

## Real Time Detection of Driver Cognitive Distraction Using Machine Learning Classifiers

Dr.A.Sumaiya Begum<sup>1</sup>, Ms.P.Poonkuzhali<sup>2</sup>, Jalashree J<sup>3</sup>, Damini K<sup>4</sup>, Deepa Prabha S M<sup>5</sup>,  
<sup>1,2</sup>Associate Professor/Department of ECE, R.M.D Engineering College,  
Tamilnadu 601206

[Asb.ece@rmd.ac.in](mailto:Asb.ece@rmd.ac.in)<sup>1</sup>, [poonkuzhali.ece@rmd.ac.in](mailto:poonkuzhali.ece@rmd.ac.in)<sup>2</sup>  
<sup>4,5,6</sup>UG Students/Department of ECE, R.M.D Engineering College, Tamilnadu 601206

**Abstract**— According to researchers, 5 to 25% of car accidents occur due to the driver's distractions. Distracted driving is an activity that makes the driver lose his concentration in driving. Several researches have identified that distraction is of three types, which include visual distractions, manual distractions and cognitive distractions. Our approach deals with the cognitive distractions like eye closure, blink rate and yawning using machine learning and deep learning concepts which helps in classifying the image of driver into the different classes and predict if the driver is in distracted state or not. Moreover, the aim of this paper is to detect the state of the driver and avoid road accidents and to ensure safer roads.

**Keywords**---Driver drowsiness; eye detection; yawn detection; blink pattern; fatigue

### I. INTRODUCTION

Drowsiness is a state of near sleep, where the person is desiring for sleep. It has two distinguished meanings, one referring the usual state before falling asleep and the other is chronic condition that is independent of human's daily pattern. Sleepiness might be dangerous when performing tasks that requires constant concentration throughout the task, such as driving a vehicle, they will experience drowsiness and this leads to abnormal rise of road accident.

The main aim of the paper is to build a simulation of real-time detection of driver's drowsiness. The aim is to design a system that will precisely monitor the closed or open state of the driver's eyes and mouth. When the eyes is being monitored continuously, the driver's state of drowsiness can be identified in early stage and driver can be warned about his distraction by which the accidents can be avoided. Yawning detection is a method for estimating the driver's fatigue. When a person is tired, they keep yawning to ensure if there is enough oxygen for the brain consumption before going to sleeping state. The fatigue and drowsiness can be detected by observing a series of images of face and identifying if the eyes and mouth are in opened or closed states and their corresponding duration in that state.

The study of face images is a popular research area with applications such as face recognition, and human identification and tracking for security systems. This project aims on locating the position of the eyes and mouth by focusing on the entire image of the face, and locating the position of eyes and mouth, by applying the existing methods in image processing algorithm. Once the eyes and mouth are located, the system is designed such that it determines whether the eyes and mouth are opened or closed, and detect driver's drowsiness and fatigue.

In section II, existing system are discussed. In section III, proposed system are discussed. In section IV, discuss about software specifications. In section V, discuss about conclusion of this project.

### II. RELATED WORK

A number of assessments on the theme of driver distraction detection methods were done before as projects reports or, as research papers using Physiological distraction detectors.

Multimodal drowsiness detection using machine learning algorithms- Y. Kim, M. Yeo, I. Sohn, and C. Park[1]

A one-channel electrocardiogram (ECG) and a single-channel electroencephalogram (EEG) were simultaneously recorded. The ECG and EEG features were extracted and fed into machine learning, random forest, multilayer perceptron, and support vector machine algorithms. Various feature parameters were utilized to train the algorithms, and random forest yielded the best performance at about 90% accuracy, precision, and recall, with 10-second epochs in the ECG and EEG.

A hybrid approach to detect driver drowsiness utilizing physiological signals to improve system performance and wearability- M. Awais, N. Badruddin, and M. Driberg [2]

The study measures differences between the alert and drowsy states from physiological data collected from 22 healthy subjects in a driving simulator-based study. A monotonous driving environment is used to induce drowsiness in the participants. Various time and frequency domain feature were extracted from EEG including time domain statistical descriptors, complexity measures and power spectral measures. Features extracted from the ECG signal included heart rate (HR) and heart rate variability (HRV), including low frequency (LF), high frequency (HF) and LF/HF ratio. Furthermore, subjective sleepiness scale is also assessed to study its relationship with drowsiness. We used paired t-tests to select only statistically significant features ( $p < 0.05$ ), that can differentiate between the alert and drowsy states effectively. Significant features of both modalities (EEG and ECG) are then combined to investigate the improvement in performance using support vector machine (SVM) classifier

