

International Journal of

Railway Research



Train driver drowsiness detection using deep learning approach

Parsa Fakhri¹, Ramin Fathi², Ahmad Mirabadi^{3*}

- ¹MSc. Student, School of Railway Engineering, Iran University of Science & Technology
- ²MSc. Student, School of Railway Engineering, Iran University of Science & Technology
- ^{3*} Assistant Professor, School of Railway Engineering, Iran University of Science & Technology

ARTICLE INFO

ABSTRACT

Article history:

Received: 02.13.2024 Accepted: 13.06.2024 Published: 24.06.2024

Keywords: (TNR 9 Bold)

YOLO Deep learning CNN

Train driver

drowsiness detection

Safety is one of the most important priorities of the rail transport industry. Train driver sleepiness is a major threat to railway safety as it can lead to accidents and irreparable losses. In recent years, artificial intelligence and machine learning have emerged as promising approaches for developing driver drowsiness detection strategies. In this article, both deep learning (DL) and convolutional neural network (CNN) approaches are used, and methods to detect train driver drowsiness using YOLOv7 and YOLOv8 models are presented. YOLOv7 and YOLOv8 are the most recent object detection models that are effective for various tasks, including driver drowsiness detection. Our studies show that YOLOv8 outperforms YOLOv7 in terms of accuracy, speed of processing and learning, and required memory in detecting train driver drowsiness. Real-time detection features using YOLOv8 models are also demonstrated. These features can be used to detect drowsiness in real time, which can help prevent accidents.

1 Introduction

Rail transportation is one of the four main modes of transportation in the world, and it ranks first in terms of safety. This has given rail transportation special importance over other modes of transportation [1]. Train driving is one of the occupations in the rail industry, and excessive fatigue can have a significant impact on its quality [2]. Fatigue and distraction among train drivers can directly impact their attention, cognitive ability, and judgment. These factors can lead to impaired performance in train drivers, which is one of the leading causes of railway accidents. These accidents can be classified as human-related accidents [3].

Kecklund et al. [4] conducted an analysis of 79 train accidents involving a train driver and found that fatigue or drowsiness was a factor in approximately 17% of the accidents. In addition, a sample report of 246 British rail accidents has identified fatigue as a contributing factor in 21% of the accidents under review [5]. Furthermore,

studies have shown that drivers with high levels of fatigue are more susceptible to external factors, and their efficiency in controlling the train and optimizing fuel consumption decreases. Drivers are also more likely to commit severe braking and speeding violations.

In 2006 and 2008, two railway accidents occurred in the United Kingdom and the United States. In both cases, fatigue was the main factor. In the first accident, a train in the United Kingdom passed a red signal and derailed. In the second accident, due to the driver's hard work shift, a train at the O'Hare station in Chicago moved at an excessive speed, derailed, and entered the station platform. These accidents show that fatigue can be a serious threat to the safety of rail transport [6], [7].

Visual monitoring of facial states is one of the most effective methods for detecting drowsiness in train operators. This method can identify drowsiness by examining facial features such as eyes, mouth, and facial expression. Akhlaq et al.

Email address: mirabadi@iust.ac.ir

[8] indicate that eye state recognition is considered the best indicator of fatigue. In this article, using the YOLO (You Only Look Once) algorithm, which is a popular method in deep neural network image data processing, image data related to different types of people with different characteristics and types of eye states were trained and compared. Then, by using drowsiness detection algorithms, the sleepiness of the train driver was accurately diagnosed. According to Figure 1, in this study, the new version of the YOLO algorithm, namely YOLOv8, was examined. The increase in its accuracy and speed compared to the previous version, YOLOv7, was evaluated. Additionally, in this paper, a drowsiness detection program for train operators was developed using the YOLOv8 algorithm, and its real-time performance was investigated.

Deep learning has been extensively adopted in various domains, including railways, due to its proficiency in learning from diverse data and high generalizability. Applications of deep learning in railways encompass energy optimization [9], battery status diagnosis [10], obstacle detection on tracks [11], rail fracture detection [12], and more. One of its most prominent applications is in machine vision.

