Real-Time Driver Drowsiness Detection System

Minor Project (CC3270)

Report

Submitted in the partial fulfillment of the requirement for the award of

Bachelor of Technology

in

Computer and Communication Engineering

By:

Sanket Deb
209303135

Under the guidance of:

Dr. Somya Goyal

MANIPAL UNIVERSITY JAIPUR

April 2023

Department of Computer and Communication Engineering
School of Computer and Communication Engineering
Manipal University Jaipur
VPO. Dehmi Kalan, Jaipur, Rajasthan, India – 303007

Department of Computer and Communication Engineering
School of Computer and Communication Engineering, Manipal University Jaipur,
Dehmi Kalan, Jaipur, Rajasthan, India- 303007

STUDENT DECLARATION

I hereby declare that this project Real-Time Driver Drowsiness Detection System is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma of the University or other Institute, except where due acknowledgment has been made in the text.

Place: Manipal University Jaipur
Date: 19.04.2023

Sanket Deb
209303135
B.Tech (CCE) 6th Semester

Department of Computer and Communication Engineering
School of Computing and Communication Engineering, Manipal University Jaipur,
Dehmi Kalan, Jaipur, Rajasthan, India- 303007

Date: 19.04.23

CERTIFICATE FROM GUIDE

*This is to certify that the work entitled "**Real-Time Driver Drowsiness Detection System"** submitted by **Sanket Deb** (209303135) to **Manipal University Jaipur** for the award of the degree of **Bachelor of Technology** in **Computer and Communication Engineering** is a bonafide record of the work carried out by him/ her under my supervision and guidance from January 2023 to April 2023.*

Dr. Somya Goyal
Department of Computer and Communication Engineering
Manipal University Jaipur

ACKNOWLEDGEMENT

This project would not have been completed without the help, support, comments, advice, cooperation, and coordination of various people. However, it is impossible to thank everyone individually; I am hereby making a humble effort to thank some of them. I acknowledge and express my deepest sense of gratitude to my internal supervisor Dr. Somya Goyal for her constant support, guidance, and continuous engagement. I highly appreciate her technical comments, suggestions, and criticism during the progress of this project “Real Time Driver Drowsiness Detection System”.

I owe my profound gratitude to Dr. V S Dhaka Director, School of Computer and Communication Engineering, for her valuable guidance and facilitating me during my work. I am also very grateful to all the faculty members and staff for their precious support and cooperation during the development of this project.

Finally, I extend my heartfelt appreciation to my classmates and peers for their help and encouragement.

Sanket Deb

209303135

ABSTRACT

Driver drowsiness is a significant cause of road accidents, which can result in injuries and fatalities. With the advancement in technology, there is a growing interest in developing systems that can detect driver drowsiness in real-time. In this project, we propose to study and analyze the different methods of drowsiness detection especially focusing on image based measures where techniques such as convolutional neural networks (CNNs) are used to detect driver drowsiness. The proposed system will use a well-balanced dataset to train a CNN model. The CNN model is trained on a large dataset of images to accurately classify the driver's state as alert or drowsy. The system can be designed to provide real-time alerts to the driver when drowsiness is detected, allowing them to take corrective action. Our results show that the proposed system achieves high accuracy in detecting driver drowsiness, and thus has the potential to significantly reduce the number of accidents caused by driver fatigue.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Driver drowsiness is a major concern worldwide, and it can have severe consequences if not addressed in time. According to the World Health Organization (WHO), approximately 1.35 million people die each year as a result of road traffic accidents, and between 20% to 30% of these accidents are caused by driver fatigue. Furthermore, driver fatigue has been identified as a significant cause of road accidents in various countries, specially developed and developing nations.

Driver drowsiness is characterized by a decreased ability to maintain alertness and concentration while driving, leading to slower reaction times, reduced vigilance, and impaired judgment. This can result in lane departures, erratic driving, and an increased risk of accidents. Therefore, detecting and preventing driver drowsiness is crucial in ensuring road safety.

In recent years, various technologies and methods have been developed to detect driver drowsiness, including physiological measures, vehicle-based measures, and visual-based measures. These technologies range from wearable sensors to eye-tracking systems to machine learning algorithms. Physiological measures, such as electroencephalography (EEG), electrooculography (EOG), and heart rate variability (HRV), are commonly used to detect changes in the driver's physiological signals that indicate drowsiness. Vehicle-based measures, such as steering wheel movements, vehicle speed, and lane deviation, are used to detect changes in driving behavior that indicates drowsiness. Visual-based measures, such as eye movements and facial expressions, are used to detect changes in the driver's visual attention that indicate drowsiness.

Despite the advancements in driver drowsiness detection technologies, there are still challenges that need to be addressed. For instance, many of the current detection systems are prone to false alarms or are affected by external factors such as lighting conditions and weather. Additionally, there is a need for more accurate and reliable methods that can detect drowsiness in real-time, as well as methods that can adapt to different driving conditions and individual drivers.

Therefore, this project aims to review the current state of driver drowsiness detection research, including the different techniques used and their effectiveness in real-world situations. Additionally, this project proposes a novel method for driver drowsiness detection that combines technology with medical history to improve accuracy and reliability. The proposed method uses machine learning algorithms, to detect driver drowsiness in real time. The system is designed to be adaptive to different driving conditions and individual drivers, and it is evaluated using a dataset of real-world driving scenarios.

In summary, this project will contribute to the growing body of literature on driver drowsiness detection by providing a comprehensive review of the current state of research and proposing a novel approach to improve the accuracy and reliability of these systems. The goal is to promote the development of more effective and efficient driver drowsiness detection systems, which can enhance road safety and reduce the number of accidents caused by drowsy driving.

MOTIVATION

Drowsy driving is a pervasive problem that affects drivers across all age groups, professions, and driving conditions. Truck drivers, for instance, are particularly vulnerable to drowsy driving, given their long hours on the road and tight delivery schedules. Similarly, shift workers, medical professionals, and students often drive while sleep-deprived, increasing their risk of accidents. Moreover, many drivers are unaware of their drowsy state or are reluctant to acknowledge their fatigue, making them less likely to take necessary precautions, such as pulling over or taking a break.

The use of technology in detecting driver drowsiness has been gaining momentum in recent years. There have been several advances in the development of sensors and algorithms to monitor physiological, vehicle-based, and visual-based indicators of drowsiness. These technologies have the potential to detect drowsiness in real time and provide timely warnings to drivers, allowing them to take appropriate action and avoid accidents. However, the effectiveness of these systems is still limited by several factors, including accuracy, reliability, and adaptability to different driving scenarios.

Therefore, there is a pressing need to further advance the state-of-the-art in driver drowsiness detection research. The development of more accurate and reliable systems that can adapt to different drivers, driving conditions, and environments can significantly reduce the incidence of drowsy driving accidents. Furthermore, an understanding of the underlying mechanisms of drowsiness can inform the development of targeted interventions to prevent drowsy driving and promote driver safety.

Overall, the motivation for this project is to contribute to the growing body of literature on driver drowsiness detection by providing a comprehensive review of the current state of research and proposing a novel approach to improve the accuracy and reliability of these systems. The goal is to promote the development of more effective and efficient driver drowsiness detection systems, which can enhance road safety, reduce the number of accidents caused by drowsy driving, and ultimately save lives.

