2023 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE) | 979-8-3503-3826-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCIKE58312.2023.10131730

AI-Based Driver Drowsiness and Distraction Detection in Real-Time

Anna Titu Kurian
School of Enginnering and Physical Sciences
Heriot – Watt University
Dubai, United Arab Emirates
Email: atk2000@hw.ac.uk

Prashant Kumar Soori
School of Engineering and Physical Sciences
Heriot – Watt University
Dubai, United Arab Emirates
Email: p.k.soori@hw.ac.uk

Abstract— **This paper proposes a solution to combat risks associated with road accidents namely drowsiness and distractions which have been established to be the prominent causes of accidents worldwide. The proposed methodology integrates camera vision and mathematical computations to accurately detect driver drowsiness and distracted driving. The eye aspect ratio and mouth aspect ratio are utilized to recognize drowsiness characteristics while the eye tracking methodology is adopted to identify distracted behavioral factors. On the detection of the mentioned risk factors, alerts are provided to the driver in visual and audio formats by use of the Raspberry Pi microprocessor, LCD display and buzzer. The developed system was tested under an experimental setup and exposed to various lighting conditions. The results suggested that the approach is capable of recognizing drowsiness and distractions with an accuracy of 94.1% and 89% respectively during both day and night conditions and provide warnings as required.**

Keywords—camera vision; distraction detection; driver drowsiness; eye aspect ratio; eye tracking; mouth aspect ratio

I. INTRODUCTION

Crash injuries has been reported to be the eighth leading cause of fatalities worldwide and the primary cause of fatalities in children and young adults. Approximately 3700 deaths occur on a daily basis rooting from road accidents [1]. In addition, low-to-average income regions are more prone to road crashes due to inability to provide safe roads and infrastructures [2]. The prominent human-related factors of accidents comprise of drowsiness/fatigue and distractions.

As per surveys conducted, 60% of the general population experience drowsiness while driving at least once in a month. Fatigue driving affects basic cognitive functionality including response speeds, decision-making capabilities, and memory. Occasionally, the driver can also slip into a temporary state of unconsciousness hence posing an imminent threat to passengers within the vehicle and neighbouring vehicles. It has been estimated that about 100,000 road crashes worldwide stem from driver drowsiness and moreover, it has been reported that statistics, in reality, could be higher than this value due to drowsiness being a highly subjective reasoning [3]. The critical symptoms of drowsiness include drooping eyes and yawning. Driver fatigue is also dependent on factors such as time of day, health factors and lack of sleep among others [4].

On the other hand, distracted driving accounts for any action that hinders the driver's attention and focus from the road and the driving activity. During the year 2020, distracted driving attributes accounted for 8% of fatal accidents, 14% of injury crashes and 13% of all motor vehicle road accidents [5]. The term distracted driving is very broad and can be primarily classified as:

- Auditory – distractions rooting from audio.
- Visual – distractions that leads to the driver gazing away from the road.
- Cognitive – distractions caused by lack of focus on driving.
- Manual - distractions that results in the driver's hands not on the steering wheel.

Visual distractions are one of the most predominant causes of accidents and could be a result of one's cell phone, conversing with a passenger or a sign on the street to mention a few. Essentially, the actions that fall under visual distractions, regardless of the cause, results in the driver's gaze direction deviating from the centre.

In order to combat the rising global road accident crisis, the United Nations have set one of the 16 Global Goals to halve road injuries and deaths (Target 3.6) [6]. Although road safety is deemed of high priority and road accident statistics are at an exponential rise in the recent decades, limited practical implementations are available to approach the concerns in this sector. However, extensive research has been conducted in this sector and various proposals to monitor driver behavioural characteristics has come to light. The proposed methodologies for drowsiness detection can be primarily divided into two sections as follows:

- Internet of Things (IoT) based: methods utilize sensors such as infrared sensor for eye bink rate measurement, tilt sensors to measure head tilt and heart rate measurements to identify changes in biometrics [7, 8].
- AI based: procedures make use of camera vision to monitor driver's actions. Monitoring can be carried out by object detection method utilizing neural networks like Convolutional Neural Network (CNN) or Support Vector Machine (SVM). Face detection and eye and mouth feature detection is carried out prior to classifying the state of the eyes or mouth as 'closed' or 'open' [9, 10].

