Enhancing Road Safety: Eye Movement Analysis for Advanced Driver Fatigue Detection

Boby Dubey[1], Devesh Kumar[2], Utkarsh Kumar[3] ,Dr Mandeep Singh[4]

[1] University Institute of Engineering, Apex Institute of Technologies - Computer Science & Engineering, Chandigarh University, Punjab, India

[2] University Institute of Engineering, Apex Institute of Technologies - Computer Science & Engineering, Chandigarh University, Punjab, IndiaI

[3] University Institute of Engineering, Apex Institute of Technologies - Computer Science & Engineering, Chandigarh University, Punjab, India E-mail:

[1]Kumardeveshbsb@gmail.com

Abstract: In image processing or computer vision, image segmentation is a vital issue for applications such as scene understanding, medical image evaluation, robotic perception, video surveillance, increased reality or compression, etc. Every year in road accidents caused because of human mistake, the numbers of dead and injured are rising. Drowsiness and driving are particularly risky and difficult to recognize. The second leading cause of road crashes in drowsiness after alcohol. Detecting driver drowsiness is a technology of safety for vehicles that helps placed an end to driver injuries that are dozy. One of the main causes of road accidents is driver drowsiness. It is a very serious issue for road safety. We have presented various methods for detecting the drowsiness of the driverin this paper and the comparisons among such methods are extremely challenging. For this purpose, we have compared machine learning methods based on facial expression,especially on eye state. Apart from eye detection, it performed experiments on mouth detection and face detection as well. This paper explores several methods for machine learning, like SVM, CNN, or HMM. From the analysis, we have found that the HMM model achieved more accurate results in comparisonto others.

1. Introduction

Drowsiness or fatigue is one of the key risks to road safety which causes serious injury, death, and expense. The increased drowsinessworsens driving performance. The unconscious shift from waking to sleep leads to a lack of alerting which leads to many major road injuries. Drowsy driving has resultedin over 100,000 road injuries and over 1,500 death rates each year, documented in U.S. National Highway Traffic Safety Administration (NHTSA). Fatigue of driver can have many factors, including sleep loss, long journey, restlessness, alcohol consumption & mental pressure. Each will lead to a severe disaster. Today, road rage has been in many parts of the past, causing drivers strain. The previous transport system is therefore not adequate to deal with these threats on the roads. Therefore, most deadly injuries can be avoided by embedding automatic fatigue detection systems in vehicles. The drowsiness detection system (DDS) analyzes drivers' attention and alerts the drivers before reaching the serious road safety threat [1].

In our life, emotions play a crucial role. During our lives and interactions with others, there are many types of emotions. Recent research is done on emotion recognition. If this is applied in the social networking setting, it can be an important way to understand how individuals and social circles, neighborhoods, and towns feel about recent or different events [2,3].

Machine learning has become one of the major stays of IT. There is a strong purpose to assume that intelligent data processing will become much more widespread as an essential factor for technical purposes, considering the growing amount of information available. Most of the science of ML is about solving these needs and providing the solution [4].

The organization of this paper as follows Section II introduces image segmentation including image classification techniques. In Section III we describe the drowsy driver detection system in this paper. We present a detailed description of numerous Computer vision methods to detect driver drowsiness in Section V. Next Machine learning techniques are described in Section VI. After this, we provide a detailed study of related work. Conclusions are presented in the last section.

2. Classification of Segmentation Algorithms

Segmentation algorithms havebeen created to segment imagesfrom segmentation; its dependent on 2 key characteristics, discontinuity & similarity. Division and subdivision are predicated on discontinuity depending on the abrupt intensity or gray picture levels. Our interest in this approach is primarily to define individual points, lines, and boundaries. The predicted pixel community in heterogeneous attributes contains approaches such as thresholds, regional expansion, and regionsplitting as well as merging.

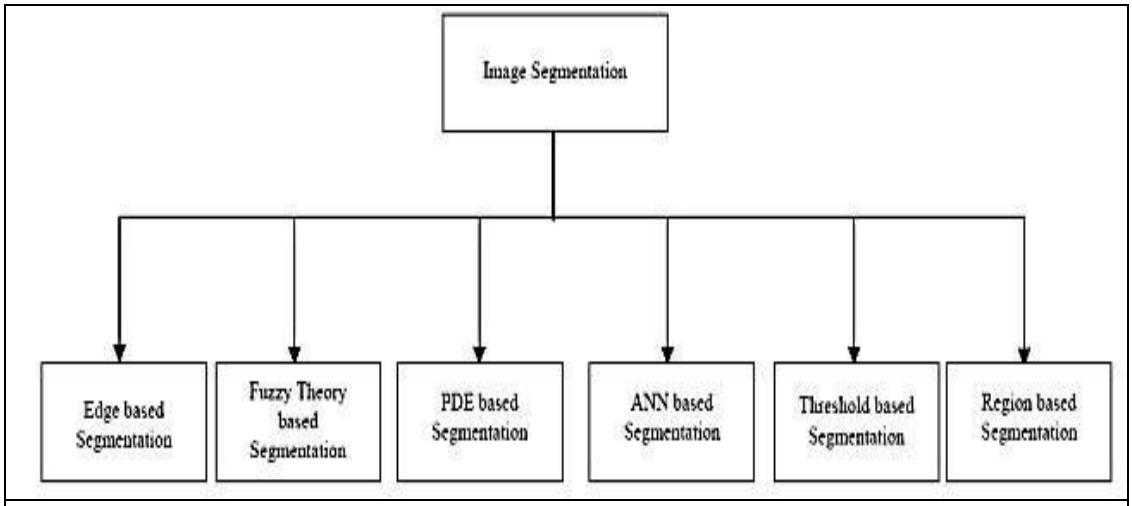


Figure 1. Image segmentation Methods.

2.1. Segmentation by Edge Detection

Edge detection is a fundamental step in the segmentation process. The image splits into the object as well as its context. Divide the image by the edge detectionprocess to detecting the change in intensity or image pixels. Two key methods for edge detection embedded in the segmentation are found in the Gray histogram and Gradient. The canny edge detector is a second-order derivative operator that is split into two types of edge detection operators, as first-order derivative & secondorder derivative operators. The operators with a second-order derivative provide reliable results.

2.2. Segmentation by Thresholding

Image segmentation is one of the easiest methods for the image which is dependent on levels of intensity named as a threshold. Thresholding can be implemented either globally or locally. Global threshold differentiates object & context pixels by comparing them to the selected threshold value and by using a binary partition in the image section. An oftencalled adaptive threshold is the local thresholding strategy. Adaptive thresholding strategyis threshold value varies according to the locale characteristics of the regions in the image and histogram thresholding is usedfor the segmentation of the specified image; some preprocessing & post-processing approaches are essential for threshold segments. Mean technique, Pstyle method, histogram-based method, Edge maximization technique & visual method are primary thresholding techniques suggested by various researchers.

2.3. Segmentation by Region-based

The segmentation region observed should be closed. Also, the region-based segmentation is called the similarity primary based segmentation. Due to the lack of edge pixels, there is no difference in the region primarily based on segmentation the boundaries are known for segmentation. The sting flow is regenerated into a vector after distinctive alteration in the color and texture. The sides are detected for further segmentation.

2.4. Segmentation by Feature

- Based Clustering
- Segmenting by K Means Clustering[5]

3. Driver Drowsiness

Drowsiness is simply referred to as "fatigue-related near-sleep." It's technically distinct from fatigue, distinct as a “refusal to carry out the task at hand”. There are the same effects of sleepiness and fatigue. Fatigue has an effect on mental alertness, reduces the ability of the person to drive a vehicle safely & raises the risk of human error, which can lead to damage and injury. Sleepiness slows down the time of response, decreases awareness, and undermines judgment. Both transport operators, such as pilots, truck drivers, and railway engineers, have been impacted byfatigue and sleep deprivation. In these cases, the driver cannot focus on the main driving task, which will increase the risk of an accident. This problem would deteriorate further with ever-growing traffic conditions [6].

3.1. Drowsy Driver

Drowsy driving is the riskiest groupings of sleepiness or fatigue. Usually, driving sleep deprivation has become a growing risk of an accident; because the driver was not properly asleep, often owing to sleep syndromes, alcohol effects, few medicines, and even day & night activities.
If the driver is drowsy it increases the risk of a collision more often than normal driving. Nowadays it is very difficult to know the exact moment when sleep comes over. Also, the risk is raised during the night.

3.2. Factors Cause Driving Drowsiness

Drowsiness driver affects driving greatly and raises the risk of crashes. More than 30% of driver sleepiness or fatigue is responsible for road accidents.There are 2 key reasons for driver fatigue:

- If sleep quality &quantity are insufficient.
- Start driving when sleeping is usually finished [7].

3.3. Drowsy Driver Detection System

Many different methods can be used to assess the drowsiness of the driver. These methods are image recognition-based techniques, Electroencephalograph techniques, and techniques based on artificial neural networks. Image processing techniquesmay be split intothree types. These are the technique of matching the template, the technique of eye blinking and yawning. These methods are focused on the interpretation of images through computer vision. In computer vision technology the driver uses facial features e.g. blindness of the eyes and head movements to detect driver drowsiness.

Table 1. Techniques for drowsy detection

No.	Technique used	Strongpoint	Weak point
1.	Brain-computer	Very efficient in health care	Electrodes outside of the skull interface [8] can detect very few electric signals from the brain
2.	Geometry based [9]	Small database, recognition rate 95%	A large number of features are used.
3.	Template-based [10]	Recognition rate 100%	Complex
4.	Color-based [11]	Simple and small database	Limited performance
5.	Support vector machine [12]	Flexible	Lack of transparency in result
6.	Iris recognition [13]	Produce high accuracy result in less time	Expensive

3.3.1. Image processing-based Techniques

Drivers face images are used for the analysis of IP techniques so that their states are detected. From the face image, one can see that driver is awake or sleeping. The driver is asleep or drowsiness with the same images because the eyes of the driver are closed in the face image. And the facial image can also detect other signs of drowsiness. These strategies can be divided into three sub-categories:

1) Eye Blinking Based Technique

The blink rate andduration of the eye closure are measured to detect the drowsiness of the driver. Since eye blinkers and gaze between the eyelids differs from normal situations when the driver is sleepy at the time, so they detect drowsiness easily. The position of iris and eye conditions in this system is tracked over time to measure the duration of blink and the near life of the eye. And a remotely placed camera uses this sort of system to achieve video and computer vision techniques, such that visual, eyes, and eyes are sequentially located in the closure ratio. The driver's drowsiness can be identified by this eye closing and blinking ratio.

2) Template Matching Technique

In this technique, the eye state should be used for some particular timei.e. when the driver closes the eye/s then the system produces an alert. And, there is a driverthat has both open and closed eyestemplatein this system. It can also be used to open and close eye templates of the drivers.

3) PERCLOS Technique

PERCLOS is a fixed parameter for drowsiness detection level. The PERCLOS score is intended to determine if the driver is in a drowsy condition or not (the percentage of time an eye is closed at a certain time).

4) Yawning Based Technique

Yawn is drowsiness symptoms. yawn is ostensibly shown on a big vertical opening of the mouth. The mouth is wide openis wider than in the yawncompared to the speech. They can detect yawn by using face tracking and mouth tracking [14,15].

4. Computer Vision Method to Detect Driver Drowsiness

Computer vision techniques for identification of changes in facial expressions of the driver.

Computer vision approach to detect driver drowsiness using machine learning algorithms to detect the closure and opening of the eyes. There was no fully automatic model previously built. Second, a Lowcost solution has not been existingdriver drowsiness detection. Thirdly, no model worked for different luminance conditions (produced by sunlight & during dim light conditions similar to bad weather). Thus,there is an enormous need for a low costfor the fully automated driver drowsiness detection model, which operates in different luminance conditions.

