Recognition of Unsafe Driving Posture Using Machine Learning

Bachelor of Technology

In

Information Technology

Submitted by

Harsh Meena (IIT2019005)
Hritik Sharma (IIT2019020)
Shubham Netam (IIT2019028)
Karan Kumar Sah (IIT2019046)
Ravikant Vaishnav (IIT2019048)

Under the supervision of

Dr. Navjot Singh

Assistant Professor

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY

ALLAHABAD - 211015, (UP) INDIA

DECLARATION

We hereby declare that the work presented in this thesis report entitled "**Recognition of Unsafe Driving Posture Using Machine Learning** " submitted towards the fulfillment of the 6th Semester Project Report (2022) of **B.Tech** in **Information Technology** at the INDIAN INSTITUTE OF INFORMATION TECHNOLOGY ALLAHABAD, U.P., INDIA is an authenticated record of our original work carried out from Jan 2022 to May 2022 under the supervision of **Dr. Navjot Singh** Assistant Professor, Department of Information technology, IIIT-Allahabad. The thesis is being accomplished in full compliance with the requirements and constraints of the prescribed curriculum.

Proper citation for all those resources which are not our original contribution (websites, research papers, ideas, graphics, experiments) has been delivered to the respective original authors or sources. We affirm that no portion of our work is plagiarised. In the event of a complaint of plagiarism, we shall be fully responsible.

Place: Allahabad, Prayagraj
Date : 06-05-2022

Harsh Meena (IIT2019005)
Hritik Sharma (IIT2019020)
Shubham Netam (IIT2019028)
Karan Kumar Sah (IIT2019046)
Ravikant Vaishnav (IIT2019048)

CERTIFICATE

It is certified that the work contained in the project titled **"Recognition of Unsafe Driving Posture using Machine learning"** has been carried out under my supervision and that this work has not been submitted elsewhere.

Signature of Supervisor

Date: 06-05-2022

Dr. Navjot Singh

Assistant Professor

ACKNOWLEDGEMENT

We would like to offer our thanks to the leniency of Almighty, that invigorated us with enough strength to finish this project work. We are extremely grateful to our Instructors **Dr. Navjot Singh and Dr Satish Kumar Singh ,**Assistant Professor, Department of Information Technology, IIIT-Allahabad to direct us all through the undertaking work and always inspired us to work hard. The consistent discussions and invaluable remarks aided us to polish up our work and it's significant impact can be seen in our project.

We are also additionally thankful to all the faculty members and research scholars who have been exceptionally supportive in providing assets to finish this work.

In the end, we also want to thank every one of our companions who have forever been an incredible help to us and a wellspring of inspiration to advance further.

Harsh Meena (IIT2019005)
Hrithik Sharma (IIT2019020)
Shubham Netam (IIT2019028)
Karan Kumar Sah (IIT2019046)
Ravikant Vaishnav (IIT2019048)

Contents

1. Origin of Proposal

With road transport becoming more accessible than ever before, road safety has become one of the biggest challenges faced by humanity in the 21st Century. Most estimates point at horrifying statistics when it comes to road accidents, with approximately 2 people dying, and nearly 100 being injured or permanently disabled every minute, worldwide, due to road accidents. [1] According to a WHO status report on road safety from 2018, road traffic injuries are the leading cause for deaths in children and young adults aged 5 to 29 [2]. Apart from causing a major loss in human life, it causes huge economic losses to countries all around the globe, costing countries of the world approximately 3% of their GDPs every year.

Furthermore, 93% of these road accidents occur in countries that are considered to still be developing, or underdeveloped, even though these nations account for only 60% of the world's automobiles. [2] These stats are a huge cause for concern, and are reason enough for the nations of the world to invest intensively in researching ways to combat traffic accidents.

1.1. The Human Factor

A majority of road accidents are, without a doubt, down to human error. In fact, some of the major causes of road accidents, like mobile phone usage, drug habit and addiction to alcohol use, distraction and wild driving, often end up going unreported in most records of road accidents. [3] Fortunately, vision based solutions to perform behaviour recognition on drivers, is a technology that has advanced significantly over the past few years.

One of the most common techniques, which we will be exploring in our research, is the use of convolutional neural networks on motionless image datasets, subjected to real driving conditions. [4] Being able to eliminate the human factor through this technique can help significantly advance research in the area of road safety.

2. Problem Statement

The Problem Statement of our project is, **given images of different perspectives of a driver we have to create a machine learning model that recognises and classifies the unsafe and safe driving postures.**

In our approach, we will put emphasis on one of the most important areas of Machine Learning, that is, image recognition and computer vision. The problem which we will be working on is related to the anticipation of safe/unsafe driving posture when driving a vehicle in real driving conditions.

