

Driver Drowsiness Detection System Using Machine Learning

Rahul Thakur, Shivam, Shubham Raj, Subhanshu Pandey

Delhi Technological University, New Delhi, India

rahulthakur@dtu.ac.in

Abstract- Drowsy driving is one of the leading causes of traffic accidents all over the world. Driving in a monotonous manner for an extended amount of time without stopping causes tiredness and catastrophic accidents. Drowsiness has the potential to ruin many people's lives. As a result, a real-time system that is simple to create and configure for early and accurate sleepiness detection is required. In this study, a real-time vision-based system called Driver Drowsiness Detection System has been developed utilizing machine learning. In this study, the Haar Cascade classifier was used to recognize the driver's face characteristics and functions present in OpenCV library to detect the region of the face. The following step is to examine the open/close state of the eyes, followed by sluggishness depending on the sequence of ocular conditions. The non-intrusive and cost-effective nature of this vision-based driver tiredness detection is its distinguishing attribute.

Keywords— Face Detection, Eye Tracking, Eye Detection, Haar Classifier, Template Matching, Driver Drowsiness Detection.

1. INTRODUCTION

Distracted driving is a dangerous and hazardous fatal flaw of most of the drivers on road. This mostly happens when a driver has not had enough sleep which is also known as drowsy driving, but can also be a cause of just being ignorant or insincere for a while.

Distracted Driver is a major cause for concern as being alert and cautious on roads is of utmost importance. Negligence on the road can end up in altering one's life completely by death of a close person, or someone due to your own negligence. About a lakh accident were reported by police according to NHTSA[1-4] each year due to drowsy driving. The reaction time of the driver gets dampened due to distractions.

Getting distracted and falling asleep suddenly is a common mistake and can happen to anyone at any time due to monotony of the driving or just tiredness. It can be prevented by taking precautions by the drivers themselves. But often people misjudge even the fact about their level of fatigue. Hence developing an efficient driver distraction detecting system is of prime importance to prevent road accidents and mishaps. As part of our major project, we explored the various techniques in use to detect driver getting distracted and developed an efficient distracted driver detection system.

According to the National Sleep Foundation[5-8], about 50% of the adult population that are eligible to drive admitted to driving while feeling drowsy. Often people get behind wheels even after feeling tired and sleepy and could easily fall asleep behind the wheels. Driving under influence can be easily deciphered by tests. However, the level of fatigue can't be measured as simply as that. A majority of accidents occur due to negligence of

the driver. The need or greed for a little extra money pushes drivers to drive even after 5 being fatigued and sleepy which is hazardous for not only the driver but even the passengers and other people on the road [9-14].

Our aim is to create a method that will detect the drowsiness and at that time only warn the driver by making some kind of noise and also to slow the speed of the car. It is thought that by monitoring the eyes, It is possible to recognize weariness early enough to alert an automobile collision.

General architecture of drowsiness detection is shown in Fig.1

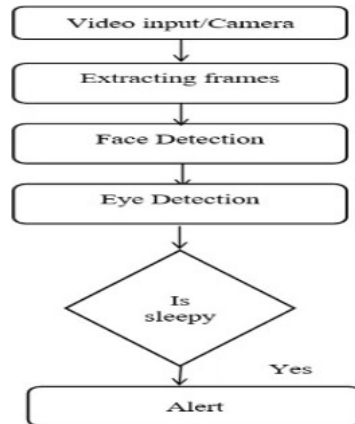


FIG 1: General architecture of drowsiness detection

2. Related Work

Various researchers used different techniques to detect or classify images for preventing distraction while driving. Below are the following techniques which can be used for this purpose: -

Eye Blinking Technique[15]: It is a pre-requisite condition of sleepiness detection. The score is used to determine whether or not the driver is drowsy [16].

Yawning Technique[17]: One of the signs of exhaustion is yawning. It is believed that yawning is modeled as a wide open mouth. When you yawn, your mouth is wider than when you speak. Yawning[17] can be detected by monitoring the face and then tracking the mouth. The speed at which the mouth opens and the degree of change in the mouth contour area are used to identify yawns on paper, and when yawn is detected the driver is warned.