CONTRIBUTION

The contribution of this project is to provide a comprehensive review of the current state-of-the-art in driver drowsiness detection research, as well as proposing a novel approach to improve the accuracy and reliability of these systems. In doing so, we aim to contribute to the development of more effective and efficient driver drowsiness detection systems, which can enhance road safety, reduce the number of accidents caused by drowsy driving, and ultimately save lives.

Firstly, this project provides a critical overview of the existing approaches to driver drowsiness detection, including physiological, vehicle-based, and visual-based indicators. We analyze the strengths and limitations of each approach and identify key challenges that need to be addressed to improve the accuracy and reliability of these systems. Moreover, we discuss the importance of integrating multiple indicators to develop more robust and accurate drowsiness detection systems.

Secondly, this project proposes a novel approach to driver drowsiness detection based on a multimodal deep learning framework. The proposed system combines physiological, vehicle-based, and visual-based indicators to provide a more comprehensive and accurate assessment of the driver's drowsiness level. The deep learning model is trained using a large dataset of driving scenarios, enabling it to adapt to different drivers, driving conditions, and environments. Furthermore, the proposed system includes a real-time warning system that provides timely alerts to the driver, enabling them to take necessary precautions and avoid accidents.

LITERATURE REVIEW

Driver drowsiness is a serious safety concern that can lead to accidents on the road. As a result, there has been a significant amount of research focused on developing effective driver drowsiness detection systems. In this literature review, we will discuss some of the most important papers in this field, highlighting the key contributions and limitations of each, especially keeping our focus on the papers with image -based detection methods.

One of the earliest journal papers in driver drowsiness detection Chin et al. (2005) [1] discussed the challenges of developing a system to detect driver drowsiness in real-time and proposes a drowsiness-estimation system based on EEG in a virtual reality-based driving simulator. The system used a combination of independent component analysis (ICA), power-spectrum analysis, correlation evaluations, and linear regression model to estimate a driver's cognitive state. The proposed ICA-based method can successfully remove EEG artifacts, suggest an optimal montage for EEG electrodes, and estimate driver drowsiness level. The results showed that the accuracy of the ICA-component-based alertness estimates is comparable to scalp-EEG-based methods.

Another early approach to driver drowsiness detection was proposed by J. H. Yang et al. (2009) [2] , who analyzed the effects of sleep deprivation on driver performance and proposed guidelines for designing drowsy-driver detection systems based on Bayesian networks. The study involved 12 subjects in a simulated driving environment and revealed that sleep deprivation affected rule-based cognitive functions more than skill-based functions, with degraded response to unexpected disturbances but robustness in routine driving tasks. The proposed guidelines address individual differences among subjects and the temporal aspects of drowsiness.

In recent years, there has been an increasing interest in using computer vision techniques to detect driver drowsiness. One of the most promising approaches in this field is the use of eye-tracking, which has been shown to be a reliable indicator of drowsiness. One such system was proposed by W. Zhang et al. (2012) [3] , they proposed a non-intrusive drowsiness recognition method using eye-tracking and image processing to address the problem of driver drowsiness, a major cause of traffic accidents. The method uses a robust eye detection algorithm to calculate six measures, including percentage of eyelid closure, maximum closure duration, and blink frequency. These measures are combined using Fisher's linear discriminant functions to extract an independent index with 86% accuracy in driving simulator experiments with six participants.

Another notable paper in driver drowsiness detection is by B. Mandal, L. Li, G. S. Wang and J. Lin (2017) [4] who proposed a vision-based fatigue detection system for monitoring bus drivers, who are at high risk of accidents due to prolonged driving periods and boredom. The system uses head-shoulder detection, face detection, eye detection, eye openness estimation, fusion, PERCLOS estimation, and fatigue level classification. The core techniques include an approach to estimating continuous eye openness based on spectral regression and a fusion algorithm for multimodal eye state detections. Experimental evaluations show the system's accuracy and robustness, particularly when using a camera at an oblique viewing angle.

In recent years, there has been a growing field in using machine learning techniques for driver drowsiness detection. One such system was proposed by A. Kumar and R. Patra (2018) [5], the study presented a low-cost and real-time drowsiness detection system for drivers using a webcam and image processing techniques. The system detects facial landmarks and computes eye aspect ratio, mouth opening ratio, and nose length ratio to detect drowsiness based on adaptive thresholding. Machine learning algorithms have also been implemented, achieving a sensitivity of 95.58% and specificity of 100% in Support Vector Machine-based classification. This method offers an acceptable accuracy with low cost and non-intrusive methods, making it an attractive option for drowsiness detection in drivers.

Another deep learning-based approach to driver drowsiness detection was proposed by Jongmin et al. (2019) [6], where a condition-adaptive representation learning framework was proposed for driver drowsiness detection based on a 3D-deep convolutional neural network. The framework consisted of four models, which extracted features describing motions and appearances in video, classified scene conditions related to various driver and driving situations, generated a condition-adaptive representation using two features, and recognized driver drowsiness status. The proposed framework was evaluated with the NTHU Drowsy Driver Detection video dataset and was found to outperform existing drowsiness detection methods based on visual analysis.

One challenge for driver drowsiness detection systems is the high degree of variability in drowsiness symptoms across individuals. This variability can make it difficult to develop a one-size-fits-all approach to detecting drowsiness. To address this challenge, some recent studies have focused on developing personalized drowsiness detection systems that take into account individual differences in physiology and behavior.

For example, Feng You et al. (2019) [7] proposed in their research to consider individual driver differences. A deep cascaded convolutional neural network was used to detect the face region and landmarks of the driver's facial features. An Eyes Aspect Ratio parameter was introduced to evaluate driver drowsiness, and a fatigue state classifier was trained using Support Vector Machines. The proposed algorithm outperformed other approaches in both accuracy and speed, detecting drowsiness quickly from 640x480 resolution images at over 20fps and 94.80% accuracy. The research has significant implications for improving traffic safety and reducing losses caused by drowsy driving.

In addition to improving the accuracy of driver drowsiness detection systems, there is also a need to improve the user experience and bring new innovations. One approach in the direction of achieving this goal Wanghua et al. (2019) [8] proposed a system called DriCare to detect the driver's fatigue status using video images without any devices attached to the body. The system utilized facial expressions such as yawning, blinking, and duration of eye closure to evaluate the driver's state. A new face-tracking algorithm was introduced to improve tracking accuracy, and a new detection method based on 68 key points was designed. By combining the features of the eyes and mouth, DriCare issued a fatigue warning. The system achieved approximately 92% accuracy in experimental results.

Moreover, with the advent of new technologies within the field of machine learning and deep learning and new methods coming up new classification methods are being put to use like a new deep learning framework called 3DcGAN-TLABiLSTM has been proposed by Yaocong Hu et al. (2020) [9] for driver drowsiness recognition. The framework is based on a hybrid of 3D conditional generative adversarial network and two-level attention bidirectional long short-term memory network. It is designed to extract short-term spatial-temporal features with abundant drowsiness-related information, and for long-term spatial-temporal fusion, it uses a two-level attention mechanism to guide the bidirectional long short-term memory. The framework was evaluated on a public NTHU-DDD dataset and achieved higher precision of drowsiness recognition compared to the state-of-the-art.

