ISSN 2007-9737

Distraction Detection to Predict Vehicle Crashes: a Deep Learning Approach

Reda Bekka[1], Samia Kherbouche[1], Houda El Bouhissi[2]

[1] University of Bejaia,
Faculty of Exact Sciences,
Algeria

[2] University of Bejaia,
Faculty of Exact Sciences,
LIMED Laboratory,
Algeria

{reda.bekka, kherbouchesamia28}@gmail.com, houda.elbouhissi@univ-bejaia.dz

Abstract. The Road safety is a major issue, both in terms of the number of road casualties and the economic cost of these accidents at the global, regional and national levels. Combating road insecurity is a priority concern for every country, as travel continues to increase and, despite the measures taken in many countries to improve road safety, much remains to be done in order to reduce the number of deaths and fatalities. In this paper, we review the most applied approaches in the detection of driver distractions. Furthermore, we propose a novel approach for preventing road crashes in the context of intelligent transport. Preliminary results indicate that the proposed methodology is efficient and provides high accuracy.

Keywords. CNN, distraction, detection, deep learning, drowsiness, intelligent transport, OpenCV, transfer learning, VGG16.

1 Introduction and Motivation

Driver distraction is the diversion of attention away from activities for safe driving toward a competing activity and occurs, when a driver is delayed in the recognition of information needed to safely accomplish the driving task, because of some event, activity, object, or person within or outside the vehicle compels or induces the driver's shifting attention away from the driving task [7].

Traffic accidents are considered as the most serious social, safety and economic problems for the most nations of the world, both developing and developed countries. Road traffic accidents remain a global scourge. According to the World Health Organization (WHO) global status report, road traffic accidents cause 1.35 million deaths each year. This is nearly 3700 people dying on the world's roads every day [19].

Every year, thousands of people die in highway-related crashes and millions are injured. By 2030, the crashes are predicted to be the 5th leading cause of death in the world [4].

Algeria is the 98th in the world road accident ranking and 42nd on the African continent. The number of road accidents in Algeria is increasing and the human factor is the main cause of these accidents. In the first quarter of 2020, 918 people died and 11,919 were injured.

In recent years, many researchers studied the traffic accidents influencing factors, focusing on the environment, people, cars, roads. They stated that one of the most important factors in road accidents is driver fatigue, drowsiness and distraction, which reduce the driver's perceptions and decision-making ability to control the vehicle.

Recently, the Algerian government made laws and strategies for traffic reform concerning drivers,

pedestrians and vehicle to improve road traffic security and reduce road accidents.

According to the Algerian law, traffic violations result in the payment of penalties ranging from 2,000 to 4,000 Algerian dinars and the withdrawal of a driving license for one month or more.

Preventive measures to reduce traffic accidents involve increasing the number of police checkpoints and cameras to track drivers who exceed the limited speed or violate the traffic laws.

Most of the people travel for long distances without any sleep and using mobile phones while driving this results to the issue of tiredness and as a result to the drowsiness. This can be avoided just by alerting the driver when there is any such case of occurrence. So we are proposing a system which can alert the driver using a alarm when the driver gets distracted or feels drowsy.

Main detection distraction researches focused on facial recognition such as landmarks and face detection, which is not enough to detect whether the driver is using his phone or no, he is talking with the passengers or not or whether he is tired or not.

Distracted driving detection methods are mainly based on the driver's facial expression, head operation, line of sight or body operation [10].

In addition, current works focused only on fatigue detection. Detecting drivers' movements remains a difficult task. To the best of our knowledge, small effort has been spent to detect driver movements.

Our work is part of smart transportation systems and aims to propose a real-time detection of public transport drivers' movements in order to reduce road accidents.

The aim of this paper is to review the most important works related to driver distraction using smartphones to detect inattentive and distracted driving or related aspects and to propose a new approach to decrease the number of accidents and save lives.

Thus, the main contributions of our work can be summarized as follows:

— Detecting facial expression if the driver wears glasses or not.

— Improving the efficiency of distraction detection algorithms.

— Implementing a software tool which deals with both motion and facial expression detection.

The rest of this paper is structured as follows: Section 2 provides an overview of the basic concepts used in this paper. Then, in section 3, we review the most important approaches related to our proposal. In section 4, we present in detailed our approach. An empirical study of the proposed approach is presented in section 5, in order to assess its performance and efficiency. In Section 6, the results are presented along with discussion, and Section 7 concludes the paper and establishes the opportunity for future work.

2 Theoretical Background

This section covers the key techniques from the Artificial Intelligence (AI) domain used in the literature for accident prevention and the driver behavior.

2.1 Deep Learning

Deep Learning (DL) [16] is a sub set of the Machine Learning (ML) field which was used by authors to predict traffic flows. Recently, Deep learning have demonstrated impressive performance in automatically extracting image features for computer vision tasks and has gained more attention in distraction detection.

At present, DL is widely used for image recognition, natural language processing and automatic driving and achieved good performance. In particular, with the rapid development of Internet of Things, a large number of data are collected, providing rich data for the establishment of efficient DL model.In spite of the advances of DL approaches dominating those tasks, the effectiveness of them is mostly tested and evaluated in data sets of exceptional qualities [11].

2.2 Convolutional Neural Networks

Convolutional Neural Networks (CNN) are a DL algorithms that are currently paving new avenues in the field of image analysis and computer vision [9].

