

Drowsiness Detection Based on Yawning Using Modified Pre-trained Model MobileNetV2 and ResNet50

Hepatika Zidny Ilimadina¹, Muhammad Naufal², Dega Surono Wibowo¹

¹Politeknik Harapan Bersama, Tegal, Indonesia

²Universitas Dian Nuswantoro, Semarang, Indonesia

Article Info

Article history:

Received February 24, 2023

Revised May 03, 2023

Accepted May 30, 2023

Keywords:

Drowsiness detection

Facial-expression change

MobileNetV2

ResNet50

Yawn detection

ABSTRACT

Traffic accidents are fatal events that need special attention. According to research by the National Transportation Safety Committee, 80% of traffic accidents are caused by human error, one of which is tired and drowsy drivers. The brain can interpret the vital fatigue of a drowsy driver sign as yawning. Therefore, yawning detection for preventing drowsy drivers' imprudent can be developed using computer vision. This method is easy to implement and does not affect the driver when handling a vehicle. The research aimed to detect drowsy drivers based on facial expression changes of yawning by combining the Haar Cascade classifier and a modified pre-trained model, MobileNetV2 and ResNet50. Both proposed models accurately detected real-time images using a camera. The analysis showed that the yawning detection model based on the ResNet50 algorithm is more reliable, with the model obtaining 99% of accuracy. Furthermore, ResNet50 demonstrated reproducible outcomes for yawning detection, considering having good training capabilities and overall evaluation results.

Copyright ©2022 The Authors.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Muhammad Naufal, +6282329622754,
Faculty of Computer Science,
Dian Nuswantoro University, Semarang, Indonesia,
Email: m.naufal@dsn.dinus.ac.id

How to Cite:

I. Hepatika Zidny Ilimadina, N. Muhammad, and W. Dega Surono, "Drowsiness Detection Based on Yawning Using Modified Pre-trained Model MobileNetV2 and ResNet50", *MATRIK: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer*, Vol. 22, No. 3, pp. 419-430, Jul, 2023.

This is an open access article under the CC BY-SA license (<https://creativecommons.org/licenses/by-sa/4.0/>)

1. INTRODUCTION

The vital cause of road accidents is human error, with drowsiness due to various fatigue contributing to up to 20% of serious accidents [1]. Unknowingly the driver may experience brief episodes of cessation of responses to stimuli, such as closing their eyes or yawning. This deviation of attention can interfere with the driver's ability to detect or respond to stimuli at a critical event [2]. The main determining microsleep parameter can be detected through the electrocardiogram signal from the drivers. Studies have found changes in brain regions during wakefulness and fatigue. If the driver were tired, the brain area would be more prominent in the parietal lobe in the alpha and beta frequency bands [3]. To conduct this research, it is clear that the cost of the device used is very high. Early detection of tired drivers is also based on specific driving behaviors by Li Wei et al. [4]. An exhausted driver will show the results of steering wheel rotation offset from the road pattern. The selected algorithm works without direct contact with the driver's body, causing the algorithm to be not robust in determining tired drivers due to different driving habits, road conditions, and vehicle models [5]. So that it is possible to develop a fatigue driver detection method based on computer vision. This method is easy to do and does not affect the driver's vehicle handling. The computer vision-based method is carried out in several ways, including collecting the driver's facial expressions [2], detecting the open state of the driver's eyes and mouth, and calculating the closing of the eyelids, blinking, and yawning frequency [1]. Research in this scope requires time and method complexity, so how to increase the performance and speed of the algorithm has become a hot topic and a difficult point in recent years. This study aims to improve the accuracy of microsleep detection techniques that have previously been researched using the driver's blinking signal. The methods include eye aspect ratio, namely by manually setting thresholds and the minimum time for the eyes to close or blink, with several methods such as the Support Vector Machine and Convolutional Neural Network (Vanilla) methods [6–10]. Automatic eye blink detection with a combination of SVM implementation has obtained an accuracy of 99% and high sensitivity of close to 1 [8]. Eye blinking has its weakness in detecting a drowsy driver because the minimum threshold for closing the eyes, usually blinking with drowsiness, is challenging to determine. Humans naturally blink in their daily lives. In a drowsy state, he has some unique facial and body features, and yawning is essential evidence of fatigue. This is the difference between the research from previous studies.

