Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS 2020)
IEEE Xplore Part Number:CFP20K74-ART; ISBN: 978-1-7281-4876-2

Driver Alertness System using Deep Learning, MQ3 and Computer Vision

Aashreen Raorane
Computer Engineering
D. J. Sanghvi College of Engineering
Mumbai, India
aashreenraorane@gmail.com

Hitanshu Rami
Computer Science
Illinois Institute of Technology
Chicago,United States of America
hrami@hawk.iit.edu

Pratik Kanani
Computer Engineering
D. J. Sanghvi College of Engineering
Mumbai, India
pratik.kanani@djsce.ac.in

Abstract— **Being alert at the wheel, is a pressing problem concerning road safety, that many people face in their daily life. This project proposes developing technologies that can help to detect how alert the driver is. It addresses seven different use cases, to provide better feedback to the driver and therefore helps in reducing the number of accidents. The prototype performs Emotion Analysis through facial expressions, Visual Analysis of traffic signboards, calculates the driver's Eye Aspect Ratio to detect drowsiness and mobile distractions and also calculates object proximity from the car using techniques of Computer Vision and Neural Networks. Drunk driving is detected using an alcohol sensor module MQ3. It also provides monthly alerts to the car owner about his vehicle maintenance and service. Every use case has a unique response which aims at alerting the driver to tackle the vital issue of safety at the wheel.**

Keywords— Emotion Analysis, Eye Aspect Ratio, Lane Detection, Proximity Sensing, Computer Vision, Neural Networks

I. INTRODUCTION

Deep Learning can be used to solve a variety of issues in the fields of healthcare and finances [1-2]. Computer Programming combined with hardware can also be used to find solutions to numerous problems [3]. This paper aims to combine Deep Learning techniques and Hardware to develop a system that reduces the accidents on roads. An increasing number of road accidents lead to research on the causes of these accidents. On an average 13 people die every day in India due to the cause of drunk driving [4], making it an important aspect of road safety. The prototype detects the presence of alcohol in the driver's seat of the car and will not permit the car engine to start if the alcohol detected does not lie in the legal limit.

Furthermore, it is proved that lack of sleep for 17 hours has the same effect on driving as would a Blood Alcohol Concentration of 0.05 would have [5]. And being awake for 24 hours has the same effect as a BAC of 0.1 would, which is double the legal limit. Studies say that the primary cause of more than 20% of road fatalities is Drowsy Driving, thus developing an immediate need to ensure the person behind the wheel is not drowsy. This prototype can detect the drowsiness of the driver by constantly calculating the eye aspect ratio and alerting when the EAR does not qualify the required threshold. This ratio can also be used to classify whether or not the driver is distracted by the use of mobile phones.

Road Rage is another of the major causes of road accidents. The prototype performs emotion analysis of the driver to classify if he/she is in a state fit to drive the car. If the person is angry for a long time it is advised not to drive. In this situation, a guardian is informed via a text message to look into the matter and possibly placate or convince the person to not drive. The guardian can access the live feed of the driver via a WebRTC from anywhere around the world and can subsequently communicate with the driver.

A certain number of rules are followed on the road, some of which include the lane system [6]. Drivers are not supposed to cross lanes unless necessary and if done so, need to indicate it to other drivers of its future actions. But when changing lanes often in high speed becomes a habit it could be called as rash driving. The prototype can detect if a vehicle is driving within a certain lane and also notifies the driver in case he/she is unintentionally overriding the limits of the particular lane. In addition to this, Rash Driving can also be detected if the vehicle crosses lanes too often in a certain period.

The prototype also provides the feature of prompting a road sign in the form of audio in case the driver misses it to ensure road safety and provide a better and safer driving experience as the driver will have fewer distractions and more focus on the road. It can also recognize other cars and people on the road and the proximity of these objects is calculated. The driver is alerted via audio if the person or car is too close.

