COMMUNICATIONS IN
SCIENCE AND TECHNOLOGY
Homepage: cst.kipmi.or.id

Evaluating the effectiveness of facial actions features for the early detection of driver drowsiness in driving safety monitoring system

Yenny Rahmawati[a], Kuntpong Woraratpanya[b], Igi Ardiyanto[a], Hanung Adi Nugroho*[a]

[a]*Department of Electrical and Information Engineering, Universitas Gadjah Mada, Indonesia*
[b]*Faculty of Information Technology, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand*

Article history:
Received: 2 December 2024 / Received in revised form: 15 June 2025 / Accepted: 17 June 2025

Abstract

Traffic accidents caused by drowsiness remain a serious threat to driving safety. Many of these accidents can actually be prevented with an early warning system that detects the early signs of driver drowsiness. This study proposes a non-invasive system to detect drowsiness based on visual features extracted from videos recorded by a dashboard camera. The system uses facial landmarks generated by a facial network detector to identify key areas such as eyes, mouth, and head. The eye aspect ratio (EAR), mouth aspect ratio (MAR), and head rotation angle were calculated as the main features. These features were fed into three classification models: 1D-CNN, LSTM, and BiLSTM. Evaluation was conducted using 87 videos from the YawDD dataset for training and 20 videos from custom data for testing. During training, the 5-fold cross-validation was used to ensure model generalization and reduce the risk of overfitting. In addition to accuracy, other metrics such as precision, recall, and F1-score were used to provide a more comprehensive overview of the system performance. The results showed that the combination of the three facial features (EAR, MAR, and head rotation) provided a better performance than did the use of a single feature or a combination of two features, with an accuracy improvement of 5–8%. The BiLSTM model showed the best performance, with a training accuracy of 99% on the YawDD dataset and a testing accuracy of 98% on the custom data.

Keywords: Drowsiness detection; facial action features; EAR; MAR; head pose; 1DCNN; LSTM; BiLSTM

1. Introduction

Driver drowsiness detection (DDD) is an important research area with a major impact on public safety, particularly in transport and healthcare sectors [1]. Real-time DDD systems are capable of providing early warning to drivers, helping to reduce the potential for accidents. Transportation plays a crucial role in human life, and a significant portion of a country's economy is based on this industry. Despite its contribution to safe and efficient travel, driver inattention, fatigue, and drowsiness can cause personal injury [2]. Driver fatigue is a leading cause of many accidents worldwide. Various studies indicated that 20-50% of accidents are caused by fatigue and drowsiness of drivers on certain roads [3]. Drowsiness occurs when a person feels dizzy or falls asleep involuntarily, frequently due to the lack of sleep or mental and physical exhaustion. This is particularly dangerous in situations that require high alertness, such as industrial work, mining, and driving, to avoid potentially life-threatening events [4].

Various methods for DDD have been developed, including physiological, behavioral, and visual approaches. Physiological approaches involve monitoring changes in heart rate (ECG) [5], brain activity (EEG) [6], as well as skin conductance (EMG) [7]. However, these methods require the use of equipment with direct physical contact. The trends of recent research have increasingly focused on developing more reliable and non-invasive DDD techniques.

Behavior-based approaches focus mainly on the analysis of driving patterns, such as lane deviation [8] and steering wheel movements. Meanwhile, visual methods that use facial landmark analysis have emerged as a promising solution due to their effectiveness and convenience. This approach uses the monitoring of eye movements and facial expressions to identify the signs of driver drowsiness.

In recent years, facial landmark-based DDD is an active research area in computer vision and has attracted much attention due to its effective and non-invasive nature. It involves the identification of facial landmarks and the analysis of their move-

*Corresponding author. Telp.: +62-274-552305
Email: adinugroho@ugm.ac.id
https://doi.org10.21924/cst.10.1.2025.1594

ments to recog- nize the signs of fatigue in drivers. Researchers in the field of visual feature analysis have conducted various in-depth studies related to this method.

Panda *et al,*[9], discussed a method for detecting drowsiness in drivers using facial features and hand gestures. The system is based on cameras installed in the car to analyze the driver's eyes, mouth, and hand movements. This paper proposed a method integrating the Eye Aspect Ratio (EAR), the distance between the upper and lower lips to detect yawning, and the detection of hands covering the mouth when yawning. When detectings drowsiness or yawning, the system will gives a voice alert to the driver. Nevertheless, the weakness of this paper is that the dataset used to detect yawning with hands covering the mouth is self-generated and consists of only 2300 images for the yawning class and 2500 images for the normal condition.

Meenatchi *et al,* [10], in their, paper discussed a drowsiness de- tection system for drivers based on eye blink rate monitoring using computer vision technology. The system uses a webcam installed on the vehicle dashboard to detect eyes and monitor facial expressions. If the driver's eyes remain closed for a certain period, the system will activate a warning alarm to wake up the driver. The drawback of this paper is that it only relies on eye blink (EAR) without other factors., In addition, the paper did not compare the results of this system with other approaches, such as more sophisticated deep learning methods (CNN, LSTM).

Moa *et al,* [11], discussed the development of a driver drowsiness detection system using a convolutional neural network (CNN). The system works by analyzing eyes and mouth conditions as the main indicators of sleepiness. The developed CNN model was tested with a dataset containing the images of opened eyes, closed eyes, yawning mouth, and normal mouth, and managed to achieve 97.23% accuracy. The drawback of this research is related to the relatively minimal dataset used.

Osmani and Wawage [12], discussed a realtime driver drowsiness detection system using Vision Trans- former (ViT). The system focuses on the analysis of eyes' condition (open/closed) as a key indicator of drowsiness. The Vision Transformer model was trained with a large dataset of 84,900 eye images, achieving 98.8% accuracy, higher than previous CNN-based methods. However, the system is still limited to eye detection only.

Modi *et al,* [13], proposed a CNN-LSTM hybrid model for a real-time driver drowsiness detection. The model combines facial (EAR, MAR) and behavioral (head movement, eye blink, and yawn) feature analysis to improve the accuracy of drowsiness detection. This method is superior to CNN-based models alone, as CNN captures the spatial features of the face, while LSTM analyses the temporal patterns of changes in facial expressions and head movements. The accuracy of the model reached 98.89%. However, the model still needs to be tested in real-world conditions and can be improved with the integration of additional sensors or optimization for devices with limited computing power.

Image-based approaches are preferred for not interferinge with the driver's activities. Previous studies have focused on image-based methods, including yawning frequency [14], eye closure frequency [15], head bobbing [16], eyelid closure percentage [17], eye aspect ratio (EAR) [18], and head movement analysis [19]. Feature fusion combinations have also been explored [20].

Driver actions play a critical role in maintaining road safety, for both driver and other road users. Given its importance in saving lives, the detection of driver drowsiness has recently gained an increasing attention. Various studies in the literature have explored methods for detecting driver alertness using facial features such as head pose, eyes movements, and other facial expressions [21], [22]. Though these studies showed promising progress, the key challenges such as ensuring accurate and real-time detection of drowsiness remain unsolved still. Drowsy driving leads to millions of accidents, injuries, and fatalities each year, underscoring a need for systems with the high levels of accuracy and precision. The effective detection of drowsiness is deemed essential for preventing road accidents. This study aims to determine whether using a combination of facial features eyes movements, mouth movements, and head pose is more effective for detecting driver drowsiness rather than using only one or two of these features.