Deep Neural Networks (DNNs), capable of performing tasks such as image pattern recognition, have been widely used in computer vision [13]. One type of DNN is convolutional neural networks (CNNs), which are more commonly used in computer vision to analyze images in a similar way to human vision [14].

CNNs can learn complex patterns from images and signals and are therefore used in a variety of computer vision applications, such as object recognition and detection and physiological signal identification. Driver-drowsiness detection systems are often developed using pretrained CNNs on relevant datasets [15].

YOLO is one of the CNN algorithms that, by arranging convolutional layers, reframes object detection as a single regression problem, directly from image pixels to bounding box coordinates and class probabilities. With our system, you only look once (YOLO) to predict what objects

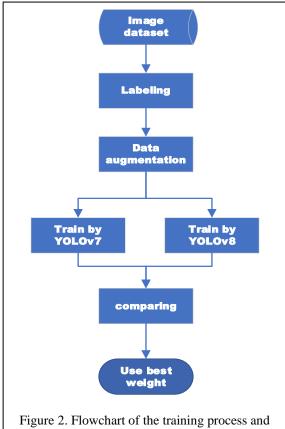
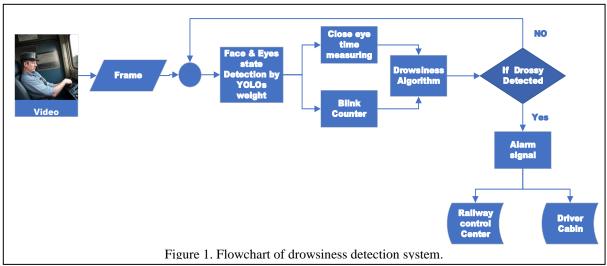


Figure 2. Flowchart of the training process and comparison.



are present and where they are. YOLO has different versions, and in this paper, as shown in Figure 2, the performance of versions 7 and 8 in this application is studied and compared.

2 Literature review

In recent years, researchers have conducted various research and experiments and developed various techniques to determine driver drowsiness in real time. Systems with different methods have been designed to detect the lack of awareness of drivers, which can prevent possible dangers that limit drivers and passengers.

2.1 Physiological methods

A physiological approach to the detection of driver drowsiness involves measuring electrical signals from the human body. These signals include electroencephalography (EEG) or brain signals [16], electrocardiography (ECG) or heart signals [17], electromyography (EMG) or muscle cell electrical signals [18], and electrooculography (EOG) signals that record eye movements [19]. Electrodes are attached to specific parts of the body to measure these signals. The signals are then analyzed to determine the driver's drowsiness status. Physiological signals are weak and can be easily distorted by noise.

In Zhang et al. [20], a wearable system is proposed to detect fatigue in train operators using deep learning. The system uses an EEG system to collect brain signals from the driver. The signals are then processed using SVM and FFT algorithms to determine the operator's state of consciousness. If the operator's state of alertness is below a certain threshold, the system activates an alarm to prevent fatigue [20].

Karuppusamy et al. [21] present a hybrid system that uses EEG, gyroscope, and image processing data to detect and predict driver fatigue. EEG signals are used to detect the transition between wakefulness and drowsiness based on alpha waves and other frequency bands. Gyroscope data is used to track the driver's head movement based on the rotation angle and field of view. A module called Vision is also used, which uses image processing to detect driver facial expressions such as yawning, closed eyes, and closed eyes while yawning.

