# Drowsiness Detection Using Transfer Learning and Mobile Net as a Base Model

Sadir Ahmed Zidan, Md. Raufur Rahim, Md. Shahadot Hossain Shakib, Abu Shaleh Md. Kaium
*Department of Computer Science and Engineering, American International University Bangladesh*

*Abstract*—**Drowsiness is a state of strong inclination to sleep and poses a significant risk to safety, in activities such as driving or operating heavy machinery.** It **can significantly affect cognitive abilities such as attention, concentration, and decision-making. It can lead to decreased alertness, slower reaction times, and reduced ability to process information effectively. Drowsy driving accidents have serious repercussions including the possibility of fatal injuries and substantial economic losses, so to prevent mishaps and secure the health of individuals it's crucial to identify and alleviate sleepiness in real-time. Machine learning techniques like transfer learning and deep neural networks are facilitating new prospects for detecting drowsiness in recent times and by using Mobile Net architecture as a base model and incorporating transfer learning method we are aiming to construct an efficient and accurate drowsiness detection system. In this paper we have used eye images for dataset as two states for open and close eyes of 37 individuals and using Mobile Net as a base model and some custom layers are added to adapt the features. The result shows that the model works quite well in this dataset for both training and validation data and got a good accuracy in test data or new data. Also got good accuracy in real time detection using webcam.**

*Keywords*—**Drowsiness detection; Mobile Net; Object Detection; Transfer Learning**

## I. INTRODUCTION

A person can be identified as drowsy by using some techniques like eyes detection, yawning, and nodding. But some people can sleep without yawning and nodding. Another method is to use physiological sensors like biosensors. Here the disadvantages are like the driver may hesitate to wear them or he may forget to wear them. That could be a reason for why, detecting drowsiness through eye detection is best compared to the remaining techniques.[2]

Drowsiness can hinder performance in various domains, including academic, work, and sports. It can impact productivity, learning, and skill execution. Tasks that require focus, precision, and quick thinking may be compromised, leading to suboptimal outcomes. The aim of developing this model is to detect drowsiness in individuals using eye images taken by a camera or other imagining devices. And we used a dataset [1] for eye pictures of close and open eyes and Haar Cascade Classifier for detecting face and eye frame. The model utilizes a pre-trained mobile Net architecture (which is designed to be lightweight and efficient for mobile and embedded devices) and fine-tunes the model on the dataset. The model will learn to distinguish between open and closed eyes and predict whether a person is drowsy or not. While deploying the model, it will take input from the camera as a continuous stream of images, process it and output the prediction in real time.

The objectives of our system are:

- To develop a deep learning model based on the Mobile Net architecture that can classify between open and closed eyes for the purpose of detecting drowsiness in drivers.
- To evaluate the performance of the Mobile Net model in terms of accuracy, precision, recall, and F1 score.
- To investigate the impact of various hyperparameters such as learning rate, regularization rate, batch size on the performance of the proposed model for drowsiness detection.
- To explore the feasibility of real-time drowsiness detection.

## II. RELATED WORK

The goal of driver drowsiness detection research is to reduce traffic accidents brought on by drowsy driving. To create efficient drowsiness detection systems, different approaches and techniques have been investigated. Eyelid closure is one of the most accurate signs of sleepiness. To record face and eye movements for drowsiness analysis, researchers have used various kinds of camera systems, including normal cameras, infrared (IR) cameras, and stereo cameras. Some research has used convolutional neural networks (CNN) to identify drowsiness. These CNN-based models learn and represent drowsiness-related characteristics using deep learning techniques. One such model has a drowsiness detection accuracy of 78%. Based on eye closure patterns, one approach is used to determine drowsiness. It is known as PERCLOS (percentage of eye closure over time). These technologies can detect when a driver is about to go to sleep and sound alerts to warn them by monitoring eye movements. The detection of driver drowsiness has been suggested with vision-based approaches that use technologies such as AHOG, SVM, and OpenCV. These techniques have produced positive outcomes in terms of identifying drowsiness.[3]

