Real-time Driver's Drowsiness Detection by Convolution Neural Network (CNN) of Deep Learning Approach

Prashant Gosai, Usha Barad

Abstract— Statistics have shown that 20% of all road accidents are fatigue-related, and drowsy detection is a car safety algorithm that can alert a snoozing driver in hopes of preventing an accident. This work will propose real-time drowsiness detection; this approach is based on Convolution Neural Network (CNN) of Deep Learning. Which is aimed to implement driver's behavior-based drowsiness detection scenario. Convolution Neural Network (CNN) for learning effective features or facial landmark input to detecting drowsiness by given an input video of driver. A common global face which is not capable enough to extracting effective facial landmarks and features, like facial movements and head gestures, which are strictly important for learning. This proposed work consists Convolution Neural Network (CNN) for attaining well-aligned facial movements and head gestures important for reliable detection. The output of neural network is integrated and feed to classifier for drowsiness detection.

Index Terms— road accident, accuracy, prediction, random forest, ML, security, traffic, severity prediction, computer vision, CNN deep network.

I. Introduction

Sleepiness also referred as Drowsiness can be defined as "the basic need to fall asleep". This affects normal human biological cycle, which comprised of sleep-weak cycles. This cycle is regulated by both hemostatic and circadian factors. Homeostasis relates to neurological need to sleep; long period of wakefulness puts more pressure to sleep and more difficult to resist it. The circadian factor of human body is an internal clock which maintains cycle of wakefulness and sleepiness approximately in every 24 hours and Homeostatic factors regulate circadian factor that regulates the timing of Sleepiness and Wakefulness. The main purpose of this research is to reduce the road accidents due to drowsiness of driver, especially in India, stated by the Indian Ministry Of Road Transport And Highway Research Wing's survey, accidents have been increased by 3.86% from 2018 to 2019. From the analysis of road accidents translate that an average of 1230 accidents and 414 deaths every day and 32% are drowsiness related. [1,2,3]

A. Problem Statement:

Available driver's drowsiness detection system generally lies into two categories: (i) very costly systems, limited to specific high-end car models; and (ii) affordable solutions that lack robustness. Our aim is to focus on to make a drowsiness detection system that makes bridge the gap between them by balancing availability and affordability with functionality to low budget vehicle also. The analysis of previous research in the field of driver's drowsiness detection we emphasize on difficulty and complexity of the problem and three basic challenges found that need to be tackled, which are: reliability, accuracy and speed. The aim of our approach is to overcome these challenges by building a mobile, real-time, dynamic, adaptive system that leverages, whenever possible, readily available computer vision tools.

B. Drowsiness Detection and Measurement Methods:

There are many different ways to detect and measure driver drowsiness (or fatigue mood, sleepiness). They are normally classified into five categories: physiological, vehicle-based, hybrid, subjective and behavioral. Below section provides a brief details of driver drowsiness detection methods in every category.

Subjective Methods:
The subjective method is tool for measuring and analyzing the sleepiness or drowsiness level and measure the particular tendency of the person's drowsiness. Drowsiness is a psychological need that resistance the fatigue of the human body. The more system is awake the stronger need for sleep. Many scientific organizations like "Laboratory for sleep, Division of Sleep Disorder of US" [4] with association of "Professional Sleep Societies" provided various descriptive scale of drowsiness levels.
Subjective methods are used for the assessment of drowsiness are based on questionnaires and electro-phycological measures of drowsiness.

- Epworth Sleepiness Scale (ESS): In this measurement method the drowsiness levels are measured by ongoing activity of the individuals, like watching TV, doing normal work, doing exercise etc. and put the individuals on a scale from 0 (no chance) to 3 (high chance) of drowsiness. Individuals then score between 0 to 24 of total score.