Tuncer et al. [22] have proposed two novel methods for driver fatigue detection based on

EEG signals. The first method is a hybrid and multi-level feature generator that uses two lightweight and simple feature extraction functions: binary pattern (BP) and statistical feature extraction. The second method is a feature selector called RFINCA that is used to select the most informative and discriminative features from the extracted feature set. The two methods were evaluated on a publicly available EEG driving fatigue dataset, and the results showed that they perform well. The hybrid and multi-level feature generator extracted 1315 features, and RFINCA selected 55 of the most valuable ones. These features were used as input to 18 shallow classifiers, including k-nearest neighbors (k-NN), support vector machines (SVMs), and decision trees, and a classification accuracy of 100% was achieved. These results demonstrate the success of the proposed cognitive strategy for driver fatigue detection based on EEG.

2.2 Behavioral method

The first parameter that has attracted the most attention is PERCLOS, or Percentage of Eye Closure. In the paper by Daskupta et al. [23], a system is presented that uses cameras to measure driver attention. In this system, face detection is performed using Haar-like features, the face is tracked using a Kalman filter, and then eye state classification is performed using SVM. Finally, the PERCLOS metric is used, which is the ratio of the time that the driver's eye lids are closed by more than 80%. The results of the experiments show that this system can accurately monitor the driver's alertness level and activate an alarm in case of drowsiness.

Zhao et al. [24] propose a hybrid approach for driver fatigue detection using CNNs. The approach consists of two stages: the first stage detects the face and landmarks on the face using the multitask cascaded convolutional network (MTCNN) architecture, and the second stage detects the eye and mouth states using EM-CNN. The approach evaluates driver fatigue using the PERCLOS and POM (i.e., mouth opening degree) parameters. The proposed approach performs very well in detecting driver fatigue. Its accuracy and sensitivity are 93.62% and 93.64%, respectively. The authors claim that their proposed approach can improve driver fatigue detection by reducing the impact of environmental factors, achieving richer facial information, and improving the adaptability and flexibility of the approach.

Akrout et al. [25] propose a non-invasive system for detecting driver hypovigilance. The system uses three main analyses: yawn index, blinking index, and head pose estimation. First, the mouth area is identified. Then, the degree of mouth opening is measured at regular time intervals using a spatiotemporal feature. The proposed spatiotemporal feature is motion guided by optical flow. Experimental results show that this method performs well, with recall and precision rates of 84% and 85%, respectively. Akrout et al. [25] also identify the iris and two eyelids of the driver. Then, a normalized geometric descriptor is calculated, and finally, a non-stationary, non-linear signal is generated that describes the spatial-temporal variation of these features. The analysis of the non-stationary, non-linear signal is based on a decomposition into intrinsic functions using the empirical mode decomposition (EMD) and back-propagation (BP) algorithms. The major advantage of this decomposition is that the basic functions are obtained from the signal itself. Finally, a new approach is used to estimate the 3D position of the head section in the network structure. This approach is based on a combination of head geometric features and optical flow features. Experimental results show that this approach has high accuracy in estimating the 3D position of the head.

3 Methodology

Finding the face and eyes of a driver within an image is a key step in detecting driver drowsiness. In this paper, we use approximately 6115 images of faces in different states from the roboflow.com website [26]. In this dataset, the locations of the driver's face and eyes on the images are marked and labeled with bounding boxes.

3.1 YOLOv7

This version uses a convolutional neural network for object detection and has been improved in terms of both accuracy and speed. YOLOv7 uses a deeper neural network that can learn more complex features from images. These more complex features help the algorithm more accurately detect objects in the image. YOLOv7 uses a new method for training the neural network called "trainable bag-of-freebies." This method allows the algorithm to use several improvement techniques accuracy simultaneously. Additionally, YOLOv7 uses a set of pre-made boxes (anchor boxes) to predict the position and size of the object [27].

YOLOv8

YOLO-v8, the latest version of the YOLO family, was released by Ultralytics in January 2023. Initial comparisons with previous versions show that YOLO-v8 has made significant progress over previous versions in terms of speed and accuracy. YOLO-v8 improved its backbone by introducing C2f and replacing C3. The stem's first 6x6 convolution was replaced by a 3x3, the main building block was changed, and C2f replaced C3. In C2f, all the outputs from the Bottleneck module (a fancy name for two 3x3 residual connections) with concatenated. In C3, however, only the output of the last Bottleneck was used. YOLOv8 uses anchor-free object detection, in which the model directly predicts the center of the object. This makes the model more flexible and able to detect more unusual objects, as well as improving its accuracy and efficiency. One of the other features of YOLOv8 is closing the mosaic augmentation, which is used in the training of object detection models. This method involves stitching four images together to form a single image. This helps the model recognize objects in new locations against different surrounding pixels [28].

