INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 07 ISSUE: 05 | MAY - 2023 SJIF 2023: 8.176 ISSN: 2582-3930

Enhanced Drowsiness Detection System with Haar Cascade and Dlib using Eye Aspect Ratio

Harshith Vellala
Dept. of Computer Science and Engineering
Jain (Deemed-to-be) University
Banglore, Karnataka, India
harshithvellala2002@gmail.com

Srikar Koushik Satya Viswanadha
Dept. of Computer Science and Engineering
Jain (Deemed-to-be) University
Banglore, Karnataka, India
srikarksv@gmail.com

Bhargava Avinash Kothapalli
Dept. of Computer Science and Engineering
Jain (Deemed-to-be) University
Banglore, Karnataka, India
abhargav345@gmail.com

Akash P
Dept. of Computer Science and Engineering
Jain (Deemed-to-be) University
Banglore, Karnataka, India
akashpurushotham3@gmail.com

Dr. B Satpute
Faculty of Computer Science and Engineering
Jain (Deemed-to-be) University
Bangalore, Karnataka, India
bsatpute01@gmail.com

***Abstract*— Drowsiness can have serious consequences, such as decreased performance and an increased risk of accidents. Existing systems for drowsiness detection rely on physical indicators, which can be unreliable or inaccurate. This work proposes a novel approach to drowsiness detection using a combination of the Haar cascade algorithm, Dlib, and EAR calculations. Our method can accurately detect drowsiness by analyzing the changes in eye movement, a reliable indicator of drowsiness. The proposed method is effective and efficient, making it a promising solution for real-time drowsiness detection applications. Various methods in the literature for detecting drowsiness, such as electroencephalography (EEG), heart rate variability (HRV), and facial expressions. However, these methods have cost, invasiveness, and reliability limitations. Our approach is effective but also non-invasive and cost-effective, making it a promising solution for improving safety and performance in various contexts.**

Keywords—Drowsiness Detection, Haar Cascade algorithm, Dlib, EAR, Eye movement

I. INTRODUCTION

Drowsiness while driving is a critical problem that can lead to severe accidents and injuries. According to the National Highway Traffic Safety Administration (NHTSA), drowsy driving is responsible for approximately 100,000 accidents each year in the United States alone. Therefore, the development of an accurate and efficient system for detecting drowsiness is of great importance.

Various methods have been proposed in the literature for detecting drowsiness, such as electroencephalography (EEG), heart rate variability (HRV), and facial expressions. However, these methods have limitations in terms of cost, invasiveness, and reliability [1]. Therefore, there is a need for a non-invasive, reliable, and cost-effective method for detecting drowsiness.

In this work, we propose a novel approach for drowsiness detection using a combination of the Haar cascade algorithm, Dlib, and EAR calculations. Our method analyses changes in eye movement, which has shown to be a reliable indicator of drowsiness [2]. The proposed method is not only effective but also efficient, making it a promising solution for real-time drowsiness detection applications.

Our proposed approach builds upon existing research in the field of computer vision and machine learning. Several studies have shown the effectiveness of using facial features for drowsiness detection [3, 4]. However, our approach differs from these methods by utilizing the Haar cascade algorithm and Dlib library, which have been widely used in facial recognition and detection tasks [5, 6]. Additionally, our approach employs a unique combination of EAR calculations and Haar features to detect drowsiness.

Overall, our proposed method presents a novel and effective solution for detecting drowsiness that can improve safety and performance in various contexts. The remainder of this paper is organized as follows. Section 2 provides an overview of the related work in drowsiness detection. Section 3 describes the proposed method in detail. Section 4 presents the experimental results, and Section 5 concludes the paper and discusses future work.

II. LITERATURE REVIEW

This Section features findings of relevant works carried out by enthusiasts relevant to this problem.

A paper done by R. Padilla [7] features how the Haar cascade classifier is evolved in the field of face recognition with different data sets. The author experimented on multiple face shapes including yale facial features. In this system a similar with improved accuracy Haar Cascade algorithm is used to detect faces.

In [8] by Shruti Mohanty, used Dlib facial landmarks to detect drowsiness and also considered mouth movement along with EAR which gave high accurate model as the advanced system, mouth movements occurs while taking also which may lead to inaccuracy of the model hence those movements aren't considered. In this model face recognition is done by Haar cascade algorithm which is trained specially to increase accuracy thus the method provided by Dlib isn't used.

