State Farm Distracted Driver Detection

Akanksha Shrimal
MT20055

Prabal Jain
MT20115

Akhil Mahajan
MT20107

Abstract

Driving a car is a complex task, and it requires complete attention. Distracted driving is any activity that takes away the driver's attention from the road. Approximately 1.35 million people die each year as a result of road traffic crashes.
Road traffic crashes cost most countries 3% of their gross domestic product So, our aim/goal in this project is to detect if the car driver is driving safe or performing any activity that might result in an accident or any harm to others, by using various Machine Learning Models to classify the provided images into different categories of Distraction.
Furthermore, we can extend this work into comparing various Machine Learning Models to determine the accuracy based on respective models.

1 Importance of Project

Many states now have laws against texting, talking on a cell phone, and other distractions while driving. We believe that by applying Machine Learning algorithms, we can detect the risk of accident by classifying the driver images in one of the distracted classes and hence prevent accidents caused by distracted driving. If this information can be known at real time then a lot of accidents can be reduced with proper implementation.

2 Literature Survey

This section summarises review of some of the relevant and significant work from literature for distracted driver detection.

In paper [1], images are resized for training and testing purposes and traditional Machine Learning algorithms are applied namely Linear SVM,Softmax,Naive Bayes,Decision Tree,Two-Layer Neural Network. The paper[1] is further extended to compare the performance of traditional machine learning algorithms and advance CNN based techniques such as ResNet and VGG.

In paper[2] images are resized and feature extraction techniques are applied like Pixel,HOG , Sobel,Clustering. After this, feature extraction for dimensionality reduction PCA is applied. The classification is performed using algorithms like SVM , Decision Tree, Random Forest , 2-Layered Neural network and CNN. A comparative analysis is done for each model with different feature extraction techniques and accuracy is obtained individually using each feature extraction technique. For CNN a grid search is performed for individual parameters to obtain the best model. At the end, transfer learning algorithm VGG-19 is used. Paper[2] is further extended to obtain the results using modern Deep Learning Network Architectures and combined feature vectors are used for Image Representation.

Paper[3] preprocesses the images as skin-segmented images, face images, hands images, and "face+hands" images. These along with the raw images are trained using a weighted ensemble of CNN.

3 Dataset

The dataset used is State Farm Distracted Driver Detection taken from *https://www.kaggle.com/c/state-farm-distracted-driver-detection/data.* The dataset contains 22424 driver images in total and has 10 classes. The 10 classes are Safe driving , Texting(right hand) ,Talking on the phone (right hand),Texting (left hand) ,Talking on the phone (left hand), Operating the radio ,Drinking , Reaching behind , Hair and makeup , Talking to passenger(s).Each image belongs to one of the classes above and are taken in a car with a driver doing something in the car. The images are coloured and have 640*480 pixels each as shown below. For the training and testing

purposes the images are resized to 64*64 coloured images.
Stratified splitting is used to split the dataset into 80:10 Training-Testing ratio. The training dataset is further split into 90:10 Training-Validation set.

Figure 1: Data Visualization

4 Data Analysis

Images are resized to 64*64 coloured images for training and testing purposes.Following feature extraction techniques are applied LBP,HOG,color Histograms,KAZE,SURF.The result of feature extraction can be visualized in figure 2. Normalization is performed over the extracted features.
Dimensionality reduction techniques like PCA and LDA are used to reduce the dimensions and avoid ‘Curse of Dimensionality’. For deciding the n_components of PCA, variance-components graphs are used(Figure 3).
All the features are stacked together to get complete image representation and ML algorithms are applied to obtain the accuracy.