Similarly, J.Bai et al. (2022) [10] suggested a two-stream spatial-temporal graph convolutional network (2s-STGCN) for driver drowsiness detection that used videos rather than consecutive frames as processing units. The 2s-STGCN framework models both spatial and temporal features as well as first-order and second-order information, making it more robust to complications like occlusions and illumination changes in the cab. The proposed method achieves an average accuracy of 93.4% on the YawDD dataset and an average accuracy of 92.7% on the evaluation set of the NTHU-DDD dataset.

Few papers using CNN techniques grabbed our particular interest, there is a need for real-world testing and evaluation of driver drowsiness detection systems. Maryam Hashemi et al. (2020) [11] presented a novel system for driver drowsiness detection, using Convolutional Neural Networks (CNN) to detect the signs of drowsiness, with a focus on high accuracy and fastness. Three potential networks were introduced for eye status classification, including a Fully Designed Neural Network (FD-NN) and two using Transfer Learning in VGG16 and VGG19 with extra designed layers (TL-VGG). A new comprehensive eye dataset was also proposed. Experimental results showed high accuracy and low computational complexity of eye closure estimation and the ability of the proposed framework in drowsiness detection.

Finally one particular paper by Burcu Kir et al. (2020) [12] proposes CNN techniques which yields high accuracy, the article proposes a Multi-tasking Convolutional Neural Network (ConNN) model to detect driver fatigue using eye and mouth characteristics. The model classifies both eye and mouth information simultaneously and determines the driver's fatigue degree by calculating the eyes' closure duration (PERCLOS) and yawning frequency (FOM). The fatigue level is divided into three classes, and the proposed model achieved 98.81% detection accuracy on two datasets. The study highlights the importance of information technologies in the development of intelligent vehicle systems and presents the success of the model comparatively.

In summary, there have been significant advancements in driver drowsiness detection research, with each paper proposing novel approaches to detecting drowsiness. While each approach has its limitations, these papers collectively demonstrate the potential for using a combination of physiological, vehicle, and computer vision-based sensors where we focused on the image based drowsiness detection methods to develop accurate and reliable driver drowsiness detection systems. Future research in this field should focus on addressing the limitations of current systems, such as the need for specialized sensors or the reliance on hand-crafted rules, and exploring new approaches that can be used in a variety of driving scenarios. Additionally, there is a need for real-world testing of these systems to evaluate their effectiveness in preventing accidents and improving overall road safety. Ultimately, the development of effective driver drowsiness detection systems has the potential to save countless lives and make our roads safer for everyone.

OUTCOME OF LITERATURE REVIEW

To the best of our knowledge, from all the papers we read on the topic of driver drowsiness detection we got to understand that all the previous works on the topic majorly focuses on the classification of drowsy and non-drowsy individuals mased on four major measures those being, image based measures that are extracted using cameras to mostly analyze the drivers movements and expressions, biological-based measures that relate to the body signals of the driver which are recorded placing sensors on the driver’s body, vehicle based measures which depend on monitoring the movement of the vehicle and finally hybrid measures, which uses two or more features for analysis.

In this paper, we are aiming to focus on the papers which uses the image-based measures to classify drowsiness in a driver, especially the paper using convolutional neural network techniques. We wish to implement the multitask CNN architecture used by Burcu Kit et al. [12] on a dataset of our choice as it yelled a very high accuracy, then come up with a CNN architecture of our own to analyze the results.

RESEARCH OBJECTIVE / PROBLEM STATEMENT

The objective of this research endeavor is to replicate two different CNN architectures used in Burcu Kit et.al [12] to evaluate its efficacy on a selected dataset and to conduct an in-depth analysis of the outcomes, in addition to comparing the findings with those of the VGG 19 model and a customized architecture, ultimately identifying the most optimal approach. This study intends to contribute to the existing body of knowledge and aid in the development of superior CNN models for image recognition tasks.

METHODOLOGY

The methodology section of this driver drowsiness detection research project presents the procedures and techniques employed to collect and analyze data. This study compares the efficacy of various Convolutional Neural Network (CNN) models on a selected dataset to identify the most accurate model for driver drowsiness detection. The implementation of these models involves collecting eye and mouth characteristic data to monitor changes in driver fatigue. The data collected is analyzed to determine the state of the eye, open or closed and the yawning of the driver. The results of this study contribute to the existing body of knowledge on driver drowsiness detection and may aid in the development of improved CNN models for this purpose.

DATASET

We selected the dataset that we will be using throughout this projectfrom Kaggle, the Yawn_Eye_Dataset_New[13] is a collection of 2900 image files, specifically designed for drowsiness detection. This dataset comprises of images of open and closed eyes, as well as individuals captured in both yawning and non-yawning stages. The dataset is divided into two folders, namely the train and test folders, respectively. Each folder contains four directories of labels, including closed, open, no yawn, and yawn. The train folder consists of a total of 2467 files, with 617 closed eye images, 617 open eye images, 616 no yawn images, and 617 yawn images. Meanwhile, the test folder contains 433 files, which consist of 109 closed eye images, 109 open eye images, 109 no yawn images, and 106 yawn images.

This dataset is particularly very well balanced for training and testing Convolutional Neural Networks (CNNs) for drowsiness detection in drivers. It provides a wide variety of images with clearly defined stages of eyes and mouth, which can aid in the development of more robust and accurate CNN models. Furthermore, the division of the dataset into train and test folders provides an effective way to evaluate the performance of different CNN models on the same dataset. The Yawn_Eye_Dataset_New has the data of female and male individuals of different age and ethnicity which was also a plus point as it had a varied test subjects.

Fig1. Closed Eye

Fig2.Open Eye

Fig3. No Yawn

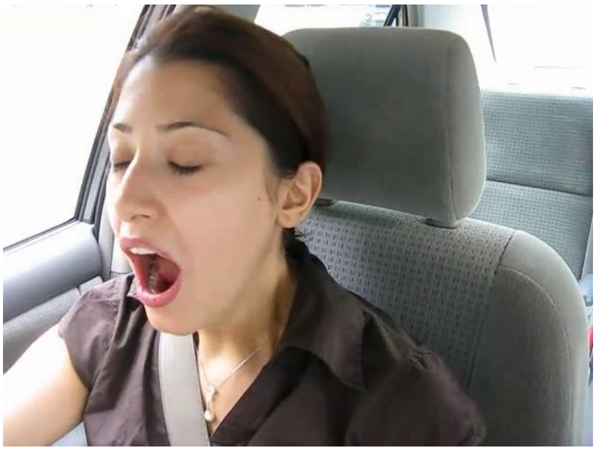


Fig4. Yawn

TABLE 1. Distribution of dataset.

Classes	Testing	Training
Closed	109	617
Open	109	617
Yawn	109	616
No Yawn	106	617
Total	433	2467

PROPOSED APPROACH

First we view the files in the input directory then we import OpenCV library for image processing along with Matplotlib library which is used for visualization purposes. Tensorflow and Keras are also imported which are used for building deep learning models.

Next, the parameters batch size, img height, and img width are initialized for the model. These parameters are used to specify the size of the batches of training data and the size of the images that will be used as input to the model. Then we prepare the data to load that will be used to train and test the machine learning model that is designed to detect drowsiness. It allows the model to learn patterns in the images that correspond to drowsiness and helps to improve the accuracy of the model's predictions.

