Classification of Distracted Driving using Transfer Learning and Deep Neural Network

Deepthi M. Pisharody
Research Scholar, Department of Computer Science
Sullamussalam Science College, Areekode

Binu P. Chacko, PhD
Associate Professor, Department of Computer Science
Prajyoti Niketan College, Pudukad

Mohamed Basheer K.P., PhD
Professor, Department of Computer Science
Amal College of Advanced Studies

ABSTRACT
Distracted driving is a significant contributor to traffic accidents. In order to improve road safety, it is critical to not only detect instances of driver distraction but also to identify the core causes of these distractions. In this study, description about a complete technique to classifying instances of distracted driving that fully incorporates transfer learning technology. Our process is predicated on a precisely produced dataset that has been thoroughly annotated to assure ac- curacy. This dataset is augmented further, and feature extraction is carried out using a wide selection of transfer learning models, including VGG16, Resnet50, Inception, Densenet, and Xception. Following that, the collected features are fed into a DDD classifier which classifies and identifies distraction types. Our experimental results indisputably show that the DNN model, after feature extraction via the Resnet50 transfer learning model, outperforms all other models in the context of distracted driving classification.

General Terms
Pattern recognition, deep neural network

Keywords
Driver Distraction, Deep Learning, Feature Extraction, Transfer Learning models, Driver Distraction Detection (DDD Classifier)

1. INTRODUCTION
Distracted driving may be defined as including in secondary activities like using mobile phones, eating, drinking, taking to co passengers etc. Research by the National Highway Traffic Safety Administration (NHTSA) which is a US based organisation for setting and enforcing standards for road safety and the Virginia Tech Transportation Institute (VTTI), US found that driver distraction, including texting, talking on a cell phone, and other activities, causes 80% of road accidents [1].

In this article driver distraction is defined as secondary activities involved during driving like using cell phones, left and right hands, texting in mobile phones using left and right hands, engaging in hair make up while driving, adjusting car audio system or using dashboard, drinking and eating while driving, turning back and talking and talking to copassengers.

This paper proposes a comparative study of various transfer learning networks on our dataset. Later extracted features from dataset using transfer learning models and applied this feature to a customised DNN.

The further part of this paper includes literature survey, overview of transfer learning models, optimizers and loss functions used, proposed work, experiments and results and conclusion part.

2. LITERATURE REVIEW
Qunfang Xiong et.al [2] outlines a deep learning-based technique for detecting a driver's cell phone use. This approach uses a convolutional neural network-based technique to detect the calling behaviour after first using the Progressive Calibration Networks algorithm to recognise faces and monitor the candidate region. The experimental findings demonstrate that the algorithm's accuracy is 96.56%.

Guofa Li et.al [3] proposed a new discriminant deep learning method to identify distracted driving. They used publicly available dataset like State Farm, 3MDAD and origin for their work. ResNet performed well on 3 datasets.

In our previous work we used a public data set and performed a study on it using various machine learning techniques. SVM gave us a better result. [4]

Uzzol Hossain et.al performed a study on StateFarm dataset using transfer learning techniques. They adopted various transfer learning models like VGG 16 and Resnet50 on data set and performed a comparison on it [5].

Using the Semi-Supervised Extreme Learning Machine technique, Tianchi Liu e.al [6] found driver distraction. Since unlabeled data was used, this method was less expensive. Utilising Laplacian support vector machines and semi-supervised extreme learning machines, two driver states—cognitively distracted and attentive—are recognised. As a result of using unlabeled data, this method has the issue of using data from the same driver in a specific driving scenario to detect attention.

To detect eating/drinking distraction and texting distraction, Atiquz- zaman et. al [7]presented a method using the SVM, LDA, LR, and RF (Random forests) algorithms. Classification accuracy could be improved us- ing RF algorithm. The accuracy for detecting texting distraction is 85.38 percent, and the accuracy for detecting eating distraction is 81.26 percent.

