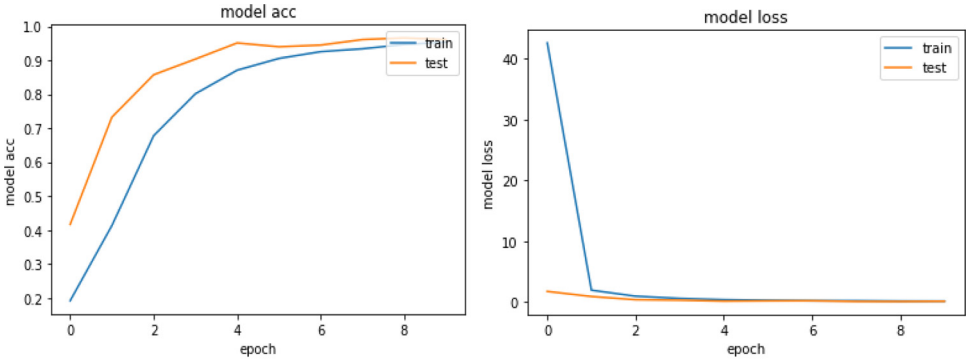
Fig. 10. Image Preprocess Pipeline.

5.4. Results of experiment −1

Fig. 11, 12, 13 and 14 depicted the training and testing accuracy and loss respectively of this experiment. Table 3 summarizes the performances of different architectures. From the table, it is clear that MobileNetV2 provides highest accuracy around 99.68% while exhibits the lower training loss among them.

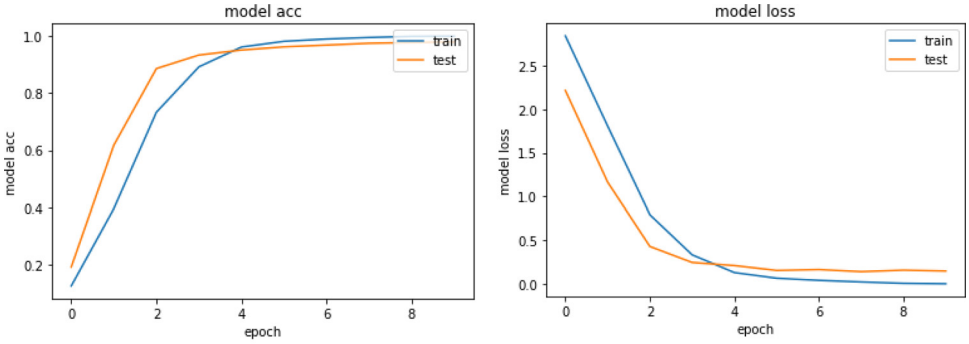
5.5. Results of experiment −2

In this experiment, we evaluate our model using 6000 images from the dataset. Fig. 15, 16, 17 and 18 depicted the model accuracy and loss respectively. Table 4 analyzes the average accuracy of the models. The table shows that MobileNetV2 provides highest training and testing accuracy around 99.98% and 98.12% respectively.



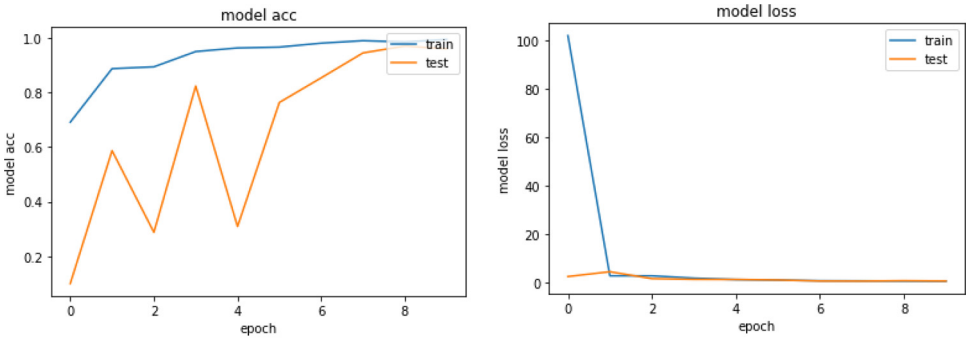
(a) Accuracy (b) Loss

Fig. 11. Performance Results of Simple CNN.