This study aims to analyze the effectiveness of combining various facial features in improving the accuracy of driver drowsiness detection. The main contributions of this research include improving the detection method by considering the limitations of previous studies, which generally only use one facial feature and face constraints in the availability and diversity of datasets:

a) This research proposes a new approach by combining three facial features eyes movement, mouth movement, and head position to improve the accuracy and effectiveness of driver drowsiness detection. In addition, this research evaluated the performance of the combination of three features by comparing them against the use of one or two facial features.

b) The dataset used in training were taken from the YawDD public dataset, while testing was conducted using a dataset collected by the authors themselves to overcome the limitations of the available dataset.

c) Feature evaluation was conducted by applying three deep learning models 1D-CNN, LSTM, and BiLSTM to assess the effectiveness of the proposed feature combination in accurately detecting drowsiness.

2. Materials and Method

This research proposes a study of drowsiness detection using an artificial neural network model to detect specific facial features from images. Fig. 1 provides an overview of the proposed approach. The study consists of five main processes: data collection, preprocessing, feature extraction, feature selection, and classification. It used two data sources: the public YawDD dataset and the custom dataset. The YawDD dataset consisteds of videos of drivers displaying various facial expressions while driving, including when being drowsy. The custom dataset were collected in a controlled manner involving 20 subjects under normal lighting conditions inside a stationary vehicle. Each video was approximately 2 minutes long, recorded at a resolution of 640×480 pixels at 30 frames per second (fps) with help of a standard dashboard camera.

Data preprocessing was performed by extracting video frames every 7 frames (equivalent to 4.3 fps) to reduce redundancy and computational load. From each frame, facial landmarks were extracted using a MediaPipe-based facial detection network. Three main features were calculated: eye aspect ratio (EAR), mouth aspect ratio (MAR), and head rotation based on Euler angles (roll, pitch, yaw) derived from the nose landmark position.

The thresholds for EAR and MAR were adopted from previous research [23], namely EAR <0.2 to detect closed eyes, MAR >0.5 to detect yawning, and MAR <0.1 to detect sleeping with the closed mouth. To ensure suitability for the custom dataset, an analysis of the distribution of EAR and MAR values was firstly conducted, and these thresholds were retained as they showed consistency with the sleepy behavior as observed visually in the custom data.

The training data consisteds of 22,500 samples, comprising 8,200 ‘sleepy’ labels and 14,300 ‘not sleepy’ labels, indicating the class imbalance. To address this, balancing techniques were applied under the sampling of the majority class, and additional experiments were conducted with class weighting during the training process. No data augmentation was performed as all features used were numerical extractions, not raw images.

The three features (EAR, MAR, and head rotation) were incorpo-rated into three different model architectures: 1D-CNN, LSTM, and BiLSTM. The models were trained using the 5-fold cross- validation to ensure generalization and avoid overfitting. Meanwhile, model performance was evaluated using accuracy, precision, recall, and F1-score metrics.

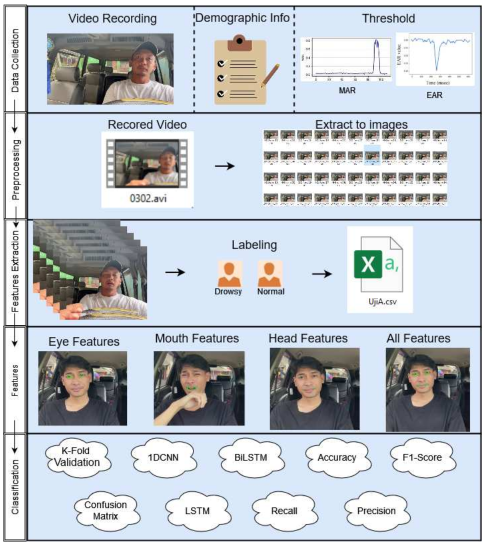


Fig. 1. An overview of the driver drowsiness detection process

2.1. Data Collection

Two data sets were used for the experiments: (i) YawDD dataset [24]: A publicly available dataset containing videos of drivers showing various indicators of drowsiness, such as yawning, eyes closure, and head movements. (ii) Customized data set: Additional datasets collected specifically for testing and validation. The data included various driver behaviors, including yawning, eye closure, and head bowing. The thresholds were then set to classify the driver’s status on the basis of these observed behaviors. For data preparation, we split the data set into training and testing sets. The model was trained on 80% of the combined dataset, with features extracted from each frame (EAR, MAR, head pose) used as input. Standard data augmentation techniques were applied to reduce overfitting. The remaining 20% of the data set was used for testing. The custom dataset served as an additional validation set to ensure the model generalized well to unseen data.

2.2. Preprocessing

In preprocessing stage, videos were extracted into images. This extraction process generated individual frames from the video footage, allowing for the separatione analysis of each frame. By isolating these frames, specific details, such as changes in facial expressions or head movements, could be identified and processed, which are crucial for detecting and classifying the driver’s condition.

This process useds facial landmark points to recognize the driver’s face. Fig. 2a shows the facial geometry solution used to estimate 468 landmark points in three dimensions as the, instrument in detecting the eyes, mouth, and nose. This solution determined the location of 68 key points on the face, forming a map representing the main structure of the face. Fig. 2b shows the analysis of the eye structure, while Fig. 2c illustrates the analysis of the mouth structure. This method wasis applied to detect and extract the areas of eyes and mouth.

In the estimation of head pose, a MediaPipe face mesh solution [25] was used, which also predicted 468 facial landmark points in three dimensions. The X and Y coordinates of the face mesh solution were normalized based on the frame size, while the Z coordinate representeds the depth of the face mesh, reflecting the distance of the head from the camera. To determine the pose of the head in the video, the initial coordinates of the nose were firstly extracted and used as a reference to determine the position and movement of the head in subsequent frames.

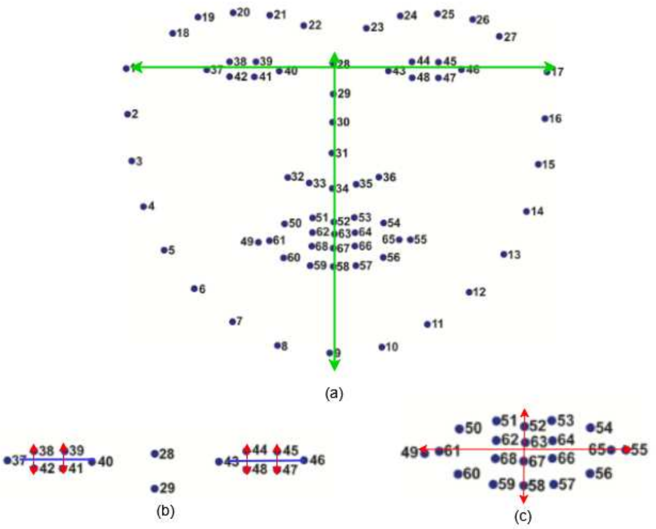


Fig. 2. Driver drowsiness detection analysis (a) mean face, (b) eyes analysis, and (c) mouth analysis

2.3. Feature Extraction

Various human and vehicle features have been used to model different drowsiness detection systems. However, this study focused on modeling drowsiness detection using EAR and MAR metrics, along with the estimation of head pose.