All the above previous studies used publicly available datasets. Here in our study, created a customized dataset. The dataset was created using a mobile phone camera mounted on the dashboard of the vehicle to gather video footage of drivers in various scenarios. These movies were shot with both male and female drivers, using a variety of car models. The videos were recorded by placing mobile phones in front of automobile above the steering. Each of these recorded videos was painstakingly annotated manually. The features extracted using pretrained networks. But after that for classification purpose these features were added into a customised Deep Neural Network. This work is a comparative study of how transfer learning models worked on bench mark dataset like StateFarm data set and how it is working on our dataset.

3. 3.METHODOLOGY

This study tries to find best model for distracted driving classification from video dataset. The dataset contains 100 videos which falls under 10 classes of distraction during driving. Each Video is converted into image frames which is then fed to a preprocessing stage. Channel shifting, Zooming and Rotation operations are performed on image frame to augment images. Features were extracted using several transfer learning models like Resnet50, Inception, Densenet, Xception, and VGG16. Those features were fed into a customised Deep neural network to perform classification tasks. A comparison on accuracy and loss were analysed on each models.

Figure 1 shows the pipeline of the work proposed. The flow of work includes creating of our own dataset, preprocessing the dataset to make it fit for experiment, performing data augmentation for training datset, extracting features by inputting dataset to transfer learning models and training and testing dataset using a customised Deep Neural Network.

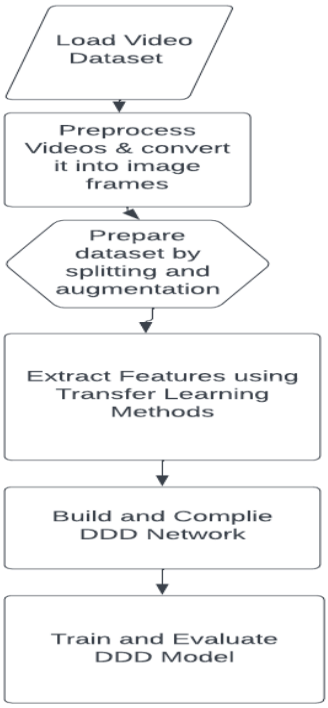

Fig 1: Flowchart of the DDD (Driver Distraction Detection)

3.1 Contributions and Dataset

Created a Video dataset that were shot on male and female drivers using variety of car models.

A well-structured dataset is required for any model to be trained properly. Then compiled a varied dataset containing multiple instances of driver distractions to suit this need. A mobile phone camera mounted on the dashboard of the vehicle to gather video footage of drivers in various scenarios. These movies were shot with both male and female drivers, using a variety of automobile models. The videos were recorded by placing mobile phones in front of automobile above the steering. Each of these recorded videos was painstakingly annotated manually. A picture dataset was methodologically generated from these video recordings by turning the films into individual image frames. Each video was meticulously turned into a 50-frame sequence, and comments were meticulously correlated with each frame. It's worth noting that a balanced dataset throughout the procedure. In the context of our work," balanced" means that each class within the dataset contains roughly an equal amount of data samples, ensuring that no class is overrepresented in comparison to others.

Our dataset has ten separate classes, each of which represents a different driving scenario. Safe driving, making phone calls with the right hand, making phone calls with the left hand, texting on the phone with the left hand, texting on the phone with the right hand, eating or drinking while driving, turning to converse with fellow passengers, engaging in conversations with fellow passengers, using the dashboard or adjusting the radio while driving, and performing makeup or hairdressing activities while driving are the classes in our dataset. Classification was done using a customised Deep Neural Network.

3.2 Preprocessing

The further preprocessing step includes both image scaling and data augmentation. Since deep learning requires large amounts of data and it is not practical to collect huge amount of data, data augmentation plays a crucial role in the process. It aids in expanding the dataset and adding variability to the dataset. In this work data augmentation is applied to training data set by performing channel shifting, rotating and zooming image frames.

Each video is converted to image frames of dimensions 1920*1080. Created nearly 250 frames from a video. It is then resized to 224*244 sized images. In the next step those RGB images are converted to BGR frames.

