Driver's Fatigue Recognition Using Convolutional Neural Network Approach

Samer Abdullah Deeb Abbas[1], Sew Sun Tiang[1, *], Wei Hong Lim[1], Li Sze Chow[1], Chin Hong Wong[2,3]
([1]UCSI University, Malaysia) ([2]Fuzou University, China) ([3]Maynooth University, Ireland)
E-mail: 1001851168@ucsiuniversity.edu.my, tiangss@ucsiuniversity.edu.my, limwh@ucsiuniversity.edu.my, chowls@ucsiuniversity.edu.my, chinhong.wong@mu.ie
http://www.ucsiuniversity.edu.my/

Abstract

Drowsy driving is a serious issue that has been leaking in our communities since long time, the definition of drowsy driving is when the driver is not aware enough to proceed with driving the vehicle causing catastrophic accidents. Multiple methods were found to approach this complication across the years. Convolution Neural Network has approved to be a reliable approach to treat this issue by using face feature detection. In this paper, the effect of key parameters of the trained framework based on the driver's fatigue recognition model are analyzed, and the accuracy of the driver's fatigue recognition model is investigated, as well as a driver's fatigue recognition is studied under different conditions using CNN. Transfer learning is used to develop a reliable method for detection, Mediapip Face Mesh model is used to extract the features from the face. MAR (Mouth Aspect Ratio) as well as EAR (Eyes Aspect Ratio) are obtained through the detection, these terms are responsible for detecting the eye and mouth closure ratio, the model has proved to work with accuracy of 98.3% and in different light conditions with accuracy of 94.7% outperforming several past models.

Keywords: Drowsy detection, Eye detection, Mouth Detection, Eye Aspect Ratio, Mouth Aspect Ratio.

1. Introduction

World Health Organization stated that, road accidents cost the lives of millions of individuals every year [1]. Statistics have shown that most deadly accidents are caused by driver drowsiness and carelessness. According to the American Automobile Association [2], drowsy drivers are responsible for 7% of all accidents and 21% of fatal traffic accidents. In another research done by the Foundation for Traffic Safety in 2017, 42.4% of drivers travel without getting at least one day of sleep or less than six hours of sleep in a normal week [3]. For the majority of individuals (87.9%), these difficulties are serious, and they perceive what they observe as improper behavior (95.2%). However, nearly one-third of ten drivers (30.8%) admit to driving when too fatigued to keep their eyes open in the previous months. As stated by the American Automobile Association's Road Safety Foundation, driver weariness is to blame for 16–21% of traffic accidents. The risk of a traffic collision induced by driver drowsiness is 46 times higher than while driving normally [3].

This paper aims to develop a CNN model that can detect fatigue drivers to decrease the accidents that may occur as a consequence of drowsy driving. In addition will aim to provide a CNN model which can be used in the industry, with few conditions to expand the field of analyzing face detection under different circumstances, looking into the fact that the day goes by, and the night will come, and according to most drowsy driving occur between midnight and 6 am [4]. Lastly key parameters of the trained model will be supervised to enhance the predication output of the model while keeping an eye on the accuracy to ensure high detection efficiency [5]. Studies have showed different approaches to resolve this issue and they can be categorized as shown in Table 1. When a driver is fatigued, the blinking frequency is dramatically increased compared to usual. Additionally, because the driver's head posture changes depending on whether he or she is awake or weary, the frequency with which the driver nods over some time may be utilized as a measure for assessing fatigue [6].

Table 1. Summary of the Past Approaches

Methods	Used by	Limitation
Physiological Measures	Detecting bioelectric signals using sensors mounted on the head [5].	Uncomfortable for the driver. Expensive equipment
Vehicle-Based Measures	Analyzing lane offset, SWA, and vehicle speed, among other vehicle based variables [6].	Affected by External Factors. Different driving styles affects accuracy
Behavioral Measures	Determining the driver's fatigue level based on the fluctuation of the driver's head, eyes, lips, and other characteristics [7].	Can be affected by surrounding environment.

