PREDICTING DISTRACTED DRIVING

A.Rajsaranya[*], M.Nisha[**], S.Sivaranjani[***],LawrenceXavier[****]

[*] Department of computer science, Loyola icam college of engineering and technology

DOI: 10.29322/IJSRP.X.X.2018.pXXXX
http://dx.doi.org/10.29322/IJSRP.X.X.2018.pXXXX

ABSTRACT:

Distracted driving is the leading cause of motor vehicle crashes all over the world.The detection of a distracted driver is a key study field for decreasing road accidents.With deep neural networks, this research focuses on a methodology for reducing accidents caused by distracted drivers.The behaviors of the driver are developed using a CNN-based algorithm using the driver picture collection, which is then used to classify distracted drivers into distinct groups.The focus of this research is on an approach for lower the number of collisions caused by distracted drivers Deep neural networks are a type of artificial intelligence. In this research, we propose a set of computer vision approaches for processing picture data captured from inside automobiles to automatically detect when drivers are distracted. We accomplish this by using pictures to detect and localize a variety of things in automobiles that contribute to distracted driving (e.g., hands, cellphones, radio, etc.).Then, inside a picture, we use machine learning techniques to analyze the relative placements of these items in order to develop predictions about distracted driving.A CNN-based method is utilized to develop a driver's activities from a driving picture dataset, which is then used to categorize distracted drivers into different groups.A state farm dataset was used to create a deep neural network model, which includes 10 activities in 26 different subjects such as texting, cell phone use while driving, late arrival, normal driving, and alcohol use. The results from six epochs demonstrate that all of the experiments have above 75 percent accuracy, with the greatest result being 98 percent.

Keywords - distracted, drivers, driving, picture, research, neural, deep learning

Index Terms—Distracted driver, road accident, action recognition, deep learning, convolutional neural network, vanilla CNN,

I. INTRODUCTION

According to data released by the Ministry of Road Transport and Highways, more than three lakh road accidents were recorded in India in the calendar year 2020 by various states and union territories (UTs), killing 1,31,714 deaths and injuring 3,48,279 people. Another survey from the United States, published by the National Highway Traffic Safety Administration (NHTSA), said that distracted driving was responsible for 64.4 percent of fatalities. In addition, 94 percent of automobile accidents are caused by driver error. Furthermore, 86 percent of drivers admit to using one or more of the following while driving: answering phone calls, responding to messages, checking radio, conversing with a fellow passenger, eating or drinking while driving, checking maps, watching video, grooming, and surfing the web**.** We create computer vision algorithms to evaluate images from a dataset in order to detect distracted driving while keeping the context of the discovery. The state farm distracted driving dataset from Kaggle is the dataset we utilize for this issue. In this data collection, almost 1000 photographs were taken , with an in-car camera positioned just above the passenger side window and facing the driver window. The photos are separated into nine categories, one for safe driving and nine for distracted driving. Throughout the image collection, real drivers imitated different aspects of distracted and safe driving while the automobile stayed still (for safety reasons). This dataset is well-known, and it's frequently used in peer-reviewed research. For this study, we used photos from all nine distracted driving sessions as well as images from a single safe driving lesson. Talking on the phone (with both hands), texting (with both hands), drinking while driving, adjusting the car radio, interacting with a side passenger, applying cosmetics, and reaching for the back are all examples of distracted driving.The eleventh class is all about driving without being distracted.

In this study, we suggest a two-tiered technique to identify a picture as distracted or safe driving. We construct a Vanilla Convolutional Neural Network algorithm in the first layer to scan an image and then recognise and locate critical components that might contribute to distracted driving. Our dataset includes smartphones and bottles (external devices), as well as steering wheels, radios (internal to the car), and the left hand, right hand, look straight, look right, and keep looking back (human-centric). Following the identification and localization of these objects of interest in a photo, less advanced machine learning techniques are used to process the relative locations of the localized objects and categorize samples based on their proximity. The total contribution of our approaches in this study will create better and contextual inputs to the end user, which (we believe) will lead to greater driver comprehension and distracted driving correction, which is not done in this paper. We go into further information regarding these consequences later in the research.