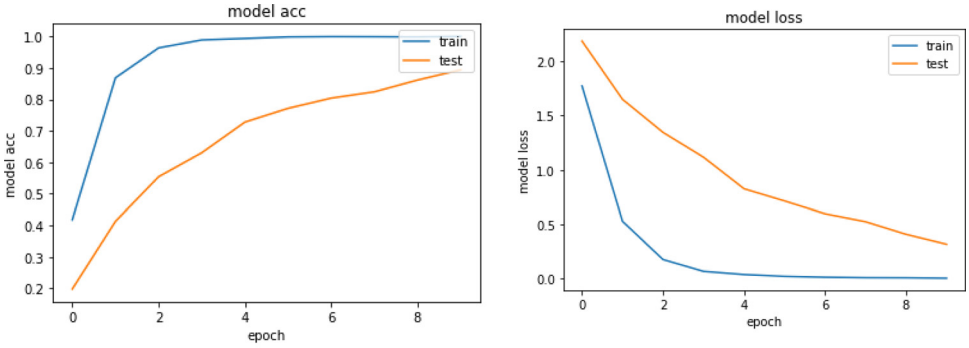
(a) Accuracy (b) Loss

Fig. 12. Performance Results of VGG-16.



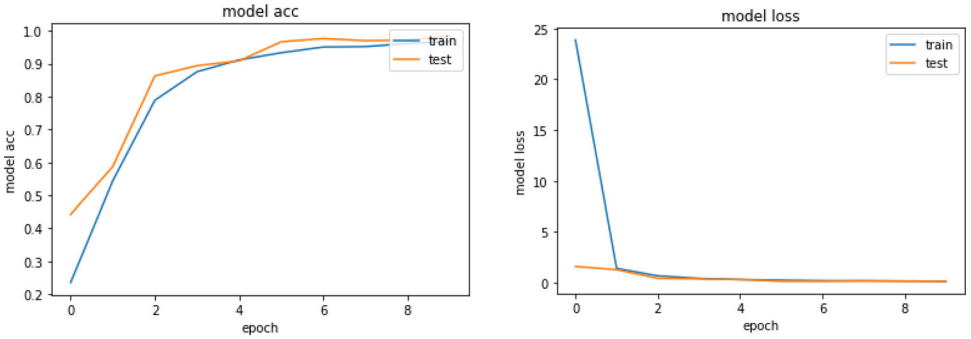
(a) Accuracy (b) Loss

Fig. 13. Performance Results of ResNet50.



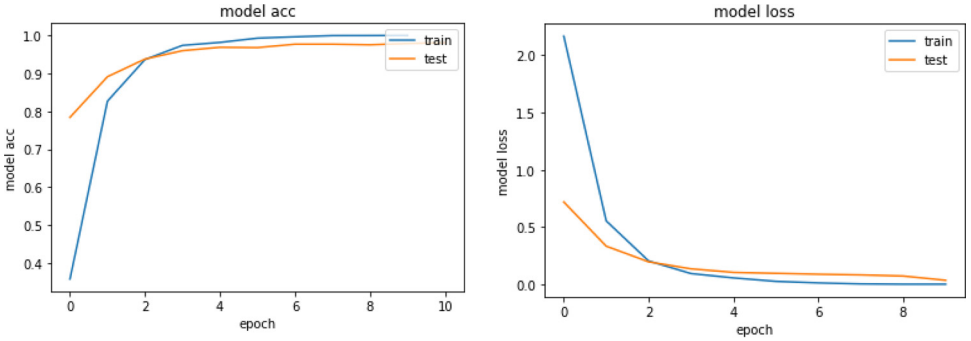
(a) Accuracy (b) Loss

Fig. 14. Performance Results of MobileNetV2.