CNN are able to reduce the images into a form which is easier to process, without losing features which are critical for getting a good prediction.

Furthermore, CNN can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other.

The self-learning capabilities of CNN are a great advantage in terms of computational efficiency and automation since a feature design process is not required [2].

2.3 VGG16

VGG16 is a CNN model proposed by Simonyan and Zisserman [15] for image recognition in the 2014 ImageNet large scale visual recognition challenge (ILSVRC). The model achieves a test accuracy of 92.7% in the top 5 in ImageNet.

3×3 filters are used for all convolutional layers. The network accepts the input image with a dimension of 224 × 224. The image is passed through a sequence of 16 convolutional layers. A multi layer perceptron (MLP) classifier including three fully-connected (FC) layers is used in addition to the convolutional layers to perform the classification. Rectified Linear Unit (ReLU) layers and max-pooling layers are used in the whole network to prevent the over-fitting problem [13].

2.4 Transfer Learning

Transfer learning (TL) is a research problem in machine learning (ML) that focuses on storing knowledge gained while solving one problem and applying it to a different but related problem. The latter relies on neural networks because they are stacked in layers, each one learning from the previous one.

The transfer learning approach shows better overall performance than the approach without it. This is because transfer learning approach pulls out more information from the image [3].

This is how a CNN will have its first layers specialized in the recognition of simple shapes (horizontal lines, vertical lines, diagonals, ...), its next layers is dedicated to the recognition of slightly more complex shapes (circle, square, triangle, ...), followed by next layers oriented towards, for example the recognition of faces, the body parts recognition, etc. Moreover, the last layers will be dedicated to what is being learned in this network.

3 Related Work

Detecting distracted driving is receiving significant amount of attention from the research community and the industry. There is a wealth of data that has been made available to the research.

As the problem is an important point of discussion, there is already research done on the same topic in previous years. In the upcoming section, different studies will be discussed which have been already carried out. Based on the high accident rate caused by driver distraction, researchers have proposed various methods to detect distraction. We list here the most related methods.

A matching method for head rotation detection and a fuzzy expert system to estimate machine vision hypo vigilance for the driver's face monitoring is proposed by [14]. This approach could detect driver hypo vigilance (fatigue and distraction) through movement of the eye and face regions. The acquisition of face detection images is the first step in the process. Hypo vigilance symptoms are extracted from the facial image. However, an explicit eye detection step was not used to determine the eye of the face, but some of the important symptoms related to the eye region (the segment of the upper half of the face) are extracted. A template-matching method is used to detect head rotation. The method also used a fuzzy expert system to estimate driver hypo vigilance. This approach is effective for different individuals with different facial and eyelid behaviors in real time by reducing the risk of crashes.

Another method is presented in [5] that aims at using the visual attention mechanism to detect unusual motion for vision-based driver

assistance.The authors propose a real-time unusual movement detection model that involves two steps: detection of protruding areas and detection of unusual movements. They use the temporal attention model, the scale-invariant feature transform (SIFT) and the optical flow technique. Their approach used this unusual motion detection technique to detect the risk of collision for the image points under consideration and, most importantly, to effectively and efficiently detect areas of unusual motion. The model can estimate the unusualness for an output of warning messages to the driver to avoid vehicle collisions.

Researchers in [13], use DL algorithms, so they tested the following algorithms: VGG16, GoogleNet, Alexnet and Resnet which are all based on CNN. Driver assistance systems use this technique to detect driver distractions. In order to evaluate and validate the CNN models for distraction detection, these authors conducted experiments on the assisted driving simulator. They used images of drivers in normal and distracted driving postures as inputs. The algorithms are then implemented and evaluated on an integrated GPU platform.

An interesting approach is proposed by [12]. The authors use a supervised DL algorithm named as DriveNet to determine the distraction of drivers. This algorithm comprises of CNN network and Random forest. The DriveNet architecture was compared against two other machine learning techniques namely recurrent neural network (RNN) and multi-layered perception (MLP). The dataset used was picked from the Kaggle competition which was publicly available. According to the authors, the proposed model attained an accuracy of around 95% which was far better than the earlier results from the competition. Also, the DriveNet architecture proved to be a better model over the other two compared models.

Another study is proposed in [6] which uses hierarchical weighted randomized forest classifiers (WRF) for safe real-time driving. The proposed technique was used for driver facial expression recognition. WRF classifiers were used to obtain a more accurate classification. The first WRF classifier was used to learn to distinguish between fear, happiness and another group of expressions. Indeed, the three emotions anger, disgust and sadness have similar facial characteristics and can therefore be classified more accurately in the second level. In the second level, the second WRF classifies the anger, disgust and sadness of the group to obtain a more precise classification. The two types of WRF classifiers are learned separately using two vectors of characteristics, in terms of inputs. This approach used FER (Facial Emotion Recognition), CK (Cohn Kanade) and MMI (Man Ma- chine Interface) databases. The authors state that the results obtained on facial expressions from the confusion matrix were very satisfactory.