3. Objective

Our objective is to build a system which applies deep learning to automatically learn and classify driver postures, in order to ensure the practice of safe driving. Our system will monitor driver position with information extracted to predict safe/unsafe driving posture.

4. Literature Survey

4.1 International Status

Over the past few years CNN for driver posture recognition has been a popular research topic, with a significant amount of research being done in order to try and devise an accurate model for the problem. Researchers tend to use a CNN architecture for the problem, with a few common datasets being used in order to create a model.

- **Chao Yan and Bailing Zhang** came up with a solution which uses a CNN architecture and focuses on characterization of a driver's posture that mainly depends on a driver's hand position. Each layer of the network creates a working model by using sliding filters, with the map sizes decreasing track by track and eventually becoming more complex and globalised. Finally, the output is fed into a multilayer perceptron in order to classify it. The researchers try to classify the poses into 4 different types and achieve an accuracy of 99.78%. [9]

- **Teng, Yuxuan & Yan, Shiyang & Zhang & Smith, Jeremy, Bailing** proposed a solution wherein, from a picture, skin-like parts are identified Gaussian Mixture Model, which is then send to the deep convolutional neural networks model, called R*CNN, to produce labels. They used SEU dataset and corresponding distribution for the dataset which contains six categories of behaviour: break, call, eat, wheel, mobile-play, smoke. The skin tests collected are switched over completely to YCbCr colour space for further GMM trainingFurther in testing, pictures are changed to YCbCr space and afterward demonstrated by GMM, the result trademark maps are post-handled by binarization and widening to make last skin-like locales.. To assess the framework execution, the AP is processed for each activity classification, the accuracy review bend (PR bend) is charted for each class. The region under the PR curve that is AP, is computed likewise. Their system performance computed a mean AP of 97.76. [4]

- **Yan, Chao; Coenen, Frans; Zhang, Bailing:** in 2015 discusses the importance of separation of the driver's behaviour in a reduced state for car accidents. This paper introduces a novel program that uses a CNN architecture to automatically guess among pre-characterized driving conditions. The fundamental intasion is to screen the location of the driver's hand with discriminatory information issued to guess unsafe/safe driving position. In the works of the writers, the CNN model was firstly trained sub-filtration, and then neatly organised into phases. This method is verified using SU's driving status database, which contains video clips, which includes four driving modes, including normal driving, answering calls, smoking,and food. Compared to other many popular methods with descriptions of various images and methods of classification, the authors' plan accomplishes excellent performance with an overall accuracy of 99.77%. Performance testing again worked normally in real-world situations, the method was tested and two more were used, specifically data sets designed for poor lighting and various road conditions, to find an absolute accuracy of 99.3 and 95.77%, respectively. [5]

- **Researchers from Central South University, Changsha, China,** note that CNN models have shown higher accuracy in object detection and recognition. They devise a way to figure out the fatigue levels in drivers. During the fatigue detection process, images can have uneven grayscale levels and may produce faulty results during the detection process. To deal with this issue, gamma rectification is included in the frame to diminish the impact of uneven grey distribution on the picture and to increase outline picture quality. State Recognition Network and Location Detection Network are suggested for the implementation of drowsiness detection. A driver tiredness guidelines with three levels were given to finish up whether the driver is driving in an exhausted state. A YawDD dataset and a self-built dataset were used. The researchers obtained an accuracy of 93.83% using gamma correction and 78.96% without it. [10]

- **Mahmood Fathy,Mohammad Sabokrou, Reza Berangi,Mohammad Shahverdy,, from Iran proposed a new and efficient way for examining the behaviour of drivers.** The method they proposed used driving signals, including speed,Revolutions Per Minute (RPM), throttle,gravity, and acceleration to identify different types of driving styles, that have 5 types which are: distracted,drunk driving,
drowsy, aggressive, and normal . They trained a 2D CNN on pictures created from driving signals which depended on the repeat plot technique,as deep neural networks perform very well on images. [14]

- **Chien-Yi & Huang,Cheng-Rong,Peng-Yu & Lin,chen, Ju-Chin & Lee.**
. This set of researchers wrote that motor vehicle crashes are one of the top causes of road accidents. According to them, we need to improve the system for detecting driver interference in automatic vehicles. Promoted by two CNN models, they developed a driver behaviour analysis system using CNN analysis, sequential inputs, and feature maps from the last convolution layer released as spatial and temporal features for continuous separation. Unlike previous studies using manual defined weights or instruments obtained through a fine-tuning process, a fusion type network was created to combine the attributes of the separation status. In addition, a self-generated database of ten activities in the car was invented. From the test results, the mentioned system could be scaled up; the rate of accuracy was about 30% compared to the original CNN model of the two streams. Next, a complete network configuration can be designed for highly identifiable and varied actions scene challenges, such as night time behaviour recognition. [12]