II. PROPOSED SOLUTION

Fig 1. shows Driver Case Diagram of the system to delineate the various functions the system performs. The two main actors in the system are Driver and Car Owner. The driver needs to be alerted whenever necessary in the form of audio. The system specifically does not use any other form of output such as images, videos or vibrations as it may, in turn, lead to distraction of the driver from the road. As mentioned earlier, the system performs various tasks such as Analyzing alertness of the driver, detection of the car lane changes, detection of alcohol, calculating proximity of the objects on the road, analyzing sentiments of the driver, detecting traffic board signs and calculating the maintenance date of the vehicle. After performing these functions, the system must be able to take appropriate steps to increase safety on the road. In the cases where the driver needs to be alerted regarding the situation, an alarm is played. Considering the case in which a lane crossing is detected. If for a long time, the driver is not driving in the proper lane it may be due to the driver's lack of attention or negligence. In such cases, the driver needs to be informed that he/she may not be driving in proper lanes.

If alcohol is detected by the system, the prototype locks

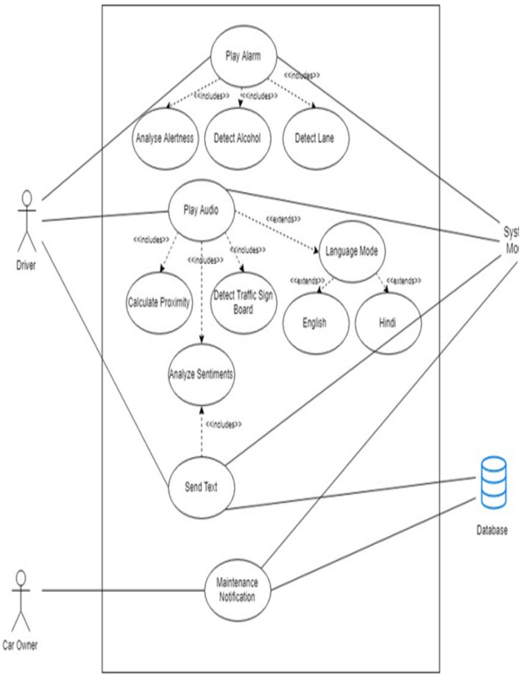

Fig.1. Driver Case Diagram

the car engine which will not allow the car to start. Whereas, if the driver is found to be distracted or not alert due to drowsiness, an alarm is played in order to alert the driver. Therefore, the use case 'Play Alarm' includes the use case 'analyze alertness', 'detect alcohol' and 'detect lane'.

Similarly, the use case 'Play Audio' includes 'Calculate proximity', 'Analyze sentiments/emotions' and 'Detect traffic sign boards'. In these cases, the driver specifically needs to be told about what has been detected which calls for the use of audio. If proximity of a car or a person is calculated the driver is informed about the object being too near. If a traffic sign board is detected, the driver is informed about the sign board. If the driver is found to be angry, a text message is sent to the guardian or the specified number which is fetched from the database. The guardian can then contact the driver by a video call where the live feed of the driver's camera is available to the guardian and the guardian can speak to the driver to possibly placate him.

The prototype also provides the user to choose the language mode. The user/driver can choose from among the languages available. Most of the drivers in India lack a proper education which is why the prototype has the feature to play the audio in Hindi, the national language of India. Currently the prototype is designed to operate in Hindi as well as English.

Typically, a car needs to undergo maintenance after every 35 days not only to ensure proper car condition but also reduce the number of accidents on road due to lack of proper maintenance. The prototype saves the previous maintenance details of the car in the system database along with the upcoming maintenance date and notifies the car owner once his/her maintenance is due.

III. WORKING

A. Drowsiness Detection

In the proposed solution, Drowsiness of the driver is predicted with the help of EAR (Eye Aspect Ratio). EAR is an estimation of what percent of the eye is open. In Real Time Eye Blinking Using Facial Landmarks [7], Soukupová and Čech have come up with an equation to find the Eye Aspect Ratio.