2.3.1. Eye Aspect Ratio (EAR)

Rosebrock [15] stated that eye blink detection using EAR features has a number of significant advantages over traditional image processing-based detection methods. In traditional methods, the process begins with the eye localization, followed by applying a threshold to identify the white part of the eye in the image. An eye blink is then detected based on the loss of that white area. In contrast, the use of the Eye Aspect Ratio (EAR) metric require no complex image processing, thereby reducing memory usage and speeding up computation time. EAR is calculated based on the ratio of the distance between facial landmarks around the eye area; this then makes, it a simple yet efficient method for visual analysis. Commonly, the EAR metric calculates the horizontal and vertical distance ratios of six coordinate points of the eye markers, as shown in Fig. 3. These coordinates are numbered, started from the left eye corner at p1 and rotating clockwise to p6. The six coordinate points from p1 to p6 are two-dimensional. When the eyes are open, the Eye Aspect Ratio (EAR) value tends to be stable. However, when the eyes are closed, the difference in position between points p3 and p5, as well as p2 and p6, becomes insignificant, causing the EAR value to drop dramatically to near zero.

To extract the Eye Aspect Ratio (EAR) feature, Equation (1) was used. In this equation, the numerator calculates the Euclidean distance between the vertical eye landmark points, while the denominator calculates the distance between the horizontal landmarks,] then multiplied by two to maintain the ratio balance. The EAR value was calculated for each frame in the video and then stored in a list for further analysis.

$$EAR = \frac{|p_2 - p_6| + |p_3 - p_5|}{2 \times |p_1 - p_4|} \tag{1}$$

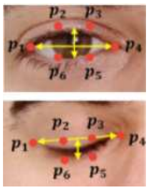


Fig. 3. Variation of the EAR value over time

2.3.2. Mouth Aspect Ratio (MAR)

Similar to EAR, the MAR is used to measure the degree of openness of the mouth. The mouth is represented by 20 coordinates on the facial markers (points 49 to 68). However, in this study, we focused on points 61 to 68, as shown in Fig. 4, to determine the degree of mouth openness. Using these coordinates, the distance between the upper and lower lips is calculated using Formula (2) to assess whether the mouth is open [26]. In Formula (2), the numerator represents the calculation of the vertical distance between specific coordinates, while the denominator calculates the horizontal distance. As in Formula (1), the value in the denominator is multiplied by two to maintain proportional balance in the ratio calculation.

$$MAR = \frac{|p_{64} - p_{68}| + |p_{63} - p_{67}| + |p_{62} - p_{66}|}{3 \times |p_{61} - p_{65}|} \tag{2}$$

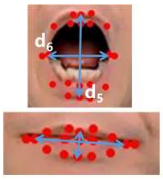


Fig. 4. Variation of the MAR value over time

2.3.3. Determination of EAR and MAR threshold values

The determination of the threshold values for the eye and mouth boundaries is based on the literature in [23]. To evaluate the effectiveness of the approach proposed in that study, experiments were conducted on three different scenarios with various difficulty levels, from the easiest to the most challenging. In this analysis, the MAR and EAR metrics were used as defined in Equations (1) and (2) to assess the level of driver drowsiness.

The first experiment was conducted using short videos recording various individuals showing the signs of drowsiness in a controlled laboratory environment. As shown in Fig. 5, an eye closure event was represented by the EAR curve, where the EAR value droppeds below the threshold of 0.2. Similarly, yawning events were seen in the MAR curve when the value exceededs the threshold of 0.5. In addition, MAR values below the threshold of -0.1 indicated an asleep state. Interestingly, two troughs were observed in the curve corresponding to the moment when a person falls asleep with their head down. This indicates that the MAR design effectively considers extreme head postures, including the asleep state.

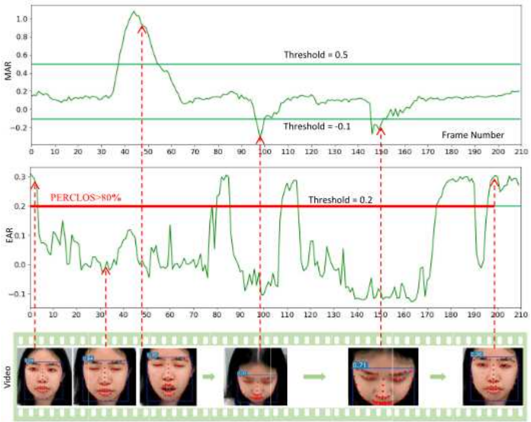


Fig. 5. EAR and MAR threshold values [23]

2.3.4. Head Pose

In this study, the head rotation angle was used to estimate the orientation or position of the head in each frame. The rotation angle refers to the amount of rotation of an object relative to a fixed point, known as the rotation point. To calculate the head rotation angle, the center of the nose landmark from MediaPipe was used as a reference for the head position, as explained in the preprocessing stage. Subsequently, the X and Y coordinates of the nose landmark were normalized by multiplying each value by the width and height of the frame. Based on these coordinates, the rotation angles along the X and Y axes were calculated to detect head movements up or down, using a series of thresholds defined in this study [26] :

1. Head pose up: if X of angle 7 degree
2. Head pose up: if X of angle -7 degree

2.4. *Data Labeling*

Blinking is a rapid movement in which the eyes close and reopen, normally lasting about 100 to 400 milliseconds. Yawning, on the other hand, is a quick action of opening and closing the mouth, usually lasting about 4 to 6 seconds. A drowsy head pose is characterized by random head movements caused by severe drowsiness, often accompanied by eye closure, and it can last for several seconds. The patterns of blinking, yawning, and head pose vary among individuals, dependent upon factors such as duration, degree of opening or closing, angle of head tilt, and speed of movement. Additionally, a single data capture of EAR, MAR, and the X and Y coordinates of the nose per frame is insufficient to fully capture events like blinking, yawning, or head movements associated with drowsiness. Therefore, to detect various action patterns indicating drowsiness, we then added a time range in the drowsiness classification.

In this study, labels ‘1’ and ‘0’ were used to differentiate between facial features that indicated drowsiness and those that appeared normal. Label ‘1’ markeds facial features indicating drowsiness, while label ‘0’ marked features considered normal. Table 1 provides a description of the facial features of each driver, and Table 2 presents the labeling results for the training and testing data.

Table 1. Dataset Description

Driver’s Behaviors	Description
Sleepy-eyes	When eyes slowly close due to drowsiness
Yawning	When mouth open wide due to drowsiness
Nodding	When head falls forward when drowsy
Stillness	When normally driving

If the image showed signs such as eyelid closing, mouth yawning, or head nodding, it was then labeled “1”. A sleepy eye condition was defined as an eye threshold value of less than 0.2, while a mouth threshold value of more than 0.5, indicating the sleepy driver. In contrast, if the image did not show these signs, it was labeled “0”, where the eyes were considered non-drowsy if the threshold value of the eye was greater 0.2 and the threshold value of the mouth was less than 0.5, indicating that the awake driver.