- Another group of researchers, come up with a solution consisting of a weighted cluster of CNN’s. The CNN’s are trained on hand images, raw images, face images, skin-segmented images, and“face+hands' ' images. An AlexNet network had trained and benchmarked for those 5 images, a ResNet network with fifty layers, an Inception V3 , and a VGG-16 network. For all these networks, they fine-tuned a trained Image Net model. Then by using a genetic algorithm, they evaluated the weighted sum of all network’s outputs that gives the final distribution of class. The datasets used were from StateFarm and Southeast University (SEU). Their most efficient model was using a genetically weighted ensemble of CNNs' to achieve a classification accuracy of 90%. They also showed a model that only uses AlexNet that can work in real-time and sustain satisfactory classification results and good accuracy. [13]

4.2 Indian Status

- **Rajamohanan S.P. Et. al. used a hybrid of biLSTM and CNN to detect a driver's sleepiness.** They noted that due to complexity in the human face and there are many variations, it is difficult to do facial recognition with better accuracy,so recognition of the eye feature is preferred as it is more advantageous in that case . Through eye recognition, they estimated the facial features and to some extent also estimated the drowsiness level of the driver, which proves to be a good advantage during drowsiness detection. They used shape predictor dlib for receiving content from facial image's real time data . They used opencv and dlib to create facial landmarks. CNN was for taking out the needed traits from the data sets. The method utilising CNN and biLSTM gave the authors an accuracy of upto 96% [15]

4.3 Importance in the context of current status

There's been progressing interest in creating deep learning models for various vision based tasks lately, with it being applied in a vast range of fields, from document recognition to action recognition[5]. Training such models using both unsupervised and supervised learning approaches has resulted in significant advancements in the fields of natural language processing [6], speech recognition [7], and image classification [8] to name a few.

- Perhaps the most popular model used in Deep Learning is the CNN model. CNN stands for convolutional neural networks, and is a hierarchical multilayered neural network which is able to learn patterns directly from an image's pixels. Little fixes of the picture or image are generally fed into the input layer of the neural network, and in each further layer, trainable filters and neighbourhood pooling tasks are taken advantage of to create remarkable features or highlights for the data observed. [5]

In our research, we will attempt to create a CNN architecture in order to detect and classify postures in drivers into various categories of safe and unsafe. We believe that this kind of technology can be used in effectively combating the dangers of road traffic, which is why further advancements are needed in order to make it as efficient as possible.

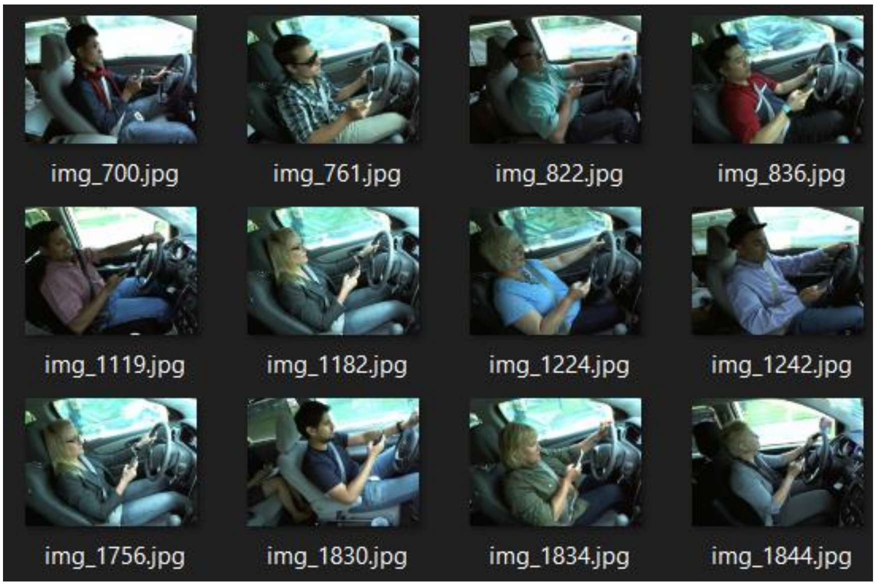
4.4 Area of Operation

A camera will be attached inside the vehicle in the top corner of the front passenger seat, facing towards the driver seat. It will take input and will send it to the Driving Posture Recognition system for the evaluation of safe driving.