The authors in [18] proposed a CNN-based method for the recognition of drivers' facial expressions for the analysis of facial expression in a driver assistance system. In this context, a model based on ShufeNet has a total number of 871849 parameters and the output layer is a softmax log. The model was formed end-to-end using the NLLLoss (pytorch function) and the Adam optimizer. In another project on the face identification, [18] formed a network based on a MobileNetV2 backbone of size 128×128 and was significantly smaller with only 84871 parameters. He replaced the last layer and reworked the model using ArcFace loss and the Adam optimizer. The proposal provides promising results with the ShuffleNet daemon.

A driver distraction detection system is presented in [1]. The proposed methodology uses a combination of three of the most advanced deep learning techniques, the Residual Network (ResNet), Hierarchical Recurrent Neural Network (HRNN) and inception architecture (deep neural network architecture). In this model, a ResNet (Residential Energy Services Network) block and 2 HRNN layers were integrated in the inception module, followed by 2 dense layers and finally the softmax classifier. In order to evaluate the results of the proposed method, they used the state farm Distracted Driver Detection dataset as provided by Kaggle. They concluded that their model learns richer representations with a small number of parameters. They divide their data into test and training sets according to percentages, took 10%, 20%, 30% for training and the rest for

testing. The proposed method gave promising results, by combining several techniques.

Zhao et al. [20] use the method based on physical feature fusion and the method based on DL to detect driver fatigue. The proposal focuses on the method of fatigue detection based on CNN. The authors state that the experimental results in various situations provide the possibility for the realization of the driver's fatigue detection system. They perform fatigue classification based on the faces detected by the SSD network. The proposal results show that eyes and mouth are important features that play an important role in fatigue detection. According to the authors,the method of combining eyes and mouth achieves 95% accuracy. At the same time, the authors made the NTHU-DDD data set. The detection method based on the VGG16 network has 91.88% accuracy on this data set, which is about 5% higher than the original method. In addition, the authors state that the experiments prove that their method has better accuracy. The result of this work is that it can be used in a driving assistance system for high-precision, high-safety driver fatigue detection, and has a wide range of applications in a driving assistance system.

The authors in [8] also developed a conversational alert system that warns the driver in real time when he or she is not concentrating on the driving task. As a result, they create realistic driving experiences. The authors state that the experimental results show that the proposed approach is more effective than the basic approach. In addition, the results also indicate that GoogleNet is the best of the four models for distraction detection on the driving simulator test bed.

Finally, the authors in [17] proposed a data augmentation method for driving position area with the faster R-CNN module. The convolutional neural network CNN classification model was used to identify ten distracting behaviors in the AUC dataset, reaching the top-1 accuracy of 96.97%. The authors state that the extensive results carried out show that the proposed method improves the accuracy of the classification and has strong generalization ability. according to the authors,The experimental results showed that the proposed method was able to extract key information. This provided a path for the pre-processing stage of driving behavior analysis.

To address the driver and passenger safety needs, various researchers have focused on this problem and have tried developing different models based on DL for detecting driver distractions.

In the most of the proposed methods, the authors obtained satisfactory results on face detection and recognition. Indeed, the authors with the detection of driver fatigue and drowsiness, were able to reduce the number of accidents.

However, despite the encouraging results, we note some shortcomings:

— The face tracking method proposed by [14] is imprecise and very complex from a computational point of view, as the capture of the face is done in an inaccurate way and the complexity of the algorithm is very high.

— In the proposal of [5], we note the lack of successive frames consideration and the proposed technique is not robust enough.

— In [13], the proposed system does not take into account the detection of facial expressions such as fatigue and drowsiness, in addition, the chosen algorithms were based on GoogleNet which is not enough accurate to perform the motion detection task.

— With regard to the work of [6] and despite its good facial expression detection, the authors did not obtain good results when the face is turned or partially obscured by objects.

— In the work of [18], the performance of MobileNetV2 is lower due to the reduced complexity of the model and the reduced size of the input data.

— Finally, in the work of [1], although the results prove that it can maintain real-time performance, the model requires a lot of computing time.