Table 2. Result of labelling training and testing datasets

Status	Type	Num Face	
		Training	Testing
Drowsy	Images	8,200	1,320
Non-drowsy	Images	14,300	4,000
Both	Videos	87	20

Each driver video was extracted into 280 images, with labelling performed on each frame. The resulting data included various features such as EAR (Eye Aspect Ratio), MAR (Mouth Aspect Ratio), head pose, as well as corresponding labels, which were then saved in CSV format for further analysis. Table 2 presents the image extraction results for the training and testing datasets.

2.5. *Classification*

After labeling the extracted values, it was continued with data classification. This research employed three deep neural network architectures to evaluate behavioral features, including. 1D-CNN architecture, LSTM architecture, and BiLSTM architecture.

2.5.1. 1D-CNN Architecture

The 1D-CNN architecture was used to automatically learn and ex- tract temporal patterns from the sequence of behavioral features, such as the movement of the eyes, mouth, and head pose, for the detection of drowsiness. It efficiently processeds sequential data by applying convolutional filters across one dimension, making it suitable for tasks involving time-series data. Table 3 presents the details of the proposed 1D-CNN architecture. The proposed model consisted of several layers. The first layer was Conv1D with 16 filters and an output size of (None, 1, 16), resulting in a total of 80 parameters trained. The second Conv1D layer had 32 filters with an output size of (None, 1, 32), which added 544 parameters. Following this, Batch Normalization was applied to the layer with an output size of (None, 1, 32), which resulted in 128 additional parameters. Then, dropout was used with a fixed output size (None, 1, 32) to prevent overfitting, although this layer did not add any parameters. Next, the flatten layer changed the output dimension to a one-dimensional vector of size 32. The next Dense layer had 64 units, resulting in 2112 parameters, and ended with the last Dense layer with 2 output units, adding 130 additional parameters. Overall, the model consisted of several convolution, normalization, dropout, and fully connected layers used to generate the final prediction.

Table 3. Architecture of proposed 1D-CNN model

Layer (type)	Output Shape	Parameter
conv1d_4 (Conv1D)	(None, 1, 16)	64
conv1d_5 (Conv1D)	(None, 1, 32)	544
batch_normalization_2	(None, 1, 32)	128
dropout_2 (Dropout)	(None, 1, 32)	0
flatten_2 (Flatten)	(None, 32)	0
dense_4 (Dense)	(None, 64)	2,112
dense_5 (Dense)	(None, 2)	130

2.5.2. LSTM Architecture

The LSTM architecture was designed to capture long-term dependencies in sequential data [27], making it suitable for detecting temporal patterns associated with driver drowsiness. By using memory cells, the LSTM model can efficiently process the sequences of EAR, MAR, and head pose features over time. The details of the proposed LSTM architecture are presented in Table 4.

The model started with an LSTM layer that had 100 output units with an output shape of (1, 100), resulting in 41,600 trainable parameters. Afterwards, a dropout layer was applied with the same output shape (1, 100) to help to reduce overfitting, without increasing the number of parameters. A flatten layer was then used to convert the output into a one-dimensional vector of size 100. Following this, the model was fitted with the first Dense layer, containing 128 units, which added 12,928 parameters. The second Dense layer had 64 units, contributing 8,256 additional parameters. Finally, the model ended with a dense layer with 3 output units, adding 195 parameters. This layer was responsible for making the final prediction with 3 output classes.

Table 4. Architecture of proposed LSTM model

Layer (type)	Output Shape	Parameter
LSTM	(None, 1, 100)	41,600
dropout_2 (Dropout)	(None, 1, 100)	0
flatten_2 (Flatten)	(None, 100)	0
dense_4 (Dense)	(None, 128)	12,928
dense_5 (Dense)	(None, 64)	8,256
dense_6 (Dense)	(None, 2)	195

2.5.3. BiLSTM Architecture

The BiLSTM (Bidirectional LSTM) architecture was designed to capture both past and future context in a sequence, making it more effective for tasks where the full context is important, such as detecting drowsiness patterns over time. By processing data in both forward and backward directions, BiLSTM provided a more comprehensive understanding of sequential features. Table 5 presents the details of the proposed BiLSTM architecture. The model began with a bidirectional LSTM layer, combining two directions of LSTM, resulting in an output of size (1, 200) and a parameter count of 84,000. This layer processed sequential information from both directions (forward and backward) to better understand the context of the data. After this layer, a Dropout layer was applied with the same output size (1, 200) to prevent overfitting without increasing the number of parameters. Next, a Flatten layer was used to convert the output into a one-dimensional vector of size 200. Then, a Dense layer with 128 units is applied, adding 25,728 parameters, followed by another Dense layer with 64 units, contributing an additional 8,256 parameters. Finally, a dense layer with 2 output units is used to make the final prediction, adding 130 parameters.

Table 5. Architecture of proposed BiLSTM model

Layer (type)	Output Shape	Parameter
bidirectional_1	(None, 1, 200)	84,000
dropout_2 (Dropout)	(None, 1, 200)	0
flatten_2 (Flatten)	(None, 200)	0
dense_1 (Dense)	(None, 128)	25,728
dense_2 (Dense)	(None, 64)	8,256
dense_3 (Dense)	(None, 2)	130

2.5.4. Justification for Selecting a Deep Learning Model

The selection of deep learning models in this study was based on efficiency, capability to capture temporal patterns, and balance between precision and computational requirements. The 1D-CNN model was chosen for its ability to extract features from sequential data with less care than 2D or 3D CNNs, mak- ing it more processing efficient and resource efficient. Mean- while, LSTM was used for its ability to handle time-based data by retaining information from previous frames, crucial in detecting gradual drowsiness patterns. BiLSTM was chosen to improve accuracy by processing information from both directions, allowing the model to capture changes in facial expressions with a broader time context.

Compared to other models, standard CNNs such as 2D-CNN or 3D-CNN are more suitable for the spatial analysis of static images, whereas the drowsiness detection task relies more on analyzing changes in facial features within a video sequence. In addition, 3D-CNN has higher computational requirements, making it less efficient than the combination of 1D-CNN and LSTM/BiLSTM. Meanwhile, Transformer-based models such as Vision Transformer (ViT) or TimeSformer, while having potential in sequential data processing, require much larger datasets and high computational power to perform optimally. Therefore, the selection of a combination of 1D-CNN, LSTM, and BiLSTM in this study provided a more balanced solution in terms of accuracy, computational efficiency, and temporal pattern capture capability, making it suitable for video-based drowsiness detection in real-time driver monitoring systems.

2.6. *Evaluation Metrics*

To evaluate model performance, the following metrics were used: (i) Accuracy: to measures the overall accuracy of the predictions. (ii) Precision and recall: Precision evaluates how many the predicted drowsy states are correct, while recall measures refers to how many actual drowsy states are identified and. (iii) F1 score: combining precision and recall to provide a balanced performance metric. The hyperparameters used in this study are described in Table 6.