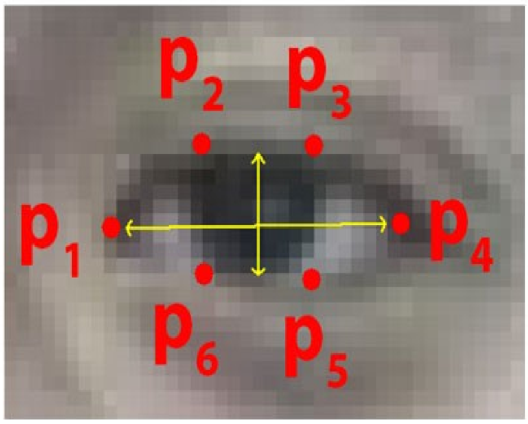

Fig.2.Open Eye with marked positions [7]

Based on Fig 2, EAR can be calculated as:

$$EAR = \frac{||p2-p6||+||p3-p5||}{2||p1-p4||} \tag{1}$$

$$EAR_{th} = 0.3 \tag{2}$$

Normally, for an open eye, EAR is maintained above 0.3. Hence, the value of $\mathrm{EAR_{th}}$ as 0.3 is initialized. To calculate the EAR ,first localize the face and eye region of the driver. A pre-trained Histogram of Oriented Gradients + Linear Support Vector Machine Object Detector for Facial Detection is utilized. It is based on "Histograms of Oriented Gradients for Human Detection" paper by Navneet Dalal and Bill Triggs[8].

The general algorithm using HOG to train a detector works in the following way:

1. Extract Histogram of Oriented Gradients [HOG] for x positive samples of the training data, i.e. which includes the object to detect.
2. Similarly, extract HOG for y negative samples from the training data, i.e. which doesn`t include the object to detect.
3. Now train an SVM on the x and y samples.
4. Using the sliding window protocol on the negative y samples the classifier records the false positives and its associated feature vector.
5. The false positives and their associated feature vector are sorted according to their probability and use this to retrain the classifier.
6. After training the classifier, use it to detect the objects want in our dataset and store the coordinates of the bounding box with the highest probability.

Now, to find the coordinates of the eye on the face detected using the aforementioned method and use the algorithm proposed by Kazemi and Sullivan in their paper "*One Millisecond Face Alignment with an Ensemble of Regression Trees*(2014)" [18] Their algorithm begins by training a regression tree on a dataset which contains images of faces which are labelled to give the coordinates of the surrounding facial regions. The regression tree then can detect the facial landmark without any feature extraction but from pixel intensity itself. Our model has trained on the 68 points iBUG 300-W dataset which can be visualized as shown in Fig. 3.

The live video feed is taken from the driver's camera and is converted into frames. Using dlib library[9], localization of the eye region is performed. Then, using Euclidean distance for each frame, calculation of the EAR can be performed. The model is then coded to play an alarm if the EAR values are recorded to be below the fixed threshold EAR_{th} for a continuous period of at least 35 frames. Furthermore, this can also be used to detect distraction with the use of a mobile phone. As the person looks down to use his mobile, the EAR falls below EAR_{th} which triggers the alarm.

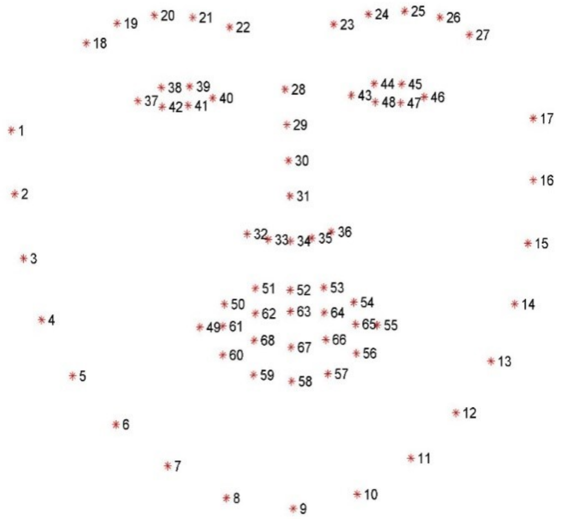

Fig.3. The 68-point mark-up used for annotations. [20-22]

