Identifying Driver Distraction Using Deep Neural Networks

MSc Research Project
MSc Data Analytics

Pushkar Dashpute
X18180124

School of Computing
National College of Ireland

Supervisor: Dr. Manaz Kaleel

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Pushkar Anil Dashpute

Student ID: x18180124

Programme: MSc Data Analytics **Year:** 2019-2020

Module: MSc Research Project

Supervisor: Dr. Manaz Kaleel

Submission Due Date: 28th September 2020

Project Title: Identifying Driver Distraction Using Deep Neural Networks

Word Count: 7101 **Page Count** 19

I hereby certify that the information contained in this (my submission) is information pertaining to the research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Pushkar Anil Dashpute

Date: 27th September 2020

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on the computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Identifying Driver Distraction Using Deep Neural Networks

Pushkar Dashpute
x18180124

Abstract

With the rise in globalization, Distracted Driving induces many deaths in road accidents and has become an increasingly relevant matter to discuss in recent traffic safety research. Driver Distraction has a major effect on people's safety and is an important subject for a variety of applications, ranging from automated driving assistance to insurance firms and research. The aim of the research in this paper is to design a system that can detect the distraction of drivers caused due to driving. The dataset used for the research was selected from Kaggle which was publicly accessible. This dataset was from State Farm Insurance company. This data was later transferred into binary form by data structure automation using Python. Further, the proposed design exploited Deep Learning models namely Convolutional Neural Networks (CNN), Xception, VGG19, and Thin MobileNet. These models are further assessed based on evaluation metrics namely accuracy and computational time required by each model. It was observed that Thin MobileNet outperformed all the other models in terms of accuracy which was 98.39% and computational time of 45 min for the training period of 10 epochs.

1 Introduction

Driver distraction is a significant threat to the younger generation's public health and safety for those who drive on public highways which gives rise to the mortality rate of teenagers. This distraction caused during driving is nothing but inattention of a driver to some secondary activity while driving. Major factors that are responsible for fatalities caused by distraction from drivers include the use of music devices, getting phone calls, making hair, and communicating with passengers. Through the growing use of social media, awareness of distractions caused by driving behaviours such as texting while driving is widespread. Few other reasons that contribute to driver distractions are lack of sleep and fatigue among the drivers. These distractions are mainly categorized as visual, cognitive, and manual. The visual distractions are those which are caused due to eyesight getting off the roads while the cognitive distractions include daydreaming or mentally not present in driving the vehicle. And finally, the manual form of distraction which includes other actions such as texting on the phone, getting calls, consuming, or drinking food, and many others. Drivers must be aware of the pitfalls incurred while driving because of any such activities[1].

[1] https://www.cdc.gov/motorvehiclesafety/distracted_driving/index.html

As per the survey done by the World Health Organization (WHO), road accidents estimate to almost 1.35 million deaths yearly among which about 50 million of them suffer from non-fatal injuries due to road traffic crashes[2]. Adding more to this, as reported by the national highway traffic safety administration (NHTSA), distracted driving caused more than 3000 deaths in the year 2017. In a tenure of 5 years i.e. from 2012 to 2017, about 20,000 people lost their life due to distracted driving[3]. This mortality rate can be compared to deaths caused by crashes that take place due to about seven Boeing aircraft. This increasing rate of deaths due to car accidents is not a good sign for the upcoming generation with the increasing use of vehicles every day. Moreover, accidents due to driving cause immense damage to property.

The above-mentioned stats and to overcome such challenges motivate to work on this research where the distraction of drivers can be identified and reduced. Deep learning has proved to perform significantly well in the field of object detection and image classification. A future could be foreseen where the smart cars may identify and handle such distractions and alert the driver and further take preventive action. A detection system designed can help the laws and regulations to recognize the distractions caused on the highway by applying models developed using deep learning which can further enable to implement penalties on certain kinds of distraction.

Further, the research here aims at identifying the driver distraction by classifying the images which consist of both i.e. the driver's in the state of normal driving as well as driver's carrying out secondary activities while driving. Deep learning models will be used in this paper for the same purpose which includes Convolutional Neural Networks (CNN), Transfer learning Models, and finally applying an Ensemble technique. Neural networks have proved to perform well in classification related problems. Additionally, using the Ensemble technique has improvised the results obtained from transfer learning architectures. Thus, this inspired me to make use of ensemble techniques along with transfer learning models to classify whether a driver is distracted or not while driving.

Research Question:

RQ: "How well can Deep Neural Networks using Transfer learning identify distracted drivers into safe or distracted driving?"

The later part of this research is discussed in the following sections. The first section consists of a critical review of various studies done concerning different techniques implemented in the detection of driver distraction. The next section briefs the research methodology followed for the research study which is a Cross-industry standard process for data mining (CRISP-DM) followed by detailed information on the implementation of different models used. And finally, the future work and conclusion are discussed in the last part of the paper.

Research Objective:

[2] https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries
[3] https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death

Table 1 Research Objectives

Objective	Description	Metrics
Objective 1	A critical review of the study related to the Detection of driver distraction.	Identify state-of-the-art.
Objective 2	Pre-processing of data and selection of features to classify the drivers as distracted or not distracted.	-
Objective 3	Implementation of deep learning models used for detecting the driver's distraction by classifying them as distracted or not distracted.	Accuracy Computational Time
Objective 3.1	Implementation, Evaluation, and results of Convolutional Neural Network (CNN).	
Objective 3.2	Implementation, Evaluation, and results of Xception	
Objective 3.3	Implementation, Evaluation, and results of VGG19.	
Objective 3.4	Implementation, Evaluation, and results of Thin MobileNet.	
Objective 4	Comparison of models developed.	-

These objectives are implemented in the sections below.

2 Related Work

Driver Distraction has been a major problem since recent years thus resulting in fatalities. 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.

2.1 Application of CNN based models in identifying driver distraction.

Lorem Many of the accidents are caused due to sleepiness in the drivers. Thus, it is important to detect drowsiness among drivers. In this paper, the use of representation-based learning along convolutional neural networks is done which classifies whether the driver is drowsy or not. The dataset was split into 5 folds by applying cross-validation where 4 of them were used for training while one was used for validation. The images in the dataset were then resized to 48×48. A validation accuracy of 92.33% was obtained by the model. Although the results obtained are satisfactory, the author mentions of applying 3D convolutional neural networks for reliable driver identification of sleep state (Dwivedi, Biswaranjan and Sethi, 2014).

Majdi *et al.*, (2018) has presented a supervised deep-learning 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 perceptron (MLP). The dataset used was picked from the Kaggle competition which was publicly available. Further, 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.

