Real-Time Computer Vision-based System for Monitoring & Detecting Driver's Drowsiness & Fatigue

Parul Singh R[1], Anil Kumar Gupta[2]
[1]*CDAC ACTS, Department of Computer Engineering, Vellore Institute of Technology, Chennai, Tamil Nadu, India – 600127*
[2]*Senior Member IEEE, CDAC ACTS Pune*

Abstract **- Usually, Driver's eye vision during driving plays a vital role in order to avoid any kind of road accidents, which is a huge number these days in the society we all live in [25]. NCBI claims that the total number of road accidents was 464,910 in India, and it causes 405 deaths and 1,290 injuries daily from 1,274 accidents in our country. So, this is the main reason that the alertness of the drivers during driving is the primary matter of concern to our society. The drivers' attentiveness is affected due to many reasons, like getting a fatigue attack during driving or Alcohol consumption, etc. If such cases exist, there is a likely chance of the vehicle to meet with an accident. So, In, order to avoid such kinds of accidents on the road, there should be some internal safety devices inbuilt in the vehicles initially. These devices should be capable of alerting and protecting the vehicle from the accident by warning about the danger of the particular vehicle to a responsible person. So, the system is developed to determine the alertness of the Driver via an eye region of the specific person using CNN Machine learning algorithm and OpenCV models and deploying the Machine Learning model in a mobile application.**

Index Terms **- face recognition; eye; consciousness detection; CNN; machine learning; OpenCV; Mobile App; Flutter; accident prevention.**

I.INTRODUCTION

THIS paper aims to resolve and prevent the vehicle from the road accidents caused by the lack of attention of the drivers or may the Driver fall sick while driving or when the Driver consumes toxic alcohol and drive. So, to achieve this, a novel system is designed using the Convolution Neural Network algorithm with the OpenCV design.

In the proposed system Machine Learning Algorithm is used to do the Face recognition of the drivers in order to determine whether the Driver is stable and conscious or not for the driving. The stability of the Driver can be determined by focusing on the eye of the Driver's image. If not, then the speed of the vehicle decreases gradually, and the location of the particular vehicle location will be sent to some particular initially feed location (phone number) to rescue the Driver. Here the camera is used for capturing the face and the eye retina of the Driver while driving. After getting the data from the camera, the Convolutional Neural Network algorithm is applied to the data then can determine the positioning of the centre of an eye. With the help of the bare eye, it can determine if the person is alert while driving or not. In this process, the camera from the mobile can be used to process the input data. After getting input data, the processing of the data can be done in the flutter-based mobile application on mobile phones. If the person is detected with some kind of illness in the mobile application, then the alert message about the vehicle will be sent to the Driver's known person or the person who is responsible for that specific vehicle at that time. This, system design of this research will be very much beneficial to determine the consciousness level of the drivers while they are driving instantly with real-time data streaming. And in this system, the integration of the three models has been done in order to increase the accuracy level to determine the consciousness level of the drive while driving. This makes the system design in a way that it can help to alert the drivers during driving and can prevent the vehicle from an accident.

II. LITERATURE REVIEW

Andrew [1] has developed the facial recognition system using facial expressions of Angry, sad, happy, surprised, fear, and neutral. With the help of a few specific features of expressions, the Convolutional

Neural Networks method (CNN) is used to determine the facial expressions of a person.