B. *Emotion Analysis to prevent Road Rage*

The dataset Kaggle-fer2013 [10] is used to train the model to perform Emotion Analysis. The dataset contains grayscale images of 48x48 pixels containing faces. The faces are approximately centered and occupy the almost equal area in every image. The CSV file contains the column "emotion" which is a numeric value between 0 and 6 inclusive. Each of the numeric value corresponds to one emotion. 0 equals Angry, 1 equals Disgust, 2 equals Fear, 3 equals Happy, 4 equals Sad, 5 equals Surprise, 6 equals Neutral. Using advance image processing one can classify the emotions of the driver by taking input as a human face. The model is created based on the architecture proposed by Octavio Arriago, Paul G.Plogger and Matias Valdenegro as shown in Fig. 5.

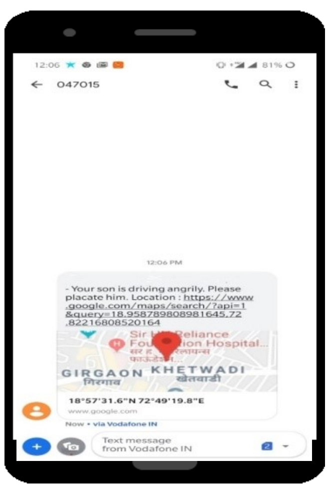

Fig.4. Screenshot of safety message being sent to the guardian

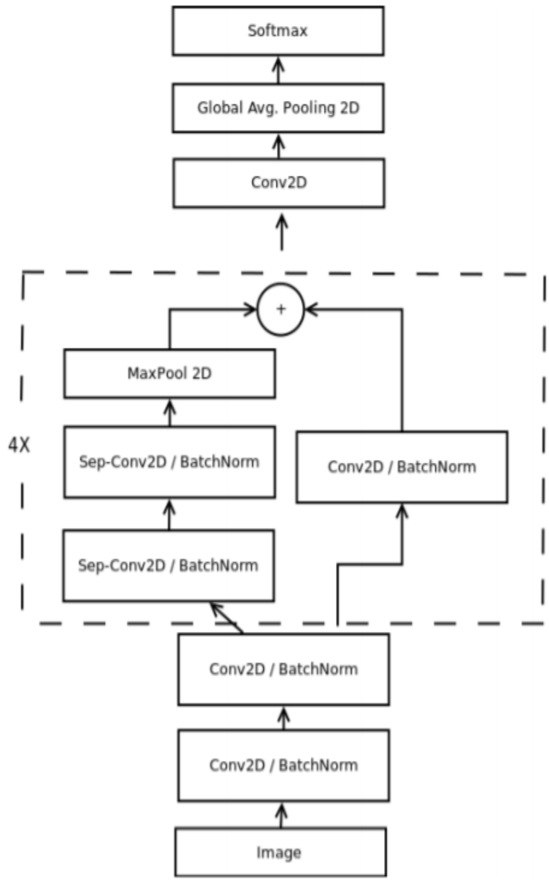

Fig.5. Model used for Real time classification of emotions [13]

The model clings on a scheme to stamp out any fully connected layers. Elimination of fully connected layers is done by global average pooling. It is a standard fully-convolutional neural network that consists of 9 convolution layers, ReLUs[11], Batch Normalization[12] and Global Average Pooling [13]. Adam optimizer was used to train the model which helps us to classify in real time which emotion does the face depict.

A number of accidents can be reduced if road rage is detected soon and necessary steps are taken. The system takes input from the camera and converts it into frames, similar to Drowsiness Detection and makes a prediction on the emotion of the driver. If the system predicts angry for more than 20 frames, a SMS notification is send to the driver`s guardian. The guardian can call up the driver to placate him and convince him not to drive in such conditions. Also, an audio notification is made to the driver that prompts to him that an SMS has been sent. The guardian can use the WebRTC feature provided in the system to directly connect the driver via a live feed.