Manuscript revised on March 19, 2023 and published on March 30, 2023
Prashant Gosai, *Department of Computer Engineering Gokul Global University, Siddhpur, Gujarat, India. prashantgosai11@gmail.com*
Usha Barad, *Head of Department, Department of Computer Engineering Gokul Global University, Siddhpur, Gujarat, India*

doi: 10.32622/ijrat.111202307

Subjects are scoring less than 10 are considered mild sleepy and 15 points or more indicate severe sleepy.

- Multiple Sleep Latency Test (MSLT): This experiment-based method that classified the subject that asleep faster are drowsier deprived. This method measures the tendency to fall drowsy by allowing the subject for fragmented sleep and subject are instructed to fall asleep and logs the timing to fall asleep.
- Maintenance of Wakefulness Test (MWT): This is also experiment-based method to measure the drowsiness level of subject. Drowsiness level is measured during heavily drowsy subject is instructed to stay awake least to 20-minute time period and record the duration of fall in asleep during the 20 minutes of time period.
- Stanford Sleepiness Scale (SSS): This method is standardized method to scale the sleepiness. In this method record the seven-measurement statement and rate the current state of subject's alertness.

Physiological Methods:

This method measures the drowsiness level by measuring the electric signal of human body, Body electrical signals such as electrooculogram (EOG), electroencephalogram (EEG)and electrocardiogram (ECG). They are briefly defined and explained below.

- Electrocardiogram (ECG): This method records the human heart electrical signal. As the heart is beat at certain frequency and whenever human body in different state the heart rate and frequencies are change. In drowsy mode heart rate is low and in normal state heart rate constant. Electrocardiogram measures the heart's electrical signal, and analysis of this rate in different state is what this method does.
- Electroencephalogram (EEG): As like ECG method EEG records the human brain's electrical signals. Whenever the subject in different state the Alpha signal is vary and in drowsy state the alpha signal's frequency is low. Only this method can precisely describe human's alertness level.
- Electrooculogram (EOG): This method records the light signals between the cornea and the retina of eye. When human in drowsy state the eye performs the action for fall in sleep. Whenever the human fall in sleep the iris of eye is contract, the beam of Infrared is targeted to iris through cornea energy level of infrared is drop and by this energy drop this method records the energy levels of IR.

  Vehicle-Based Methods:

Vehicle-Based Method is the analysis of past record of road crashes, road accident events and driving pattern of the driver. Some methods are given below:

- Higher speed with little or no breaking**.** The method analyzes the driving pattern during sleepy condition of subject by real-time simulation.
- A vehicle leaves the roadway**.** This method is action-oriented pattern analysis of past recorded accidents. Data of the accident is gathered form road surveillance camera and the pattern of driving condition is recorded.
- The crash occurs on a high-speed road**.** This method is also similar to the "A vehicle leave the roadway" where past record of accident evidences are analyzed.
- The driver does not attempt to avoid crashing**.** This is the tool to measure the alertness of driver behavior by the vehicle indicator applied during the driving condition. This improvises the vehicle safety indicators and safety Modes.
- The driver is alone in the vehicle**.** This is the condition where the driver is driving vehicle alone and report states 82% of drowsy road crash is occur during a single occupant.

Behavioral Methods:

These methods were developed for detecting the drowsy state by computer system. Basic and major methods are described below:

- Head or eye position**.** Whenever the driver is going in drowsy state, some of the muscles of the body is relax, leads to nodding and change in head or eye position. This expression can be recognized by the computer system by computer vision [5].
- Yawning. Constant yawning is behavioral pattern of drowsy state, but the yawning does not always occur before driver goes in drowsy state. This cannot be used as stand-alone feature; it needs to be combined with other indicators [2].
- Eye state. The main focus of this method is to detect the state of eye by determine the eye is open or closed. In particular the frequency of eye blinking is observed and ratio of the eye closure is determined [3].
- Multiple Facial Actions. This method uses multiple facial expression like the position of head, eyebrow, lip and jaw lining combined all together.