4.1. Facial Expression

Three facial detection approaches are available. They are focused on features, templates, and appearance. The feature-based approach detects invariantface features but it is difficult to extract the featurein complex contexts. A predefined standard face pattern is used in the template-based approach and the correlation is used to identify the face. But it is difficult to expand different scales with this method. Face and non-face are observed in appearance-basedapproaches. But only with a simple background, this approach will produce reliable results. Facial expression characteristics are retrieved using the LG expression system. Six simple facial expressions are happiness, surprise, sadness, anger, hate, and fear.

4.2. Facial Tracking

The face tracking system must be robust for head movements, rotation, changes in illumination, and variation. For this purpose, a method has been suggestedfor the simultaneous use of face recognition & object tracking systems. This grouping provides us the chance to take advantage of 2 different programs together.

4.3. Eye Detection

It is difficult to find an eye because of various variables such as lighting, posture, facial shadow, etc. Different measures could be measured with the eyelid closure percentage, maximal closing time, blink theoretical area that focuses systematically on theory, results, & properties of education systems & algorithms. It draws on several diverse areas of ideas: optimization theory, artificial intelligence, psychology, knowledge processing, optimal controls, cognitive sciences, also other branches in research, science, and mathematics. It is a strongly interdisciplinary area. The machine learning area is usually categorized into 3 sub-domains: supervised, unsupervised, & reinforcement learning [16].Let us discussed the various types of machine learning algorithms:

5.1. Supervised Learning

In supervised learning, we learn anobjective function that can be applied to predict the values of approved or not approved discrete class features. In a specific sample set, a machine learning algorithm makes predictions, whereas a supervised learning algorithm looks for designs to assign the data point labels. This algorithm is anoutcome variable from which a certain number of predictorscan be estimatedi.e. independent variables. Will create a function that maps input to desired outputs using this set of variables. The training phase continues until the model achieved accuracy in terms of training data. This whole procedure aims to reduce manual checks and coding expenses. Supervised learning examples are NNs, Regression, Decision Tree, NN, SVM, Naive Bayes, etc.

5.2. Unsupervised Learning

The learning useful structure is referred to as unsupervised learning with the labeled classes, criteria for the optimization, input signal, or other knowledge beyond the raw data. We don't have any target variable in this algorithm for calculating mean and we do not have a data points label or can say the training data class label is unknown here. This algorithm is applied to collect data into cluster group to explain the configuration of the data i.e. the cluster of the data which shows essential partitions and hierarchies. Examples: K-means, Clustering Fuzzy, Clustering Hierarchy Data are not labeled as well as the result is not known.

5.3. Reinforcement Learning

The machine is trained to make specific decisions using this algorithm. Based on each data point, these algorithms choose an action and later learn how to achievea successful decision. This is an environment in which the machine trains [17].

4. Proposed Approach

Almost all drivers have experienced this drowsiness problem while driving [7]. Youngsters and professional drivers are mostly affected by this drowsy driving because of continuous hours of driving without any rest. In many cities, auto drivers and cab drivers drive continuously overtime sometimes to complete their targets or at times to get bonus profit. Many of the poor workers in order to meet their daily expenses and for the sake of their loved families tend to work in night shifts for long time, this can be one of the main reasons for accidents taking place because of drowsy driving.

Therefore, a driver drowsiness monitoring system has been developed in this paper. The block diagram of the proposed method for drowsiness monitoring is shown in figure 1.

A webcam has been used to record the video of the driver. The webcam is arranged in such a way that it captures the front facial image of the driver [8]. Once the video capturing is done, the recorded frames are then pulled out to get the 2-Dimensional images [9]. The object (Face) in the frames is detected by HOG and SVM algorithm.

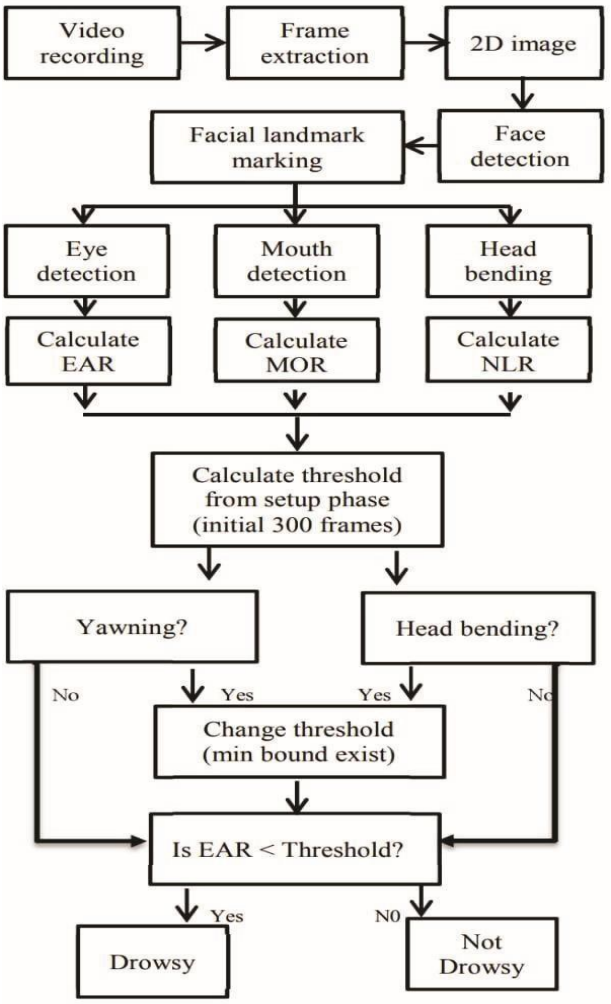


Fig.1 The block diagram of drowsiness detection system

5. Implementation

5.1 Acquiring Data:

At first, the video is captured and recorded using a laptop webcam and then the frames are released and proposed in a laptop. Then after the completion of extraction of frames, the picture implementation techniques are proposed on the 2-dimensional picture of the recorded video. Now, the required driver information is produced. The volunteers are asked to focus on the laptop webcam and perform activities like continual eye blinking & closing, mouth yawning, and head bending. And the webcam is adjusted to capture the video for 20 to 30 min.

5.2 Face Identification: The driver's face is identified first after the frames have been extracted. In this paper, HOG & SVM algorithms [10] are used for face extraction. In this detection, only positive examples of the stable window size are taken for photographs and histogram of oriented gradients (HOG) descriptors are calculated on them. Following that, the negative examples of same size are considered for HOG descriptors and then results are evaluated. Typically, the count of negative examples is more than positive examples. Then after getting the characteristics of both the groups, a direct SVM algorithm is used for classifying the required task. To get more detection and better accuracy for SVM algorithm, strong negatives are used. In this detection, after the guidance, the SVM classifier is investigated on labeled data and the incorrect positive example characteristic values are reused for guidance purpose. To test the picture, the window used for positive examples are rendered over the image and then the required result or output will be classified for each window location. Then based on the results obtained and considering the various sample results, the higher value samples are considered for drowsiness identification and the boundary outline is drawn on the face. This minimum elimination steps will reduce the overlapping and redundant bounding in the bounding box.

5.3 Locating Face Points:

After the detection of faces, the further step is to locate the points on the human face like eyes, mouth, nose etc. Then the photograph that used for face detection has to be normalized to minimize the distance factor between camera and the driver. Therefore, the face photograph is resized with the width of 500 pixels and converted into gray scale pictures. Then normalization is done for regression trees [11]. And it will be approximate the positions of location points on the face from pixel intensities. In this process, the square error loss is reduced by using gradient boost learning. Different priors are used to discover various structures. By considering this procedure, the location of the boundaries of eyes, mouth and nose are noted and shown in Table I. Then the location points on face are marked and shown in fig.2. The red points on the figure are used for further identification.

Parts	**Landmark Points**
Mouth	[13-24]
Right eye	[1-6]
Left eye	[7-12]
Nose	[25-28]

Table I: The location points on face

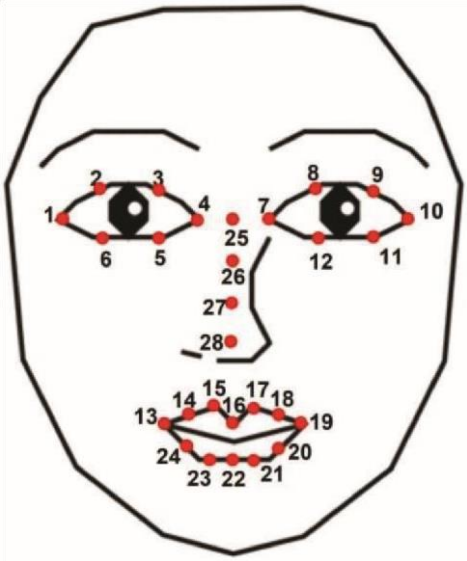


Fig.2 The landmark points on face

5.4 Feature Extraction:

After marking the points on face, the drowsiness features are calculated as given below.

Eye Aspect Ratio (EAR): From the boundary points on the eyes, the EAR is evaluated as inverse ratio of width of eyes to the height of eyes. The mathematical formula for EAR is given as,

$$= \frac{(p_2 - p_6) + (p_3 - p_5)}{2(p_4 - p_1)}$$

where i is the point marked as i in facial landmarking (i − j) gives the distance between the points i and j. When the eyes are open EAR is max and when the eyes are closed, the EAR is approximated to zero. So, the decreasing value of EAR indicates the closing of eyes and identifies the drowsy behavior of the driver.

Mouth Opening Ratio (MOR): Mouth Opening Ratio is defined as the identification of yawning of the driver and ultimately drowsiness detection is done. The mathematical formula for calculating the MOR is given as,

$$= \frac{(p_{15} - p_{23}) + (p_{16} - p_{22}) + (p_{17} - p_{21})}{3(p_{19} - p_{13})}$$

The MOR ratio is maximum when the mouth of driver is open and if it remains same for some time, then yawning condition is indicated. And if it decreases towards zero, then the condition is considered as normal, and not yawning.

Nose Length Ratio (NLR): When the head of the driver is down along with vertical axis, then by using the bending angle of head the drowsiness alert is identified. The focal plane of the webcam is directly proportional to head bending, and by using this head bending is calculated.

Normally, the nose makes an acute angle with the webcam. The acute angle increases when head moves upward and vice versa. Therefore, the nose length ratio is given as nose length to average length of nose, which also measures the head bending. The mathematical formula for NLR is given as,

$$= \qquad h\,(-)$$

The facial points for nose length ratio and head bending are shown in figure.2.

5.5 Classification:

After obtaining the values of eyes, mouth and nose, the next step is to identify the drowsiness from the SVM frames. The flexible value of threshold is required for the drowsiness. Then, algorithms like SVM are used for the classification of the information.

For obtaining the threshold values for eyes, mouth and nose the initial condition of the driver is assumed to be in normal state. This step is known as setup phase. The EAR for nearly 100 to 300 frames will be recorded, and from this an average of 150 to 200 higher values are recognized as strong threshold values for EAR. The values which are high i.e., in which eyes are not closed is considered. If threshold is more than the tested value, then it is identified as eyes closed. For every person, the size of eye will be different; so, this makes the impact to decrease the setup for person to person. For computing MOR values, the threshold value of frames is calculated based on the condition that the mouth is not open. If the threshold is less than the test value i.e., when the mouth is open, the yawning is identified. Using the head bending feature, the angle between access and head can be determined in terms of nose length ratio. NLR values range from 0.8 to 1 in normal condition of head and varies with head bending up and down. The average of the NLR is measured as mean of lengths of nose in the setup assuming that head is not bend. After obtaining the threshold values, it is tested. Then, if at least one of the eyes, mouth and nose identifiers are not satisfied, then drowsiness alert is indicated. In practical, for example out of 75 frames at least 70 frames satisfy drowsiness conditions for one or more features, then the overall system is identified as drowsiness detection and the driver is alerted with the alarm sound.

$$\frac{nose \qquad p_{28} \quad 25}{average\ nose\ length}$$

To overcome the thresholding problem, a single value of threshold is considered, and this threshold value depend on EAR. To obtain the average value of EAR, 150 higher EAR values from 300 frames are considered. If the threshold value is more than EAR value, then the driver is at danger. By considering yawn and head bending, this EAR threshold will be increased, and it is dispersed into more frames. The yawning and head bending frames are combined to get flexible value of threshold. If EAR value is more than obtained threshold value, then that condition is treated as drowsiness, and this indicates to alert. In head bending situation, if the head is down, then the frames are considered for a drowsiness alert. Table II shows the calculated parameters.