Recently, deep learning systems, particularly those based on Convolutional Neural Networks (CNNs), have gained importance in addressing difficult categorization issues. Most of them are ground-breaking advances in a variety of Computer Vision tasks, including scene segmentation, emotion identification, object detection, and picture classification. The advancement of machine learning technology (particularly CNN) enables increased accuracy and performance. This technique applies to a broad number of applications, including driver awareness testing [8]. PERCLOS, MAR and EAR are presented as an index parameter for measuring exhaustion and developed a matching fatigue detection system, in which the closure of the eyelids and frequent yawing might partially represent the fatigue condition. The Carnegie Mellon Institute has frequently proved through experimentation and demonstration that the physical quantity "PERCLOS" indicates drowsiness. Additionally, pupil features, eye gaze direction, and blink frequency all contribute to the identification of exhaustion[8]. The ability of computer vision models to detect the existence of human faces in digital pictures means that the system can recognize the presence of a human face in an image or video and distinguish it from other things [9].

The system begins by analyzing the input image before identifying the face of the user, locating their eyes and mouth from there, and detecting the movement of their eyelids. In order to distinguish between voluntarily and involuntarily blinking, the eyelids are first detected. Then, the Eye Aspect Ratio is determined. Multiple (x, y) coordinates are used to represent each eye, beginning at the left corner and moving clockwise around the rest of the area [10]. Similar to how the eye aspect ratio is dependent on these variables to determine when the eye is open and closed. the mouth aspect ratio can be determined using the coordinated information from the area around the upper and lower lip. To create the detection system, OpenCV and Mediapipe face mesh libraries were used. The libraries are used to analyses photos and live stream video to recognize specific objects, including faces, hands, and feet.

2. Related Works

2.1. *Methodology*

The system architecture mainly outlines how data is used, how it is changed, and how the results are affected by these changes. The suggested method starts by taking a video frame with the system's camera as shown in Fig. 1.

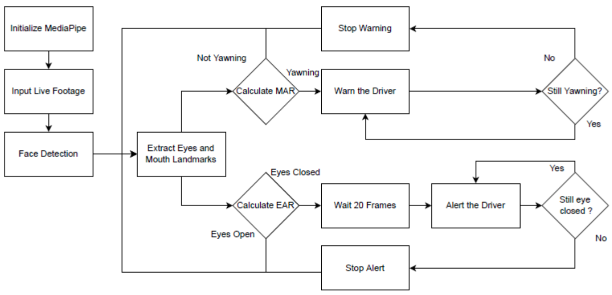

Fig. 1. Flowchart of the Fatigue Detection System.

2.2. *Frame Capturing*

OpenCV (Open Source Computer Vision Library) is a programming function library geared mostly at real-time computer vision. It's helpful in obtaining live footage through the camera, as well as its compatible with important libraries like Dlib and Mediapipe, also it is used for real-time activities like plotting the obtained data into a graph form. OpenCV is very crucial for the performance of the model, it's used to get the input live footage which will be processed to acquire the important terms related to the study like MAR and EAR. OpenCV also provides a way to monitor the MAR and EAR and make it user-friendly, by using a few lines of code EAR and MAR will be plotted into *x* and *y* graphs which makes it easier to determine when the driver hits the threshold and also to monitor the driver's behavior as well.

2.3. *Mediapipe Landmark*

Shape prediction includes identifying landmarks on the face. When an input image is given, the Shape Predictor seeks to locate the localization points around the structure, which is critical for a variety of facial analysis tasks. The Facial Landmark Algorithm is used to regionalize the face. In terms of facial landmarks, our goal is to identify the facial structures on the face using shape prediction methods. Using plot 468-point system[11] as shown in Fig. 2, the Mediapipe Face attributes are identified using the face mesh library [11]. In face landmarks, transfer learning is used to train a network with multiple objectives: the network simultaneously predicts 2D semantic contours on annotated real-world data and 3D landmark coordinates on artificially rendered data. The developed network allowed the models to predict 3D landmarks with high accuracy using both synthetic and real-world data.

