JOIV : Int. J. Inform. Visualization, 7(4) - December 2023 2489-2497

INTERNATIONAL JOURNAL
ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

Drowsiness Detection System through Eye and Mouth Analysis

Bey-Ee Belle Lim [a], Kok-Why Ng [a,*], Sew-Lai Ng [a]

[a] *Faculty of Computing and Informatics, Multimedia University, Selangor, 63100, Malaysia*

*Corresponding author: *kwng@mmu.edu.my*

***Abstract*— Traffic jams are one of the serious issues in many developed countries. After the pandemic, many employees were allowed to travel interstate to work. This contributes to more severe jams, especially in the capital and nearby states. Long-distance driving and congestion can easily make the drivers sleepy and thus lead to traffic accidents. This paper aims to study and analyze facial cues to detect early symptoms of drowsy driving. The proposed method employs a deep learning approach, utilizing ensemble CNNs and Dlib's 68 landmark face detectors to analyze the facial cues. The analyzed symptoms include the frequency of eyes opened or closed and yawning or no yawning. Three individual CNN models and an ensemble CNN structure are built for the classification of the eyes and mouth yawn. The model training and validation accuracy graph and training loss and validation loss graph are plotted to verify that the models have not been overfitted. The ensemble CNN models achieved an approximate accuracy of 97.4% from the eyes and 96.5% from the mouth. It outperforms the other pre-trained models. The proposed system can immediately alert the driver and send text drowsy messages and emails to the third party, ensuring timely intervention to prevent accidents. The proposed method can be integrated into vehicles and transportation systems to ensure driver's safety. It can also be applied to monitor the driving behavior of those who drive long distances.**

***Keywords*—Drowsiness detection system; driver monitoring; facial expression recognition; ensemble CNN; image processing.**

Manuscript received 5 Dec. 2022; revised 10 Jul. 2023; accepted 30 Aug. 2023. Date of publication 31 Dec. 2023.

I. INTRODUCTION

Drowsiness is feeling sleepy, tired, or unable to keep one's eyes open. Drowsy people may fall asleep when they do not want to or when their safety is seriously impacted. Even if a driver is aware of the regulations and safety protocols, drowsiness causes brain and nerve recklessness, which results in accidents and unforeseen circumstances. Many road accidents are mainly due to driver drowsiness. The major concern caused by massive traffic is a significant rise in traffic accidents. Traffic accidents are undoubtedly a national problem in our country. The World Health Organization Launches the Global Plan for the Decade of Action for Road Safety 2021–2030 reported that traffic accidents result in nearly 1.3 million fatalities and injuries in its worldwide road safety status report. Malaysia ranks third in the WHO Western Pacific Region.

Drowsy driving results in the driver losing control of the car. As a result, the driver may suddenly veer off the road, strike an object, or collide with another car, turning it over. Drowsy driving can cause some symptoms, including frequent yawning, difficulties keeping your eyes on the road, head nodding, and abnormal driving habits, including driving too fast, drifting out lane, tailgating, or disobeying traffic signals. Drivers should be mindful of their situations before getting behind the wheel. In order to solve this issue, the system will employ face recognition to identify the symptoms of drowsy driving and detect their early warning indications. The system will generate a sequence of stimuli to get the driver's attention and encourage him to stop the car and take a break.

A. Driver Drowsiness Detection Using Convolutional Neural Networks (CNN) Techniques

Jabbar et al. [1] developed a CNN model for a system to detect driver drowsiness in real-time. The system can extract and identify facial landmarks from pictures taken with a device utilizing the Dlib Library OpenCV's Haar Cascade from the image before passing it to a Deep Learning CNN model for training on the NTHU Driver Drowsiness Detection dataset. The model is built using the five-layer CNN technique. The model managed to keep a relatively compact model size of no more than 75KB while averaging 83.33% accuracy across all categories.

Besides, Rafid et al. [2] proposed a solution using a convolutional neural network (CNN) model to detect drivers' drowsiness. This paper combines the CenterFace method with the Haar-Classifier to extract eye feature information and detect faces. A CNN architecture that consists of 3 convolutional layers has been developed and trained using the MRL eye dataset. The model has been evaluated using three datasets created by the authors. The CenterFace technique produced promising results with and without mask scenarios. Regarding this dataset, it was impressive that the proposed technique's validation accuracy was 98%, and the accuracy for real-time is 94.5%.

Moreover, Elham et al. [3] proposed a CNN model using the MRL dataset to determine the opening and closing of the driver's eyes. Detection of edges, conversion to grayscale, and dilation are a few of the image processing methods the proposed model used to process the images in the dataset. Then, the MediaPipe Face mesh model from Google was used to determine facial landmarks from image sequences. The authors applied dilation to the images after performing noise removal and Canny edge detection to enhance the system. As a result, the proposed CNN model achieved a 95% accuracy rate.

In another work, Suresh et al. [4] proposed using the vision-based methodology to detect drowsiness. This method relied on eye blinks to identify driver drowsiness. The Viola-Jones technique, commonly called the Haar cascades algorithm, is used to identify faces and eyes in input images. In this study, a CNN is built and used to extract features during the learning phase. In the suggested method, categorical cross entropy is employed as the model's training loss function, and adaptive moment estimation (Adam) is used as an optimizer. The suggested approach is assessed using a sizable portion of 48000 pictures of the MRL eye dataset. The proposed model achieved a test accuracy of 86.05%.

Furthermore, in this study by Chirra et al. [5], a framework for detecting driver drowsiness focusing on eye movement is presented. The Viola-Jones face detection technique is employed to identify faces in the images. Stacked deep convolutional neural networks are created for the learning phase to extract features. Four convolutional layers in a CNN are used to extract the deep features, which are subsequently passed on to a fully connected layer. The SoftMax layer in CNN classified the images' drowsy and awake categories. As a result, the proposed method had an accuracy of 96.42%.

Next, Faisal et al. [6] proposed a CNN model to predict driver drowsiness and offer a warning system to prevent accidents caused by driver drowsiness. The CNN model's hyperparameters are systematically optimized using pre-processed images from the Haar cascade algorithm. The first step used the object detection approach based on Viola-Jones for image extraction and pre-processing. Four hyperparameters were examined at this stage to find the model that achieves the best performance: the size of the kernel, the pooling size, the learning rate, and the number of epochs. Lastly, test results correctly classified the drowsy driver with 97.98% accuracy.