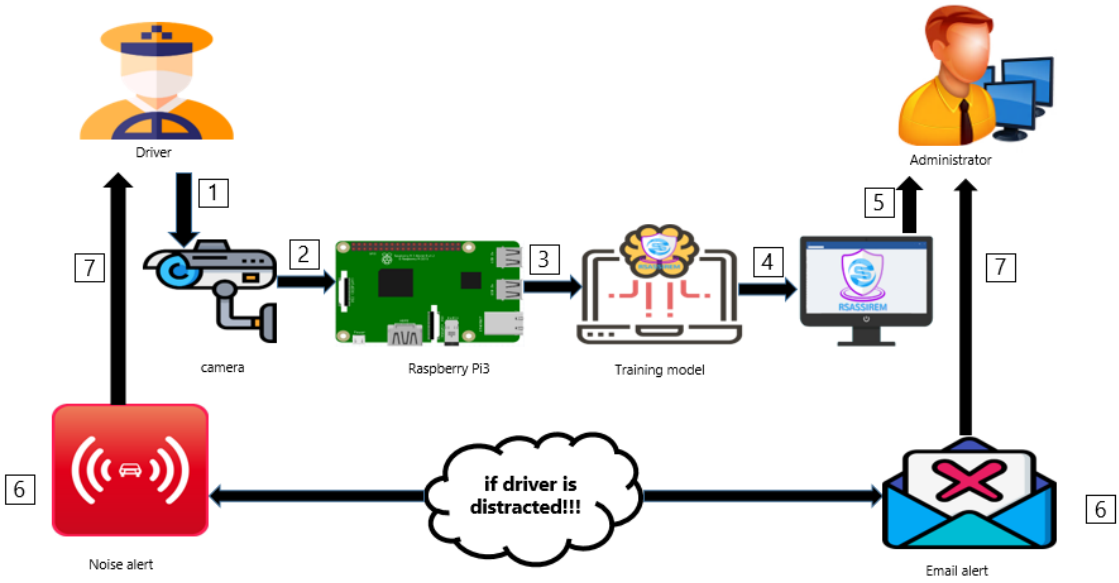

Fig. 1. System architecture

In order to overcome the above shortcomings, we focus on identifying the state of the driver (distracted or not). This technology could be applied in smart cities to automatically detect driver distraction and then send a warning message to the driver to prevent accidents.

For this purpose, we focus on detecting manual distractions where driver is engaged in other activities than safe driving and also identify the cause of distraction. The proposal involves forming a model for the driver distraction detection, using different CNN-based algorithms to improve the performance of the driver distraction detection behaviors.

The methodology is inspired by the related works but attempts to reduce the computational complexity and memory requirement while maintaining good accuracy which is desirable in real time applications.

The proposed system is supported by a software tool that processes data and sends an alert warning to the driver when he is a distracted state and it can be part of the intelligent transport system which will effectively monitor the state of transport while it is driving.Our proposal.

In the next section, we will present our proposal in detail.

4 Proposed Approach

Distracted driving is an activity that makes the driver lose his concentration in driving. A driver is considered to be distracted when there is an activity that attracts his/her attention away from the task of driving. Mainly, there are three types of driving distractions [6] :

— Manual distraction: The driver takes his/her hands off the wheel such as drinking or eating.

— Visual distraction: The driver looks away from the road, such as watching the phone.

— Cognitive distraction: The driver's mind is not fully focused on the driving task, for example, talking with passenger, or thinking.

The goal of the present work is to build an assisted-driving system by classifying the images that can detect distracted driving behaviors and alert the driver to focus on the driving task.

Our work deals with the cognitive distractions in different ways and predicts if the driver is in distraction state or not to ensure safer roads.

For this purpose, we propose an effective technique by creating a predictive model, based on CNN which covers more layers than a simple CNN, using transfer learning called VGG16 to solve driver distractions in different cognitive distraction ways such as: “talking to passengers”, “talking on the phone”, “putting on make-up”, etc.

In addition, we develop a warning system software that alerts the driver in real-time when he/she does not focus on the driving task.

The overall architecture of the proposed approach is shown in Fig 1 and it mainly involves five steps. The first step ”Image acquisition” consists in capturing the driver’s image.

The second step ”Feature Extraction” involves image processing and retrieval of the main features from the posture images and feeds them to the training model.

The third step ”Transfer learning on the VGG16 model” that learns the spectral correlations among the feature maps to predict the driver’s posture. The fourth step ”Image classification” which purpose is to categorizes the image according to the used Dataset. And finally, the five step ”Alert system activation” that avoids the driver if a distracted state is detected.

The whole proposal can be summarized as follows: a posture image is first captured, and then the image is processed to detect all the movements of the driver. According to the safe state image, if any change is observed that leads to distraction, a real-time alert is made to the driver using the alarm to get his attention while driving.

The system we propose uses the concept of tracking, which consists in monitoring persistent or long-term distractions, to be able to react in an anticipatory way and warn the driver progressively and smoothly, even before the driving situation becomes critical. Our system can act as a recommendation system in the context of intelligent transportation dedicated to road safety. In addition, it offers a significant response time. Below, we describe each step in detail.

4.1 Image Acquisition

This step is defined as the fact of extracting a posture image from a hardware source. It is the foremost stage in the workflow because, without an input, no prediction is possible.

We use an in-vehicle camera to stream the driver’s video while driving. We acquire image and video data from the in-vehicle camera and import it directly for visualization and further processing.

4.2 Feature Extraction

In this step, all the images captured by the camera are firstly resized to fulfill the input size requirements of each CNN model.

The images are resized to 224×224 and per channel the average RGB planes are subtracted from each pixel in the image. The initial layers of the CNN act as a feature extractor and the last layer is a softmax classifier that classifies the images into one of the predefined categories.

However, the original model has 1000 output channels corresponding to 1000 ImageNet object classes. The last layer is exploded and is replaced by a dense layer with 10 classes corresponding to the 10 classes present in our dataset (further, more details about dataset). Here, the cross-entropy loss function is used for performance evaluation.

The main operation of CNN is kernel convolution. We feed the input image from the training samples to the input layer of CNN and then apply kernel at different resolutions to convolve the image.

The Convolutional layers represent convolution is followed by a ReLU activation layer, pool layers represent pooling layers and fc layers represent fully connected layers than a Relu activation. The new addition to the model is the dropout layer that replaced the previous standards layers.

Dropout is mainly introduced to handle regularization by introducing noise into the model and thus reducing the number of overturning. The noise introduced by this function is done, with some flexibility, by disabling some perceptron to prevent it from storing models that are specific only to training data. Finally, another layer of convolution and pooling are included to add complexity to the model, since training is done on a GPU and computation is not such a major constraint.