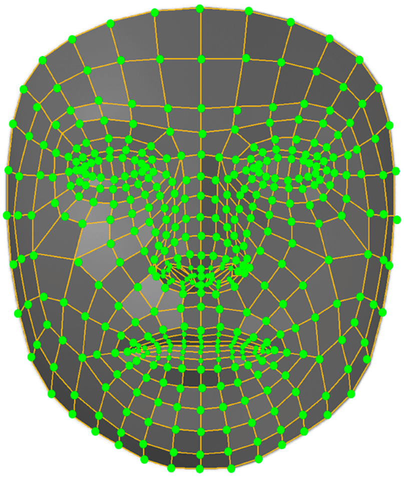

Fig. 2. The 468-point system.

2.4. *Face Detection, Eye and Mouth Key Points*

A facial landmark technique is used to identify a face inside a frame in order to detect faces. Only facial structures are shown, and all other foreign objects are disregarded by this algorithm. To complete this stage, Mediapipe Face Mesh is used to find and reflect notable facial characteristics. The model generates 468-xy points using the Mediapipe Face Mesh facial landmark detector. To label these 468 spots, the shape predictor approach is employed. Using these locations, the facial region can be identified. According to the 468 points surrounding the face, a placement of the points is computed in Fig. 3 as shown.

Fig. 3. Face detected using 468-point system.

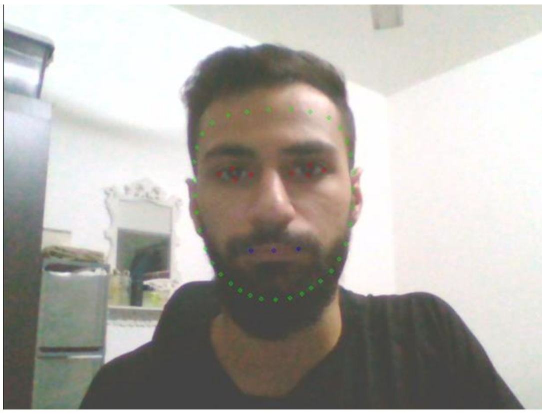

Fig. 4. Eye and Mouth detected.

An important step in this method is localization of the area around the lips and eyes. Eye and mouth areas were observed and used for eye and mouth monitoring and detection from the face which was captured. Based on the 468-point system the indexes for the left eye [362, 382, 381, 380, 374, 373, 390, 249, 263, 466, 388, 387, 386, 385,384, 398] and for the right eye [33, 7, 163, 144, 145, 153, 154, 155, 133, 173, 157, 158, 159, 160, 161, 246], and for lips [61, 146, 91, 181, 84, 17, 314, 405, 321, 375,291, 308, 324, 318, 402, 317, 14, 87, 178, 88, 95, 185, 40, 39, 37,0 ,267 ,269 ,270 ,409, 415, 310, 311, 312, 13, 82, 81, 42, 183, 78]. The convex hull computed for the right eye, left eye, and mouth is displayed in Fig. 4.

2.5. *Eye Aspect Ratio and Mouth Aspect Ratio*

Fig. 5 shows the eye vertical and horizontal coordinates. Once the left eye and right eye coordinates have been extracted, they are used to calculate the Eye aspect ratio based on the formula:

$$\delta = \frac{||E_2 - E_6|| + ||E_3 - E_5||}{2||E_1 - E_4||} \qquad (1)$$

where δ represent EAR and E_n represent the landmark on the eyelids and n represents the location of the landmark.