Hybrid Methods:

All above mentioned methods have some pros and cons. Vehicle based method only determine the driving pattern and provides guideline accordingly. Where physiological methods give highly accurate performance but it is more expansive and inconvenient for driving condition. While Behavioral methods are cost effective and accurate as well as the convenient for driving condition. Below comparative table shows the method relative issues and conveniences:

doi: 10.32622/ijrat.111202307

Table 1: Comparative table of different detection and measurement methods

Issue	**Detection Methods**		
	Physiological	**Behavioral**	**Hybrid**
Operating Mode	Electronics sensors	Camera	Combination
Convenience	Inconvenient	Convenient	Convenient
Light condition	Not affect	Affect	Less affect
Cost	Ineffective	Effective	Less effective
Detection stage	Earlier	In time	Can be both
Accuracy	Highest	Medium	Moderate
Complexity	More complex	Simple	Medium

doi: 10.32622/ijrat.111202307

II. Literature Review

Table 2: Comparative analysis of literature survey

	José Solaz, José Laparra-Hernández[9]	**Tereza Soukupov**[13]	**Charlotte Jacobé**[8]	**Rateb Jabbar** [12]	**Wanghua Deng, Ruoxue Wu**[23]	**Younes Ed-Doughmi, NajlaeIdrissi**[24]
Year of publish	2016	2016	2018	2018	2019	2020
Method used	Kinect processing method	EAR(Eye Aspect Ratio) SVM	AdaptiveANN	Multi-layer Perceptron	CNN and LSTM	3D CNN and RNN
Dataset	Drivers Breathing and Plethysmography band	300-VWdataset and ZJU Eyeblink Database	Custom dataset of 21 participants	NTHU	Custom	NTHU
Source of detection	Breathing rate and cameras	TV camera, Web camera	Physiological, Behavioral and Car data	Camera	Camera	Camera
Performance	90% accuracy	86% accuracy	80% accuracy	81% accuracy	92% accuracy	92% accuracy
Embedded solution	No	No	No	Yes (Android)	No	No
Advantage	non-invasive and able to handle illuminated roads conditions	Able to detect level of eye openness and robust to low image quality	Different data collected for each driver, that gives more accurate result subjectively	Lower size of trained model	Provide efficient MC-KCF algorithm over KCF	Higher accuracy
Disadvantage	Only 5 subject's data was collected, and required custom camera system to detect motion	Blinkduration for all subjects was assumed	Limited persons dataset used that will not give generalized result	Cannot differentiate left/right eyes, lower performance	Moderate processing time which leads low frame rate (12-16 FPS)	High processing time which leads low frame rate (12-16 FPS)

doi: 10.32622/ijrat.111202307

III. Proposed Methodology

In our study we adjust the base methodology algorithm and Deep Neural Network (DNN) model. In this study we proposed new Convolution Neural Network (CNN) model as the base method used Multi-Laired Perceptron (MLP) though the CNN is proven effective for image processing than MLP. Figure 4.1-1 shows the proposed framework.