Research in determining drivers' drowsiness based on yawning was carried out using several techniques, namely, the geometric features of the mouth change obviously. This method is to verify the location of the segmented mouth in case of yawning detection. Several studies also carried out representations of facial geometry [11, 12]. The detection of a yawning drowsy driver using geometric features has several areas for improvement; namely, it is difficult to detect the edges of the lips precisely. Recent research on the in-depth study of facial expression representation with MB-LBP and Adaboost classifier was conducted. However, this development is still centered on deepening the facial characteristics of drowsy drivers and machine learning classification. Similar research with a machine and deep learning algorithms can be found in [13, 14, 2]. The novelty of this research aims to build on previous research by producing a transfer learning model which can detect drowsy drivers based on yawning. This research's contribution that differentiates from previous studies is also by combining with the Haar Cascade Classifier to detect the location of the driver's mouth more accurately than in previous studies. The pre-trained models that will be used are MobileNetV2 and ResNet50. These two pre-trained models have been developed in computer vision applications for the scope of object detection properly. This research implemented the Haar Cascade Classifier to initially detect the driver's face and mouth. Then, the detected mouth will be extracted more to classify using the best-proposed method. The concept of transfer learning is carried out using weights from the pre-trained model MobileNetV2 and ResNet50, which is used to identify drivers who yawn. This research still uses some of the layers of the pre-trained model MobileNetV2 and ResNet50 to train a dataset that consists of two classes yawning (2528 images) and non-yawning (2591 images) (license: CC BY-NC-SA 4.0).

2. RESEARCH METHOD

This section explains our proposed method in research, which includes the process of dataset collection, pre-processing, training, testing, and evaluation to the implementation stage in real-time. This research will combine the Haar Cascade Classifier as object detection (driver's mouth) and drowsiness classification using the MobileNetV2 and ResNet50 transfer learning methods. The detailed stages of the research of each step will be further explained in the following Figure 1.



Figure 1. Driver Drowsiness Based on Yawning Detection Process

For image acquisition, this research acquired a dataset consisting of two classes, namely yawning (2528 images) and non-yawning (2591 images). The next step is pre-processing by augmenting the dataset (shear, rotation, flip, zoom) and resizing the dataset to the same size. Before the next stage, the dataset is divided into 60% training data, 20% validation data, and 20% test data. This research trained the dataset using ImageNet weights and trained on modified MobileNetV2 and ResNet50 layers. The new weight generated from the training process is then stored and tested. Finally, the overall evaluation results will be obtained in the testing process, including the accuracy, precision, and recall obtained through the confusion matrix.

2.1. Data Set

This study acquired a dataset consisting of two classes of yawning (2528 images) and non-yawning (2591 images) (license: CC BY-NC-SA 4.0) with varying image sizes [15], as illustrated in Figure 2. The model will later use this dataset to recognize drowsy drivers based on yawning detection. Variations in the dataset include mouth indicating yawning and neither male nor female gender.

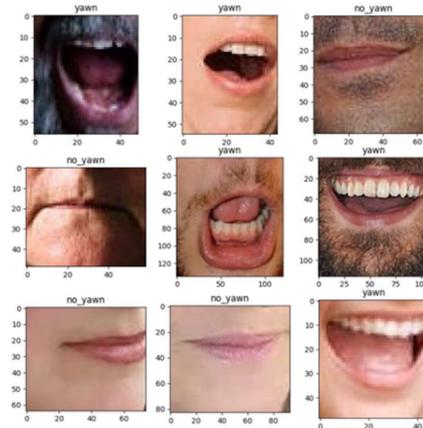


Figure 2. Various Sample of Yawning and Non-yawning Dataset Varying Image Sizes

2.2. Pre-Processing

At this stage, some data preparation is carried out, which will be used in the model training process later. Then, several pre-processing steps were carried out, including resizing the dataset to the same sizesplitting the dataset into 60% training, 20% validation, and 20% testing data. Next, the training data is augmented before studying the model further. Finally, data augmentation was carried out to provide enough model information to study the drowsiness sign feature based on yawning drivers.