5. Work Plan

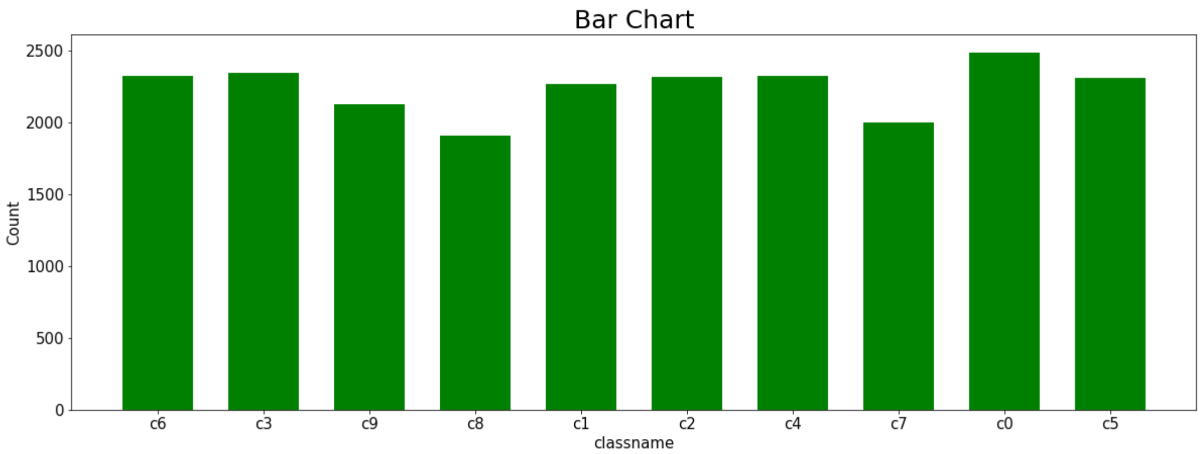
5.1 Dataset and Preparation

The dataset we are using is the State Farm Distracted Driver dataset. The dataset contains various images of drivers performing various actions, such as texting, eating, doing nothing etc. The dataset is divided into training and testing datasets, and no driver who appears in one appears in the other. This will ensure that we are able to test accurately and prevent overfitting the model on the training dataset.



The actions being performed by the driver are divided into 10 classes:

- c0: safe driving
- c1: texting - right
- c2: talking on the phone - right
- c3: texting - left
- c4: talking on the phone - left
- c5: operating the radio
- c6: drinking
- c7: reaching behind
- c8: hair and makeup
- c9: talking to passenger



The dataset is taken from kaggle.

5.2 Proposed Methodology

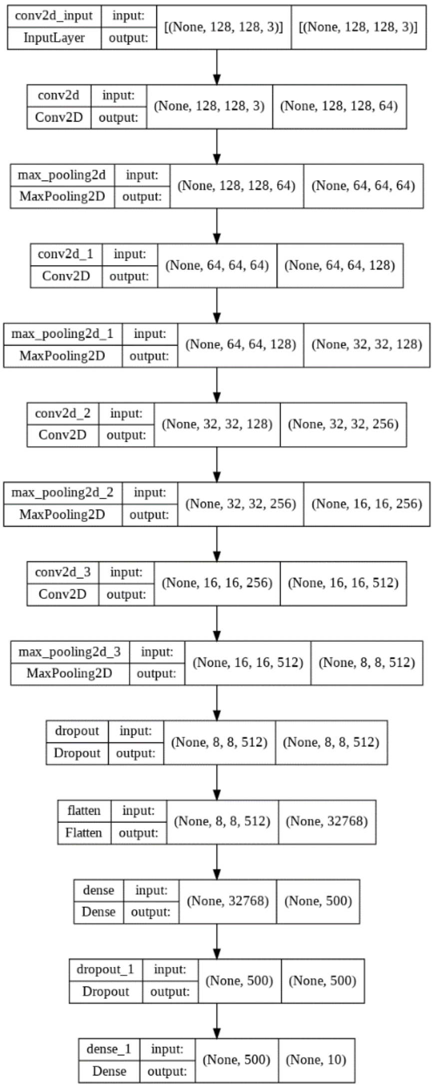
5.2.1 CONVOLUTIONAL NEURAL NETWORKS

Convolutional Neural Networks are deep learning architectures that are particularly suited to identifying patterns in images, in order to recognise objects like faces, scenes, alphanumeric values, etc. A CNN generally consists of several layers, each trained in learning different aspects or features of a given image.

A few of the various types of layers that we have used in our architecture are:

- Convolutional Layers - They are the main building blocks of a convolutional neural network. They generally consist of a set of filters, of a given size known as the kernel size. Each filter consists of parameters that need to be learnt through training.
- Max Pooling Layers - Max pooling is a form of pooling operation in which the maximum value from a region covered by the filter is picked, in order to obtain a convolved image that covers the most prominent features of the original image.
- Dropout Layers - A dropout layer is essentially a kind of mask that negates the impact of some neurons of a neural network, on each iteration. The number of neurons impacted by this is determined by the dropout rate, which is a non-learnable parameter. This is a regularisation technique used to avoid overfitting.