The machine learning algorithms and the threshold factors are used for the identification of driver's drowsiness, from the values obtained from EAR, MOR, and NLR. Earlier, these features were used for the analysis of classification from feature space to individual. However, here it is used for the principal component analysis [12].

Table II. Calculated parameter values of threshold

EAR from setup phase (average of 150 maximum values out of 300 frames)	0.34
Threshold = EAR - offset	0.34 - .045 = 0.295
At Yawning, (MOR > 0.6)	Threshold = Threshold +0.002 *Max bound exist
At Head Bending, (NLR < 0.7 OR NLR > 1.2)	Threshold = Threshold +0.001 *Max bound exist

After converting the values obtained from threshold, whether the features are significant for classes or not is tested. If three factors give five percent significance, this classification based on Bayesian Classifier and SVM algorithm is used [12].

6. Results And Discussion:

The webcam of the laptop is turned ON and the face of the driver is recorded and captured. The captured video is transferred to the processing block to identify the drowsiness detection. If mishap is detected, then alarm will buzz a sound to alert the driver. The initial state of driver condition is given in Fig. 3. For this frame the values are given as,

EAR= 0.350, MOR= 0.341, NLR= 1.030

Fig.3 Initial condition of driver

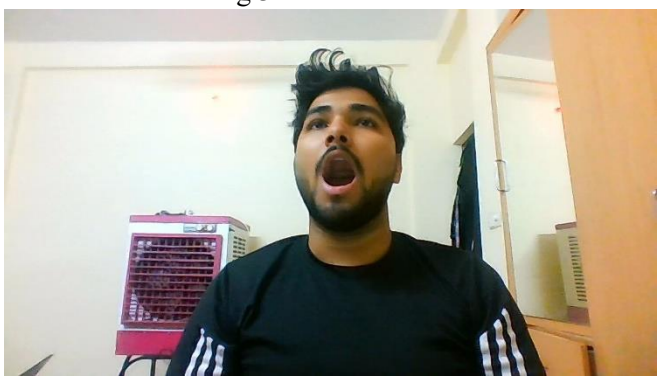


Fig. 4(i) Drowsy identification based on Yawning

Fig. 4(ii) Drowsy identification based on closed eyes

Fig. 4(iii) Drowsy identification based on bent head

Various drowsiness situations are shown in Fig. 4. Figure 4(i) shows the example of alert for drowsiness due to yawning and Fig. 4(ii) shows the example of drowsiness alert due to eye closing. Fig. 4(iii) shows an example of drowsiness due to head bending. Table III shows the sample values of the parameters for various states.

Table III: Sample values of distinct parameters & states

State	EAR	MOR	NLR
Normal	0.35	0.34	1.003
Yawning	0.22	0.77	0.76
Eye Closed	0.15	0.419	0.876
Head Bending	0.15	0.577	0.66

The system can also detect the driver's drowsiness along with spectacles as shown in Fig. 5.

Fig. 5 Identification of eyes with glasses

The algorithms are used for the testing on the INVEDRIFAC dataset [13].

2.1. Facial Landmarks for Eye Blink Detection

Deep-learning-based facial landmark detection systems have made impressive strides in recent years [32,33]. A cascaded convolutional network model, as proposed by Sun et al. [34], consists of a total of 23 CNN models. This model has very high computational complexity during training and testing. To detect and track important facial features, identification of facial markers must be performed on the subject. As a result of head movements and facial expressions, facial tracking is stronger for rigid facial deformations. Facial landmark identification is a computer vision job in which we try to identify and track key points on the human face using computer vision algorithms [35]. Multi-Block ColorBinarized Statistical Image Features (MB-C-BSIF) is discussed in [36]; this is a novel approach to single-sample facial recognition (SSFR) that makes use of local, regional, global, and textured-color properties [37,38].

Drowsiness can be measured on a computational eyeglass that can continually sense fine-grained measures such as blink duration and percentage of eye closure (PERCLOS) at high frame rates of about 100 fps. This work can be used for a variety of problems. Facial landmarks are used to localize

and represent salient regions of the face, including eyes, eyebrows, nose, mouth, and jawline. Blinking occurs repeatedly and involuntarily throughout the day to maintain a certain thickness of the tear film on the cornea [39]. The act of blinking is a reflex that involves the fast closure and opening of the eyelids in rapid succession. Blinking is also known as blepharospasm. The act of blinking is performed subconsciously. The synchronization of several different muscles is required for the act of blinking one's eyes.

While keeping the cornea healthy is an important function of blinking, there are other benefits as well [40], and this is supported by the fact that adults and infants blink their eyes at different rates. A person's blink frequency changes in response to their level of activity. The number of blinks increases when a person reads a certain phrase aloud or performs a visually given information exercise, whereas the number of blinks decreases when a person focuses on visual information or reads words quietly [41].

In our investigation, we made use of the 68 facial landmarks from Dlib [42]. Estimating the 68 (x,y)coordinates corresponding to the facial structure on the face was carried out with the help of a pretrained facial landmark detector found in the Dlib library. Figure 1 displays that the jaw points range from 0 to 16, the right brow points range from 17 to 21, and the left brow points range from 22 to 26. The nose points range from 27 to 35, the right eye points range from 36 to 41, and the left eye points range from 42 to 47. The mouth points range from 48 to 60, and the lip points range from 61 to 67. Dlib is a library that helps implement computer vision and machine learning techniques. The C++ programming language serves as the foundation for this library.

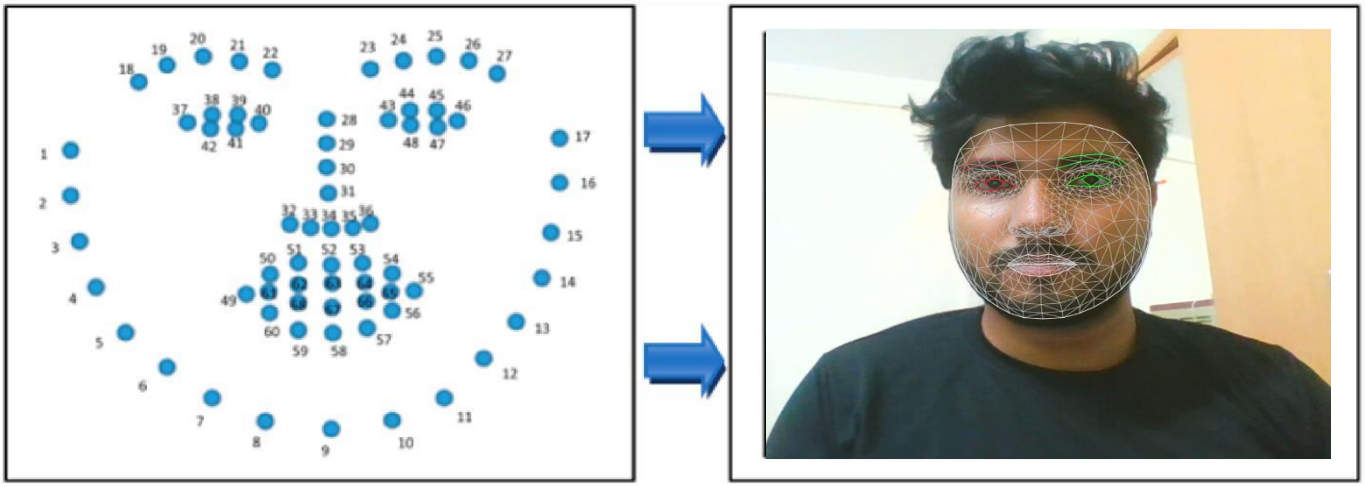


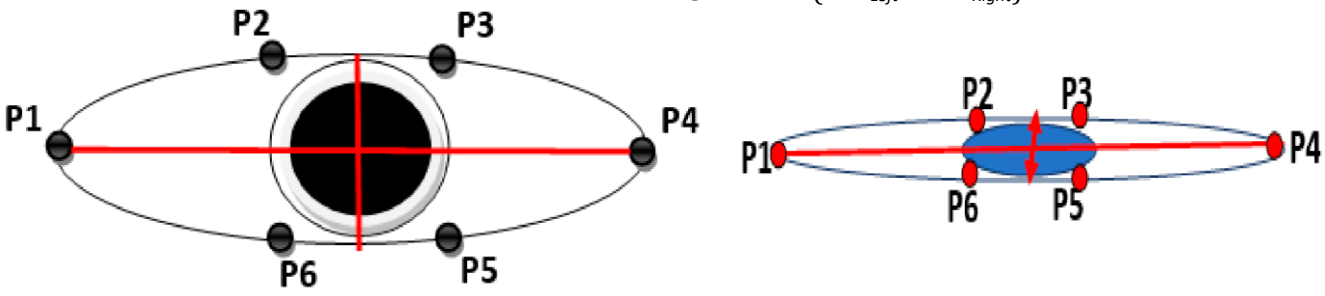
Figure 1. Eye identification made possible by employing facial landmarks (right eye points = 36–41, left eye points = 42–47).

The process of locating facial landmark points with the use of Dlib's 68-landmark model consists of two stages, all of which are described in the following order: (1) The first way to locate a human face, face detection, is done by returning a value in the form of x, y, w, and h coordinates, which together form a rectangle. (2) A facial landmark: Once we have determined the location of a face within an image, we must then place points within the rectangle. This annotation is included in the 68-point The Dlib framework can be utilized to train form predictors on input training data, regardless of the dataset that is selected to be trained on.

2.2. Eye Aspect Ratio (EAR)

$$EAR = \frac{\| P2 - P6 \| + \| P3 - P5 \|}{2 \| P1 - P4 \|} \quad (1)$$

$$AVG\ EAR = \frac{1}{2}(EAR_{Left} + EAR_{Right}) \quad (2)$$



(**a**) (**b**)

Figure 2. Open and closed eyes with facial landmarks (P1, P2, P3, P4, P5, P6). (**a**) Open eye. (**b**) Close eye.

The EAR equations are described by Equation (1), where P1 through P6 stand in for the locations of the 2D landmarks on the retina. P2, P3, P5, and P6 were utilized to measure the height of the eye, whereas P1 and P4 were utilized to measure the breadth of the eye. This is depicted in Figure 2. When the eyes are closed, the EAR value quickly drops to virtually zero, in contrast to when the eyes are open, in which case the EAR value remains constant. This behavior is seen in Figure 2b.

2.3. Research Workflow

Our system architecture is divided into two steps, namely, data preprocessing and eye blink detection, as described in Figure 4. In the data preprocessing step, the video labeling procedure using Eyeblink Annotator 3.0 by Andrej Fogelton [46] is shown in Figure 3. OpenCV version 2.4.6 is utilized by the annotation tool. Both video 1 and video 3 were recorded at a frame rate of 27.97 frames per second. Video 2 was captured with 24 fps. Video 1 has a length of 1829 frames, totaling 29.6 MB. Video 2 has a length of 1302 frames and a file size of 12.4 MB. Next, video 3 has 2195 frames and a file size of 38.6 MB. The Talking Face and Eyeblink8 datasets contain 5000 frames and 10,712 frames, respectively. Video information is explained in Table 1.