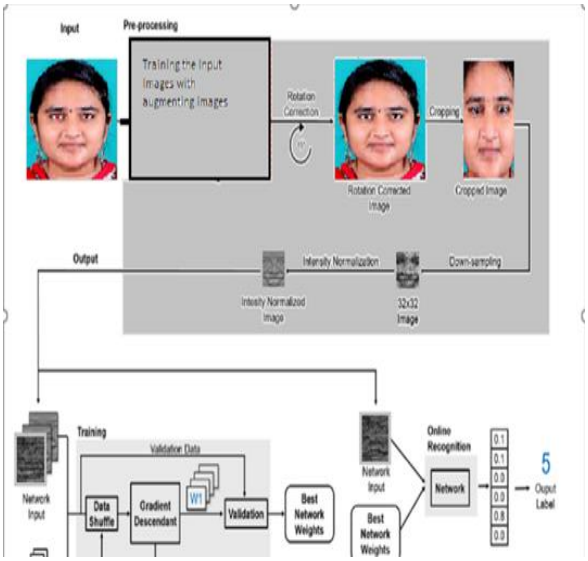

Figure 1:Image processing
image source: researchgate.net/publication

Paweł Tarnowski [2] has built a system that recognizes the emotion of the person with the help of facial expressions. The method uses a 3-dimensional Face Model, and for extracting the features, the classification is done using the KNN classifiers and the MLP classifiers.

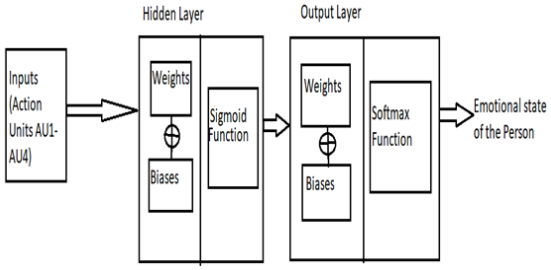

Figure 2: Neural Network Structure
(Image source: researchgate.net/publication)

In this paper, Jiaxing Li [3] used faster R-CNN to identify facial expressions. There are many advantages of facial expression recognition. In this Model, the original image was used as the input. And in the process of feature extraction, facial expression recognition is usually avoided. The features in this Model are extracted by a network from the training dataset by default. Region Proposal Networks (RPNs)were used to generate an efficient and accurate region proposal.

DJude Hemantha has done research work [4] on the proposed approach; a high excitement with positive emotions tends to good emotion. If both the parameters of emotions are weak, then the emotion is negative in nature. In this work [4], the four major emotions are considered for the analysis of the emotion of the subjects.

Bin li [5] solves a most effective method for detecting and analyzing an eye, which has a stage called the candidate region. Then the eye centre position is found.

In Figure 4. In the 1st step, the local extreme points and the gradient values in the full facial image has been calculated, then several candidate eye regions were quickly generated while taking these feature points as the centres; in the 2nd step, these candidate centre eye regions were determined by the 1st set of CNN's to determine the eye regions and the eye class (left or right). In the 3rd step, the 2nd CNNs were useful to locate the eye centre

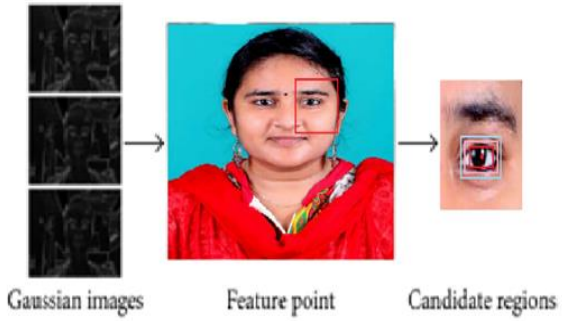

Figure3 Feature point extraction.
(Image source:
https://doi.org/10.1155/2018/1439312)

From the above dataset its been seen that the facial feature points and the candidate eye centre regions.

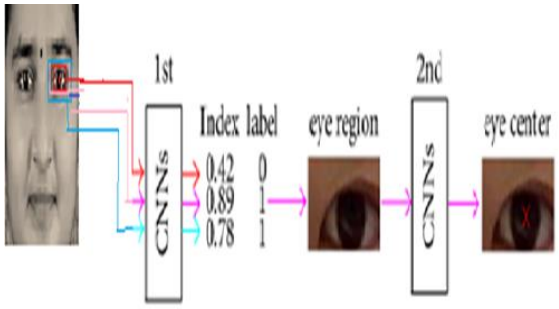

Figure4:1st and 2nd set of CNN
(Image source:
https://doi.org/10.1155/2018/1439312)

In figure 4, the CNNs' output has been displayed. The green point region represents the selected eye candidate feature points. The indexes show the regions predicted by the first set of CNN's. The eye region was fed into the second set of CNN's output as the accurate

eye centre. The centre of the eye represents the red cross.

