

REAL-TIME DROWSY FACE DETECTION FOR ONLINE LEARNING BASED ON RANDOM FOREST AND DECISION TREE ALGORITHMS

Ani Dijah Rahajoe^{*1}, Mohamad Rafli Agung Subekti², Muhammad Suriansyah³

^{1,2}Computer Science, Universitas Pembangunan Nasional Veteran Jawa Timur, Indonesia

³Information System, Universitas Diponegoro, Indonesia

Email: ¹anidijah.if@upnjatim.ac.id, ²18081010111@student.upnjatim.ac.id,

³muhammadsuriansyah@students.undip.ac.id

(Article received: November 29, 2023; Revision: December 12, 2023; published: Desember 26, 2023)

Abstract

In the current era, technology regarding artificial intelligence has developed rapidly and has been used in various areas of life. Face detection is one of the applications of Artificial Intelligence. This research aims to detect students' faces during the online learning process and succeeded in getting positive feedback when tested on students. Student detection includes drowsy and alertness. The method is via webcam in real-time so that the screen will show whether the student is drowsy or alert. In the trial, the teacher can find out who is in a drowsy and alert condition. On the other hand, students can find out that they fall into the drowsy or alert category. So that both parties immediately respond to what should be done based on the classification results. The algorithms used are Decision Tree and Random Forest. The accuracy results of the Random Forest algorithm are better than the Decision Tree algorithm, namely 65 percent, while the Decision Tree algorithm is 58 percent. The division of training data and test data uses a Kfold of 5. When Kfold is equal to 2, both algorithms have the highest accuracy, where Random Forest has an accuracy of 85 percent, and Decision Tre has an accuracy of 65 percent.

Keywords: *decision tree, eye aspect ratio, mouth aspect ratio, online learning, random forest.*

1. INTRODUCTION

After the pandemic, teaching in the form of online or offline is still being carried out. In the current era, even though it is not a pandemic, distance teaching modeling has still been implemented for several years. Distance teaching allows teaching to be carried out without considering regional or national boundaries. The advantages of distance teaching include flexible access and students can access material from anywhere. This allows access from remote areas. For time flexibility, students can set their study schedule. Variations in learning materials: online learning can involve various materials such as text, videos, and interactive animations, which can enhance students' learning experience. Social interaction and collaboration, evaluation, and monitoring are easier for teachers and allow students to continue learning to improve their skills and knowledge without physically attending class. The weaknesses of distance learning include a lack of involvement and motivation due to not gathering with teachers and friends. Lack of supervision and lack of social involvement are also disadvantages of online learning. It's just that current technological developments require long-distance face-to-face meetings.

This research aims to detect sleepiness in students attending lectures so that it can provide opportunities for lecturers or instructors to change

their way of learning so that what is taught can be accepted and understood by students.

Several studies regarding drowsiness detection are used to recognize drowsiness in drivers [1][2][3]. Meanwhile, this research will detect students who take online lectures because if the driver is for passenger safety, then the students are for effectiveness and understanding during the teaching process. There is a difference in drowsiness in drivers and drowsiness in students. In drivers, it is more due to fatigue and long driving duration [4][6][7]. The relationship between student sleepiness and effectiveness and understanding during lectures is an interesting and relevant topic in the context of higher education. Several factors can influence this relationship, namely sleep quality. Students who don't get enough sleep the night before tend to experience sleepiness during lectures. The teaching method is an interactive and engaging teaching approach that allows students to stay engaged and avoid drowsiness. The relevance of the material deemed relevant and important by students tends to keep them awake and engaged. In addition, mental health problems such as stress and anxiety can disrupt sleep and cause drowsiness in class. If the teacher finds that many students are sleepy through this research, the teacher can change the atmosphere or create icebreakers to encourage them during teaching. So that students can focus more and understand the teaching material presented by the teacher.