2.3. Haar Cascade Classifier

The success of computer vision applications first began in 2001, when Paul Viola and Michael Jones proposed the first Object Detection framework in real-time video detection contained in this research paper first introduced entitled "Rapid Object Detection using a Boosted Cascade of Simple Features" (2001) [16]. This algorithm has a working system that requires positive images and negative images in its classification. Haar cascade can detect a part of an image because it is one of the edge detection methods, but it can be trained to identify almost any object. The cascade classifier generally trains images with four main stages: selecting features with haar, creating integral images, classifying objects with Adaboost training, and detecting objects with a cascading classifier [8, 17].

2.4. MobileNetV2 and ResNet50

The first proposed model in this study is the pre-trained MobileNetV2 [18] with the pre-trained weights of ImageNet loaded from Tensorflow. The original architecture of MobileNetV2 (as shown in Table 1) contains the initial fully convolution layer with 32 filters, followed by 19 residual bottleneck layers described in Table 1. In addition, the architecture includes ReLU6, dropout layers, average pooling layers, and batch normalization.

Table 1. The Original Architecture of MobileNetV2

Input	Operator	t	c	n	s
2242 x 3	conv2d	-	32	1	2
1122 x 32	Bottleneck	1	16	1	1
1122 x 16	Bottleneck	6	24	2	2
562 x 24	Bottleneck	6	32	3	2
282 x 32	Bottleneck	6	64	4	2
142 x 64	Bottleneck	6	96	3	1
142 x 96	Bottleneck	6	160	3	2
72 x 160	Bottleneck	6	320	1	1
72 x 320	Bottleneck	-	1280	1	1
72 x 1280	Bottleneck	-	-	1	-
1 x 1 x 1280	conv2d 1x1	-	k	-	-

The second proposed model is ResNet50 [19] which was formed to defeat quandaries in deep learning training that learning a residual function concerning the input layer was more efficient than learning layer parameters without referring to inputs. This network consists of four residual blocks and has 50 layers implemented a bottleneck technique: firstly, there is a 1 x 1 filter followed by 3 x 3, then followed by a 1 x 1 filter, where the 1 x 1 layers are responsible for reducing and restoring, leaving the 3 x 3 layer a bottleneck with smaller input/output dimensions described in Table 2. In addition, this architecture includes max and average pooling, fully connected, and softmax layers.

Table 2. The Original Architecture of The ResNet50

Layer	Output size	50-layer
conv1	112x112	7x7, 64, stride 3 3x3 max pool, stride 2
conv2.x	56x56	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1 \end{bmatrix} \times 3$
Conv3.x	28x28	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$
Conv4.x	14x14	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 4$
Conv5.x	7x7	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 4$
	1x1	Average pool, 1000-fc, softmax
FLOPs		3,8x10 ⁹

The base layers from MobileNetV2 and ResNet50 are frozen and eliminated to prevent the loss of the previously learned feature from ImageNet. This research eliminates the last three layers of the original MobileNetV2 and ResNet50, then adds three more trainable layers to the network in both last layers, as illustrated in Figure 3.

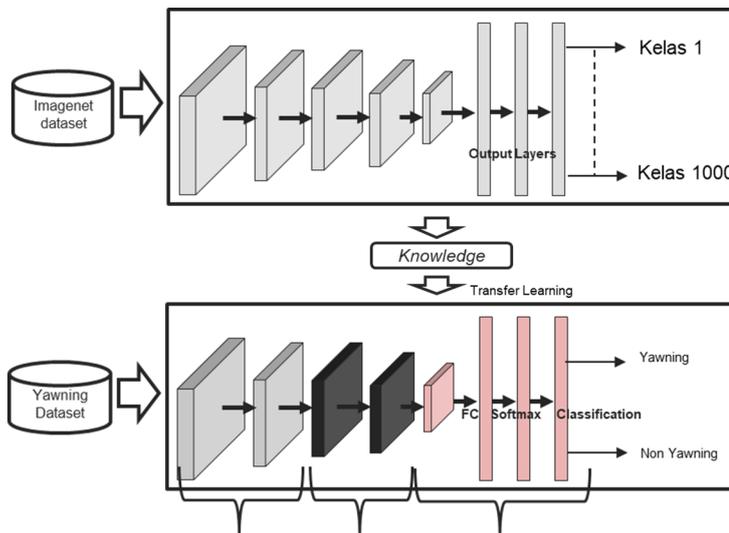


Figure 3. Illustration of the Implementation of the Pre-trained Model

Figure 3 illustrates the transfer learning process by cutting the original architecture’s last three layers and replacing them with customized layers. The three layers added here are the dropout layer, fully connected layers, and a final dense layer. The fully connected and final dense layers contain 128 and 2 neurons, respectively. In addition, the activation function softmax was used in the final layer. The architecture and parameters used in this study are described in Table 3 and Table 4.