3.3 **Drowsiness detection algorithms**

The human blinking rate varies depending on the situation and the level of concentration. The average blinking rate for an adult is approximately 15-25 blinks per minute [29], but this rate can decrease to 4-5 blinks per minute when performing a task that requires high concentration [30]. For example, train drivers, who are required to focus on the track and onboard equipment at all times, blink approximately 8–10 times per minute. This is because blinking can interfere with visual attention and reduce the driver's ability to see potential hazards [30].

The average duration of a human blink is 100– 400 milliseconds [29]. This distinguishes blinking from eye closure, which is typically defined as a blink that lasts for more than 0.5 seconds.

4 Results of trained algorithms

In this study, pre-trained weights were used to initialize the weights of each model. The batch size was set to 32, and the image size was set to 380 for training. The models were trained on a GPU using Google Colab. In addition, data augmentation techniques, including rotation, salt and pepper noise, brightness augmentation, and dropout, were used in the pre-training phase to prevent overfitting and improve the accuracy of the system. Additionally, 400 images from the 6115-image dataset were used for validation and 125 images for testing. Each version of the model was trained for 25 epochs, which resulted in significant improvements in accuracy and speed.

The YOLOv8 model has 168 layers and 3,006,428 parameters. The results of training this model show that YOLOv8 has a mean average precision (mAP50) of 97.1% and a training time of 26 minutes and 4 seconds. The results of YOLOv7 training also show that YOLOv8 has made significant progress over YOLOv7. The training time for all epochs is approximately 93 minutes and 18 seconds.

The YOLOv7 model has 415 layers and 37,218,132 parameters. This model has a lower processing speed than YOLOv8, processing each image in 4.37 milliseconds. This is due to the larger number of layers and parameters in YOLOv7 than in YOLOv8. The mAP50 of YOLOv7 in all classes is 94.2%. In Table 1, the comparison of the results of two algorithms is clearly stated.

$$mAP = \frac{1}{n} \sum_{k=1}^{k=n} AP_k$$

$$AP_K = \text{the AP of class } k$$

$$n = \text{the number of classes}$$

Based on the comparison results in Table 1, it can be concluded that the YOLOv8 model has higher accuracy and speed than the YOLOv7 model. However, YOLOv7 has a larger number of parameters, which means that it requires more

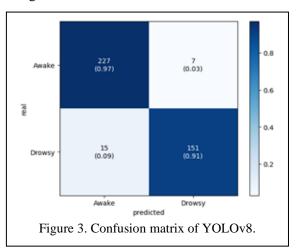
Table 1. Comparison of results obtained from two versions of the YOLO algorithm. Processing Percision Recall mAP50% time (sec) YOLOv7 93% 92.4% 94.2% 4.37 YOLOv8 96.1% 94.3% 97.1% 1.6

memory and training time, resulting in slower learning speed.

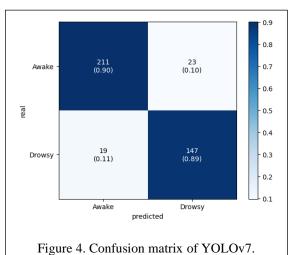
Recall measures the ability of a model to correctly identify positive instances.

$$Recall = \frac{TP}{TP + FN} = \frac{true \ object \ detection}{all \ ground \ truth \ boxes}$$

In a validation set of 400 images, a YOLOv8 model is shown in Figure 3. The algorithm-based network correctly classified 91% of drowsy images (151 out of 166 images). This means that 9% of drowsy images (15 out of 166 images) were incorrectly classified as awake images. Additionally, the model correctly classified 97% of awake images (227 out of 234 images). However, 3% of awake images (7 out of 234 images) were incorrectly classified as drowsy image.