II. RELATED WORKS

One of the difficulties that is being researched is finding the distracted driver using machine learning techniques. This subject has been the subject of several studies.

Table 1 lists the many studies on driver distraction and their findings.

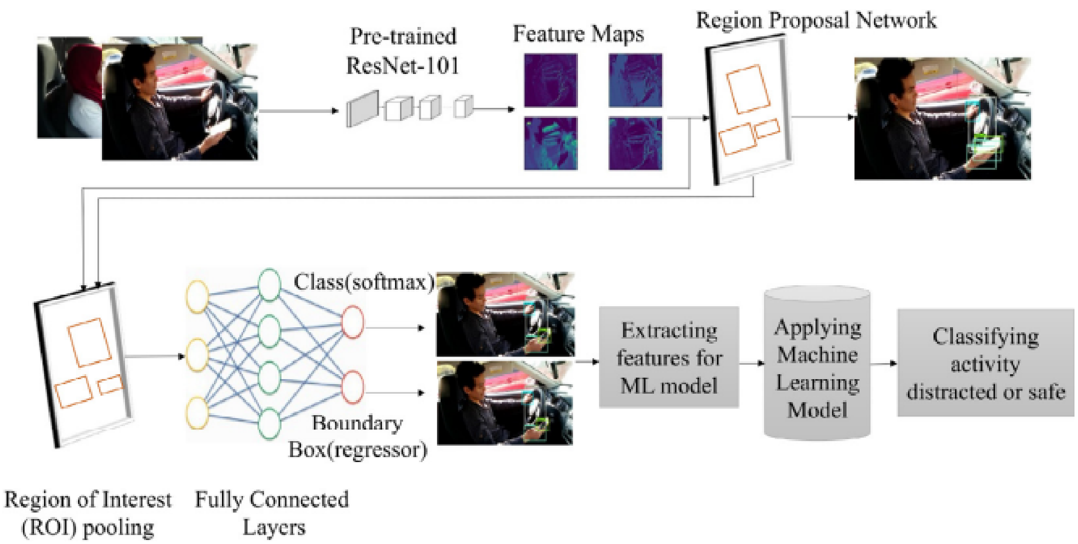
Table 1. Literature Survey

PAPER NAME	YEAR	ALGORITHMS USED	DATASET	AUTHOR	REFERENCE
Real time detection of distracted driving	2020	Vanilla CNN	State farm Distracted driving dataset	Abdul Jamsheed V Dr. B Janet Dr. U Srinivasulu Reddy	https://doi.org/10.1109/ICSSIT48917.2020.9214233
Context-driven detection of distracted driving using images from in-car cameras	2021	CNN	American University of Cairo (AUC) distracted driving dataset	Arup Kanti Dey* , Bharti Goel, Sriram Chellappan	https://doi.org/10.1016/j.iot.2021.100380
Machine Learning Techniques for Distracted Driver Detection	2019	CNN	State farm distracted driving dataset	Demeng Feng Yumeng Yue	http://cs229.stanford.edu/proj2019spr/report/24.pdf
Distracted Driver Detection Based on a CNN With Decreasing Filter Size	2021	CNN	AUC Distracted Driver (AUCD2) and State Farm Distracted Driver Dataset	Binbin Qin , Jiangbo Qian , Yu Xin, Baisong Liu, and Yihong Dong	https://doi.org/10.1109/TITS.2021.3063521