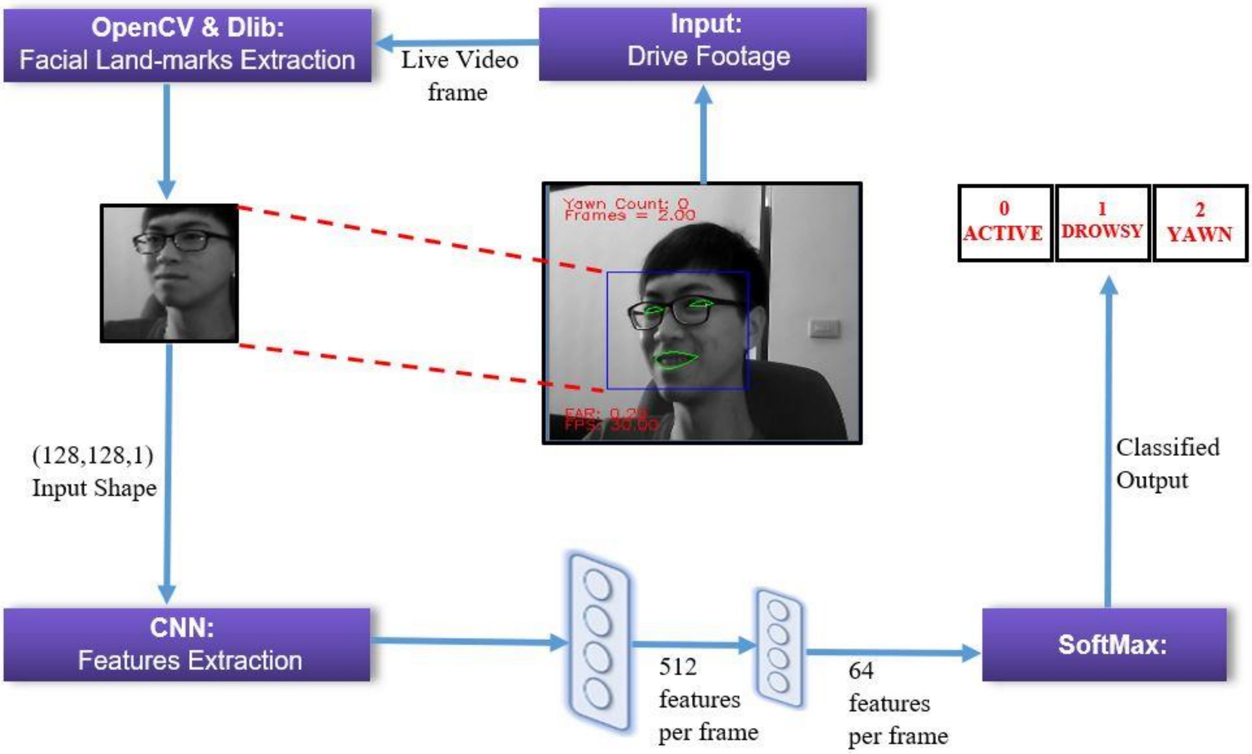


Figure 1: Proposed framework

A. Proposed Algorithm:

Input**:** Facial feature image
Output**:** Trained Model and Predicted labels

1. Loading training and testing data
2. Defining the neural network model
3. Adding layers and dropout to the model:
   a) Average-Pooling2D layer with pool size 2, reduce the input size thus heavy computation is reduced.
   b) The first Conv2D layer with 256 number of kernels. Of 3x3 filter size with Activation function and Max-Pooling2D with pool size 3.
   c) The second Conv2D layer with 128 number of kernels. Of 3x3 filter size with Activation function and Max-Pooling2D with pool size 2.
   d) The first Dense layer with 512 neurons.
   e) The second Dense layer with 64 neurons.
   f) The last Dense layer with activation function SoftMax to get output class label probabilities. The number of neurons is 3
4. Determine consecutive drowsy state by predicting class labels.

   The overall framework comprises of steps as illustrated in Figure 4.1-1. Its step-by-step process in which the 1st step is to extracting facial feature from NTHU [12] dataset including 22 subjects with different ethnicity. In the 2nd step only required facial feature is extracted by Dlib [13] library and reshape the shape in size of 128x128 by OpenCV. 3rd step is to feed reshaped image to the trained model as an input and the model predict the class that image belongs. The 4th step is to determine the consecutive drowsy state i.e., 20 frames that predicted as drowsy. Trained model has three classes "Active", "Drowsy" and "Yawning" and predict labels of drowsiness level.

B. Input Facial Features:

From dataset we extract images frame by frame from subject's videos, by Dlib we determine the facial landmark and by OpenCV we reshape the extracted image. The purpose of reshaping image is that the image contains unnecessary data also so by reshaping image we eliminate the unwanted data.

doi: 10.32622/ijrat.111202307

IV. Implementation

A. Dataset and Preprocessing:

A very important thing for this research is driving condition-based video which is dataset for this research. A standard dataset provided by National Tsing Hua University (NTHU) [12] Driver Drowsiness Detection Dataset is used for this research; this standard dataset is also used by previous research works.