People who wear glasses and have relatively tiny eyes are represented in our dataset in a unique way. The people who operate automobiles make up the environment. This dataset may be utilized for additional research endeavors. Based on what we know, it is difficult to locate a dataset of persons who have tiny eyes, wear spectacles, and drive cars. We have the footage from the dashboard camera installed in a vehicle in the Wufeng District of Taichung, Taiwan. We have verified that informed consent was received from each individual who participated in the video dataset collection. Our data collection includes five films and one individual performing a driving scenario. The annotations start with line "#start," and rows consist of the following information:

frame ID: blink ID: NF: LE_FC: LE_NV: RE_FC: RE_NV: F_X: F_Y: F_W: F_H: LE_LX: LE_LY: LE_RX: LE_RY: RE_LX: RE_LY: RE_RX: RE_RY. An example of a frame with a blink is: 118: −1: X: X: X: X: X: 394: 287: 220: 217: 442: 369: 468: 367: 516: 364: 546: 363. The eyes may or may not be completely closed during a blink. According to the website blinkmatters.com, the range of fully closed eyes during a blink is between ninety percent to one hundred percent [46]. The row will be like: 415: 5: X: C: X: C: X: 451: 294: 182: 191: 491: 362: 513: 363: 554: 365: 577: 367. In this particular study, our experiments were only interested in the blink ID and eye completely closed

(FC) columns; as a result, we ignored any additional information that may be provided. Table 2 provides an explanation of the features included in the dataset.

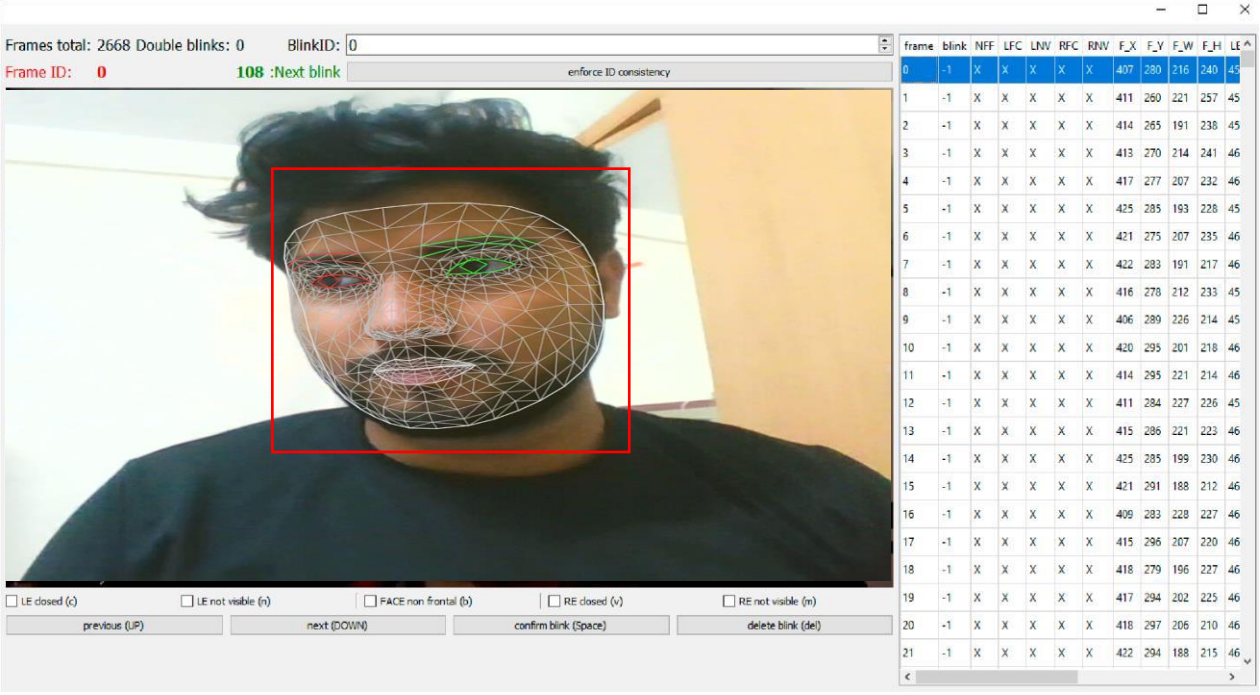


Figure 3. Video labeling process with Eyeblink Annotator 3.0.

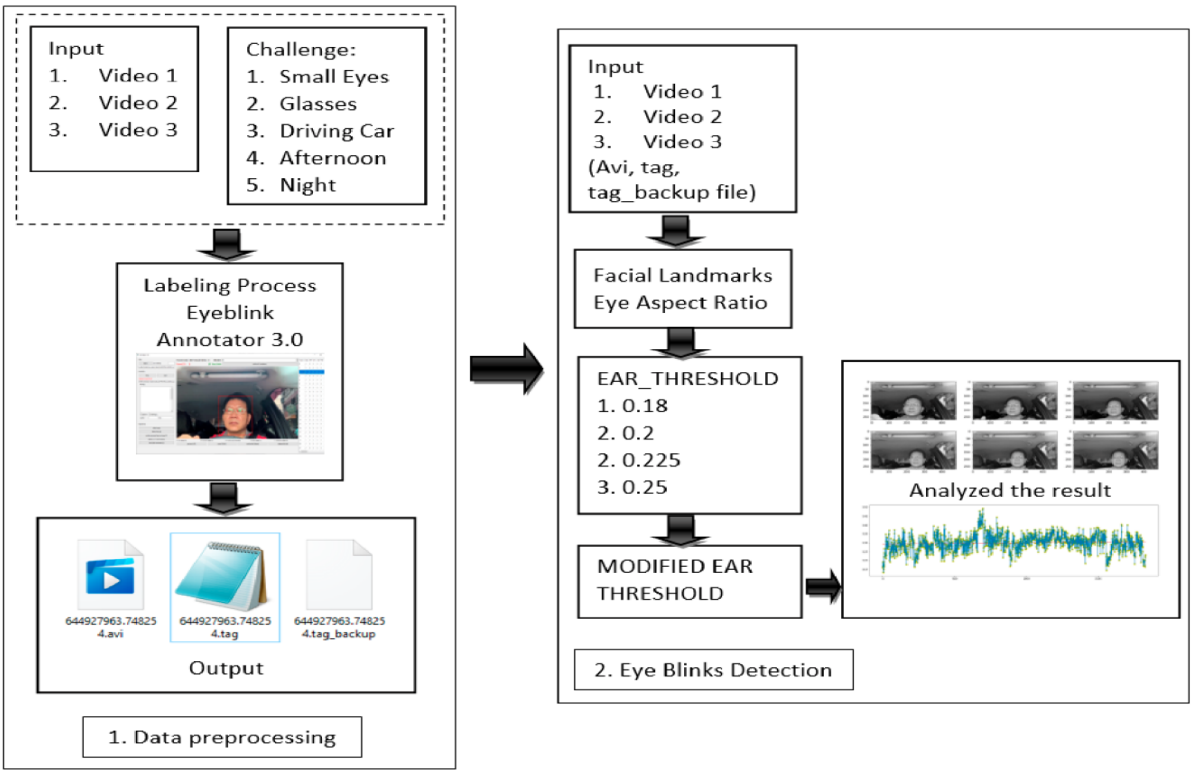


Figure 4. The system architecture.

Table 1. Video dataset information.

Video Info	Video 1	Video 2	Video 3	Talking Face	Eyeblink8 Video 8
FPS	29.97	24	29.97	30	30
Frame Count	1829	1302	2195	5000	10,712
Durations (s)	61.03	54.25	73.24	166.67	357.07
Size (MB)	29.6	12.4	38.6	22	18.6

Table 2. Dataset features.

No	Description	Features
1	Alternatively, a frame counter may be used to get a timestamp in a different file.	frame ID
2	A unique blink ID is defined as a sequence of blink ID frames that are all the identical. The time between two consecutive blinks is measured in terms of a sequence of identical blink ID frames.	blink ID
3	A change from X to N occurs in the provided variable while the person is looking sideways and blinking.	non frontal face (NF)
4	Left Eye.	left eye (LE),
5	Right Eye.	right eye (RE),
6	Face.	face (F)
7	The given flag will transition from X to C if the subject's eye closure percentage is between 90% and 100%.	eye fully closed (FC)
8	This variable changes from X to N when the subject's eye is covered (by the subject's hand, by low lighting, or by the subject's excessive head movement).	eye not visible (NV)
9	x and y coordinates, width, height.	face bounding box (F_X, F_Y, F_W, F_H)
10	RX (right corner x coordinate), LY (left corner y coordinate)	left and right eye corners positions

The Eyeblink8 dataset is more complex because it includes facial expressions, head gestures, and staring down at a keyboard. According to [46], this dataset has a total of 408 blinks across 70,992 video frames at 640 × 480-pixel resolution. This clip has an average length of between 5000 and 11,000 frames and was shot at 30 frames per second. There is only one video of a single participant chatting to the camera, so to speak, in the Talking Face dataset. In the video, the person can be seen smiling, laughing, and doing a "funny face" in a variety of situations. In addition, the frame rate is 30 frames per second, the resolution is 720 by 576, and 61 blinks that have been labeled.

2.4. Eye Blink Detection Flowchart

Figure 5 illustrates the blink detection method, and the frame-by-frame breakdown of the video is the initial stage. The facial landmarks feature [47] was implemented with the help of Dlib to detect the face. The detector used here is made up of classic histogram of oriented gradients (HOG) [48] feature along with a linear classifier. In order to identify face characteristics, including the ears, eyes, and nose, a facial landmarks detector was built inside Dlib [49,50]. Moreover, with the help of two lines, our research was able to identify blinks. The lines dividing the eye are drawn in two directions: horizontal and vertical. Blinking is the act of briefly closing the eyes and shifting the eyelids from one side to the other. Blinking is a natural thing to happen.

The eyes are closed or blinking when the eyeballs are not visible, the eyelids are closed, the upper and lower eyelids are fused, and the upper and lower eyelids are not connected. Further, when the eyes are opened, the vertical and horizontal lines are almost the same, but the vertical lines narrow or almost disappear when the eyes are closed. We may consider eye blinking if the EAR is less than the modified EAR threshold for three seconds. To perform our experiment, we used four alternative threshold values: 0.18, 0.2, 0.225, and 0.25. Additionally, we experimented with different EAR cutoffs and both video datasets.