The method proposed here is monitoring the driver's distraction by means of facial features such as nose, eyes, and mouth. The system introduced here is based on the CNN network where a minor variation is done in order to implement a method for face detection, localizing landmarks, and head-positioning. The dataset used here consists of faces that are used for face detection. This system used the AlexNet framework which was further passed to the second phase enhancement. It was observed that the second stage refinement had substantially improved precision from the face detectors as compared to stage one. The method here got a good rate of detection near to 86%. The work done in this paper would deliver a strong base for higher advancement in the facial analysis of the driver's distraction state (Yuen, Martin, and Trivedi, 2016).

Deep Convolutional Neural Networks have proven significant success in the classification of images in recent decades. In this research, the author has performed a comparison of two different architectures i.e. Inception v3 and Xception to recognize ten various types of driver behaviors. In the pre-processing stage, normalization and resizing of the images were done. Further, using transfer learning along with the Xception model, a validation accuracy of about 85% was achieved where initialized ImageNet weights were used. Later, this accuracy was increased to 99% for the same architecture the difference being the training data was split randomly rather than splitting into subjects. In the future, the author mentions of using an ensembled approach combining both Xception and Inception v3 which could offer more precise results (Varaich and Khalid, 2019).

Gjoreski *et al*., (2020) provided an analysis to determine which could be the Machine learning (ML) methods that can better identify different driver distraction behaviors. The statistical study proved that the most helpful feature in order to detect the distraction of a driver is by its type. The experimental study carried out included 7 standard ML and Deep-learning (DL) models which were tested on 10 different subjects. The extreme gradient boosting (XGB) achieved the best results when used with AUs outpacing the deep learning models. Nevertheless, in the DL models, the Resnet (spectro-temporal) performed the best among all others. Moreover, the author identified a few problems such as label jitter, overfitting which needs to be looked after for enhancing the results in the future.

The distractions caused by car behaviors such as talking with passengers, running music systems and smartphone texting are crucial to identify. Conversely, the impact of lighting conditions and skin color of the driver's hands makes it very hard to recognize biomechanical disturbances. The framework used in the study had two main sub-models, i.e. localizing hand and face using a layered autoencoder, and using the Convolutional Neural Network to classify each class of biomechanical distraction. The number of convolutional layers applied in the second sub-model was 6, while the polling layers were 4. The performed experiment achieved an accuracy of 98.68% which outperformed the previous state of the art. For the future, the study is aimed at focusing on the driver's visual distractions that have created significant problems caused by improper-attention to driving (Asefa and Wenhong, 2019).

Baheti, Gajre, and Talbar (2018) aimed to build a CNN-based framework that not only senses distraction from the driver but also recognizes the reason for distraction. For this particular task, VGG-16 architecture is altered, and numerous regularization methods were implied with a view to improving performance. Later, all images were resized to 224 / 224 in the pre-processing phase and each pixel in an image was deduced from the channel mean of RGB planes. The results gained a high accuracy of around 96% which were far better than the results achieved in previous findings. The study intends to extend the current work by reducing the measurement time and number of parameters which can minimize the misclassification caused and enhance the accuracy. The author also hopes to construct a system that can detect cognitive and visual distractions along with manual distractions.

The chain of crash accidents in mainly because of harsh driving thus it is important to monitor the necessary information. Kim, Choi, and Jang (2018) suggested a model that can detect both the drowsiness and distraction of a driver. The model which was used comprised of a single CNN which further uses MobileNet with 27 layers. It involved 3 distinct steps in the preprocessing step, i.e. Random image crop with the bounding box, Random brightness adjustment, contrast and image saturation, and Random flipping of images to double the training cases. Furthermore, the findings obtained from the study appeared substantially enhanced with addition to a reduced fall in frames per second. The work aims at detecting driver emotions in the future in reference to the amount of study that has been already done on the detection of human emotions.

2.2 Used cases of different transfer learning models used in driver distraction detection.

Oliveira and Farias (2018) presented a comparison of various transfer learning methods applied to the classification of images of distracted drivers. Four different architectures were mentioned by the research namely VGG19, Inception v3, Densenet161, and Resnet152. Each of the algorithms was trained using end-to-end optimization of transfer learning. There were three different transfer learning approaches discussed which included transfer learning with: end-to-end refinement (TL1), fully connected layer tuning (TL2), and feature extraction (TL3). When testing was performed using all the three techniques, the results showed that Densenet161 performed well for the TL1 and TL2 while for the TL3, VGG19 performed best among all. The author tends to expand its research in understanding the results obtained and gain knowledge in more depth for the discussed techniques.

Koesdwiady *et al.*, (2017) suggested an end-to-end deep learning method for recognizing the driver's distraction during driving where a transfer learning architecture VGG-19 was used. The model here performed well with an overall accuracy of 95% for testing, overcoming the various challenges such as lighting conditions, positioning of the camera, or gender. The model also outpaced the existing XGBoost by an accuracy difference of 7%. The major challenge faced by these end-to-end frameworks is the complexity of tuning the neural networks because it takes a quite long time and also a large amount of resources. The future work could include working on increasing the dataset size and further using the ensemble techniques to enhance accuracy.

The author proposed an approach that detects the distraction of drivers by use of transfer learning. The ResNet-50 model with pre-trained weights on ImageNet was used which was trained on data that was acquired by the use of a driving simulator. However, in order to increase the size of the dataset, more images were collected from the internet. Also, it was noticed that the combination of these datasets gained the best results. In addition to this, the accuracies that were achieved separately for the internet and simulation dataset were 95% and 98% respectively. The system's reliability was demonstrated by testing it under various lighting conditions. This test results demonstrated that even with a fairly large shift in luminance, the proposed method was able to retain a strong classification accuracy (Ou *et al.*, 2018).

A pre-trained transfer learning model i.e. AlexNet is fine-tuned in order to identify the driving distraction issue. In the pre-processing stage, the images are segmented using GMM (Gaussian mixture model) which eliminates the unrelated objects from the image which could further help in improving the accuracy. Further, the transfer learning architecture AlexNet was used which identified the seven different driver distraction classes with an accuracy of about 79%. However, if the CNN based model used as a binary classifier it would give better accuracy of nearly 94%. In the future, the author states of using more distinct models namely VGG or GoogleNet which may help in improving the detection accuracy as it was seen in the current study that some of the distraction behaviors were not correctly identified (Xing *et al.*, 2018).

The paper presented by Tran *et al.*, (2018) sorely discusses the various deep learning architectures to overcome the driver's distraction activities. Four deep Convolutional Neural Networks (CNNs) were employed and assessed on an embedded GPU network, namely AlexNet, GoogleNet, and ResNet, VGG-16. The data used in this study was developed using an assisted testbed of driving where the models can be evaluated. The results obtained revealed that GoogleNet performed the best among four of the architectures. Furthermore, an alert system was also developed which would warn any driver while carrying out any activities other than driving. In later work of research, the author suggests of improving the misclassifications occurring in the current implemented model which could be done by the use of classifiers like SVM.