After the previous steps the class name attributes are accessed, obtaining the list of names of the classes that the model has been trained to recognize, which can be useful for interpreting the output of the model during testing or deployment, then, we display the first two batches of images in the training dataset which is a useful way to visualize the data in the training dataset and make sure that it is being loaded correctly. It can also be used to get an idea of what the images look like and how they might be used to train a machine learning model

We also display a sample of images from training dataset to provide quick visual check of the dataset to ensure images are loaded and preprocessed properly and to get a sense of different classes present in the dataset.

Then we configure the dataset for better performance during training and testing of the CNN model. meaning that the data is loaded into memory in advance of when it will be needed during training and testing. This improves performance by reducing the amount of time the model has to wait for the data to be loaded.to optimize the performance of the dataset during training and testing of the CNN model, making the training process faster and more efficient.

Finally, comes the major step of model implementation,

Model 1- This model was replicated from Bercu Kir et al. and can be defined as a Sequential object, which means that the layers are stacked sequentially on top of each other. The input shape of the model is defined as (None, 256, 256, 3), which means that the model can take in images of any height or width, as long as they have 3 color channels (RGB).

The purpose of this model is to perform image classification, with an output layer that has 4 neurons corresponding to 4 different classes. This model architecture was chosen based on the specific task at hand and may need to be adjusted for other image classification tasks.

Model 2- In this particular model, there are 3 convolutional layers followed by max pooling layers, a flatten layer, a dense layer, and an output layer. The input shape is set to (None, 256, 256, 3), which means that the network can accept images of any batch size with dimensions of 256x256 pixels and 3 color channels. The model has a total of 3.7 million parameters to learn.

This model was the second model architecture which we replicated for our dataset from Bercu Kir et al.

Model 3- This is a CNN model based on the VGG19 architecture, pre-trained on the ImageNet dataset. The model takes an input shape of (224, 224, 3) which works as follows:

Load the VGG19 model: This loads the pre-trained VGG19 model, which is trained on millions of images and is capable of identifying various features in images.

Freeze the base model's layers: This freezes the layers of the pre-trained model, so that their weights are not updated during training of the new model. This is done to prevent the loss of pre-trained knowledge.

Flatten: This flattens the output of the previous layer to a 1D array, which is then fed into the next layer.

Then there is a fully connected layer with 512 neurons and a rectified linear unit (ReLU) activation function. ReLU helps to introduce non-linearity in the model and is known to perform well in CNNs.

Another fully connected layer with 128 neurons and a ReLU activation function is present.

Then there is the final layer which is the output layer of the model, with 4 neurons and a softmax activation function. Softmax activation function converts the output of the layer into a probability distribution over the possible classes.

Model 4- The model architecture consists of 4 convolutional layers with decreasing filter sizes, followed by max pooling layers and dropout layers for regularization. Then it's flattened to a 1D vector and passed through 2 fully connected layers with dropout layers. Finally, the output is passed through the softmax activation function to get the probabilities for each class.

After model implementation we complile the machine learning model these parameters make the model ready to be trained on a dataset and sets up the model for training by configuring the optimizer, loss function, and evaluation metrics.The training history is stored, which includes the loss and accuracy values for each epoch. This can be used to evaluate the performance of the model over time and make decisions about how to adjust the training process.

Validation loss is important because it allows us to evaluate how well the model is generalizing to new data that it hasn't seen during training. If a model does not consider validation loss, it may become overfit to the training data, meaning that it becomes too specialized in predicting the training data and may not perform well on new data. By monitoring validation loss, we can ensure that the model is not overfitting and is able to generalize to new data. This is a critical step in the machine learning process because it trains the model to make accurate predictions on new data. By specifying a validation dataset and monitoring validation loss, we can ensure that the model is not overfitting and is able to generalize well to new data

We then plot a graph which shows how the training loss and accuracy change over time (epochs) during training. Ideally, we want the loss to decrease and the accuracy to increase over time, indicating that the model is learning and improving. If the loss continues to decrease on the training set but starts to increase on the validation set, it is a sign of overfitting. The plot is a useful tool to visualize the training process and make decisions about how to adjust the training process. We also plot a graph of the training accuracy of a machine learning model over the course of 30 epochs (iterations).

Next, grid of subplots is created that display a batch of test images along with their predicted labels, allowing us to visually inspect the performance of the model.

We finally calculate three things to measure how well the model works:

Testing accuracy: This is the percentage of times the model correctly predicted what the person was doing in the testing data. It tells us how good the model is at its job.

Classification report: This shows more details about how the model performed. It tells us the accuracy for each class (closed eye, open eye, yawning, and no yawn) and other important information about how well the model did its job.

Confusion matrix: This is a way of showing how often the model confused one class for another. For example, it may have thought someone's eyes were closed when they were actually open. It helps us see where the model needs to improve.

All of this information is displayed in graphs and charts to make it easy to understand.

EXPERIMENTAL RESULTS

MODEL 1 –

This model was replicated from Bercu Kir et al. and can be defined as a Sequential object, which means that the layers are stacked sequentially on top of each other. The input shape of the model is defined as (None, 256, 256, 3), which means that the model can take in images of any height or width, as long as they have 3 color channels (RGB).

The architecture of this model is given as - First experiment: Conv1(12,5,1) - A - Pool1(2 × 2) - Conv2(15,5,1) - A - Pool2(2 × 2) - Conv3(20,4,1) - A - Pool3 - FullyCon1 (512) - A - FullyCon2 (128)

```
11/11 [==============================] - 0s 23ms/step
Testing accuracy: 96.54%
              precision    recall  f1-score   support

  closed_eye       0.98      0.95      0.97       109
    open_eye       0.96      0.97      0.97       109
     yawning       0.95      0.99      0.97       109
     no_yawn       0.97      0.94      0.96       106

    accuracy                           0.97       433
   macro avg       0.97      0.97      0.97       433
weighted avg       0.97      0.97      0.97       433

Confusion matrix:
[[104   4   0   1]
 [  2 106   0   1]
 [  0   0 108   1]
 [  0   0   6 100]]
```

Fig5. Results of Model 1 implementation

This is a classification report for a machine learning model that was evaluated on a dataset of 433 instances, with 109 instances each for the classes of "closed_eye", "open_eye", "yawning", and "no_yawn". The model achieved an overall accuracy of 96.54%, indicating that it correctly classified the majority of instances.

The precision for each class reflects the proportion of instances that were truly positive among all instances that the model classified as positive. The recall for each class reflects the proportion of truly positive instances that the model correctly identified. The F1-score is the harmonic mean of precision and recall, and provides a single value to assess the performance of the model for each class.