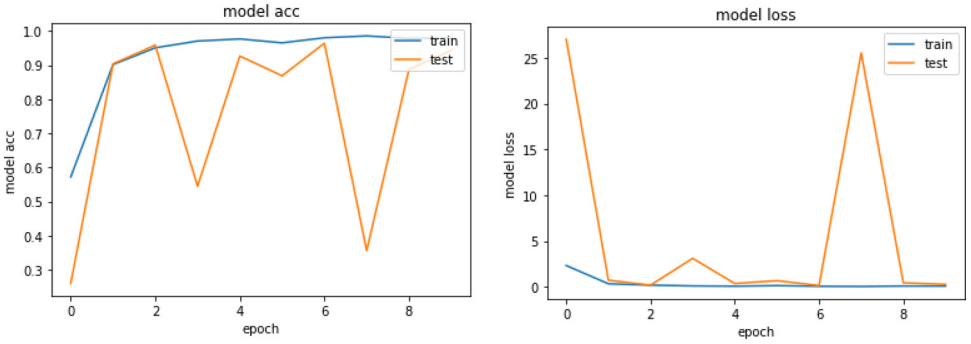
(a) Accuracy (b) Loss

Fig. 15. Performance Results of Simple CNN.



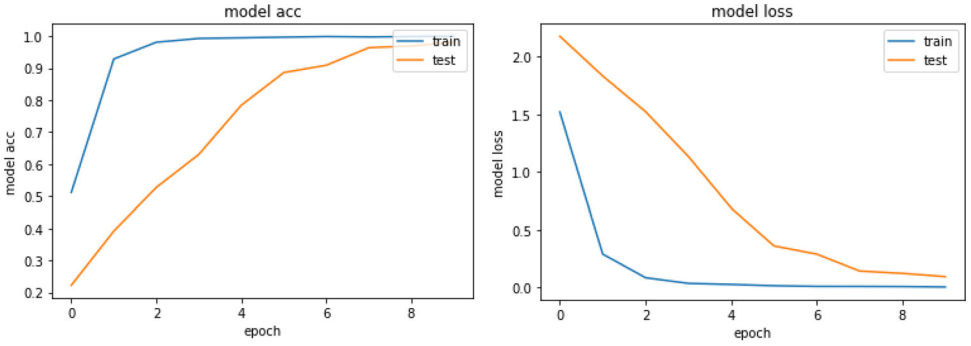
(a) Accuracy (b) Loss

Fig. 16. Performance Results of VGG-16.



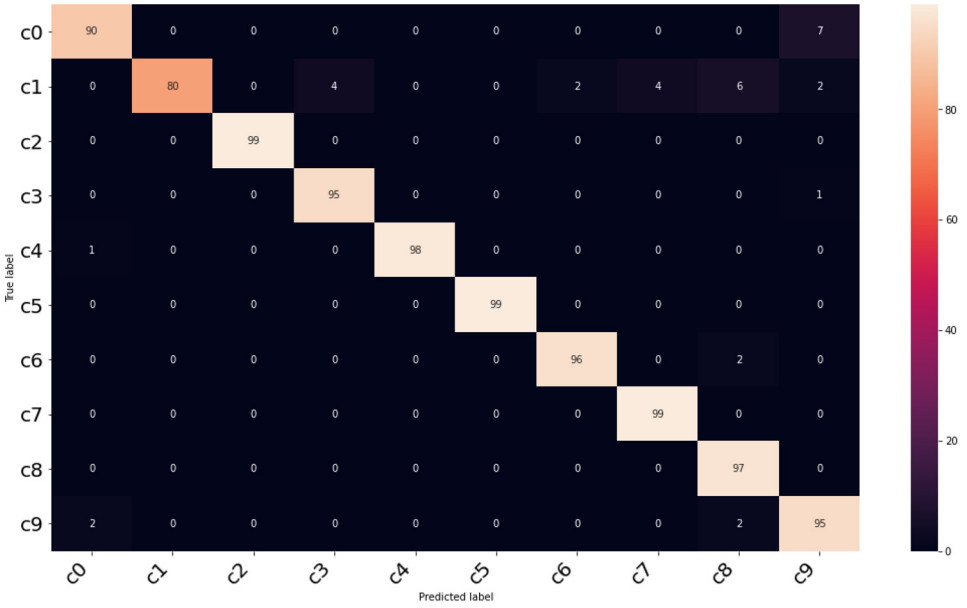
(a) Accuracy (b) Loss

Fig. 17. Performance Results of ResNet50.