Distracted Driver Detection with Deep Convolutional Neural Network	2019	CNN	Statefarm distracted driving dataset	O. G. Basubeit, D. N. T. How, Y. C. Hou, K. S. M. Sahari	http://dx.doi.org/10.35940/ijrte.D5131.118419
Distracted driver detection by combining in-vehicle and image data using deep learning	2020	CNN	data collected from real-world drives.	Furkan Omerustaoglu, C. Okan Sakar * , Gorkem Kar	https://doi.org/10.1016/j.asoc.2020.106657
distracted driver detection	2020	CNN	State farm distracted driving dataset	Prof..Manya Gidwani Deep Ruparel Abhay Rajde Sahil Shah	https://www.jetir.org/papers/JETIR2010371.pdf
Distracted Driver Classification Using Deep Learning	2019	CNN,Alexnet	AUC Distracted Driver (AUCD2) and State Farm Distracted Driver Dataset	Bandar Alotaibi Munif Alotaibi	https://www.researchgate.net/publication/336685346_Distracted_Driver_Classification_Using_Deep_Learning
Recognition of driver distractions using deep learning	2018	CNN	State farm distracted driving dataset	Leonel Cuevas Valeriano; Paolo Napoletano; Raimondo Schettini	https://doi.org/10.1109/ICCE-Berlin.2018.8576183

III. SYSTEM STUDY

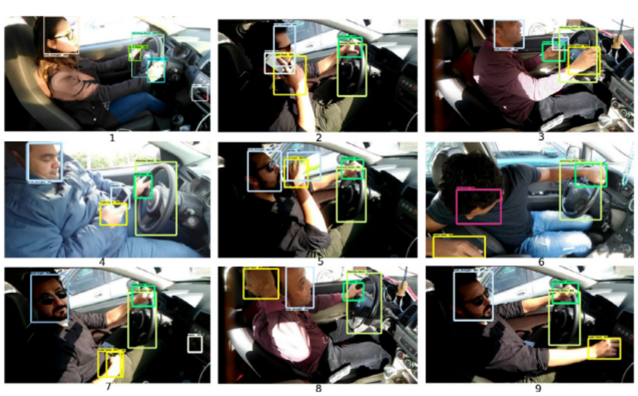
III A). Existing System:



They utilised the AUC distracted driving dataset in the previous system. They isolated 150 unseen photos from the image collection divided across all classes under consideration for testing reasons (nine distracted and one safe driving class). Based on intuition and observation of the AUC dataset, the nine elements of interest in this present system are left hand, right hand, smartphone, bottle, radio, steering wheel, face facing straight, face looking behind, and face looking to the passenger side. They look for these things of interest in a picture and then locate them in the image to estimate their relative locations. They then use their relative locations to judge if the image is suggestive of inattentive driving.They used image processing to see if any of these things of interest are present, and then they localised them in the image to see where they are in relation to one another. They determine whether or not the image is suggestive of inattentive driving. The Faster R-CNN architecture for recognising and localising objects of interest in an image is one of the approaches used in their proposed distracted driving detection systems. They use the Faster R-CNN (Faster Regional-Convolutional Neural Networks) approach for object recognition and localization. For their challenge, they chose 1317 photos for training and validation that were evenly dispersed throughout all 10 classes of interest in their investigation. For training the ten courses, there were 102 photographs from six classes and 103 images from four classes, totaling 1024 images. They then chose 293 additional photos for algorithm validation (equally distributed across all 10 classes). Initially, the picture size for training was 1920 x 1080 pixels, but they lowered it to 960 x 540 pixels for faster training. They utilised the ResNet-101 network to extract features from training photos, and they now have the fundamental feature maps for the complete image dataset. They used numerous bounding boxes to encompass the full picture, some of which had foreground and some of which contained background. They created and optimised a simplified CNN to distinguish between boxes having foreground and background components. They used the idea of ROI pooling to turn these variable-sized feature maps into fixed-sized feature maps, in which the variable-sized feature maps are broken into smaller portions of fixed-sized feature maps to ease processing without sacrificing accuracy. Finally, they created a regressor to tighten the localised object, since the narrower the box is, the better our ability to discern relative locations of many items of interest in an image for contextual categorization of distracted driving will be. They then analyse the identified and localised items of interest (i.e., foreground classes) to reach a final conclusion on whether they are distracted driving or not. To get the greatest results, they devised a basic random forest-based method. The characteristics they detected and processed for this categorization task. They constructed and evaluated four different machine learning models to identify distracted driving from safe driving after localising items of interest and determining attributes. They sought to categorise merely distracted driving or not using the same localization technique, followed by characteristics retrieved, in addition to multi-class classification of distracted driving.On average, their approach required 200ms to process one image when run end-to-end on a PC (2.5 GHz Quad-Core Intel Core i7 CPU and 16 GB of memory). This included localization of items of interest, categorization of the image as suggestive of distracted driving, and so on.