The study focuses on VGG16, a version of the well-known convolutional neural network called VGG-Net. The cov1 layer input is a fixed size 224 × 224 RGB image. The image was passed through a stack of convolutional layers, where the filters were used with a very small receiver field In one of the configurations, we used 1×1 convolution filters,which can be seen as linear transformation of the input channels (followed by non-linearity). The convolution stride is fixed at 1 pixel;the spatial fill of conv.The input of the layer is such that the spatial resolution is preserved after convolution, i.e. the fill is 1 pixel for 3×3 convolutional Layers.Spatial pooling is achieved by five maximum pooling layers, which follow part of the conv. Layers (not all convolutional layers are followed by maximum pooling). Maximum pooling is performed over a 2×2 pixel window.

Three Fully Connected Layers follow a stack of convolutional layers which have different depths in different architectures. The first two has 4096 channels each,the third performs a 1000 channel ILSVRC classification and contains 1000 channels (one for each class). The final layer is the soft-max layer. The configuration of the fully connected layers is the same in all networks.

All hidden layers are equipped with grinding non-linearity (ReLU).We note that none of the networks (except one) contains Local Response Normalization (LRN) does not improve performance on the ILSVRC dataset but results in increased memory consumption and computing time. VGG16 learns 138.357.544 parameters. To find this number, we count the weights of all convolution and fully-connected layers, not forgetting the bias parameters.

4.3 Transfer Learning on the VGG16 Model

TL generally refers to the process in which a model trained on one problem and is used for a second, related problem by using the TL method, we can simply attach our dataset to an already trained model,since it is perfect and efficient.

We complete the construction of the VGG16 model by removing the last dense which involves 1000 categories, we take two categories as example.

```
model.add(Dense(2,activation='softmax'))
```

```
model.summary()
```

Layer (type)	Output Shape	Param #
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1180160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2359808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2359808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2359808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
flatten (Flatten)	(None, 25088)	0
fc1 (Dense)	(None, 4096)	102764544
fc2 (Dense)	(None, 4096)	16781312
dense (Dense)	(None, 2)	8194

```
Total params: 134,268,738
Trainable params: 8,194
Non-trainable params: 134,260,544
```

Fig. 2. Model summary VGG16

To finish the realization of the preformed model, we add the dense layer with two categories as an example,as we define the activation function for the model which is the softmax function (2).

4.4 Image Classification

We use the VGG16 network for the distracted driver image classification. This network was pre-formed on the ImageNet dataset. The purpose is that the pre-trained network would have learned characteristics from the previous data that could be useful for predicting distracted drivers.

This would not work right away, because the network has been trained to predict things like cats and dogs, and do not transfer images directly to drivers. The thought is to disconnect the fully connected layers of the VGG16, as they contain very specific weights to the original ImageNet dataset. In order to improve the model to match the distracted driver dataset,a new sequential model was created with a fully connected classifier similar to that of the single CNN.

4.5 Alert System Activation

The final step consists of alerting the driver if the system identifies a distracted state. The purpose of driver alerts is to bring the driver's attention back to the driving task after being distracted by an event, activity, object or person inside or outside the vehicle to predict roads safety.

Our system, in order to improve the storage space, allows to erase all the captured images once the driver quits the car. However, with the driver's approval, these images can be uploaded to the cloud storage to assist the model for further training and improve the response time.

An interesting feature of the proposed system is that the trained model will be implemented directly on the in-vehicle camera system for real-time driving detection. This feature provides reduced response time to alleviate the technical problems such as network latency and high power consumption.

5 Experiments

To validate the proposed approach and gain insights about its fullness and perfection, we implement a software tool in python under the operating system Ubuntu LTS 20.04 called RS-ASSIREM (R : Reda , S : Samia the software developers and Amazigh language word that means hope) which collects data from an inward facing camera and mobile device to perform driver behavior classification.

RS-ASSIREM is designed using Python language which is one of the languages that is witnessing incredible growth and popularity year by year. Python is also considered one of the best programming languages for ML since it includes different libraries that provide a simple way to implement ML algorithms.

The tool provides a user-friendly interface and simple to employ and includes different modules whose purpose is to perform the driver distraction behavior classification.

In the experiment, we use the Keras DL framework (version 2.2.4) based on Tenserflow (Version 1.12.0) to develop and implement our algorithms.

In addition, installation of cameras and sensors inside vehicles has helped us in observing driver behavior.

For the acquisition card we chose the Raspberry Pi card the latter is a small computer the size of a credit card, which can easily be connected to the Internet and serve as an interface to many electronic components. Indeed, despite its small size, it is as powerful as a smartphone. Raspberry Pi 3, includes a 4-core processor as well as 1GB of RAM. You can install a real operating system, such as Raspbian, Ubuntu or Windows.

With 40 GPIO pins, we can easily connect our board with many sensors and electronic components. Plus anyone can use a Raspberry Pi board. All we have to do is download an operating system, write it to a microSD card, connect our Raspberry Pi to a display. The card can easily be connected to the Internet and serve as an interface for many electronic components, such as the camera.

5.1 Dataset

The Dataset used in this study is provided by State Farm Insurance Company of the United States published on Kaggle[1]. This Dataset is the most commonly used Dataset for detecting the driver distraction and has been applied in many studies.