C. *Lane Detection*

The model detects the current lane of the vehicle and alerts the driver in case of any deviations. Rash driving can also be detected if too many lane changes occur within a particular time period. The system uses sobel and gradient filters for detection of the lines. Radius of curvature value is used to decide the threshold. Sliding window is used to detect lane changes.

The model uses three-colour scheme to detect the rash driving cases: Green- Normal driving, Blue- Threshold crossed once, Red- If the threshold is crossed continuously for the next 50 frames. If the color turns Red, the driver will be notified. The reason for introducing the blue level is because a driver may normally want to change the lane and an unnecessary alarm would be annoying so keeping this intermediate level is necessary. If the driver changes lane for more than 5 times in a period of 50 frames, the driver is declared to be driving rash.

The task is to detect the lanes on the road. Each frame from the real-time video feed is converted to HLS space, and then it puts double weight on S channel to filter the image which makes the yellow and white lines on the road more visible which helps us find the extremes of road. Then it makes a quadrilateral using these endpoints of the road. It is then converted into a warped binary image. Finally, it takes a histogram of the lower columns of the image and the two most prominent peak will give us the x-coordinate of the lane lines. After that, a sliding window protocol can be used to find and follow the lane lines [7]. A 2nd order polynomial fit to both left and right lane lines is calculated.

Using the lane lines and position of the center of the car to it, it is now possible to calculate the radius of curvature of the lane and position of the vehicle to the lane.

$$Radius\ of\ Curvature = \frac{\left[1+\left(\frac{dy}{dx}\right)^2\right]^{\frac{3}{2}}}{\left|\frac{d^2y}{dx^2}\right|} \qquad (3)$$

Assuming the camera is placed exactly at the centre of the vehicle, the centre of the input image can be assumed to be the centre of the car. To calculate the offset of the vehicle, the centre of the lanes is calculated. Centre of the lane is mean of the bottom x value of the left lane and the bottom x value of the right lane. Then the offset of the vehicle is calculated by subtracting lane's centre value from the vehicle's centre value.

D. Proximity Sensing, Traffic Sign Detection and Audio Alert:

This feature will keep the driver alert about the distance with objects and other cars, thus avoiding crashes. It will give different audio warnings to the driver whenever a vehicle or pedestrian is detected in close vicinity. The audio output is available in different languages and the driver has a choice to select a language as per his preference.

The proximity of the object from the camera is calculated by the triangle similarity. The triangle similarity is at first used to calculate the focal length of the camera used in the prototype. Then the calculated focal length is used in real-time to calculate the proximity of objects such as cars and humans.

To calculate the focal length of our camera, let us consider that have an object with a known width W. Then,

Fig.6. Output of Lane Detection

place the object D distance away from the camera and measure the apparent width P of the object in pixels. Hence, it is possible to calculate the focal length of the camera F by the triangle similarity as,

$$F = (P \times D) / W \qquad (4)$$

Now, using this derived focal length one can estimate the distance D' of an object from the camera as,

$$D' = (F \times W) / P \qquad (5)$$

The average width of any car is 6 ft. and at any given point the apparent width P is known. Thus the distance D' of a car can be estimated by substituting these values in the above formula which will give us the proximity of the vehicle. When the distance of our car from the object exceeds the safe limit, the prototype plays an alarm to alert the driver about the obstacle in front of him.

Traffic signs maybe sometimes ignored when the driver isn't paying attention. This feature gives an audio warning to the driver whenever any important traffic sign is detected and thus not only increasing the concentration on the road but also increasing road safety. Currently, the model is trained only to stop sign. About 180 images of stop sign were fed to a Single Shot Multibox Detector (SSD) [15] along with MobileNet [16] as the base network to detect the traffic signboards on roads

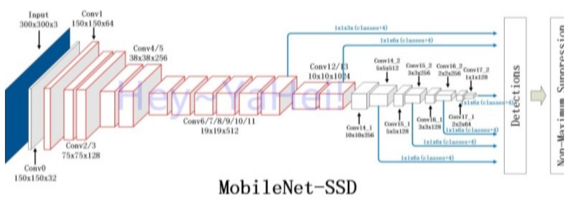

Fig.7. Architecture of SSD used with MobileNet [17]