The development of effective drowsiness detection systems is necessary since driver drowsiness plays a significant role in both industrial and car accidents. Among other approaches, using eyelid movement patterns gathered by Electrooculogram (EOG) data has shown promise. The study's findings showed how the SVM model may be used to determine sleepiness levels with accuracy. So, the model was able to correctly identify every case of "very sleepy" scenarios. However, there were 16 false alarms, or a certain number of false alarms. The results of this study offer insightful information on the efficacy of an SVM model for drowsiness detection that makes use of characteristics linked to eyelid movement. The high accuracy in 'very sleepy' condition classification and good performance in "sleepy" state detection point to the validity of this technique. To improve the applicability of the developed technique, the study first acknowledged the necessity of integrating data from alert driving circumstances. It's important to remember that research using a simulator cannot be an exact copy of real-world driving situations. The study utilized several measures, such as the Karolinska Sleepiness Scale (KSS), Karolinska Drowsiness Score (KDS), and hits of the strip, as references for drowsiness levels.[4]

Drowsy driving has a significant impact on road safety because it increases the risk of accidents. Researchers are looking at electroencephalography (EEG) signals as a potential way to identify driver drowsiness to address this problem. The goal of this literature review is to provide an overview of the developments and discoveries in EEG-based drowsy driver identification. In several studies, EEG data from drivers in simulated driving environments—like virtual reality or computer-based driving games—has been the main subject of study. Preprocessing of the gathered data included filtering and labeling of alertness and drowsiness states. Extracting features from EEG data has been the first step in creating efficient drowsy driver detection systems. Statistical tests, such as computing p-values, have been used to assess the importance of the retrieved characteristics. The outcomes of these assessments have demonstrated that the retrieved EEG characteristics, notably the logarithm of energy, showed remarkable capacities to distinguish between alert and drowsy states. These results show the potential for drowsy driver detection systems based on EEG to increase road safety by recognizing tired drivers and preventing accidents.[5]

The face-based approach and the ROI-based approach are the two methods under consideration. Additionally, the categorization of driving behavior and the drowsiness confirmation model are covered. The model used in the face-based drowsiness detection method is trained using data from 16 participants from the NTHU. When using different optimizers for situations including raw and pre-processed face frames, the best results are obtained with 256 neurons in the FC layer. The noise produced by the face's surroundings causes the ROI-based strategy to perform better than the face-based approach. The Kaggle eye and mouth dataset is used to train classifiers for the eye and mouth ROI-based technique. The Kaggle dataset has the best classification accuracy results, with the eye and mouth classifiers performing better than those using the NTHU dataset. The Kaggle dataset makes it easier to train the full ROI, which promotes model performance and detection accuracy. The LSTM algorithm is used in the categorization of the driving behavior-drowsiness model. The drowsy class performs slightly lower than the not-drowsy class when it comes to both recall and F1-score. Macro-Average and Weighted-Average metrics are used to evaluate precision, recall, and F1-score, with comparable precision values but different recall and F1-score results. Particularly when trained on the Kaggle dataset, the eye and mouth classifiers achieve great accuracy.[6]

In this study of the literature, we will look at the drowsiness detection methods and research that has already been done, with a focus on the model that was put out in the paper. Real-time eye movement analysis makes it feasible to spot tiredness early on and launch efficient treatments. Major automakers like Volvo and Mercedes-Benz have already installed drowsiness detecting technologies in their cars. These technologies are less widely available to the public due to limitations, including high prices and implementation ineffectiveness. As a result, there is a rising demand for solutions that are both accessible and readily adopted. The device wants to identify early indicators of fatigue and prevent possible accidents by continually taking eye scans and using image processing algorithms. An IoT-based monitoring system and LabVIEW programming are both used to make this possible. The system's coding was developed in LabVIEW, enabling the ongoing collection and examination of eye pictures. Drowsiness is recognized, and an alarm is set off to warn the driver of their level of exhaustion. The system offers real-time monitoring capabilities and prompt alarm production through the integration of IoT technology with LabVIEW programming, ensuring quick action during instances of driver intoxication.[7]