Accuracy of different model for binary classification.

Model	Accuracy(%)
Random Forest	98
K-Nearest Neighbor	97
SVM	90
Decision Tree	97



III B). Proposed System

B.I. DATASET:

Data collection is the systematic process of acquiring and measuring information on variables of interest in order to answer research questions, test hypotheses, and make conclusionsand In this study, the dataset we used is the Statefarm Distracted Driving Dataset . The collection contains 1000 tagged photos of 26 persons of various colors, ethnicities, actions, and ages. These subjects were used to classify them into ten categories, including normal/safe driving, talking on the phone with both hands, controlling the radio, text messaging with left hand, text messaging with right hand, talking to a co-passenger, drinking while driving, hair and makeup, reaching behind, and so on. The action class of each picture is labelled.The current study made use of the StateFarm distraction-detection dataset.

This is the most widely used dataset for detecting driver attention, and it has been utilized in several research.Each class has around 2,200 RGB pictures at a size of 640x480 pixels.In this data set it shows the number of photos for each class.To build fresh training and testing subsets from the initial training set, the holdout split approach was used to create 10% and 30% testing sets.

Dataset details with number of images per class

CLASSES	DRIVER ACTION	IMAGES
C0	safe driving	2489
C1	texting - right	2267
C2	talking on the phone - right	2317
C3	texting - left	2346
C4	talking on the phone - left	2326
C5	operating the radio	2312
C6	drinking	2325
C7	reaching behind	2002
C8	hair and makeup	1911
C9	talking to passenger	2129
SUM		22424

B.II. ALGORITHMS AND TECHNIQUE:

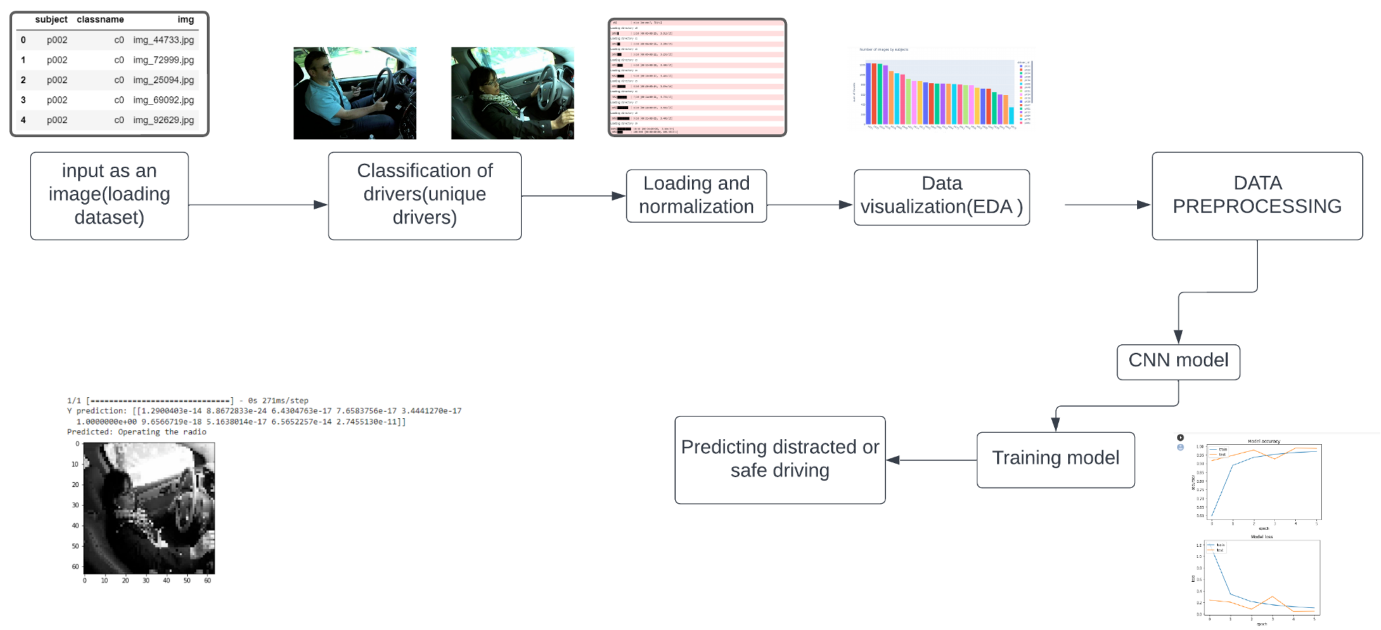
VANILLA CNN :