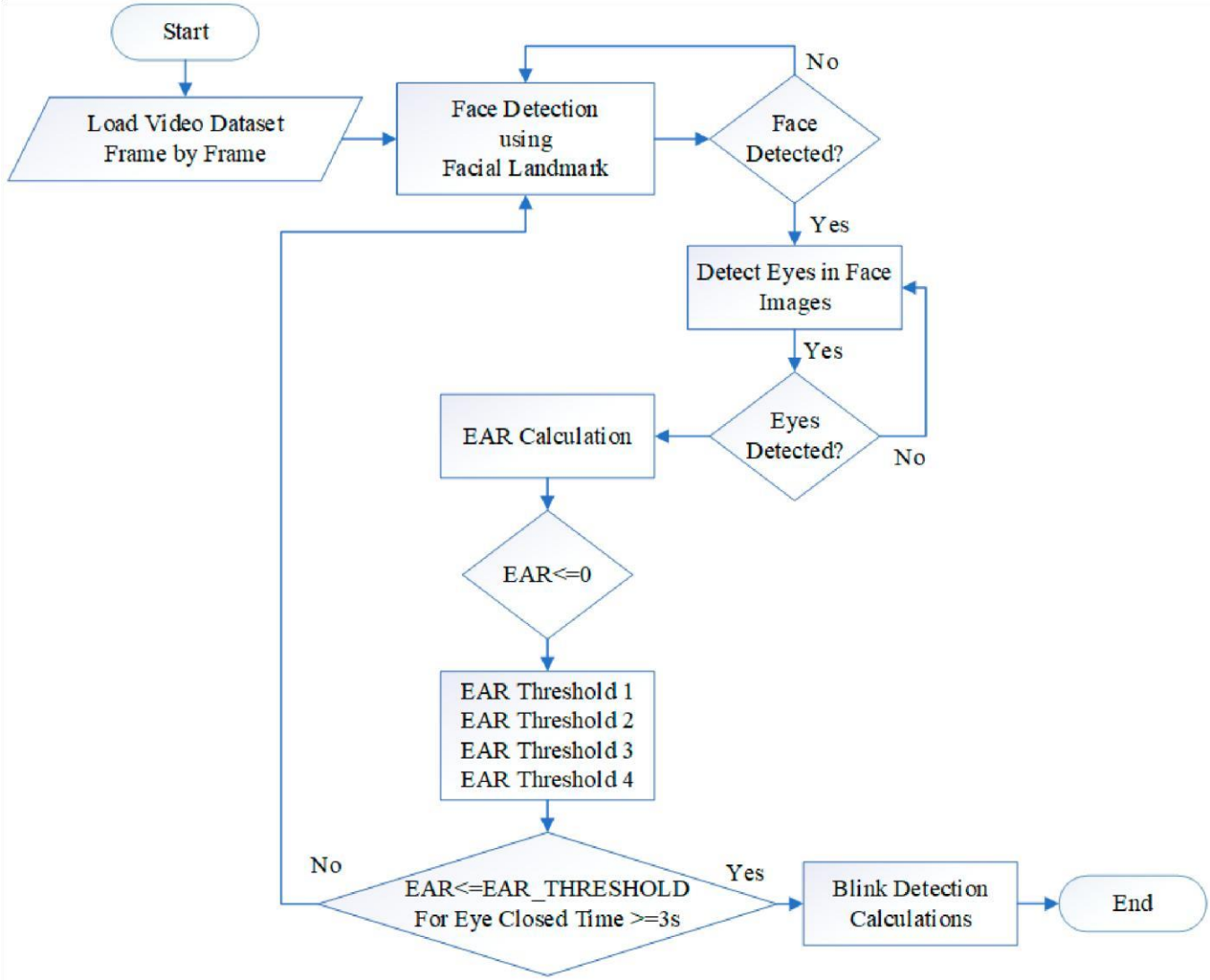


Figure 5. Eye blink detection flowchart.

3. Results

Table 3 summarizes the statistics for the predictions and test sets of video 1, video 2, and video 3. For the EAR threshold of 0.18, the total frame count for the prediction set of videos 1 is 1829, the number of closed frames analyzed was 23, and the number of blinks found was 2. On the other hand, the statistics for the test set state that there are 58 closed frames, and there are 14 blinks. This experiment exhibited an accuracy of 95.5% and an area under the curve (AUC) of 0.613. Furthermore, video 3 has 1302 frames with 182 closed frames. The maximum accuracy was obtained while implementing the 0.18 EAR threshold: 86.1%. Moreover, video 3 contains 2192 frames, totaling 73.24 seconds. Using an EAR threshold of 0.18 resulted in 89% accuracy for this dataset. In the third video, we see the minimal results of 47.5% accuracy and 0.594 AUC for an EAR threshold of 0.25 being used. Moreover, Table 4 describes the statistics for the prediction and test sets of Talking Face and Eyeblink8 datasets. Talking Face has 5000 frames with 227 closed frames. The optimum accuracy was achieved while employing the 0.18 EAR threshold: 97.1% accuracy and 0.974 AUC. For Eyeblink8, video 8, the highest accuracy was obtained when using the 0.18 EAR threshold: 86.1% accuracy and 0.732 AUC. This dataset has 1302 frames with 182 closed frames and 18 blinks.

Table 3. Statistics for prediction and test sets (Eye Blink Dataset).

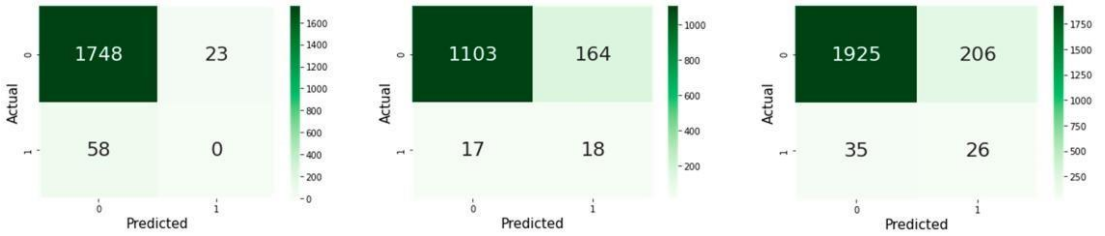
Dataset	Video 1				Video 2			
EAR Threshold (t)	0.18	0.2	0.225	0.25	0.18	0.2	0.225	0.25
Statistics on the prediction set are Total Number of Frames Processed 1829		1829	1829	1829	1302	1302	1302	1302
Number of Closed Frames	23	56	131	281	182	342	614	884
Number of Blinks	2	6	9	16	18	39	73	65
Statistics on the test set are Total Number of Frames Processed	1829	1829	1829	1829	1302	1302	1302	1302
Number of Closed Frames	58	58	58	58	35	35	35	35
Number of Blinks	14	14	14	14	9	9	9	9
Eye Closeness Frame by Frame Test Scores								
Accuracy	0.955	0.938	0.897	0.820	0.861	0.74	0.543	0.340
AUC	0.613	0.581	0.528	0.501	0.732	0.692	0.654	0.591

Table 4. Statistics on prediction a nd test (Talki g Face and E yeblink8 datas ets).

Dataset	Talking Face				
EAR Threshold (t)	0.18	0.2	0.225	0.25	0.18
Statistics on the prediction set are Total Number of Frames Processed	5000	5000	5000	5000	10,663
Number of Closed Frames	227	292	352	484	404
Number of Blinks	31	42	49	59	37
Statistics on the test set are Total Number of Frames Processed	5000	5000	5000	5000	10,663
Number of Closed Frames	153	153	153	153	107
Number of Blinks	61	61	61	61	30
Eye Closeness Frame by Frame Test Scores					
Accuracy	0.971	0.968	0.959	0.933	0.970

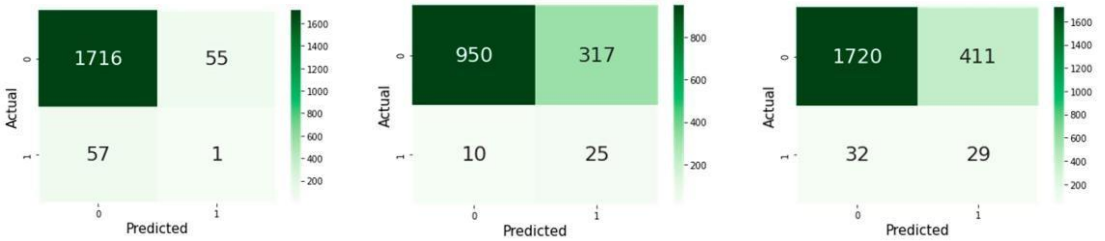
The best EAR threshold in our experiment was 0.18. This value provided the best accuracy and AUC values in all experiments. Hence, 0.25 is the worst EAR threshold value because it obtained the minimum accuracy and AUC values. Based on the experimental results, it can be concluded that the higher the EAR threshold, the lower the accuracy and AUC performance. In previous studies, it was said that the EAR threshold of 0.2 is the best value, but it was not for our experiment. In fact, 0.18 was the best EAR threshold in our work. Our dataset is unique because of the small eyes. The size of the eyes will certainly affect the EAR and EAR threshold values. Therefore, our dataset also has some challenges, namely, people driving cars and people wearing glasses.

Figure 6 shows the confusion matrix for the video 1, video 2, and video 3. Figure 6a describes the false positive (FP) value of 58 out of 58 positive labels (1.0000%) and false negative (FN) rate of 23 out of 1771 negative labels (0.0130%) for video 1 and EAR threshold 0.18. Figure 6c explains the false positive (FP) rate of 35 out of 61 positive labels (0.5738%) and false negative (FN) rate of 206 out of 2131 negative labels (0.0967%).



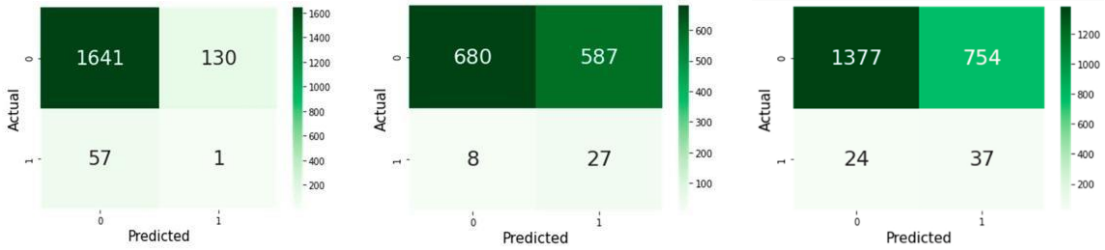
Video 1 Video 2 Video 3

0.18



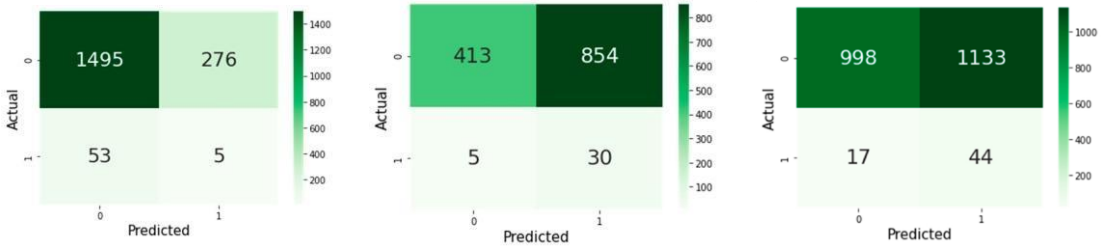
(**a**) (**b**) (**c**)

0.2



(**d**) (**e**) (**f**)

0.225



0.25

(**g**) (**h**) (**i**)

(**j**) (**k**) (**l**)

Figure 6. Confusion matrix (Eye Blink dataset). (**a**) Video 1 with 0.18 EAR threshold, (**b**) Video 2 with 0.18 EAR threshold, (**c**) Video 3 with 0.18 EAR threshold, (**d**) Video 1 with 0.2 EAR threshold, (**e**) Video 2 with 0.2 EAR

threshold, (**f**) Video 3 with 0.2 EAR threshold, (**g**) Video 1 with 0.225 EAR threshold, (**h**) Video 2 with 0.225 EAR threshold, (**i**) Video 3 with 0.225 EAR threshold, (**j**) Video 1 with 0.25 EAR threshold, (**k**) Video 2 with 0.25 EAR threshold, (**l**) Video 3 with 0.25 EAR threshold.

In our experiment, we analyzed videos frame by frame and identified eye blinks every three frames, as shown in Figure 7. The results of the experiment only show the blinks at the beginning, middle, and end frames. For instance, Figure 7a illustrates the 1st blink started in the 3rd frame, the middle of the action was in the 5th frame, and it ended in the 7th frame. Next, Figure 7b describes that the 2nd blink started in the 1555th frame, the middle of action was in the 1556th frame, and it ended in the 1557th frame. Moreover, Figure 7c explains the 3rd blink started in the 1563rd frame, the middle of the action was in the 1564th frame, and it ended in the 1565th frame.