3.3 Data Augmentation

Image classification is one area of computer vision where machines are now more accurate at classification than humans. Nonetheless, the main drawback of image classification is the massive volume of data required. If you train your model on a large number of images, it is more likely to do classification tasks with high accuracy. One potential solution to the aforementioned problem is data augmentation. By employing this technique, more data can be added to the model's training set without adding any new data. The most often used techniques for enlarging the data size over the images include flipping, rotating, padding, and cropping. But occasionally, we may not have enough data for model training A tensor of the dimensions (height, width, and colour channels) is typically used to encode digital image data. An additional tactic that is quite doable is performing augmentations in the colour channels space. Isolating a single colour channel, such as R, G, or B, is one very basic method of colour enhancement. A matrix can be isolated and two zero matrices from the other colour channels added to it to quickly transform an image into its representation in one colour channel. Further- more, the RGB values can be readily changed using basic matrix operations to change the brightness of the image.

By adjusting the [R,G,B] channels of the input image, channel shift modifies the colour saturation level (e.g. light Red/dark Red) of pixels. To let the model learn color-based features regardless of the saturation value of the dataset, channel shift is utilised to introduce color augmentation. The image is rotated to the right or left along an axis between 1 and 359 to perform rotation augmentations. The rotation degree parameter has a significant im- pact on the safety of rotation augmentations. Figure 2 represents data augmentation phases and sample images used in this work.

Fig 2: Data Augmentation Phases and Sample images

3.4 Feature Extraction

Feature engineering entails changing the data's present shape to one that better suits the purpose of the problem being solved. It is difficult to manually engineer features, especially when there is a lot of data, hence automatic feature extraction via transfer learning is the preferable method. Images are pre-processed in order to extract pertinent features using the pre-trained model, or a component of the model. In this work transfer learning models like ResNet50, VGG16, MobileNet V2, InceptionV3, Xception, DenseNet121 are used for extracting features from the dataset.

The model's capabilities are strengthened and learning is accelerated through the application of transfer learning. Through the machine learning process known as transfer learning, a model developed for one task is utilised as the foundation for another. Transfer learning methods enhance performance and save time. When a model cannot be fully trained from start due to a lack of data in the dataset, transfer learning is usually employed. A few layers from a pre-trained model are used in transfer learning models. In order to avoid losing any of the data they hold for future training sessions, freeze them. Place a few new, trainable layers on top of the frozen layers. They will learn how to forecast utilising the prior features on a new dataset. Get the new layers trained on your dataset to extract features from it.

3.5 DDD Classifier

When loading a particular model, the fully-connected output layers of the model that are used to create predictions are not loaded, enabling the addition and training of a new output layer. The features are then taken from the transfer learning model, examined for the removal of undesirable features, and lastly used to train an artificial neural network (ANN).

In this experiment a sequential DNN model is created who has input layer and number of neurons in the input layer matches the size of feature extracted array. A set of dense layers where added whose activation function is Relu. The output layer has softmax as activation function. Adam optimiser used for compiling this model and the loss function used is categorical entropy function.

ReLU is the default activation function and the most often used activation function in neural networks, particularly CNNs.

$$F(x) = max(0, x) \qquad (1)$$

When the function is given a negative value as input, it returns 0, but when it is given a positive value, it returns x. The output therefore has a range from 0 to infinity.

A vector of K real values can be transformed into a vector of K real values that total to 1 by using the softmax function. The softmax turns the input values, which can be positive, negative, zero, or higher than one, into values between 0 and 1, allowing them to be understood as probabilities. The softmax converts little or negative inputs into small probabilities, and big or positive inputs into large probabilities, although it will always fall between 0 and 1.

The multi-class logistic regression or softargmax function is another name for the softmax function. This is because the softmax is a logistic regression generalisation that may be utilised for multi class classification.

To update network weights during training, the Adam optimisation ap- proach is a further refinement of stochastic gradient descent. Adam optimizer modifies the learning rate for each network weight separately, unlike SGD training, which maintains a single learning rate. Adam uses both the first and second moments of the gradients to adjust learning rates rather than just the first moment (mean) as it does in RMS Prop.

When there are two or more output labels in a multi-class classification model, categorical cross entropy is used as the loss function. One-hot cate- gory encoding value in the form of 0s and 1 are allocated to the output label. If the output label is an integer, it is changed to a categorical encoding.

Figure 3 shows the architecture of DNN used for training and testing and Figure 4 shows the diagrammatic representation of feature extraction and training.