Real-time driver drowsiness detection for embedded system using model compression of deep neural networks- B. Reddy, Y.-H. Kim, S. Yun, C. Seo, and J. Jang[3]

Accidents occur because of a single moment of negligence, thus driver monitoring system which works in real-time is necessary. This detector should be deployable to an embedded device and perform at high accuracy. In this paper, a novel approach towards real-time drowsiness detection based on deep learning which can be implemented on a low cost embedded board and performs with a high accuracy is proposed. Main contribution of our paper is compression of heavy baseline model to a light weight model deployable to an embedded board. Moreover, minimized network structure was designed based on facial landmark input to recognize whether driver is drowsy or not. The proposed model achieved an accuracy of 89.5% on 3-class classification and speed of 14.9 frames per second (FPS) on Jetson TK1.

Driver fatigue detection based on eye state recognition - F. Zhang, J. Su, L. Geng, and Z. Xiao[4]

An effective fatigue detection method based on eye status with pupil and iris segmentation. The segmented feature map can guide the detection to focus on pupil and iris. A streamlined network, consisting of a segmentation network and a decision network, is designed, which greatly improves the accuracy and generalization of eye openness estimation. Specifically, the segmentation network that uses light U-Net structure performs a pixel-level classification on the eye images, which can accurately extract pupil and iris features from the video's images. Then, the extracted feature map is used to guide the decision network to estimate eye openness. Finally, the detection method is test by the National Tsing Hua University Drowsy Driver Detection (NTHU-DDD) Video Dataset and the precision of fatigue detection achieves 96.72%. Experimental results demonstrate that the proposed method can accurately detect the driver fatigue in-time and possesses superior accuracy over the state-of-the-art techniques.

A review on EEG-based automatic sleepiness detection systems for drive- R.P. Balandong, R. F. Ahmad, M. N. Mohamad Saad, and A. S. Malik[5] Electroencephalography-based sleepiness detection system (ESDS) is a brain-computer interface that evaluates a driver's sleepiness level directly from cerebral activity. The goals of ESDS research are to estimate and produce a timely warning to prevent declines in performance efficiency and to inhibit sleepiness-related accidents. We first, review different types of measures used in sleepiness detection systems (SDSs) and

presents this advantages and drawbacks. Second, the review includes several techniques proposed in ESDSs to optimize the number of EEG electrodes, increasing the sleepiness level resolution and incorporation of circadian information. Finally, the review discusses future direction that can be considered in the development of ESDS.

I. METHODOLOGY

In this proposed system, the CNN ResNet-50 is used for image classification to warn the fatigue condition of the driver. In this the camera takes in the video as input and Viola Jones algorithm is implemented to extract features. These obtained features are matched against the trained database that is trained using ResNet-50. The segregated feature is checked for the percentage accurate of matching and the evaluated result is displayed. If the result is adverse, it gives a beep sound. Following that the driver is made alert of his drowsy condition, thus avoiding the accident caused due to this. The method's success is totally relied on the training of the Convolutional Neural Network layers. Further work can be made by including the heart-rate and other distraction detectors so that a safer road is made.