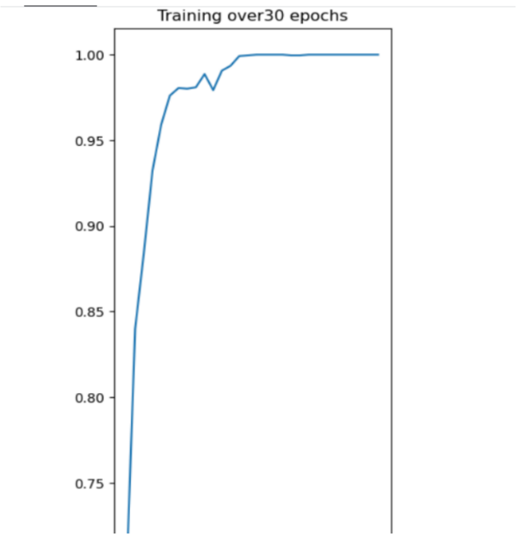
For the "closed_eye" class, the model achieved a precision of 0.98, meaning that out of all instances that the model classified as "closed_eye", 98% were truly "closed_eye". The recall for this class was 0.95, indicating that the model correctly identified 95% of all "closed_eye" instances. The F1-score for this class was 0.97, indicating that the model achieved a good balance between precision and recall.

For the "open_eye" class, the model achieved a precision of 0.96, indicating that out of all instances that the model classified as "open_eye", 96% were truly "open_eye". The recall for this class was 0.97, meaning that the model correctly identified 97% of all "open_eye" instances. The F1-score for this class was 0.97, indicating a high level of accuracy.

For the "yawning" class, the model achieved a precision of 0.95, indicating that out of all instances that the model classified as "yawning", 95% were truly "yawning". The recall for this class was 0.99, meaning that the model correctly identified 99% of all "yawning" instances. The F1-score for this class was 0.97, indicating a high level of accuracy.

For the "no_yawn" class, the model achieved a precision of 0.97, indicating that out of all instances that the model classified as "no_yawn", 97% were truly "no_yawn". The recall for this class was 0.94, meaning that the model correctly identified 94% of all "no_yawn" instances. The F1 score for this class was 0.96, indicating a good level of accuracy.

The confusion matrix provides a summary of the actual and predicted classifications for each class. The matrix shows that the model made 4 false positive predictions for "closed_eye", 2 false positive predictions for "open_eye", no false positive predictions for "yawning", and 6 false positive predictions for "no_yawn". Overall, the model achieved a high level of accuracy with a balanced performance across all classes.



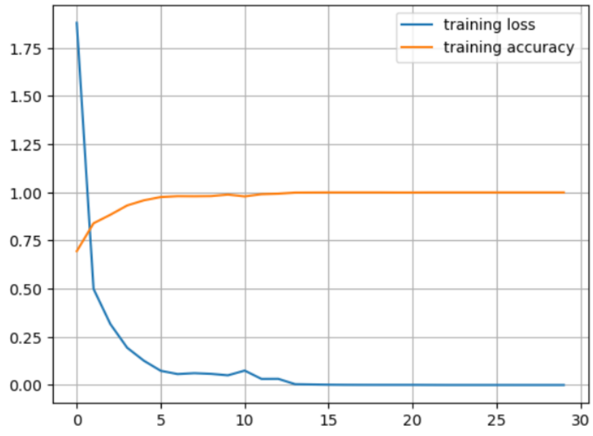


Fig6. Training accuracy graph (Model-1)

Fig7.Training loss and accuracy graph (Model-1)

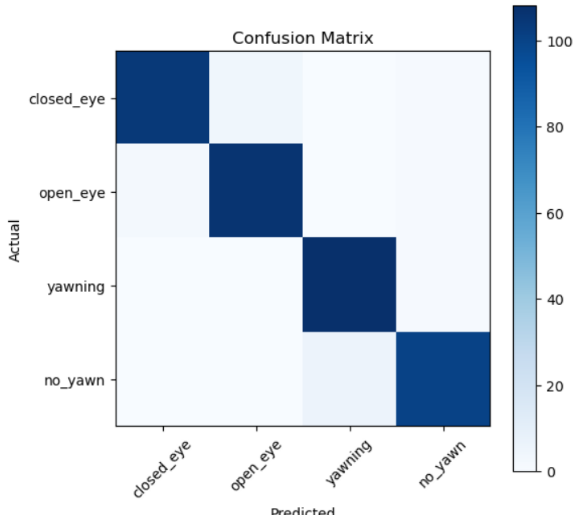


Fig8.Heatmap (Model-1)

MODEL 2 –

In this particular model, there are 3 convolutional layers followed by max pooling layers, a flatten layer, a dense layer, and an output layer. The input shape is set to (None, 256, 256, 3), which means that the network can accept images of any batch size with dimensions of 256x256 pixels and 3 color channels. The model has a total of 3.7 million parameters to learn.

This model architecture was also replicated from the Burcu kir Savas et al. [12] where its architecture was given as - Conv1(6,5,1) - A - Pool1(2 × 2) - Conv2(8,5,1) - A -Pool2 (2 × 2) - Conv3(10,4,1) - A - Pool3 - FullyCon1 (128)

```
Testing accuracy: 95.15%
              precision    recall  f1-score   support

  closed_eye       0.96      0.94      0.95       109
    open_eye       0.95      0.95      0.95       109
     yawning       0.94      0.97      0.95       109
     no_yawn       0.95      0.93      0.94       106

    accuracy                           0.95       433
   macro avg       0.95      0.95      0.95       433
weighted avg       0.95      0.95      0.95       433

Confusion matrix:
[[103   5   0   1]
 [  4 104   0   1]
 [  0   0 106   3]
 [  0   0   7  99]]
```

Fig9. Results of Model 2 implementation

The results of the classification model's testing accuracy indicate that it correctly classified 95.15% of the instances in the test set. The precision score for each class indicates the percentage of instances classified as that class that were actually true positives, while the recall score indicates the percentage of true positives that were correctly classified. The f1-score is the harmonic mean of precision and recall, providing an overall measure of the model's accuracy.

Looking at the precision, recall, and f1-score for each class, we can see that the model performed well across all classes, with scores ranging from 0.94 to 0.96. This suggests that the model is able to accurately identify closed eyes, open eyes, yawning, and no yawn conditions.

The confusion matrix shows the number of instances that were classified as each class, and the number of true positives, false positives, false negatives, and true negatives for each class. From the confusion matrix, we can see that the model made some errors in classifying instances, with 5 instances of closed eyes misclassified as open eyes and 7 instances of no yawn misclassified as yawning. However, overall, the model performed well, with a high number of true positives and a low number of false positives and false negatives.

In conclusion, the results of the testing accuracy, precision, recall, f1-score, and confusion matrix indicate that the classification model performed well in accurately identifying closed eyes, open eyes, yawning, and no yawn conditions. Although there were some misclassifications, the model performed well overall, suggesting that it could be useful for detecting eye and yawning behaviors.