2.3 Ensemble Technique

A robust vision-based system was introduced which identifies distracted postures of a driver. An ensemble technique was implemented consisting of convolutional neural networks with an ensemble of genetic weights. This model was trained on images of hands, face as well as a combination of both. The algorithms used for these image sources were AlexNet, ResNet, InceptionV3, and VGG-16. Further, a weighted sum of all of these networks generating a genetic algorithm was evaluated which performed with great accuracy of 90%. The author aims at using the Fast-RCNN over the current CNN to produce better detection enabling to locate both faces as well as hands in one shot (Eraqi *et al.*, 2018).

Object detection in dynamic backgrounds with abrupt image variations is a complicated subject in computer vision. The author presents a novel approach that includes a deep learning-based architecture with a framework of on-line AdaBoost. A stacked denoising autoencoder (SDAE) was used to learn multi-level features from a set of images. Moreover, good results were achieved with the two-fold approach that consisted of deep learning architecture and boosting algorithm i.e. Adaboost. In the future, the research expects to use a convolutional neural network which has lately shown greater effectiveness in extracting features of images.

Presently, the DNN takes a long time without GPU i.e. it requires high computing cost thus reducing it in future work. An ensemble of deep neural networks (DNN) built from the SDAE layers are combined with the framework of on-line Adaboost to classify the actual target from the background (Zhou *et al.*, 2014).

Table 2 Comparative study of work related to the Detection of Driver Distraction

Name	**Year**	**Algorithm Used**	**Dataset**	**Evaluation Metrics**	**Reference of paper**
Drowsy driver detection using representation learning	2014	Convolutional Neural Network	Dataset created using 30 different subjects	Accuracy	https://ieeexplore.ieee.org/document/6779459
Driver Distraction Detection Using Semi-Supervised Machine Learning	2016	Laplacian Support Vector Machine	Data collected in an on-road experiment	Accuracy, Sensitivity, Specificity and G−mean	https://ieeexplore.ieee.org/document/7347425
Driver distraction detection using a single convolutional neural network	2017	Single Convolutional Neural Network.	ILSVRC2012 Dataset	Accuracy & Processing speed.	https://ieeexplore.ieee.org/document/8190898
Detection of Distracted Driver Using Convolutional Neural Network	2018	Convolutional Neural Network. VGG-16	Y. Abouelnaga, H. M. Eraqi, and M. N. Moustafa. Real-time distracted driver posture classification. CoRR, abs/1706.09498, 2017.	Accuracy & Confusion Matrix	https://ieeexplore.ieee.org/document/8575304
Detecting Distraction of Drivers Based on Residual Neural Network	2019	Residual Neural Networks	StateFarm dataset	Accuracy	https://ieeexplore.ieee.org/document/8942692
HCF: A Hybrid CNN Framework for Behavior Detection of Distracted Drivers	2020	Convolutional Neural Network, ResNet50, Inception V3, Xception.	StateFarm dataset	Accuracy & Training time	https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9113267

3 Research Methodology

The methodology adopted for the purpose of this research is CRISP-DM which will enable us to understand the research in more depth with respect to the subject matter. The CRISP-DM methodology is found to be really useful when using an analytical approach to solve business-related issues. Further, a modified approach of CRISP-DM will be followed concerning identifying the distraction of drivers while driving which will include a detailed process flow. Following Figure 1 shows the graphical representation of CRISP-DM with respect to research.

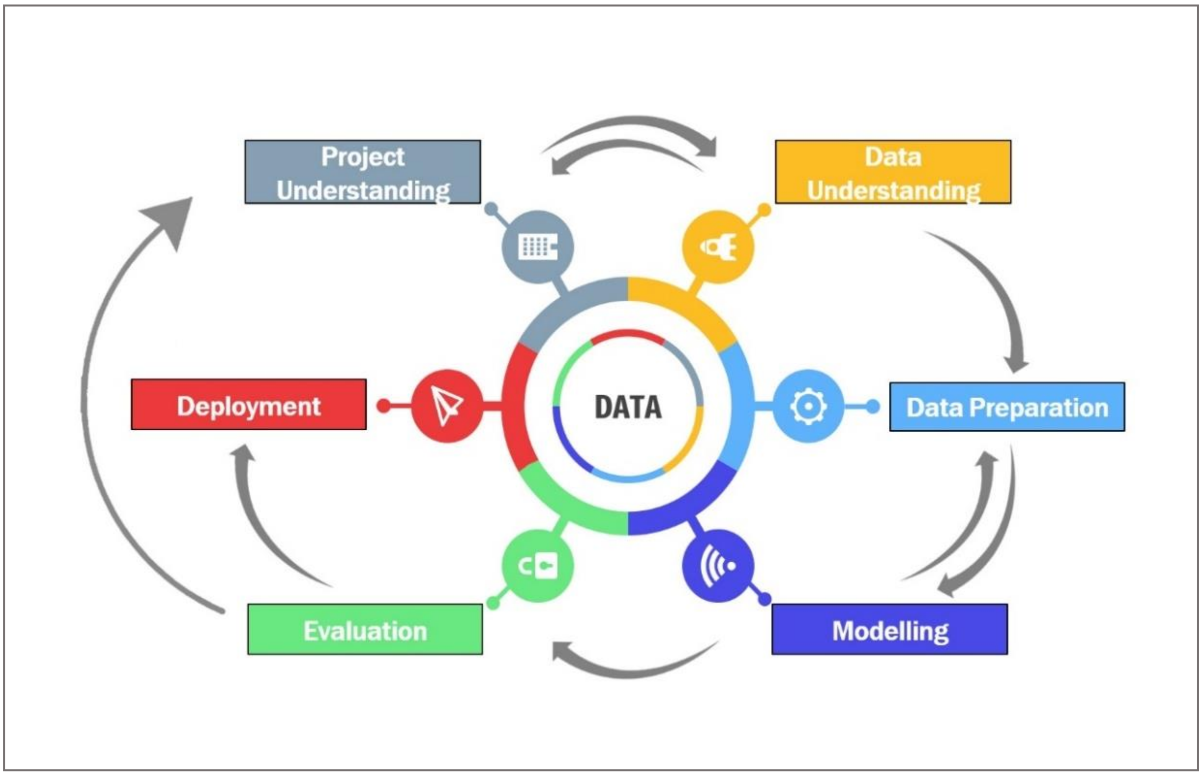

Figure 1 Driver Distraction Detection Methodology

3.1 Project Understanding

This section of the research addresses the requirements of projects also the objectives which are considered in the study. In consideration of identifying the driver distraction, a monitoring system can help to reduce the number of fatal accidents taking place. This can be done using deep learning techniques that would be able to classify whether a driver is distracted or not while driving. The various CNN based transfer learning models along with ensemble technique is used for resolving this purpose of distraction caused.