The entire component of NTHU dataset such as testing and training data comprise of 22 subject Figure 5.1-1 shows different ethnicities, data contains videos of different light condition. Video is recorded by Infrared (IR) camera and each videos resolution is 640x480 in AVI format. At first step, the video is extracted from the NTHU dataset, we use 18 subjects for training and 4 subjects are used for evaluation.

By OpenCV we extract image frame by frame at 30 Frame Per Second (FPS) as the video has same frame rate. By using Dlib we extract the facial feature frame and by OpenCV we reshape the image with the size of 128x128. Different images are stored according to its class. Figure 5.1-2 shows the extraction process of facial feature.

Figure 2: NTHU dataset including 22 subjects with different of ethnicities.

doi: 10.32622/ijrat.111202307

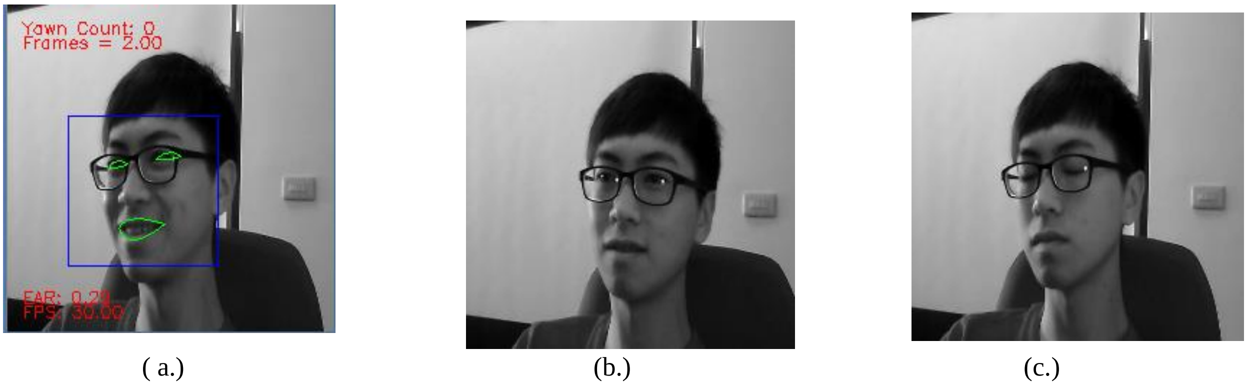


Figure 3: Facial features extraction a.) shows capturing image(frame) from video using OpenCV and b.) and c.) is the classified images one is drowsy and normal behavior.