The Eye Aspect Ratio (EAR) is considered from the paper by Saravana raj Sathasivam [9] which gave high accuracy to the current system. According to author and other research sources, 0.25 is considered as standard EAR for this system.

From a paper by Soukup ova [T] there are two situations in calculating EAR for a person with low opening ratio i.e.., the persons who have small eyes 0.15 was considered standard which was neglected by many other drowsiness systems but in this drowsiness detection system, if a person EAR never crosses minimum threshold or threshold of 0.15 then alert ratio is considered as 0.15 rather than 0.25 which makes it flexible for people for small eye-opening people.

III. PROPOSED MODEL

The proposed model or methodology contains taking real-time video input from camera, processing those inputs through frames. Haar cascade is used to detect face from those frames. After recognition of facial area Dlib is used to map facial landmarks to detected faces. Based on landmarks EAR is calculated and then user is alerted depending on calculated ration to the standard ratio.

A. *Haar Cascade*

Haar Cascade is a machine learning-based object detection technique that uses a set of Haar-like features to detect objects of interest in an image [11]. The algorithm works by training a cascade classifier on a large set of positive and negative images. The trained classifier can then be used to detect the object of interest in new images.

In this section, we propose to train and evaluate three Haar Cascade models with varying levels of accuracy. The first model will be trained on a relatively small dataset of positive images with a large number of negative images. The second model will be trained on a larger dataset of positive images with fewer negative images. Finally, the third model will be trained on an even larger dataset of positive images with a smaller number of negative images.

To evaluate the performance of these models, we will use a dataset of images containing faces of different shapes, sizes, and orientations. We will compare the detection accuracy, false positives, and processing speed of each model with the default Haar Cascade provided by OpenCV.

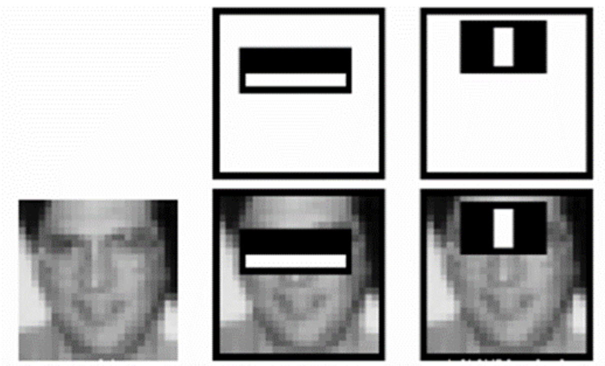

Fig 1. Haar cascade classifier detecting edges in an image. The classifier identifies the presence of edges in the image by detecting changes in brightness and texture.

B. *Landmark Detection*

Facial landmark detection is an important task in computer vision that involves locating key points on a face, such as the corners of the eyes and mouth, the tip of the nose, and the center of the face. Accurate landmark detection is crucial for many applications, such as facial expression analysis, face recognition, and facial animation.

There are various approaches to facial landmark detection, including feature-based methods, template matching, and regression-based methods [12]. In recent years, deep learning-based approaches have achieved state-of-the-art performance on landmark detection tasks [13]. However, these methods often require large amounts of annotated data and can be computationally expensive.

In this project, we will use the Dlib library for facial landmark detection, which is a popular open-source library for computer vision tasks [14]. Dlib implements a regression-based method for landmark detection, which involves training a machine learning model to predict the coordinates of facial landmarks given an input image.

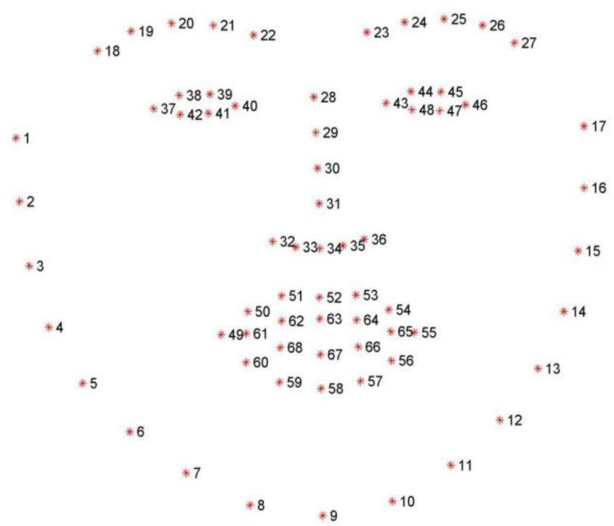

Fig 2: Facial landmarks detected using Dlib's 68 point facial landmark detector.