The following is the CNN architecture being used by us on our problem dataset:



As can be seen from the above illustration, the model consists of 4 convolutional layers, each followed by a max pooling layer. This set of layers leads into a dropout layer, followed by a flatten layer and dense layer, followed by another dropout layer, and finally ending with the output layer.

The convolutional layers and dense layers all use ReLu as the activation function, except for the output layer which naturally uses softmax, since it is a multi class classification problem.

5.2.2 BENEFIT OF CNN OVER STANDARD ANN

Standard neural networks are nowhere near as powerful as variations like convolutional neural networks. They are not capable of detecting important features in the dataset without the presence of human supervision. CNNs can provide an ample platform for computer vision and image recognition tasks which require the models to identify unique features between images, sometimes having to differentiate between two images based on very fine details.

5.2.3 BAYESIAN CNN

Bayesian CNN is a probabilistic model that we can use to analyse the effectiveness of our model. In Bayesian CNN we mainly focus on estimating uncertainties, which can be of two types, one that measures for variation of data and the second that measures for the model.

$$\underbrace{\frac{1}{T}\sum_{t=1}^{T}\mathrm{diag}(\hat{p}_t)-\hat{p}_t^{\otimes 2}}_{\text{aleatoric}}+\underbrace{\frac{1}{T}\sum_{t=1}^{T}(\hat{p}_t-\bar{p})^{\otimes 2}}_{\text{epistemic}},$$

The first term in the equation is for measures for variation of data and the second measures variation of the models. After calculating these two terms we can also convert them in probability to output the results.

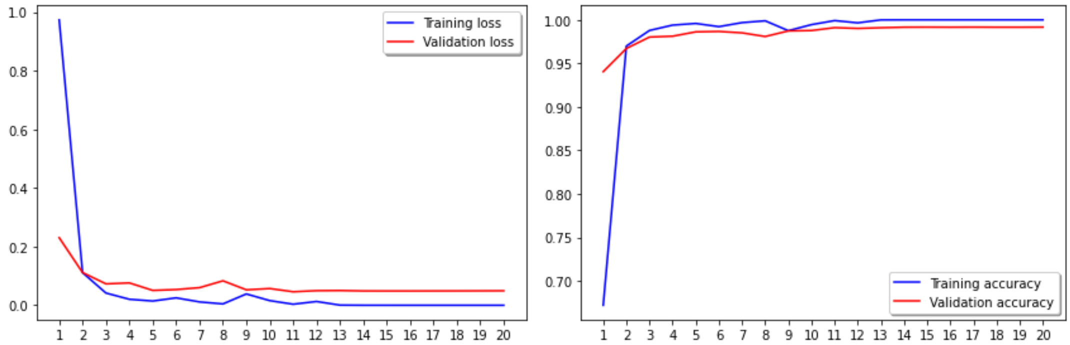
We divide our methodology to solve this problem into the following steps:-

- Training a few standard CNN architectures and analysing their performance.
- Training a unique CNN architecture based on the needs of our problem.
- Fine tuning the model to achieve maximum accuracy.
- Performing Bayesian analysis on our model.

5.3 Experimental setup and Metrics

We first trained a few well known CNN architectures on our dataset, in order to get some benchmarks of how well they would work for our problem statement.

The first model that we trained is a popular CNN architecture called AlexNet. It is considered to be one of the most influential and powerful deep learning models in the world when it comes to image recognition problems. The following is the progression of the loss and accuracy values over 20 iterations of training, with a batch strength of 40, and using rmsprop as our optimizer, and categorical crossentropy as the loss function, with a 80-20 split of the dataset for training and testing:

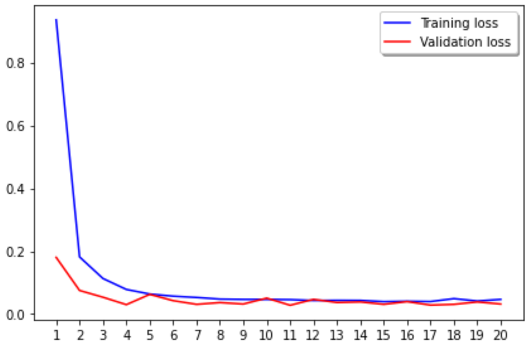


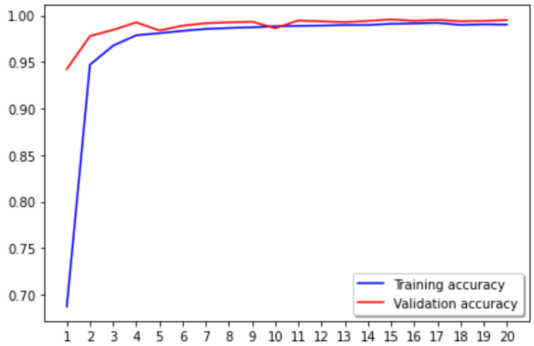
The accuracy scores are as follows:

Accuracy = 99.16%
Precision = 0.991624
Recall = 0.991616
F1 score = 0.991615

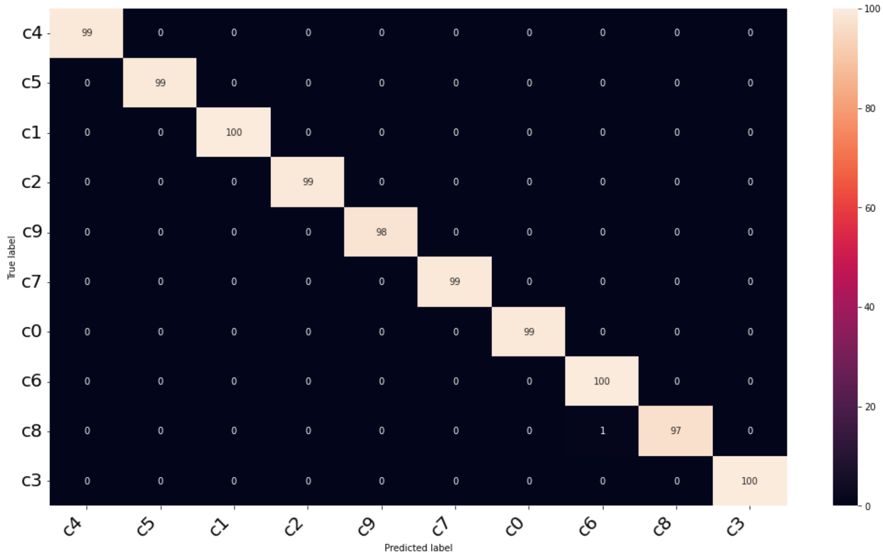
We then proceeded to trial a few more architectures, including some used in other papers, and then followed that up with our own architecture. The full comparison table with accuracies of all models can be found after the following section.

The results of our own model after training over 20 iterations, under the same conditions as the previous model, were as follows:





The confusion matrix while classifying over our test dataset is as follows:



Finally, the accuracy scores were as seen in the snapshot below:

```
Accuracy: 0.995095
Precision: 0.995113
Recall: 0.995095
F1 score: 0.995084
```

A comparison of the performance of our model with that of a few other architectures, including those found in other papers based on our problem, is as follows:

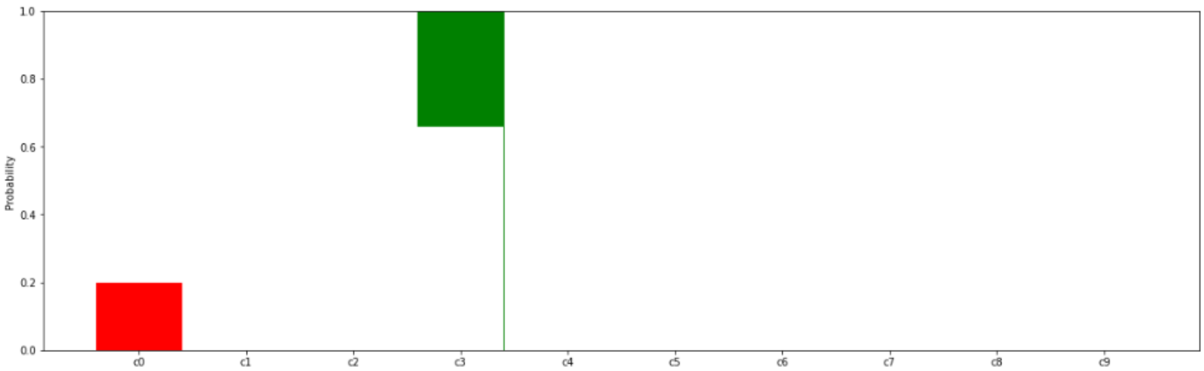
Model	Source	Accuracy(on our dataset)
Our model	-NA-	99.5%
AlexNet	[16]	99.1%
LeNet-5	[17]	98.9%
VGG-16	[18]	99.1%
Yan, Chao; Coenen, Frans; Zhang, Bailing	[5]	99.4%

Further we performed some Bayesian Analysis on our model. The results of that are as follows:

Image I:

Expect result: c3 (Texting-left)

Predicted result:



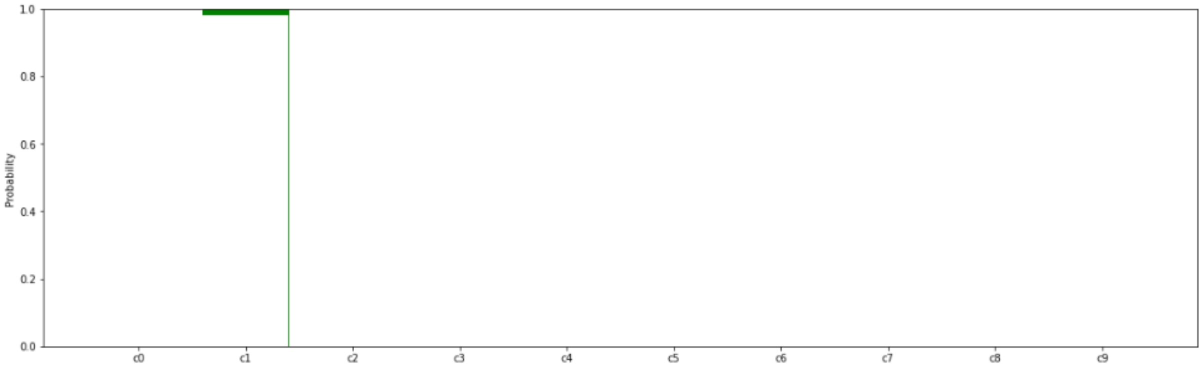
Here we can see the probability distributions of what the model considers the given image to be. The red bar indicates an incorrect prediction, while the green bar indicates a correct prediction. The probability distribution of each predicted class is given over a set of 400 predictions. As we can see, while making the correct prediction of c3, our model is anywhere between 100% to nearly 60% confident, while the incorrect prediction is done with a low confidence of nearly 0% to 20%.

Similarly for other images.

Image II:

Expect result: c1 (Texting-right)

Predicted result:

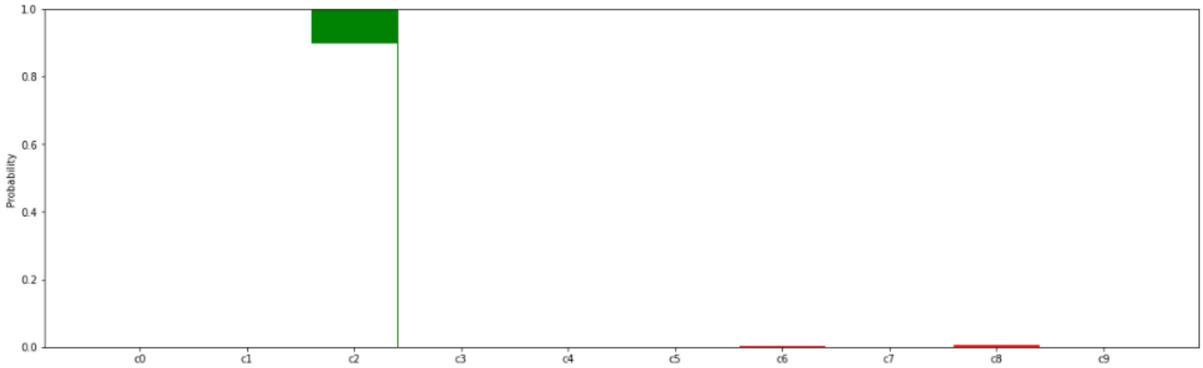


As we can see, the model is nearly always close to 100% confident in making the correct prediction of c1.

Image III:

Expected result: c2 (Talking on the phone- right)

Predicted result:



In this case, the model does make a few incorrect predictions with very low confidence levels (probabilities) while it is fairly confident whenever it makes a correct prediction.

6. Conclusion and Future Scope

Being able to recognize a driver's behaviour and physical state without having any physical presence in the vicinity of the driver can prove to be extremely crucial in the domain of vehicular transport moving forward. The list of domains our model can be applied in is endless. It can help conduct driving tests for corporations and the government alike. It can help monitor proper driver behaviour for corporations offering taxi/delivery services to help ensure safe practices from the employees' side.

The scope for research in this domain is endless, it can help solidify the already growing presence of tech in vehicular transport, by helping cover one of the most crucial aspects of the industry :- safety. The next step to further this project would be to implement an application for practical use in real life. Also the problem must be solved from an Indian perspective, with the driving wheel on the right side of the car. This should be possible by mirroring the images, but must be verified by gathering a dataset from an Indian perspective.