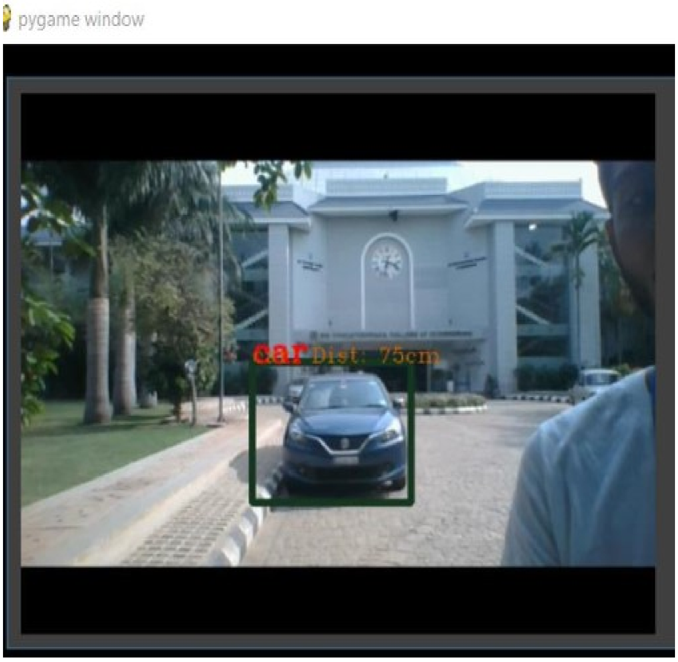

Fig.8. Output of Proximity Detection for a Car

Fig.9. Output when the system detects a person

In both cases, the real-time output is in the form of audio to alert the driver as "Vehicle Ahead" and "Person ahead" respectively.

E. *Reminder for maintenance:*

While considering the safety on the road, maintenance of the car might be ignored by most. But in reality, it plays a major role in the causes of various car accidents. Roughly 2% of vehicle crashes occur due to improper maintenance that leads to the vehicular problem [18]. Thus it is equally important to alert the car owner whenever a vehicle is due for its maintenance.

For this purpose, a database is created which consists of driver's login credentials, basic information about vehicle and driver, Emergency Contact, Scheduled maintenance date. As per the scheduled date, an alert is sent to the driver to remind about the maintenance. Once the vehicle maintenance is done, the database automatically updates the date for the next cycle. Difference between two maintenance cycles is ideally considered to be 35 days.

F. *Alcohol Detection*

Drink and Drive are illegal and prohibited for the safety of the driver as well as his surroundings. So not only do need a system which tells the driver he is drunk and cannot drive but also prevents him from doing so. The best way to do is if he is disallowed to start the car. This paper asserts a basic hardware and software system to achieve the above goal.

The system uses the MQ3 sensor and STM32 microcontroller to detect alcohol vapors in the driver's breath. An external switch is connected to control the motion of the servo motor. Servo motor is used to represent the engine system in our prototype. Every time, a user will try to switch on the circuit, the sensor will detect the alcohol amount and if it is below the permissible level it allows the servo motor to start. When alcohol amount greater than the legal limit is detected, the car won't start and the driver will receive an LED indication.