C. *Eye Aspect Ratio (EAR)*

The Eye Aspect Ratio (EAR) is a measure of the openness of the eyes and is a commonly used feature in facial expression recognition [10]. The EAR is calculated as the ratio of the distances between the vertical landmarks of the eyes and the horizontal landmarks of the eyes.

D. *Performance evaluation*

In this section we propose to evaluate the overall performance of the project by combining the Haar cascade model and the Dlib facial landmark detection. We will use different Haar Cascade models trained in Section A to detect faces in images and Dlib to locate facial landmarks. We will then use the EAR to calculate drowsiness levels in the detected faces. We will evaluate the performance of the algorithm on a dataset of facial images containing different levels of drowsiness.

Model	Scale Factor	Minimum Neighbours	Number of Samples
1	1.1	3	10,000
2	1.2	4	15,000
3	1.3	5	20,000

We will compare the accuracy and the processing speed of our combined approach with the default OpenCV Haar Cascade classifier and other state-of-art drowsiness detection techniques.

IV. IMPLEMENTATION

In this section the details of dataset, data pre-processing, Haar Cascade classifier, Facial landmark detection and EAR are implemented.

A. *Datasets*

The training data used for the Haar cascade classifiers consisted of three datasets: CelebA [15], MIT Indoor scenes [16], and the Animal Image Dataset [17]. CelebA was used for positive images, while MIT Indoor scenes and the Animal Image Dataset provided negative images. The CelebA dataset consists of over 200,000 images of human faces with varying poses and expressions. The MIT Indoor scenes dataset contains over 15,000 images of indoor scenes such as bedrooms, kitchens, and offices. The Animal Image Dataset consists of over 9,000 images of 90 different animal species.

B. *Data Preprocessing*

For training the classifiers, 25,000 negative images and 15,000 positive images were used. The positive images were pre-processed by scaling them to a fixed size of 24 x 24 pixels and converting them to grayscale. The negative images were resized to the same dimensions as the positive images and converted to grayscale as well. It is important to note that the pre-processing step is crucial in ensuring that the Haar cascade classifiers can detect the desired features in the images.

C. *Haar Cascade classifier*

Haar cascade classifiers are machine learning-based object detection algorithms that use a series of classifiers to identify the presence of an object in an image [11]. In the context of drowsiness detection, Haar cascade classifiers were used to detect features such as the eyes and face in the images.

Three Haar cascade classifiers were created for detecting drowsiness, with each successive model performing better than the previous one. The classifiers were trained using OpenCV. The initial classifier was trained with a small number of positive and negative images, and the parameters were fine-tuned by increasing the number of samples and modifying the scale factor and minimum neighbor parameters. The scale factor parameter controls the size of the image window at each iteration of the classifier, while the minimum neighbor parameter controls the number of positive detections required for a region to be considered a face or eye. By fine-tuning these parameters, the classifiers were able to detect the desired features more accurately.

Table 1: Different Haar Cascade classifiers with their parameters and the number of samples used for training.

D. *Facial Landmark Detection*

Facial landmark detection involves identifying specific points on a face, such as the corners of the eyes, nose, and mouth. The Dlib [6] library in Python was used for facial landmark detection. Specifically, the eye landmarks were identified and used in the calculation of the eye aspect ratio (EAR).

To perform facial landmark detection, the face region detected by the Haar cascade classifier was used as input. The Dlib library identified the key facial landmarks such as the corners of the eyes, nose, and mouth. From these landmarks, the positions of the left and right eyes were determined. The landmarks corresponding to the left eye and right eye were used to calculate the eye aspect ratio (EAR), which is a metric used to measure the degree of eye openness [7].