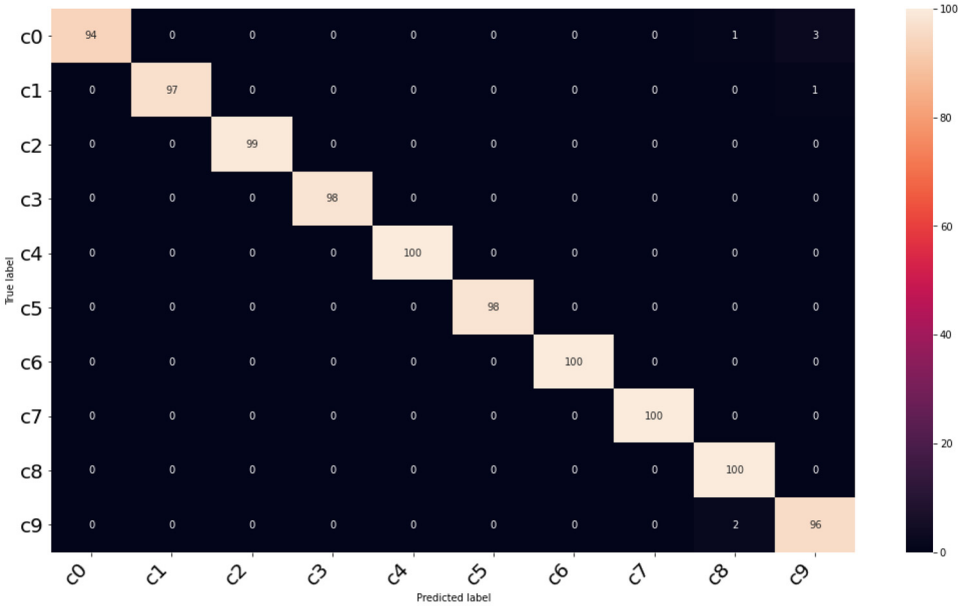


(a) Accuracy (b) Loss

Fig. 18. Performance Results of MobileNetV2.



(a) Confusion Matrix of ResNet50



(b) Confusion Matrix of ResNet50

Fig. 19. Class-wise Confusion Matrix.

Table 5 and 6 represent the class-wise average accuracy. Finally, Fig. 19 represents the confusion matrix of the experiment for ResNet50 and MobileNetV2.

5.6. Suitableness for embedded application

MobileNets are lightweight deep neural networks ideal for mobile and embedded vision applications. MobileNets are low-latency, low-power models that have been parameterized to satisfy specific use cases' resource constraints. As a results, this model outperforms ResNet50 and VGG-16 in terms of speed. In addition, model generates a light weight that is 32 times less than VGG-16. Furthermore, in real-world applications such as autonomous vehicles or robotic sights, object detection can be accomplished on a computationally constrained platform. MobileNetV2, a network for embedded visual applications and mobile devices, was developed to address this problem. As a result, MobileNetV2 is preferable for mobile and embedded vision applications with minimal resources.

Table 2
Distracted behavior class distribution.

Class	Bahavior
C-0	Safe driving
C-1	Texting on right hand
C-2	Cell-phone talking on right-hand
C-3	Texting on left Hand
C-4	Cell-phone talking on left-hand
C-5	Operating the radio
C-6	Drinking
C-7	Reaching Behind
C-8	Hair and makeup
C-9	Talking to passengers

Table 3
Accuracy Results of Experiment-1.

Model Name	Training Loss	Training Accuracy	Testing Loss	Testing Accuracy
Simple CNN	0.1770	95.38%	0.1521	96.06%
VGG-16	0.0021	98.27%	0.4488	92.45%
ResNet50	0.0313	99.31%	0.2606	95.10%
MobileNetV2	0.0035	99.68%	0.3177	89.38%

Table 4
Accuracy Results of Experiment 2.