B. Driver Drowsiness Detection Using Ensemble Neural Network Model

Dua et al. [7] proposed four modes: VGG-FaceNet, AlexNet, ResNet, and FlowImageNet. These models divide the characteristics into four categories: no drowsiness, drowsiness with yawning, nodding, and eye closed. The output from the models is passed to the ensemble algorithm through a SoftMax classifier to produce the result. For ensembles, a basic averaging method is implemented in this paper. The driver is drowsy if the average of all model outputs exceeds the threshold value 0.24. The proposed system exhibits robustness under all circumstances and has an accuracy of 85%.

In another work by Ahmed et al. [8], they proposed an ensemble neural network that used the eyes and mouth features samples from the NTHU dataset. The MTCNN has been used in the proposed model to obtain apriori positions for the areas of focus: the eyes and mouth. The proposed ensemble model consists of two InceptionV3 CNNs to extract the features of the eyes and mouth data samples from the face images using the MTCNN. Each ensemble's weights are given training on a feed-forward neural network with input from the output of the two CNNs. As a result, the ensemble architecture outperformed individual CNN models. The model's accuracy on the testing dataset was 97.1%, with training and validation accuracy of 99.65% and 98.5%, respectively.

Moreover, Park et al. [9] proposed a deep learning network comprising three different neural networks: VGG-FaceNet, FlowImageNet, and AlexNet. The outputs from the three networks are combined to create a single prediction. They used independently averaged architecture (IAA) and feature-fused architecture, two different ensemble strategies (FFA). The proposed system achieved an accuracy of 73.06% on the NTHU dataset.

C. Driver Drowsiness Detection Using Hybrid Deep Learning Model

Kumar et al. [10] developed a hybrid deep learning model that combined modified InceptionV3 networks with long short-term memory (LSTM) networks. InceptionV3 has been customized for spatial robustness by adding a global average-pooling layer. Following that, the output of the modified InceptionV3 was fed into LSTM for drowsiness detection. The NTHU-DDD dataset is used to assess the performance of the proposed drowsy detection model. Over the NTHUDD dataset, the proposed model achieved an accuracy of 91.36%.

Furthermore, Guo and Markoni [11] proposed a hybrid of CNN and long short-term memory (LSTM) model to detect driver drowsiness. However, the temporal analyzer, part of the suggested method, is a new strategy combining LSTM. The suggested method's temporal analyzer combined LSTM and used a time skip feature to enable each LSTM to catch a distinct temporal signal for processing. The model's output predicts the drowsiness and will be fed back into the refining step. The system's performance was evaluated using the dataset from the ACCV 2016 competition, and an accuracy of 84.85 % was achieved.

In another work from Najla and Reshma [12], they presented a hybrid deep learning model consisting of two deep CNs: a temporal convolutional neural network, a spatial convolutional neural network, and a deep belief network (DBN) to detect drowsiness. The outputs are concatenated and provided as the input to the fusion network created with a deep belief network after training with both spatial and temporal CNNs. Multiple layers of neural networks, each with a visible and a hidden layer, make up the DBN. The DBN captures the nonlinear interactions between spatial and temporal networks. The current dataset is retrained after receiving a new image and predicting the eye's condition using tensor flow. A threshold for eye closure is established after several frames of observation. An alert sound lets the driver know they are too sleepy to drive as soon as the threshold value is reached.

D. Drowsiness Detection System Based on Machine Learning Model

Öztürk et al.[13] proposed the Viola-Jones algorithm to identify the areas of the eyes and face. Using a machine learning technique, the eye area detection is classified as open or closed. Finally, drowsiness conditions are identified using a decision tree, Support Vector Machine (SVM), and kNN classifiers. The models were evaluated with an accuracy of 99.77%, 94.35%, and 96.62%, respectively on seven real people.

Moreover, in this work by Mehta et al. [14], an advanced system called AD3S (Advanced Driver Drowsiness Detection System) was created using an Android application. The facial landmarks are used to derive several parameters, such as Nose Length Ratio (NLR), Eye Aspect Ratio (EAR), and Mouth Opening Ratio (MOR) based on the adaptive threshold, which can detect driver drowsiness. Mehta et al. [14] proposed SVM, Random Forest, Naive Bayes, Boosting, Voting, and Bagging as machine learning classifiers to evaluate the AD3S's accuracy. With the Bagging classifier, the system obtained approximately 98% accuracy.

This research by Lin et al. [15] also developed a driver drowsiness detection system based on machine learning and gradient statistics. The suggested system consists of four components: detection of the face, detection of the eye bridge, detection of eyes, and detection of eyes closure. The proposed system design employed the Haar-Adaboost-based face detection algorithm. When a driver wears glasses, the accuracy is 91.49%, and when the driver does not wear glasses, the percentage is 95%.

Besides, Dey et al. [16] proposed a driver drowsiness surveillance system and compared it using three different methods: FLDA, Bayesian classifier, and SVM. A feature descriptor trained by linear SVM and HOG is used to extract the nose, mouth, and eyes position. These details are then contrasted with the value threshold taken from the aspect ratio data set for sleeping or drowsy face models. FLDA and SVM have outperformed the Bayesian Classifier with an accuracy of 93.3% and 96.4% respectively.

In [17], four machine learning models were used to build the assessment model of drowsiness. Open-source C++ library Dlib is used to detect facial landmarks that have been trained in a huge image dataset. Logistic regression produced results with an accuracy of 83.7% and 85.4%, respectively, using cross-validation.

Furthermore, Maior et al. [18] also developed Machine Learning based models to detect early symptoms and drowsiness. The proposed method estimates EAR using face landmarks and then uses the model to categorize the user's state in real-time and at a low cost. The SVM has the highest accuracy of 94.44% out of MLP, Random Forest, and SVM.

E. Deep Learning-Based Driver Drowsiness Detection System

In [19], Magán et al. proposed two deep-learning models of driver drowsiness detection systems. The first model was proposed using a combination of a recurrent neural network with a CNN, whereas the second model was proposed using the pre-trained CNN model and fuzzy logic. The CNN is built using the EfficientNetB0 architecture, which offers a compact model with accelerated processing. On the other hand, the second model's fuzzy inference system (FIS) produced a numerical output that represents the driver's estimated level of drowsiness. Both models achieved a similar accuracy level of approximately 65% and 55-65% on training and test datasets. Considering that only two balanced classes (awake and drowsy) exist, these results are quite poor.