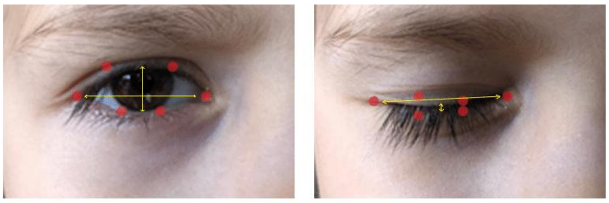

Fig. 5. Eye vertical and horizontal coordinates

This equation's denominator, which is weighted correctly because there is only one set of horizontal points but two sets of vertical points, computes the distance between horizontal eye landmarks while the numerator computes the distance between vertical eye landmarks. When a blink occurs, the eye aspect ratio will quickly drop below the threshold even if it is roughly constant when the eye is open. The ratio of eye landmark distances can be used to assess whether someone is blinking by using this straightforward equation. The average EAR is determined after the individual EARs for each eye are obtained; in this approach, the EAR threshold is set at 4.

Fig. 6 shows the mouth vertical and horizontal coordinates. The mouth aspect ratio can be calculated using the formula:

$$\delta = \frac{||P_2 - P_6|| + ||P_3 - P_5||}{2||P_1 - P_4||} \qquad (2)$$

where δ represent MAR and E_n represent the landmark on the lips and n represents the location of the landmark.

Fig. 6. Mouth vertical and horizontal coordinates

Similar to EAR, the ratio of mouth landmark distances can be used to assess whether someone is yawing by using this straightforward equation. The average MAR is determined after the individual MARs for mouth are obtained; in this approach, the MAR threshold is set at 1.8 if the ratio was below the threshold the individual will be considered yawing.

2.6. *Face Detection Training*

As transfer learning is used the model for detecting facial landmarks is a pre-trained that is able or detect 468 landmarks on the face. the model was trained using 17 evenly distributed samples (based on the United Nations geo-scheme) to perform a fairness evaluation [12]. Table 2 shows the dataset used for training the model and the distribution among 17 geographical regions.

Table 2. Dataset Used for Training

Dataset	Samples	Distribution
Dataset I	720 Samples	40 images per area plus 40 images without faces
Dataset II	800 Samples	400 Male + 400 Female
Dataset II	425 Samples	350 photos with one face on each, and 75 photographs for each category of skin tone

3. Results and Discussion

The distance between the camera and the user and the brightness of the room's lighting are the two factors that determine the system output. A laptop camera with a resolution of 640 x 480 at 30 frames per second is employed in this setup. The number of frames set for each warning is 20 frames, and these values are intended to prevent needless notifications when a user's eyes are naturally closing due to blinking, or any other facial emotion [13]. Fig. 7, Fig. 8, and Fig. 9 illustrate the angles which are best for detection. As a result, face roll and pitch (tilt) angles cannot deviate from the straight alignment by more than 45 degrees. Yaw (pan) angles must not be greater than 90 degrees. These angles are best fitted to detect the driver's behavior, while maintaining the range of these angles to ensure the accuracy increases.