Table 6. Hyperparameters Models

Hyperparameter	Nilai
Learning Rate	0.001
Batch Size	32
Epochs	100
Optimizer	Adam
Loss Function	Categorical Crossentropy
K-fold Validation	5

3. Results and Discussion

This research proposes a more effective and accurate method for detecting driver drowsiness by combining three facial features, then compared with the use of one or two facial features alone. To evaluate the performance of the proposed feature combination, we used three deep learning models to ensure its accuracy and efficiency. Seven experiments were conducted on the complete dataset. The results of each experiment are discussed in this section. These experiments aimed to achieve the highest classification accuracy and optimal performance in other quantitative measures to detect drowsiness in drivers. To measure the impact and performance of the deep learning architecture in predicting driver yawning, performance was assessed using standard metrics such as accuracy, precision, recall, and F1 score. These measures were used to compare our solution with any existing relevant literature

3.1. Experiment Focused on a Single Facial Feature

In the first experiment, we focused on detecting driver drowsiness by utilizing a single facial feature. The YawDD public dataset was used for both training and validation. Three sepa- rated experiments were conducted: (i) detecting driver drowsiness using eye features, (ii) detecting driver drowsiness using mouth features, and (iii) detecting driver drowsiness using head posi- tion features. These experiments were evaluated using three deep learning models 1D-CNN, LSTM, and BiLST to assess the effectiveness of each model in detecting drowsiness. The evaluation aimed to compare the performance of these models in terms of detection accuracy.

The accuracy of the models in the first experiment is shown in Fig. 6 for the eye feature. The training accuracy using the 1D-CNN model is presented in Fig. 6(a), the LSTM model in Fig. 6(b), and the BiLSTM model in Fig. 6(c). Furthermore, Fig. 7 presents the accuracy graph for the mouth feature. Fig. 7(a) shows the accuracy of the 1D-CNN model, Fig. 7(b) displays the accuracy of the LSTM model, and Fig. 7(c) shows the accuracy of the BiLSTM model. Finally, Fig. 8 presents the accuracy graph for the head feature. Fig. 8(a), Fig. 8(b) and Fig.8 (c) respectively show the accuracy of the 1D-CNN model, the LSTM model, and the BiLSTM model.

Table 7 shows the training and testing comparison results of three deep learning models applied to detect driver drowsiness based on one facial feature. All three models achieved 0.99 accuracy during training. However, in the test data, the eye feature provided the best results with 0.96 accuracy in the LSTM model, followed by BiLSTM and 1DCNN with 0.95 accuracy. Meanwhile, the testing accuracy for the mouth feature only reached 0.71, and the head pose feature obtained an accuracy of 0.69.

This analysis indicated that eye features were more dominant and sensitive in detecting the signs of sleepiness compared to mouth and head pose features. The lower performance of the mouth and head pose features could be due to the greater variability in facial expressions and head movements, which may be inconsistent or less directly reflective of sleepiness levels. In contrast, the eyes tended to be more stable as an indicator of sleepiness, especially through metrics such as EAR which indicated prolonged eye closure. This supports the use of eye features as a key focus in the development of more accurate and reliable sleepiness detection models. In addition, these results showed the importance of using a combination of features to improve detection performance in real applications.

3.2. Experiment Focusing on a Two-Facial Feature

In the second experiment, we focused on detecting driver drowsiness by utilizing a two-facial feature. The YawDD public dataset was used for both training and validation. Three sepa- ratinge experiments were conducted: (i) detecting driver drowsiness using eye and mouth features, (ii) detecting driver drowsiness using mouth and head pose features, and (iii) detecting driver drowsiness using head pose and eye features. These experiments were evaluated using three deep learning models 1D-CNN, LSTM, and BiLSTM to assess the effectiveness of each model in detecting drowsiness. The evaluation aimed to compare the performance of these models in terms of detection accuracy.

Fig. 9 illustrates an accuracy graph on the eye and mouth features. Figs. 9(a), 9(b) and 9(c) show the 1D-CNN accuracy model, the LSTM accuracy model, and the BiLSTM model accuracy model, respectively. Furthermore, Fig. 10 depicts an accuracy graph on the eye and head pose features. Fig. 10(a), 10(b) and 10(c) respectively show the 1D-CNN accuracy model, the LSTM accuracy model, and the BiLSTM model accuracy model. Finally, Fig. 11 illustrates an accuracy graph on the mouth and head pose features. Fig. 11(a), 11(b) and 11(c) shows the 1D-CNN accuracy model, the LSTM accuracy model, and the BiLSTM model accuracy model, respectively.

Table 8 shows the training and testing results of the three deep learning models used to detect driver drowsiness based on a combination of two facial features. Each model achieved 0.99 accuracy on training data with different combinations of two features. The combination of eye and mouth features proved to be superior to the combination of head-mouth pose and head-eye pose. Detection accuracy with the combination of eye and mouth features reached 0.98 on the test data, followed by the combination of eye and head pose features with 0.97 accuracy, while the combination of mouth and head pose features only achieved 0.73 accuracy.

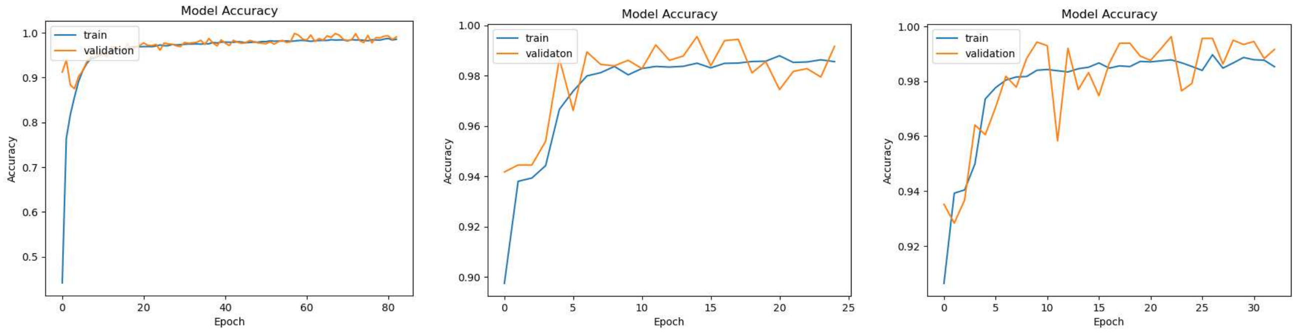
These results indicated that the features of the eyes and mouth were more relevant and effective in detecting drowsiness compared to the features of the head pose. The combination of eye and mouth provided richer information on the signs of sleepiness, such as eye closure (through eye aspect ratio, EAR) and mouth movements that may indicate yawning. In contrast, the characteristics of the head pose tended to be less sensitive in capturing sleepiness signals, possibly due to head movements that can be affected by other factors, such as sitting position or road conditions.

The lower accuracy in the combination of mouth and head pose (0.73) suggested that signals from mouth and head pose alone were not enough to consistently detect sleepiness. Head pose features, such as rotation or tilt, may be better used as supporting rather than main features in a drowsiness detection system. Meanwhile, the high accuracy in the combination of the eye and mouth confirmed the importance of multifunctional processing to improve the reliability of drowsiness detection models.