Nevertheless, the results of training the YOLOv7 algorithm under the same conditions were slightly worse, Figure 4. The network correctly classified 89% of drowsy images (147 out of 166 images). This means that 11% of drowsy images (19 out of 166 images) were



rigare i. Comusion matrix of 102017.

incorrectly classified as awake images. Additionally, the model correctly classified 90% of awake images (211 out of 234 images). However, 10% of awake images (23 out of 234 images) were incorrectly classified as drowsy images.

A comparison of the two paragraphs presented shows that the YOLOv8 model has better performance than the YOLOv7 model in classifying drowsy and awake images. The YOLOv8 model has an accuracy of 91% in classifying drowsy images, while the YOLOv7 model has an accuracy of 89%. Similarly, the YOLOv8 model has an accuracy of 97% in classifying awake images, while the YOLOv7 model has an accuracy of 90%. Overall, the YOLOv8 model can classify drowsy and awake images with higher accuracy than the YOLOv7 model. This suggests that the YOLOv8 model may be more suitable for applications that require high accuracy in classifying drowsy and awake images.

5 Software program

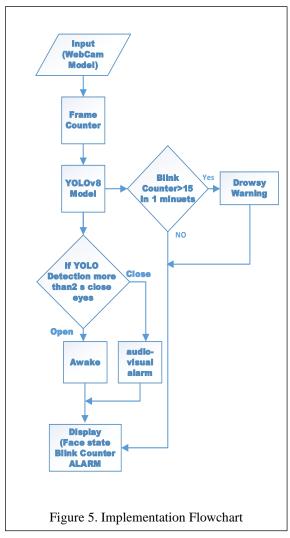
YOLOv8 is a more suitable model for train driver drowsiness detection due to its higher accuracy and speed. This model is capable of detecting faces and the state of eye closure in images. To display the results of the trained model in real time, images must be sequentially fed into the network. In this study, a laptop webcam and the OpenCV and ultralytics libraries in the Python programming language were used to perform this process.

This program was run on a system with a base processing speed of 1.8 GHz on an Intel Core i7-10510U processor with a frame rate of 15 frames per second. This means that each frame is processed in 66.6 milliseconds. Since each frame takes 66.6 milliseconds, a blink must occur within two to five frames. In other words, if a blink starts at second 1, it must end at least at second 3 and at most at second 6.

5.1 Implementation

According to the definition provided in the Methodology section and Figure 5, if the eyelid is closed for more than 500 milliseconds, it is no longer considered a blink and is considered a closed eye. In addition to drowsiness, other factors can also cause the eyelid to close for a longer period of time. For example, the sudden

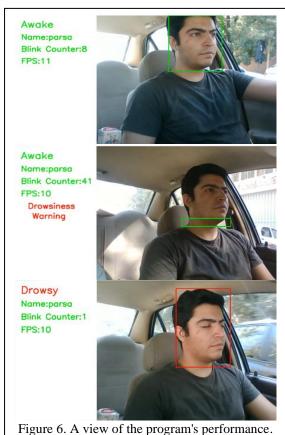
entry of dust from the locomotive window or eye sensitivity can cause the driver's eye to close. However, due to the sensitivity of the train driver's job and the risk of irreparable accidents,



the driver cannot pay attention to the front and onboard equipment for a long time. Therefore, a threshold was defined that would alert the driver with an audible and visual warning if the eyelid was closed for more than two seconds.

As previously mentioned, the frequency of blinking decreases to 4-5 times per minute when a person is focused. Additionally, based on studies conducted, drivers blink between 8 and 10 times per minute. Considering that the speed of the network overlaps with the speed of human blinking, a blink counter was used. Given that factors other than drowsiness can temporarily increase blinking, a threshold was defined: if the number of blinks exceeds 25 blinks per minute, the system predicts driver drowsiness and issues a visual alert for the driver while sending a message to the railway control center.