This research uses image data from the UTA Real-Life Drowsiness Dataset to calculate the Mouth Aspect Ratio (MAR) and Eye Aspect Ratio (EAR). In some applications, the aspect ratio of the mouth is used to detect facial expressions, emotions, or even lip movements for speech recognition and human emotion recognition. The use of image recognition and image processing technology, as well as mouth aspect ratio, can provide important data for various analytical purposes for detecting sleepiness in students during the learning process. Likewise, with the eyes, the eye area, which is a sign of sleepiness, will affect the learning process.

This research uses live video to detect whether students are classified as alert and withdrawn. Thus, if the teacher knows that many students are sleepy, the teacher can provide a different teaching method that is more interactive. In [1][2], the research used PERCLOS to break down the driver's face and eyes, whereas EAR and MAR used the KNN algorithm to compare. The HAAR method and LBPH technique are used for drowsiness detection [5]. The Haar detection method is a popular technique in object recognition and image processing to detect objects or patterns in images or videos, and the LBPH (Local Binary Pattern Histogram) technique is a feature extraction and pattern recognition method in images that can be used in various applications, including facial recognition and texture analysis [5]. This method was discovered by two researchers, Paul Viola and Michael Jones, and introduced in their paper in 2001.

Detection of drowsiness in drivers using the SVM algorithm was carried out by [9]. The research uses GPS (Global Positioning System), which is connected to a Raspberry Pi via a USB serial port. The algorithms used by researchers are mostly CNN, DNN (deep neural network) SVM, XGBoost, and Fuzzy logic [10][11][12].

Eye contour extraction was performed by [13] by calculating Sclera Extraction (the white part of the eye). However, this research only detects faces, not driver detection. The use of driver data is better in Real-time Drowsiness and Distraction Detection than traditional methods [8][11][14]. Detection of driver drowsiness was also researched by [7]. Only without going through a classification process [7][15]. In this proposed research, classification is carried out because the drowsiness and alert boundaries are different for one face from another, and using the CNN algorithm for this detection is faster and more accurate [16].

This research proposal uses test data from students at a university who are taking online lectures. It is different from the research [17] even though the data used is almost the same. A study [17] uses the Ear Aspect Ratio and is based on Natural Language Processing (NLP). Meanwhile, the research proposal uses MAR and EAR as well as real-time video to detect sleepiness during the teaching process, and the

algorithms used are Decision Tree and Random Forest.

The first contribution of this research is detecting live videos of sleepy students using the Decision Tree and Random Forest algorithms, where the Random Forest algorithm is still not widely used. The second contribution is to analyze the accuracy of the results of the decision tree and random forest algorithms. Chapter I discusses the introduction and related works. Chapter 2 discusses the methods and materials, and Chapter 3 discusses the results and discussion during the research. Chapter 4 discusses the research results. Chapter 5 concerns conclusions from research results.

2. METHODS AND MATERIALS

This section will discuss the methods used and the dataset used for the classification process.

2.1. Mouth Aspect Ratio and Eye Aspect Ratio

Mouth Aspect Ratio is a measure used to measure the shape and expression of a person's mouth in the context of image analysis and image processing. This ratio refers to the ratio between the width of the mouth and its height. In the context of face detection and facial expressions, the mouth aspect ratio can provide information about the shape of the mouth, such as whether the mouth is smiling, open, or closed. This ratio is often used in emotion recognition applications, speech analysis, and various other image-based detection systems. Many benchmarks for detecting faces include EAR (Eye Aspect Ratio) and percentage of eye closure time (PERCLOS). The use of MAR is because we look at the mouth to see how focused students are taking part in online sessions. If it is too open, it means that something is making the student unable to focus on understanding what the teacher is saying. The points used in this research are the point in the middle of the mouth and the distance between the left lip corner and the right lip corner. This research uses the Classic Histogram of Oriented Gradient (HOG) for the Dlib face detector.

2.2. Facial Landmark

In this research, facial landmark detection was used. Figure 1 is a 16-point model of the face. If you look at Figure 1, the distance calculated is point 12 and point 16 as well as the distance between points P13 and P19, P15 and P17. The reason for choosing these points is that if you are sleepy or lack concentration, the student will open his mouth wide [18]. When the mouth is open, the distance between the points extending in the middle of the lips varies. In this research proposal, these points are used for the machine learning process on the dataset.