The State Farm Dataset consists of 22,424 training and 79,727 testing color images that showed drivers either captured in distracting activities or driving safely, it includes ten classes called categories, and each image is classified

[1] https://www.kaggle.com/c/state-farm-distracted-driver-detection

Fig. 3. Sample images representing different actions

among these categories. These categories are presented in Table 1.

There are approximately 2,200 RGB images for each class and each image's resolution is 640×480 pixels. Fig 3 provides Sample images representing different actions in the state farm dataset.

There are many types of distracted driving behaviors in real world scenarios. In this study, we consider distracted driving behaviors presented in table 1.

The dataset was divided into training and testing data, where 70% of the data are dedicated to training and 30% for testing and validation.

5.2 Model Formation

The purpose is that the trained network would learn features from the previous data that could be useful in predicting the images of distracted drivers.

In order to expand the model to match the distracted driver dataset, a new sequential model was created with a fully connected classifier similar to the single CNN.

The training and validation images were propagated through the VGG16 network only once, the characteristics of the output bottlenecks were recorded in files and then fed into the top model of the newly created fully connected classifier.

We note that only the upper model was trained on the new dataset, and only the learned characteristics are present in the distracted driver dataset. Due to the reduced training time, we increase the image size to 224×224, which allows to remain lot of information.

5.3 Model Training

The training model was carried similarly to the training method applied in the Simple CNN model. The validation set was used with a patience of 60

Table 1. Dataset details with the number of images per class

Class	Description	Images
C0	Driving Safety	2,489
C1	Texting with right hand	2,267
C2	Talking with right hand	2,317
C3	Texting with left hand	2,326
C4	Talking with left hand	2,326
C5	Operating the radio	2,312
C6	Drinking	2,325
C7	Reaching behind	2,002
C8	Hair and Makeup	1,911
C9	Talking to the passenger	2,129

and worked with 50 epochs. After 14 epochs, there were no (positive) changes in the accuracy of the validation set, so the training was terminated, and the best weights were retained. The model overall accuracy is 99.396%. These results are quite promising as the training take only a few minutes.

After training the VGG16 model, the comparison between epoch 1 and epoch 50 shows that the error rate decreases when the accuracy increases. This means that our model was well trained and fulfills the neural network definition that the deeper the neural network, the better its performance.

We note that all images were taken from the same angle, at the same environmental conditions (i.e. during the day), which is both a good and bad feature for model training.

However, it can be considered as a positive fact if a subset of the data is used as a test set, the model will most likely perform well since the test images have the same conditions. However this is an issue when the model is generalized to images that do not necessarily match to the same conditions as the original data set.

6 Results and Discussion

To evaluate the experimental results, a standard performance metrics is ”Accuracy” which is well used . This evaluation metric is computed by using the following equation:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}, \quad (1)$$

in which TP is the number of true positives samples in the testing set, which were correctly detected, and TN is the number of true negatives, which truly gives the number of negative samples in the testing set. FP is the number of false positives, which give the number of false positives in the testing set that are actually incorrectly detected. FN is the number of false negatives, which gives the number of false negative samples in the testing set that are incorrectly detected.

Accuracy is used to calculate the portion of samples that are detected correctly during the testing phase with all the data sets.

The VGG model has overhead costs when creating the bottleneck functions, but can be stored offline and loaded at any time, so the subsequent formation of fully connected layers took much less time. This model is good because it not only provides the highest accuracy, but also allows training in the shortest time possible. It was so powerful that it can even be trained on an ordinary laptop with a very reasonable training time.

Let’s analyze the model in more detail by looking at the receiver operating characteristic (ROC) curve shown in Figure 4, by comparing the actual positive rate (TPR) and the fictitious positive rate (FPR). An optimal prediction would have an RPR of 1 and an RPF of 0, which corresponds to the upper left corner of the graph.

A stronger and faster increase in the upper left corner indicates a solution close to the optimum, as shown in Figure 5. Another feature in the graph is its strong curvature, which indicates a very clear and almost perfect separation.

The change in accuracy and categorical cross-entropy (called the loss) can be seen in Figure 6, as the model is formed through the different epochs.

The training and validation accuracy gradually increase over the epochs in Figure 6 until they reach a certain threshold, indicating that no further improvements are made and leading to the end of training. Similarly, in Figure 5 the loss of validation