Figure 7. Video 1 eye blink prediction results frame by frame (threshold = 0.18). (**a**) The 1st blink started in the 3rd frame, the middle of the action in the 5th frame, and it ended in the 7th frame. (**b**) The 2nd blink started in the 1555th frame, the middle of the action was in the 1556th frame, and it ended in the 1557th frame. (**c**) The 3rd blink started in the 1563rd frame, the middle of the action was in the 1564th frame, and it ended in the 1565th frame.

Figure 8 exhibits the Video 3 eye blink prediction result frame by frame with the EAR threshold of 0.18. The 1st blink started in the 69th frame, the middle of the action was in the 71st frame, and it ended in the 72nd frame as shown in Figure 8a. The 2nd blink started in the 227th frame, the middle of the action was in the 230th frame, and it ended in the 232nd frame, as described in Figure 8b. Next, Figure 8c illustrates that the 3rd blink started in the 272nd frame, the middle of the action was in the 273rd frame, and it ended in the 274th frame.

Figure 8. Video 2 Eye Blink Prediction result frame by frame (Threshold = 0.18). (**a**) The 1st blink started at the 69th frame, the middle of the action was in the 71st frame, and it ended in the 72nd frame. (**b**) The 2nd blink started at the 227th frame, the middle of the action was in the 230th frame, and it ended in the frame. (**c**) The 3rd blink started at the 272tnd frame, the middle of the action was in the 273rd frame, and it ended in the 274th frame.

Tables 5 and 6 present the specific results that were obtained for each video dataset. These tables include the precision, recall, and F1-score measurements. In our testing, the best EAR threshold was 0.18, which was excellent. In all tests, this number yielded the highest accuracy and AUC values, respectively. As mentioned, 0.25 was the worst EAR threshold value, since it only achieved the bare minimum in terms of accuracy and AUC. Researchers normally choose 0.2 or 0.3 as the EAR threshold, despite the fact that not everyone's eye size is the same. As a result, it is better to recalculate the EAR threshold to detect whether the eye is closed or open in order to identify the blinks more accurately. For video 1 we achieved 96% accuracy, followed by video 3, with 89% accuracy, and video 2, with 86% accuracy, as shown in Table 5. Using Talking Face and Eyeblink8 datasets, we obtained the same accuracy, 97%, by employing the 0.18 EAR threshold, as shown in Table 6.

Table 5. Precision, recall, and F1-score (Eye Blink dataset).

Evaluation	Video 1				Video 2				Video 3			
	Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
	EAR Threshold (t) = 0.18				EAR Threshold (t) = 0.18				EAR Threshold (t) = 0.18			
0	0.97	0.99	0.98	1771	0.98	0.87	0.92	1267	0.98	0.90	0.94	2131
1	0.00	0.00	0.00	58	0.10	0.51	0.17	35	0.11	0.43	0.18	61
Macro avg	0.48	0.49	0.49	1829	0.54	0.69	0.55	1302	0.55	0.66	0.56	2192
Weight avg	0.94	0.96	0.95	1829	0.96	0.86	0.90	1302	0.96	0.89	0.92	2192
Accuracy			0.96	1829			0.86	1302			0.89	2192
	EAR Threshold (t) = 0.2				EAR Threshold (t) = 0.2				EAR Threshold (t) = 0.2			
0	0.97	0.97	0.97	1771	0.99	0.75	0.85	1267	0.98	0.81	0.89	2131
1	0.02	0.02	0.02	58	0.07	0.71	0.13	35	0.07	0.48	0.12	61
Macro avg	0.49	0.49	0.49	1829	0.53	0.73	0.49	1302	0.52	0.64	0.50	2192
Weight avg	0.94	0.94	0.94	1829	0.96	0.75	0.83	1302	0.96	0.80	0.86	2192
Accuracy			0.94	1829			0.75	1302			0.80	2192
	EAR Threshold (t) = 0.225				EAR Threshold (t) = 0.225				EAR Threshold (t) = 0.225			
0	0.97	0.93	0.95	1771	0.99	0.54	0.70	1267	0.98	0.65	0.78	2131
1	0.01	0.02	0.01	58	0.04	0.77	0.08	35	0.05	0.61	0.09	61
Macro avg	0.49	0.47	0.48	1829	0.52	0.65	0.39	1302	0.51	0.63	0.43	2192
Weight avg	0.94	0.90	0.92	1829	0.96	0.54	0.68	1302	0.96	0.65	0.76	2192
Accuracy			0.90	1829			0.54	1302			0.65	2192
	EAR Threshold (t) = 0.25				EAR Threshold (t) = 0.25				EAR Threshold (t) = 0.25			
	0	0.97	0.84	0.90	1771	0.99	0.33	0.49	1267	0.98	0.47	0.63	2131
	1	0.02	0.09	0.03	58	0.03	0.85	0.07	35	0.04	0.72	0.07	61
Macro avg	0.49	0.47	0.47	1829		0.51	0.59	0.28	1302	0.51	0.59	0.35	2192

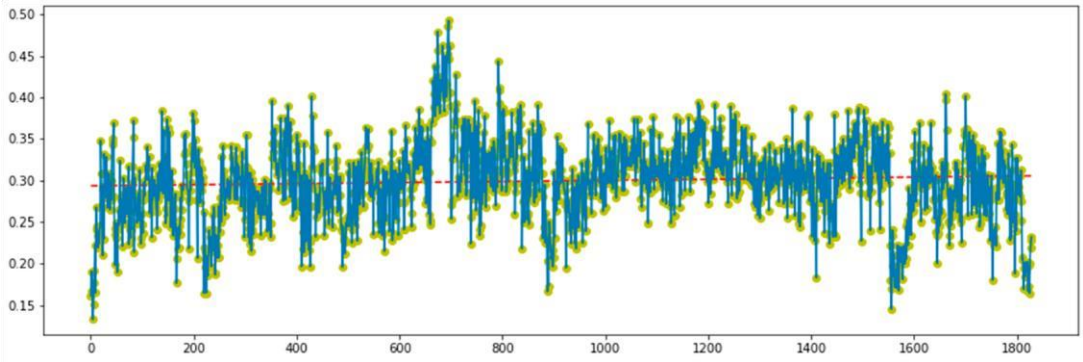
Table 6. Precision, recall, and F1-score (Talking Face and Eyeblink8 datasets).

Evaluation	Talking Face				Eyeblink8 Video 8			
	Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
		EAR Threshold (t) = 0.18				EAR Threshold (t) = 0.18		
0	0.99	0.98	0.99	4847	1.00	0.97	0.98	10,556
1	0.52	0.77	0.62	153	0.24	0.92	0.38	107
Macro avg	0.76	0.87	0.80	5000	0.62	0.94	0.68	10,663
Weight avg	0.98	0.97	0.97	5000	0.99	0.97	0.98	10,663
Accuracy			0.97	5000			0.97	10,663
		EAR Threshold (t) = 0.2				EAR Threshold (t) = 0.2		
0	1.00	0.97	0.98	4847	1.00	0.96	0.98	10,556
1	0.49	0.93	0.64	153	0.19	0.96	0.32	107
Macro avg	0.74	0.95	0.81	5000	0.60	0.96	0.65	10,663
Weight avg	0.98	0.97	0.97	5000	0.99	0.96	0.97	10,663
Accuracy			0.97	5000			0.96	10,663
		EAR Threshold (t) = 0.225				EAR Threshold (t) = 0.225		
0	1.00	0.96	0.98	4847	1.00	0.91	0.95	10,556
1	0.43	0.99	0.60	153	0.10	1.00	0.18	107
Macro avg	0.71	0.97	0.79	5000	0.55	0.96	0.57	10,663
Weight avg	0.98	0.96	0.97	5000	0.99	0.91	0.95	10,663
Accuracy			0.96	5000			0.91	10,663
		EAR Threshold (t) = 0.25				EAR Threshold (t) = 0.25		
0	1.00	0.93	0.96	4847	1.00	0.91	0.95	10,556
1	0.32	1.00	0.48	153	0.10	1.00	0.18	107
Macro avg	0.66	0.97	0.72	5000	0.55	0.96	0.57	10,663
Weight avg	0.98	0.93	0.95	5000	0.99	0.91	0.95	10,663
Accuracy			0.93	5000			0.91	10,663

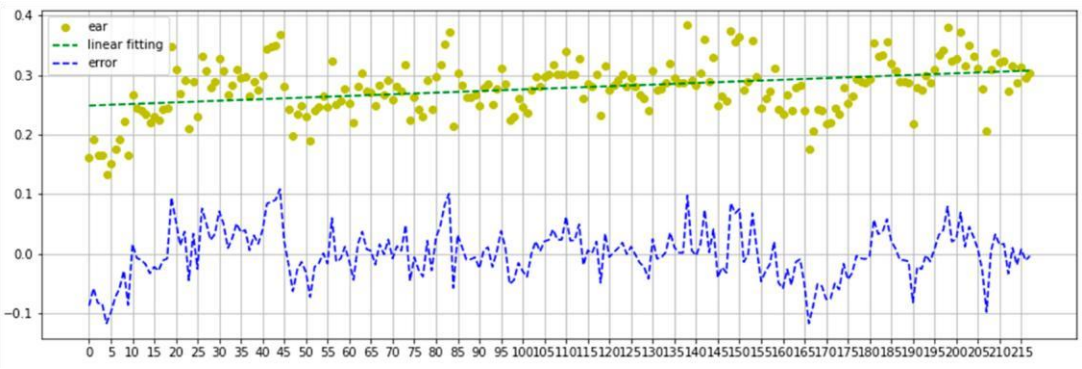
4. Discussion

The EAR and error analysis of the video 1 dataset is presented in Figure 9. The EAR threshold for this dataset was set to 0.18. Assume that the linear regression's optimum slope is m ≥ 0. All the data from video 1 were plotted

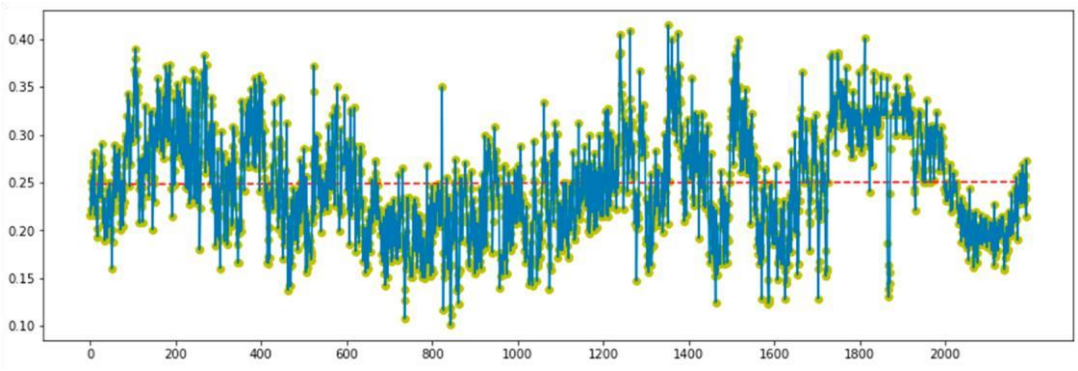
$$= \frac{}{\text{Precision} + \text{Recall}}$$



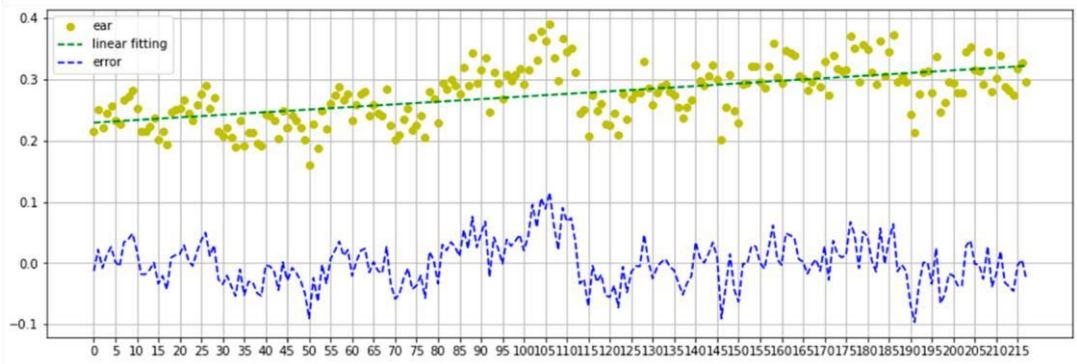
(**a**)



(**b**) **Figure 9.** EAR and error analysis (Video 1, EAR threshold = 0.18). (**a**) Average EAR. (**b**) Error and EAR.