## III. Proposed model

In this study, we propose a novel deep learning model for drowsiness detection based on the Mobile Net architecture. The objective of our model is to accurately classify eye images and detect signs of drowsiness in real-time. Drowsiness detection is a critical task in various domains, including transportation safety, workplace productivity, and healthcare.

Our proposed model leverages the powerful features of the Mobile Net architecture, which is a lightweight and efficient convolutional neural network (CNN) designed for mobile and embedded devices. By using the Mobile Net as the base model, we benefit from its pre-trained weights on the ImageNet dataset, which enables effective feature extraction from eye images. To adapt the base model for drowsiness detection, we introduce additional layers on top of the Mobile Net. The output from the base model is fed into a Flatten layer, which converts the multi-dimensional feature maps into a one-dimensional array. This allows for efficient information aggregation.

| Type / Stride | Filter Shape | Input Size |
|---|---|---|
| Conv / s2 | $3 \times 3 \times 3 \times 32$ | $224 \times 224 \times 3$ |
| Conv dw / s1 | $3 \times 3 \times 32$ dw | $112 \times 112 \times 32$ |
| Conv / s1 | $1 \times 1 \times 32 \times 64$ | $112 \times 112 \times 32$ |
| Conv dw / s2 | $3 \times 3 \times 64$ dw | $112 \times 112 \times 64$ |
| Conv / s1 | $1 \times 1 \times 64 \times 128$ | $56 \times 56 \times 64$ |
| Conv dw / s1 | $3 \times 3 \times 128$ dw | $56 \times 56 \times 128$ |
| Conv / s1 | $1 \times 1 \times 128 \times 128$ | $56 \times 56 \times 128$ |
| Conv dw / s2 | $3 \times 3 \times 128$ dw | $56 \times 56 \times 128$ |
| Conv / s1 | $1 \times 1 \times 128 \times 256$ | $28 \times 28 \times 128$ |
| Conv dw / s1 | $3 \times 3 \times 256$ dw | $28 \times 28 \times 256$ |
| Conv / s1 | $1 \times 1 \times 256 \times 256$ | $28 \times 28 \times 256$ |
| Conv dw / s2 | $3 \times 3 \times 256$ dw | $28 \times 28 \times 256$ |
| Conv / s1 | $1 \times 1 \times 256 \times 512$ | $14 \times 14 \times 256$ |
| $5\times$ Conv dw / s1 | $3 \times 3 \times 512$ dw | $14 \times 14 \times 512$ |
| $5\times$ Conv / s1 | $1 \times 1 \times 512 \times 512$ | $14 \times 14 \times 512$ |
| Conv dw / s2 | $3 \times 3 \times 512$ dw | $14 \times 14 \times 512$ |
| Conv / s1 | $1 \times 1 \times 512 \times 1024$ | $7 \times 7 \times 512$ |
| Conv dw / s2 | $3 \times 3 \times 1024$ dw | $7 \times 7 \times 1024$ |
| Conv / s1 | $1 \times 1 \times 1024 \times 1024$ | $7 \times 7 \times 1024$ |
| Avg Pool / s1 | Pool $7 \times 7$ | $7 \times 7 \times 1024$ |
| FC / s1 | $1024 \times 1000$ | $1 \times 1 \times 1024$ |
| Softmax / s1 | Classifier | $1 \times 1 \times 1000$ |

**Fig. 1.** Mobile Net Architecture.

We then incorporate a fully connected layer with 64 units and apply the rectified linear unit (ReLU) activation function. This dense layer enables learning of higher-level representations and helps capture complex patterns in the eye images. To prevent overfitting, we apply L2 regularization with a strength of 0.001 to the weights of this dense layer. To further enhance the generalization capability of our model, we include a dropout layer with a dropout rate of 0.5. This layer randomly deactivates a portion of the units during training, reducing interdependencies between neurons and enhancing the model's ability to generalize to new, unseen eye images.