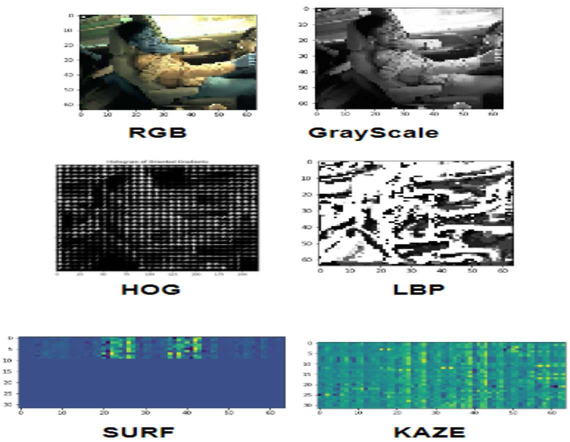


Figure 2: Feature Extraction Techniques

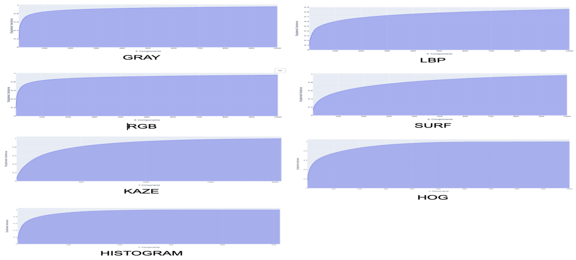


Figure 3: PCA variance and n_components graph

5 Methods and Results

5.1 Traditional ML Models

The following traditional ML algorithms are used along with feature extraction and dimensionality reduction.

- **Decision Tree:**
  Decision Tree is an important non-parametric method for image mining, This method uses several simple decision rules of image to provide a path for image classification.
- **Support Vector Machine:**
  SVMs are supervised learning models with associated learning algorithms that analyze data for classification and regression analysis.
- **KNN:**
  The KNN algorithm assumes that similar things exist in close proximity. In other words, similar things are near to each other KNN captures the idea of similarity (sometimes called distance, proximity, or closeness)

5.2 Ensembling Methods

- **XGBoost**
  XGBoost is a decision-tree-based ensemble Machine Learning algorithm that uses a gradient boosting framework. In prediction problems involving unstructured data (images, text, etc.) artificial neural networks tend to outperform all other algorithms or frameworks.

- **Bagging**
  Bootstrap aggregating, also called bagging (from bootstrap aggregating), is a machine learning ensemble meta-algorithm designed to improve the stability and accuracy of machine learning algorithms used in statistical classification and regression. It also reduces variance and helps to avoid overfitting.

- **ADABoost**
  AdaBoost algorithm, short for Adaptive Boosting, is a Boosting technique that is used as an Ensemble Method in Machine Learning. It is called Adaptive Boosting as the weights are re-assigned to each instance, with higher weights to incorrectly classified instances.

Model	Optimal Hyperparameters
DT	criterion = 'entropy', max-depth = 20
SVM	C=10 and kernel='rbf'
KNN	n-neighbours = 5
XGB	max-depth = 6, eta = 0.5
Bagging	n-estimators=40
Adaboost	n-estimators=200

Table 1: Hyperparmeter Tuning

Model	Precision	Recall	F1 Score	Acc
DT	0.8221	0.8213	0.8214	0.822
SVM	0.9973	0.9973	0.9973	0.997
KNN	0.9872	0.9870	0.9870	0.987
XGB	0.9856	0.9849	0.9852	0.985
Bagging	0.7927	0.7848	0.7861	0.789
Adaboost	0.7197	0.6957	0.7010	0.693

Table 2: PCA

Model	Precision	Recall	F1 Score	Acc
DT	0.9753	0.9752	0.9751	0.974
SVM	0.9881	0.9876	0.9876	0.987
KNN	0.9922	0.9924	0.9923	0.992
XGB	0.9912	0.9912	0.9912	0.991
Bagging	0.9825	0.9826	0.9825	0.982
Adaboost	0.5160	0.5785	0.5191	0.574

Table 3: LDA

Model	Precision	Recall	F1 Score	Acc
DT	0.9638	0.9639	0.9638	0.964
SVM	0.9799	0.9791	0.9795	0.979
KNN	0.9806	0.9793	0.9799	0.979
XGB	0.9757	0.9756	0.9756	0.976
Bagging	0.9720	0.9710	0.9714	0.971
Adaboost	0.6880	0.6634	0.6304	0.658