Table 3. The Architecture of MobileNetV2 in Research

Layer (type)	Output Shape	Param
mobilenetv2_1.00_128_input (InputLayer)	[(None, 128, 128, 3)]	0
mobilenetv2_1.00_128 (Functional)	(None, 1280)	2257984
flatten_8 (Flatten)	(None, 1280)	0
dense_16 (Dense)	(None, 128)	163968
dropout_8 (Dropout)	(None, 128)	0
dense_17 (Dense)	(None, 2)	258
Total params: 2,422,210		
Trainable params: 2,385,538		
Non-trainable params: 36,672		

Table 4. The Architecture of ResNet50 in Research

Layer (type)	Output Shape	Param
resnet50_input (InputLayer)	[(None, 128, 128, 3)]	0
resnet50 (Functional)	(None, 2048)	23587712
flatten (Flatten)	(None, 2048)	0
dense (Dense)	(None, 128)	262272
dropout (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 2)	258
Total params: 23,850,242		
Trainable params: 23,797,122		
Non-trainable params: 53,120		

This research trains these newly added layers on the selected datasets so that the model can fix the features to detect yawning drivers. The use of the MobileNetV2 architecture is because this model has been developed a lot in the implementation stage of localization (position) of object detection using a device, as has been done in previous research on palmprint recognition and face-mask position recognition [20–22]. ResNet50 is also more widely used for image classification than other models and has shown great results in computer vision applications, such as high-resolution optical object detection [23–25].

2.5. Evaluation

In this study, the measurement of the performance evaluation of the model using a confusion matrix to measure accuracy and sensitivity (recall) [8]. Accuracy is used to understand how precisely the model is in predicting/detecting yawning and not yawning drivers correctly, as seen in Equation (1). The sensitivity value is used to measure how precise the model is in detecting drowsy drivers, indicated by yawning, as explained in Equation (2). The sensitivity value is essential in detecting false negatives that will have a dangerous impact if they occur. If a drowsy driver (yawning condition) is not detected, it will be dangerous if the warning does not appear later. Where TP, FP, TN, and FN are the number of classified cases of true positives, false positives, true negatives, and false negatives, respectively.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

2.6. Real-Time Testing

Along with evaluating the model, this research also saves the best models weight to build a detection system. It will be built to detect drowsy drivers based on yawning features in real-time. For this condition, this research uses the implementation of the Haar Cascade Classifier in performing mouth detection. When the mouth is well detected, the model predicts whether the class is yawning or not.

3. RESULTS AND ANALYSIS

The experiments were performed on 2.0GHz Intel Core i5 MacBook Pro with 16GB memory using Python with Jupyter Notebook. Implementation using Python and Keras library with a Tensorflow backend and also uses supported libraries such as NumPy, sklearn, matplotlib, and pandas. Resizing the dataset to 224x224 was applied in pre-processing. This research also split the dataset into 60% training, 20% validation, and 20% test data.

3.1. Implementation of Pre-Trained Model MobileNetV2 and ResNet50 in Yawning Detection

Implementing the MobileNetV2 and ResNet50 pre-trained model begins with fine-tuning to take the weight of the previously trained network by adding modified layers. This research has empirically added three new layers: flatten, dropout, dense, and fully connected layers on both based models. Moreover, our experiment also increases batch size to 32 to avoid overfitting the pre-trained model. In implementing the modified pre-trained models, this research also uses the adam optimizer with a learning rate of 0.0001, categorical cross-entropy as the lost function, and the activation function using softmax. This research gets performance evaluations within 100 epochs with early stopping patience of 10 epochs for both proposed methods. The overall training, validation, and testing processes take approx. 2-6 hours. Before the model is implemented in real-time, this research conducts a performance evaluation using a confusion matrix, and Table 5 shows the overall performance of our proposed model.