Next, Hashemi et al. [20] also developed a driver drowsiness detection system to categorize the eye status (TL-VGG) using Transfer Learning in VGG16 and VGG19 as well as Fully Designed Neural Networks (FD-NN) with additional layers. The absolute distance between points 37 and 40, 43 and 46 is measured to locate the eye landmarks when using the Viola-Jones algorithm to detect faces. Additionally, the challenging lighting condition is overcome by adjusting eye contrast using the histogram equalizer. The ZJU dataset and the author's extended dataset are used to measure drowsiness. FD-NN is superior to other proposed networks due to its simplicity and speed for real-time tasks. The experimental findings showed that the FD-NN network achieved 99.8% AUC and 98.15% accuracy. The required time for classifying eye states in the FD-NN network is 1.4 milliseconds, making it also reliable for real-time tasks.

II. Material and Method

Figure 1 shows the flowchart of the proposed drowsiness detection system. First, the system will find the face and eye movements from the web camera to determine the driver's state. The facial cues of a drowsy driver will be analyzed by looking at the symptoms shown by the driver, including the frequency of eyes opening/closing and yawning/ no yawning. To identify the face of a human, a face detector will be used to locate a person's face in an image. Once the position of the face has been identified, features of the human face can be extracted using facial landmark points detection through Dlib's 68 Model.

Dlib’s 68 landmark face detector uses 68 coordinates that match the facial structures on the face to estimate their location on the face. Dlib uses methods such as Support Vector Machine (SVM) and Histogram of Oriented Gradients (HOG) to identify and construct the 68 landmarks of a face in an image, such as the eyes, nose, mouth, and jawline. As seen in Figure 2 above, each facial feature is mapped with a series of points. The points from 49 to 68 can be used to locate a mouth in the face. Therefore, Dlib's 68 landmark model is used to crop this paper's eyes and mouth images.

Furthermore, after the process of extracting the eye and mouth as a region of interest, it is fed to the trained CNN models for classification. CNN is one of the most popular deep learning algorithms due to its effectiveness in image classification [21], [22]. Additionally, CNN requires substantially less pre-processing than other classification methods [23]. Three different Convolutional Neural Networks (CNN) and Ensemble CNN techniques are decided to be used on each of the eyes and mouth classification models. Since there is a lack of a dataset of drowsy drivers, building and training the eyes and mouth features individually is preferable.

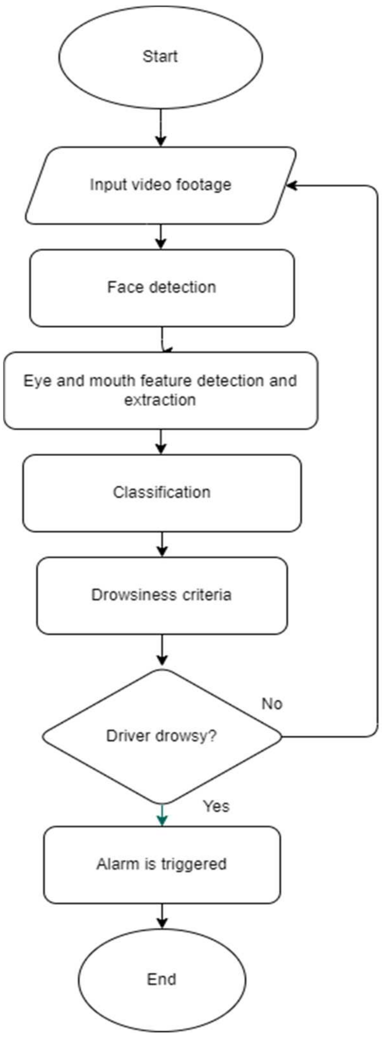

Fig. 1 Flowchart of the drowsiness detection system

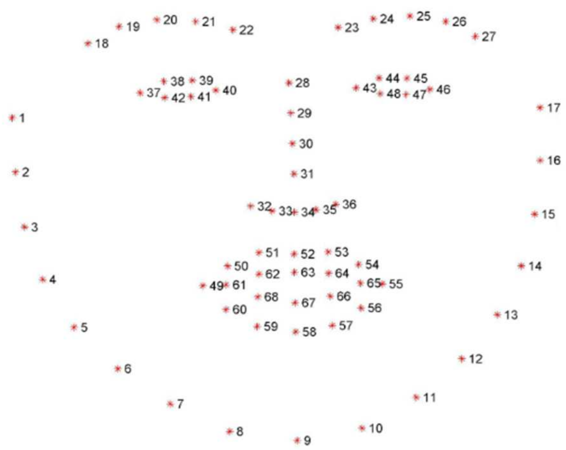

Fig. 2 Dlib's 68 Facial Landmarks [24],[25]

Consequently, the prediction CNN model will use the image dataset to determine whether the driver is yawning or not and whether their eyes are open or closed [26]. If the driver's eyes are predicted to be closed, or his mouth is predicted to be yawning, an alarm is triggered to regain the driver's concentration.

A. Convolutional Neural Network (CNN) models for Eye Classification

Figure 3 below shows the process to detect if the person's eyes are open or closed in the drowsiness detection system. First, the eye dataset from MRL is used to train the eye model. The dataset is then subjected to data pre-processing by performing a few steps, such as grayscale conversion and resizing, as the images are in different sizes. After the data pre-processing, three individual CNN models and an ensemble CNN model are built for training. Evaluation metrics are conducted to measure the classification performance [27]

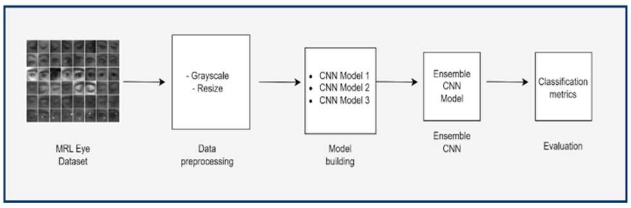

Fig. 3 Overall Process of building and training the Eyes Classification Model

B. Convolutional Neural Network (CNN) models for Mouth Yawn Classification

The process to detect if the person is yawning or not yawning in the drowsiness detection system is shown in Figure 4. First, the Yawn dataset is used to train the mouth model. The dataset is then subjected to data pre-processing by performing a few steps, such as grayscale conversion and resizing, as the images are in different sizes. Data Augmentation is required to prepare the dataset for model training as the yawn dataset obtained is relatively small. After the data pre-processing, three individual CNN models and an ensemble CNN model are built for training. Lastly, evaluation metrics are conducted to measure the classification performance.