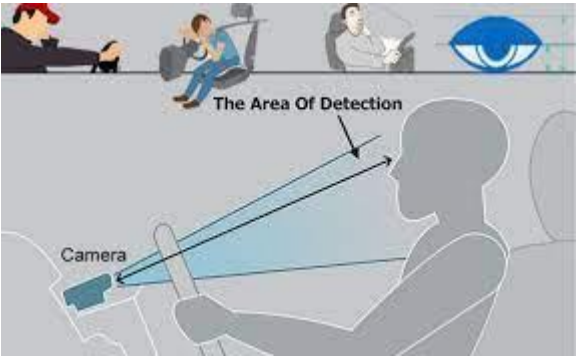


Fig 2: Actual setup

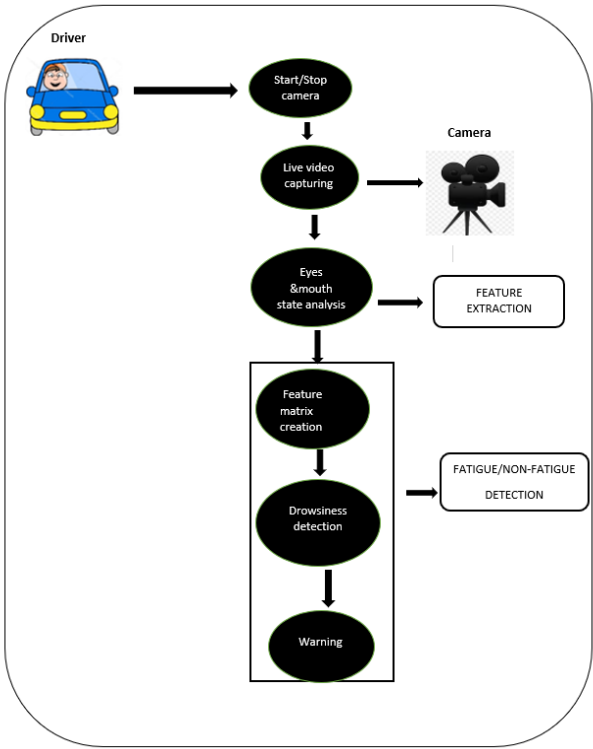


Fig1:Block Diagram

#### A. *Software Requirements*

- MATLAB R2021a
- MATLAB R2021a image processing tool
- CNN library
- Viola Jones Feature Detector
- CNN residual network

#### B. *Hardware Requirements*

- Laptop with basic hardware with integrated webcam

### II. SOFTWARE PROCESS

#### A. *MATLAB*

MATLAB is a high-performance language for technical computing. It coordinates visualization, computation and programming in an easy-to-use environment where problems and solutions are expressed in famous mathematical notation. Typical applications include:

- For computation and Mathematics
- Algorithm development
- Modeling, simulation, signal processing and prototyping
- Statistics and machine learning (ML) □
- □ engineering graphics and Scientific management
- Application development that included Graphical User Interface building

The basic element of MATLAB is an array sequence which does not require any dimensioning. MATLAB is also an interactive system that allows one to solve as many technical computational problems, mainly those of vector and matrix type, in a fraction of second by writing a program in non-interactive scalar language like Fortran or C.

#### B. *Image Acquisition*

It is defined as the act of extracting an image from the source (webcam), usually a hardware source for processing. It is the foremost stage in the workflow because, without an input, no preprocessing or acquisition is possible. MATLAB allows to acquire image and video data from hardware and import it directly into for visualization and processing. Image Acquisition Toolbox provides functions and blocks to connect industrial and scientific cameras to MATLAB and Simulink. It helps simplify the acquisition process by providing a consistent interface across operating systems, hardware devices and vendors

#### C. *Preprocessing*

Improving the image's attributes and removing undesired deformities in readiness for potential yield. Image processing is the use of algorithmic logics to accomplish image processing on digitalized images in computer science. It enables a wider selection of algorithms to be applied to analyze the data (image) and can ignore problems such as unwanted noise and signal deterioration

AHE (adaptive histogram equalisation) is a PC picture handling technique that improves picture differentiation. It differs from standard histogram levelling in that the flexible technique processes many histograms, each corresponding to a distinct region of the image, and uses them to redistribute the image's brightness estimates. It is useful for enhancing neighborhood differentiation and updating the definitions of edges in each district of a picture in this way. Nonetheless, AHE has a tendency to exaggerate clamour in a picture's normally homogeneous districts of picture

Ordinary histogram equalization uses the similar transformation derived from the image histogram to transform all pixels. This works well when the distribution of pixel values is same throughout the image input. However, when the image contains regions that are significantly lighter or darker than most of the image, the contrast in that region will not be sufficiently enhanced.

Ordinary histogram equalization transforms all pixels using a similar transformation derived from the image histogram. This works well when the pixel value distribution is consistent in the image input. The contrast in that area will not be properly enhanced if the image includes regions that are noticeably lighter or darker than the rest of the image. The transformation functions are derived from the histograms in the same way that the ordinary histogram equalization is done: The transform function is proportional to the cumulative distribution function (CDF) of pixel values in the immediate vicinity.

Since their neighborhood does not lie entirely within the image, pixels near the image boundary must be viewed differently and more carefully. This can be solved by mirroring pixel lines and columns with respect to the image boundary to expand the image. Copying the pixel lines on the boundary isn't a good idea because it'll result in a neighborhood histogram with a lot of peaks.