Model Name	Training Loss	Training Accuracy	Testing Loss	Testing Accuracy
Simple CNN	0.1170	96.79%	0.1175	97.45%
VGG-16	0.0016	97.08%	0.0360	94.01%
ResNet50	0.0881	97.93%	0.2753	94.28%
MobileNetV2	0.0035	99.98%	0.0937	98.12%

Table 5
Class-wise Accuracy Results for ResNet50.

Class	Experiment-1	Experiment-2
C-0: Safe driving	98.06%	51.29%
C-1: Texting on right hand	95.50%	90.59%
C-2: Cell-phone talking on right-hand	100.00%	100.00%
C-3: Texting on left Hand	99.95%	85.56%
C-4: Cell-phone talking on left-hand	100.00%	98.15%
C-5: Operating the radio	100.00%	100.00%
C-6: Drinking	99.98%	100.00%
C-7: Reaching Behind	99.85%	100.00%
C-8: Hair and makeup	98.93%	99.28%
C-9: Talking to passengers	95.51%	90.59%

Table 6
Class-wise Accuracy Results for MobileNetV2.

Class	Experiment-1	Experiment-2
C-0: Safe driving	73.11%	99.41%
C-1: Texting on right hand	55.68%	99.82%
C-2: Cell-phone talking on right-hand	99.81%	100.00%
C-3: Texting on left Hand	99.55%	99.98%
C-4: Cell-phone talking on left-hand	98.27%	94.62%
C-5: Operating the radio	92.67%	100.00%
C-6: Drinking	62.91%	79.12%
C-7: Reaching Behind	39.48%	99.59%
C-8: Hair and makeup	98.51%	99.80%
C-9: Talking to passengers	84.73%	99.98%

6. Conclusion

Driver distraction is one of the major problems worldwide. We have designed a convolutional neural network-based architecture for detecting the driver distraction behaviors and the cause of distraction. Various architectures such as Simple CNN, VGG-16, ResNet50 and MobileNetV2 are adopted as the training models. Extensive experiments have been conducted with an ideal dataset to verify the model. We have found that the ResNet50 and MobileNetV2 provide higher accuracy of 94.50% and 98.12% respectively. The proposed model results can be used to design real-time driver distraction detection systems. Though the proposed model provides high accuracy, a real-time test-bed implementation can make the work more effective. In future we are planning to develop an android application to detect distracted driver in real time.

Author statements

The role(s)/contributions of all authors are enlisted using the relevant categories as follows:

Md.Uzzol Hossain, Md. Ataur Rahman Data collection, methodology and simulation

Md. Manowarul Islam Conceptualization, Supervision, methodology, Editing

Md. Ashraf Uddin Conceptualization, Writing- Reviewing

Arnisha Akter Writing- validation

Bikash Kumar Paul Visualization, Investigation, Editing

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abouelnaga, Y., Eraqi, H.M., & Moustafa, M.N. (2017). *Real-time distracted driver posture classification*.

Agrawal, U., Mase, J.M., Figueredo, G.P., Wagner, C., Mesgarpour, M., & John, R.I. (2019). Towards real-time heavy goods vehicle driving behaviour classification in the United Kingdom. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)* (pp. 2330–2336).

Ahamed, K.U., Islam, M., Uddin, A., Akhter, A., Paul, B.K., Yousuf, M.A., et al. (2021). A deep learning approach using effective preprocessing techniques to detect covid-19 from chest CT-scan and X-ray images. *Computers in Biology and Medicine*, Article 105014.

Ahmed, N., Ahammed, R., Islam, M.M., Uddin, M.A., Akhter, A., Talukder, M.A.-A., et al. (2021). Machine learning based diabetes prediction and development of smart web application. *International Journal of Cognitive Computing in Engineering, 2*, 229–241.

Aksjonov, A., Nedoma, P., Vodovozov, V., Petlenkov, E., & Herrmann, M. (2017). A method of driver distraction evaluation using fuzzy logic: Phone usage as a driver's secondary activity: Case study. In *2017 XXVI International Conference on Information, Communication and Automation Technologies (ICAT)* (pp. 1–6).