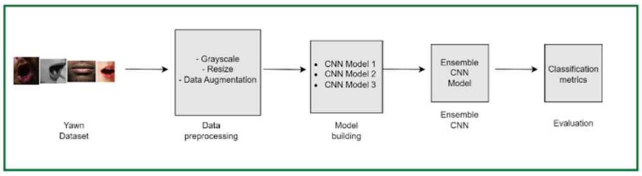

Fig. 4 Overall Process of building and training the Mouth Classification Model

C. Ensemble Learning

As mentioned in the previous section, the ensemble model is proposed as the method for the final classification models for both the eyes and mouth yawn classification. Therefore, a simple yet powerful technique for the ensemble is used, averaging the output from the individual classification models. The average prediction from all the models is used as the final prediction [28]

Figure 5 shows the overall idea of averaging the outputs from the models to obtain the final prediction. The predictions from all the models are combined to produce more accurate and generalizable results despite ensemble modeling involving multiple models, which typically results in longer training and computational times. It is possible to develop a model that can learn from the training data more effectively and produce better results using an appropriate aggregation model.

Apart from that, a single model also has issues with variance, bias, and noise. Therefore, the ensemble model can improve reliability and robustness by lowering bias and variance and giving weight to more features.

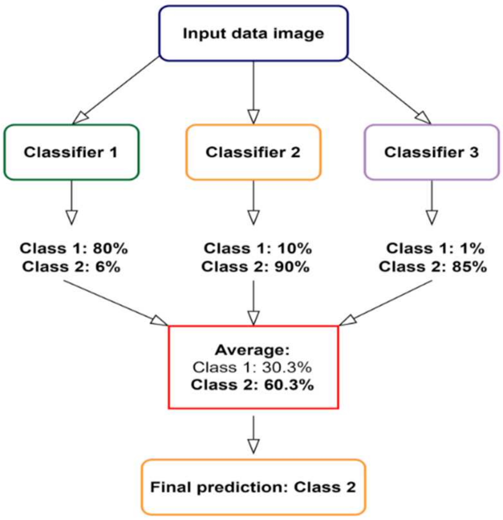

Fig. 5 Averaging the outputs from the models to obtain the final prediction

D. Data Pre-processing

The MRL Eye Dataset by Fusek [29] and Yawn Dataset [30] by Vazquez are used as the training data for the driver drowsiness prediction model. For the MRL eye dataset, the images are separated into two classes: eyes open and eyes closed. The images had been cropped only to include the section of the eyes using a histogram of oriented gradients (HOG) and the SVM classifier. A total of 76,598 infrared images is used to train the models, and the images were captured in low and high-resolution using a variety of cameras and equipment under a range of lighting conditions.

The Yawn mouth image dataset is labeled into two classes: ' Yawn' and 'No Yawn'. There are 5052 colored or grayscale images in the dataset, which have varying resolutions. The face images were retrieved from a variety of sources, including Google.

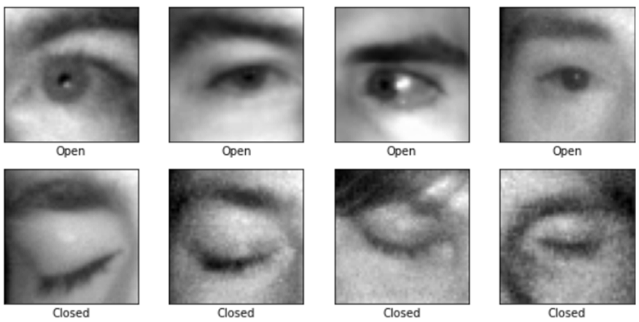

Fig. 6 Sample images of Open and Closed Eyes after the data pre-processing

Furthermore, both eyes and mouth image datasets were split into training and testing datasets with a ratio of 80:20. Next, the training and testing data were converted to grayscale, resized to 50 x 50 pixels, and the pixel values were normalized by being divided by 255 to make computations simpler and faster. Figure 6 shows Open and Closed Eye sample images after the data pre-processing. Figure 7 shows the sample images of Yawn and No Yawn after the data pre-processing.

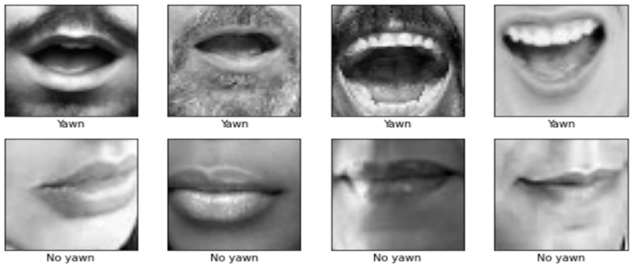

Fig. 7 Sample images of Yawn and No Yawn after the data pre-processing

E. Building of Convolutional Neural Network Models for Eye Classification

Three individual CNN models and an Ensemble CNN structure are built for the eye's classification. Figure 8 shows the ensemble CNN structure applied in this paper.

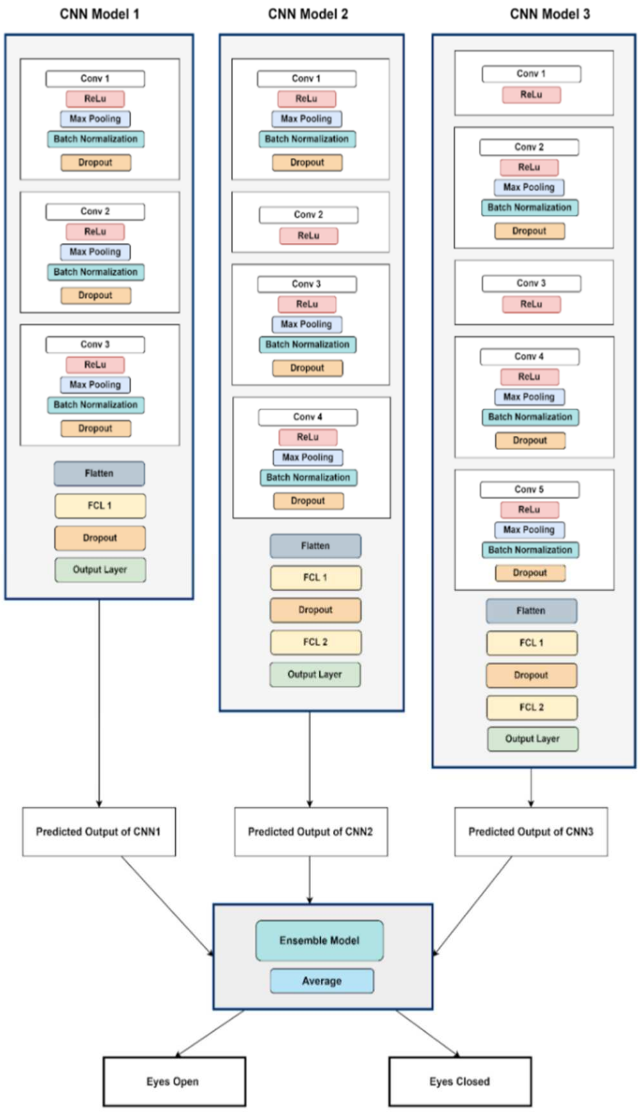

Fig. 8 Ensemble CNN structure of Eyes Classification

The first Eyes CNN model 1 structure is made up of three convolutional layers, and each layer is followed by ReLU activation function with max-pooling layers, batch normalization, and drop-out layers. Then, a flattened layer is added to convert the data into a 1-dimensional array as input to the next layer. Lastly, one fully connected layer with a drop layer is implemented to get the final output.

