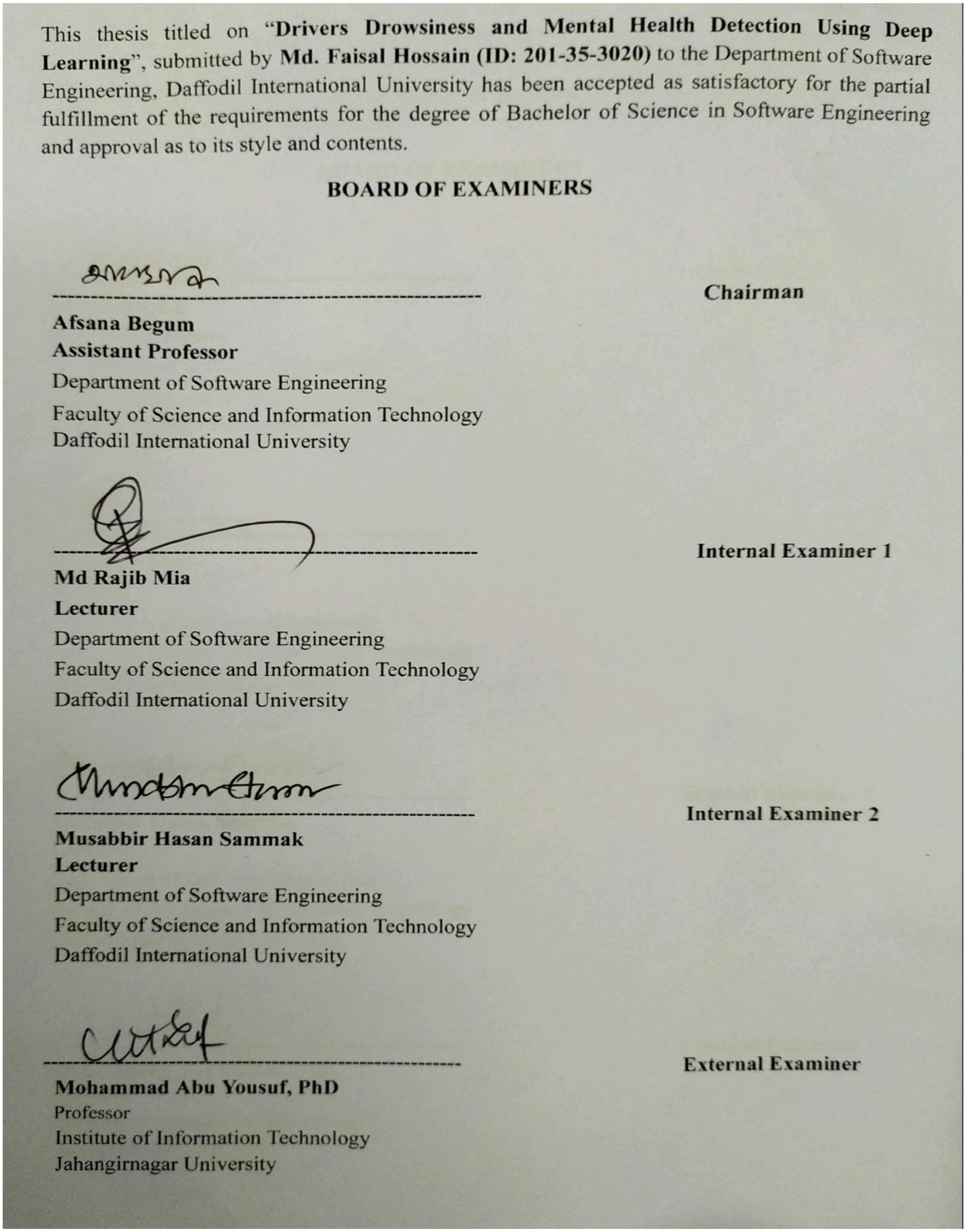
Drivers Drowsiness and Mental Health Detection using Deep Learning