Disadvantages with respect to the IoT based procedures is its intrusiveness as the sensors must be attached directly to the user. In the case of an infrared sensor, often the driver is required to wear specialized goggles that has the infrared sensor fitted on. However, this can cause a blind spot in the driver's vision leading to greater dangers while on the road. On the other hand, predictions made by machine learning models are accurate with considerable training however has its own downsides. The classification states are limited to either 'closed' or 'open' suggesting that detection is only made once the eyes are completely closed therefore warning might be delayed in practical settings. Moreover, a paper has reported errors regarding the inability to detect accurately when openness of mouth is relatively larger [9].

The term distracted driving accounts for a vast range of driver actions however existing methodologies in this field is considerably limited to the detection of certain actions such as phone usage by a trained machine learning model or determining whether hands are positioned on the steering wheel by use of Bluetooth enabled devices available in the vehicle [11,12].

The novelty of this paper proposes a vision-based solution integrated with mathematical computations to identify sleepiness, yawning and distracted driving factors in real time. The openness of one's eyes and mouth are analysed by computing ratios to identify sleepiness and yawning factors. The ratios are capable of tracking the proportion by which the eyes are open or closed hence early warnings can be generated when the driver begins to depict drowsiness factors. On the contrary, the driver's gaze direction is determined by tracking the iris in order to verify if the driver's attention is on the road. Hence, the eye tracking methodology identifies a broader spectrum of distractions. Early warnings and alerts are provided to the driver via the LCD display and buzzer controlled by the Raspberry Pi microprocessor. Thus, in its entirety, this proposal requires lower computational burden relative to neural networks.

II. METHODOLGY

The proposal aims to detect drowsiness as well as possible distractions while driving in real time to provide early alerts and prevent the occurrence of a possible accident.

This is carried out by processing the visual input from a camera placed on the vehicle dashboard with a machine learning model. Fig.1 depicts the system flowchart. The machine learning model is programmed to firstly carry out face detection followed by the identification of locations of all facial features. The Haar cascade face detection model provided by OpenCV, and Dlib's 68-point Face Landmark predictor is utilized respectively for face detection and facial feature identification.

The landmarks from Dlib's face landmark predictor is utilized to obtain the eye location for a certain user. In order to recognize drowsiness, the eye aspect ratio and mouth aspect ratio is computed. The values are compared to a pre-set threshold to verify if drowsiness characteristics are present. On the contrary, distracted driving attributes are detected via tracking the driver's iris to determine the gaze direction. The direction is analysed to ensure if the driver's vision is focused on the road.

Fig. 2 displays the components and the relationship between them within the system. In the case that either of the mentioned risk factors are detected, the Raspberry Pi-4 microprocessor transmits a signal to the Liquid Crystal Display (LCD) and active buzzer to provide warnings.