In figure 5, Eye region detection samples, classification of regions, and eye centre estimation of the three different types of the datasets, there are BioID, GI4E, and their datasets find out the results. And the found-out results are displayed by the red cross on the above figure.

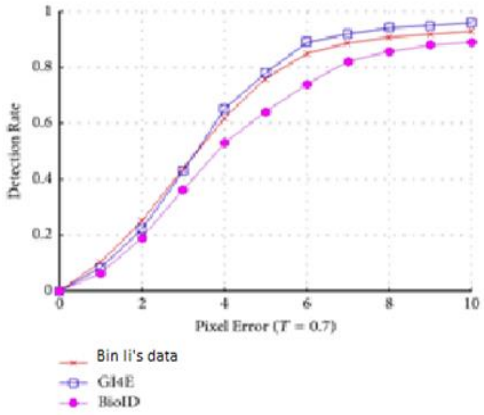

Figure5: Comparison of datasets
(Image source:
https://doi.org/10.1155/2018/1439312)

Comparing the three different datasets with pixel error and detection rate helps us to determine the proper estimated positioning of the centre of an Eye.

In this particular[6] paper, Capito talks about the behaviour of the person after drunk or intoxicated and this paper states that when the person is drunk then he or she has the more anger emotion developed in them usually they will become very much hyperactive in nature, and since they become very much hyperactive in emotions they will be having a very high level of the facial expression developed with themselves, when the person found intoxicated they will be having a rapid change of the facial emotion expressions. An Intoxicated person shows less panic and fear kind of emotions, and this shows that while the person is intoxicated, then they'll become less attentive to the external environment.

This research work [7] of D.G.Jha is done with the detection of alcohol when the person is driving the vehicle. In this work, the detection of the alcohol has been done prior and prevents the Driver from driving when the particular person is found drunk. In this process, the IoT and the Embedded system have been used to get done with the process. The ultimate aim of this work is to prevent a person from driving a vehicle or a car when the person is found drunk or intoxicated. This work of Dada Emmanuel Gbenga[8] has been done in a prototyped model of the device, which uses an IoT and Embedded system and implements a model which will be mainly working with breath analysis of the person(Driver). The only limitation of this work is that it can only detect the abnormalities that occur in the person due to the alcohol only, but in the real-time application of this work, the abnormalities can occur in the person due to many other reasons and may be due to other intoxications. So, In order to prevent the accident, there should be some prediction of the consciousness of the person with respect to all intoxication usually occurs in real-time.

In this paper, Bikaramjit Mann [9] established that alcohol consumption before driving could cause injury crashes. This research shows that the more the person during driving is alcoholic or intoxicated, the more dangerous the injury to the person and the vehicle may occur.

In this research, Karl kim[10] states that the person's emotions and behaviour during driving can be the main factor to predict the accidents or the type of accidents. So in this research regarding these points and facts, the statistical survey in this research has been done.

In this research, Tereza Soukupova[11] has used computer vision technology in order to determine the blink in the given input video to the Model of the system. They took different types of input video data and then compared the value of the outputs altogether.

In this paper, K.P.Yao[12] found a system to determine a vision system for monitoring a driver's consciousness. The level of consciousness is determined by integrating many facial parameters. In order to determine these parameters, the facial features of the eyes, mouth, and head are located in the input video. The located facial features are then tracked over the subsequent images. Facial parameters are estimated during facial feature tracking.

The estimated parametric values are collected by K.P.Yao and analyzed every fixed time interval to provide a real-time consciousness level of the Driver. A series of a demonstration on real-time video sequences were demonstrated to reveal the results of the proposed system: the in-vehicle vision system, Driver's alertness monitoring, facial parameters, fuzzy reasoning.

In this research, Qiang Ji[13] uses Bayesian Networks to estimate car accidents on the road. In this research, Qiang Ji monitors the eye region and the face posture of the subject, and then with the help of the fuzzy logic information fusion Bayesian Networks with the data of the input detection model, the Bayesian Network gives the output, and according to that output, the alert to the system is done.