(**a**)



(**b**) **Figure 10.** EAR and error analysis (Video 3, EAR threshold = 0.18). (**a**) Average EAR. (**b**) Error and EAR.

There is a possibility that if the driver does not blink for a long time and his EAR value decreases without any blinking in the initial period, the algorithm will not return an error. Further, our work calculates errors as *errors = calibration – linear* and *cumsum ()*. This function will return the cumulative sum of the elements along a given axis. A new array holding the result, in which case a reference to out is returned, is returned unless out is specified. The result has the same size and shape as if the axis were none or a 1d array. Cumulative errors are not very important for blinking, as their effects are delayed. However, typical errors can be exploited to detect anomalies. Figure 10 describes the EAR and error analysis of the video 3 dataset with an EAR threshold of 0.18. The average EAR value for video 3 was 0.25, as shown in Figure 9b. This average value is slightly different from the average value in Figure 10a, which is close to 0.30. During our tests, we found that the optimal EAR threshold is 0.18. This

figure yielded the greatest accuracy and AUC values in all tests, both excellent results. The statistics are listed in Tables 5 and 6.
Furthermore, Table 7 describes the evaluation of the proposed method in comparison to existing research. Our proposed method achieved peak average accuracies of 97.10% with the TalkingFace dataset, 97.00% with the Eyeblink8 dataset, and 96% with the Eye Video 1 dataset. We improved on the performances of previous methods.

Table 7. Evaluation of the proposed method in comparison to existing research.

7. Problems Associate in Driver Drowsiness Detection

Security is our primary priority during travels or driving. A driver'smistake can lead to severe physical injury, death, or severe economic loss. Many systems, such as navigation systems, sensors, etc. are currently available on the market to make the driver work easy. There are many causes, in particular human mistakes that lead to road crashes. After recent years,sources show that there has been a massive increase in road injuries in our country. The main cause of traffic crashes is the driver's sleepiness and fatigue when driving. An effective technique is required to detect drowsiness as soon as the driver is sleepy. This will save a lot of accidents.

One of the major causes of road accidents is driver drowsiness. This is a major issue for road safety. If drivers have warned of these crashes before they were too drowsy to drive in safety. To detect drowsiness reliably, it relies on the timely warning of drowsiness. Drowsy is one of the main reasons behind traffic accidents which place the driver significantly at higher risk of accidents than driving alertly.

8. Measures and Techniques Used for Detection Drowsiness

Numerous facial features may be drawn from the face to deduce the degree of drowsiness. These compriseeye blinks, head movements & yawning. Nonetheless, it is a task to build a drowsy detection method that offers consistent & accurate results as detailed and robust algorithms are required. A

Reference	Dataset Accuracy (%)		
	Talking Face	Eyeblink8	Eye Video 1
Drutarovskys et al. [7]	92.20	79.00	-
Fogelton et al. [46]	95.00	94.69	-
Proposed Method	97.10	97.00	96.00

variety of methods fordrowsy detection have been examined in the past time. The recent increase in deep comprehension requires revisions of such algorithms to detect drowsiness accurately. This paper explores techniques for machine learning that include SVM, CNN, and HMM in the context of the detection of drowsiness.Table 2 and the included are briefly described the approaches below.

8.1. Support Vector Machine (SVM)

SVMs have supervised classification and regression approaches to learning. Boser, Guyo&Vapnik initially introduced SVMs in 1992. SVMs aim to find a hyperplane that divides training data into predefined paths. In the field of driver drowsiness, SVMs are mainly applied to identify the various driver state from the defined data. There have been many efforts in the detection of drowsiness using the SVMs. A variety of measures were used to assess driver drowsiness rate using SVMs. Table 2 presents a comparison of these steps and briefly describes the methods below.

8.2. Hidden Markov Model (HMM)

HMMs are a mathematical model used to simulate hidden states based on detected probabilitydefined parameters. Leonard Baum and colleagues developed HMMs at the end of the 1960s and early 1970s. Today, HMMs are commonly used for applications like facial expressionrecognition, gene annotation, DNA sequence error modeling, and computer virus classification. Table 2 indicates different features and methods applied by HMM-DDs, but Zhang et al. and Choi et al. omitted details essential to compare their results. The Authors suggested a new facial feature using wrinkle changes that are detected by measuring the intensity of the local edge on the face. They have also applied an IR camera to avoid changes in light to allow including day-night operation. Fortunately, if peopleused this older system could give false results since the wrinkles are deeper. By contrast with HMM, methods for eye-tracking based on color and geometric features have been applied. Authors use a two-stage Lloydmax calculation to extract light to ensure that changes in lighting are robust. This machine is unfortunately built for enclosed operating conditions & if the driver is not onward then fails to recognize the face.

8.3. CNN

CNN's are the same as a normal neural network and often consist of neurons for learning weights. CNN uses spatial layers that are suitable for images and that represent effective spatial correlations.
CNN has proved effective in the areas of imagerecognition, video analysis, and classification. Yann LeCun and YoshuaBengio were the first to provide a computer vision in CNN. Table 2 presents a short review of the CNN-based drowsiness detection methods [26].

8.4. Viola-Jones Algorithm

Viola-Jones technique is based on input image processing or extracts pixel-based information to aid in the detection of features. Optimization of this window detects image faces of different sizes. It is an invariant detector in sizes also and runs through the image with different sizes each time. The required component issimilar no. of calculations, regardless of the image size, because it is invariant in scale. Viola-Jones is a cascade detector. The first phase contains detectors that only remove those image parts that do not face. The complexity of detectors has been further extended in later phases to further analyze the characteristics. A face is only detected as the entire cascade is passed through [27].

Table 2. SVM, HMM & CNN Techniques on Drowsiness Detection.

Author	Year	Metric	Methods	Classifiers	Accuracy%
F. Zhang, J. Su [28]	2017	Eye state	AdaBoost, LBF & PERCLOS	CNN	95.18
A. George and A. Routray [29]	2016	Eye state	Viola & Jones algo	CNN	98.32
A. Punitha, M. K. Geetha [30]	2014	Eye state	Adaboost& HMM	SVM	93.5
M. Sabet [31]	2012	Eye state	LBP-SVM	SVM	98.4

B. Zhang [32]	2012	Eye state	Cooccurrence matrix of oriented gradients (CMOG)	HMM	95.9
A. Bagci and R. Ansari [33]	2004	Eye state	Baum-Welch algo	HMM	99.7
Luo, R. C.[34]	2020	Eye state	Perspective-nPoint (PNP)	multimodel fusion	75.447%

Then,the SVM was trained to class as well as to trigger an alarm when the eyes open or closed. The developers of [30,31] suggested a method to detect drowsiness and distraction of the driver as well. So, this Viola & Jones algorithm was used with local binary patterns (LBP) to detect faces and color histograms to track faces over frames. The system produced an accuracy of 100% in face recognition, but the expected low frame rate that can lead to missing facial expressions which is potential decreases. The variety of features and approaches used by HMM drowsiness detectors are seen in Table 2 but Zhang et al. [32] omitted the details needed to compare its results and it is not part of the meta-analysis stage. Instead [33], introduced HMM eye detection methods based on color and geometric features. The authors used a 2-level Lloydmax empirical to be stable for illumination changes [33]. This system is predictably conceived to be enclosed and fails to identify the face of the driver who doesn't look ahead. Table 2 also provides a brief overview of CNN & multi-modal fusion detection methods.

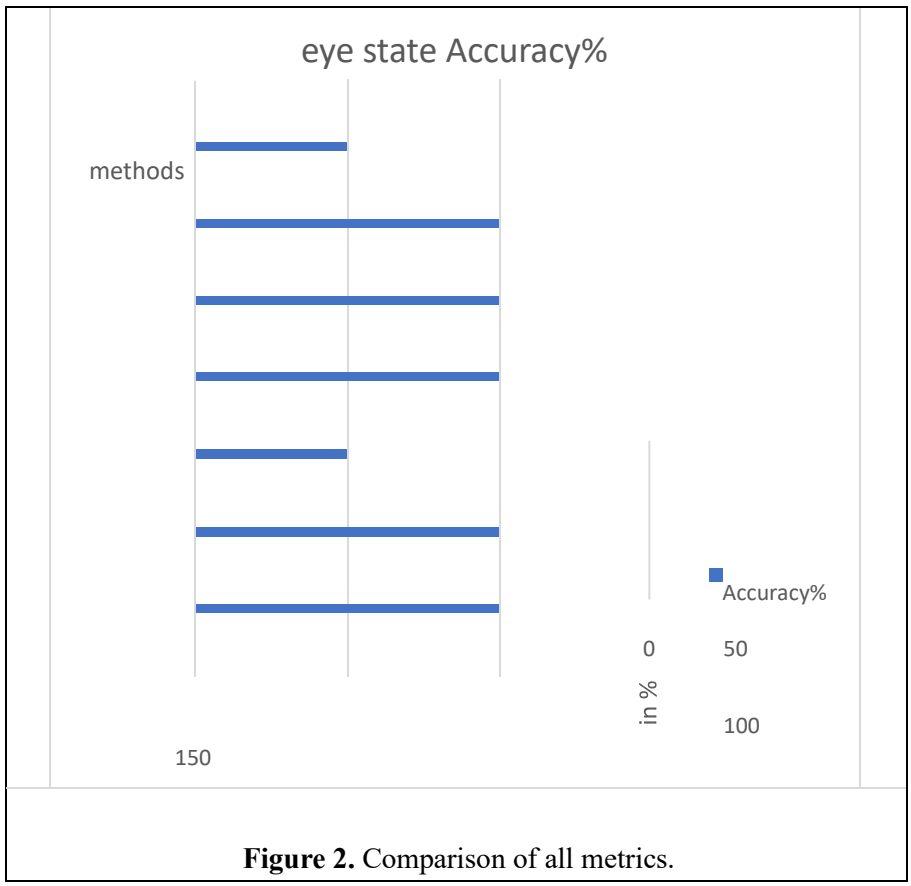


Figure 2. Comparison of all metrics.

As an outcome, it is difficult to compare methods by only estimating reported accuracies. An analysis of measures creating the driver's drowsiness detection system is now explored on machine learning techniques to identify various levels of drowsiness.

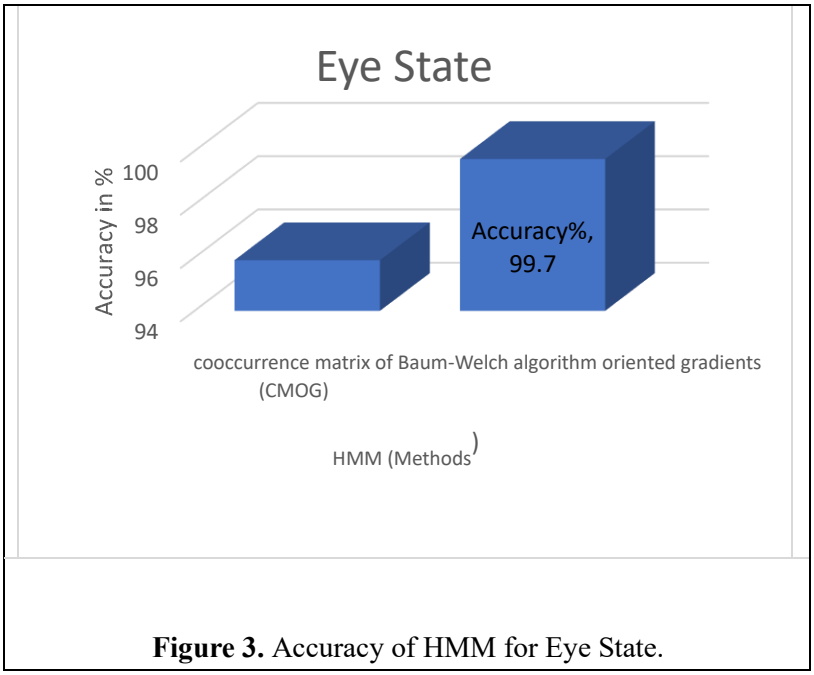


Figure 3. Accuracy of HMM for Eye State.