Vehicle Movement: In this technique a close eye is kept on the motion of the vehicle[12]. So for taking the reading of the motion of the vehicle we connect the various components of the vehicle by the sensor like connecting accelerator pedal and steering wheel and then analyzing the data to assess levels of sleepiness[18].

ECG, EOG, and head motion[19][20]: This is an example of an invasive approach. Some of these decisions required drivers to wear helmets while driving. Another approach used heart rate sensors on the steering wheel and seat backs. Most of the time, these were also unreliable. As a result, the general public has not been warm to this strategy. Proposed System

3. Dataset [18]

This paper makes use of the following datasets. Figure 2 shows the image annotation of the proposed dataset.

The following properties were annotated in the dataset (in the following order):

- subject ID**; In the dataset, we conducted surveys from 37 multiple parties (33 men and 4 women)
- gender** [0 - man, 1 - woman]; Here the image of different gender is used mainly male and female.
- glasses** [0 - no, 1 - yes]; Information on whether the eye picture contains eyeglasses is also supplied for each image (with and without the glasses)
- eye state** [0 - closed, 1 - open]; This property keeps track of two different eye states (open, close)
- reflections** [0 - none, 1 - small, 2 - big]; Based on the size of the reflections, we identified three different sorts of reflection states (none, small, and big reflections)
- lighting conditions** [0 - bad, 1 - good]; Each image has two states based on the amount of light available during video shooting (poor, good).
- sensor ID** [01 - RealSense, 02 - IDS, 03 - Aptina]; Photos from three different sensors have now been added to the collection (IDS Imaging sensor with 1280 x1024 resolution, Aptina sensor, Intel RealSense)



FIG 2: Image annotations of the proposed dataset

4. DEEP CNN BASED DROWSINESS DETECTION MODEL

1. Algorithm Proposed

- Viola Jones[10] method is used to detect face in the required section or in the model predicted area, it is then passed as an input to the Viola-Jones algorithm[10].
- The recognition of face[19] is the first step and then algorithm extract the eyes using the Viola jones[10] eye detection approach and send it to CNN.
- To extract the deep feature a CNN with 4 convolutional layer is used, which later on treat as an input to a fully connected layer.

CNN's soft max layer categorises photographs as sleepy or non-sleepy. There are three stages to the recommended paradigm.

- Pre-processing stage
- Extraction of characteristics
- Deep Learning-based CNN Classifier

Proposed system architecture and transfer learning have been shown in Figure 3 and 4.

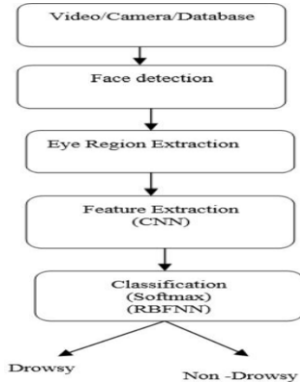


FIG 3: Proposed system architecture

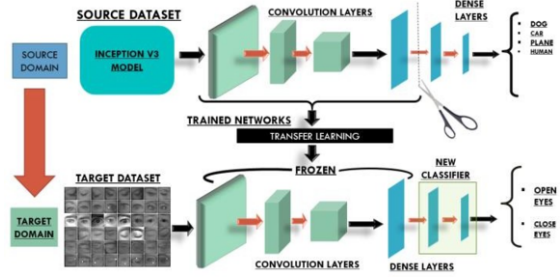


FIG 4: Transfer Learning

2. CNN Architecture

A. Convolution Layer: Convolution is a method of combining two sources of data in an orderly manner; it is a transformation from one function to another. The main and very important task of this layer is to make the image blurry and sharpened, it is not confined to this only it has many uses. CNNs is used to set up the communication patterns between neurons in adjacent layers as shown in Fig. 5A. Features like edges are available or not in the input provided image are detected by CNNs which is having Kernels/Filters. Four primary operations:

- Convolution
- Non Linearity (ReLU)
- Pooling or Sub Sampling
- Classification

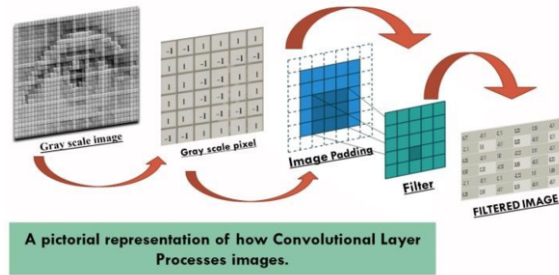
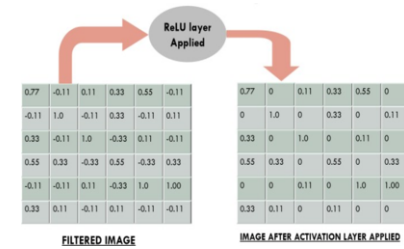


FIG 5(A): Convolution Layer

B. Activation Layer: Negative values in the filtered image on this layer are removed and replaced with zeros to avoid summing the values to zero. The activation function is one of the most important aspect of the neural network architecture. Activation Function present in hidden layer performs the task that how much efficient is our model and it will provide good results at real cases or not. This function or layer used in output layer helps in determining the model's capability of producing best prediction. Figure 5(B) and Figure 5(C) demonstrates the same.

x	$f(x)=x$	$F(x)$
-3	$f(-3)=0$	0
-5	$f(-5)=0$	0
3	$f(3)=3$	3
5	$f(5)=5$	5

FIG 5(B): Activation Layer Table**FIG 5(C):** Removing Negative Values

C. Pooling Layer: In this layer, the resultant image after applying the activation layer will be shrunk to smaller image.

Steps involved in this process: Take a window size (usually 2 or 3)

- Take a stride
- Take your window across your filtered image
- Take the maximum value from each window

D. Dense Layer: Dense layer is very fundamental layer where every neuron plays an important role and each neuron collects input from all the neurons of previous layer therefore it is known as Dense layer. It is used in classification of images based on CNN. In contrast, the dense layer performs a matrix-vector multiplication.

3. Classification and Feature Extraction

A variant of dimensionality reduction is feature extraction. The most important parts of the image are represented by feature vectors. In this article, information from images of the eye area is extracted using a CNN. This is because each sleepy image requires a feature vector that can be compared with existing features in the database to determine if the image is sleepy or not. Preprocessing is usually required since CNN demands definite pixel/size value to process. Pre-processing is the process of collecting important frames from video and storing them in a database based on temporal variations. Feature vectors are formed from these stored pictures in CNN's convolution layers.

These feature vectors are then developed to determine if the driver is jaded or not. The convolution layer uses kernels (filters), each with a width, depth, and height. Feature map is produced by the scalar product/multiplication of Kernels and the local areas of the input image. In order to speed up calculations, CNN employs pooling layers to cut the size of the feature maps (Max or Average). This layer separates the input picture into several portions, with actions carried out separately on each section. This layer divides the input picture into many sections, with actions performed on each part independently. Max Pooling chooses the highest value for each region and inserts it in the output at the appropriate location. ReLU is a nonlinear layer (Rectified Linear Units). Any of the values in the input data are multiplied by the max function, and the ReLU

layer turns all negative values to zero.

$$f(x)=\max(0,x) \quad (1)$$

From the activations used for categorization, the fully-connected layers produced 463 class ratings.

The suggested approach employs four convolutional layers and one fully linked layer. Convolution layer.1 receives key pictures that are 128 X 128 pixels in size (Conv2d 1). The input image is convolved using 84 3*3 filters in Conv2d 1. The architecture includes convolution, batch normalisation, non-linear transformation ReLU, and Max pooling over 2 x 2 cells, followed by 0.25 percent dropout. Conv2d 1 requested that 840 parameters be used. In batch normalisation, 336 parameters are utilized. The output of convolution layer 1 is sent to convolution layer 2. (Conv2d). The input is convolved using 128 5x5 filters in Conv2d 2. With stride 2 and 0.25 percent dropout over 2 2 cells, convolution, batch normalisation, non-linear transformation ReLU, and MaxPooling were used. A total of.268927 parameters were required by Conv2d 2. Batch normalisation 2 necessitates the use of 512 parameters. Convolution layer 3 receives the output of convolution layer 2. (Conv2d 3). The input is convolved in Conv2d 3 with 256 5*5 filters. Conv2d 3 required 819456 parameters after convolution, batch normalisation, and non-linear transformation. 0.25 percent dropout after ReLU, MaxPooling over 2 2 cells with stride 2. 1024 parameters are required for batch normalisation 3. Conv2d 4's convolution layer-3 output is sent on to Conv2d 4's convolution layer-4. In Conv2d 4, the input is convolved with 512 filters of size 5x5 filters each. After convolution, Batch Normalization, and non-linear transformation, Conv2d 4 required 3277312 parameters. 0.25 percent dropout, ReLU, and Max Pooling across 2 x 2 cells with stride 2. Batch normalisation 4 needed a total of 2048 parameters. 8388864 parameters are required for the dense 1 completely connected layer. A CNN model with 12,757,874 trainable parameters was proposed by the researchers. Because the classifier's output is two states, the output layer only has two outputs. For optimization, the Adam method is applied. For classification, the softmax classifier is utilised. The deep features derived from input eye images are the 256 outputs of the fully linked layer in our proposed CNN system. The proposed deep CNN has been shown in Figure 6.