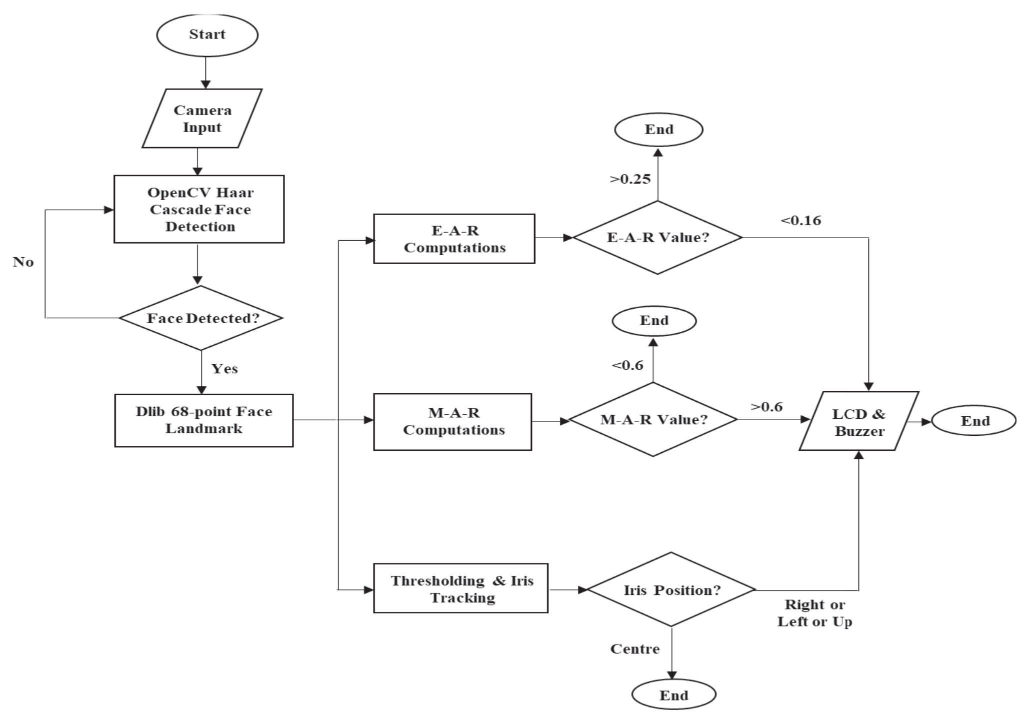


Fig.1 System Flowchart

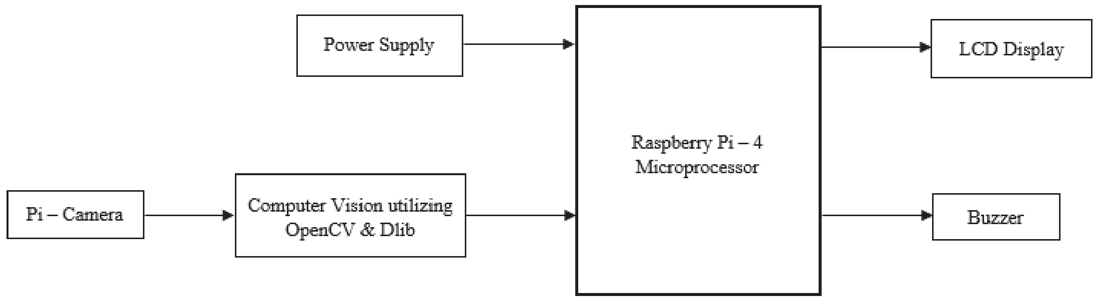


Fig. 2 System Block Diagram

A. *Haar Cascade Face Detection*

The Haar Cascade Face Detection model utilizes haar features obtained from training models with both positive (images that contains a face) and negative (images that do not contain a face) data. The haar feature, shown in Fig. 3, essentially consists of only black or white pixels and hence is a binary image. Each feature represents a section on the face and multiple such Haar features are developed. To verify if a face is present, a detection window slides through the entire image and compares each section with the obtained Haar features. If the Haar features are present in a detection window, it is ruled out as the presence of a face [13].

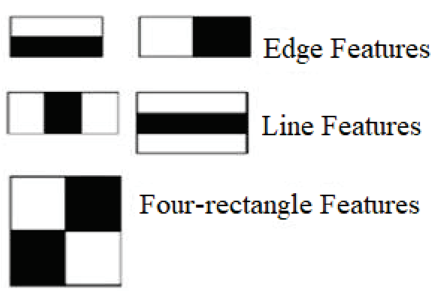


Fig. 3 Haar Features.

Within this project, the pre-trained model provided by OpenCV for face detection is adopted. The provided XML files are imported and included in the program code utilizing the 'CascadeClassifier' function part of the OpenCV library. The input to the machine learning model is the live feed from the camera. Hence, frames are captured from the camera at certain time intervals. The frames are converted to grayscale and resized to increase processing speed of the program. The face detection model is then applied to each frame and in the case that a face is detected, the output comprises of four coordinates (x, y, w, h). Table I depicts the representation of each of the output coordinates. Input parameters such as the scale factor, minimum neighbours, minimum size, and flags are configured to increase both performance speed and accuracy within the detection model.

TABLE I. OUTPUT COORDINATES OF FACE DETECTION MODEL

Coordinates	Definition
x	x-coordinate of a rectangular edge
y	y-coordinate of a rectangular edge
w	horizontal length of rectangle
h	vertical length of rectangle

The coordinates ultimately form a rectangle around all detected faces. For the project requirements, only the driver face is to be detected therefore the CV_HAAR_FIND_BIGGEST_OBJECT flag is included which allows only the largest face in a certain frame to be detected.

B. *Dlib's 68-point Face Landmark*

The pre-trained 68 – point Facial Landmark predictor provided by the Dlib library locates all facial features by mapping each of the landmarks at a specific point on one's face. The numbering is uniform and unique to each landmark as shown in Fig. 4.