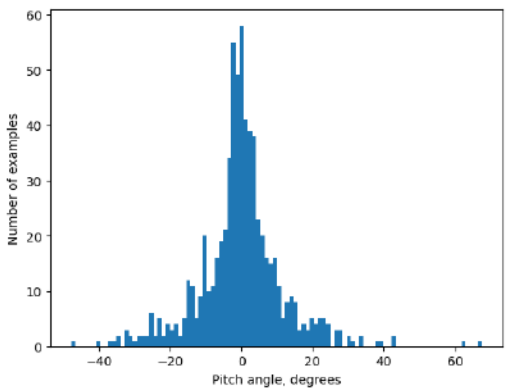

Fig. 7. Detection Pitch Angle

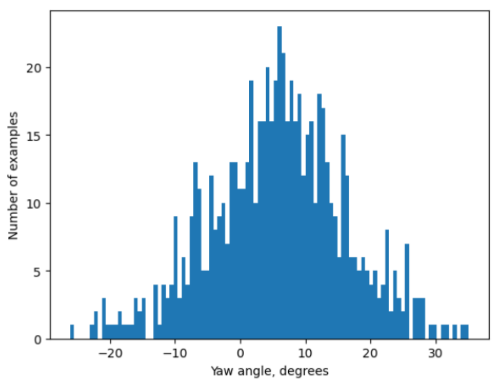

Fig. 8. Detection Yaw Angle

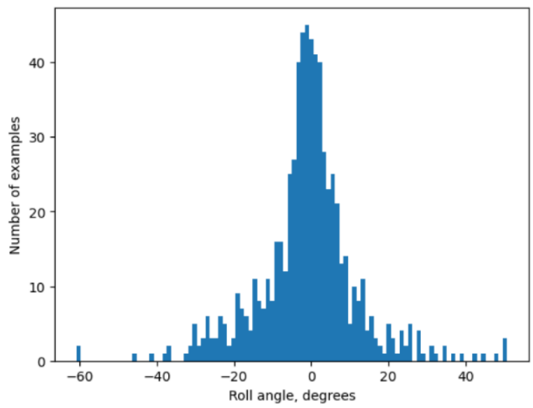

Fig. 9. Detection Roll Angle

The EAR was monitored to check the Eye closure, if the person's eyes closure exceeds the threshold the model will hold for 20 frames, and then alert the driver. After supervising the EAR, Fig. 10 shows the EAR value when the eyes are open, the values are ranged from 2.8 to 3.2, which considered as standard to open eyes. MAR was observed to check if the Mouth was considered yawning or not yawning, if the person's MAR ratio dropped below the threshold the model warns the driver and prompt the driver to take a rest. After supervising the MAR, when the mouth is closed the MAR value is undefined using the formula, however it can vary from a person to another due to different face features. Fig. 11 shows the values while the person is talking giving a range above 1.8.