The vanilla CNN model is made up of three convolutional layers, a flatten layer, and three dense layers in total. Layer 1 of the convolutional network has 60 filters, each with a 3x3 kernel size. The relu activation function is used with the same padding as before, and the weights are set to 0.0001. This layer receives a 0.3 dropout. A total of 90 filters, each measuring 3x3, are employed in the second convolutional layer. This layer also has a 0.3 dropout applied to it.In the third convolutional, a total of 200 filters with a dropout of 0.5 are utilized in a similar way. A flatten layer and three dense layers with 512, 128 and 10 filters each are utilized after three convolutional layers. Loss is also calculated using the RMS-Prop optimizer and category cross entropy. ReLU is the activation function for three convolutional layers, ReLU is the activation function for the first dense layer, and softmax is the activation function for the last dense layer.

Model configuration for improved vanilla CNN is shown below

LAYER	OUTPUT SHAPE
Convolutional Layer	(None, 62, 62, 32)
Max pooling	(None, 31, 31, 32)
Convolutional Layer	(None, 31, 31, 64)
Max pooling	(None, 16, 16, 64)
Convolutional Layer	(None, 16, 16, 128)
Max pooling	(None, 8, 8, 128)
Flatten	(None, 8192)
Dense	(None, 512)
Dense	(None, 128)
Dense	(None, 10)

B III.SYSTEM ARCHITECTURE :



The photos are collected from the state farm distracted driving dataset as input. The drivers are then categorized using the provided dataset. Following that, we must go through the loading and normalization procedures. We must first load the training data set. The pictures are divided into 10 classes, numbered c0 through c9, with each class including various driver actions and graphics. The dataset must then be validated. further Using the EDA principle, we must visualize the data. EDA stands for exploratory data analysis, which is a method of analyzing and summarizing the primary properties of a dataset, such as class distribution, size, and distribution. In our project, for example.In our project, we utilized the aforementioned categories to indicate the quantity of photographs by category and the driver's ID who is distracted. In addition, we analyze the data and create an image of the driver in their mode of distraction. The primary architecture of the vanilla CNN model is made up of three convolution layers: a flat layer, two dense layers with relu and dropouts, and one dense layer for classification using softmax. We can discover the total number of parameters using the technique. Which parameters can you train and which can't you? training For the batches, the model is complete. Accuracy, loss, validation loss, and validation accuracy are all measured in each batch. We'll plot the validation accuracy when we've trained all of the batches.We'll plot the validation accuracy and validation loss across epochs once we've trained all the batches. Then we can anticipate the driver's actions. To forecast the activities of drivers, we concentrate on distractions, which are classified into ten categories for the drive greyscale photos in the data trained and verified dataset.