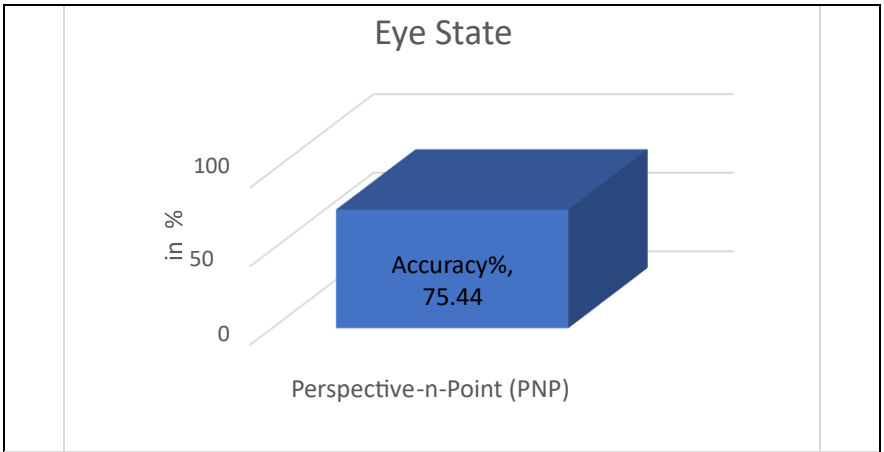


Figure 4. Accuracy of Multimodel Fusion for Eye State.

Table 3. Accuracy of eyes, Face, Mouth detection [35].

Metric	Total frame	Correct frame	Error frame	Miss frame	Accuracy
Face detection	3778	3758	3	17	99.47 %
Eyes detection	3761	3548	89	124	94.33 %
Mouth detection	3761	3664	97	0	99.80 %

Table 3 shows the accuracy rate for Face detection, Eyes detection, and Mouth detection metrics on the frame numbers.

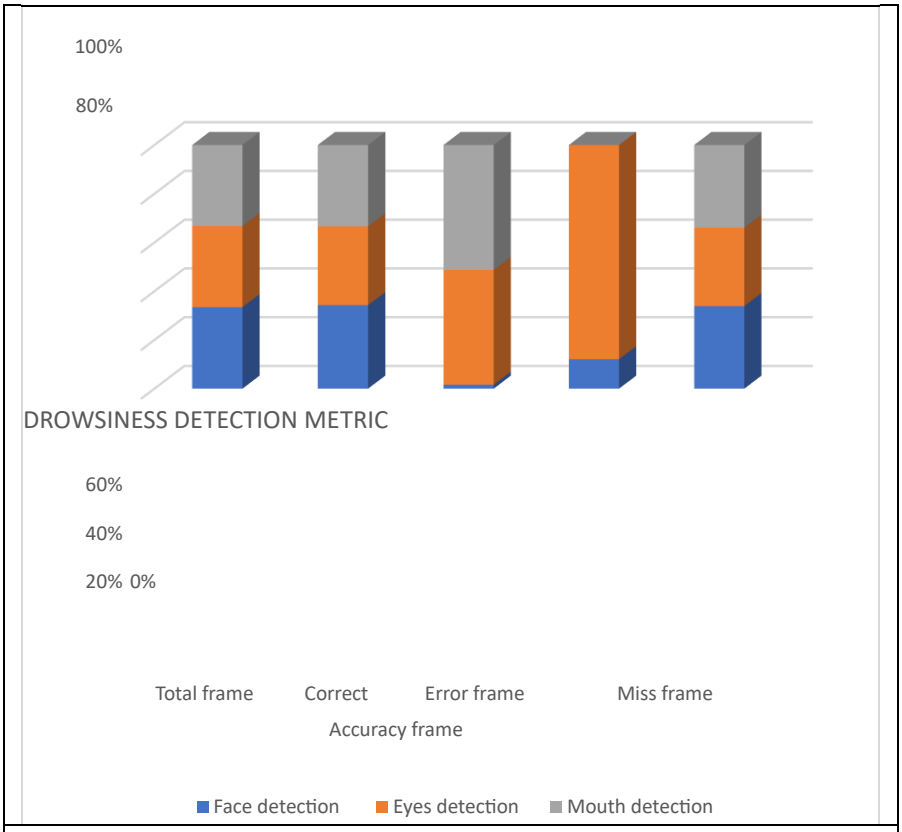


Figure 5. Comparison graph of Eyes, Face, and Mouth detection for Drowsiness Detection.

9. Conclusion

Many vision applications need precise and effective image segmentation methods and classification processes to analyze visual information & perform real-time decision making. Drowsy driving may be almost as fatal as drunk driving. The drivers' drowsiness doesn't only lay itself but it risks for everyone else. Tired & sleepy drivers have delayed reactions & have made poor decisions. One of the main advantages of the drowsy driver detection system is promotingsafety. Several technologies are obtainable to identify the drowsiness of the driver and every approach through its limitations. In this work, we have discussed and analyzed driver drowsiness with their approaches. There are various machine learning methods available to detect driver drowsiness. Machine learning is a paradigm that could apply to learning from previous experience to improve future success (in this case prior data). This paper consists of a support vector machine, convolution neural network, Hidden Markov model, and multi-model fusion classifiers. Such all classifiers usedseveral different methods and hybrid methods to detect drowsiness using eye state metric.It revealed that SVM technology is the most widely applied method to detect drowsiness, and recently CNN is the most suitable classifier, but HMM is performed better than the other two. It also tested face detection and mouth detection metrics with eye detection metric on the frame numbers and we found that drowsiness can detect more accurately (i.e. 99.80%) using mouth detection metric.

10. References

[1] Ramzan M, Khan H U, Awan S M, Ismail A, Ilyas M and Mahmood A 2019 *IEEE***7** pp 6190461919

[2] Kumar K S 2020 *Int. Res. J. of Eng. and Technol. (IRJET)* **07**3 [3] DhargaveS C 2015 *Int. J. of Current Eng. and Scient. Res. (Ijcesr)***2** 2 [4] Kaur S and Jindal S 2016 *Int. J. of Innov. Res. in Adv.*

*Eng. (IJIRAE)***3**11 [5] MariyammalC 2018 *Int. Adv. Res. J. in Sci. Eng. and Technol (IARJSET)***5**11 [6] Chisty 2015 *Int. J. of Comp. Sci. Tren. and Technol. (IJCST)***3** 4 *[7]* Kumari B M K and KumarP R 2017 *Int. Conf. on Big Data Anal. and Comput. Intell. (ICBDAC)*

[8] Dunbar J, Gilbert J E and Lewis B2020 Exploring differences between self-report and electrophysiological indices of drowsy driving: A usability examination of a personal braincomputer interface device *J. of Safety Research***74**pp 27-34

[9] Farahmand B and BoroujerdianA M 2018 Effect of road geometry on driver fatigue in monotonous environments: A simulator study *Transportation Research Part F: Traffic Psychology and Behaviour***58**pp 640-651

[10] Falahpour M, Chang C, Wong C W, Liu T T 2018 Template-based prediction of vigilance fluctuations in resting-state fMRI *NeuroImage***174**pp 317-327

[11] Panning A, Al-Hamadi A and Michaelis B2011 A color-based approach for eye blink detection in image sequences*IEEE International Conference on Signal and Image Processing Applications (ICSIPA)* pp 40-45

[12] Hu S and Zheng G 2008 Driver drowsiness detection with eyelid related parameters by Support Vector Machine *Expert Systems with Applications* **36**4 pp 7651-7658.

[13] Perez C A, LazcanoV A and Estevez P A 2007 Real-Time Iris Detection on Coronal-AxisRotated Faces I*n IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)***37**5 pp 971-978

[14] Nair I R 2016 *Int. J. of Eng. and Comp. Sci.***5** 11 pp. 19237-19240 [15] Choudhary P 2016 *Int. Res. J. of Eng. and Technol. (IRJET)* **03** 12 *[16]* QiuJ et al. 2016 *EURASIP J. on Adv. in Sign. Pro.* [17] Kaur S and Jindal S 2016 *Int. J. of Innova. Res. in Adv. Eng. (IJIRAE)***3** 11 [18] Hachisuka S 2013 *Int. Conf. on Biometrics and Kansei Eng.* pp 320-326 [19] Manu B N 2016 *12th Int. Conf. on Innova. in Info. Tech. (IIT)* pp 1-4 [20] Yan J, Kuo H, Lin Y and Liao T 2016 *Int. Sympos. on Comp. Consu. and Cont. (IS3C)* pp 243246 [21] Hussein W and El-SeoudM S A 2017 *Euro. Conf. on Elec. Eng. and Comp. Sci. (EECS)* pp 261265

[22] Fouzia, Roopalakshmi R, Rathod J A, Shetty A S and SupriyaK 2018 *Second Int. Conf. on Invent. Comm. and Comp. Technol. (ICICCT)* pp 1344-1347

[23] Baek J W, Han B, Kim K, Chung Y and Lee S 2018 *Tenth Int. Conf. on Ubiqu. and Fut. Net. (ICUFN)* pp 73-75

[24] Eraldo B, Quispe G, Chavez-Arias H, Raymundo-Ibañez C and Dominguez F 2019 *IEEE 39th Cent. Amer. and Panama Conv. (CONCAPAN XXXIX)* pp 1-6

[25] R K M, R V, and Franklin R G 2019 *Int. Conf. on Vis. towards Emer. Tren. In Comm. and Net. (ViTECoN)* pp 1-5 [26] Ngxande M, Tapamo JR and Burke M 2017 *Patt. Recog. Ass. of South Africa and Robo. and Mech. (PRASA-RobMech)* [27] BangalS, Bangal S and Sagvekar V2017 *Int. J. of Engg. Res. & Technol. (IJERT)*5 1

[28] Zhang F, Su J, Geng L and Z. Xiao Z 2017 *Proc. Int. Conf. Mach. Vis. Inf. Technol. C.* pp 105– 110

[29] George A and RoutrayA 2016 *Int. Conf. Sign. Process. Comm.* pp 1–5

[30] Punitha A, Geetha M K and SivaprakashA 2014 *Int. Conf. Cict., Power Comp. Technol. (ICCPCT-2014)* pp 1405–1408

[31] Sabet M, Zoroofi R A, Haghighi K S and SabbaghianM 2012 *20th Iran. Conf. Elec. Eng. (ICEE2012)*pp 1247– 1251

[32] Zhang B, Wang W and Cheng B 2015 Adv. Mech. Eng. **7** 2

[33] BagciA and Ansari R 2004 *Recognition ICPR***3** pp 2–5

[34] Luo R C, Hsu C H and Wen YC 2020 *Int. Conf. on Adv. Robo. and Intel. Syst.(ARIS)*pp 1-6

[35] Tipprasert W, Charoenpong T, Chianrabutra C andSukjamsri, C. 2019 *First International Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics (ICA-SYMP)* pp 61-64