Aksjonov, A., Nedoma, P., Vodovozov, V., Petlenkov, E., & Herrmann, M. (2018a). Detection and evaluation of driver distraction using machine learning and fuzzy logic. *IEEE Transactions on Intelligent Transportation Systems, 20*, 2048–2059.

Aksjonov, A., Nedoma, P., Vodovozov, V., Petlenkov, E., & Herrmann, M. (2018b). A novel driver performance model based on machine learning. *IFAC-PapersOnLine, 51*, 267–272.

Alvarez, A.D., Garcia, F.S., Naranjo, J.E., Anaya, J.J., & Jimenez, F. (2014). Modeling the driving behavior of electric vehicles using smartphones and neural networks. *IEEE Intelligent Transportation Systems Magazine, 6*, 44–53.

Baheti, B., Gajre, S., & Talbar, S. (2018). Detection of distracted driver using convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 1032–1038).

Bejani, M.M., & Ghatee, M. (2018). A context aware system for driving style evaluation by an ensemble learning on smartphone sensors data. *Transportation Research Part C: Emerging Technologies, 89*, 303–320.

Benoit, A., Bonnaud, L., Caplier, A., Ngo, P., Lawson, L., Trevisan, D.G., et al. (2009). Multimodal focus attention and stress detection and feedback in an augmented driver simulator. *Personal and Ubiquitous Computing, 13*, 33–41.

Boril, H., Omid Sadjadi, S., Kleinschmidt, T., & Hansen, J. (2010). Analysis and detection of cognitive load and frustration in drivers' speech. In *Proceedings of the 11th annual conference of the international speech communication association* (pp. 502–505).

Craye, C., & Karray, F. (2015). *Driver distraction detection and recognition using rgb-d sensor*.

Dash, A.K. (.2019). Vgg16 architecture. https://iq.opengenus.org/vgg16/ Accessed: 2021-03-20.

Eraqi, H.M., Abouelnaga, Y., Saad, M.H., & Moustafa, M.N. (2019). Driver distraction identification with an ensemble of convolutional neural networks. *Journal of Advanced Transportation*, 1–12.

Feng, D., & Yue, Y. (2019). Machine learning techniques for distracted driver detection.

Fernández, A., Usamentiaga, R., Carús, J.L., & Casado, R. (2016). Driver distraction using visual-based sensors and algorithms. *Sensors, 16*, 1805.

Galarza, E.E., Egas, F.D., Silva, F.M., Velasco, P.M., & Galarza, E.D. (2018). Real time driver drowsiness detection based on driver's face image behavior using a system of human computer interaction implemented in a smartphone. In *International conference on information technology & systems* (pp. 563–572).

Islam, M.M., Uddin, M.A., Islam, L., Akter, A., Sharmin, S., & Acharjee, U.K. (2020). Cyberbullying detection on social networks using machine learning approaches. In 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering *(CSDE)* (pp. 1–6).

Iversen, E.B., Møller, J.K., Morales, J.M., & Madsen, H. (2016). Inhomogeneous Markov models for describing driving patterns. *IEEE Transactions on Smart Grid, 8*, 581–588.

Jabbar, R., Al-Khalifa, K., Kharbeche, M., Alhajyaseen, W., Jafari, M., & Jiang, S. (2018). Real-time driver drowsiness detection for android application using deep neural networks techniques. *Procedia Computer Science, 130*, 400–407.

Jain, A., Singh, A., Koppula, H.S., Soh, S., & Saxena, A. (2016). Recurrent neural networks for driver activity anticipation via sensory-fusion architecture. In *2016 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 3118–3125).

Kaggle (2016). State farm distracted driver detection. URL: Https://www.kaggle.com/c/state-farm-distracted-driver-detection Accessed: 2021-03-20.

Kamruzzaman, M. (2020). Bangladesh: Alarming rise in road crashes. URL: Https://www.aa.com.tr/en/asia-pacific/bangladesh-alarming-rise-in-road-crashes/1692643?fbclid=IwAR2N9Cr35HVbWl3YxUUPSg9TTTWCYQUFjvK5fc3AqEKn9qs7AlBohYKTMXc Accessed: 2021-03-20.