In this paper, V Sharath[14] uses a Haarcascade Algorithm to determine the consciousness of the person while determining an eye region of a particular subject. In this research, it has been mentioned that if the blinks occur for the duration of 3 to 4 secs, then the person is in normal condition, but if the person exceeds this time range, then there might be some kind of chances of the person is not well or has a symptom of fatigue. Haarcascade Algorithm classifies the face region from the given input image; then, with the help of the output of the Haarcascade classifier, the system determines the condition of the person if the person is in the fatigue condition or a normal condition.

In this research, Mrs Radha Rathod[15] uses the Haarcascade classifier with Raspberry Pi attached with all the sensors in that. In this research, Mrs Radha Rathod used an alcohol sensor for calculating the Intoxication level of the particular subject during driving. Mrs Radha Rathod uses a proper Embedded system for the intoxication detection of the person during driving. When the person is found intoxicated, then with the help of the buzzer, the system provides an alert to the person.

There is another research of Windsor Kwan-Chun Ting [16] about the study of the eye movement of the subject to determine the consciousness level of the subject at that time. In this research, Windsor Kwan-Chun Ting uses the different stages of the subject to get the proper results by using the proper clinical methods for getting a proper result of this research.

In this research of Javier Sánchez Sierra [17], the application for blind people has been created in order to help visually challenged people. In this research, there is an integration of the embedded system with the android and iOS applications.

There is another research of Gabriel Hermosilla [18] et al. about the Infrared method to classify the difference between drunk subjects and the normal subjects with the help of few more techniques like face recognition and classification. In this Model, Gabriel Hermosilla uses the input as thermal Images while creating its new database of drunk person images to train its Model. After recognition of the face image of the subject, the classification of the image done while comparing the drunk subject images and the normal subject images.

III. METHODOLOGY

The proposed system consists of four Algorithms; the first Algorithm is about applying CNN in the face image to extract the facial features than in the second Algorithm, with the help of the EAR threshold values, the consciousness of the subject is found, and in the third Algorithm is all about Haarcascade method to determine the consciousness level of the person, and the Fourth Algorithm is about the usage of the Infrared Images to classify the alcohol drunk person from the normal subject person in this paper. The ultimate idea of this research is all about identifying the consciousness level of the drivers while he/she is driving the vehicle just to avoid the automobile crashes that occur on the roads due to the distraction of the drivers during driving. The overall workflow of this research,

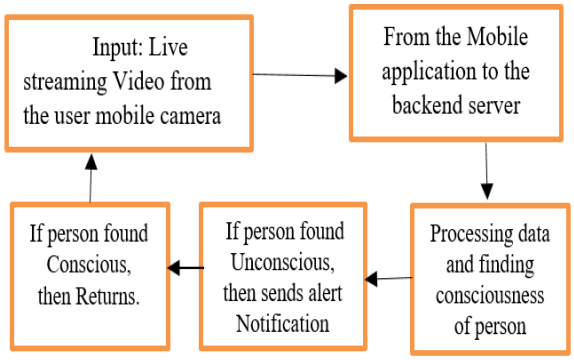

Figure 7, Representing the workflow of this research.

The above figure 7 represents the whole idea of this research; the main idea of this research is to capture the input via the mobile camera of the subject just to make the Model flexible to use by the users while driving, and the main advantage of this would be that it won't cost any additional hardware setup for the user to use this Model while driving. Then the input data sent to the backend server to process the input data finds if the person is currently conscious or not. And if the Model finds the person unconscious, then the mobile application would be sending an alert notification to the Driver and sends the location of that particular vehicle to the related or responsible person

of that vehicle. And if the Model did not find anything, then the process remains the same and then returns.

[A] Algorithm-I (APPLY CNN)

This project determines the conscious level of that particular person via determining the centre of the positioning of an eye with the help of a CNN algorithm. The consciousness of the person(Driver) is determined in order to control or adjust the speed of the vehicle just to prevent the vehicle from an accident. Thus, the overview of this project is further explained in this flow chart.