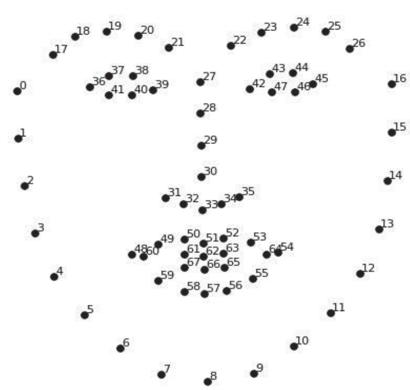


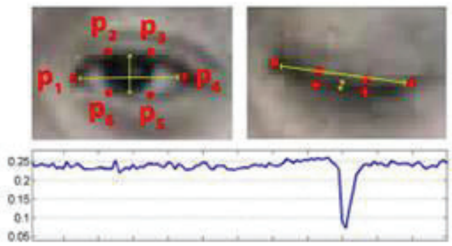
Fig. 4 Dlib's 68-Point Facial Landmark.

Once a face is detected and face location coordinates within the frame are obtained, the coordinates are input into the landmark predictor which in turn provides the x-y coordinates of each landmark. The resulting landmarks are then converted into an array to point to specific landmarks for further computations. The landmarks required for the realisation of the proposed procedure are the following:

- 36 – 41: Left eye
- 42 – 47: Right eye
- 60 – 67: Inner lining of mouth.

C. Eye Aspect Ratio (EAR)

The eye aspect ratio (EAR) classifies the state of an individual's eye as open or closed. Equation (1) represents the eye aspect ratio formula [14]. The definitions of the parameters introduced in (1) is displayed in Fig. 5 where parameters p_1 to p_6 corresponds to the respective eye landmarks obtained from Dlib's 68-point facial landmark



predictor.

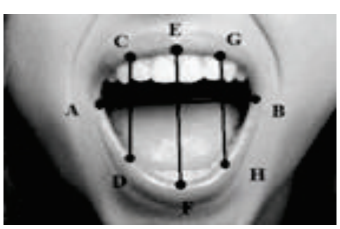
Fig. 5 Eye Aspect Ratio Computation Variables

$$\text{EAR} = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\times\|p_1 - p_4\|} \qquad (1)$$

This ratio is essentially the relation of the average vertical length to the maximum horizontal length. The plot in Fig. 5 displays the variations in the eye aspect ratio during a single blink. It is observed that the ratio decreases at a steep gradient when the eyes are closed. Hence, for EAR values lesser than 0.16, it can be concluded that the driver presents drowsiness characteristics.

D. Mouth Aspect Ratio (MAR)

The mouth aspect ratio (MAR) is a similar concept to the eye aspect ratio. It provides an insight into the openness of an individual's mouth and hence can detect



yawning features.

Fig. 6 Mouth Aspect Ratio Computation Variables.

$$\text{MAR} = \frac{|EF|}{|AB|} \qquad (2)$$

Equation (2) is the ratio of the maximum vertical length to the maximum horizontal length (represented by EF and AB in Fig. 6). A higher MAR value corresponds to more openness of the mouth. The pre-set threshold magnitude has been fine-tuned and set to 0.6 where for MAR values greater than 0.6, it is identified as the driver presents yawning characteristics.

E. Eye Gaze Tracking

Eye Tracking is a methodology implemented to track the iris and determine the gaze direction. The landmarks provided by Dlib's library is utilized to distinguish the eye from the rest of the face. To extract the eye image, the 'fillPoly' and 'bitwise AND' functions provided by the OpenCV library is implemented where the fillPoly function fills the eye region with white pixels to be utilized as a mask while the bitwise AND operation compares the grayscale frame to the above mask and extracts the eye regions.

Thresholding is then performed where each pixel on the eye image is assigned either a HIGH (black) or LOW (white) value based on a set threshold and a binary image consisting of only white or black pixels is obtained. The maximum and minimum points of the extracted eyes are then computed, and the maximum vertical and horizontal lengths of the eye are obtained. The maximum lengths are generally positioned towards the mid-points within the eye. The left and right eye regions are then divided individually into three equal parts both vertically and horizontally. In order to determine gaze direction, the accumulation of black pixels within each divided region is evaluated and compared to verify which region provides the highest accumulation. This region represents the gaze direction.

However, an exception exists in the case that the driver's gaze is directed downwards. Due to the positioning of the camera, the iris might not be visible and hence accuracy of the gaze tracking procedure decreases. This issue is tackled by integrating the eye aspect ratio concept where if an eye seems to be closed to a certain extent, then it is considered as a downward gaze.