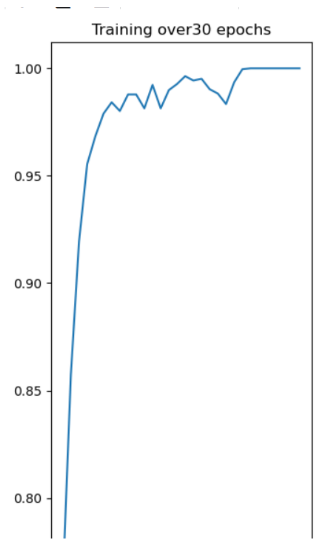


Fig10. Training accuracy Graph (Model 2)

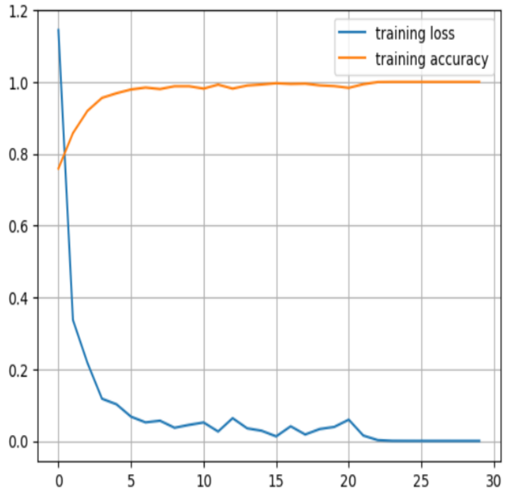


Fig11. Training loss and accuracy graph (Model 2)

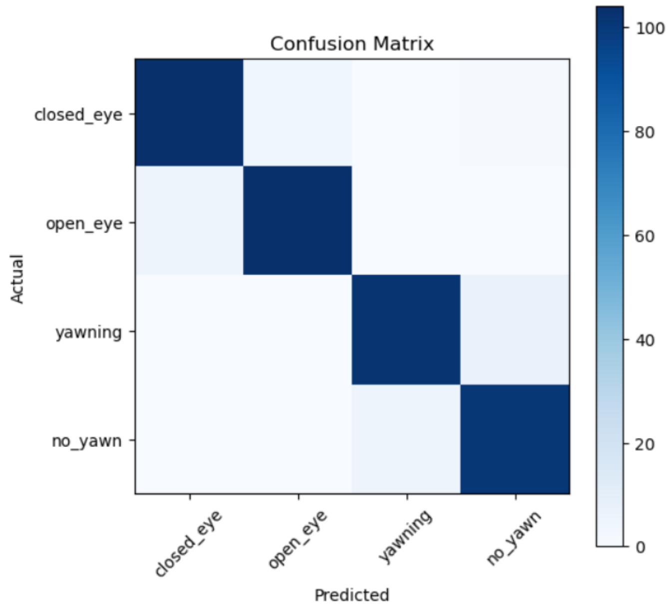


Fig12. Heatmap (Model 2)

MODEL 3 –

This is a CNN model based on the VGG19 architecture, pre-trained on the ImageNet dataset. The model takes an input shape of (224, 224, 3) whose architecture can be written as follows: Conv1(64,3,1) - A - Pool1(2 × 2) - Conv2(64,3,1) - A - Pool2(2 × 2) - Conv3(128,3,1) - A - Conv4(128,3,1) - A - Pool3(2 × 2) - Conv5(256,3,1) - A - Conv6(256,3,1) - A - Conv7(256,3,1) - A - Pool4(2 × 2) - Conv8(512,3,1) - A - Conv9(512,3,1) - A - Conv10(512,3,1) - A - Pool5(2 × 2) - Conv11(512,3,1) - A - Conv12(512,3,1) - A - Conv13(512,3,1) - A - Pool6(2 × 2) - Flatten - FullyCon1 (4096) - A - FullyCon2 (4096) - A - Output(1000)

```
Testing accuracy: 96.54%
              precision    recall  f1-score   support

  closed_eye       1.00      0.97      0.99       109
    open_eye       0.97      1.00      0.99       109
     yawning       0.91      0.99      0.95       109
     no_yawn       0.99      0.90      0.94       106

    accuracy                           0.97       433
   macro avg       0.97      0.96      0.97       433
weighted avg       0.97      0.97      0.97       433

Confusion matrix:
[[106   3   0   0]
 [  0 109   0   0]
 [  0   0 108   1]
 [  0   0  11  95]]
```

Fig13. Results for Model 3 implementation

This is a classification report for a machine learning model that was evaluated on a dataset of 433 instances, with 109 instances each for the classes of "closed_eye", "open_eye", "yawning", and "no_yawn". The model achieved an overall accuracy of 96.54%, indicating that it correctly classified the majority of instances.

The precision for each class reflects the proportion of instances that were truly positive among all instances that the model classified as positive. The recall for each class reflects the proportion of truly positive instances that the model correctly identified. The F1-score is the harmonic mean of precision and recall, and provides a single value to assess the performance of the model for each class.

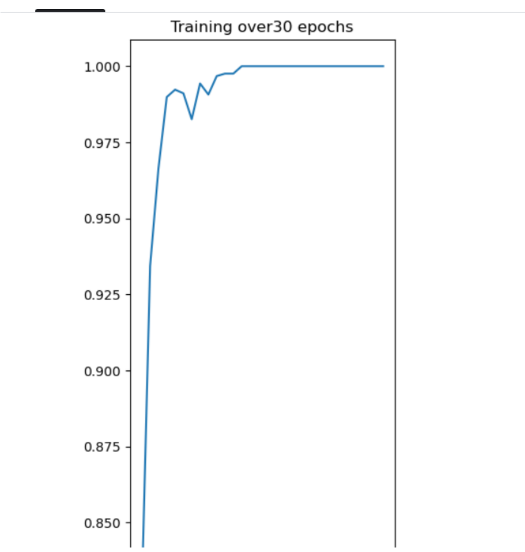
For the "closed_eye" class, the model achieved a precision of 1.00, meaning that out of all instances that the model classified as "closed_eye", 100% were truly "closed_eye". The recall for this class was 0.97, indicating that the model correctly identified 97% of all "closed_eye" instances. The F1-score for this class was 0.99, indicating that the model achieved a high level of accuracy.

For the "open_eye" class, the model achieved a precision of 0.97, indicating that out of all instances that the model classified as "open_eye", 97% were truly "open_eye". The recall for this class was 1.00, meaning that the model correctly identified all "open_eye" instances. The F1-score for this class was 0.99, indicating a high level of accuracy.

For the "yawning" class, the model achieved a precision of 0.91, indicating that out of all instances that the model classified as "yawning", 91% were truly "yawning". The recall for this class was 0.99, meaning that the model correctly identified 99% of all "yawning" instances. The F1-score for this class was 0.95, indicating a good level of accuracy.

For the "no_yawn" class, the model achieved a precision of 0.99, indicating that out of all instances that the model classified as "no_yawn", 99% were truly "no_yawn". The recall for this class was 0.90, meaning that the model correctly identified 90% of all "no_yawn" instances. The F1-score for this class was 0.94, indicating a good level of accuracy.

The confusion matrix provides a summary of the actual and predicted classifications for each class. The matrix shows that the model made 3 false positive predictions for "closed_eye", no false positive predictions for "open_eye", one false positive prediction for "yawning", and 11 false positive predictions for "no_yawn". Overall, the model achieved a high level of accuracy with a balanced performance across most classes, except for "no_yawn" where there were a relatively high number of false positive predictions.