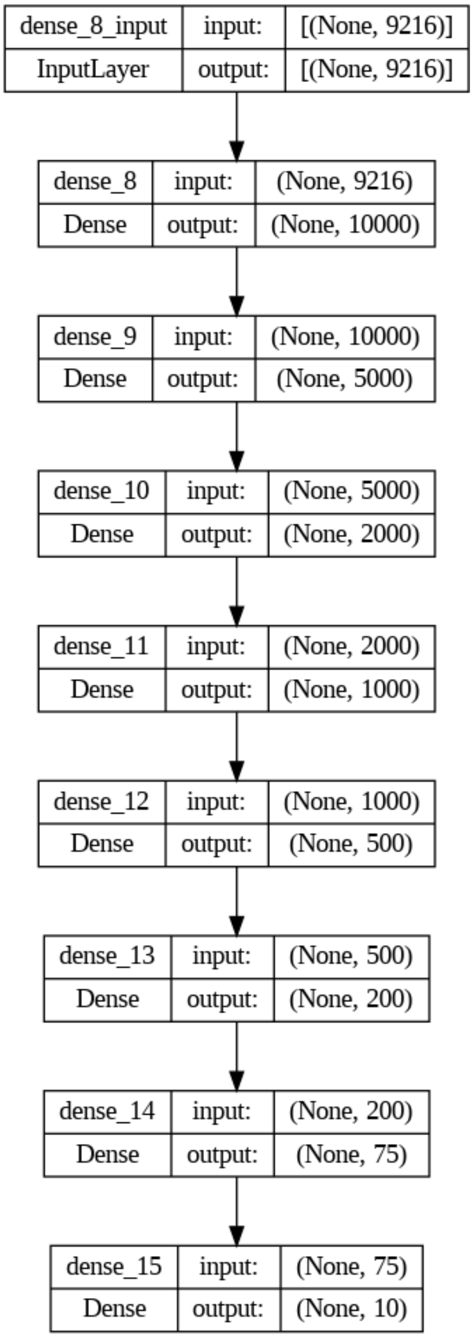

Fig 3: Architecture of DDD Classifier

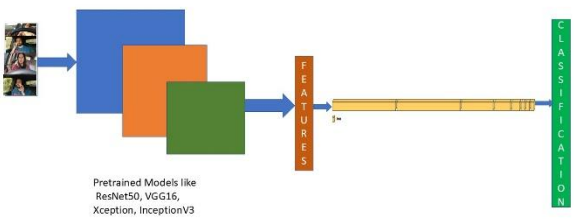

Fig 4: Flow of Feature Extraction and Classification

4. RESULTS AND DISCUSSIONS

The experiment was conducted on Colab Pro. It can give you access to a V100 or A100 Nvidia GPU, depending on availability. Python coding was used to conduct experiments and plot graphs. A balanced dataset consisting of 450 images in each 10 classes were used for the conduct of this experiment. Balanced means that each class has roughly equal shares of the data over the whole dataset and that no class has significantly more samples than any other. For training and testing split dataset into 4:1 ratio, that 80% is used for training and 20% for testing. Various graphs like Accuracy- Validation Accuracy, Loss- Validation Loss were plotted for each category of experiments. Confusion matrix also was plotted for each.

The graphs and confusion matrix after extracting features using Resnet is plotted as figure 5 to 7.

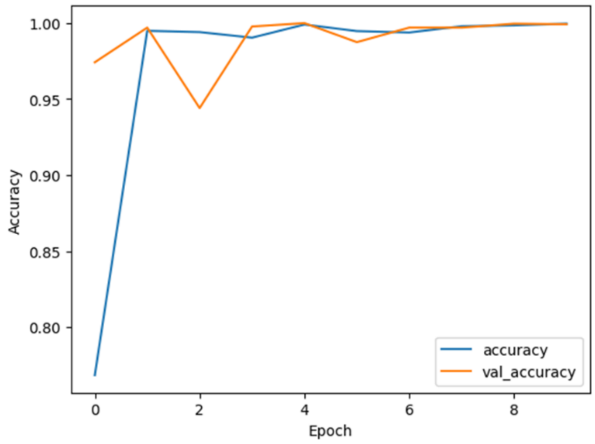

Fig: 5 Graph Plotting Relation between Accuracy and Validation Accuracy after Extracting Features using ResNet50