Next, the second Eyes CNN model 2 structure comprises four convolutional layers, each with a ReLU activation function. The first, third, and fourth layers consisted of a max pooling layer, batch normalization, and a drop-out layer. A flattened layer, two fully connected layers, and a drop-out layer are added to obtain the final output.

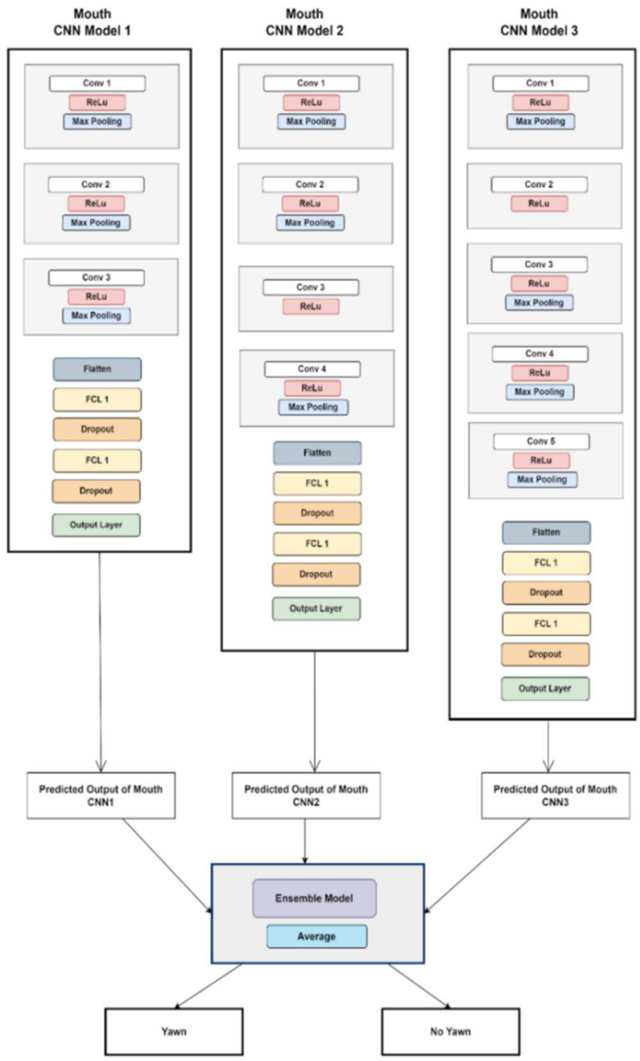

Fig. 9 Mouth Yawn Ensemble CNN structure applied to this paper

Moreover, the third Eyes CNN model 3 has five convolutional layers with a ReLu activation function for each layer. The second, fourth, and fifth layers consisted of a max pooling layer, batch normalization, and a drop-out layer. Next, a single flattened layer and two fully connected layers are the last stages of CNN in order to get the final result. Finally, all three individual CNN models are trained using the MRL eyes training dataset.

F. Building of Convolutional Neural Network Models for Mouth Yawn Classification

Three individual CNN models and an Ensemble CNN structure are built for the mouth yawn classification. Figure 9 shows the Mouth Ensemble CNN structure applied in this paper. The first Mouth CNN model 1 structure is made up of three convolutional layers, and the ReLU activation function with max-pooling layers follows each layer. A flattened layer is added to convert the data into a 1-dimensional array as input to the next layer. Lastly, two fully connected layers with a drop layer are implemented to get the final output. Next, the second Mouth CNN model 2 structure is made up of four convolutional layers where each layer has a rectified linear units (ReLU) activation function and a max pooling layer. A flattened layer and two fully connected layers followed by a drop-out layer, are added to obtain the final output.

Moreover, the third Mouth CNN model 3 has five convolutional layers with a ReLu activation function for each layer and a max pooling layer. Lastly, a single flattened layer and two fully connected layers are the last stages of CNN in order to get the final result. Next, all three individual Mouth CNN models are trained using the Yawn training dataset and loaded to perform model averaging.

III. Results and Discussion

A. Eyes Classification

The model training and validation accuracy graph and training loss and validation loss graph were plotted to verify that the models have not been overfitted. Based on the plotted graphs shown in Figures 10, 11 and 12 below, the training and validation loss increase and decrease in similar ways, and they do not differ significantly. As a result, we have concluded that three of the CNN models were not overfitting.

For the CNN Model 1, the model stopped training at the 31st Epoch as the validation loss did not improve to avoid overfitting. The training time for the CNN Model 1 is 5 minutes and 5 seconds. As for the CNN Model 2, the validation loss did not improve, and it stopped at the 20th Epoch, and training ended at 4 minutes and 4 seconds. Lastly, the training of CNN Model 3 ended at the 19th epoch, while the training duration was 4 minutes and 18 seconds.

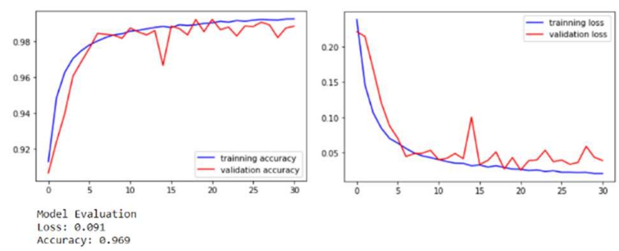

Fig. 10 Training and validation Accuracy, Training and validation loss and Model Evaluation of Eyes CNN Model 1.

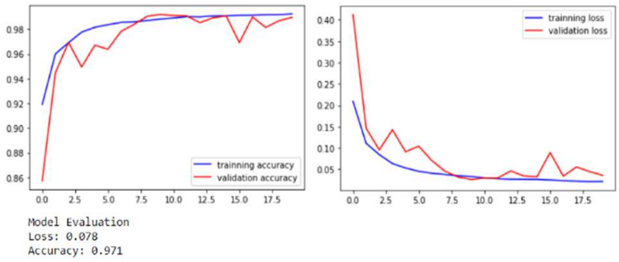

Fig. 11 Training and validation Accuracy, Training and validation loss and Model Evaluation of Eyes CNN Model 2.