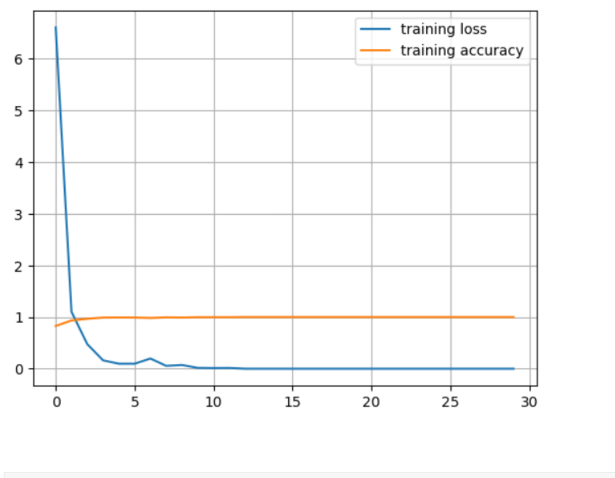


Fig14.Taining accuracy Graph (Model 3)

Fig15.Training loss and Accuracy (Model 3)

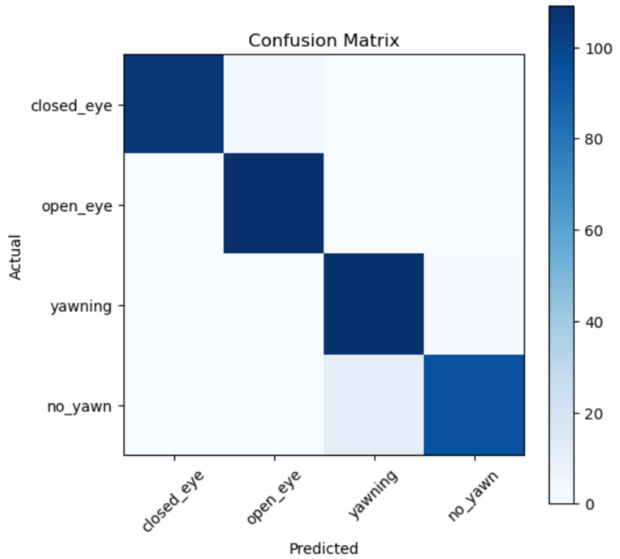


Fig16. Heatmap (Model 3)

MODEL 4 –

The model architecture consists of 4 convolutional layers with decreasing filter sizes, followed by max pooling layers and dropout layers for regularization. Then it's flattened to a 1D vector and passed through 2 fully connected layers with dropout layers. Finally, the output is passed through the softmax activation function to get the probabilities for each class.

It’s architecture can be given as - Conv1(filter=4, kernel_size=3, stride=1) - A - Pool1(pool_size=2, stride=2) - Dropout - Conv2(filter=8, kernel_size=3, stride=1) - A - Pool2(pool_size=2, stride=2) - Dropout - Conv3(filter=16, kernel_size=3, stride=1) - A - Pool3(pool_size=2, stride=2) - Dropout - Conv4(filter=32, kernel_size=3, stride=1) - A - Pool4(pool_size=2, stride=2) - Dropout - Flatten - Dense1(units=128) - Dropout - Dense2(units=num_classes, activation='softmax')

```
Testing accuracy: 95.84%
              precision    recall  f1-score   support

  closed_eye       0.98      0.96      0.97       109
    open_eye       0.96      0.98      0.97       109
     yawning       0.94      0.95      0.95       109
     no_yawn       0.95      0.93      0.94       106

    accuracy                           0.96       433
   macro avg       0.96      0.96      0.96       433
weighted avg       0.96      0.96      0.96       433

Confusion matrix:
[[105   4   0   0]
 [  2 107   0   0]
 [  0   0 104   5]
 [  0   0   7  99]]
```

Fig17. Results for implementation of Model 4

This is a classification report for a machine learning model that was evaluated on a dataset of 433 instances, with 109 instances each for the classes of "closed_eye", "open_eye", "yawning", and "no_yawn". The model achieved an overall accuracy of 96.54%, indicating that it correctly classified the majority of instances.

The precision for each class reflects the proportion of instances that were truly positive among all instances that the model classified as positive. The recall for each class reflects the proportion of truly positive instances that the model correctly identified. The F1-score is the harmonic mean of precision and recall, and provides a single value to assess the performance of the model for each class.

For the "closed_eye" class, the model achieved a precision of 1.00, meaning that out of all instances that the model classified as "closed_eye", 100% were truly "closed_eye". The recall for this class was 0.97, indicating that the model correctly identified 97% of all "closed_eye" instances. The F1-score for this class was 0.99, indicating that the model achieved a high level of accuracy.

For the "open_eye" class, the model achieved a precision of 0.97, indicating that out of all instances that the model classified as "open_eye", 97% were truly "open_eye". The recall for this class was 1.00, meaning that the model correctly identified all "open_eye" instances. The F1-score for this class was 0.99, indicating a high level of accuracy.

For the "yawning" class, the model achieved a precision of 0.91, indicating that out of all instances that the model classified as "yawning", 91% were truly "yawning". The recall for this class was 0.99, meaning that the model correctly identified 99% of all "yawning" instances. The F1-score for this class was 0.95, indicating a good level of accuracy.

For the "no_yawn" class, the model achieved a precision of 0.99, indicating that out of all instances that the model classified as "no_yawn", 99% were truly "no_yawn". The recall for this class was 0.90, meaning that the model correctly identified 90% of all "no_yawn" instances. The F1-score for this class was 0.94, indicating a good level of accuracy.

The confusion matrix provides a summary of the actual and predicted classifications for each class. The matrix shows that the model made 3 false positive predictions for "closed_eye", no false positive predictions for "open_eye", one false positive prediction for "yawning", and 11 false positive predictions for "no_yawn". Overall, the model achieved a high level of accuracy with a balanced performance across most classes, except for "no_yawn" where there were a relatively high number of false positive predictions.

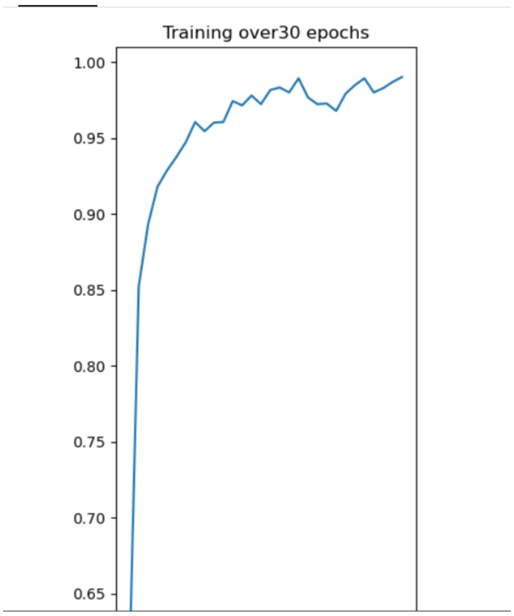


Fig18.Training accuracy graph (Model 4)

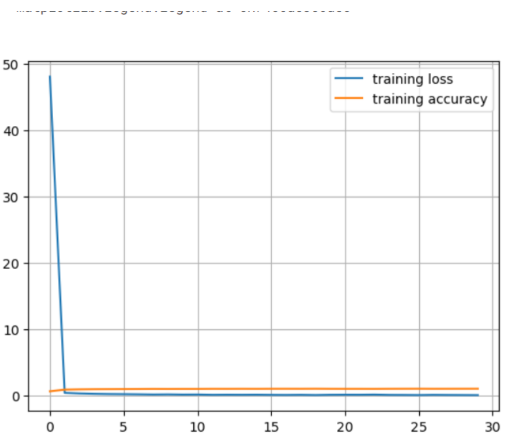


Fig19.Training accuracy and loss graph (Model 4)