3.3. Experiment Focusing on a Three-Facial Feature

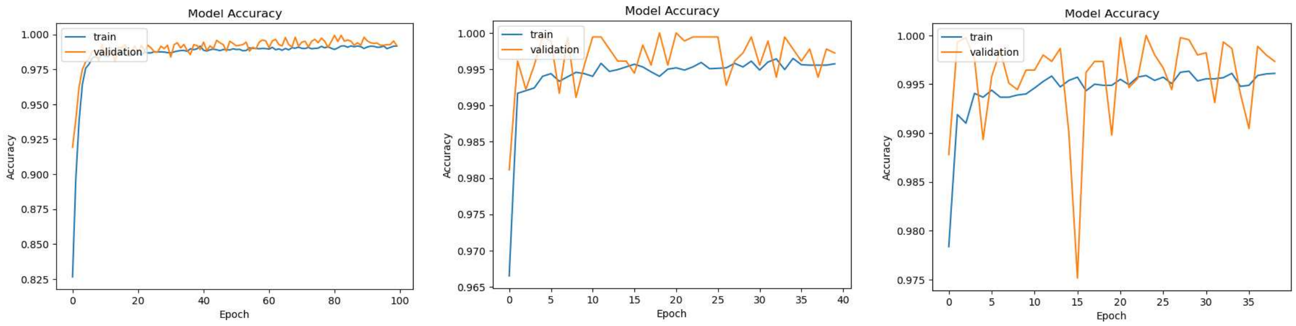
In the third experiment, we focused on detecting driver drowsiness using three facial characteristics. The YawDD public dataset was used for training and validation. This experiment incorporated three facial features:, eyes, mouth, and head pose. It was evaluated using three deep learning models: 1D-CNN, LSTM, and BiLSTM to assess the effective- ness of each model in detecting drowsiness. This evaluation aimed to compare the performance of these models in terms of detection accuracy.

Fig. 12 displays the accuracy graph based on the features of the mouth. Fig. 12(a) to Fig.12 (c) respectively show the accuracy of the 1D-CNN model, the LSTM model, and the BiLSTM model. The last test was conducted by combining three facial characteristics of the driver:, closed eyes, open mouth, and head pose as an indicator of drowsiness. Table 9 displays the training and testing accuracy results of the three deep learning models used. The results showed that the training accuracy reached 0.99, and the same accuracy was also achieved in the test data,



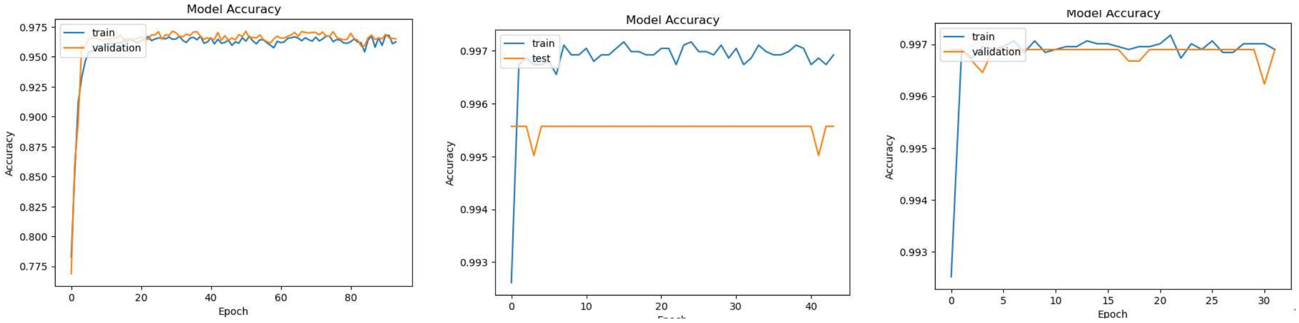
(a) (b) (c)

Fig. 6. Training and validation accuracy for the eye feature: (a) 1D-CNN results, (b) LSTM results, and (c) BiLSTM results



(a) (b) (c)

Fig. 7. Training and validation accuracy for the mouth feature: (a) 1D-CNN results, (b) LSTM results, and (c) BiLSTM results



(a) (b) (c)

Fig. 8. Training and validation accuracy for the head pose feature: (a) 1D-CNN results, (b) LSTM results, and (c) BiLSTM results

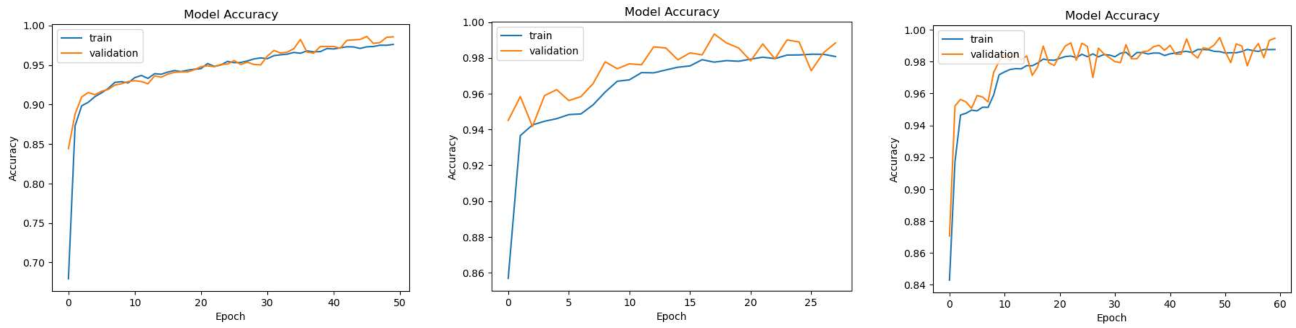
Table 7. Comparison of training and testing accuracy results on a single feature

Model	Eye Feature		Mouth Feature		Head Pose Feature	
	Training	Testing	Training	Testing	Training	Testing
1D-CNN	0.99	0.95	0.99	0.71	0.99	0.67
LSTM	0.99	0.96	1.00	0.71	1.00	0.69
BiLSTM	0.99	0.95	0.99	0.71	0.99	0.69

with a value of 0.99.

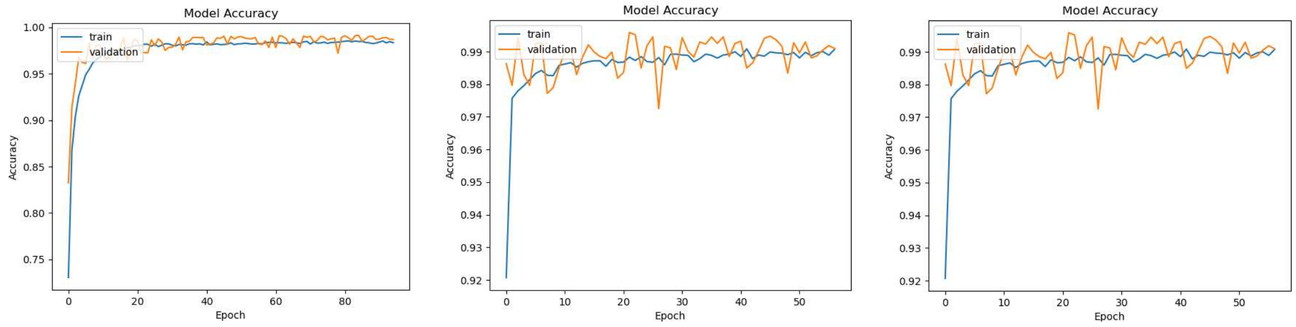
The very high accuracy in training and testing (0.99) indicated that the three models 1D-CNN, LSTM, and BiLSTM were able to generalize very well in detecting drowsiness when using the combination of closed eyes, open mouth, and head pose features. Together, these three characteristics provided a clear and comprehensive indicator of sleepiness.