#### *D. FeatureExtraction*

It's a form of dimensionality reduction in which the most interesting sections of an image are efficiently represented as a compact feature vector. When image sizes are large and a reduced feature representation is needed to complete tasks like image matching and retrieval quickly, this method is useful

##### *I. Viola-JonesAlgorithm*

The sums of image pixels inside rectangular areas are used in all of the detection framework's functionality. Viola - Jones' features, on the other hand, are typically more complex since they all depend on more than one rectangular field. Using an image representation known as the integral image, is a data structure and algorithm for quickly and efficiently generating the sum of values in a rectangular subset of a grid. A summed area table, also called an integral image, is a data structure and algorithm for generating the number of values in a rectangular subset of a grid quickly and efficiently.

Rectangular features can be measured in real time, giving them an advantage over their more advanced relatives in terms of speed.

The learning process generates powerful classifiers, but the assessment isn't quick enough to run in real time.. A cascade ordering is used in order of their complexity, where each classifier stronger than its predecessor is trained only on the select candidates which have passed the classifiers prior to them. If at any stage in the cascade a classifier rejects the sub-window under inspection, no further processing is performed and continues on searching the next sub-window.

#### **Face detection:**

This function takes one frame at a time from the frames given by the frame grabber and attempts to detect the face in each frame. This is accomplished by employing a collection of pre-determined Haar-cascade samples.

#### **Eyes detection:**

If no faces are detected after the face detection function, the Viola Jones algorithm's eyes detection function attempts to detect the eyes. If there aren't any faces, check for eyes.

#### **Mouth detection:**

The aim of detecting mouth is to find the drowsiness symptom of yawning. The cascade object detector, which employs the Viola-Jones algorithm, was used to detect the mouth by detecting

objects in a rectangle shape.

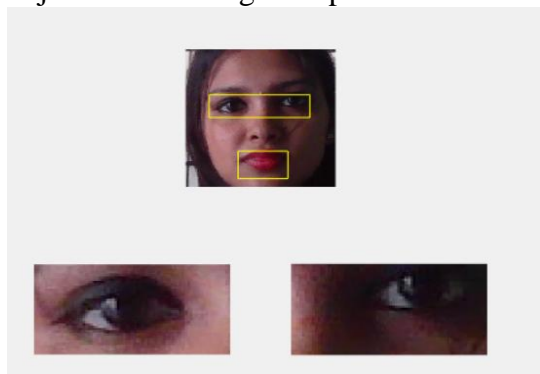


Fig 3: Features being Extracted using Viola Jones algorithm

### E. Image Classifier

For image classification, we use ResNet-50 for Image classification. ResNet-50 is a convolutional neural network which is 50 layers deep. By loading pretrained version of the network, trained on more than a million of images from the ImageNet database. The pretrained network can classify images into 1000 object categories. As a result, the network has learned rich and wider range of feature representations. As a result, the network has learned rich and wider range of feature representations.

BLOCKS IN RESNET 50:

#### 1. Predict

. The Predict block uses the trained network defined by the block parameter to predict responses for the data at the input. This block helps you to load a pretrained network from a MAT-file or a MATLAB feature into a Simulink model. There are input and output ports on it.

PARAMETER:

- Network — database for trained network
- File path — MAT-file containing trained and categorized network
- Predictions — Output predicted responses
- Activation — Output network activations for a specific type of layer

#### 2. Image classifier

The Image Classifier block anticipates the class labels for the incoming data with the help of the trained network framed through these block parameters. The block allows piling of a pretrained network onto a Simulink model from a MAT-file or from a MATLAB function. This has also got ports for input and output.

The code is tested with some examples and here are the outcomes

#### A. Test Samples

Test ID	Test condition	Right eye	Left eye	Mouth	Expected result	Actual result
T01	STRAIGHT FACE, GOOD LIGHT	OPEN	OPEN	CLOSE	NON-FATIGUE	NON-FATIGUE
T02	STRAIGHT FACE, GOOD LIGHT	OPEN	CLOSE	OPEN	FATIGUE	FATIGUE
T03	STRAIGHT FACE, GOOD LIGHT	CLOSE	OPEN	OPEN	FATIGUE	FATIGUE
T04	STRAIGHT FACE, GOOD LIGHT	CLOSE	CLOSE	OPEN	FATIGUE	FATIGUE

Note: Testing is performed manually

## B. Outcomes



Fig 4: Test samples

## II. Modelling

After this algorithm has been developed during the experimentation part, it needs improvement for the system to meet the purpose of this project. Figure 5 below shows the flow process of the improved algorithm. This algorithm successfully detects the features like mouth and eyes in the video; hence the result can be obtained.