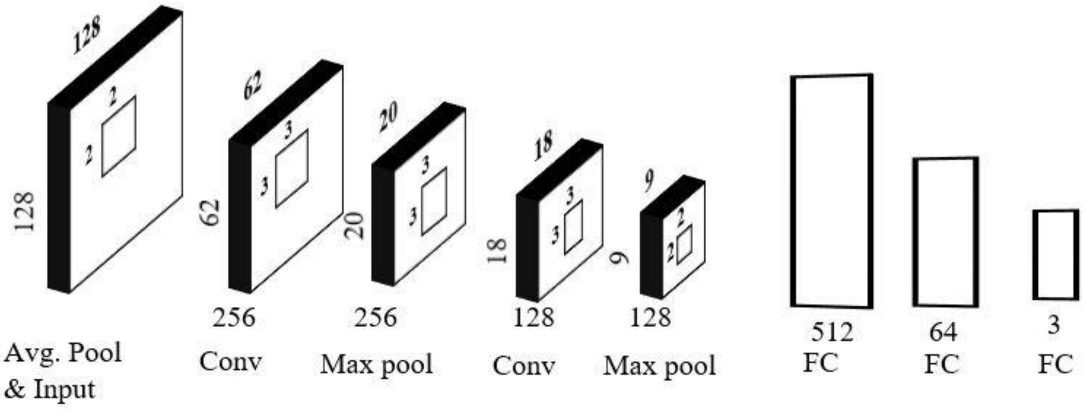


Figure 4: Model architecture

B. Training Deep Learning Model:

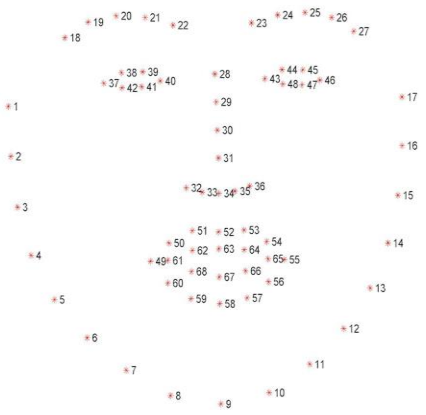
The deep learning model as showed in Figure 5.1-3 is implemented by suing TensorFlow Keras. The model is comprised of two convolution Layer and three Dense layers. As the model's first layer is Average pooling layer, is used purposely because to reduce the input feature size to reduce heavy computation. As the large number of images required more memory for computation so at the very first, we use Average pooling layer that cut down the input image in half. Second reason for use of Average pooling layer is that the large number of images cannot be scale down so the batch-by-batch Average pooling can reduce the input size.

V. EXPERIMENTAL RESULTS

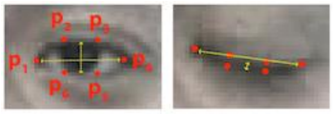
A. Facial Feature extractions:

Though the Dilib facial recognition model is powerful enough and accurate on some data, but we have to face a problem to extract facial expression images. We have to change the standard threshold to achieve accurate images. The threshold for eve aspect ratio (EAR) for open or close eye is 0.4 but, in some videos, we have to varies to 0.25 to 0.45 as some different ethnic group have different facial features.

doi: 10.32622/ijrat.111202307



(a)



$$\text{EAR} = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\|p_1 - p_4\|}$$

b)

Figure 5: a.) Facial landmarks and b.) eye aspect ratio (EAR)

B. Training of Model:

Ones the images are extracted from the dataset we had to arrange the images belonging to its class. Then model is prepared according to input size and how learnable parameter reacts on input images.