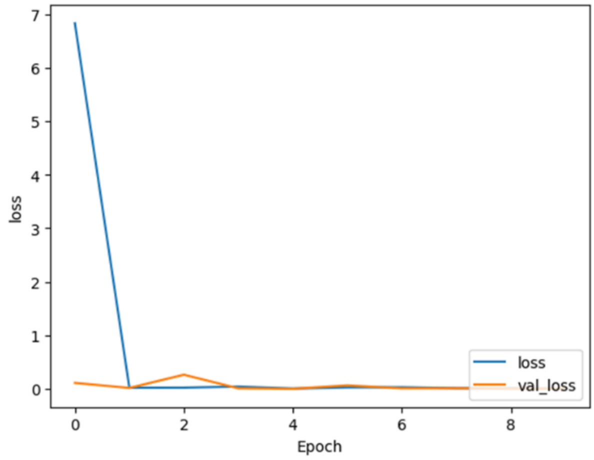

Fig 6: Graph Plotting Relation between Loss and Validation Loss after Extracting Features using ResNet50

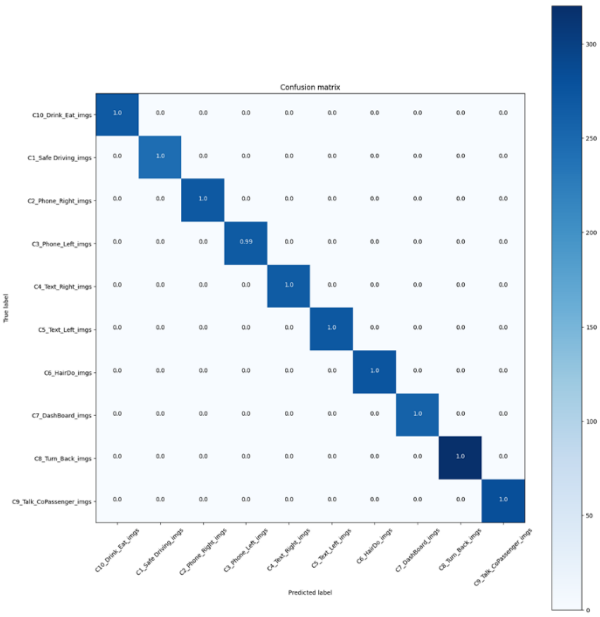

Fig 7: Confusion Matrix after extracting features using ResNet50

Table 1 shows comparative study of testing accuracy obtained after extracting features using various transfer learning techniques. Since a preprocessed balanced dataset was used in this work, most modes gave a better results.

The Figure 20 plots testing accuracy of using various models. From results it is clear that when trained using DNN after extracting features using ResNet50 transfer learning mechanism provides highest accuracy of 99.99%. While observing Confusion Matrix, it is noted that the classification that is misclassified in few cases using phone using left hand and driving when features were extracted using ResNet50.

When data set was trained using DNN after extracting features using ResNet50 transfer learning mechanism provides highest accuracy of 99.99%. Confusion matrix was generated for each model.

Table 1: Comparative study of testing accuracy obtained after extracting features using various transfer learning techniques.

Model	Testing Accuracy in DDD dataset	Testing accuracy in StateFarm Dataset
ResNet50	99.99	99.31
VGG16	99.96	98.27
Xception	99.22	98.99
Inception V3	96.21	96.11
DenseNet121	97.28	97.06