Table 5. Overall Performance Evaluation of Our Proposed Methods

Method	Class	Recall	Accuracy	Time Computation
Baseline 1 yawn [11]	yawning (cold voxels)	-	71%	-
	yawning (hot voxels)	-	87%	-
Baseline 3 yawn [13]	yawning	-	95%	-
Baseline 2 yawn [2]	yawning	-	96,5%	-
MobileNetV2	0 (non yawn)	99%	98%	02 hrs 10 mins
	1 (yawn)	97%		
ResNet50	0 (non yawn)	99%	99%	06 hrs 38 mins
	1 (yawn)	99%		

The fine-tuning pre-trained MobileNetV2 model achieves 98%, and ResNet50 obtained 99% test accuracy, better than the baseline research method of machine learning and deep learning [11, 13, 2]. Each proposed model performs well because of the stability of the evaluation results during training, validation, and testing. The best model's weight will then be stored to be combined with the Haar Cascade Classifier in detecting yawning as a sign of drowsiness.

3.2. Real-Time Testing

The performance of the best-proposed model was evaluated with the implementation of practical applications in a natural environment. So, this research combines the pre-trained model ResNet50 and the Haar Cascade Classifier. When our system captures video from the camera, as shown in Figure 4, it predicts the mouth position and is marked by a bounding box around the face, along with the predicted class and the confidence score.

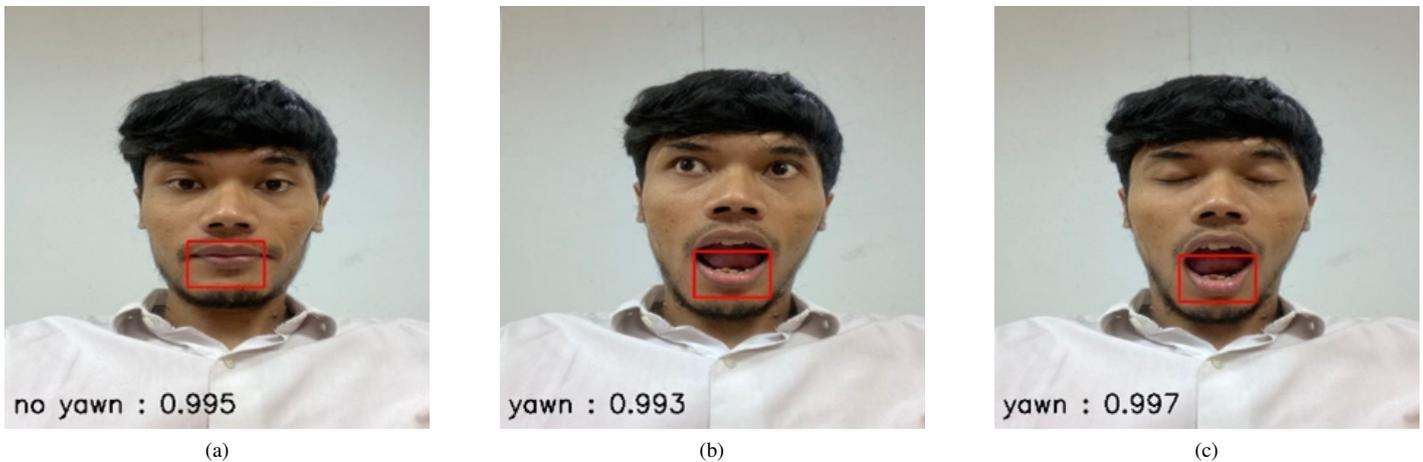


Figure 4. Implementation of Facial Expression Change of Yawning on Drivers. From Left to Right: (a) Non-yawning Condition; (b) Yawning Condition with Awake Eyes; (c) The Condition of Yawning and Closing Eyes

In Figure 4, this research can see the results of implementing the proposed model in several driver conditions. First, when the driver's eyes are awake and not yawning, they will not be detected, as shown in Figure (a). Then in Figure (b), this research sees that the system will still detect this condition with the yawning state even though the driver is awake. Finally, in condition (c), the system can also detect facial change expressions of yawning when the driver closes his eyes.