The final layer of our model consists of a dense layer with a sigmoid activation function, producing a single output representing the probability of drowsiness. This allows us to classify each eye image as either indicating drowsiness or not. To ensure the effectiveness of our proposed model, we freeze the weights of the base Mobile Net layers during training. By doing so, we retain the knowledge encoded in the pre-trained Mobile Net weights and focus on fine-tuning the additional layers specific to drowsiness detection.

TABLE I
PROPOSED MODEL ARCHITECTURE

| Layer (type) | Output Shape | Param # |
|---|---|---|
| input_1 (Input Layer) | [(None, 80, 80, 3)] | 0 |
| mobilenet_1.00_224 (Functional) | (None, 3, 3, 1024) | 3228864 |
| flatten (Flatten) | (None, 9216) | 0 |
| dense (Dense) | (None, 64) | 589888 |
| dropout (Dropout) | (None, 64) | 0 |
| dense_1 (Dense) | (None, 1) | 65 |

Total params: 3,818,817
Trainable params: 590,953
Non-trainable params: 3,227,864

The model architecture combines the power of the Mobile Net base model for feature extraction with additional dense layers for higher-level representation learning and classification. The freezing of the base model layers allows for transfer learning, utilizing the pre-trained weights to improve performance on the drowsiness detection task.

For training our model, we utilize a dataset of eye images specifically collected for drowsiness detection. The dataset is divided into training and validation sets, enabling us to monitor the model's performance and prevent overfitting. Eye images of 37 candidates, Split them to open eyes and close eyes folder. Then convert them to train and test data folder. From the total, 80% data used for training and validation and 20% used for testing and evaluation. Also used the pre-trained Haar Cascade for detecting face and eyes while taking real time image stream for detecting drowsiness.

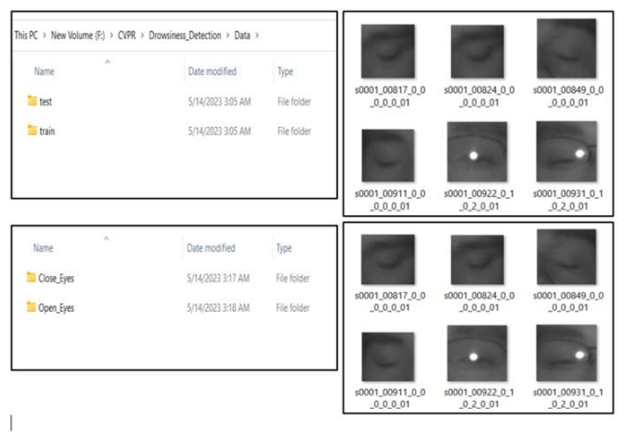


**Fig. 2.** Example of dataset and folder format

We employ the Adam optimizer, a popular optimization algorithm known for its efficiency and robustness. The model is trained using binary cross-entropy loss, which is suitable for binary classification tasks like drowsiness detection. The accuracy metric is used to evaluate the performance of the model during training and validation.

To prevent overfitting and ensure early stopping if the model's performance plateaus, we incorporate several callback functions. These include model checkpointing to save the best performing model based on validation loss, early stopping to halt training if the validation loss does not improve for a specified number of epochs, and dynamic learning rate adjustment using ReduceLROnPlateau to fine-tune the learning process. We train the proposed model over a fixed number of epochs, with mini-batch training using a batch size of 8. The steps per epoch and validation steps are determined based on the dataset size and batch size to ensure comprehensive coverage of the data during training and validation. The proposed model aims to provide an accurate and efficient solution for real-time drowsiness detection using eye images. Through its unique architecture and training methodology.