B.IV. SYSTEM CONFIGURATION:

INTRODUCTION:

Design is a process that takes into account the structure of software, the details of how it works, and the interface between different modules. The design process translates requirements into a presentation of software that can be accessed for quality assurance before coding begins. Software design is constantly evolving as new methods and better analysis and understanding of borders emerge.At this early stage of the software design revolution, there is much room for improvement. Therefore, software design methodology is not as deep, flexible, and quantitative as more traditional engineering disciplines. There are techniques for designing software, and there are criteria for good design. Design notation can be used to help describe design patterns.

SOFTWARE REQUIREMENTS:

- Operating System : Windows 7,8,10 (64 bit)
- Software : Python
- Tools : Anaconda (Jupyter notebook IDE)
- Hard Disk : 500GB and above
- RAM : 8 GB and above
- Processor : I5 and above
- Python libraries : OpenCV, Tensorflow, Scikit Learn, Pandas,...
- Computer vision techniques : Vanilla CNN, Data Augmentation, Transfer Learning

PYTHON:

Python is a high-level, general-purpose programming language that is interpreted.Python's design philosophy prioritizes code readability, as seen by its extensive use of indentation.Its language elements and object-oriented approach are aimed at assisting programmers in writing clear, logical code for both small and large-scale projects.Python is garbage-collected and dynamically typed.It supports a variety of programming paradigms, including structured (especially procedural) programming, object-oriented programming, and functional programming.Because of its extensive standard library, Python is frequently referred to as a "batteries included" language.

Python has the following features:

Free and Open Source,OOPs concept,Support for GUI Programming,High level language,

Extensible feature,Python is Portable language,Python is Integrated language,Interpreted Language

Large Standard Library,Dynamically Typed Language ,Easy to use etc

Open CV:

OPENCV (Open Source Computer Vision Library) is a programming library primarily for real-time computer vision.It was created by Intel and then sponsored by Willow Garage and Itseez (which was later acquired by Intel).The Apache 2 License makes the library cross-platform and free to use.Since 2011, GPU acceleration has been available in OpenCV for real-time processing.

TENSORFLOW:

TensorFlow is an open source machine learning platform that runs from start to finish.This session focuses on using a specific TensorFlow API to develop and train machine learning models. TensorFlow is a rich framework for managing all parts of a machine learning system; however, this class focuses on using a specific TensorFlow API to develop and train machine learning models.

SCIKIT-LEARN

Scikit-learn (previously scikits.learnand also known as sklearn) is a Python machine learning library that is available for free.It includes support vector machines, random forests, gradient boosting, k-means, and DBSCAN, among other classification, regression, and clustering techniques, and is designed to work with the Python numerical and scientific libraries NumPy and SciPy.

PANDAS

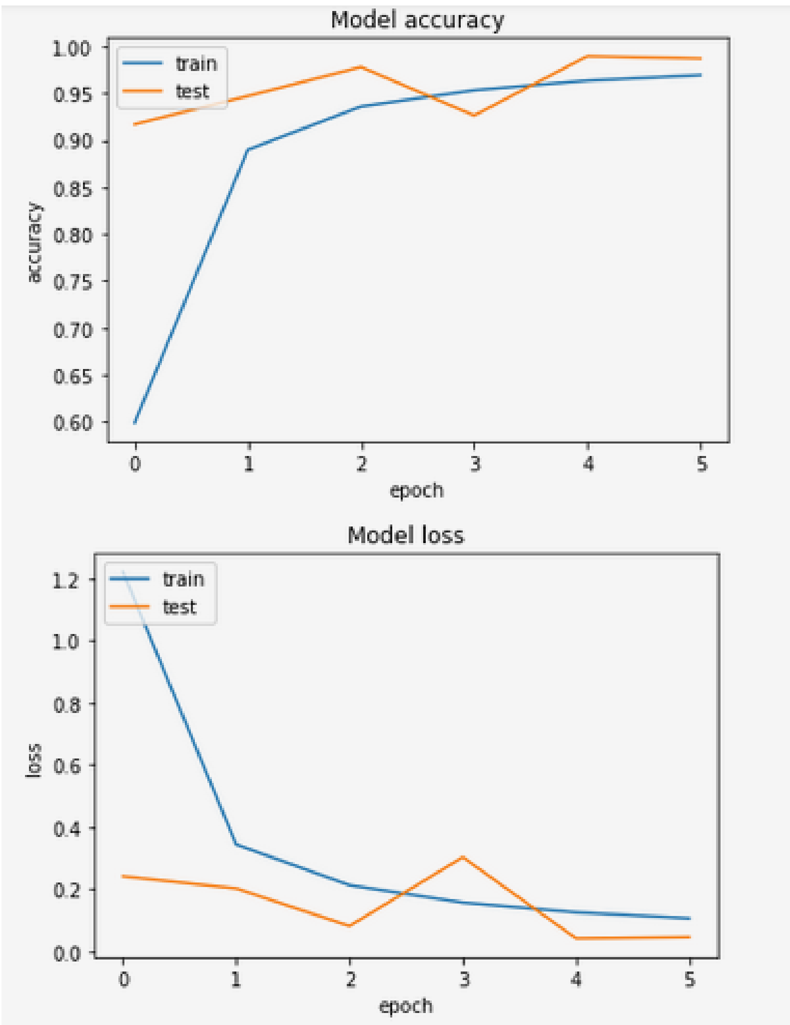
PANDAS is a data processing and analysis software package created in the Python computer language. It includes data structures and methods for manipulating numerical tables and time series, in particular.