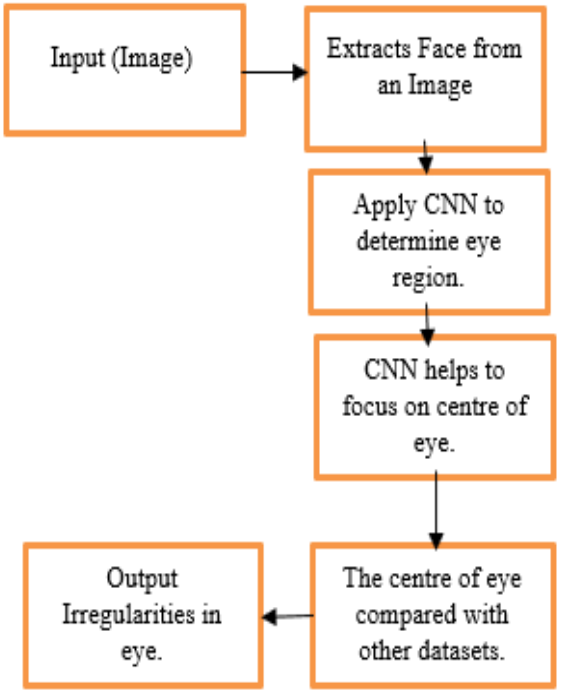

Figure 6, the workflow of the Algorithm-I Procedure

With using the MTCNN (Multi-Tasking Convolution Neural Network) Algorithm, the face region of a model has been determined, and in the second step, the facial landmarks of a model are marked, which can be seen in figure 7

The working Algorithm of the MTCNN method. There are three networks used in this Model; there are P-net, R-net, O-net. Basically, these three networks have the same application, but they will be working while eliminating the wrong values and picking up the right values. So initially, the input is given as an image with the particular level of the pixel values, then the image scaled in Multiple copies of different sizes. Now the scale image has been fed to the P-net, then the output of the P-net is saved, and the Padding of the output has been done. (Padding is the process of filling arrays with the zeroes). Then the Padded image fed to the R-net, and the same process repeated with the O-net as well.

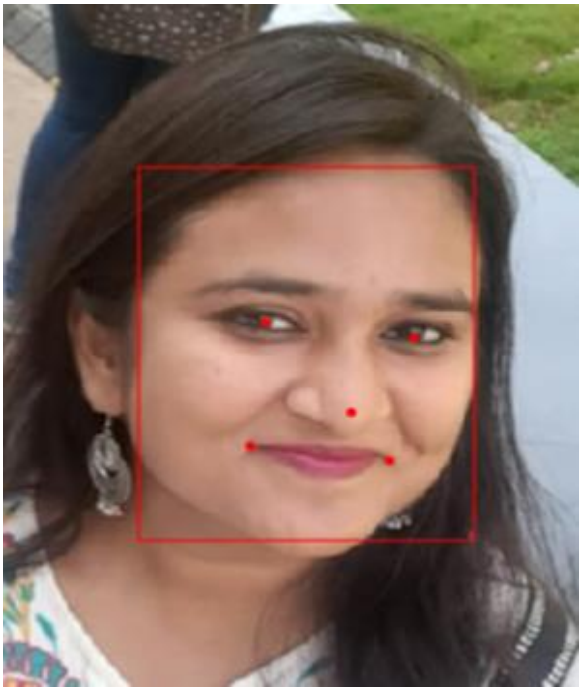

Figure 7 MTCNN algorithm Output

This MTCNN algorithm helped in this research to localize the facial landmarks in order to determine the exact pixel locations of the facial landmarks in the given particular facial image of a model. So with the help of these exact facial landmarks in the image, this will be very much beneficial in order to especially focus on the pixel values of an eye. So basically, the MTCNN algorithm works specifically to determine the facial landmarks from the given input image and determines the pixel values of the facial landmarks.