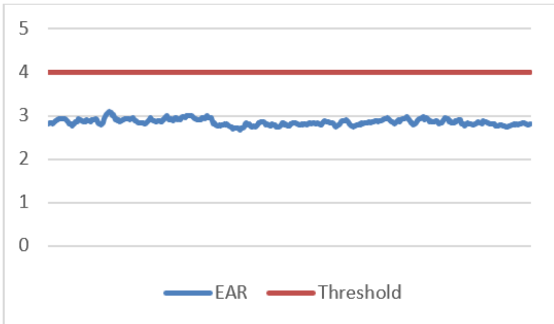

Fig. 10. Eye Aspect Ratio Monitoring

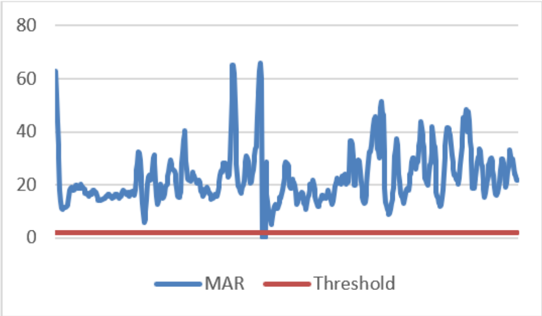

Fig. 11. Mouth Aspect Ratio Monitoring

The last stage is the communication between the system and the user. As the EAR and MAR computed, based on its results the system will voice out alerts to the user. Since the driver is drowsy sending out voice signals is very crucial to the state to warn the driver and the surrounding of the situation. The resulting eye or mouth reaction will be divided into 2 alerts once the driver yawn the system will send out warning by using pyttxs3 library which transforms writing into a speech sound, telling the driver to rest (Drowsy Warning … Take rest), the second alert is when the driver has closed his eyes for 20 frames counter then the system will voice out alerts to wake up the driver (Drowsy Alert … Wake up) the action will be repeated as long the driver is considered drowsy. As the day goes the model will adapt to the surrounding luminance around the driver, the accuracy of the detection is dropped to 91% at night due to the luminance while the accuracy will increase to 98.3% in good light condition. Table 3 displays the system's detection accuracy results when the ambient light level is 300 lux or more, which indicates that the interior of the automobile is sufficiently lit. A lux meter is used to evaluate the brightness of the surrounding environment. According to the results, accuracy is highest between 80 and 100 cm, then decreases as distance rises. 94.33% accuracy as the average detection is achieved through different lux settings. Fig. 12, Fig. 13, and Fig. 14 show the detection of the model tested on a different set of light.

Table 3. Detection Accuracy with Different Light and Distance

Luminance in Lux	Accuracy
> 300 lx	98.3%
100 > lx >300	95%
< 100 lx	91%

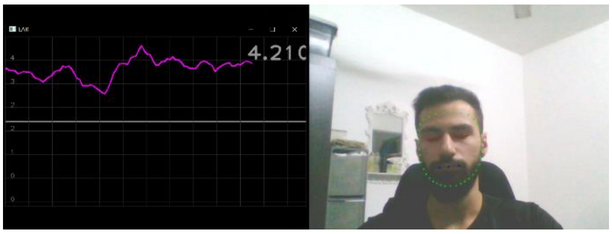

Fig. 12. Detection in > 300 lx.

Fig. 13. Detection in 100 > lx > 300.

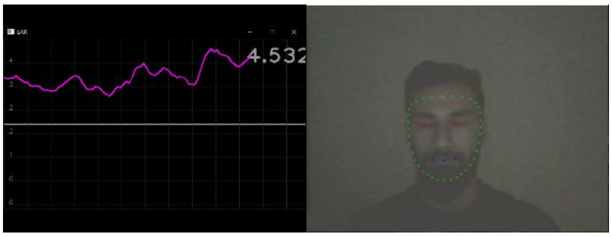

Fig. 14. Detection in < 100 lx.

A confusion matrix is a special table structure that enables visualization of the performance of an algorithm. It is sometimes referred to as an error matrix. In Fig. 15 and Fig. 16 illustrate the eye confusion matrix results. Table 4 shows the evaluation metrics of the model. Table 5 shows the results compared between our model, and a model trained using YawDD dataset [13] to detect mouth and eyes. Our model has the highest set of accuracy and precision; however, our model is considered the least among the other models if comparing the recall score.

Table 4. Evaluation Model Results

Class	Accuracy	Precision	Recall	F1 Score
Eyes	99%	99.5%	77.4%	87%
Mouth	97.7%	98.6%	87.6%	92.7%

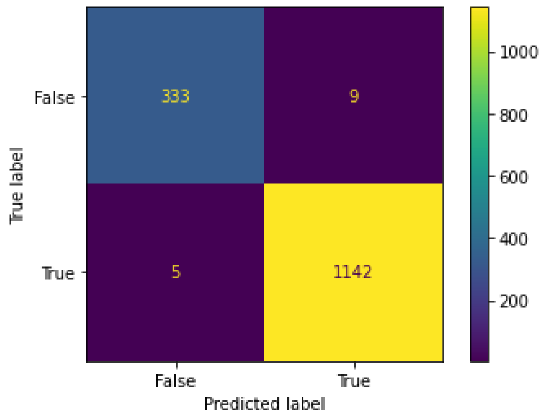

Fig. 15. Eyes Confusion Matrix

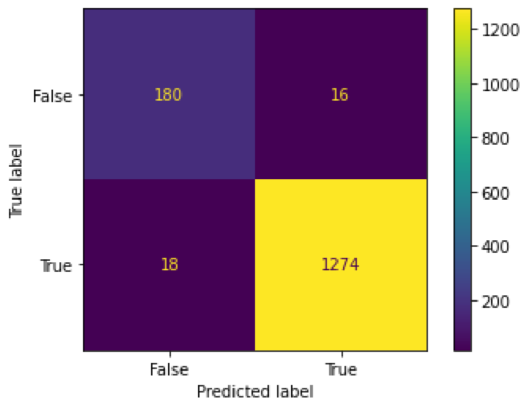

Fig. 16. Mouth Confusion Matrix

Table 5. Comparing Detection in Frame Per Second.