F. Alert System

The alert system implemented comprises of both visual and audio warnings. The auditory warnings are implemented via a buzzer while the visual warnings are implemented through an LCD display. For each risk factor namely drowsiness or distractions, the warnings are only provided when a certain number of consecutive frames is detected with a risk factor. The auditory alert effectively grabs the driver's attention while the visual display provides clarification on the reasoning behind the warning.

III. RESULTS & DISCUSSION

The developed system was tested for accuracy measurements. The experiment was set up under various practical conditions and results were recorded and analysed.

A. Experimental Setup

The experimental setup aims to take into consideration various dependencies that affect the detection accuracy namely the positioning of camera, ambient conditions, and tilt of an individual's face.

The camera is positioned to scan for reckless driving factors. Hence, it is essential the camera is placed to obtain a clear view of the driver's portrait. For the purpose of the experiment, the camera is located at an average level where it is generally perpendicular to the driver's face. The experiment is repeated with 4 participants to analyse adaptability. In addition, the system is also tested in 'day' and 'night' conditions to verify for results with limited lighting conditions.

B. Accuracy Results

The developed machine learning based algorithm was tested and sample results are as displayed in Fig. 7, Fig. 8 and Fig. 9. The algorithm accurately detects and provides a pop-up thus allowing for the verification on accident risk factor detection.

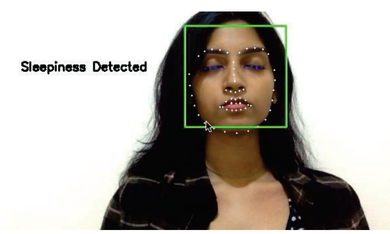


Fig. 7 Testing of Sleepiness Detection Algorithm.

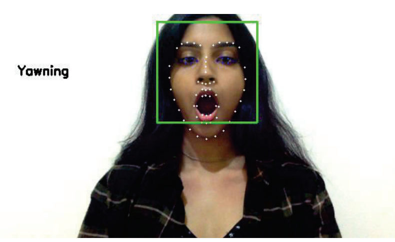


Fig. 8 Testing of Yawn Detection Algorithm.

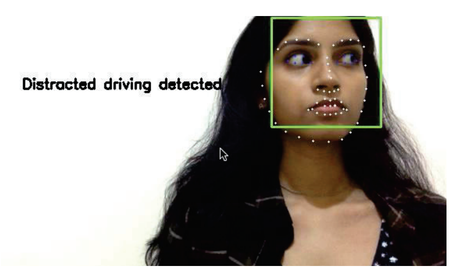


Fig. 9 Testing of Distraction Detection Algorithm

The displayed figures also depict the workings of the Haar Cascade face detection model as well as Dlib's 68-point face landmark. The results for each of the testing completed can be found in Table II for sleepiness, Table III for yawning and lastly Table IV for distracted driving.

TABLE II. E-A-R BASED SLEEPINESS DETECTION RESULTS

Sample No.	*Participant*	*E-A-R*	*Drowsiness Detected?*	*Real Drowsy?*
1	A	0.116	Yes	Yes
2	A	0.307	No`	No
3	B	0.119	Yes	Yes
4	B	0.298	No	No
5	C	0.131	Yes	Yes
6	C	0.312	No	No
7	D	0.176	No	Yes
8	D	0.322	No	No

TABLE III. M-A-R BASED YAWN DETECTION RESULTS

Sample No.	*Participant*	*M-A-R*	*Yawn Detected?*	*Real Yawn?*
1	A	1.036	Yes	Yes
2	A	0.1	No	No
3	B	0.965	Yes	Yes
4	B	0.027	No	No
5	C	0.595	No	Yes
6	C	0.0	No	No
7	D	1.071	Yes	Yes
8	D	0.0263	No	No

TABLE IV. EYE TRACKING BASED DISTRACTION DETECTION RESULTS

Sample No.	*Participant*	*Distraction Detected?*	*Real Distraction?*
1	A	Yes	Yes
2	A	Yes	Yes
3	B	Yes	Yes
4	B	No	No
5	C	Yes	No

Sample No.	Participant	Distraction Detected?	Real Distraction?
6	C	No	Yes
7	D	Yes	No
8	D	No	No

It can be concluded that the proposed methodology provides optimal accuracy and reliability. However, the system does exhibit a limitation in the case that a face is not clearly visible in the camera frames. If one's head is tilted to the point where facial features are not apparent to the camera, the detection accuracy decreases considerably.