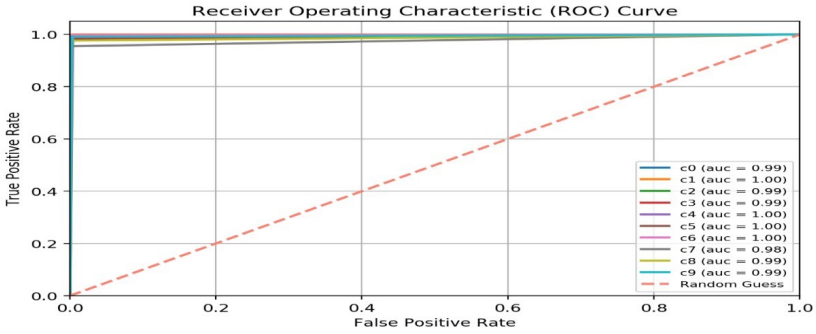

Fig. 4. ROC curve of driver distraction predictions

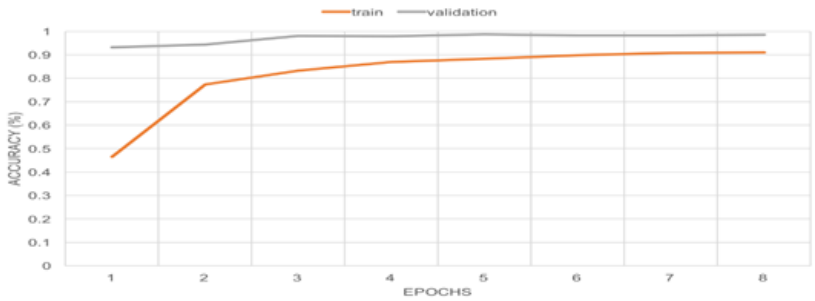

Fig. 5. Model accuracy

and training decreases as the model learns, until it finally ends when the model begins to over-fit.

The approach of [1] provide good results regarding detect driver distractions while driving, however our approach detects not only the 10 classes of distractions mentioned in Table 2 and provides an accuracy that exceeds 99% in a limited period of time thanks to the use of transfer learning.

Our proposal provides also an opportunity to broaden the scale of our work taking into account the detection of driver hypo vigilance which is the one of the major road accidents causes, this system seems effective because it detects the ocular regions in different light conditions and allows an alarm to be triggered when the driver begins to fall asleep. The proposed system can be part of the intelligent transport system which will effectively monitor the state of public transport while it is driving.

Another interesting evaluation measure that we have used is the confusion matrix shown in figure 7. Most predicted labels have an accuracy close to 100%. However, c7 stands out with an accuracy of 0.96, which is relatively low compared to the rest of the classes. This class corresponds to ”Getting in tune”, as shown in table 1.

The confusion matrix allows us to conclude that the class c7 prediction is the worst ranked or most difficult to predict.

In addition, the confusion matrix shows that c7 is most often mislabeled as c1, which means that tuning up is usually due to text mislabeled with the right hand. This is understandable, since drivers who stand behind usually perform this action with their right hand raised.

Thus, the model perceives that the drivers are talking on the phone with their right hand if they are captured in a specific movement frame.

Finally, we were able to record the model’s weight, display previously unseen images of distracted drivers, and display the model’s five most important predictions.

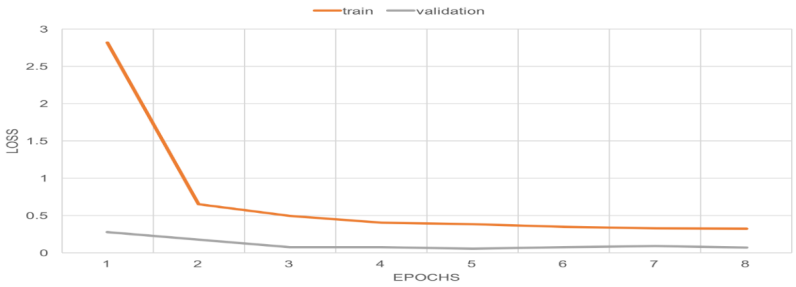

Fig. 6. Model loss

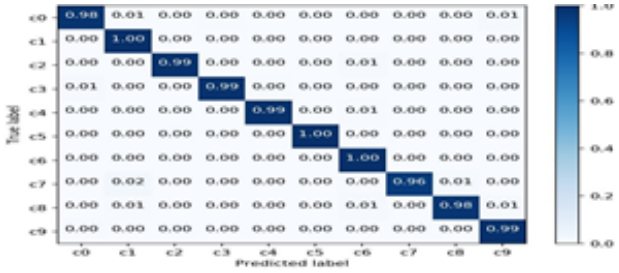

Fig. 7. Model Confusion Matrix

Another interesting feature provided by our system is the Processing Time. Performance analysis is performed to ensure that software application will perform well under their expected workload. Most performance problems revolve around speed, response time, load time and poor scalability.

Response time is often one of the most important attributes of an application. A slow running application will lose potential users. Performance testing is done to make sure an app runs fast enough to keep a user's attention and interest. The total average query processing time taken by the proposed system with maximum workload is 2.6592 ms.

The approach of [1] provides good results regarding detect driver distractions while driving, however our approach detects not only the 10 classes of distractions mentioned in Table 1 and provides an accuracy of 99% in a short time through the use of transfer learning.

In summary, the proposed system contributes efficiently and is able to update the model and provide high accuracy for simulated custom data sets representing real-life applications.

7 Conclusion and Future Work

Driver distraction is a major issue which leads to a consistent increase in the number of accidents vehicle accidents. Usually, drivers can become distracted by different activities such as talking to a passenger, talking on the phone, drinking, eating, etc. Therefore, it is essential to build a system that detects such activities and alerts the driver in order to avoid or reduce road accidents.

In this paper, we first overview the most related works according to the driver distraction detection. Further, we introduce our model for detecting driver distraction and drowsiness to avoid or reduce the number of traffic accidents and fatalities.

We use a pre-trained VGG16 network, the model is able to achieve an accuracy over 99% on the data test. Our model is able to fulfill its role as a distraction detector with great success. We have also implemented a drowsiness detection model using OpenCV and Dlib.