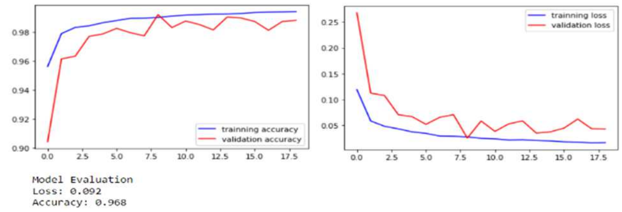

Fig. 12 Training and validation Accuracy, Training and validation loss and Model Evaluation of Eyes CNN Model 3.

TABLE I
SUMMARY EVALUATION OF EYES CLASSIFICATION MODELS

Features	Eye Model		
	CNN Model 1	CNN Model 2	CNN Model 3
Training Accuracy	0.9921	0.9923	0.9939
Validation Accuracy	0.9884	0.9896	0.9889
Evaluation Accuracy	0.9690	0.9710	0.9680
Training Loss	0.0209	0.0213	0.0160
Validation Loss	0.0393	0.0356	0.0427
Evaluation Loss	0.0910	0.0780	0.0920
Training Model Time	5 min 5s	4min 4s	4 min 18s
Training Model Epoch	31	20	19

Based on the model evaluation using the test data (Table 1), the loss for Eyes CNN Model 1 was 0.091, and the accuracy was 0.969. Moreover, a loss of 0.078 and an accuracy of 0.971 were obtained in Eyes CNN Model 2. Lastly, the loss for Eyes CNN Model 3 was 0.092, and the accuracy was 0.968. Three of the individual CNN models are loaded to perform model averaging. Using the test data to evaluate the model, a loss of 0.066 and an accuracy of 0.974 was observed. A conclusion can be made that the ensemble modeling did help to improve the overall performance of the model and overcome various issues such as generalization error and large variance.

B. *Mouth Yawn Classification*

For the Mouth CNN Model 1, the model stopped training at the 76th Epoch as the validation loss did not improve to avoid overfitting. The Mouth CNN Model 1 training time is 5 minutes and 50 seconds. As for the Mouth CNN Model 2, the validation loss did not improve, and it stopped at the 72nd Epoch, and training ended at 5 minutes and 23 seconds. Lastly, training of CNN Model 3 ended at the 64th Epoch while the training time duration is 4 minutes and 49 seconds.

The model training and validation accuracy graph and training loss and validation loss graph were plotted to verify that the models have not been overfitted. Based on the plotted graphs shown in Figures 13, 14 and 15 below, the training and validation loss increase and decrease in similar ways, and they do not differ significantly. As a result, we have concluded that three of the CNN models were not overfitting.

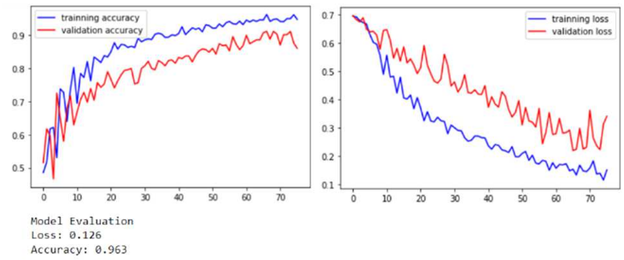

Fig. 13 Training and validation Accuracy, Training and validation loss and Model Evaluation of Mouth Yawn CNN Model 1.

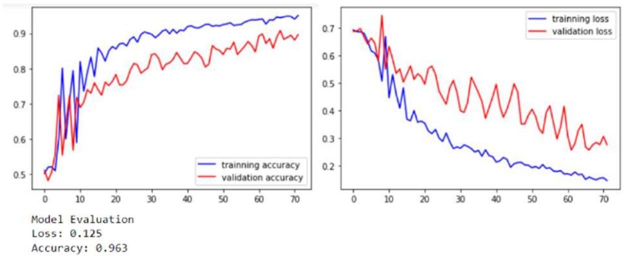

Fig. 14 Training and validation Accuracy, Training and validation loss and Model Evaluation of Mouth Yawn CNN Model 2.

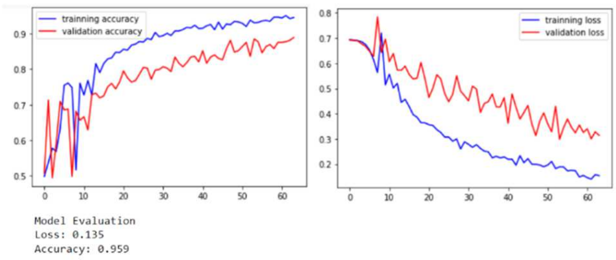

Fig. 15 Training and validation Accuracy, Training and validation loss and Model Evaluation of Mouth Yawn CNN Model 3.

Based on the model evaluation using the test data (Table 2), the loss for Mouth CNN Model 1 was 0.126, and the accuracy was 0.963. Moreover, a loss of 0.125 and an accuracy of 0.963 were obtained in Mouth CNN Model 2. Lastly, the loss for CNN Model 3 was 0.135, and the accuracy was 0.959. Three of the individual Mouth CNN models are loaded to perform model averaging. A loss of 0.122 and an accuracy of 0.965 was tested using the test data to evaluate the model.

TABLE II
SUMMARY EVALUATION OF MOUTH YAWN CLASSIFICATION MODELS

Features	Eye Model		
	CNN Model 1	CNN Model 2	CNN Model 3
Training Accuracy	0.9463	0.9507	0.9448
Validation Accuracy	0.8594	0.8965	0.8887
Evaluation Accuracy	0.9630	0.9630	0.9590

Features	Eye Model		
	CNN Model 1	CNN Model 2	CNN Model 3
Training Loss	0.1506	0.1470	0.1540
Validation Loss	0.3409	0.2770	0.3136
Evaluation Loss	0.1260	0.1250	0.1350
Training Model Time	5 min 50s	5 min 23s	4 min 49s
Training Model Epoch	76	72	64

C. *Transfer Learning*

In the analysis section, we compare the performance of different state-of-the-art methods for transfer learning using various pre-trained models such as ResNet50, VGG16, DenseNet, and our proposed ensemble model. The evaluation metrics used to assess the performance are accuracy and loss.