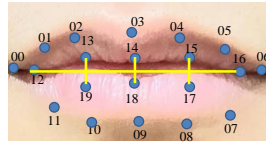


Figure 1. Dlib model in real object for MAR

MAR calculations use the equation below [10]:

$$MAR = \frac{\|P_{13}-P_{19}\|+\|P_{15}-P_{17}\|}{2\|P_{12}-P_{16}\|} \quad (1)$$

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

As input to MAR, equation 1. The placement of the points is adjusted to Figure 1, so point P1 = point 12, P2 = point 13, P3 = point 15, P4 = point 16, P5 = point 17, and P6 = point 19. In the EAR calculation, the input points have been adjusted to Figure 2. The use of EAR at these points is used to detect alert eyes, if the distances P2-P6 and P3-P5 are open, and vice versa.

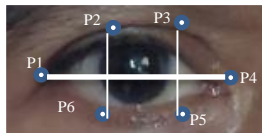


Figure 2. Dlib model in real object for EAR

2.3. Methods Used

This research used MAR and EAR to detect labels, namely drowsy drowsiness and awake alert, namely drowsy drowsiness, and awake alert. The classification algorithms used are Random Forest and Decision Tree. The choice of these two algorithms is because there is still not much discussion about face detection research using these two algorithms. Next, we will discuss the results of these two algorithms. Figure 3 is the method used in this research.

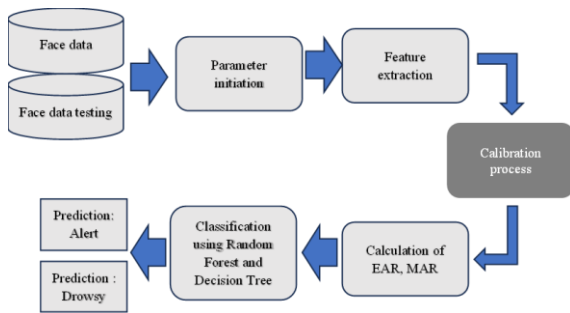


Figure 3. Machine learning flow for classification.

2.4. Dataset

The training dataset is taken from the UTA Real-Life Drowsiness Dataset (UTARLDD). The training data used was 180 RGB videos. Each video is 10 minutes long. This dataset has 9 levels to determine whether you are awake, low alert, or sleepy. These 9 levels include extremely alert, very alert, alert, rather alert, neither alert, some signs of

sleepiness, sleepy, but no difficulty remaining awake, sleep-some effort to keep alert, and extremely sleepy-fighting sleep. In this study, we only used wakefulness, namely extremely alert, very alert, and alert, and drowsiness (drowsy), namely Sleepy- some effort to keep alert and extremely sleepy- fighting sleep. Thus, not all data is taken from the source. Images are taken every 300 milliseconds. This means that an average of 3 pictures are taken from the video in 1 minute. The total dataset used as training data and test data is 2,352 images. In this study, we tested whether sleepy or awake was taken from a camera on or in real life. Then, every 300 milliseconds, an image is taken and then transformed from RGB (Red, Green, Blue) to Grayscale, then features are extracted, and MAR-EAR is calculated. The next step uses the Random Forest and Decision Tree classification algorithms for the classification process. The prediction results are displayed on the monitor screen, where they were previously changed from Grayscale to RGB. This flow is in Figure 4.

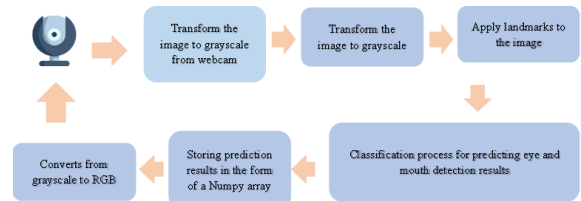


Figure 4. Method used during testing.