Models	Accuracy	Recall	Precision	F1-Score
[13]	95%	95%	95%	94.5%
Trained YOLOv5 Model	89.7%	93.7%	92.2%	92.9%
Our Model	98.3%	82.5%	99%	89.85%

Table 6 shows the comparison between other models, while Mohammad's [12] and Jonathan's [13] model achieved higher accuracy, our model has taken into accountability different environments which affects the accuracy. Table 7 compares between the detection in real-time, as shown our model did not have the highest fps between other models, because the camera used to test the model were limited to 30 – 40 fps. Better camera can achieve up 275 fps which makes the model significantly faster in detection than other models.

Table 6. Comparison with other methods

Models	Method	Accuracy
Walizad, et. al. [13]	CNN	95%
Monroy, et. al. [14]	SS-CNN	98.95%
Our Model	Mediapipe, OpenCV	98.3%

Table 7. Comparing Detection in Frame Per Second.

Authors	Detection Per FPS
Liu, et. al. [15]	60 FPS
Li, et. al. [16]	58 FPS
You, et. al. [17]	20 FPS
Our Model	30 FPS

Based on the mentioned data, the system's overall detection accuracy is 98.%. taken into consideration different light condition the accuracy decreases to 94.73%. the model did not have a high accuracy detecting glasses due to the reflection of light. In addition, better camera resolution can aid when the room's brightness is less than 300 lx. and a camera with higher frame rates will be able to detect the eye closure and mouth yawing much quicker up to 275 FPS as mentioned by the authors of MediaPipe model, producing better results with greater precision.

4. Conclusions

In this paper, OpenCV and Mediapipe are used to create a system for detecting drowsy driving. The system's goal is to provide the drivers with a constant warning for their behavior in on spot situation. The method is built around the detection of faces using a 468-point system and the detection of eyes and mouth using the eye aspect ratio and mouth aspect ratio. Results from the system revealed an average 90% percent accuracy in detection. A camera with faster frame rates and sharper quality can increase the accuracy percentage.

Acknowledgements

This work was supported by the Ministry of Higher Education Malaysia under the Fundamental Research Schemes with project codes of Proj-FRGS/1/2019/TK04/UCSI/02/1 and the UCSI University Research Excellence & Innovation Grant (REIG) with project code of REIG-FETBE-2022/038.