Scientifically, the closed-eyes feature is often used in drowsiness detection systems because sustained eye closure is one of the most significant signs. An open mouth, which can indicate yawning, is also a common signal of sleepiness. Head pose, specifically tilt or rotation, can be an additional indicator, as the head often leans or falls when a person is sleepy. By combining these three features, the model has access to a variety of comple-



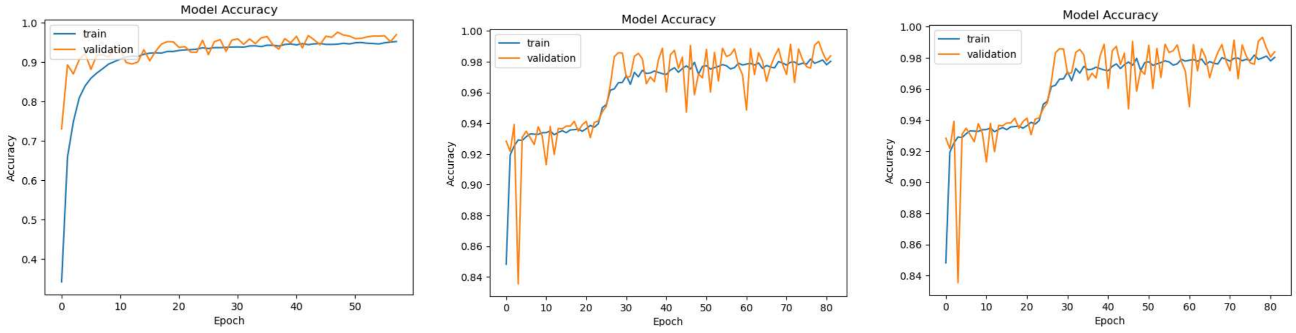
(a) (b) (c)

Fig. 9. Training and validation accuracy for eyes and mouth features:(a) 1D-CNN results, (b) LSTM results, and (c) BiLSTM results



(a) (b) (c)

Fig. 10. Training and validation accuracy for mouth and head features:(a) 1D-CNN results, (b) LSTM results, and (c) BiLSTM results



(a) (b) (c)

Fig. 11. Training and validation accuracy for head and eye features: (a) 1D-CNN results, (b) LSTM results, and (c) BiLSTM results

Table 8. Comparison of training and testing accuracy results on two features

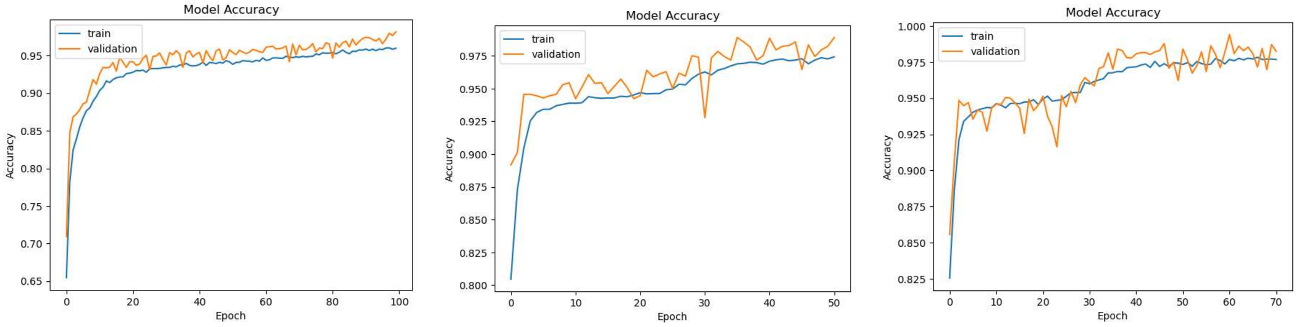
Model	Eye-Mouth Features		Mouth-Head Features		Head-Eye Features	
	Training	Testing	Training	Testing	Training	Testing
1D-CNN	0.99	0.98	0.98	0.73	0.97	0.95
LSTM	0.99	0.98	0.99	0.72	0.98	0.97
BiLSTM	0.99	0.98	0.99	0.73	0.97	0.96

mentary information, which explains why the accuracy results are so high.

3.4. Discussion

On testing a single facial characteristic, as presented in Table 7, all three models 1D-CNN, LSTM, and BiLSTM showed high scores in predicting training data. However, the prediction performance decreased on the head pose feature test data, with the accuracy reaching 0.69. However, when the three facial features were compared, the eye feature showed the best performance with an accuracy rate of 0.98.

The same is true for tests with two facial features together, as shown in Table 8. The test prediction for the combination of mouth and head features is lower, with an accuracy of 0.73. The



(a)

(b)

(c)

Fig. 12. Training and validation accuracy for eyes, mouth, and head pose features: (a) 1D-CNN results, (b) LSTM results, and (c) BiLSTM results

Table 9. Comparison of training and testing accuracy results on three features

Model	Eye-Mouth-head pose Features	
	Training	Testing
1D-CNN	0.99	0.98
LSTM	0.99	0.98
BiLSTM	0.99	0.98

best result in accuracy in testing both facial features is obtained from the combination of eyes and mouth, with prediction accuracy reaching 0.98. Finally, in the test combining the three facial features, as shown in Table 9, all three models showed excellent performance with accuracy reaching 0.98. Based on the test results, the eye feature was found able to makes the most significant contribution in the detection of drowsiness, with very high accuracy in all models tested. The head feature, although being useful, showed lower prediction performance than the other features. The combined features of the eyes and the mouth were also shown to be very effective, showing that the combination of information from these two features could improve the accuracy of the prediction. When all three features were combined, the results showed that all deep learning models could achieve very high accuracy, suggesting that multimodal approaches are more effective in detecting drowsiness comprehensively.

This analysis underscored the importance of selecting relevant features for deep learning models in drowsiness detection applications, especially in the context of using real-time video data. In addition, significance tests were performed to determine whether the EAR and MAR characteristics in the public data set have significant differences compared to the private data set of the researcher. A performance test was also applied to evaluate the reliability of the features in detecting drowsiness on both private and public data. The results of the ANOVA analysis showed a significant difference in the use of the public dataset and the private dataset of the researcher.

In addition, the computational efficiency of each model was evaluated to determine its feasibility in a real-time drowsiness detection system. The results of the analysis showed that the 1D-CNN model had a faster inference time compared to the LSTM and BiLSTM, making it more suitable for real-time implementation on devices with limited resources. In contrast, although the LSTM and BiLSTM models were able to capture sequential patterns better, the higher processing time can be challenging in applications that require fast responses. Therefore, the selection of the model in real-time systems should consider the balance between accuracy and computational efficiency.

4. Conclusion

This study demonstrated that combining facial features such as eye closure (EAR), mouth opening (MAR), and head pose (rotation) is an effective approach for detecting driver drowsiness. The three deep learning models evaluated in this research, namely 1D-CNN, LSTM, and BiLSTM achieved high performance, with BiLSTM yielding the best results. Specifically, the model achieved an accuracy of 0.99 and an F1-score of 0.98 on the YawDD dataset, and an accuracy of 0.98 with an F1-score of 0.96 on a custom dataset. These results confirmed that integrating multiple facial cues significantly improves the performance of visual-based drowsiness detection systems, especially for real-time applications.