The proposed model will detect the driver distraction and will alert with a beep sound the driver according to the duration of the distraction.

In addition, we built a user friendly software tool which detects all types of distraction and drowsiness. If the driver is distracted an alert signal will be send from the Raspberry to warn the driver of the different types of risks in accidents.

We use the existing State Farm Distraction Dataset for fine-tuning our models. The dataset is divided into training and testing data, where 70%

of the data are dedicated to training and 30% for testing and evaluation.

We believe that our proposal is helpful because it gives law-officers the ability to identify a distracted driver and monitor them with radar and cameras.

Our future work includes building a real-time detection system using wireless techniques to impose paying penalties based on driver distraction. Thus, this proposal, reminds drivers of the seriousness of distraction and reduces the different types of distraction as much as possible.

In addition, the detection of drowsiness is a topic of interest in several research projects. For future work, we plan to improve a distraction detection system by integrating a driver drowsiness detection model. It is easy to discern that the subject of automatic sleepiness detection while driving include face detection, eye detection, eye aperture percentage and other drowsiness situation.

References

1. **Alotaibi, B., Alotaibi, M. (2020).** Distracted driver classification using deep learning. Signal Image and Video Processing, Vol. 14. DOI: 10.1007/s11760-019-01589-z.
2. **Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M., Al-Amidie, M., Farhan, L. (2021).** Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. Journal of Big Data, Vol. 8.
3. **Dash, S. K., Sureshchandra, Y. V., Mishra, Y., Pakray, P., Das, R., Gelbukh, A. F. (2020).** Multimodal learning based spatial relation identification. Computación y Sistemas, Vol. 24, No. 3.
4. **Faria, M. V., Baptista, P. C., Farias, T. L., Pereira, J. (2020).** Assessing the impacts of driving environment on driving behavior patterns. Transportation, Vol. 47, No. 3, pp. 1311–1337.
5. **Fu, L.-H., Wu, W.-D., Zhang, Y., Klette, R. (2015).** Unusual motion detection for vision-based driver assistance. International Journal of Fuzzy Logic and Intelligent Systems, Vol. 15, No. 1, pp. 27–34.
6. **Jeong, M., Ko, B. C. (2018).** Driver's facial expression recognition in real-time for safe driving. Sensors, Vol. 18, No. 12.
7. **Kaiser, C., Stocker, A., Papatheocharous, E. (2021).** Distracted driver monitoring with smartphones: A preliminary literature review. Proceedings of the XXth Conference of Open Innovations Association FRUCT, volume 29, pp. 169–176. DOI: 10.23919/FRUCT52173.2021.9435545.
8. **Kattenborn, T., Leitloff, J., Schiefer, F., Hinz, S. (2021).** Review on convolutional neural networks (cnn) in vegetation remote sensing. ISPRS journal of photogrammetry and remote sensing, Vol. 173, pp. 24.
9. **LeCun, Y. (2015).** Deep learning & convolutional networks. Hot Chips Symposium, pp. 1–95.
10. **Liu, Y., Zhang, Y., Li, J., Sun, J., Fu, F., Gui, J. (2013).** Towards early status warning for driver's fatigue based on cognitive behavior models. **Duffy, V. G.**, editor, Digital Human Modeling and Applications in Health, Safety, Ergonomics, and Risk Management. Healthcare and Safety of the Environment and Transport, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 55–60.
11. **López-Monroy, A., Aldana, D., Miranda, A., Carmona, J., Espinosa, H. (2021).** Deep learning for language and vision tasks in surveillance applications. Computación y Sistemas, Vol. 25. DOI: 10.13053/cys-25-2-3867.
12. **Majdi, M. S., Ram, S., Gill, J. T., Rodríguez, J. J. (2018).** Drive-Net: Convolutional network for driver distraction detection. 2018 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI), pp. 1–4. DOI: 10.1109/SSIAI.2018.8470309.
13. **Sheng, W., Tran, D., Do, H., Bai, H., Chowdhary, G. (2018).** Real-time detection of distracted driving based on deep learning. IET Intelligent Transport Systems, Vol. 12. DOI: 10.1049/iet-its.2018.5172.
14. **Sigari, M.-H., Fathy, M., Soryani, M. (2013).** A driver face monitoring system for fatigue and distraction detection. International journal of vehicular technology, Vol. 2013.
15. **Simonyan, K., Zisserman, A. (2014).** Very deep convolutional networks for large-scale image recognition. arXiv 1409.1556.
16. **Simonyan, K., Zisserman, A. (2015).** Very deep convolutional networks for large-scale image recognition.
17. **Wang, J., Wu, Z., Li, F., Zhang, J. (2021).** A data augmentation approach to distracted driving detection. Future internet, Vol. 13, No. 1, pp. 1.

18. **Wilhelm, T. (2019).** Towards facial expression analysis in a driver assistance system. 2019 14th IEEE International Conference on Automatic Face Gesture Recognition (FG 2019), pp. 1–4. DOI: 10.1109/FG.2019.8756565.
19. **World Health Organization (2018).** Global status report on road safety 2018: Summary. Technical report.
20. **Zhao, G., He, Y., Yang, H., Tao, Y. (2021).** Research on fatigue detection based on visual features. IET Image Processing.

Article received on 23/11/2020; accepted on 20/09/2021.
Corresponding author is Houda El Bouhissi.