Model is also train using TensorBoard TensorFlow's visualization toolkit to track and visualizing matrices. We have train two different models to achieve suitable model for training. Figure 6.1-2 shows the summary of best suitable model.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
average_pooling2d (AveragePo (None, 64, 64, 1)         0
_________________________________________________________________
conv2d (Conv2D)              (None, 62, 62, 256)       2560
_________________________________________________________________
activation (Activation)      (None, 62, 62, 256)       0
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 20, 20, 256)       0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 18, 18, 128)       295040
_________________________________________________________________
activation_1 (Activation)    (None, 18, 18, 128)       0
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 9, 9, 128)         0
_________________________________________________________________
flatten (Flatten)            (None, 10368)             0
_________________________________________________________________
dense (Dense)                (None, 512)               5308928
_________________________________________________________________
dense_1 (Dense)              (None, 64)                32832
_________________________________________________________________
dense_2 (Dense)              (None, 3)                 195
=================================================================
Total params: 5,639,555
Trainable params: 5,639,555
Non-trainable params: 0
_________________________________________________________________
```

Figure 6: Model summary

doi: 10.32622/ijrat.111202307

C. Hardware Configuration:

GPU: NVIDIA GeForce GTX 1060 with 6GB dedicated VRAM.
CPU: Intel Core i5-8300H 8th generation @2.3GHz
RAM: ~24 GB Available
Disk: ~320 GB Available

D. Experimental Result:

By training different models, we choose the best model out of other models and gives the 91.02% accuracy. Figure 6.1-3 shows the Validation accuracy over the Validation loss. Best model is trained over 318 epoch and can be embed with Long-Short Memory (LSTM) model for better learning.

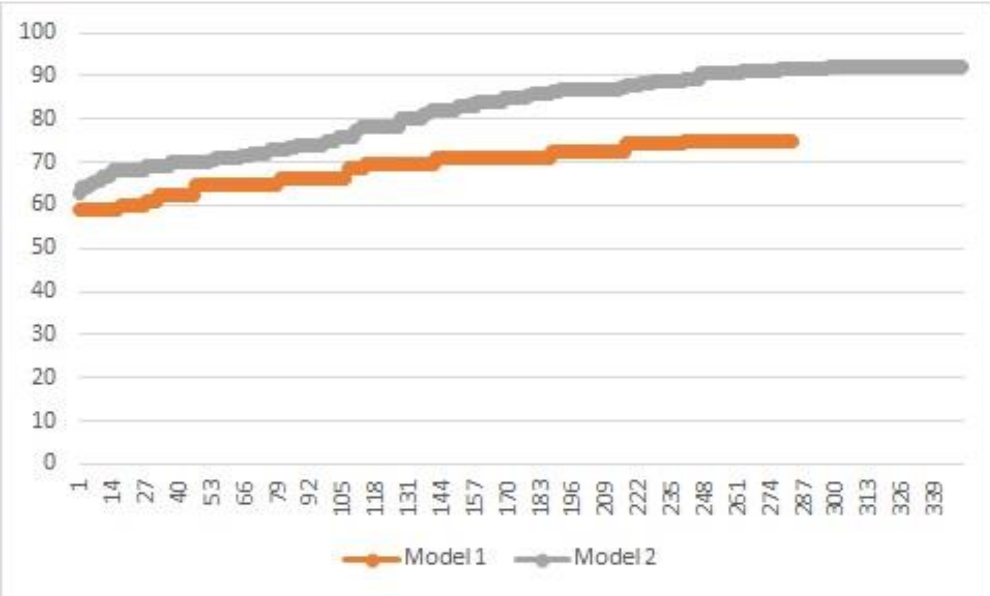


Figure 7: Validation accuracy

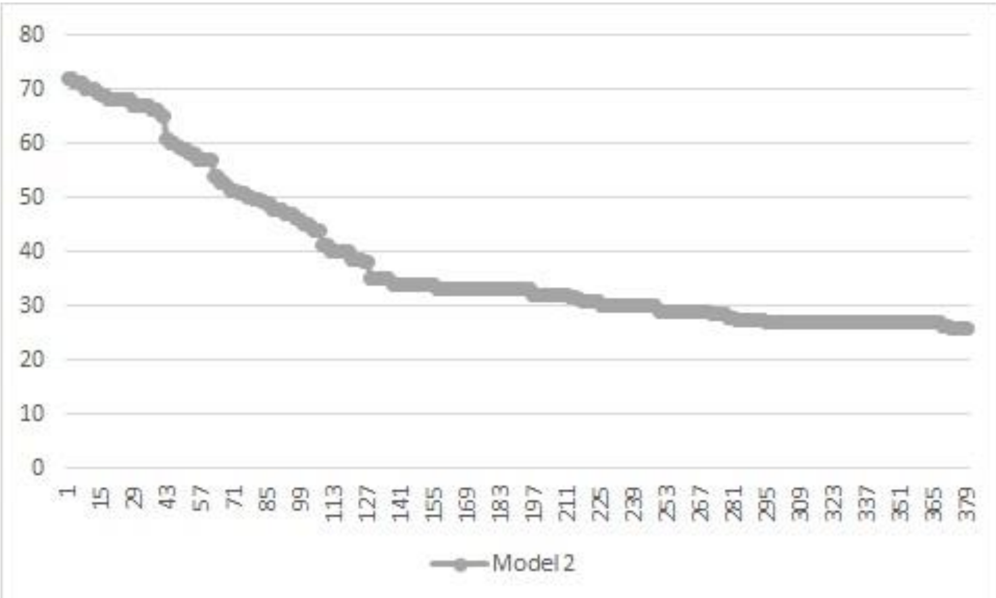


Figure: 8 validation loss over step

doi: 10.32622/ijrat.111202307

VI. Conclusion

From previous work we conclude that the performance and accuracy of system can be improve by changing in algorithm and Deep model of base approach. And by manually selecting the input data can also be improve the overall performance of the system. Eliminating and sorting the wrong input data manually which was not detected by Dlib, reflect in accuracy and performance of the model.

References

[1] Data from MORTH, Government of India, available at https://morth.nic.in/sites/default/files/RA_Uploading.pdf