Supervised By	**Submitted By**
Md. Shohel Arman	Md. Faisal Hossain
Assistant Professor	ID : 201-35-3020
Department of SWE, FSIT	Department of SWE, FSIT
Daffodil International University	Daffodil International University

This Thesis report has been submitted in fulfillment of the requirements for the

Degree of Bachelor of Science in Software Engineering. Fall-2023

APPROVAL



This thesis titled on "**Drivers Drowsiness and Mental Health Detection Using Deep Learning**", submitted by **Md. Faisal Hossain (ID: 201-35-3020)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

Chairman

Afsana Begum
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 1

Md Rajib Mia
Lecturer
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

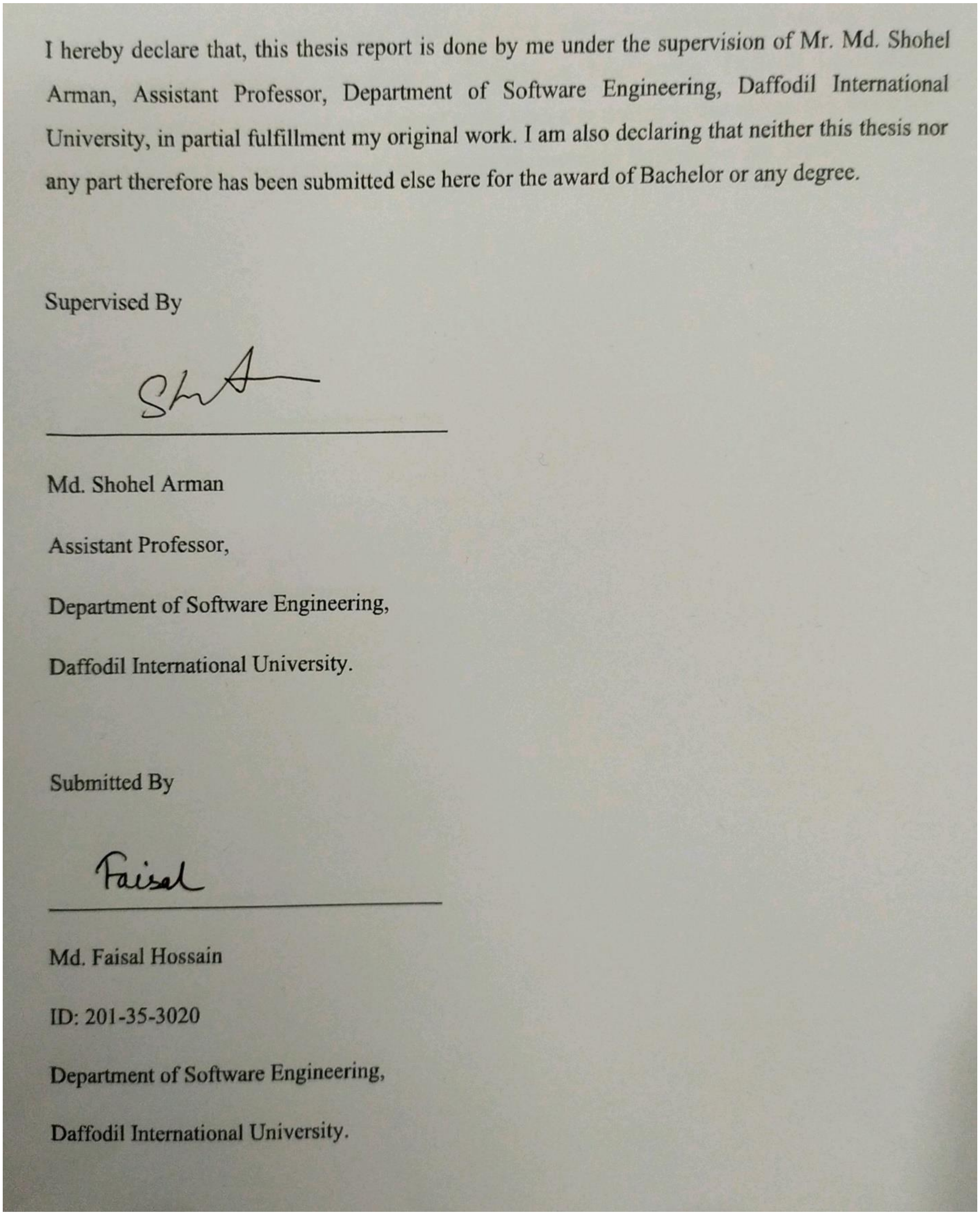
Internal Examiner 2

Musabbir Hasan Sammak
Lecturer
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

External Examiner

Mohammad Abu Yousuf, PhD
Professor
Institute of Information Technology
Jahangirnagar University

THESIS DECLARATION



I hereby declare that, this thesis report is done by me under the supervision of Mr. Md. Shohel Arman, Assistant Professor, Department of Software Engineering, Daffodil International University, in partial fulfillment my original work. I am also declaring that neither this thesis nor any part therefore has been submitted else here for the award of Bachelor or any degree.

Supervised By

Md. Shohel Arman

Assistant Professor,

Department of Software Engineering,

Daffodil International University.

Submitted By

Faisal

Md. Faisal Hossain

ID: 201-35-3020

Department of Software Engineering,

Daffodil International University.

ACKNOWLEDGEMENT

First of all, I want to express my soulful thanks and gratefulness to Almighty Allah for his blessing to me to complete the final year thesis successfully.

MD. Shohel Arman, Assistant Professor, Department of Software Engineering, Daffodil International University, has a lot to thank for in my life. My supervisor's extensive background and enthusiasm for the "Deep Learning" subject made it possible for me to move on with this research. This endeavor would not have been feasible without his unending patience, intellectual direction, constant encouragement, persistent and vigorous supervision, constructive criticism, helpful counsel, and reading of several inadequate versions and subsequent corrections.

Thanks to **Dr. Imran Mahmud**, Head of the Software Engineering Department at DIU, and the other professors and staff members who assisted me in completing my research, I was able to successfully defend my thesis and graduate with honors.