Table 4: LDA on PCA

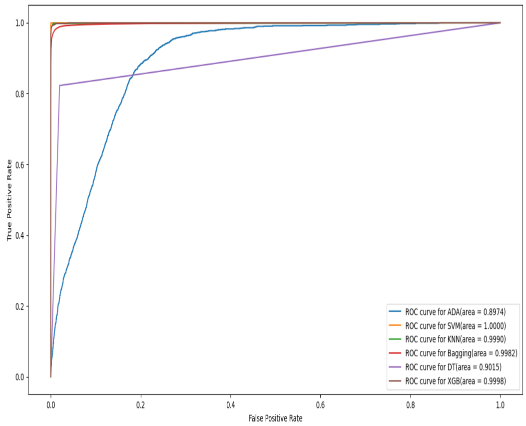


Figure 4: ROC curve for PCA

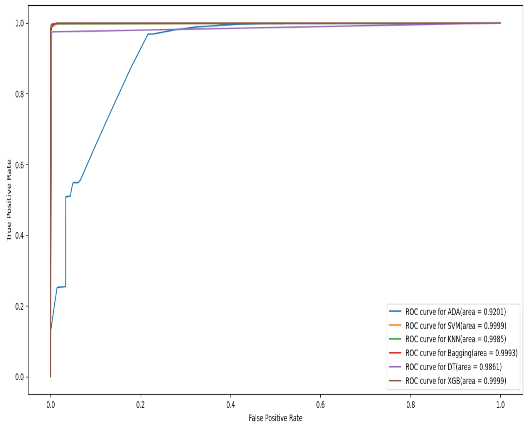


Figure 5: ROC curve for LDA

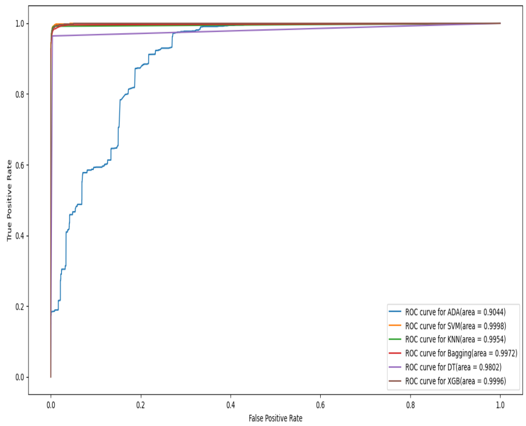


Figure 6: ROC curve for LDA on PCA

5.3 Transfer Learning

Transfer learning is a technique in deep learning where one model that is trained on a task is repurposed to fit another task.

There are three general strategies to retrain the pretrained models using the State Farm dataset of distracted drivers. The first strategy is to retrain only the last classifier layer of the pretrained model. This is the most straight forward strategy using the least amount of time and computational power in exchange for model accuracy. The second strategy is to retrain the last few layers of the model including the classifier layer. This strategy may result in improved accuracy as well as more intensive computational power and takes longer time to run. The last strategy is to retrain the entire model from scratch. This is the least desirable method because it takes the most time and computational cost. Here we evaluated ResNet-101 over Strategy 1 (S1) and Strategy 2 (S2) and evaluated the results.

5.4 Convolutional Neural Network

Convolutional Neural Network (CNN) has proven to be very effective in areas such as image recognition and classification. The advantage of using CNN is that it can automatically learn complex feature utilizing massive simple neurons and backpropagation. CNN has many types of layers. Layers of a CNN transforms one volume to another. The following subsections describe some common types of layers of a CNN. .

- **5.4.1 Input Layer** The input layer holds the pixel values of the input image. Our project uses full color images, so the input layer is 64 × 64 × 3 (for R, G, B channels).
- **5.4.2 Convolutional(CONV) Layer**

  The CONV layer's parameters are a set of learnable filters of small dimensions (e.g. 5 × 5 × 3). The filter convolves with the input volume (across width and height in 2D) to select small areas (e.g. 5 × 5) and use these small local areas to compute dot products with weights/parameters. Each filter corresponds to a slice or a depth of one in the output volume (i.e. the depth of the output volume of a CONV layer is determined by the number of filters).
- **5.4.3 Rectified Linear Units (ReLU) Layer** The ReLU layer applies an element-wise nonlinear activation function to increase nonlinearity in the model.