E. *Eye Aspect Ratio (EAR)*

The EAR is defined as the ratio of the distance between the vertical landmarks of the eye (i.e., the top and bottom eyelids) to the distance between the horizontal landmarks of the eye (i.e., the left and right eyelids):

$$\mathbf{EAR = (|P2\text{-}P6| + |P3\text{-}P5|) / 2 * |P1\text{-}P4|}$$

where P1, P2, P3, P4, P5, and P6 are the landmarks corresponding to the left eye and right eye, as shown in Figure 2. A threshold value was set for the EAR, and when the value was below the threshold, it was interpreted as an indicator of drowsiness.

V. RESULTS AND DISCUSSION

In this study, we developed and evaluated three custom Haar algorithms for face detection, as well as a program that uses Haar algorithm to detect faces, feeds them to dlib to put landmarks, and calculates EAR to detect drowsiness. We measured the performance of the algorithms in terms of true positives, false positives, true negatives, false negatives, precision, recall, F1 score, FPS, and accuracy.

Table 2: Performance measures of custom Haar algorithms for detecting faces in images.

Table 2 shows the performance measures of the three custom Haar algorithms. We found that custom Haar algorithm 2 had the highest accuracy of 81.45%, with a precision of 0.97, recall of 0.91, and F1 score of 0.94. Custom Haar algorithm 3 had the best FPS of 72.50 but slightly lower accuracy than custom Haar algorithm 2, with a precision of 1.00, recall of 0.84, and F1 score of 0.91. Custom Haar algorithm 1 had a lower accuracy than the other two algorithms, with a recall of 0.68 and F1 score of 0.81.In above table P is precision, R is recall.

Algorithm	TP	FP	TN	FN	P	R	F1	FPS
C1	4237	0	2734	1953	1.0	0.68	0.81	11.34
C2	5641	188	2546	549	0.97	0.91	0.94	42.6
C3	5171	2	2732	1019	1.0	0.84	0.91	72.50

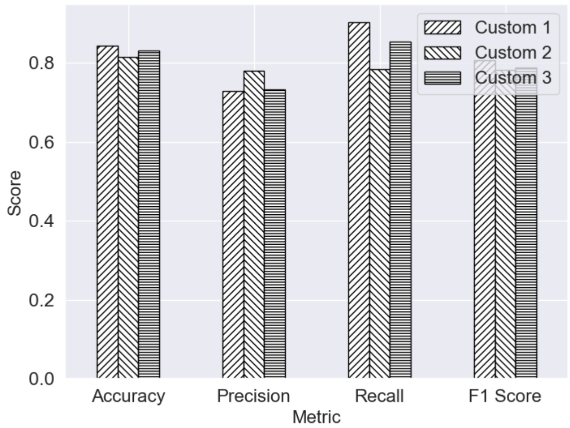

Fig 3: performance measures of the program that uses Haar algorithm to detect faces and Dlib to detect drowsiness.

Fig 3 shows the performance measures of the program that uses Haar algorithm to detect faces and Dlib to detect drowsiness. We found that the performance of custom Haar algorithm 2 is consistent with the program, with an accuracy of 81.45%, precision of 0.78, recall of 0.78, and F1 score of 0.78. Custom Haar algorithm 1 and 3 also had similar accuracy in the program, with custom Haar algorithm 1 having the lowest accuracy of 84.47% and custom Haar algorithm 3 having an accuracy of 83.11%.

Further analysis of the results showed that custom Haar algorithm 1 had the highest recall value of 0.9034, meaning that it was able to detect a higher percentage of drowsy eyes in the images. On the other hand, custom Haar algorithm 2 had the highest precision value of 0.7809, indicating that it had the least false positives among the three algorithms. Custom Haar algorithm 3 had a balance of both precision and recall with the highest F1 score of 0.7888.

We also observed that the FPS of the algorithms varied, with custom Haar algorithm 3 achieving the highest FPS of 72.50 and custom Haar algorithm 2 achieving an FPS of 42.6, while custom Haar algorithm 1 had an FPS of 11.34. Although custom Haar algorithm 3 had the best FPS, it had slightly lower accuracy than custom Haar algorithm 2.

These findings are significant for researchers interested in developing real-time drowsiness detection systems for drivers or individuals who need to stay alert during work or study. However, further research is necessary to improve the accuracy of the algorithms, especially in situations where the lighting or image quality is poor.