7. Bibliography

[1]Maqbool, Younus & Sethi, Ankit & Singh, Jagdeep. (2019). Road safety and Road Accidents: An Insight. 6. 93-105.
[2]Global Status Report on Traffic Safety, (2018), WHO
[3]Jonathan J. Rolison, Shirley Regev, Salissou Moutari, Aidan Feeney, What are the factors that contribute to road accidents? An assessment of law enforcement views, ordinary drivers' opinions, and road accident records, Accident Analysis & Prevention, Volume 115, 2018, Pages 11-24, ISSN 0001-4575, https://doi.org/10.1016/j.aap.2018.02.025.
[4]Yan, Shiyang & Teng, Yuxuan & Smith, Jeremy & Zhang, Bailing. (2016). Driver behaviour recognition based on deep convolutional neural networks. 636-641. 10.1109/FSKD.2016.7603248.
[5]Yan, Chao; Coenen, Frans; Zhang, Bailing: 'Driving posture recognition by convolutional neural networks', IET Computer Vision, 2016, 10, (2), p. 103-114, DOI: 10.1049/iet-cvi.2015.0175 **[BASE PAPER]**
IET Digital Library, https://digital-library.theiet.org/content/journals/10.1049/iet-cvi.2015.0175

[6]Sanyal, Satyaki & Hazra, Souvik & Ghosh, Neelanjan & Adhikary, Soumyashree. (2017). Resume Parser with Natural Language Processing. 10.13140/RG.2.2.11709.05607.
[7]Lin, Hung-Pang & Zhang, Yu-Jia & Chen, Chia-Ping. (2021). Systems for Low-Resource Speech Recognition Tasks in Open Automatic Speech Recognition and Formosa Speech Recognition Challenges. 4339-4343. 10.21437/Interspeech.2021-358.
[8]Prof. Khemutai Tighare, Prof. Rahul Bhandekar, Harshali Ragite, "Image Classification Between Two Animals", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 7, Issue 4, pp.373-376, July-August-2021. Available at doi : https://doi.org/10.32628/CSEIT2174103
[9]Chao Yan, B. Zhang and F. Coenen, "Driving posture recognition by convolutional neural networks," 2015 11th International Conference on Natural Computation (ICNC), 2015, pp. 680-685, doi: 10.1109/ICNC.2015.7378072. **[BASE PAPER]**
[10]Hu He, Xiaoyong Zhang, Fu Jiang, Chenglong Wang, Yingze Yang, Weirong Liu, Jun Peng,
A Real-time Driver Fatigue Detection Method Based on Two-Stage Convolutional Neural Network⁎⁎This work is partially supported by National Natural Science Foundation of China (Grant Nos. 61873353 and 61672539).IFAC-PapersOnLine,Volume 53, Issue 2,2020,Pages 15374-15379,ISSN 2405-8963,https://doi.org/10.1016/j.ifacol.2020.12.2357.
[11]Hu, Y., Lu, M., & Lu, X. (2020). Feature refinement for image-based driver action recognition via multi-scale attention convolutional neural network. Signal Process. Image Commun., 81.
[12]Chen, Ju-Chin & Lee, Chien-Yi & Huang, Peng-Yu & Lin, Cheng-Rong. (2020). Driver Behavior Analysis via Two-Stream Deep Convolutional Neural Network. Applied Sciences. 10. 1908. 10.3390/app10061908.
[13]Eraqi, Hesham & Abouelnaga, Yehya & Saad, Mohamed & Moustafa, Mohamed. (2019). Driver Distraction Identification with an Ensemble of Convolutional Neural Networks. Journal of Advanced Transportation. 2019. 1-12. 10.1155/2019/4125865.
[14]Mohammad Shahverdy, Mahmood Fathy, Reza Berangi, Mohammad Sabokrou, Driver behavior detection and classification using deep convolutional neural networks, Expert Systems with Applications, Volume 149, 2020, 113240, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2020.113240.
[15]S.P. Rajamohana, E.G. Radhika, S. Priya, S. Sangeetha, Driver drowsiness detection system using hybrid approach of convolutional neural network and bidirectional long short term memory (CNN_BILSTM), Materials Today: Proceedings, Volume 45, Part 2, 2021, Pages 2897-2901, ISSN 2214-7853, https://doi.org/10.1016/j.matpr.2020.11.898.
[16]Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1 (NIPS'12). Curran Associates Inc., Red Hook, NY, USA, 1097–1105.
[17]Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition," in Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, Nov. 1998, doi: 10.1109/5.726791.
[18]S. Liu and W. Deng, "Very deep convolutional neural network based image classification using small training sample size," 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), 2015, pp. 730-734, doi: 10.1109/ACPR.2015.7486599.