3.2 Data Understanding

This is one of the most important stages where the data used has to be understood completely with respect to its nature and eccentricity. The data was collected from a publicly available source, Kaggle which was published by State Farm Insurance company in the year 2016. This collected data was understood thoroughly with respect to the upcoming steps such as data pre-processing as well as the models to be applied.

3.3 Data Preparation

This stage includes preparing of data before applying the different models on the data. In order to understand the data in depth, few of the python libraries were used to visualize the data. Also, the image resizing, and normalization was done in the pre-processing stage.

3.4 Modelling

The data modelling stage comes after splitting of data into train and test sets and further building a variety of models in regard to data used. Following are different models which are implemented in classifying whether the driver is distracted or not while driving on road:

3.4.1 Convolutional Neural Networks

To deal with the issue of distraction caused during driving, a CNN model is implemented which has shown pretty good success in the classification of images which makes this model a perfect fit. CNN's would be used to achieve considerably higher levels of image data interpretation. The neural network consists of weights and biases in learning. The structure of CNN includes basic building blocks namely, convolution layers/filters, activation functions, pooling layers, and fully connected layers. It is generated basically by stacking certain layers one after the other. However, CNN's don’t use fully connected layers for each of the pixels instead it uses adequate weights which are used to look at little parts of the images from data. Additionally, CNN's depend more on the fact that spatial knowledge of an image is sufficiently strong, with the practical advantage of using fewer parameters, thus reducing the computation time and data needed for model training.

3.4.2 Xception

The Xception model is a CNN based architecture which is wholly based on depthwise convolutional layers. Xception is described as a starting module hypothesis that generates correlations of the cross channel and spatial interactions which can be fully decoupled within CNN feature maps. This model is considered to be an upgraded version of Inception architecture. The number of layers present in Xception architecture is 36 which forms a base of feature extraction. Following layers are configured into 14 modules. These are the depthwise layers of separable convolutional layers which are linearly stacked along with residual connections except for the first and the final module. Moreover, Xception follows a three flow-based approach where the data used firstly goes through the entry flow followed by the middle flow which gets repeated eight times, and then finally through the exit flow. This dynamic nature of the architecture makes it quite simple to define and modify by using a variety of high-level libraries namely Keras or TensorFlow (Chollet, 2017).

3.4.3 VGG19

VGG is a deep CNN based architecture which can be used in the classification of images. It is composed of multi-connected convolutional layers. One of the models used in this study is VGG19. It consists of 19 layers where 16 of them are convolutional layers while 3 are fully connected ones. The input for this model remains of constant size i.e. an image of 224 x224. The VGG19 model monitors the complexity of a network by putting 1x1 convolutions between convolution layers, which also learn to combine the resulting feature maps in a linear manner.

Later, the use of small-sized filters gives a further advantage of low computational complexity by lowering the number of parameters. The advantage of using VGG19 over VGG16 is, it uses extra ReLu units of convolution which enhances the accuracy of the model when an object detection task is carried out (Wen, Li, Li, and Gao, 2019).

3.4.4 Thin MobileNet

MobileNet is one of the most widely used models in the field of computer vision with a broad variety of applications in object detection. MobileNet has several benefits over other models i.e. it has limited parameters, less complexity and performs fast. However, there are few challenges faced by this architecture such as a limited amount of power and memory usage. Also, the model's overall accuracy usually reduces if the size and number of parameters reduced using some approach like pruning or deep compression. Further, an updated version of MobileNets will be included in this which is named as Thin MobileNet. It involves two types of blocks where one is a residual block with 1 step and another block with 2 steps for downsizing. Every block form has three layers, i.e. first layer 1 with ReLU6, the second layer includes a depthwise convolution and the final layer is again a 1×1 ReLU6 convolution that does not have non-linearity. This enhanced architecture has strengthened accuracy, and also the size and computation time of the model has been reduced instantly that will further help the model overcome the overfitting issue (Sinha and El-Sharkawy, 2019).

3.5 Evaluation

This is a stage just before the final implementation of the model is completed. In order to be confident enough to achieve the research goals properly, the evaluation and analysis of various parameters involved after building a model is carefully done in this stage.

3.6 Deployment

This is the final stage involved in a CRISP-DM approach where the implemented models are interpreted in order to gain knowledgeable insights. An extra attention is given to the research study in this step where it is monitored if the research is done can become an integral part of daily life.

4 Design Specification

Architectural design is built to develop an efficient framework that helps in classifying drivers as distracted or not distracted. The architecture shows the different tools and techniques used in order to build an efficient deep learning system. The paper presented here follows a two-tier architecture which is named as Presentation Layer and Business Logic Layer. The business logic layer describes the data source, the tools and techniques used in the pre-processing of data, the various deep learning models used for classification, and finally the evaluation of those models. On other hand, the presentation layer represents the visualization of obtained results which can be further presented to clients for better insights. Below is the design specification architecture is shown in Figure 2.