IV. RESULTS

The prototype accurately detected various cars, a person as well as a stop sign. Fig. 8 shows the detection of a car. Using the focal length derived earlier, the prototype calculated approximate distance from a car or person. Fig 9 shows the detection of a person. When the distance from the person or car reduces to lower than 30cms from the vehicle, the system plays an alarm to alert the driver. Similarly, the

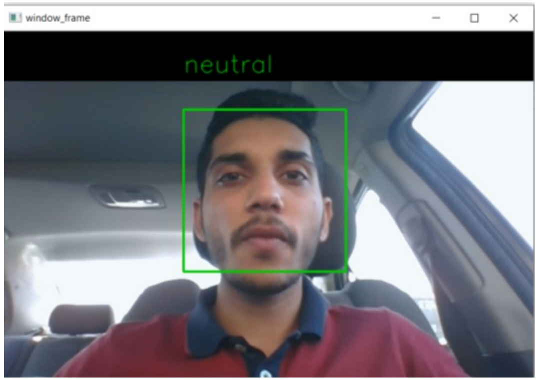

Fig.10. Output of Emotion Analysis

driver was notified of frequent lane changes and multiple lane changes were classified as rash driving. The system prevents the engine from starting if the alcohol sensor detects the alcohol level to be higher than the permitted levels. Emotion analysis was performed on the driver which classified seven emotions. If the driver is detected Angry for more than 5 seconds, a Text SMS is sent to the guardian's entry from the database along with a link that provides the

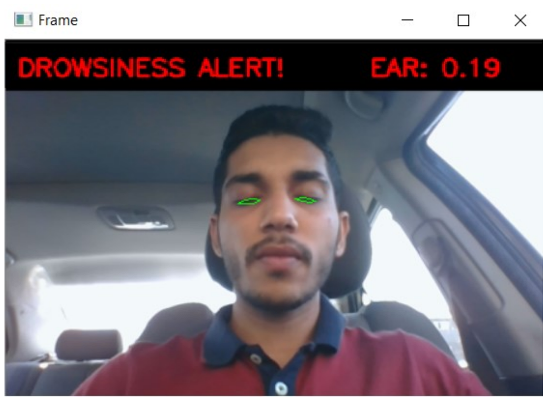

Fig. 11. Output of Driver Drowsiness Detection

precise location of the driver. Fig 10 shows the output of real-time Emotion Analysis performed by the system. The system also performs real-time drowsiness distraction which plays alarm if EAR drops below EAR_{th}. Fig 11. shows the system performing Drowsiness Detection on the driver.

V.CONCLUSION

The prototype generated is capable of being used on the roads and also helps in minimizing accidents on road due to non- alertness at the wheel. The aim of this prototype is not only to make the driver alert when in need but also to provide a better driving experience and ensure road safety in the best way possible. Various functions of the prototype have been successfully implemented and tested. It provides a non-distractive approach to notify the driver whenever necessary as well as notify the guardian whenever things might not be in control of the driver himself. Real-time analysis is performed on both the driver as well as the road to detect any important aspect that cannot be missed. This prototype thus proves to be useful whenever it is the driver's fault, that is, when the driver is distracted, angry, misses a sign or is drunk, as well as it proves to be helpful to guide the driver when it is not his fault, the cases wherein a car or person is close or lane crossing.

VI. FUTURE SCOPE

The prototype can be further improved in various ways to further increase the safety on the road and improve attendance at the wheel. Currently, the prototype is trained to detect lanes and alert the driver if the vehicle does not obey lane for a longer period or if it crosses the lanes many times in a few seconds. This can be further improved by guiding the driver about the road ahead by detecting the curvature of the road ahead. Accidents due to sharp turns can be successfully minimized by implementing this.

The prototype could also be trained in future to detect various important road signs and with the help of sensors, one can calculate the speed of the car. With the combination of detecting road signs such as speed limits to store the speed limit of a particular road and constantly calculating the speed of the car in real-time, the prototype can also be used to play an alarm if the speed of the car exceeds the specified speed limit for the road. With the help of speed and the distance of our car from a person or another car, use various algorithms to deduce if it is a safe distance and if the car will cause no harm given full brakes are applied. Furthermore, it can also be used to decide the faulty party in a road accident. Artificial Intelligence technologies can be used to build a model which could take the video of the cameras used in our prototype as an input, record and generate patterns in the tentative behavior of the driver via emotion analysis and lane detection, to further perform decision making to conclude who could be held responsible for the accident.