3. RESULTS

The prediction results are displayed on the camera screen when using new input data after implementing the machine learning process. Figure 5 shows the image processing flow when taken from the webcam. When the camera is on, every 30 milliseconds an image is taken and then processed to grayscale then the landmarks in the image are placed and the classification results appear on the screen. After the landmark process, the points are equalized and 2 is used to calculate MAR and EAR. MAR and EAR are used as features during classification.

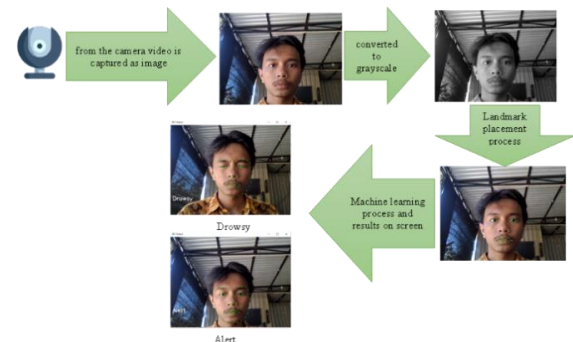


Figure 5. Application of classification results on the webcam screen

This research uses a confusion matrix to measure the performance of the classification model. The confusion matrix displays the model prediction

results compared to the actual values. True Positive (TP) is the model prediction, and the actual value is also correct. True Negative (TN) is a wrong model prediction, and the actual value is also wrong. False Positive (FP) is a correct model prediction, but the actual value is incorrect. False Negative (FN) is a wrong model prediction, but the actual value is correct. These four cells fill the confusion matrix for the Decision Tree and Random Forest algorithms. Figure 6 shows the results of the confusion matrix for Decision Tree classification.

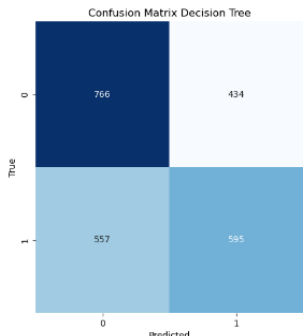


Figure 6. Confusion matrix for Decision Tree algorithm classification result.

Figure 6 has more TP and FN than TN and FP. Meanwhile, Figure 7 shows the confusion matrix results for the Random Forest algorithm. In this image, TP and FN are more than in Figure 6.

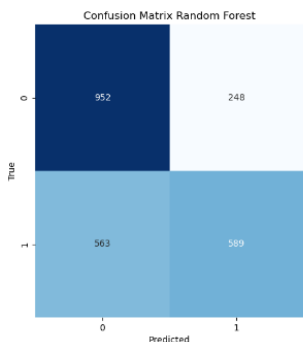


Figure 7. Confusion matrix for Random Forest algorithm classification result

A comparison of the confusion matrix between Decision Tree and Random Forest classification results shows that Random Forest is better with more True Positive and True Negative results than the Decision Tree algorithm.

The results of testing training data and testing data using Kfold Validation are between 1 and 5. Table 1 is the result of the five Kfolds and the average will be calculated.

Table 1. Calculation of classification results from the Decision Tree and Random Forest algorithms

Decision Tree Classification			
Kfold	Accuracy	Precision	Recall
1	0.4949	0.4873	0.5972
2	0.6752	0.626	0.8368
3	0.54	0.4766	0.2509
4	0.534	0.5613	0.3967

5	0.6488	0.7361	0.4858
Average	0.5786	0.5775	0.5135
Random Forest Classification			
Kfold	Accuracy	Precision	Recall
1	0.6122	0.5897	0.6849
2	0.8231	0.8358	0.7951
3	0.5638	0.5393	0.1951
4	0.639	0.7191	0.48
5	0.639	0.8054	0.3833
Average	0.6554	0.6985	0.5077

Figure 8 shows the accuracy graph of the two algorithms. It can be seen in the picture that the Random Forest algorithm has better performance than Decision Tree. Especially when Kfold 2 has maximum accuracy.