Kim, W., Choi, H.-.K., Jang, B.-.T., & Lim, J. (2017). Driver distraction detection using single convolutional neural network. In 2017 international conference on information and communication technology convergence *(ICTC)* (pp. 1203–1205).

Kouchak, S.M., & Gaffar, A. (2019). Using bidirectional long short-term memory with attention layer to estimate driver behavior. In 2019 18th IEEE International Conference On Machine Learning And Applications *(ICMLA)* (pp. 315–320).

Lanat'a, A., Valenza, G., Greco, A., Gentili, C., Bartolozzi, R., Bucchi, F., et al. (2014). How the autonomic nervous system and driving style change with incremental stressing conditions during simulated driving. *IEEE Transactions on Intelligent Transportation Systems, 16*, 1505–1517.

Majdi, M.S., Ram, S., Gill, J.T., & Rodríguez, J.J. (2018). Drive-net: Convolutional network for driver distraction detection. In 2018 IEEE Southwest Symposium on Image Analysis and Interpretation *(SSIAI)* (pp. 1–4).

Mase, J.M., Agrawal, U., Pekaslan, D., Torres, M.T., Figueredo, G., Chapman, P., et al. (2020). Capturing uncertainty in heavy goods vehicle driving behaviour. In *IEEE international conference on intelligent transportation systems*.

Morton, J., Wheeler, T.A., & Kochenderfer, M.J. (2016). Analysis of recurrent neural networks for probabilistic modeling of driver behavior. *IEEE Transactions on Intelligent Transportation Systems, 18*, 1289–1298.

Nakisa, B., Rastgoo, M.N., Tjondronegoro, D., & Chandran, V. (2018). Evolutionary computation algorithms for feature selection of eeg-based emotion recognition using mobile sensors. *Expert Systems with Applications, 93*, 143–155.

Ou, C., & Karray, F. (2019). Enhancing driver distraction recognition using generative adversarial networks. *IEEE Transactions on Intelligent Vehicles, 5*, 385–396.

Rastgoo, M.N., Nakisa, B., Maire, F., Rakotonirainy, A., & Chandran, V. (2019). Automatic driver stress level classification using multimodal deep learning. *Expert Systems with Applications, 138*, Article 112793.

Rigas, G., Goletsis, Y., & Fotiadis, D.I. (2011). Real-time driver's stress event detection. *IEEE Transactions on intelligent transportation systems, 13*, 221–234.

S Jahromi, M.N., Buch-Cardona, P., Avots, E., Nasrollahi, K., Escalera, S., Moeslund, T.B., et al. (2019). Privacy-constrained biometric system for non-cooperative users. *Entropy, 21*, 1033.

Seidaliyeva, U., Akhmetov, D., Ilipbayeva, L., & Matson, E.T. (2020). Real-time and accurate drone detection in a video with a static background. *Sensors, 20*, 3856.

Shahverdy, M., Fathy, M., Berangi, R., & Sabokrou, M. (2020). Driver behavior detection and classification using deep convolutional neural networks. *Expert Systems with Applications, 149*, Article 113240.

Shahverdy, M., Fathy, M., Berangi, R., & Sabokrou, M. (2021). Driver behaviour detection using 1d convolutional neural networks. *Electronics Letters, 57*, 119–122.

The Daily Star. (2020). 21 died on roads every day. URL: Https://www.thedailystar.net/backpage/road-accident-in-bangladeh-21-died-every-day-1852867, Accessed: 2021-03-20.

Torres, R., Ohashi, O., & Pessin, G. (2019). A machine-learning approach to distinguish passengers and drivers reading while driving. *Sensors, 19*, 3174.

World Health Organization (2020). Road traffic injuries. URL: Https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries Accessed: 2021-03-20.

Yamashita, R., Nishio, M., Do, R.K.G., & Togashi, K. (2018). Convolutional neural networks: An overview and application in radiology. *Insights into imaging, 9*, 611–629.

Yan, C., Coenen, F., & Zhang, B. (2016). Driving posture recognition by convolutional neural networks. *IET Computer Vision, 10*, 103–114.