IV. Conclusion

This paper proposed an automated methodology utilizing the Raspberry-Pi microcontroller and camera vision to efficiently identify drowsiness and distracted behaviour and effectively alert the personnel. The detection methodology comprises of face detection as well as the usage of Dlib's 68-point facial landmark predictor. Drowsiness detection was carried out by computing the eye aspect ratio and mouth aspect ratio while distracted driving detection was executed utilizing the eye tracking method. The system was designed to also provide pre-set warnings via visual and audio formats as required. Testing was conducted on the proposed system under an experimental setup to mimic practical working conditions and was also subject to various lighting conditions. Results suggested that the drowsiness and yawning detection algorithm provided an accuracy of 94.1% while the distracted driving recognition algorithm provided an accuracy of 89%. It was also concluded that lighting conditions in the surroundings as well as the distance between the driver and camera does not affect code accuracy ratings.

References

[1] "Road Traffic Injuries and Deaths—A Global Problem." CDC. https://www.cdc.gov/injury/features/global-road-safety/index.html (accessed Oct. 22, 2022).

[2] "Road traffic injuries." World Health Organisation. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries (accessed Oct. 22, 2022).

[3] R. Levi. "The dangers of Drowsy Driving." The Sleep Doctor. https://thesleepdoctor.com/drowsy-driving/ (accessed Oct. 30, 2022).

[4] "Driver fatigue." CDC. https://www.cdc.gov/niosh/newsroom/feature/driverfatigue.html (accessed Oct. 30, 2022).

[5] "Distracted driving 2020," National Center for Statistics and Analysis, New Jersey, Washington DC, USA, DOT HS 813 309, May, 2022. Accessed Date: Oct. 30, 2022. [Online]. Available: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813309

[6] "Do you know all 17 SDGs?" United Nations. https://sdgs.un.org/goals (accessed October 25, 2022).

[7] A. A. Jecko, R. Das, M. A. Ashraf, M. M. Rashed and S. Khan, "Design and Implementation of a Microcontroller Based Driver Impairment Monitoring System," *2018 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 2018, pp. 489-493, doi: 10.1109/ICSSIT.2018.8748643.

[8] K. -W. Ke, M. R. Zulman, H. -T. Wu, Y. -F. Huang and J. Thiagarajan, "Drowsiness Detection System Using Heartbeat Rate in Android-Based Handheld Devices," *2016 First International Conference on Multimedia and Image Processing (ICMIP)*, 2016, pp. 99-103, doi: 10.1109/ICMIP.2016.23.

[9] W. Tipprasert, T. Charoenpong, C. Chianrabutra and C. Sukjamsri, "A Method of Driver's Eyes Closure and Yawning Detection for Drowsiness Analysis by Infrared Camera," *2019 First International Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics (ICA-SYMP)*, Bangkok, Thailand, 2019, pp. 61-64, doi: 10.1109/ICA-SYMP.2019.8646001.

[10] S. S. I, R. Ramli, M. A. Azri, M. Aliff and Z. Mohammad, "Raspberry Pi Based Driver Drowsiness Detection System Using Convolutional Neural Network (CNN)," *2022 IEEE 18th International Colloquium on Signal Processing & Applications (CSPA)*, Selangor, Malaysia, 2022, pp. 30-34, doi: 10.1109/CSPA55076.2022.9781879.

[11] S. Carrell and A. Atapour-Abarghouei, "Identification of Driver Phone Usage Violations via State-of-the-Art Object Detection with Tracking," *2021 IEEE International Conference on Big Data (Big Data)*, Orlando, FL, USA, 2021, pp. 1236-1241, doi: 10.1109/BigData52589.2021.9671378.

[12] T. Mewborne, Y. Lee, S. Tan and J. Yang, "Distracted Driving Detection Utilizing Wearable-based Bluetooth," *2022 IEEE 19th International Conference on Mobile Ad Hoc and Smart Systems (MASS)*, Denver, CO, USA, 2022, pp. 485-490, doi: 10.1109/MASS56207.2022.00074.

[13] "Face Detection using Haar Cascades." OpenCV. https://docs.opencv.org/3.4/d2/d99/tutorial_js_face_detection.html (accessed Oct 30, 2022).

[14] J. Cech and T. Soukupova, "Real-time eye blink detection using facial landmarks," *Cent. Mach. Perception, Dep. Cybern. Fac. Electr. Eng. Czech Tech. Univ. Prague*, 2016, pp. 3:1-8.