3.3. Analysis and Discussion

In building this drowsiness detection system, supporting datasets are collected in images labeled not yawning and yawning. The dataset contains 2528 yawning class pictures and 2591 non-yawning pictures. This research uses the MobileNetV2 architecture by taking the weight from ImageNet for the training process. In this research, our proposed model will modify the last three layers with dropout, fully connected, and the last dense layer. The following process is to conduct training so the model can adequately classify sleep drivers with yawning symptoms. ResNet50 achieved the best model for drowsiness detection based on yawning conditions. This widely used model performs better in image classification than other pre-trained models [11, 13, 14]. The advantage of the pre-trained model ResNet50 is that performance does not decrease even though the architecture is modified deeper. ResNet50 can perform computations carefully by being made lighter and has good training capabilities [23–25]. This condition is shown by the computational training time on ResNet50, in which the training time is longer than the modified pre-trained MobileNetV2 model. Through accuracy in the computational process, ResNet50 can produce perfect performance evaluations, as shown in Table 3. MobileNetV2 model is also proposed in this study because it is an efficient convolutional neural network for modern object detection systems with trainable parameters and can detect objects properly. MobileNet architecture is already tiny and has low latency but is rich in features, so it supports applications that require models to be smaller and faster [20–22, 26]. Besides being fast and small, the MobileNetV2 structure is built to detect objects with training time faster (as shown in Table 2). MobileNetV2 builds on depthwise separable convolutions and is an excellent network for training images. This research answers how applications can work well but have a small computational cost. Compared to other pre-trained models, the two proposed methods can detect with good accuracy and recall, with a lighter architecture and low latency. This is a development from baseline research which only uses thermal images to detect yawning drivers [11]. In this detection, annotations are done manually for the entire training and testing video sequence. The lack of this detection occurs primarily when a person is not fully visible in the camera, i.e., only partially or is covered in hair. The results presented in Table 5 also show that the proposed method has better results than previous studies using feature and thermal images, as well as machine learning algorithms. In this detection, annotations are done manually for the entire training and testing video sequence. The yawning detection by machine learning algorithms has improved detection using only thermal images [13, 2]. The lack of this detection occurs primarily when a person is not fully visible in the camera, i.e., only partially or is covered in hair. Based on the results presented in Table 5 also shows that the proposed method has better results than previous studies using feature and thermal images, as well as machine learning algorithms. The yawning detection by machine learning algorithms has improved detection was using only thermal images. The machine learning algorithm has detected it well, but it still has an error value of 3-5%

in the testing process. Whereas in the case of detecting a drowsy driver, it requires detection results with zero tolerance (close to 0% error rate).

The proposed method based on neither the ResNet50 nor MobileNetV2 architecture is proven to detect and predict drowsy drivers correctly. This research used a test phase with validation and test data for the evaluation process, where each evaluation value obtained an accuracy of 99%. This research used the confusion matrix to evaluate the performance model. The essential evaluation values in this model are accuracy and sensitivity (recall). Accuracy is used to understand how precisely the model is in predicting/detecting yawning and not yawning drivers correctly. Moreover, the recall value is used to measure how precise the model is in detecting the real drowsy drivers based on yawning. The recall value is vital in detecting false negatives that will have a dangerous impact if they occur. If a real drowsy driver (yawning condition) is not detected, it will be dangerous if the warning does not appear later. In Table 5, the results of evaluating the best model's ability to detect the actual label correctly are 99% in the non-yawning state and 99% in the yawning state. In the case of detecting a drowsy driver, the sensitivity value should be close to 1 (zero tolerance), which means that the model should be able to detect a situation where a driver is drowsy correctly. The high recall rate is due to the effect of the combination of the Haar Cascade Classifier. Figure 4 shows that the model can detect the exact location of the mouth, which is then classified as yawning or not. Haar cascade proved to be more accurate in detecting objects as a whole in one video frame with the Adaboost classifier, which was previously trained with positive and negative images [8]. Referring to these results, the combination of Haar Cascade Classifier and ResNet50 is good at dealing with false negatives because the potential for error is minimal (only 1%). Another advantage of the ResNet50 classifier is its complex architecture but trainable lighter parameters. Therefore, this method can be implemented in devices because it only requires a small cost to build. In addition, this proposed method accurately detects drowsy drivers based on signs of yawning, concluded from 100 events; the model cannot detect one drowsy driver.