[B]ALGORITHM-2 (APPLY OPEN CV METHOD)

This method is very much useful to process the live streaming data and processing it. In this method, we use DLIB and CMAKE along with Keras in the Jupyter notebook installed with Anaconda. And here, in this method, we extract the eye region and can determine the blink rate and duration of the blink of the person with the help of the Eye Aspect Ratio (EAR), and this EAR process the data with the following aspects, I. Finding an eye localization II. Find the threshold to find the whites of the eyes III. Determine if the "white" region of the eyes disappears for a particular duration of time (that indicates a blink).

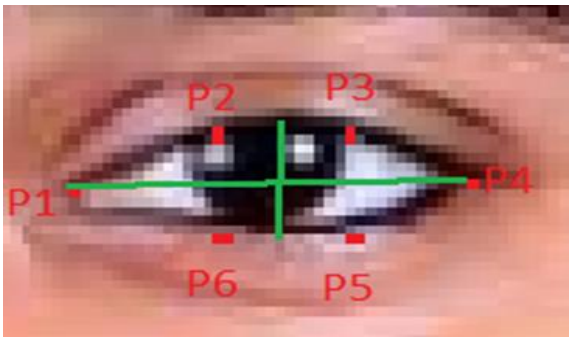

figure 8: Extracted eye region
(image source:
https://www.pyimagesearch.com/2017/04/24/eye-blink-detection-opencv-python-dlib/

On the above image, the picture of the eye is represented with the 6(x,y) coordinates. These coordinates represent the height and width values of the eye, and thus, the EAR demonstrates that there is a relationship between height and width of the eye surface because it indicates the presence of the whiteness of the eye region.

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

Where p1,p2,p3,p4,p5, p6 are 2D facial landmark locations. The value of EAR in this Model is calculated with the help of the OpenCV method; in this method, the open streaming camera data has been used in order to have the continuous consciousness level detection of the Driver while driving. This method helps to have a real-time data analysis.

So this study tries to determine the consciousness level of the person during driving, and it is usually said that when the person tends to get unconscious, their eye will be in the closed condition in the initial stages of the unconsciousness, so in such cases, the blink of the person for the longer duration time can be considered as the abnormal action or unconsciousness.

By calculating this EAR with the help of the live streaming camera data, when the EAR value will fall below the threshold value for a longer duration of time, then the person can be denoted as the unconscious person. Thus, this EAR with the OpenCV method is used to determine the consciousness level of the person.

[C]ALGORITHM-3(HAARCASCADE CLASSIFIER)

This is a classifier algorithm which usually used to determine the region of interest from the given input images. This Algorithm determines the Haar features in the human faces and then finds some of the similar properties with all datasets. The similar properties in the human face are known as the Haar features in this Haar cascade algorithm.

Usually, the eye region is darker than the other parts of the common human's face. So basically, this Algorithm finds the darker region of the eye in the given input image and extracts the eye region from the given face images. The determination of the darker region in the given image is done by finding gradients of the pixel intensities in the given images. The difference in the brighter regions and the darker regions are determined, and then the Haar region of the face is extracted in the given face images with the bounding boxes generated as an output of this method.

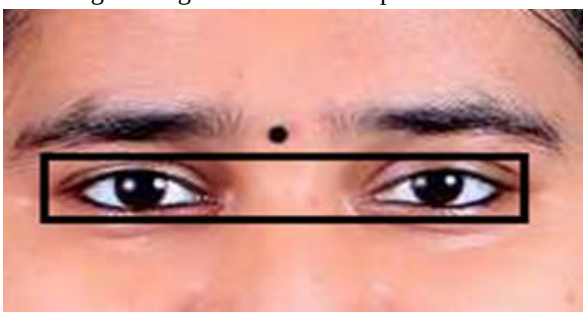

Figure 9: Haar region of the face image.

The above figure 9 demonstrates the Haar region of the face image, which can be extracted with the help of the Haarcascade Algorithm.

[D]ALGORITHM-4(INFRARED METHODS WITH BAYESIAN CLASSIFIER)

This method uses thermal images as an input image. This Algorithm works when we get proper input images from the user and then convert that normal input images in the format of the thermal image format then that particular thermal images used to find the consciousness of the person in the input image to get the result Bayesian classifiers have been used to classify the condition of the person in the given input image. The classification is done while comparing it with the intoxicated person images and then comparing both the images. This method has a very good accuracy rate, but this Algorithm needs a large number of images of the person, and it does also need a large number of images of the intoxicated subjects, which is quite difficult to get in real-time.