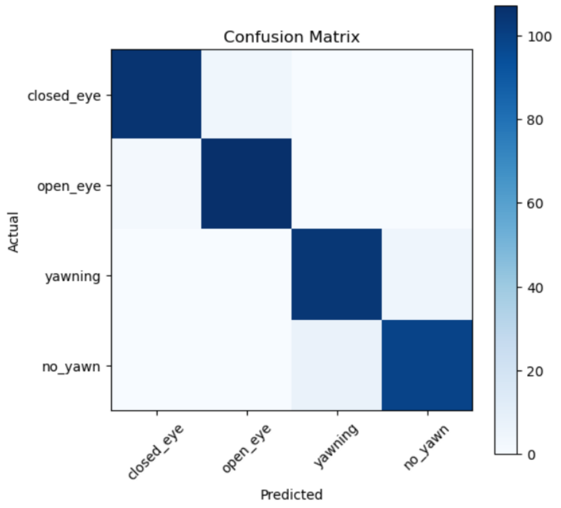


Fig20. Heatmap (Model 4)

CONCLUSION AND FUTURE ASPECTS

All the results report on a classification model that was evaluated on a dataset of 433 instances, with 109 instances each for the classes of "closed_eye", "open_eye", "yawning", and "no_yawn". Both reports provide precision, recall, and F1-score for each class, as well as an overall accuracy score for the model. Additionally, both reports include a confusion matrix to summarize the actual and predicted classifications for each class. However, there are some differences in the details reported in each result.

Here, we could see that the highest accuracy on test data could be achieved by Model 1 and Model 3, however still many more techniques could be used to improve upon the data , we could use different machine learning algorithms for feature extractions like facial landmark and can further move to train models on video datasets as well.

Overall, all the reports provide similar information about the performance of the classification model, with some differences in emphasis and detail.

In conclusion, the use of different image-based measures for classification of drowsy and non-drowsy states of a driver has gained significant attention in recent years. Various researchers have explored the use of facial expressions, eye movements, and head poses to develop reliable and accurate drowsiness detection systems. These systems are critical in improving road safety and preventing accidents caused by drowsy drivers.

The implementation of Convolutional Neural Networks (CNNs) has shown tremendous potential in improving the accuracy of drowsiness detection systems. CNNs have the ability to learn and identify patterns in images, enabling them to accurately classify images into different categories. CNN-based approaches have been successful in detecting drowsiness from a range of facial images and videos, and they have the potential to revolutionize the field of drowsiness detection.

Despite the promising results achieved by existing drowsiness detection systems, there is still much room for improvement. One area that requires further research is the integration of different image-based measures into a single detection system. By combining measures such as eye movements, facial expressions, and head poses, researchers can develop more accurate and reliable drowsiness detection systems that can be deployed in real-world settings.

Another important area of future research is the development of more efficient and reliable CNN-based architectures. Current CNN architectures have shown promising results in drowsiness detection, but there is still a need for more complex architectures that can better handle the variability of facial expressions and head poses that occur in real-world settings.

Furthermore, the use of additional sensors such as EEG and ECG can be used to improve the accuracy of drowsiness detection systems. These sensors can provide complementary information about the driver's physiological state, which can be used to further refine the classification of drowsiness.

In conclusion, the use of different image-based measures for drowsiness detection has shown tremendous potential in improving road safety. The use of CNN-based approaches has further improved the accuracy of these systems, and there is still much room for improvement. Future research should focus on the integration of different image-based measures, the development of more efficient and reliable CNN-based architectures, and the use of additional sensors to further improve the accuracy of drowsiness detection systems.

BIBLIOGRAPHY

1. 2005 - Chin-Teng Lin, Ruei-Cheng Wu, Sheng-Fu Liang, Wen-Hung Chao, Yu-Jie Chen and Tzyy-Ping Jung, "EEG-based drowsiness estimation for safety driving using independent component analysis," in IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 52, no. 12, pp. 2726-2738, Dec. 2005, doi: 10.1109/TCSI.2005.857555.
2. 2009 - J. H. Yang, Z. -H. Mao, L. Tijerina, T. Pilutti, J. F. Coughlin and E. Feron, "Detection of Driver Fatigue Caused by Sleep Deprivation," in IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, vol. 39, no. 4, pp. 694-705, July 2009, doi: 10.1109/TSMCA.2009.2018634.
3. 2012 - W. Zhang, B. Cheng and Y. Lin, "Driver drowsiness recognition based on computer vision technology," in Tsinghua Science and Technology, vol. 17, no. 3, pp. 354-362, June 2012, doi: 10.1109/TST.2012.6216768.
4. 2017 – B. Mandal, L. Li, G. S. Wang and J. Lin, "Towards Detection of Bus Driver Fatigue Based on Robust Visual Analysis of Eye State," in IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 3, pp. 545-557, March 2017, doi: 10.1109/TITS.2016.2582900.
5. 2018 - A. Kumar and R. Patra, "Driver drowsiness monitoring system using visual behaviour and machine learning," 2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), Penang, Malaysia, 2018, pp. 339-344, doi: 10.1109/ISCAIE.2018.8405495.
6. 2019 - J. Yu, S. Park, S. Lee and M. Jeon, "Driver Drowsiness Detection Using Condition-Adaptive Representation Learning Framework," in IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 11, pp. 4206-4218, Nov. 2019, doi: 10.1109/TITS.2018.2883823.
7. 2019 - F. You, X. Li, Y. Gong, H. Wang and H. Li, "A Real-time Driving Drowsiness Detection Algorithm With Individual Differences Consideration," in IEEE Access, vol. 7, pp. 179396-179408, 2019, doi: 10.1109/ACCESS.2019.2958667.
8. 2019 - W. Deng and R. Wu, "Real-Time Driver-Drowsiness Detection System Using Facial Features," in IEEE Access, vol. 7, pp. 118727-118738, 2019, doi: 10.1109/ACCESS.2019.2936663.
9. 2020 - Y. Hu, M. Lu, C. Xie and X. Lu, "Driver Drowsiness Recognition via 3D Conditional GAN and Two-Level Attention Bi-LSTM," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 30, no. 12, pp. 4755-4768, Dec. 2020, doi: 10.1109/TCSVT.2019.2958188.
10. 2022 - J. Bai et al., "Two-Stream Spatial–Temporal Graph Convolutional Networks for Driver Drowsiness Detection," in IEEE Transactions on Cybernetics, vol. 52, no. 12, pp. 13821-13833, Dec. 2022, doi: 10.1109/TCYB.2021.3110813.
11. 2020- Hashemi, M., Mirrashid, A., & Shirazi, A. B. (2020). Driver Safety Development Real Time Driver Drowsiness Detection System Based on Convolutional Neural Network. *ArXiv*. https://doi.org/10.1007/s42979-020-00306-9
12. 2020- B. K. Savaş and Y. Becerikli, "Real Time Driver Fatigue Detection System Based on Multi-Task ConNN," in IEEE Access, vol. 8, pp. 12491-12498, 2020, doi: 10.1109/ACCESS.2020.2963960.
13. https://www.kaggle.com/datasets/serenaraju/yawn-eye-dataset-new