4. CONCLUSION

A scheme for drowsy driver detection systems based on facial changes of yawning has been proposed. The proposed model consists of modified pre-trained MobileNetV2 and ResNet50. Both proposed models have been tested using a confusion matrix evaluation, and it can give us ResNet50 as the best model with 99% accuracy for test datasets. This research uses a combination of ResNet50 as a classifier and Haar Cascade for real-time detection. This research improved from previous research that only monitored faces and measured based on geometric features. ResNet50 can achieve a high recall value of 99% for a system that requires zero tolerance. The weakness of ResNet50 is the long training time, even though it takes only a short time than the other pre-trained methods. In addition, the lack of this study is that there is still a proportion of detection errors of 1% for zero-tolerance devices. The multimodal classification should overcome this. Therefore, further research intends to add a feature selection method to reduce computation. In addition, other parameters such as electrocardiogram (ECG) signal information can be developed to make an embedded system with highly accurate slowness driver detection.

5. ACKNOWLEDGEMENTS

We want to thank Harapan Bersama Polytechnic, Tegal, Central Java, Indonesia, for providing excellent cooperation and support.

6. DECLARATIONS

AUTHOR CONTRIBUTION

Hepatika Zidny Ilmadina: Conceptualization, Methodology, Validation, Formal analysis, Resources, Writing original draft, Writing - review & editing, Visualization. **Muhammad Naufal:** Conceptualization, Methodology, Software, Validation, Resources, Data curation. **Dega Surono Wibowo:** Software, Formal analysis, Investigation, Data curation.

FUNDING STATEMENT

This research was supported by Harapan Bersama Polytechnic, Tegal, Central Java, Indonesia. (Research Grant Agreement number 050.16/P3MPHB/XI/2022, dated November 18th 2022)

COMPETING INTEREST

The authors declare that we have no known competing financial interests or personal relationships that could have appeared to influence research in this paper.