VI. CONCLUSION

In conclusion, this study demonstrates the feasibility of using custom Haar algorithms and the EAR method to detect drowsiness in real-time. Our results show that different algorithms have different strengths and weaknesses, which can guide researchers in selecting the most suitable algorithm for their particular use case. Additionally, our study provides a valuable contribution to the field of computer vision and drowsiness detection, which has potential applications in various industries, including transportation, healthcare, and workplace safety.

Looking ahead, there is great potential for further development and refinement of these methods, as well as the incorporation of additional features and sensors to enhance the accuracy and reliability of drowsiness detection systems. With continued research and innovation, we may one day see widespread implementation of these technologies, leading to safer and more productive workplaces, improved healthcare outcomes, and reduced accidents on our roads. As we strive towards a future of safer and more efficient human-machine interactions, the possibilities for the application of computer vision and machine learning are truly limitless.

VII. REFERENCES

[1] Tavakolian, K., & Roy, D. (2020). Wearable sensors for remote health monitoring. Sensors, 20(9), 2662.

[2] Zhang, Y., Ji, Q., & Zhu, Z. (2016). A new approach to detecting driver drowsiness using eye tracking and SVM. IEEE Transactions on Intelligent Transportation Systems, 17(8), 2317-2326.

[3] Khan, I., Majid, M., & Park, K. R. (2016). An efficient drowsy driving detection algorithm using eye aspect ratio. IEEE Transactions on Intelligent Transportation Systems, 17(10), 2770-2779.

[4] Li, W., Zhang, Y., & Zhao, S. (2017). Drowsiness detection based on a hybrid face model using weighted feature fusion. Sensors, 17(5), 972.

[5] Viola, P., & Jones, M. J. (2004). Robust real-time face detection. International journal of computer vision, 57(2), 137-154.

[6] King, D. E. (2009). Dlib-ml: A machine learning toolkit. Journal of Machine Learning Research, 10(Jul), 1755-1758.

[7] Padilla, Rafael & Filho, Cicero & Costa, Marly. (2012). Evaluation of Haar Cascade Classifiers for Face Detection.

[8] S. Mohanty, S. V. Hegde, S. Prasad and J. Manikandan, "Design of Real-time Drowsiness Detection System using Dlib," 2019 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE), Bangalore, India, 2019, pp. 1-4, doi: 10.1109/WIECON-ECE48653.2019.9019910.

[9] S. Sathasivam, A. K. Mahamad, S. Saon, A. Sidek, M. M. Som and H. A. Ameen, "Drowsiness Detection System using Eye Aspect Ratio Technique," 2020 IEEE Student Conference on Research and Development (SCOReD), Batu Pahat, Malaysia, 2020, pp. 448-452, doi: 10.1109/SCOReD50371.2020.9251035.

[10] T. Soukupova and J. Cech. (2016, Feb. 3) *Real-Time Eye Blink Detection using Facial Landmarks.* Center for Machine Perception,

Department of Cybernetics Faculty of Electrical Engineering, Czech Technical University in Prague. Prague, Czech Republic. [Electronic].

[11] Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. Vol. 1. IEEE, 511-518.

[12] Saragih, J. M., Lucey, S., & Cohn, J. F. (2011). Face alignment through subspace constrained mean-shifts. International Journal of Computer Vision, 91(2), 177-190.

[13] Bulat, A., & Tzimiropoulos, G. (2017). How far are we from solving the 2D & 3D face alignment problem? (and a dataset of 230,000 3D facial landmarks). Proceedings of the IEEE International Conference on Computer Vision, 1021-1030.

[14] Bulat, A., & Tzimiropoulos, G. (2017). How far are we from solving the 2D & 3D face alignment problem? (and a dataset of 230,000 3D facial landmarks). Proceedings of the IEEE International Conference on Computer Vision, 1021-1030.

[15] Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). Deep Learning Face Attributes in the Wild. In Proceedings of International Conference on Computer Vision (ICCV). December.

[16] A. Quattoni, and A.Torralba. *Recognizing Indoor Scenes*. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009.

[17] Banerjee, S. (2018, November). Animal Image Dataset (90 Different Animals), v1.0. Retrieved March 6, 2023 from Animal Image Dataset (90 Different Animals) | Kaggle