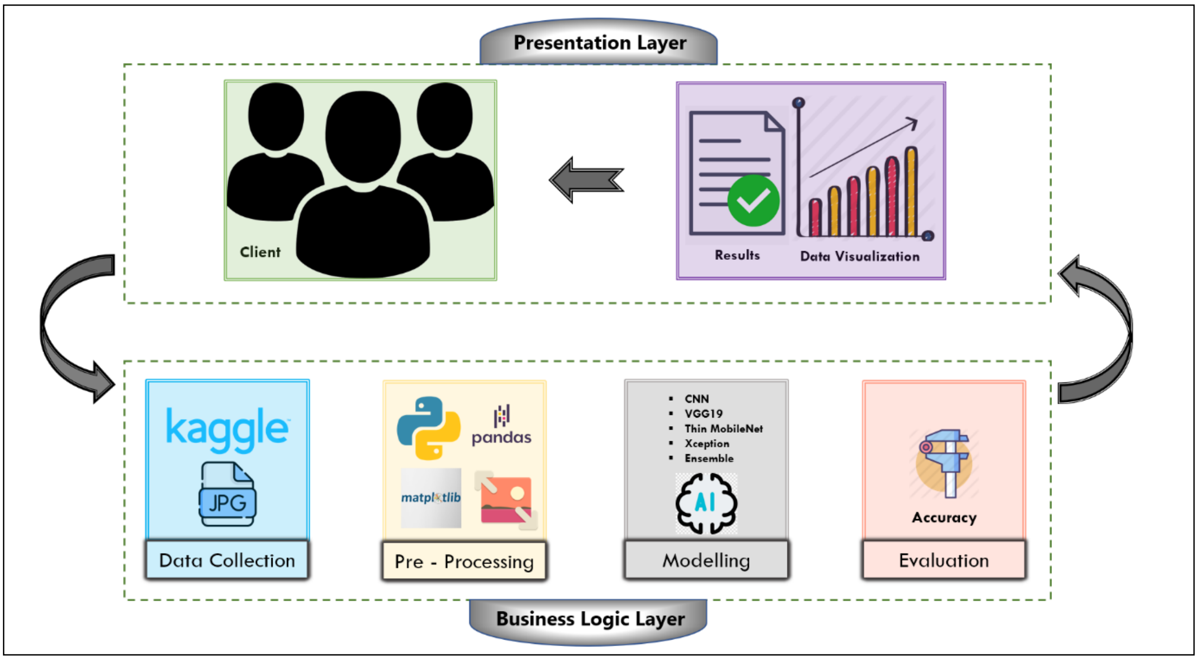

Figure 2 Design Specification

5 Implementation, Evaluation, and Results

5.1 Data Collection

The preferred framework for distracted driver identification can be evaluated using data set from State farm released by State Farm Insurance Company in 2016. This dataset was publicly available on the Kaggle competition. The dataset is composed of 22,424 images of drivers from various parts of the world which are further used for training, validation, and testing. These images were recorded from a camera-taken video installed inside the vehicle which was in the resolution of 640 × 480. The data consists of 10 different classes which include images of normal driving as well as images with different human behaviors that distract them from driving such as drinking/eating, grooming, talking with passengers, reaching behind, operating music or radio, texting, and talking on the phone. Such behaviors alluded to in the dataset can result in a high number of accidents, resulting in serious personal injuries as well as cause a great amount of economic loss.

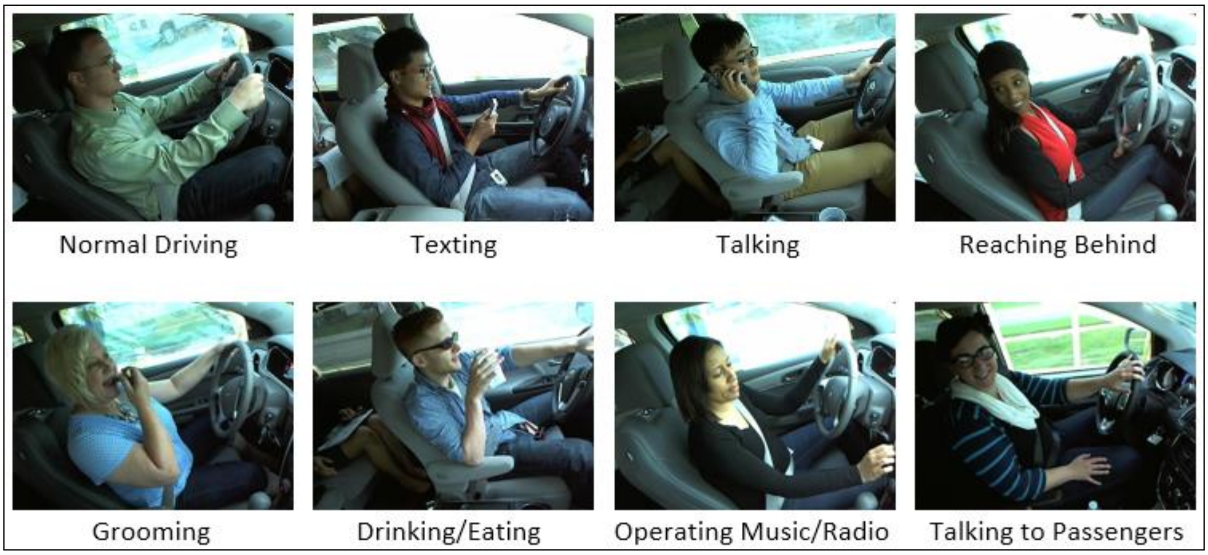

Figure 3 - Types of driver distraction (Source: State Farm, 2016)

5.2 Data Pre-processing

The primary step involved in the pre-processing of data was to convert the data structure into a binary format. The dataset had a total of 10 classes which included nine distracted driving and one safe driving class. This data was converted into two sets of classes which were named as safe_driving and distracted_driving which was done using the python automation feature. Later the data was augmented in order to balance the data as the image count in safe driving was quite less than the image count in distracted driving. Additionally, the images were captured from a video snippet, there were a lot of images that were similar thus the data augmentation could help the predictions smoothen. After the data was balanced and ready to use, it was further split into train, validation, and test folders in a ratio of 70:15:15. The images used in the dataset were in the pixel size of 640×480 which were resized to 224×224 in order to reduce the computational power used by the system and also to reduce the time consumption of each model performed for classification.

5.3 Modelling, Evaluation, and Results

The study here focuses on the comparison of various deep learning models in order to classify the distraction of drivers as safe or distracted. The deep learning algorithms used were Convolutional Neural Networks (CNN), Xception, VGG19, and Thin MobileNet. As the paper is based on the comparison, the hyperparameters passed for each of these models were kept similar to obtain a better analysis. The dataset used for training was split into train, test, and validation in a ratio of 70:15:15. The number of classes in the dataset was 2, i.e. safe and distracted driving. The images used were resized to 224×224. The learning rate used for each of the models was 0.00001. The objective function for each of the models was binary_crossentropy and the optimizer used as Stochastic Gradient Descent (SGD) optimizer. The number of epochs was kept constant to 10 with a batch size of 16.

5.3.1 Convolutional Neural Networks (CNN)

The CNN model built for the classification involved a convolutional layer with 32 feature maps along with a small-sized filter of (3,3). This layer then consisted of a 2x2 sized pooling layer whereby max pooling is used to minimize sampling and retain all significant features in a smaller size image without any information loss that further prevents over-fitting. A dropout layer is used which has a rate of 0.5. Later, the model architecture also included the flattening layer which ensures of mapping the features into a single-dimensional vector to obtain better features which will be then fed as input to CNN. Finally, the Softmax activation function is used to classify the output into safe or distracted driving. The batch size used was 16 in regard to the optimum use of computational power available within the system. As the classification being binary the loss function used was binary_crossentropy along with the SGD optimizer.

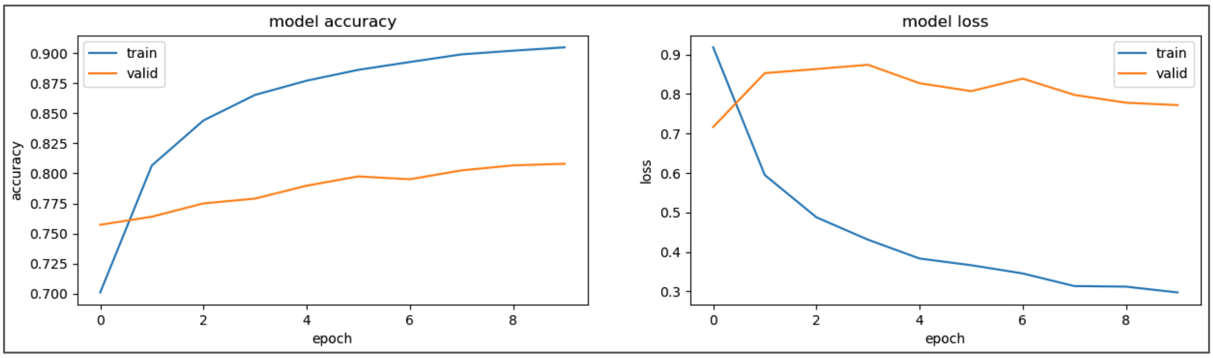

Figure 4 Model Accuracy vs Epoch & Model Loss vs Epoch - CNN

The CNN model was trained on the given dataset with 10 number of epochs which achieved a validation accuracy of about 80.80% while when performed on testing data, it gave an accuracy near to 82%. The results obtained are not that satisfying though as the graph doesn't show optimum training correctness. The model validation doesn't improve over the preceding epochs. However, the good thing about this model is with the increase in the number of epochs the accuracy is seen to be gradually increasing.