In addition to highlighting the effectiveness of visual analysis, this study emphasizes the importance of proper feature selection and fusion in building accurate and reliable drowsiness detection systems. The dynamic changes in the eyes, mouth, and head position were proven to be reliable indicators of drowsiness. While the high accuracy achieved indicated strong potential for practical use, it is important to note that the custom dataset were collected under controlled conditions. As such, further testing in real-world driving environments is necessary to validate the system's practical robustness.

For future work, more advanced deep learning architectures such as Transformers or Vision Transformers (ViT) can be explored to enhance the model's resilience to challenges such as lighting variations, camera angles, and facial occlusions. The high performance observed in this study suggested a strong foundation, but Transformer-based models may better capture complex spatial-temporal dependencies in real-time scenarios. Additionally, expanding the dataset with more diverse driving conditions, and deploying the system on lightweight hardware (e.g., edge devices or embedded systems), will be essential for real-time implementation. Integrating visual features with physiological sensors such as heart rate monitors or EEG could also further improve the reliability and robustness of drowsiness detection.

Acknowledgment

We would like to thank the University of Gadjah Mada, Indonesia, and the King Mongkut's Institute of Technology Ladkrabang, Thailand, for their cooperation and support in conducting this research. The collaboration between the two institutions has contributed significantly to the development and completion of this research. Hopefully this partnership can continue to be established and produce innovations and knowledge that are beneficial to the development of science in the future.

References

1. M. Nasr Azadani and A. Boukerche, *"Driving Behavior Analysis Guidelines for Intelligent Transportation Systems,"* in IEEE Transactions on Intelligent Transportation Systems, 23 (2022) 6027-6045.
2. A. A. Saleem, H. U. R. Siddiqui, M. A. Raza, F. Rustam, S. Dudley, and I. Ashraf, *"A systematic review of physiological signals based driver drowsiness detection systems,"* Cognitive neurodynamics, 17 (2023) 1229–1259.
3. Y. Albadawi, M. Takruri, and M. Awad, *"A review of recent developments in driver drowsiness detection systems,"* Sensors, 22 (2022) 2069.
4. K. Kanwal, F. Rustam, R. Chaganti, A. D. Jurcut, and I. Ashraf, *"Smartphone inertial measurement unit data features for analyzing driver driving behavior,"* IEEE Sensors Journal, 23 (2023) 11308–11323.
5. M. Gromer, D. Salb, T. Walzer, N. M. Madrid, and R. Seepold, *"ECG sensor for detection of driver's drowsiness,"* Proc. Comput. Sci., 159 (2019) 1938–1946.
6. Y. Jiang, Y. Zhang, C. Lin, D. Wu, and C.-T. Lin, *"EEG-based driver drowsiness estimation using an online multi-view and transfer TSK fuzzy system,"* IEEE Trans. Intell. Transp. Syst., 22 (2021) 1752–1764.
7. 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., 9 (2022) 12338–12350.
8. Z. Qiu, J. Zhao, and S. Sun, *"MFIALane: Multiscale feature information aggregator network for lane detection,"* IEEE Trans. Intell. Transp. Syst., 23 (2022) 24263–24275.
9. A. S. Agarkar, R. Gandhiraj and M. K. Panda, *"Driver Drowsiness Detection and Warning using Facial Features and Hand Gestures,"* 2023 2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN), Vellore, India, (2023) 1-6.
10. P. Baby Shamini, M. Vinodhini, B. Keerthana, S. Lakshna and K. R. Meenatchi, *"Driver Drowsiness Detection based on Monitoring of Eye Blink Rate,"* 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, (2022) 1595-1599.
11. T. Ahmed, O. Jyoti and T. H. Mou, *"Drivers' Drowsiness Detection System Enhancement Using Deep Learning: CNN-based Approach,"* 2023 26th International Conference on Computer and Information Technology (ICCIT), Cox's Bazar, Bangladesh, (2023) 1-6.
12. F. Osmani and P. Wawage, *"Real-Time Driver Drowsiness Detection System using Vision Transformer for Accurate Eye State Analysis,"* 2024 International Conference on Intelligent Systems and Advanced Applications (ICISAA) Pune, India. (2024) 1-5.
13. M. M. Desai, K. Kathad, and N. Modi, *"Real-Time Driver Drowsiness Detection using Hybrid CNN-LSTM Model with Facial Feature and Behavioral Analysis,"* Proceedings of the Fourth International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS-2024). (2024) 197-202.
14. S. T. Lin, Y. Y. Tan, P. Y. Chua, L. K. Tey, and C. H. Ang, *"Perclos threshold for drowsiness detection during real driving,"* Journal of Vision, 12 (2012) 546–546.
15. A. Rosebrock, *"Eye blink detection with opencv, python, and dlib,"* Blog in Pyimagesearch, (2017).
16. A. Moujahid, F. Dornaika, I. Arganda-Carreras, and J. Reta, *"Efficient and compact face descriptor for driver drowsiness detection,"* Expert Systems with Applications, 168 (2021) 114334.
17. A. Celecia, K. Figueiredo, M. Vellasco, and R. Gonz ́alez, *"A portable fuzzy driver drowsiness estimation system,"* Sensors, 20 (2020) 4093.
18. W. Liu, J. Qian, Z. Yao, X. Jiao, and J. Pan, *"Convolutional two-stream network using multi-facial feature fusion for driver fatigue detection,"* Future Internet, 11,(2019) 115.
19. A. Al Redhaei, Y. Albadawi, S. Mohamed, and A. Alnoman, *"Realtime driver drowsiness detection using machine learning,"* in 2022 Advances in Science and Engineering Technology International Conferences (ASET), (2022) 1–6.
20. H. Jin, S. Liao, and L. Shao, *"Pixel-in-pixel Net: Towards efficient facial landmark detection in the wild,"* Int. J. Comput. Vis., 129 (2021) 3174–3194.
21. J. Guo, X. Zhu, Y. Yang, F. Yang, Z. Lei, and S. Z. Li, *"Towards fast, accurate and stable 3D dense face alignment,"* in Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer, (2020) 152–168.
22. Q. Wu, N. Li, L. Zhang and F. Richard Yu, *"Driver Drowsiness Detection Based on Joint Human Face and Facial Landmark Localization With Cheap Operations,"* in IEEE Transactions on Intelligent Transportation Systems, 25 (2024) 19633-19645.
23. S. Abtahi, M. Omidyeganeh, S. Shirmohammadi, and B. Hariri, *"Yawdd: A yawning detection dataset,"* in Proceedings of the 5th ACM multimedia systems conference, (2014) 24–28.
24. G. Sanil, K. Prakash, S. Prabhu, V. C. Nayak, and S. Sengupta, *"2d-3d facial image analysis for identification of facial features using machine learning algorithms with hyperparameter optimization for forensics applications,"* IEEE Access, 11 (2023) 82521–82538.
25. Y. Albadawi, A. AlRedhaei, and M. Takruri, *"Real-time machine learning-based driver drowsiness detection using visual features,"* Journal of imaging, 9 (2023) 91.
26. R. A. Rajagede and R. P. Hastuti, *"Al-Quran recitation verification for memorization test using Siamese LSTM network,"* Communications in Science and Technology (CST), 6 (2021) 35-40.