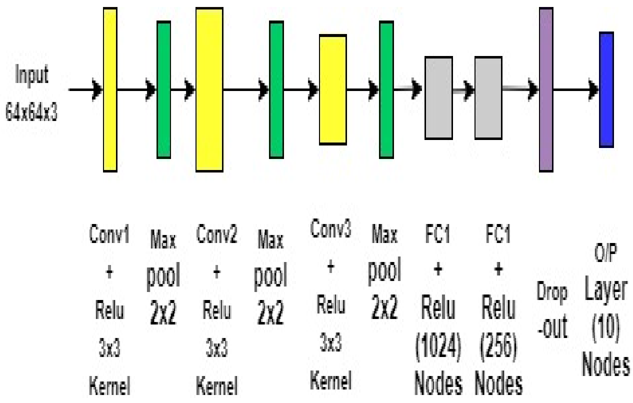


Figure 7: CNN Architecture

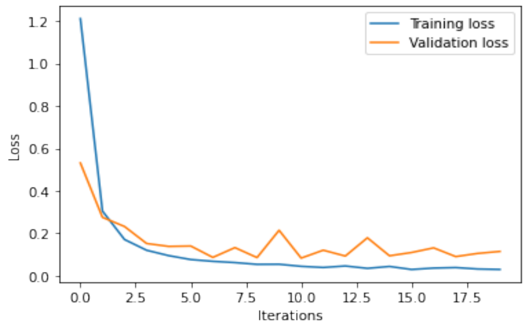


Figure 8: Training and Validation loss for CNN

- **5.4.4 Pooling Layer** A pooling layer reduces the 2D dimensions of the input volume (leaving the depth unchanged) to prevent the model from overfitting and getting too large (too many weights) to compute. It is done independently for each depth slice of the input volume by applying a small filter (e.g. 2× 2)
- **5.4.5 Dropout Layer** Adding dropout layers is a regularization method to prevent overfitting. A dropout layer randomly sets some unit activations to zero and thus removes some feature detectors by dropping out some activations. Some of the high variance due to this is removed.
- **5.4.6 Fully-connected (FC) Layer** Fully-connected layers, as the name suggests, is like ordinary NN where each neuron is connected to all the outputs from the previous layer. The last FC layer computes probabilities for each class.

Figure 7 illustrates the architecture of the CNN used in the classification process. The CNN architecture involves multiple layers of operations performed on the input image of size (64,64,3)

5.5 ResNet-101

ResNet-101 is a convolutional neural network that is 101 layers deep. The core idea of ResNet is introducing a so-called “identity shortcut connection” that skips one or more layers as the result difficulty in training a deeper network is resolved with “shortcut connections”.
Two **stratergies** used with ResNet-101. Figure shows that ResNet-101 performs better with Stratery-2(S2)

- **Stratery-1(S1)** : Retrain only the last classifier layer of the pretrained model and freeze all other parameters.
- **Stratery-2(S2)** : Retrain the last few layers of the model including the classifier layer.

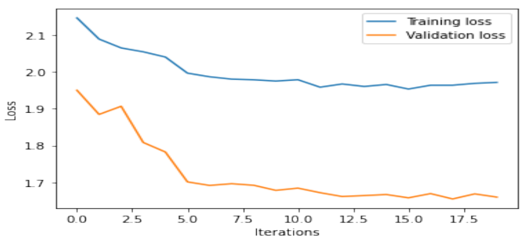


Figure 9: Training and Validation loss for ResNet-101(S1)

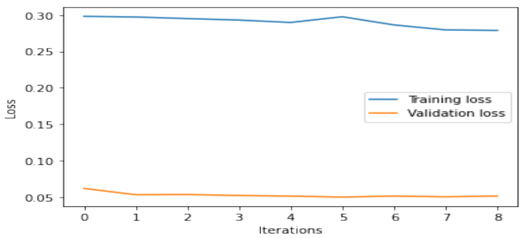


Figure 10: Training and Validation loss for ResNet-101(S2)

Figure 11: ResNet Prediction Visualisation

Model	Precision	Recall	F1 Score	Acc
CNN	0.98	0.98	0.98	0.98
ResNet-101(S1)	0.50	0.46	0.42	0.47
ResNet-101(S2)	0.99	0.99	0.99	0.99

Table 5: Deep Learning models

6 Analysis - Class Activation Mappings

Class activation maps (CAMs) is a way to highlight the regions within an image that a CNN uses to make a classification decision for that particular image. The output of a CAM is a heat map projected visually on top of the image being examined by the neural network. The areas of more intensity represent the parts of the image the neural network is focused on.