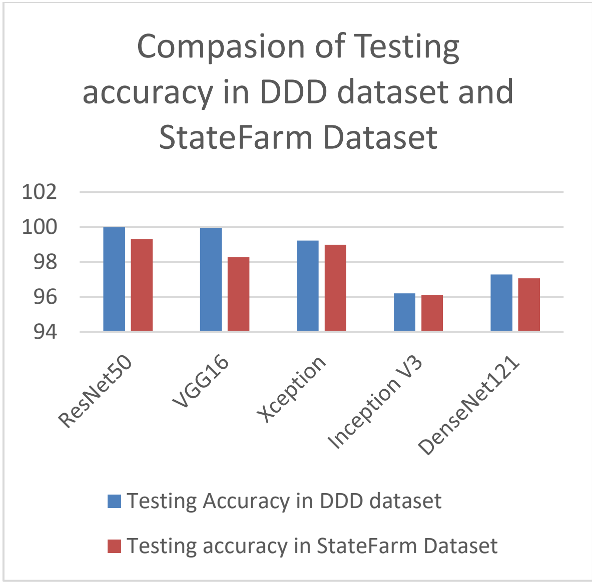

Fig 8: Testing accuracy of various Transfer Learning Models For Feature Extraction in various datasets

This technique is innovative since it uses a comprehensive approach to classify video data sets, which has been confirmed using preliminary subset experiments. The current study offers important insights and approaches that will contribute to the final benchmark dataset, even if the dataset is still being created. After the study is over, the benchmark dataset will be made available to the public, guaranteeing that the groundwork for more research on distracted driving analysis will be laid.

5. CONCLUSIONS

Distracted driving is a significant contributor to traffic accidents. In order to improve road safety, it is critical to not only detect instances of driver distraction but also to identify the core causes of these distractions. In this study, we describe a complete technique to classifying instances of distracted driving that fully incorporates transfer learning technology. Our process is predicated on a precisely produced dataset that has been thoroughly annotated to assure accuracy. This dataset is augmented further, and feature extraction is carried out using a wide selection of transfer learning models, including VGG16, Resnet50, Inception, Densenet, and Xception. Following that, the collected features are fed into a customised Deep Neural Network (DNN). Our experimental results indisputably show that the DNN model beats all other models in the context of distracted driving categorization after feature extraction using the Resnet50 transfer learning model. It classified distracted driving classes with high accuracy of 99.9%. When mobile phone was used using left hand it was misclassified when features were extracted using ResNet50 and trained. The work can be used for real time implementation of distracted driving classification and identification. This work is a comparative study to know how deep learning models works in our dataset and ad bench mark dataset. Since in this work preprocessing dataset was done well models worked well in our dataset that StateFarm dataset. The work can be extended by capturing driving videos from various orientation and trying for a better classification model.

6. ACKNOWLEDGMENTS

We sincerely acknowledge the managements of Sullamusalam Science College, Areekode unstinting support and infrastructural facilities provided. We also acknowledge all drivers who were willing to capture videos.

7. REFERENCES

[1] Distracted driving. [Internet]. https://seriousaccidents.com/legal-advice/top-causes-of-car-accidents/driver-distractions/

[2] Q. Xiong, J. Lin, W. Yue, S. Liu, Y. Liu, and C. Ding, A deep learning approach to driver distraction detection of using mobile phone, IEEE Vehicle Power and Propulsion Conference (VPPC), pp. 1– 5, IEEE, 2019. DOI:10.1109/VPPC46532.2019.8952474

[3] G. Li, Q. Liu, and Z. Guo, Driver distraction detection using advanced deep learning technologies based on images, IEEE Journal of Radio Frequency Identification, vol. 6, pp. 825–831, 2022. Doi: 10.1109/JRFID.2022.3209237

[4] Pisharody, Deepthi M., Binu P. Chacko, and KP Mohamed Basheer. Driver distraction detection using machine learning techniques. Materials Today: Proceedings 58 (2022): 251-255.doi: https://doi.org/10.1016/j.matpr.2022.02.108

[5] Hossain, M. U., Rahman, M. A., Islam, M. M., Akhter, A., Uddin, M. A., and Paul, B. K. (2022). Automatic driver distraction detection using deep convolutional neural networks. Intelligent Systems with Applications, 14, 200075. https://doi.org/10.1016/j.iswa.2022.200075

[6] J. Tang, M. Sharma, and R. Zhang, "Explaining the effect of data augmentation on image classification tasks," 2020.

[7] Q. Zheng, M. Yang, X. Tian, N. Jiang, D. Wang, *et al.*, "A full stage data augmentation method in deep convolutional neural network for natural image classification," *Discrete Dynamics in Nature and Society*, vol. 2020, 2020.