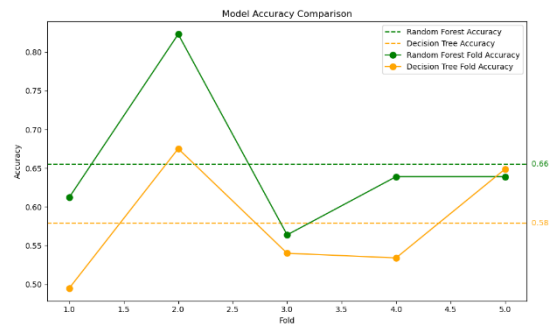


Figure 8. Comparison of the results of the fifth-fold Decision Tree and Random Forest algorithms and their averages.

4. DISCUSSION

Test results on training data and test data show that the Random Forest algorithm is better than the Decision Tree. The image on the webcam screen appears with the alert or drowsy classification results on the screen. The accuracy and speed of access affect the results of the classification. This is because it requires high access from every second of shooting. These results can detect sleepy students by appearing on the screen directly. With an online system, teachers can know what percentage of students are sleepy for various reasons.

Much research on drowsiness detection uses driver data because it is directly related to safety. Research using student data uses a different algorithm than this research [17]. Apart from that, it uses data from security guards [14]. This research uses MAR and EAR features for the calcification process. In further research, various conditions can be added as additional features for greater accuracy by connecting the conditions around the student. The result application of this research was tested on students from one of the universities and received positive feedback.

Tables 1 and 2 show the accuracy of the five Folds, where the Random Forest algorithm is better than the Decision Tree. During the machine learning process, it can be seen that Random Forest accuracy is better at each Fold stage. Graphic image 8 shows the Random Forest curve is still above the Decision Tree. In the case of face detection for online learning, the Random Forest algorithm is recommended to be

used, although other algorithms may be better. Further research on other algorithms for face detection is needed, especially those with faster access times. The speed of the machine learning process is needed to match the webcam in determining the classification results.

The feedback is in the form of a student satisfaction questionnaire in response to the teacher's positive actions to change the teaching model to increase concentration and enthusiasm to follow the lesson to completion. On the student side, when this detection is tested, students can find out whether they fall into the alert or drowsy category, so they can change their attitude and focus more on what the teacher is teaching.

5. CONCLUSION

This research uses a data set containing videos of various conditions. Testing the dataset with data from students following online criticism had positive results by knowing the classification results in real time. In testing the two algorithms, namely the Decision Tree and Random Forest algorithms, it shows that the Random Forest algorithm has better accuracy than the Decision Tree. The average accuracy of Random Forest is 65 percent, while when Kfold is equal to 2, it has the best performance, namely 82 percent. The average accuracy of the Decision Tree algorithm has a value of 58 percent and has the best performance when Kfold is equal to 2 with a value of 65 percent. The results of this research are very helpful for teachers in identifying whether students are sleepy or not, which they teach through the results that appear monitored.