It's free software distributed under the BSD three-clause licence.The word "panel data" is an econometrics term for data sets that comprise observations for the same persons over multiple time periods.Its moniker is a pun on the term "Python data analysis."While working as a researcher at AQR Capital from 2007 to 2010, Wes McKinney began developing what would become pandas.It is projected that it is faster than existing approaches for matching pathnames in directories, based on benchmarks.Apart from exact string search, we can employ wildcards ("*,?, [ranges]) with glob to make path retrieval more straightforward and convenient.

V .II. Model Evaluation and Validation:

The vanilla CNN model with the best accuracy of 98.70 percent and the lowest loss of 4.56 percent was optimized.

The Below diagram depicts a more detailed study of Improved Vanilla CNN model accuracy and loss.



V EXPERIMENTAL RESULTS

V .I. METRICS

In order to determine the model's performance, accuracy and loss are calculated.

Forming the confusion matrix allows you to calculate accuracy.

The accuracy equation is as follows:

$$\text{Classification Accuracy} = \frac{TP+TN}{TP+TN+FN+FP}$$

Where FP stands for False Positive (where we projected YES but got NO), TP stands for True Positive (where we anticipated proper classification), FN stands for False Negative (where we expected NO but got YES), and TN is for True Negative (predictions where correct).

Categorical Cross Entropy

The model's Categorical Cross Entropy is determined.For multiclass classification, this loss function is most commonly employed.

The definition of categorical cross entropy is:

$$\text{Categorical cross entropy} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} 1 y_i \epsilon C_c \log P_{model}[y_i \epsilon C_c]$$

Where N is the number of observations, C represents the number of categories, and P represents the probability predicted by the model.The model is trained over a total of six epochs.

RMSProp:

RMSProp is intended to speed up the optimization process, for example, by reducing the number of function evaluations needed to achieve the optimal solution, or to increase the capabilities of the optimization method, for example, by producing a better final result.

According to the findings, vanilla CNN provides the most accuracy with the least amount of loss.The following is a comprehensive study of the model.

VI CONCLUSION

Today, distracted driving is a serious public health threat. With the introduction of several technologies such as cellphones, smartwatches, GPS devices, tablets, and other devices that people routinely use while driving, the problem has gotten worse. We use computer vision techniques to detect distracted driving incidents using photos from a dataset in this work. Unlike other methods, our proposed method distinguishes unique drivers first, then recognises and localizes crucial elements of interest in a picture that indicate distracted driving.Left hand, right hand, smart phone, bottle, radio, driving wheel, face facing straight, face looking behind, and face looking to passenger side were among the things. We create algorithms that process the relative locations of these objects in the image to classify nine types of distracted driving, including talking on the phone (with left and right hands), texting (with left and right hands), drinking while driving, operating the car radio, talking to a side passenger, applying makeup, and reaching for the back.