As shown in Figure 6, in the normal state when the system is working, there is no warning, but in the next picture, the number of blinks was more than 25 blinks per minute, so a visual warning has been applied. In the third picture, the person's eyelid has been closed for more than 2 seconds, and an audio and visual warning have been applied.



6 Conclusion

In this paper, deep learning and real-time facial monitoring of train drivers were used to detect train driver fatigue. A number of 6115 image datasets were trained for 25 epochs using the recently released YOLOv7 and YOLOv8 algorithms. The results obtained in Figure 7 show that the accuracies of YOLOv7 and YOLOv8 are 94.2% and 97.1%, respectively. The training speed of each image in YOLOv8 is also approximately 3 times faster than in YOLOv7.

In this paper, two versions of an algorithm were compared. In general, both versions of YOLO performed well in detecting drowsiness. However, YOLOv8 shows a 2.9% increase in accuracy over the previous version due to fewer parameters.

The YOLOv8 model has also demonstrated its effectiveness, speed, and accuracy in real-time driver drowsiness detection. In addition, a system is designed that can use the trained network to alert the driver's sleep and wakefulness. The system can also predict drowsiness using a blink counter due to its high speed at detecting blinks.

Despite the good performance of the current system in detecting drowsiness, there is still room for improvement. Future work could improve the accuracy and reliability of the system by combining more sophisticated techniques and additional drowsiness indicators. In addition, the system can be converted into a comprehensive train driver supervision system by using images with different labels, which can identify violations such as smoking, drinking, and others in the train driver's cabin.

References

- [1] R. Naseri and S. Mohammadzadeh, "Nonlinear Train-Track-Bridge Interaction with Unsupported Sleeper Group," International Journal of Railway Research, vol. 7, no. 1, pp. 11– 28, 2020.
- [2] H. Iridiastadi, "Fatigue in the Indonesian rail industry: A study examining passenger train drivers," Appl Ergon, vol. 92, p. 103332, 2021.
- [3] B. S. Dhillon, Human reliability and error in transportation systems. Springer Science & Business Media, 2007.
- [4] G. Kecklund, T. Åkerstedt, M. Ingre, and M. Söderström, "Train drivers' working conditions and their impact on safety, stress and sleepiness: a literature review, analyses of accidents and schedules," National Institute for Psychosocial Factors and Health. Stress Research Report, vol. 288, 1999.
- [5] N. Bowler and H. Gibson, "Fatigue and its contributions to railway incidents," 2015.

- [6] NTSB Railroad Accident Report, "Chicago Transit Authority Train Collides with Bumping Post and Escalator at O'Hare Station, Chicago," NTSB Railroad Accident Report, Illinois, Apr. 28, 2014.
- "Derailment of a freight train at [7] Brentingby Junction, near Melton Mowbray," 2007. [Online]. Available: www.raib.gov.uk.
- [8] M. Akhlaq, T. R. Sheltami, B. Helgeson, and E. M. Shakshuki, "Designing an integrated driver assistance system using image sensors," J Intell Manuf, vol. 23, pp. 2109–2132, 2012.
- [9] R. Tang et al., "A literature review of Artificial Intelligence applications in railway systems," Transp Res Part C Emerg Technol, vol. 140, p. 103679, 2022.
- [10] S. M. Mousavi Gazafroudi, "Estimating the battery life of an electric train using the ANFIS model," International Journal of Railway Research, vol. 10, no. 2, pp. 82-93, 2023.
- [11] D. He, Z. Zou, Y. Chen, B. Liu, X. Yao, and S. Shan, "Obstacle detection of rail transit based on deep learning," Measurement, vol. 176, p. 109241, 2021.
- [12] C. Dang et al., "The Accelerated Inference of a Novel Optimized YOLOv5-LITE on Low-Power Devices for Railway Track Damage Detection," IEEE Access, vol. 11, pp. 134846-134865, 2023.
- [13] M. Alam, M. D. Samad, L. Vidyaratne, A. Glandon, and K. M. Iftekharuddin, "Survey on deep neural networks in speech and vision systems," Neurocomputing, vol. 417, pp. 302-321, 2020.