TABLE III
SUMMARY EVALUATION OF MOUTH YAWN CLASSIFICATION MODELS

Features	Eye Model			
	Resnet 50	VGG16	DenseNet	Proposed Ensemble Model
Evaluation Accuracy	0.890	0.500	0.852	0.965
Evaluation Loss	0.336	0.693	0.447	0.122

From the results as shown in Table 3, we can see that our proposed ensemble model achieved the highest evaluation accuracy of 0.965, indicating its effectiveness in capturing the underlying patterns in the data. ResNet50 also performed well with an accuracy of 0.890. However, VGG16 had a relatively low accuracy of 0.500, suggesting that it may not be suitable for this particular task. DenseNet achieved an accuracy of 0.852, which is better than VGG16 but lower than ResNet50 and our proposed ensemble model.

Lower loss values are generally desirable as they indicate better model performance. In this case, our proposed ensemble model achieved the lowest loss of 0.122, followed by DenseNet with a loss of 0.447. ResNet50 had a loss of 0.336, indicating its ability to minimize errors during training. VGG16 had the highest loss value of 0.693, suggesting that it may not be as effective in capturing the underlying patterns in the data.

Based on these results, it can be concluded that our proposed ensemble model outperformed the other pre-trained models in terms of both accuracy and loss. However, further analysis and investigation would be necessary to understand each model's specific strengths and weaknesses and determine the most suitable approach for the given task.

This study has significant implications for dealing with the problem of drowsy driving and traffic accidents brought on by heavy traffic in urban areas, particularly in Selangor, the capital, and the neighboring state of Malaysia. The article aims to deliver real-time alerts and prevent accidents by designing a system that can recognize driver tiredness and examine facial indicators.

The ensemble CNN models employed in this study have an accuracy rate of around 97%. This high level of accuracy indicates how well the deep learning method works at identifying the first indications of driving drowsiness. The method reduces the possibility of false alarms and increases its overall efficiency in detecting sleepy drivers with such dependable detection performance.

The real-time alarm system used in the study is a crucial component in accident prevention. It makes sure that drivers are aware of their drowsy status and can take appropriate actions to stay alert by promptly notifying them through visual and sound cues within the car. Figure 16 below displays an example graphical user interface snapshot for the drowsiness detection system.

Fig. 16 Sample screenshot of the drowsiness detection system graphical user interface

The system also sends alerts via emails and text messages to third-party contacts, such as family members or emergency services, adding an added measure of security and enabling prompt intervention. The system automatically sends an SMS text message and an email alert message with the driver's last known position to the driver's emergency contact when it repeatedly warns the motorist not to drive while intoxicated. This is done to help the authorities and those concerned about the driver in an emergency. The SMS and email alert messages sent to the emergency contact are shown in Figures 17 and Figure 18 below.

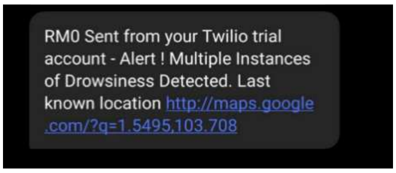

Fig. 17 SMS Alert Message to Emergency Contact

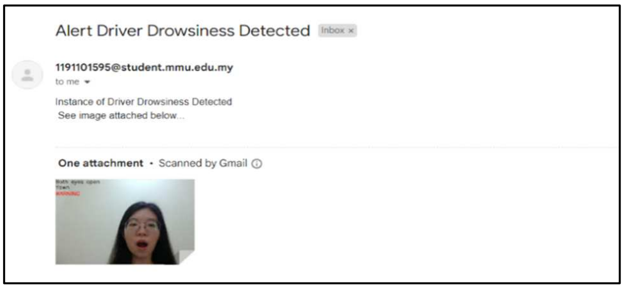

Fig. 18 Alert Message to Emergency Contact

Driver's safety may be significantly impacted by incorporating this system into automobiles and transportation systems. Implementing such technologies can drastically lower the probability of accidents brought on by fatigued driving, potentially saving lives and averting injuries. Additionally, the system's implementation in observing employees' long-distance commuters' driving habits can guarantee their safety and aid in reducing traffic collisions.

IV. Conclusion

In conclusion, developing drowsiness detection systems based on facial cues offers the potential to reduce the dangers of drowsy driving. Several challenges must be overcome to improve these systems' precision, generalizability, real-time performance, and flexibility. It is difficult to get a comprehensive dataset of drowsy drivers, so future studies should concentrate on gathering a variety of datasets with thorough annotations. Additionally, investigating a multimodal strategy that includes additional physiological and behavioral signs of sleepiness could increase detection accuracy. Future studies should focus on real-time optimization, accommodating individual variances, and integrating with vehicle systems. Long-term monitoring, predictive analytics, and real-world validation should be undertaken to further verify the viability and efficiency of sleepiness detection systems. By addressing these limitations and exploring future research directions, we can develop more robust and personalized systems that effectively prevent accidents caused by drowsy driving.

References

[1] R. Jabbar, M. Shinoy, M. Kharbeche, K. Al-Khalifa, M. Krichen, and K. Barkaoui, "Driver Drowsiness Detection Model Using Convolutional Neural Networks Techniques for Android Application," Jan. 2020, [Online]. Available: http://arxiv.org/abs/2002.03728

[2] A.-U.-I. Rafid, A. I. Chowdhury, A. R. Niloy, and N. Sharmin, "A Deep Learning Based Approach for Real-time Driver Drowsiness Detection," 2021 5th International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), Nov. 2021, doi: 10.1109/iceeict53905.2021.9667944.

[3] M. Elham Walizad, M. Hurroo, and D. Sethia, "Driver Drowsiness Detection System using Convolutional Neural Network," 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), Apr. 2022, doi: 10.1109/icoei53556.2022.9777182.

[4] Y. Suresh, R. Khandelwal, M. Nikitha, M. Fayaz, and V. Soudhri, "Driver Drowsiness Detection using Deep Learning," 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC), Oct. 2021, doi: 10.1109/icosec51865.2021.9591957.

[5] V. Chirra, S. ReddyUyyala, and V. KishoreKolli, "Deep CNN: A Machine Learning Approach for Driver Drowsiness Detection Based on Eye State," Revue d'Intelligence Artificielle, vol. 33, no. 6, pp. 461–466, Dec. 2019, doi: 10.18280/ria.330609.

[6] T. Faisal, I. Negassi, G. Goitom, M. Yassin, A. Bashir, and M. Awawdeh, "Systematic development of real-time driver drowsiness detection system using deep learning," IAES International Journal of Artificial Intelligence (IJ-AI), vol. 11, no. 1, p. 148, Mar. 2022, doi: 10.11591/ijai.v11.i1.pp148-160.