Finally, I'd want to give my parents a big thank you for being so patient and supportive throughout my life.

Drivers Drowsiness and Mental Health Detection using Deep Learning

ABSTRACT

One of the most important body part is the face which holds a lot of information. Any person's mental state is revealed through their facial expression.The purpose of the study is to develop a system that can ensure safe driving with great accuracy. The main objectives of this study is to detect whether the driver is asleep or not, avoid road accidents, eliminate reckless driving, alert drivers about their mental and emotional situation. For this research, I have collected data from two different datasets. One is FER-2013 and another one is drowsiness dataset for open eyes and closed eyes. The images of the emotion dataset contains only the face which are enough cropped and the drowsiness dataset contains only the eyes. I have used angry, fear, happy, neutral, sad and yawning for emotion classifications. I have used 4 deep learning models in this research. The 4 four models are Xception, InceptionV3, ResNet50 and VGG19. These neural network models are used for feature extraction and classification tasks. The model that gives the higher accuracy than other models is Xception. In both tasks, The Xception outperformed the competition. For eye detection, it obtained 98.97% accuracy, and for face emotion detection, 99.26% accuracy. It showed excellent accuracy and metrics when it came to classifying emotions. The model performs quite well, with an average precision, recall, and F1-score of about 0.99. Overall, Xception performed exceptionally well across a number of emotion classifications and attained 99% accuracy on the eye dataset. The weighted and macro averages both confirmed the effectiveness of the system. This study suggests an improved pretrained model based approach for detecting driver’s inattention, which will ensure safe driving with great accuracy and improve the drivers driving efficiency. In future, I will work on adding night vision capabilities, the system will be able to identify and recognize objects better and adapt to different driving circumstances more easily. Furthermore, it may be proposed that driving behaviors like speeding, safe driving, and braking suggest a more advanced system. The combination of these factors can result in an improved and more advanced system that can identify and mark drowsy drivers, improve drivers' concentration, and reduce the number of traffic accidents.