- [14] V. Buhrmester, D. Münch, and M. Arens, "Analysis of explainers of black box deep neural networks for computer vision: A survey," Mach Learn Knowl Extr, vol. 3, no. 4, pp. 966–989, 2021.
- [15] A. Bhetuwal and K. C. Siddanta. "Driver's Drowsiness Detection System".
- [16] F. Wang, S. Wu, J. Ping, Z. Xu, and H. Chu, "EEG driving fatigue detection with PDC-based brain functional network," IEEE Sens J, vol. 21, no. 9, pp. 10811–10823, 2021.
- [17] J. Halomoan, K. Ramli, D. Sudiana, T. S. Gunawan, and M. Salman, "A New ECG Data Processing Approach to Developing an Accurate Driving Fatigue Detection Framework with Heart Rate Variability Analysis and Ensemble Learning," *Information*, vol. 14, no. 4, p. 210, 2023.
- [18] Y. Fan, F. Gu, J. Wang, J. Wang, K. Lu, and J. Niu, "SafeDriving: an effective abnormal driving behavior detection system based on EMG signals," IEEE *Internet Things J*, vol. 9, no. 14, pp. 12338–12350, 2021.
- S. Murugan, P. K. Sivakumar, C. [19] Kavitha, A. Harichandran, and W.-C. Lai, "An Electro-Oculogram (EOG) Sensor's Ability to Detect Driver Hypovigilance Using Machine Learning," Sensors, vol. 23, no. 6, p. 2944, 2023.
- X. Zhang et al., "Design of a fatigue [20] detection system for high-speed trains based on driver vigilance using a wireless wearable EEG," Sensors (Switzerland), vol. 17, no. 3, Mar. 2017, doi: 10.3390/s17030486.
- [21] N. S. Karuppusamy and B.-Y. Kang, "Multimodal system to detect driver fatigue using EEG, gyroscope, and

- image processing," *IEEE Access*, vol. 8, pp. 129645–129667, 2020.
- [22] T. Tuncer, S. Dogan, and A. Subasi, "EEG-based driving fatigue detection using multilevel feature extraction and iterative hybrid feature selection," Biomed Signal Process Control, vol. 68, p. 102591, 2021.
- [23] A. Dasgupta, A. George, S. L. Happy, and A. Routray, "A vision-based system for monitoring the loss of attention in automotive drivers," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 4, pp. 1825–1838, 2013.
- [24] Z. Zhao, N. Zhou, L. Zhang, H. Yan, Y. Xu, and Z. Zhang, "Driver fatigue detection based on convolutional neural networks using EM-CNN," *Comput Intell Neurosci*, vol. 2020, 2020.
- [25] B. Akrout and W. Mahdi, "A novel approach for driver fatigue detection based on visual characteristics analysis," *J Ambient Intell Humaniz Comput*, vol. 14, no. 1, pp. 527–552, 2023.
- [26] "Drowsiness / Fatigue_Detection Detection," www.roboflow.com.
- [27] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 7464–7475.
- [28] Jacob Solawetz and Francesco, "What is YOLOv8? The Ultimate Guide.," www.roboflow.com.
- [29] J. A. Stern, L. C. Walrath, and R. Goldstein, "The endogenous eyeblink," *Psychophysiology*, vol. 21, no. 1, pp. 22–33, 1984.

[30] A. R. Bentivoglio, S. B. Bressman, E. Cassetta, D. Carretta, P. Tonali, and A. Albanese, "Analysis of blink rate patterns in normal subjects," *Movement disorders*, vol. 12, no. 6, pp. 1028–1034, 1997.