VII. REFERENCES

[1] V. Unadkat, P. Sayani, P. Kanani and P. Doshi, "Deep Learning for Financial Prediction," 2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET), Kottayam, India, 2018, pp. 1-6.

[2] P. Kanani and M. Padole, "Deep Learning to Detect Skin Cancer using Google Colab," International Journal of Engineering and Advanced Technology Regular Issue, vol. 8, no. 6, pp. 2176–2183, 2019.

[3] P. Kanani and M. Padole, "Recognizing Real Time ECG Anomalies Using Arduino, AD8232 and Java," Communications in Computer and Information Science Advances in Computing and Data Sciences, pp. 54–64, 2018.

[4] A. B. Hillel, R. Lerner, D. Levi, and G. Raz, "Recent progress in road and lane detection: a survey," Machine Vision and Applications, vol. 25, no. 3, pp. 727–745, Jul. 2012.

[5] X. Glorot, A. Bordes, and Y. Bengio. Deep sparse rectifier neural networks. In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, pages 315–323, 2011.

[6] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. International Conference on Machine Learning, pages 448–456, 2015.

[7] T. Soukupova and J. Cech, "Real-Time Eye Blink Detection using Facial Landmarks," Computer Vision Winter Workshop, Feb. 2016.

[8] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR05).

[9] O. Arriaga , M. Valdenegro-Toro, and P. Plöger, "Real-time Convolutional Neural Networks for Emotion and Gender Classification," 2017.

[10] M.Bourne article on www.intmath.com/applications-differentiation/8-radius-curvature.php

[11] Rules of the Road Regulations -1989 http://transport.bih.nic.in/Acts/The-Road-Regulations-Rules-1989.pdf

[12] A. M. Williamson, "Moderate sleep deprivation produces impairments in cognitive and motor performance equivalent to legally prescribed levels of alcohol intoxication," Occupational and Environmental Medicine, vol. 57, no. 10, pp. 649–655, Jan. 2000.

[13] Kaggle fer2013 dataset of emotions https://www.kaggle.com/deadskull7/fer2013#fer2013.csv

[14] D. E. King, "Dlib-ml: A Machine Learning Toolkit," Journal of Machine Learning Research, 2009.

[15] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single Shot MultiBox Detector," Computer Vision – ECCV 2016 Lecture Notes in Computer Science, pp. 21–37, 2016.

[16] A. Howard, M. Zhu , B. Chen, D. Kalenichenko , W. Wang, T. Weyand, M. Andreetto , and H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," 2017.

[17] W. L. Middelkoop, "Realtime human pose recognition through computer vision - Using TensorFlow and PoseNet on a video feed," Willem's Blog, 01-Dec-2019. [Online]. Available: https://willem.com/blog/2019-12-01_realtime-human-pose-recognition-through-computer-vision/. Accessed on 3 March 2020.

[18] "India Environment Portal Knowledge for change," Road accidents in India - 2017 - India Environment Portal | News, reports, documents, blogs, data, analysis on environment & development | India, South Asia, 01-Oct-2018. [Online]. Available: http://www.indiaenvironmentportal.org.in/content/459084/road-accidents-in-india-2017/.

[19] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in Computer Vision and Pattern Recognition, IEEE Conference on, 2014, pp. 1867– 1874

[20] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, M. Pantic. 300 Faces in-the-Wild Challenge: The first facial landmark localization Challenge. Proceedings of IEEE Int'l Conf. on Computer Vision (ICCV-W), 300 Faces in-the-Wild Challenge (300-W). Sydney, Australia, December 2013.

[21] C. Sagonas, E. Antonakos, G, Tzimiropoulos, S. Zafeiriou, M. Pantic. 300 faces In-the-wild challenge: Database and results. Image and Vision Computing (IMAVIS), Special Issue on Facial Landmark Localisation "In-The-Wild". 2016.

[22] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, M. Pantic. A semi-automatic methodology for facial landmark annotation. Proceedings of IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR-W), 5th Workshop on Analysis and Modeling of Faces and Gestures (AMFG 2013). Oregon, USA, June 2013.