Table 3 Accuracy - CNN

Train Accuracy	Validation Accuracy	Test Accuracy
81.95	80.8	81.58

5.3.2 Xception

Xception was another model used which is a modified depthwise separable convolution algorithm. The model was performed along with ImageNet weights where the top layer was not considered. The model comprises of three main sections namely entry flow, middle flow, and exit flow. Similar to CNN, this algorithm was also trained with 10 epochs. The batch size remains the same i.e. 16. Except for the first 15 layers, the later layers were appended to the base model and further trained to keep the layers of the model trainable.

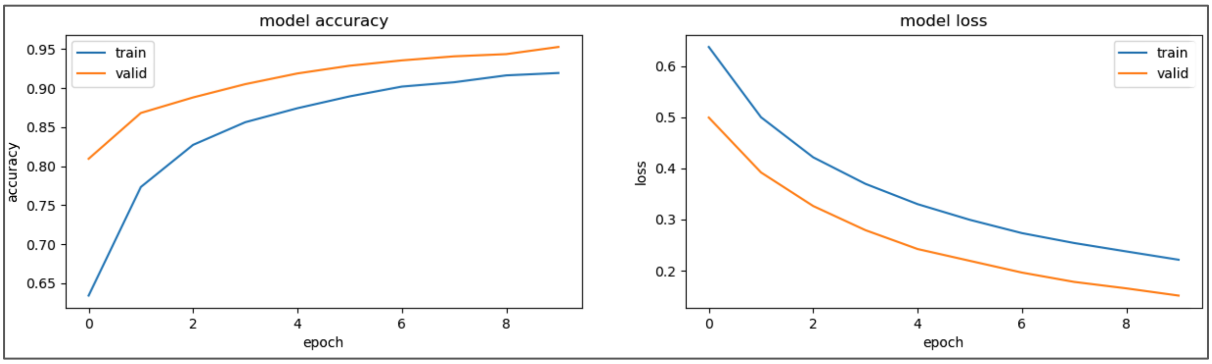

Figure 5 Model Accuracy vs Epoch & Model Loss vs Epoch - Xception

The performance of the Xception model was seen to be good as it can be seen from the plots above where both the model loss after every epoch is getting low as well as the accuracy of the model is gradually increasing. The validation and training loss can be seen moving parallel to each other which represents that model is a good fit for the given classification problem. The table below shows the accuracies obtained by the model i.e. training accuracy is 95.64% while the validation accuracy was 95.26%. The testing accuracy obtained by the model was exactly 95%.

Table 4 Accuracy - Xception

Train Accuracy	Validation Accuracy	Test Accuracy
95.64	95.26	95

5.3.3 VGG19

VGG19 was one of the other models used for classification using the 'ImageNet' weights. The architecture consisted of 19 layers as the name suggests. The pre-trained weights were used feature extraction process where the first 15 layers were not used, and later layers were appended to the base model. The binary_crossentropy loss function was used as the classification being binary. The optimizer being common for every model used, thus the SGD optimizer was used for this algorithm as well.

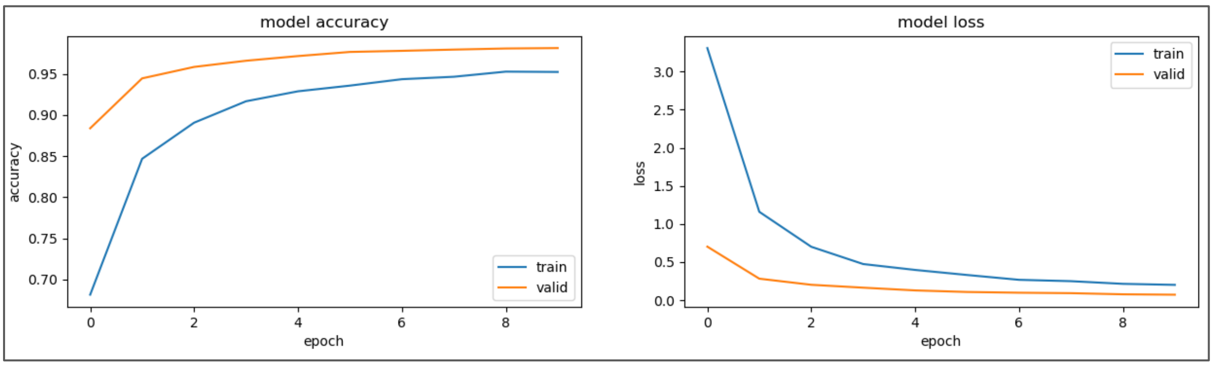

Figure 6 Model Accuracy vs Epoch & Model Loss vs Epoch - VGG19

The results obtained from this algorithm were truly decent achieving an accuracy of 98.14% for validation while 98.6%. The plots from Figure 6 show an increasing accuracy rate as well as decreasing loss which ensures that the model was able to classify the images accurately.

Table 5 Accuracy - VGG19

Train Accuracy	Validation Accuracy	Test Accuracy
98.6	98.14	98.36

5.3.4 Thin MobileNet

Thin MobileNet is an updated version of MobileNet by google. This model was proposed as a new approach to this topic and it performed quite well as compared to all the other transfer learning models used in the past. The model configuration excluded the top layer. The layers after the first 15 layers were kept trainable along with the base model which included flattening, dropout, and ReLu function. The dropout was kept to 0.5. The first layer used in the second version of MobileNet i.e. Thin MobileNet is a 1×1 convolution layer with the ReLu6 function. Following the first layer is the depthwise layer of convolution. And finally comes the third layer which is also a 1×1 convolution layer but this layer doesn't include any non-linearity. The input resolution which was feed to this network was (224,224).