Our object identification, localization, and classification accuracies are good, giving us confidence that our suggested methodologies may be used to improve driver safety, which is a pressing issue today. Furthermore, our suggested approach's contextual input to drivers is a unique feature. We're already actively collaborating with researchers across the civil, transportation, and human-computer interaction domains to develop behavior modification programmes to increase driver safety and examine how citizens react to such contextual input. We're also assessing the impact on avoiding dangerous and inattentive driving on the roadways.In this study, the model is trained to handle the problem of driver distraction. The proposed model reached 98 percent accuracy, and we attempted a new model that required substantially less training time while maintaining the same level of accuracy.With more data, this approach can still be improved. However, by focusing on the positive qualities of this method, the disadvantages may be overlooked.

VII. FUTURE WORK

We're continuing to work on reducing the amount of parameters and computation time as a result of this effort. Incorporating temporal context may aid in the reduction of misclassification mistakes and, as a result, improve accuracy. We also want to create a system that can identify visual, cognitive, and manual distractions in the future.On an **NVIDIA P5000(change your system GPU)** GPU with 8GB RAM, the system processes **42(change your system picture count)** pictures per second.Investigate further feature extraction methods.
For Image Representation, combine several feature vectors.Modern Deep Neural Network Architectures should be implemented.We may potentially speed up the CNN calculation by compressing the model, as stated in the previous section.

REFERENCES

[1] Abdul Jamsheed, V Dr. B Janet,Dr. U Srinivasulu Reddy,Real Time Detection of driver distraction using CNN,Conference,2020

[2]Arup Kanti Dey, Bharti Goel, Sriram Chellappan,Context-driven detection of distracted driving using images from in-car cameras,Journal,2021

[3]Demeng Feng ,Yumeng Yue ,Machine Learning Techniques for Distracted Driver Detection,,2019

[4]Binbin Qin , Jiangbo Qian , Yu Xin, Baisong Liu, and Yihong Dong,Distracted Driver Detection Based on a CNN With Decreasing Filter Size,Journal,2021

[5]O. G. Basubeit, D. N. T. How, Y. C. Hou, K. S. M. Sahari,Distracted Driver Detection with Deep Convolutional Neural Network,Journal,2019

[6]Furkan Omerustaoglu, C. Okan Sakar , Gorkem Kar,Distracted driver detection by combining in-vehicle and image data using deep learning,Journal,2020

[7]Prof..Manya Gidwani Deep Ruparel Abhay Rajde Sahil Shah.,distracted driver detection,Journal,2020

[8]Bandar Alotaibi Munif Alotaibi,Distracted Driver Classification Using Deep Learning,article,2019

[9]Leonel Cuevas Valeriano; Paolo Napoletano; Raimondo Schettini,Recognition of driver distractions using deep learning,Conference,2018

[10]Annual report 2020, Technical Report, Ministry of Road Transport and Highways,2020.

APPENDIX

Appendixes, if needed, appear before the acknowledgment.

ACKNOWLEDGMENT

The preferred spelling of the word “acknowledgment” in American English is without an “e” after the “g.” Use the singular heading even if you have many acknowledgments.

AUTHORS

A.Rajsaranya , B.E ,Loyola Icam college of engineering and technology,rajsaranya.22cs@licet.ac.in

M.Nisha,B.E ,Loyola Icam college of engineering and technology,nisha.22cs@licet.ac.in

S.Sivaranjani,B.E ,Loyola Icam college of engineering and technology,sivaranjani.22cs@licet.ac.in

LawrenceXavier,B.E ,Loyola Icam college of engineering and technology,lawrencexavier.22cs@licet.ac.in

Correspondence Author – Author name, email address, alternate email address (if any), contact number.