## IV. Result

Upon completion of training, we evaluate the performance of our proposed model on the validation dataset. Also, on training and test data set. This includes computing metrics such as accuracy and loss, precision, recall, and F1-score, as well as generating a confusion matrix to assess the model's ability to correctly classify eye images. After evaluating the performance of the trained model on validation, training, and test data we found the following results.

The model achieved an accuracy of 0.87 on the validation dataset, meaning it correctly identified drowsiness or alertness with an 87% accuracy rate. The loss value for the validation dataset was 0.30, which represents the error of the model during training. During the training phase, the model achieved a higher accuracy of 0.94, indicating its ability to learn and classify drowsiness accurately. The corresponding loss value for the training dataset was 0.17, suggesting that the model effectively minimized errors while learning from the training data. Lastly, the model was evaluated on an independent test dataset, where it achieved an accuracy of 0.95, indicating its ability to generalize well to unseen data. The loss value for the test dataset was 0.16, further affirming the model's robustness and low error rate.

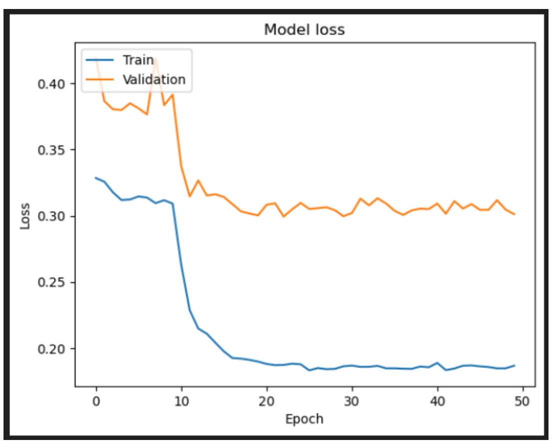


**Fig. 3.** Model loss on train and validation data

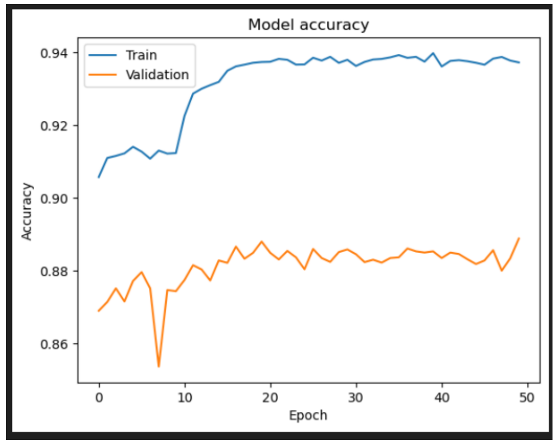


**Fig. 4.** Model accuracy on train and validation data

TABLE II
EVALUATION REPORT

| Data Set | Accuracy | Loss |
|---|---|---|
| Validation | 0.87 | 0.30 |
| Training | 0.94 | 0.17 |
| Test | 0.95 | 0.16 |

We define a confusion matrix to evaluate overall the performance of the model and to know the number of correct and incorrect predictions for each class. We can analyze the strengths and weaknesses of the model after determining the confusion matrix and its classification report.

TABLE III
CLASSIFICATION REPORT

| | **Predicted Negative** | **Predicted Positive** |
|---|---|---|
| Actual Negative | 4057 | 137 |
| Actual Positive | 292 | 4003 |

In the table, the rows represent the actual classes (negative and positive), while the columns represent the predicted classes (negative and positive). The values in the table represent the counts of instances falling into each category:

- True Negative (TN): 4057
- False Positive (FP): 137
- False Negative (FN): 292
- True Positive (TP): 4003

These values provide a breakdown of the correct and incorrect predictions made by the model for each class.