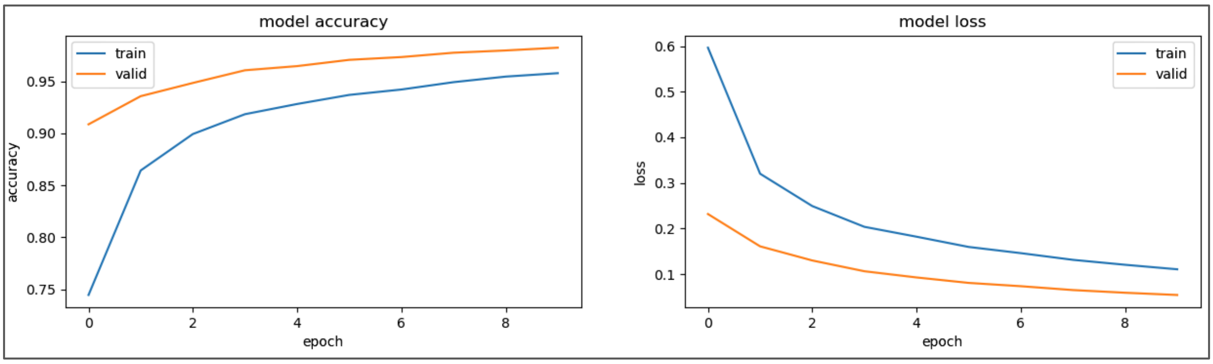

Figure 7 Model Accuracy vs Epoch & Model Loss vs Epoch - Thin MobileNet

As it can be seen from the fig 3, there are very small variations in the values of the training and validation findings which shows that the model was able to achieve a high classification capacity to classify driver as distracted or not distracted. After each epoch, it can be seen that both the validation loss as well as the training loss are reducing. Thus, this shows that the model was able to perform the classification quite robustly. Below are the obtained results of the model which states that the model achieved about 98.26% testing accuracy on images while the validation accuracy achieved was 98.39%.

Table 6 Accuracy - Thin MobileNet

Train Accuracy	Validation Accuracy	Test Accuracy
98.72	98.39	98.26

6 Model Comparison and Discussion

The research done before was mostly performed using the multiclass data. However, a different approach of the study was adopted in this project where the data was changed into a binary form and later the classification models were applied. The distinct models which were considered for the study were chosen on the basis of previous research done. The results obtained from this research study were quite satisfying and predicted the distraction of the driver significantly well. A comparison of different models that were implemented in this study is shown in Figure 8. None of the models achieved results that could be considered really bad enough to call off that particular algorithm. The algorithm with the highest validation accuracy of 98.39% was Thin MobileNet. While the model with the least computational time was CNN i.e. 28.08 min. However, one of the models Xception though obtained a good accuracy of 95.26%, it was the model that took the highest time among all models to complete the training on given 10 epochs. This time was noted and found to be almost 113 min which is almost 2 hours.

One of the main points of discussion in this research study is the use of Thin MobileNet which is nothing but an updated version of MobileNet by Google. Also called as MobileNet V2. It was the algorithm gaining the highest accuracy in terms of both validation and test. In addition to this, the time taken by the model for training was also quite good i.e. 45.10 min, which was best among all the transfer learning models used.

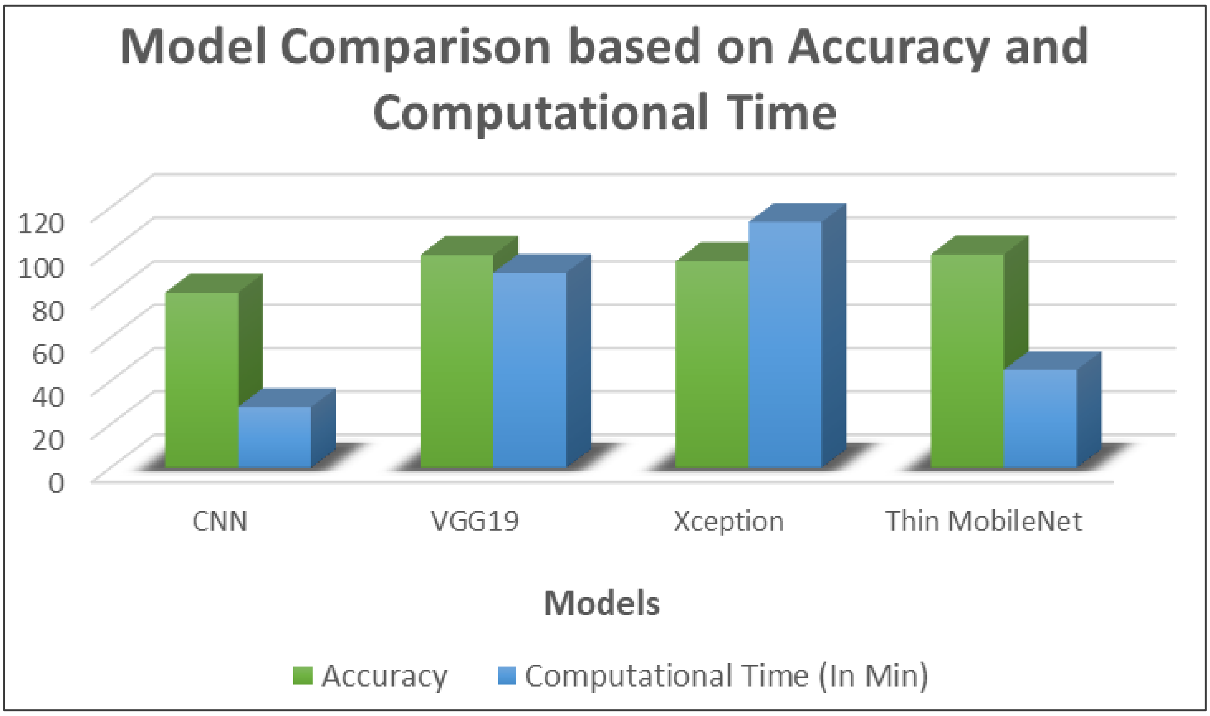

Figure 8 Model Comparison based on Accuracy and Computation Time

7 Conclusion and Future Work

The paper here discussed the use of different deep learning models to identify the distraction of drivers during driving. The chosen algorithms were based on a critical review of various studies done in the past related to driver distraction. A comparison of various deep learning algorithms namely CNN, VGG19, Xception, and Thin MobileNet was done in the study. The objective of this research was to implement a system that could detect a driver's distraction which was done by classifying the driving posture of the driver as safe or distracted. For this purpose, a dataset was picked up from Kaggle which included 10 different classes including safe driving images as well as images of secondary activities carried out during driving. This dataset was then transformed into a binary format using python library shutil (). After the transformation, the data formed was imbalanced which was later balanced by performing augmentation. The data augmentation techniques used for this purpose were horizontal flip, vertical flip, affine transformation, and color augmentation. The algorithms that were used for the study were CNN, Xception, VGG19, and Thin MobileNet. It was observed from the results that except for CNN all the models performed really well. However, Xception performed very badly in terms of computational time as compared to other models. Thin MobileNet outperformed all the used models with an accuracy of 98.39.