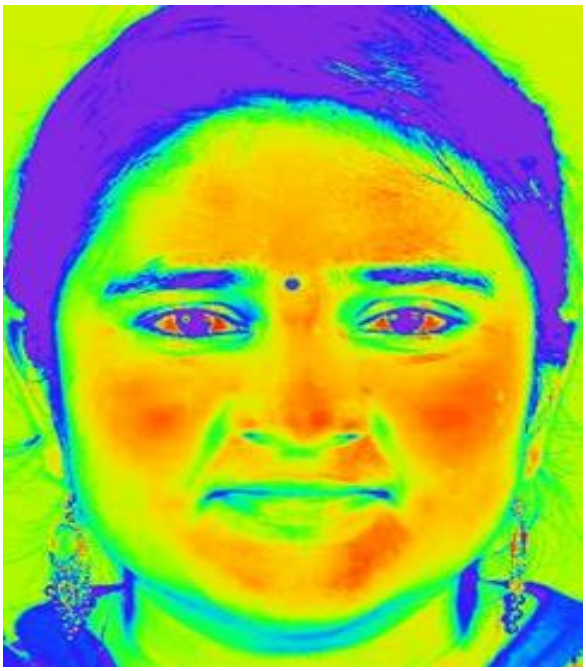

Figure no.10(Sample Thermal Image)

The sample of Infrared images has been shown in figure 10, which shows the reason why the Infrared method contains high accuracy comparatively. As clearly seen in figure 10 that Thermal images can highlight the edges of the objects in the given input image. Because of the major highlight, there are huge pixel value changes that occur in the image, which can help us to extract the exact region of interest while calculating the pixel value changes and finding the edge of each facial feature from the given input images.

[E]CAMPARING ALGORITHMS:

The main aim of this research is to determine the best suitable Algorithm for determining the eye region of the person during driving just to calculate the consciousness and intoxication level of the drivers during driving to prevent the vehicle from car accidents occurs due to the lack of consciousness of the Driver during driving. Below there are different algorithms been compared,

Algorithm Names	Advantages	Disadvantages
CNN Algorithm	This method gives the best results with a Recall value of 89.85% [22] and Precision value of 98.02 %. Which means it gives us the most accurate results.	The classifier like CNN takes more time to train the models, and it needs a large dataset. We are getting a huge number of datasets to feed in our training of Model will comparatively long time.
OpenCV Algorithm using EAR formula	This method has the live streaming input video recognition with comparatively high accuracy and recognizes the eye of the person from the given input live video	Need a proper installation of a Camera device in the place where the consciousness of the person should be found, which is in front of the steering of the car.
HaarCascade Algorithm	This method gives us a result with a Recall value of 82.60% [22] and a Precision value of 95.24%.	This method to extract face from the images works on the images and takes time for the classification process.
InfraRed Method with Bayesian Classifier	This method achieves 99% of accuracy and finds the face from the images as soon as possible.[24]	This method uses the infrared images as an input to the Model and getting a large number of infrared images to train the model is a challenging task.

Table No.1(Comparing algorithms)

III. CONCLUSION

From the above table, it has been seen that all the algorithms have their profits and consequences, but from the survey, it has been clear that using an OpenCV with EAR method will have the best suit for our application which is to prevent the vehicle from the road accidents occurs due to drivers unconsciousness.

The number of accidents caused by the drivers due to illness can be minimized by detecting the alertness of the Driver while detecting the eye of the Driver during driving. And this process can be done by using the CNN algorithm and OpenCV method; while integrating these three models, we can achieve better accuracy in consciousness determination.