[7] M. Dua, Shakshi, R. Singla, S. Raj, and A. Jangra, "Deep CNN models-based ensemble approach to driver drowsiness detection," Neural Computing and Applications, vol. 33, no. 8, pp. 3155–3168, Jul. 2020, doi: 10.1007/s00521-020-05209-7.

[8] M. Ahmed, S. Masood, M. Ahmad, and A. A. Abd El-Latif, "Intelligent Driver Drowsiness Detection for Traffic Safety Based on Multi CNN Deep Model and Facial Subsampling," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 10, pp. 19743–19752, Oct. 2022, doi: 10.1109/tits.2021.3134222.

[9] S. Park, F. Pan, S. Kang, and C. D. Yoo, "Driver drowsiness detection system based on feature representation learning using various deep networks."

[10] V. Kumar, S. Sharma, and Ranjeet, "Driver drowsiness detection using modified deep learning architecture," Evolutionary Intelligence, vol. 16, no. 6, pp. 1907–1916, Jun. 2022, doi: 10.1007/s12065-022-00743-w.

[11] J.-M. Guo and H. Markoni, "Driver drowsiness detection using hybrid convolutional neural network and long short-term memory," Multimedia Tools and Applications, vol. 78, no. 20, pp. 29059–29087, Jul. 2018, doi: 10.1007/s11042-018-6378-6.

[12] I. Journal and R. V K, "IRJET-Drowsy Driving Detection via Hybrid Deep Learning," *International Research Journal of Engineering and Technology*, 2020, [Online]. Available: www.irjet.net

[13] M. Öztürk, A. Küçükmanisa, and O. Urhan, "Drowsiness Detection System Based on Machine Learning Using Eye State," Balkan Journal of Electrical and Computer Engineering, vol. 10, no. 3, pp. 258–263, Jul. 2022, doi: 10.17694/bajece.1028110.

[14] S. Mehta, P. Mishra, A. J. Bhatt, and P. Agarwal, "AD3S: Advanced Driver Drowsiness Detection System using Machine Learning," 2019 Fifth International Conference on Image Information Processing (ICIIP), Nov. 2019, doi: 10.1109/iciip47207.2019.8985844.

[15] C.-Y. Lin, P. Chang, A. Wang, and C.-P. Fan, "Machine Learning and Gradient Statistics Based Real-Time Driver Drowsiness Detection," 2018 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), May 2018, doi: 10.1109/icce-china.2018.8448747.

[16] S. Dey, S. A. Chowdhury, S. Sultana, Md. A. Hossain, M. Dey, and S. K. Das, "Real Time Driver Fatigue Detection Based on Facial Behaviour along with Machine Learning Approaches," 2019 IEEE International Conference on Signal Processing, Information, Communication & Systems (SPICSCON), Nov. 2019, doi: 10.1109/spicscon48833.2019.9065120.

[17] Q. Cheng, W. Wang, X. Jiang, S. Hou, and Y. Qin, "Assessment of Driver Mental Fatigue Using Facial Landmarks," IEEE Access, vol. 7, pp. 150423–150434, 2019, doi: 10.1109/access.2019.2947692.

[18] C. B. S. Maior, M. J. das 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, doi: 10.1016/j.eswa.2020.113505.

[19] E. Magán, M. P. Sesmero, J. M. Alonso-Weber, and A. Sanchis, "Driver Drowsiness Detection by Applying Deep Learning Techniques to Sequences of Images," Applied Sciences, vol. 12, no. 3, p. 1145, Jan. 2022, doi: 10.3390/app12031145.

[20] M. Hashemi, A. Mirrashid, and A. Beheshti Shirazi, "Driver Safety Development: Real-Time Driver Drowsiness Detection System Based on Convolutional Neural Network," SN Computer Science, vol. 1, no. 5, Aug. 2020, doi: 10.1007/s42979-020-00306-9.

[21] X.-H. Lam, K.-W. Ng, Y.-J. Yoong, and S.-B. Ng, "WBC-based segmentation and classification on microscopic images: a minor improvement," F1000Research, vol. 10, p. 1168, Nov. 2021, doi:10.12688/f1000research.73315.1.

[22] J.-S. Ang, "Stock Market Prediction Using Deep Learning Approach," 2022.

[23] Y. Lim, K.-W. Ng, P. Naveen, and S.-C. Haw, "Emotion Recognition by Facial Expression and Voice: Review and Analysis," Journal of Informatics and Web Engineering, vol. 1, no. 2, pp. 45–54, Sep. 2022, doi: 10.33093/jiwe.2022.1.2.4.

[24] Adrian Rosebrock, "Facial landmarks with dlib, OpenCV, and Python." Accessed: Oct. 26, 2023. [Online]. Available: https://pyimagesearch.com/2017/04/03/facial-landmarks-dlib-opencv-python/

[25] Adrian Rosebrock, "Convolutional Neural Networks (CNNs) and Layer Types." Accessed: Oct. 26, 2023. [Online]. Available: https://pyimagesearch.com/2021/05/14/convolutional-neural-networks-cnns-and-layer-types/

[26] R. M. Salman, M. Rashid, R. Roy, M. M. Ahsan, and Z. Siddique, "Driver Drowsiness Detection Using Ensemble Convolutional Neural Networks on YawDD," Dec. 2021, [Online]. Available: http://arxiv.org/abs/2112.10298

[27] Z. Zhao, N. Zhou, L. Zhang, H. Yan, Y. Xu, and Z. Zhang, "Driver Fatigue Detection Based on Convolutional Neural Networks Using EM-CNN," Computational Intelligence and Neuroscience, vol. 2020, pp. 1–11, Nov. 2020, doi: 10.1155/2020/7251280.

[28] M. F. Shakeel, N. A. Bajwa, A. M. Anwaar, A. Sohail, A. Khan, and Haroon-ur-Rashid, "Detecting Driver Drowsiness in Real Time Through Deep Learning Based Object Detection," Lecture Notes in Computer Science, pp. 283–296, 2019, doi: 10.1007/978-3-030-20521-8_24.

[29] Radovan and Fusek, "MRL Eye Dataset." Accessed: Oct. 26, 2023. [Online]. Available: http://mrl.cs.vsb.cz/eyedataset

[30] David Vazquez, "Yawn Dataset," 2021, Accessed: Oct. 26, 2023. Available: https://www.kaggle.com/datasets/davidvazquezcic/yawn-dataset.