As future work, algorithms for the skin segmentation can be implemented in order to accurately identify the human poses as one of the models misclassified a few of the human postures. The implemented system can be used to study behavioral distractions in manufacturing industries where the workers/laborers work on heavy machinery which may cause harm to a human if distracted while using. Also, the designed system can be implemented in the aircraft industry where the airplane crashes take place due to human errors which may involve distractions. Lastly, the dataset could include few images with night-time which could enable further research to apply advanced deep learning models to identify the driver distraction in low light.

Acknowledgment

I am extremely grateful to my supervisor Dr. Manaz Kaleel for the untiring support and feedback throughout the research project. He was highly supportive and skillful. He guided me with every help needed regarding the documentation. Also, I would like to acknowledge the National College of Ireland for mentoring me throughout my master's program. Finally, I would like to thank my parents to help me fulfill my dream of pursuing a Masters in Data Analytics.

References

Assefa, A. and Wenhong, T. (2019) 'Bio-Mechanical Distracted Driver Recognition Based on Stacked Autoencoder and Convolutional Neural Network' in *2019 IEEE 2nd International Conference on Information Communication and Signal Processing (ICICSP).* Weihai, China, 28-30 September 2019, pp. 449-453, IEEE Xplore Digital Library. doi: 10.1109/ICICSP48821.2019.8958598.

Baheti, B., Gajre, S. and Talbar, S. (2018) ‘Detection of Distracted Driver Using Convolutional Neural Network’ in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).* Salt Lake City UT, 18-22 June 2018, pp. 1145- 11456, IEEE Xplore Digital Library. doi: 10.1109/CVPRW.2018.00150.

Chollet, F. (2017) ‘Xception: Deep Learning with Depthwise Separable Convolutions’ in the *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Honolulu, HI, 2017, pp. 1800-1807. doi: 10.1109/CVPR.2017.195.

D. Tran, H. Manh Do, W. Sheng, H. Bai, and G. Chowdhary. (2018) ‘Real-time detection of distracted driving based on deep learning’ in *2018 IET Intelligent Transport Systems*, vol. 12, no. 10, pp. 1210-1219, 12 2018, doi: 10.1049/iet-its.2018.5172.

Da Silva Oliveira, F. and Farias, F. (2018) ‘Comparing transfer learning approaches applied to distracted driver detection’ in *2018 IEEE Latin American Conference on Computational Intelligence (LA-CCI),* pp. 1-6, IEEE Xplore Digital Library. doi: 10.1109/LACCI.2018.8625214.

Dwivedi, K., Biswaranjan, K. and Sethi, A. (2014) ‘Drowsy driver detection using representation learning’ in *2014 IEEE International Advance Computing Conference (IACC),* pp.995-999, IEEE Xplore Digital Library. doi: 10.1109/IAdCC.2014.6779459.

Eraqi, H., Abouelnaga, Y., Saad, M. and Moustafa, M. (2019) ‘Driver Distraction Identification with an Ensemble of Convolutional Neural Networks’ in *2019 Journal of Advanced Transportation,* vol. 2019, pp.1-12. Hindawi Database. doi.org/10.1155/2019/4125865.

K. Yuen, S. Martin, and M. M. Trivedi. (2016) ‘Looking at faces in a vehicle: A deep CNN based approach and evaluation’ in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*. Rio de Janeiro, pp. 649-654, IEEE Xplore Digital Library. doi: 10.1109/ITSC.2016.7795622.

Kim, W., Choi, H. and Jang, B. (2018) ‘Study on Training Convolutional Neural Network to Detect Distraction and Drowsiness’ in *2018 IEEE Region Ten Symposium (Tensymp).* Daejeon, Korea, pp. 260-264, IEEE Xplore Digital Library. doi: 10.1109/TENCONSpring.2018.8692015.

Koesdwiady, Arief & Bedawi, Safaa & Ou, Chaojie & Karray, Fakhri. (2017) ‘End-to-End Deep Learning for Driver Distraction Recognition’, pp. 11-18. doi: 10.1007/978-3-319-59876-5_2.

M. Gjoreski, M. Ž. Gams, M. Luštrek, P. Genc, J. Garbas, and T. Hassan ‘Machine Learning and End-to-End Deep Learning for Monitoring Driver Distractions from Physiological and Visual Signals’ in *2020 IEEE Access*, vol. 8, pp. 70590-70603, IEEE Xplore Digital Library. doi: 10.1109/ACCESS.2020.2986810.

Majdi, M., Ram, S., Gill, J. and Rodriguez, J. (2018) ‘Drive-Net: Convolutional Network for Driver Distraction Detection’ in *2018 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI)*. Las Vegas, NV pp. 1-4, IEEE Xplore Digital Library. doi: 10.1109/SSIAI.2018.8470309.

Ou, Chaojie & Ouali, Chahid & Karray, Fakhri. (2018) ‘Transfer Learning Based Strategy for Improving Driver Distraction Recognition’, pp. 444-452. doi:10.1007/978-3-319-93000-8_50.

Sinha, D. and El-Sharkawy, M. (2019) ‘Thin MobileNet: An Enhanced MobileNet Architecture’ in 2019 *IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON),* New York City, USA, pp. 0280-0285, IEEE Xplore Digital Library. doi: 10.1109/UEMCON47517.2019.8993089.

Wen, L., Li, X., Li, X. and Gao, L. (2019) ‘A New Transfer Learning Based on VGG-19 Network for Fault Diagnosis’ in 2019 *IEEE 23rd International Conference on Computer Supported Cooperative Work in Design (CSCWD),* Porto, Portugal, 2019, pp. 205-209. doi: 10.1109/CSCWD.2019.8791884.

X. Zhou, L. Xie, P. Zhang, and Y. Zhang. (2014) ‘An ensemble of deep neural networks for object tracking’ in *2014 IEEE International Conference on Image Processing (ICIP),* Paris, pp. 843-847. doi: 10.1109/ICIP.2014.7025169.

Xing, Y., Tang, J., Liu, H., Lv, C., Cao, D., Velenis, E. and Wang, F. (2018) ‘End-to-End Driving Activities and Secondary Tasks Recognition Using Deep Convolutional Neural Network and Transfer Learning’ in *2018 IEEE Intelligent Vehicles Symposium (IV),* pp.1626-1631. IEEE Xplore Digital Library. doi: 10.1109/IVS.2018.8500548.

Z. A. Varaich and S. Khalid. (2019) ‘Recognizing Actions of Distracted Drivers using Inception v3 and Xception Convolutional Neural Networks’ in *2019 2nd International Conference on Advancements in Computational Sciences (ICACS).* Lahore, Pakistan, pp. 1-8, IEEE Xplore Digital Library. doi: 10.23919/ICACS.2019.8689131.