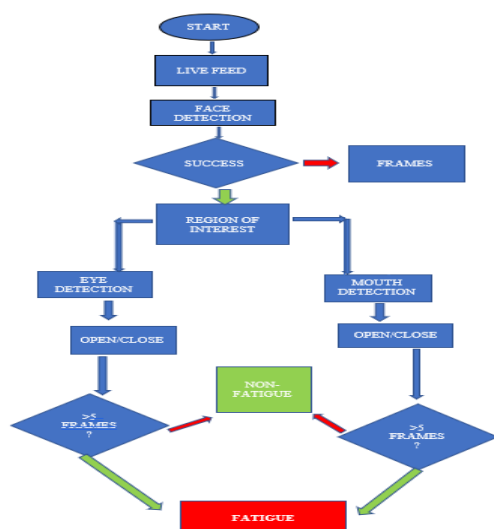


Fig 5:Flowchart of driver distraction detection

## III. SYSTEMIMPLEMENTATION

ResNet-50 is the pretrained Deep Learning model for image classification of the Convolutional Neural Network(CNN, or ConvNet), which is a category of multi-layered neural networks, most commonly applied to analyzing visual imagery. ResNet-50 is 50layers depth and is pretrained with million images over 1000 different groups from the ImageNet database. Added to this,the model has got more than 23 million trained parameters, which tells a depth of the architecture that makes it the best forrecognition of image. It enables us to teach the machine using its (ML)machine learning logics to distinguish between the various user use cases conditions like blinking, yawning or no abnormalities being detected.Using the image processing strategy and thesis of math, we cansuccessfully make use of these logics to validate the AI(artificial intelligence) conceptsavailable to recognize and categorize the events that place. Additionally, the system is can take action in accordance with the event taking place. And feature being



extracted using Viola-Jones algorithm.

*Working*

The working process is being split up into two different sections one is the training section and the other is the testing section. Initially, the datasets are formed from extracting features from the input video that is recorded using the laptop’s integrated webcam. After obtaining features like eyes and mouth, they are segregated into groups like open and closed. These groups are used for training the database and are stored as matrices in feature vector space. On other hand, the testing of Convolutional neural network ResNet-50 to achieve greater accuracy. Then comes the final stage, where the live feed is taken as input and it is equated to the trained CNN layers and their matching percentage is evaluated and the final result is displayed where it is “fatigue” or “non-fatigue”. If fatigue it gives out a beep sound in the command box.

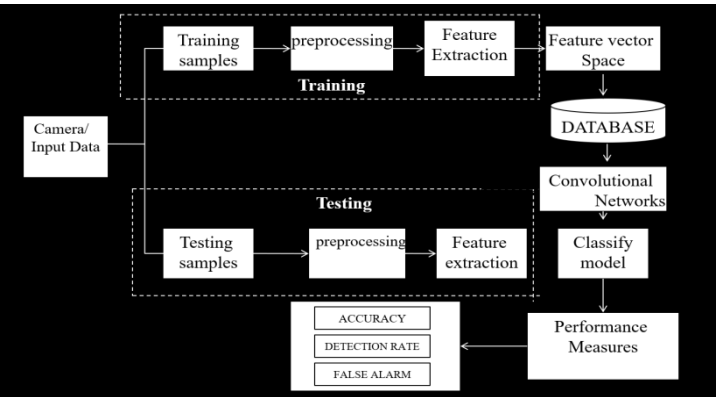


Fig6:WorkFlow

IV. RESULTSANDDISCUSSION

Followed by the learning(training) of database is the categorizing of images depending on the different features being detected. For instance, if a face is detected using the skin detector the features like eyes and mouth position are extracted. The database which already has a distinguished dataset is referred with the live feed and checks the ratio of matching features and it determines the position of the driver. Since, the input is given as video the images are acquired from it frame by frame if the detected feature maybe eyes or mouth is at the closed and opened state respectively for greater than “FIVE” successive frames, at that point the code gives a beep and it states as “FATIGUE” else it displays “NON-FATIGUE”.