REFERENCES

[1] André Teixeira Lopes, "Pattern Recognition", Elsevier15 July 2016.W.-K. Chen, Linear

Networks and Systems (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.
[2] Paweł Tarnowski, "Emotion recognition using facial expressions", Procedia Computer Science 108C (2017) 1175–1184,12 June 2017.
[3] Jiaxing Li, "Facial Expression Recognition with Faster R-CNN", ICICT2017.
[4] Jude Hemantha, "Brain signal based human emotion analysis by circular backpropagation and Deep Kohonen Neural Networks",68(2018) 170-180.
[5] Bin li et al, "Real-Time Eye Detector with Cascaded Convolutional Neural Networks", Hindawi, 22 April 2018, doi.org/10.1155 /2018/1439312.
[6] Capito et al, "Acute alcohol effects on facial expressions of emotions in social drinkers: a systematic review", Correspondence: Eva Susanne Capito Physiological Psychology, University of Bamberg, 3 Markusplatz, Bamberg 96045.
[7] D.G.Jha et al, "Alcohol Detection in Real-Time To Prevent Drunken Driving", IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661,p-ISSN: 2278-8727.
[8] Dada Emmanuel Gbenga et al, "Alcohol Detection of Drunk Drivers with Automatic Car Engine Locking System", Nova Journal of Engineering and Applied Sciences DOI: 10.20286/nova-jeas-060104.
[9] BikaramjitMann et al, "Is Blood Alcohol Level a Good Predictor for Injury Severity Outcomes in Motor Vehicle Crash Victims", Volume 2011, Article ID 616323, 6 pages doi:10.1155/2011/616323.
[10] KARL KIM et al, "PERSONAL AND BEHAVIORAL PREDICTORS OF AUTOMOBILE CRASH AND INJURY SEVERITY", Accid. Anal. and Prev., Vol. 27, No. 4, 469-481, 1995.
[11] Tereza Soukupova et al, "Real-Time Eye Blink Detection using Facial Landmarks", 21st Computer Vision Winter Workshop, Slovenia, February 3–5, 2016.
[12] K. P. Yao et al., "Real-Time Vision-Based Driver Drowsiness/Fatigue Detection System", Computer Science and Information Engineering, National Taiwan University.
[13] Qiang Ji et al, "Viusal Cues Extraction for Monitoring Driver's Vigilance", Department of Computer Science University of Nevada Reno, NV 89557.
[14] V. Sharath et al, "Driver Drowsiness Detection Using Haarcascade Algorithm", International Journal of Research in Engineering, Science and Management Volume-3, Issue-8, August-2020.
[15] Mrs. Radha Rathod et al, "Real-time Driver Drowsiness Alert and Alcohol Intoxication Detection with Vehicle Control", International Journal of Scientific Research and Review ISSN No.: 2279-543X Volume 07, Issue 05, May 2019.
[16] Windsor Kwan-Chun Ting et al, "Eye movement measurement in diagnostic assessment of disorders of consciousness", Firas H. Kobeissy, University of Florida, USA.
[17] Javier Sánchez Sierra et al, "Designing Mobile Apps for Visually Impaired and Blind Users", Centre for Computer Research in Music and Acoustics Stanford University Palo Alto, CA, USA.
[18] Gabriel Hermosilla, "Face Recognition and Drunk Classification Using Infrared Face Images", Hindawi Journal of Sensors, Volume 2018, Article ID 5813514.
[19] https://www.intechopen.com/books/reviews-refinements-and-new-ideas-in-face-recognition /thermal-infrared-face-recognition-a-biometric-identification-technique-for-robust-security-system
[20] S. Baluja and D. Pomerleau. Non-intrusive gaze tracking using artificial neural networks. Technical Report CMUCS-94-102, Carnegie Mel lon University, 1994.
[21] A. Blake, R. Curwen, and A. Zisserman. A framework for spatiotemporal control in the tracking of visual contours. Int. Journal of Computer Vision.
[22] https://datawow.io/blogs/face-detection-haar-cascade-vs-mtcnn.
[23] https://machinelearningmastery.com/how-to-perform-face-detection-with-classical-and-deep-learning-methods-in-python-with-keras/
[24] https://www.researchgate.net/publication/260003874_Infrared_face_recognition_A_comprehensive_review_of_methodologies_and_databases.
[25] https://www.ncbi.nlm.nih.gov/