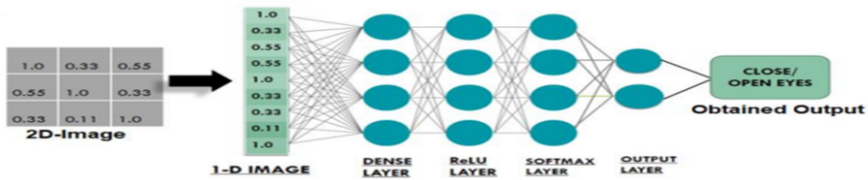


FIG 6: Proposed Deep CNN

5. RESULT AND DISCUSSION

Here, we've carried out two types of experiments. An experiment is done on a previously acquired dataset in the first category. In the second version, the experiment is videotaped. For the first experiment, we made a collection of 2850 photographs. 1450 images out of a total of 2850 are drowsy, while the rest are not. This study uses a total of 1200 images for training, 600 of whom are sleepy and 600 of which are not. A total of 500 photos are

utilized for validation, with 250 of them being asleep and the other 250 being awake. The proposed model achieved a 95.78 percent accuracy on the test dataset, which included 1150 pictures, 550 of which were sleepy and 600 non-drowsy. After the 20 epochs of cycle the model's accuracy is examined and the results are depicted as follows-

Figure 7- The training loss as a function of epochs

Figure 8- The training and validation accuracy against epochs.

Figure 9- Confusion Matrix

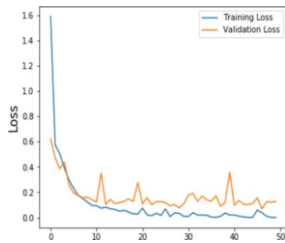


FIG 7: The training and validation losses vs the number of epochs



FIG 8: The training and validation accuracy are measured against the

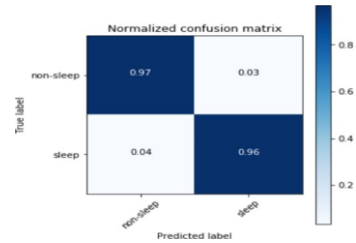


FIG 9: Confusion Matrix

So here, first the videos is acquired by the webcam for processing. After that we will divide video into frames. So here as you can see we are left with picture. As a result, we seek and detect faces in individual frames in the given photos. A new frame will be acquired if no face is recognised. Region of interest will be marked within face if a it is detected. Then, eyes are detected from the region of interest and if eyes are detected and there is no blink then counter will be decremented and a buzzer will sound when it will reach 0.

6. CONCLUSION

So here we have presented a novel approach to determine the tiredness or drowsiness of the driver with the help of their eye states. This determines whether the driver's eye is fatigued, and if it is, an alarm is started. To identify the eye and face region we have used the Viola Jones[10] detection method. And for the training stage we have used stacked deep convolution neural network. And also if the vehicle is running above a threshold speed automatic breaking system will also be applied to slow the speed of the vehicle. The proposed model provide the accurate and best detection of driver drowsiness. This discussion and design for driver alertness is provided which is used to prevent various road accidents caused by drowsiness and also helps in making driver alert or stay awake. This proposed system can be used to provide security as well. So for this we have places a video camera which will record video and simultaneously we change the external condition like lighting, persons, people seated at different angles and also we have use different camera sensor so all this factors help to increase the accuracy of the model which eventually make our mode more robust and precise. SoftMax layer determines in a CNN classifier that the driver is sleepy or non-sleepy with 95.78 percent accuracy at the current time.