[2] Drowsy Driving NHTSA reports. (2018, January 08). Retrieved from https://www.nhtsa.gov/risky-driving/drowsy-driving.
[3] Global status report on road safety 2013: Supporting a decade of action: Summary (in World Health Organization, 2013, W. H. Organization.)
[4] MERRILL M. MITLER2, KRISHNAREDDY S. GUJAVARTY, and CARL P. BROWMAN "Maintenance of wakefulness test: a polysomnographic technique for evaluation treatment efficacy in patients with excessive somnolence" Long Island Research Institute and State University of New York at Stony Brook, Long Island, N. Y.11794 (U.S.A.)
[5] MandalapuSaradadev,Dr. Preeti Bajaj "Driver fatigue detection using Mouth and Yawning," International Journal of Computer Science and Network Security, vol. 8, pp. 183-188, 2008.
[6] Charlotte Jacobé de Naurois, Christophe Bourdin, and Clément Bougard "Adapting artificial neural networks to a specific driver enhances detection and prediction of drowsiness," Accident Analysis and Prevention 121, vol. 121, pp. 118-128, 2018.
[7] José Solaz, José Laparra-Hernández, Daniel Bande, Noelia Rodríguez, Sergio Veleff, José Gerpe, and Enrique Medina "Drowsiness detection based on the analysis of breathing rate obtained from real-time image recognition," in 6th Transport Research Arena, poland, April 18-21, 2016.
[8] Rateb Jabbar, Khalifa Al-Khalifa, Mohamed Kharbeche, Wael Alhajyaseen, Mohsen Jafari, and Shan Jiang "Real-time Driver Drowsiness Detection for Android Application Using Deep Neural Networks Techniques," in 9th International Conference on Ambient Systems, Networks, and Technologies (ANT 2018), Porto, Portugal, 2018.
[9] Tereza Soukupová and Jan Cech "Real-TimeEyeBlinkDetectionusingFacialLandmarks," in 21st Computer Vision Winter Workshop, Slovenia, 2016.
[10] Wanghua Deng and Ruoxue Wu. "Real-Time Driver-Drowsiness Detection System Using Facial Features". *IEEE conference*.
[11] Younes Ed-Doughmi, Najlae Idrissi and Youssef Hbali. "Real-Time System for Driver Fatigue Detection Based on a Recurrent Neuronal Network". Journal of Imaging.
[12] Weng CH, Lai YH, Lai SH. "Driver Drowsiness Detection via a Hierarchical Temporal Deep Belief Network". *Asian Conference on ComputerVision Springer* 2016;117-133.
[13] "Dlib C++ toolkit" accessed on 09-July-2018 http://dlib.net/

AUTHORS PROFILE

Mr. Prashant Gosai, who has more than 10 years' Experience in IT sector and working as a Project Manager in reputed Government Organization. He did Bachelor of Engineering and pursuing the Master of Engineering from Gokul University of Gujarat. During professional life he had worked more than 20 projects in different languages i.e JAVA, ASP.Net, Spring Framework, Angular JS, React JS. He also contributed his technical skill on Geographical Information System and developed the different tools based on GIS system.ie. Buffer Tool, Query Tool etc.

doi: 10.32622/ijrat.111202307