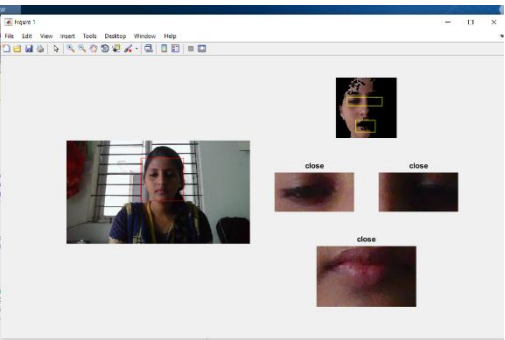


Fig 7: Eyes and mouthDetection



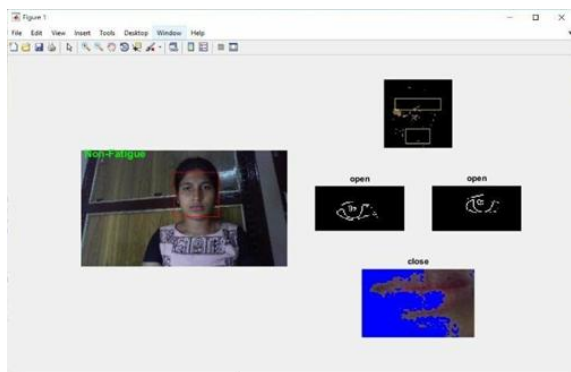


Fig8: Non-fatigue output

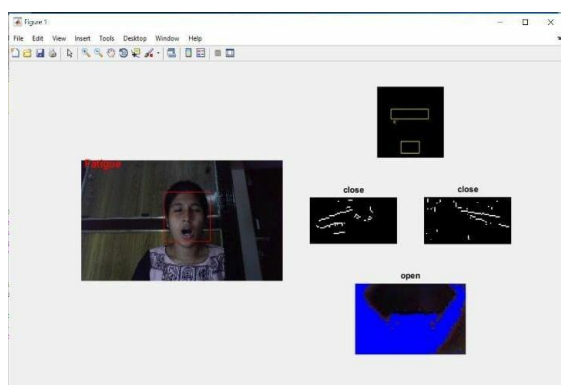


Fig 9: Fatigue output

## V. CONCLUSION

We conclude that by designing a real time drowsiness detection system by utilizing ML classifiers that measures driver's cognitive distraction. Enormous number of road accidents might then be prevented if a caution is sent to a driver about his drowsiness. A brief training phase makes the system strong, flexible and robust. In other respects, the proposed system is efficient for separate individuals with distinct types of eyelid and facial behaviors. Experiments conclude that the accuracy of proposed method for extracting the traits of driver fatigue and distraction is great. Added to this, this method can evaluate the driver's distraction and fatigue nature by personalized gauging.

## VI. REFERENCES

- [1] <https://data.gov.in/keywords/indian-road-accident-data>
- [2] Pauly, Leo, and Deepa Sankar. "Detection of drowsiness based on HOG features and SVM classifiers." 2015 IEEE International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN). IEEE, 2015.
- [3] Li, G., Lee, B. L., & Chung, W. Y. (2015). Smartwatch- Based Wearable EEG System for Driver Drowsiness Detection. Sensors Journal, IEEE, 15(12), 7169-7180.
- [4] R. Nopsuwanchai, Y. Noguchi, M. Ohsuga, Y. Kamakura, and Y. Inoue, "Driver-independent assessment of arousal states from video sequences based on the classification of eyeblink patterns," in Intelligent Transportation.
- [5] Rezaee, Khosro, et al. "Real-time intelligent alarm system of driver fatigue based on video sequences." Robotics and Mechatronics (ICRoM), 2013 First RSI/ISM International Conference

on. IEEE, 2013.

- [6] Du, Yong, et al. "Driver fatigue detection based on eye state analysis." Proceedings of the 11th Joint Conference on Information Sciences. 2008.
- [7] Choi, In-Ho, Sung Kyung Hong, and Yong-Guk Kim. "Real-time categorization of driver's gaze zone using the deep learning techniques." 2016 International Conference on Big Data and Smart Computing (BigComp). IEEE, 2016.
- [8] Tawari, Ashish, KuoHao Chen, and Mohan Manubhai Trivedi. "Where is the driver looking: Analysis of head, eye and iris for robust gaze zone estimation" Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on. IEEE, 2014.
- [9] Singh, R. K., et al. "A real-time heart-rate monitor using non-contact electrocardiogram for automotive drivers." 2016 IEEE First International Conference on Control, Measurement and Instrumentation (CMI). IEEE, 2016.
- [10] Viola, Paul, and Michael J. Jones. "Robust real-time face detection." International journal of computer vision 57.2 (2004): 137- 154.