REFERENCES

- [1] NCSDR/NHTSA Expert Panel on Driver Fatigue and Sleepiness, "Drowsy driving and automobile crashes," National Highway Traffic Safety Administration (NHTSA) report.
- [2] M. R. Rosekind, "Underestimating the societal costs of impaired alertness: safety, health and productivity risks," *Sleep Medicine*, vol. 6, pp. S21-S25, 2005.
- [3] C. Papageorgiou, A General Framework For Object Detection, In International Conference On Computer Vision, 1989.
- [4] Akerstedt, T., Peters, B., Anund, A., & Kecklund, G. Impaired alertness and performance driving home from the night shift: a driving simulator study. *Journal of Sleep Research*.
- [5] Ji Q. and Yang. X.,: Real-Time Eye, Gaze, and Face Pose Tracking for Monitoring Driver Vigilance. *Real Time Imaging*, Nr. 8, Pg.357-377, Elsevier Science Ltd.
- [6] Stephan, K., Hosking, S., Regan, M., Verdoorn, A., Young, K., & Haworth, N. The relationship between driving performance and the Johns drowsiness scale as measured by the Optalert system. Monash University Accident Research Center.
- [7] Arnedt, J. T., Wilde, G. J. S., Munt, P. W., & Maclean, A. W. Simulated driving performance following prolonged wakefulness and alcohol consumption: separate and combined contributions to impairment. *Journal of Sleep Research*.
- [8] Amodio, A., Ermidoro, M., Maggi, D., Formentin, S., Savaresi, S.M. Automatic detection of driver impairment based on pupillary light reflex. *IEEE Transactions on Intelligent Transportation Systems*.
- [9] Mardi, Z., Ashtiani, S.N., Mikaili, M. EEG-based drowsiness detection for safe driving using chaotic features and statistical tests. *Journal of Medical Signals And Sensors*.
- [10] Viola, P., Jones, M.J. Robust real-time face detection. *International Journal of Computer Vision*.
- [11] VenkataRamiReddy, C., Kishore, K.K., Bhattacharyya, D., Kim, T.H. Multi-feature fusion based facial expression classification using DLBP and DCT. *Int J Software Eng Appl*.
- [12] Huang Nice Quan, "Drowsiness Detection for car assisted driver system using image processing analysis-interfacing with hardware", Faculty of Electrical& Electronics Engineering University Malaysia Pahang.
- [13] Picot, A., Charbonnier, S., Caplier, A. On-Line Detection of drowsiness using brain and visual information. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*.
- [14] Lee, K., Yoon, H., Song, J., Park, K. Convolutional neural network-based classification of driver's emotion during aggressive and smooth driving using multi-modal camera sensors. *Sensors*.
- [15] Yang, J.H., Mao, Z.H., Tijerina, L., Pilutti, T., Coughlin, J.F., Feron, E. Detection of driver fatigue caused by sleep deprivation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*.
- [16] Dr.Suryaprasad J, Sandesh D, Saraswathi V, Swathi D, Manjunath S, "Real Time Drowsy Driver Detection Using Haarcascade Samples", PES Institute ofTechnology- South Campus, Bangalore, India.
- [17] A.N.Shewale, Pranita Chaudhari, "Real Time Driver Drowsiness Detection System", *International Journal of Advanced Research in Electrical, Electronicsand Instrumentation Engineering*.
- [18] Dataset Reference: MRL Eye Dataset | MRL (vsb.cz)
- [19] Lin, S. H., Kung, S. Y., & Lin, L. J. (1997). "Face recognition/detection by probabilistic decision-based neural network." *IEEE Transactions on Neural Networks*,
- [20] Nikhil R Pal, Chien-Yao Chuang, Li-Wei Ko, Chih-Feng Chao, Tzyy- Ping Jung, Sheng-Fu Lieng, Chin-Teng Lin, "EEG-Based Subject- and Session-independent Drowsiness Detection: An Unsupervised Approach".