References

1. WHO, "Road traffic injuries," 2022. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries.
2. AAA, "Drowsy Driving: Don't Be Asleep at the Wheel," 2018, [Online].
3. Foundation of Traffic Safety, "Prevalence of Drowsy Driving Crashes: Estimates from a Large-Scale Naturalistic Driving Study," 2018.
4. NHTSA, "Drowsy Driving," 2022. https://www.nhtsa.gov/risky-driving/drowsy-driving.
5. X. Hu and G. Lodewijks, "Detecting fatigue in car drivers and aircraft pilots by using non-invasive measures: The value of differentiation of sleepiness and mental fatigue," *J. Safety Res.*, vol. 72, pp. 173–187, Feb. 2020.
6. C. C. Ukwuoma and C. Bo, "Deep Learning Review on Drivers Drowsiness Detection," 2019.
7. F. Liu, X. Li, T. Lv, and F. Xu, "A Review of Driver Fatigue Detection: Progress and Prospect," 2019.
8. M. Ngxande, J. R. Tapamo, and M. Burke, "Driver drowsiness detection using behavioral measures and machine learning techniques: A review of state-of-art techniques," in *2017 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference*, PRASA-RobMech 2017, 2017, vol. 2018-Janua, pp. 156–161.
9. A. Kost, W. A. Altabey, M. Noori, and T. Awad, "Applying neural networks for tire pressure monitoring systems," *SDHM Struct. Durab. Heal. Monit.*, vol. 13, no. 3, pp. 247–266, 2019.
10. Z. Li, S. E. Li, R. Li, B. Cheng, and J. Shi, "Online detection of driver fatigue using steering wheel angles for real driving conditions," *Sensors (Switzerland)*, vol. 17, no. 3, Mar. 2017.
11. D. R. Carvalho, W. D. A. Fonseca, J. Hollebon, P. H. Mareze, and F. M. Fazi, "Head tracker using webcam for auralization," *Proc. INTER-NOISE 2021 - 2021 Int. Congr. Expo. Noise Control Eng.*, no. August, 2021.
12. V. Bazarevsky, Y. Kartynnik, A. Vakunov, K. Raveendran, and M. Grundmann, "BlazeFace: Sub-millisecond Neural Face Detection on Mobile GPUs," pp. 3–6, 2019, [Online].
13. M. Elham Walizad, M. Hurroo, and D. Sethia, "Driver Drowsiness Detection System using Convolutional Neural Network," *2022 6th Int. Conf. Trends Electron. Informatics, ICOEI 2022 - Proc.*, no. Icoei, pp. 1073–1080, 2022.
14. J. Flores-Monroy, M. Nakano-Miyatake, G. Sanchez-Perez, and H. Perez-Meana, "Visual-based Real Time Driver Drowsiness Detection System Using CNN," *CCE 2021 - 2021 18th Int. Conf. Electr. Eng. Comput. Sci. Autom. Control*, pp. 18–22, 2021.
15. M. Z. Liu, X. Xu, J. Hu, and Q. N. Jiang, "Real time detection of driver fatigue based on CNN-LSTM," *IET Image Process.*, vol. 16, no. 2, pp. 576–595, Feb. 2022.
16. X. Li, J. Xia, L. Cao, G. Zhang, and X. Feng, "Driver fatigue detection based on convolutional neural network and face alignment for edge computing device," *Proc. Inst. Mech. Eng. Part D J. Automob. Eng.*, vol. 235, no. 10–11, pp. 2699–2711, Sep. 2021.
17. F. You, X. Li, Y. Gong, H. Wang, and H. Li, "A Real-time Driving Drowsiness Detection Algorithm with Individual Differences Consideration," *IEEE Access*, vol. 7, pp. 179396–179408, 2019.

Authors Introduction

Mr. Samer Abdullah Deeb Abbas

He is currently pursuing Bachelor of Engineering Electrical and Electronics with Honours as final year student in Faculty of Engineering at UCSI University, Malaysia. His research interests are neural network and artificial intelligence.

Dr. Sew Sun Tiang

She is an Assistant Professor in Faculty of Engineering at UCSI University in Malaysia. She received her PhD in Electrical and Electronic Engineering from Universiti Sains Malaysia in 2014. Her research interests are optimization and antenna design.

Dr. Wei Hong Lim

He is an Assistant Professor in Faculty of Engineering at UCSI University in Malaysia. He received his PhD in Computational Intelligence from Universiti Sains Malaysia in 2014. His research interests are optimization and artificial intelligence.

Dr. Li Sze Chow

She is an Assistant Professor in Faculty of Engineering at UCSI University in Malaysia. lecturer She has received Ph.D degree in Electrical and Electronic Engineering from The University of Sheffield, UL in 2006. Her research interests are detection in brain waves and deep learning.

Dr. Chin Hong Wong

He is a Lecturer in Maynooth International Engineering College at Fuzhou University in China. He received his PhD in Electrical and Electronic Engineering from Universiti Sains Malaysia in 2017. His research interests are Energy harvesting and control system.