REFERENCES

- [1] G. A. Wani, R. P. Singh, and M. Mehra, "Analysis of Driver Drowsiness Using Convolution Neural Network Algorithms," *Int. J. Innov. Res. Eng. Manag.*, vol. 9, no. 2, pp. 84–90, 2022, doi: <https://doi.org/10.55524/ijirem.2022.9.2.11>.
- [2] V. Naren Thiruvallar and E. Vimal, "A comparative analysis on driver drowsiness detection using CNN," *Int. J. Nonlinear Anal. Appl.*, vol. 12, no. November, pp. 2008–6822, 2021, [Online]. Available: <http://dx.doi.org/10.22075/ijnaa.2021.5894>.
- [3] Z. Jie, M. Mahmoud, Q. Stafford-Fraser, P. Robinson, E. Dias, and L. Skrypchuk, "Analysis of yawning behavior in spontaneous expressions of drowsy drivers," *Proc. - 13th IEEE Int. Conf. Autom. Face Gesture Recognition, FG 2018*, pp. 571–576, 2018, doi: 10.1109/FG.2018.00091.
- [4] A. L. A. Ramos, J. C. Erandio, E. M. Enteria, N. Del Carmen, L. J. Enriquez, and D. H. Mangilaya, "Driver Drowsiness Detection Based on Eye Movement and Yawning Using Facial Landmark Analysis," *Int. J. Simul. Syst. Sci. Technol.*, pp. 1–8, 2019, doi: 10.5013/ijssst.a.20.s2.37.
- [5] C. R. Manjunath, K. S. Krishna, K. B. Sree, and K. N. Ramyatha, "Driver Drowsiness Detection Alert System Using Haar Method," *Smart Innov. Syst. Technol.*, vol. 273, no. 7, pp. 380–389, 2022, doi: 10.1007/978-3-030-92905-3_47.
- [6] Chilaka Anvitha, Mohammed Muzammil, Soundaraya. A, and Raghu Kumar, "Driver Drowsiness Detection Using Machine Learning," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 2, no. 2, pp. 314–319, 2022, doi: 10.48175/ijarsct-5456.
- [7] K. Rabiah and M. Khariol, "Driver Drowsiness Detection with an Alarm System using a Webcam," vol. 4, no. 1, pp. 87–96, 2023.
- [8] H. Pranoto, A. Adriansyah, A. Wahab, and D. Feriyanto, "Integrated Speed and Driver Fatigue Detected with Method Image Processing," *Int. J. Adv. Technol. Mech. Mechatronics Mater.*, vol. 1, no. 1, pp. 6–12, 2020, doi: 10.37869/ijatec.v1i1.7.
- [9] A. Araballi and S. Shinde, "Real Time Implementation of Driver Drowsiness Monitoring System Using SVM Classifier," *Int. J. Eng. Manuf.*, vol. 13, no. 3, pp. 48–54, 2023, doi: 10.5815/ijem.2023.03.05.
- [10] X. Chen and B. Zhang, "A XGBoost Algorithm-based Fatigue Recognition Model Using Face Detection," pp. 1–7, 2023.
- [11] P. Pattarapongsin, B. Neupane, J. Vorawan, H. Sutthikulsombat, and T. Horanont, "Real-time Drowsiness and Distraction Detection using Computer Vision and Deep Learning," *ACM Int. Conf. Proceeding Ser.*, 2020, doi: 10.1145/3406601.3406638.
- [12] G. Cao and H. Peng, "The Decision of Fatigue State Based On Characteristics' Fuzzy Fusion," no. Iccia, pp. 295–298, 2016, doi: 10.2991/iccia-16.2016.61.
- [13] Y. Ji, S. Wang, Y. Lu, J. Wei, and Y. Zhao, "Eye and mouth state detection algorithm based on contour feature extraction," *J. Electron. Imaging*, vol. 27, no. 05, p. 1, 2018, doi: 10.1117/1.jei.27.5.051205.
- [14] E. Poornima, R. P. Ram Kumar, S. Katukam, V. V. Pericherla, and S. K. Birelli, "A Real-Time IoT-based Model to Detect and Alert Security Guards'Drowsiness," *E3S Web Conf.*, vol. 391, pp. 1–6, 2023, doi: 10.1051/e3sconf/202339101151.
- [15] C. Dewi, R. C. Chen, X. Jiang, and H. Yu, "Adjusting eye aspect ratio for strong eye blink detection based on facial landmarks,"

PeerJ Comput. Sci., vol. 8, no. 2020, pp. 1–21, 2022, doi: 10.7717/peerj-cs.943.

- [16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Adv. neural Inf. Process. Syst.*, pp. 1097–1105, 2012, doi: 10.1201/9781420010749.
- [17] U. Lahoti, R. Joshi, N. Vyas, K. Deshpande, and S. Jain, “Trends in Computer Drowsiness Detection System for Online Courses,” *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 2, pp. 1930–1934, 2020.
- [18] D. Li, Z. Cui, F. Cao, G. Cui, J. Shen, and Y. Zhang, “Learning State Assessment in Online Education Based on Multiple Facial Features Detection,” *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/3986470.