REFERENCES

- [1] L. Salvati, M. D' amore, A. Fiorentino, A. Pellegrino, P. Sena, and F. Villecco, "On-Road Detection of Driver Fatigue and Drowsiness during Medium-Distance Journeys," *Entropy*, vol. 23, no. 2, pp. 1–12, feb 2021.
- [2] Z. Liu, Y. Peng, and W. Hu, "Driver Fatigue Detection Based on Deeply-Learned Facial Expression Representation," *Journal of Visual Communication and Image Representation*, vol. 71, p. 102723, aug 2020.
- [3] Y. Ma, B. Chen, R. Li, C. Wang, J. Wang, Q. She, Z. Luo, and Y. Zhang, "Driving Fatigue Detection from EEG Using a Modified PCANet Method," *Computational Intelligence and Neuroscience*, vol. 2019, pp. 1–9, jul 2019.
- [4] Z. Liu, W. Wei, H. Wang, Y. Zhang, Q. Zhang, and S. Li, "Intrusion Detection Based on Parallel Intelligent Optimization Feature Extraction and Distributed Fuzzy Clustering in WSNs," *IEEE Access*, vol. 6, no. November, pp. 72 201–72 211, 2018.
- [5] J. Xi, S. Wang, T. Ding, J. Tian, H. Shao, and X. Miao, "Detection Model on Fatigue Driving Behaviors Based on the Operating Parameters of Freight Vehicles," *Applied Sciences*, vol. 11, no. 15, pp. 1–16, aug 2021.
- [6] C. B. S. Maior, M. J. d. C. Moura, J. M. M. Santana, and I. D. Lins, "Real-time classification for autonomous drowsiness detection using eye aspect ratio," *Expert Systems with Applications*, vol. 158, p. 113505, nov 2020.
- [7] C. Ryan, B. O'Sullivan, A. Elrasad, A. Cahill, J. Lemley, P. Kieilty, C. Posch, and E. Perot, "Real-Time Face & Eye Tracking and Blink Detection Using Event Cameras," *Neural Networks*, vol. 141, no. 2021, pp. 87–97, 2021.
- [8] H. Zidny Ilmadina, D. Apriliani, and D. S. Wibowo, "Deteksi Pengendara Mengantuk dengan Kombinasi Haar Cascade Classifier dan Support Vector Machine," *Jurnal Informatika : Jurnal Pengembangan IT (JPIT)*, vol. 7, no. 1, pp. 1–7, 2022.
- [9] C. Jacobé de Naurois, C. Bourdin, A. Stratulat, E. Diaz, and J. L. Vercher, "Detection and Prediction of Driver Drowsiness Using Artificial Neural Network Models," *Accident Analysis and Prevention*, vol. 126, no. July, pp. 95–104, 2019.
- [10] S. Rajamohana, E. Radhika, S. Priya, and S. Sangeetha, "Driver drowsiness detection system using hybrid approach of convolutional neural network and bidirectional long short term memory (CNN_BILSTM)," *Materials Today: Proceedings*, vol. 45, no. 2, pp. 2897–2901, 2021.
- [11] M. Knapik and B. Cyganek, "Driver's fatigue recognition based on yawn detection in thermal images," *Neurocomputing*, vol. 338, pp. 274–292, apr 2019.
- [12] M. Kim and J. Koo, "Embedded System Performance Analysis for Implementing a Portable Drowsiness Detection System for Drivers," *Technologies*, vol. 11, no. 1, p. 124, dec 2022.
- [13] K. Chen, T. Zhu, S. Li, and Y. Shi, "Real-time Yawning Detection Based on Machine Learning Algorithm and Time Series Classification using Facial Feature Points," in *2021 International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS)*. IEEE, dec 2021, pp. 276–280.
- [14] A. Rahman, M. B. H. Hriday, and R. Khan, "Computer Vision-Based Approach to Detect Fatigue Driving and Face Mask for Edge Computing Device," *Heliyon*, vol. 8, no. 10, pp. 1–16, oct 2022.
- [15] D. Vazquez, "Yawn Dataset — Kaggle." [Online]. Available: <https://www.kaggle.com/datasets/davidvazquezcic/yawn-dataset>
- [16] R. Raj, P. Rajiv, P. Kumar, M. Khari, E. Verdú, R. G. Crespo, and G. Manogaran, "Feature based video stabilization based on boosted HAAR Cascade and representative point matching algorithm," *Image and Vision Computing*, vol. 101, p. 103957, sep 2020.
- [17] D. Yang, B. Peng, Z. Al-Huda, A. Malik, and D. Zhai, "An Overview of Edge and Object Contour Detection," *Neurocomputing*, vol. 488, no. June, pp. 470–493, jun 2022.
- [18] M. Sandler, A. Howard, M. Zhu, and A. Zhmoginov, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 4510–4520.

- [19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
- [20] A. Michele, V. Colin, and D. D. Santika, "MobileNet Convolutional Neural Networks and Support Vector Machines for Palm-print Recognition," *Procedia Computer Science*, vol. 157, pp. 110–117, 2019.
- [21] M. H. Rahman, M. K. A. Jannat, M. S. Islam, G. Grossi, S. Bursic, and M. Aktaruzzaman, "Real-time face mask position recognition system based on MobileNet model," *Smart Health*, vol. 28, no. June, pp. 1–12, jun 2023.
- [22] I. Shafi, A. Mazahir, A. Fatima, and I. Ashraf, "Internal Defects Detection and Classification in Hollow Cylindrical Surfaces Using Single Shot Detection and MobileNet," *Measurement*, vol. 202, p. 111836, oct 2022.
- [23] W. Qi, "Object detection in high resolution optical image based on deep learning technique," *Natural Hazards Research*, vol. 2, no. 4, pp. 384–392, dec 2022.
- [24] D. Sarwinda, R. H. Paradisa, A. Bustamam, and P. Anggia, "Deep Learning in Image Classification using Residual Network (ResNet) Variants for Detection of Colorectal Cancer," *Procedia Computer Science*, vol. 179, pp. 423–431, 2021.
- [25] J. Chai, H. Zeng, A. Li, and E. W. Ngai, "Deep Learning in Computer Vision: A Critical Review of Emerging Techniques and Application Scenarios," *Machine Learning with Applications*, vol. 6, p. 100134, dec 2021.
- [26] C. H. Karadal, M. C. Kaya, T. Tuncer, S. Dogan, and U. R. Acharya, "Automated classification of remote sensing images using multileveled MobileNetV2 and DWT techniques," *Expert Systems with Applications*, vol. 185, pp. 1–12, dec 2021.

[This page intentionally left blank.]