TABLE IV
CLASSIFICATION REPORT ON DIFFERENT MATRIX

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Close Eyes | 0.93 | 0.97 | 0.95 | 4194 |
| Open Eyes | 0.97 | 0.93 | 0.95 | 4295 |

In the table, each row represents a class, and the columns provide the precision, recall, F1-score, and support values for that class. The "Close Eyes" class has a precision of 0.93, recall of 0.97, F1-score of 0.95, and a support of 4194. Similarly, the "Open Eyes" class has a precision of 0.97, recall of 0.93, F1-score of 0.95, and a support of 4295.

The classification report describes the following:

**Precision**: Precision is a measure of the accuracy of positive predictions. For the "Close Eyes" class, the precision is 0.93, which means that 93% of the predicted "Close Eyes" instances are correct. For the "Open Eyes" class, the precision is 0.97, indicating that 97% of the predicted "Open Eyes" instances are correct.

**Recall**: Recall, also known as sensitivity or true positive rate, measures the proportion of actual positive instances that are correctly identified by the model. For the "Close Eyes" class, the recall is 0.97, meaning that 97% of the actual "Close Eyes" instances are correctly classified. For the "Open Eyes" class, the recall is 0.93, indicating that 93% of the actual "Open Eyes" instances are correctly classified.

**F1-Score**: The F1-score is the harmonic mean of precision and recall. It provides a balance between precision and recall. For both classes, the F1-score is 0.95, indicating a good balance between precision and recall.

**Support:** The support refers to the number of instances of each class in the test dataset. For the "Close Eyes" class, there are 4194 instances, and for the "Open Eyes" class, there are 4295 instances.

**Accuracy**: The overall accuracy of the model is 0.95, meaning that it correctly predicts the class of 95% of the instances in the test dataset.

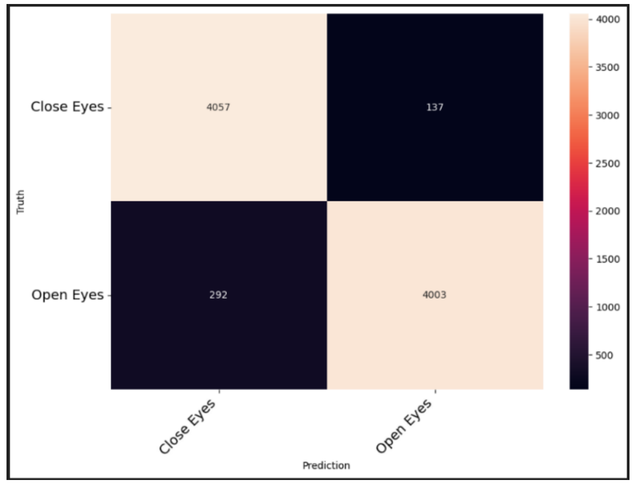

**Fig. 5.** Confusion Matrix

From the confusion matrix and classification report, it can be concluded that the model performs well in distinguishing between "Close Eyes" and "Open Eyes" classes. It achieves high precision, recall, and F1-score for both classes, indicating a balanced and accurate classification performance.