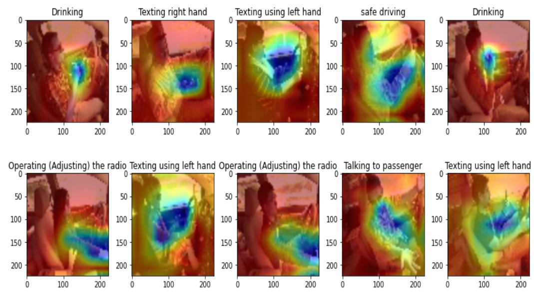
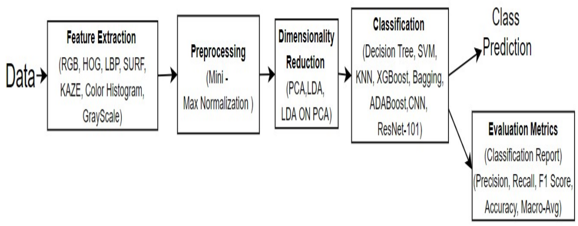


Figure 12: Class Activation Map for ResNet-101(S2)

7 Inferences and Observations

- PCA gave better results than LDA and LDA over PCA.
- Combination of features from various feature extractions gives better accuracy than taking individual features.
- Adaboost performed poorly for classification on this dataset as boosting techniques perform better on high variance datasets which is contrary to our data set which has low variance.
- After plotting the ROC curves for various traditional models, best auc-roc score was found for SVM.
- ResNet-101 S2 performed much better than ResNet-101 S1 which is also higher than the accuracy obtained using CNN.
- Data Augmentation and Data Extraction techniques help the classifier to perform better.

8 Architecture Implemented



9 Link for codes and the data which are uploaded on the Google Drive

https://drive.google.com/drive/u/1/folders/1HsxvwrOsLgszbYZue58CfCqAojGLnw_J

10 Appendix - A short explanation of project-relevant terms

- **HOG** - The histogram of oriented gradients (HOG) is a feature descriptor used in computer vision and image processing for the purpose of object detection. The technique counts occurrences of gradient orientation in localized portions of an image.
- **LBP** - Local Binary Pattern (LBP) is a simple yet very efficient texture operator which labels the pixels of an image by thresholding the neighborhood of each pixel and considers the result as a binary number.
- **SURF** - The SURF method (Speeded Up Robust Features) is a fast and robust algorithm for local, similarity invariant representation and comparison of images. The main interest of the SURF approach lies in its fast computation of operators using box filters, thus enabling real-time applications such as tracking and object recognition.
- **Color Histogram** - A color histogram is a representation of the distribution of colors in an image. For digital images, a color histogram represents the number of pixels that have colors in each of a fixed list of color ranges, that span the image's color space.
- **KAZE** - KAZE is a 2D feature detection and description method that operates completely in a nonlinear scale space.
- **Mini-Max Normalization** - In this technique of data normalization, linear transformation is performed on the original data. Minimum and maximum value from data is fetched and each value is replaced according to the following formula.
- **ROC curve** - An ROC curve (receiver operating characteristic curve) is a graph showing the performance of a classification model at all classification thresholds. This curve plots two parameters: True Positive Rate and False Positive Rate.
- **PCA** - Principal component analysis (PCA) is a technique for reducing the dimensionality of datasets, increasing interpretability but at the same time minimizing information loss.
- **LDA** - Linear discriminant analysis (LDA) is a generalization of Fisher's linear discriminant, a method used in statistics and other fields, to find a linear combination of features that characterizes or separates two or more classes of objects or events.

11 References

[1] : http://cs229.stanford.edu/proj2019spr/report/24.pdf
[2] : https://github.com/Raj1036/ML_Distracted_Driver_Detection/blob/master/CS539_DistractedDriverDetetion_FinalProject.pdf
[3] : H. Eraqi, Y. Abouelnaga, M. Saad, and M. Moustafa, "Driver Distraction Identification with an Ensemble of Convolutional Neural Networks", Journal of Advanced Transportation 2019, 1-12; doi: 10.1155/2019/4125865
[4] : https://snappishproductions.com/blog/2018/01/03/class-activation-mapping-in-pytorch.html
[5] : https://www.ijrte.org/wp-content/uploads/papers/v8i4/D5131118419.pdf
[6] : https://medium.com/@sam.bell_43711/distracted-driver-detection-using-deep-learning-ecc72
[7] : https://kharshit.github.io/blog/2018/09/07/skip-connections-and-residual-blocks
[8] : http://cs229.stanford.edu/proj2016/report/SamCenLuo-ClassificationOfDriverDistraction-report.pdf