**Real Time Detection**

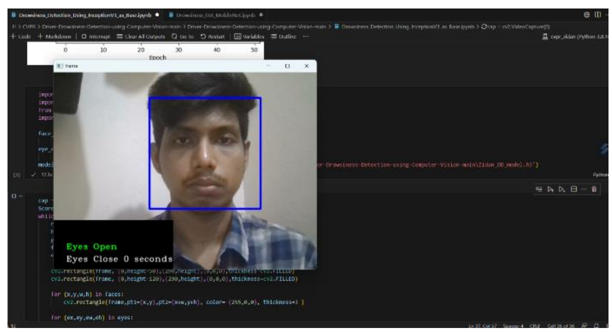

**Fig. 6.** Open Eyes detected.

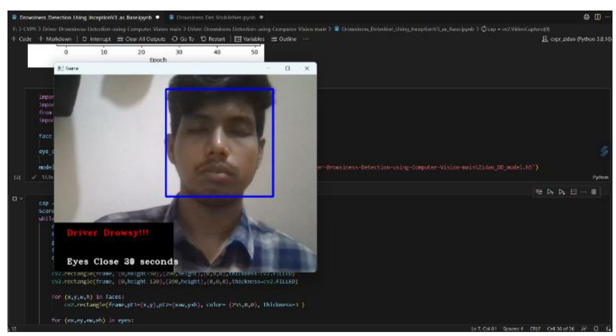

**Fig. 7.** Eyes were closed for more than 5 seconds.

## V. DISCUSSION AND CONCLUSION

In this study, we developed a model based on Mobile Net architecture for drowsiness detection using eye images. The model demonstrated promising results, achieving an accuracy of 95% on the test dataset. One of the key factors contributing to the success of our model is the utilization of transfer learning. By leveraging the pre-trained weights of the Mobile Net model trained on the ImageNet dataset, we were able to benefit from its learned features and effectively extract relevant information from the eye images. This enabled us to overcome the limitations of training a model from scratch with limited data, which often leads to overfitting.

Furthermore, the use of regularization techniques such as L2 regularization and dropout helped prevent overfitting and improve the generalization ability of the model. The L2 regularization controlled the complexity of the model by penalizing large weights, while the dropout layer acted as a regularize function by randomly dropping out a fraction of the neurons during training, thus reducing the interdependencies between neurons.

The obtained confusion matrix and classification report further support the effectiveness of our model. The high precision and recall values for both classes (Close Eyes and Open Eyes) indicate that the model can correctly classify instances from these classes with high accuracy. The overall F1-score of 0.95 indicates a balanced performance between precision and recall.

However, there are still areas for improvement and future research. Firstly, the model's performance could be further enhanced by exploring other architectures, such as Inception or ResNet, and fine-tuning their parameters. Additionally, the inclusion of more diverse and extensive datasets would help improve the model's robustness and generalization ability. Moreover, it is crucial to consider real-time implementation and deployment of the model. Further research should focus on optimizing the model's efficiency and computational requirements to ensure its suitability for real-time drowsiness detection systems.

## References

[1] "MRL Eye Dataset | MRL," mrl.cs.vsb.cz. http://mrl.cs.vsb.cz/eyedataset

[2] A. K, "Drowsiness Detection System," Analytics Vidhya, May 31, 2022. https://www.analyticsvidhya.com/blog/2022/05/drowsiness-detection-system/

[3] Md. T. A. Dipu, S. S. Hossain, Y. Arafat, and F. B. Rafiq, "Real-time Driver Drowsiness Detection using Deep Learning," International Journal of Advanced Computer Science and Applications, vol. 12, no. 7, 2021, doi: https://doi.org/10.14569/ijacsa.2021.0120794.

[4] S. Hu and G. Zheng, "Driver drowsiness detection with eyelid related parameters by Support Vector Machine," Expert Systems with Applications, vol. 36, no. 4, pp. 7651–7658, May 2009, doi: https://doi.org/10.1016/j.eswa.2008.09.030.

[5] M. Mikaili, Z. Mardi, and S. N. Ashtiani, "EEG-based drowsiness detection for safe driving using chaotic features and statistical tests," Journal of Medical Signals & Sensors, vol. 1, no. 2, p. 130, 2011, doi: https://doi.org/10.4103/2228-7477.95297.

[6] Hanane Lamaazi, A. Alqassab, R. Fadul, and Rabeb Mizouni, "Smart Edge-Based Driver Drowsiness Detection in Mobile Crowdsourcing," IEEE Access, vol. 11, pp. 21863–21872, Jan. 2023, doi: https://doi.org/10.1109/access.2023.3250834.

[7] "Drowsy Driver Warning for Accident-Avoidance System Using Image Processing," ResearchGate. Available: [Online]. Accessed: May 16, 2023.