Keywords: driver drowsiness; human-vehicle interaction; deep learning, emotion recognition; xception; inceptionv3

Table of Contents

List of Figures

CHAPTER 1

Introduction

1.1 Background:

Road accidents are a major problem all over the world (Monish, 2023). The number of people living in urban areas and the number of people using cars has led to an increase in traffic accidents, fatalities, loss of life, financial losses, and permanent physical and mental impairment (Chand & Karthikeyan, 2022). Throughout the world, millions of people are dying for irresponsible driving. Reasons behind the enlarged number of accidents and casualties are reckless driving, over speeding, overloading, overtaking, breaking laws, illegal and dangerous competition, long-distance driving without a break, drug and alcohol use (Sukhavasi et al., 2022), incompetency of the driver, hazardous road etc. All these reasons could lead to traffic accidents.

Also the person who drives the vehicle gets sleepy or his eyes get drowsy. The feeling of drowsiness may only last for a few minutes, but the effects could be serious (Singh et al., 2023). This drowsiness or sleepy feeling during driving can lead to vehicle collision as none wants to sleep during driving. The facial expressions, blinking frequency, and yawning patterns of a fatigued driver differ from those of a typical driver. This problem is faced mostly by truck drivers. They use highways regularly and it is very much possible to sleep in the middle of the driving because the roads are straight and noiseless. If drowsiness is identified, a warning (Singh et al., 2023) or alarm signal can be given to remind the driver to wake up and get out of the drowsy state.

Emotions are a crucial factor while communicating (Chowdary et al., 2021). When going through a good or difficult situation, a person's mental state might influence them (Sukhavasi et al., 2022). Mentally unstable driver poses significant risks to themselves and others. Mentally ill drivers may behave erratically, make poor decisions, lose focus, and respond unpredictablely, all of which can make it more difficult for them to drive safely. They carry the risk of endangering not only their own safety but also that of other road users, pedestrians, and passengers.

1.2 Research Motivation:

A comprehensive approach is needed to ensure road safety in order to prevent collisions and eliminate careless driving. To promote safe driving habits, this comprehensive approach includes both strict enforcement of traffic laws and instructional programs. At the same time, efforts are made to deal with drivers' psychological health, realizing that emotional and mental health affects driving safety. Cutting-edge technology, such advanced driving-assistance systems, are essential to this effort because they can identify indications of driver tiredness or drowsiness, send out alerts in a timely manner, and even take action when needed. The overarching goal is to provide a comprehensive framework that not only tackles current risks but also effectively detects potential dangers, ultimately leading to a significant reduction in the number of accidents and of terrible consequences when driving.

The statements that are provided highlight a thorough approach to road safety and promote the prevention of accidents by addressing various aspects of driving behavior. Eliminating reckless driving, increasing public awareness of drivers' mental and emotional states, putting in place technologies to identify driver drowsiness, and immediately alerting drivers to potentially dangerous circumstances that could have fatal effects are the main objectives.

1.3 Problem Statement:

Bangladesh, with a population of 162.1 million people and a land area of 1,48,610 square kilometers, is one of the world's most heavily populated countries. Road accidents are a typical occurrence in a country like Bangladesh with such a huge population, and the cause is generally due to the vehicle driver's lack of attention. Sometimes, the person who drives the vehicle gets sleepy and his eyes get drowsy. This drowsiness can lead to a serious problem like accidents. In this study, we are going to propose a method that ensures safe driving considering the driver's drowsiness problem.

Several important issues related to driving safety are addressed by the proposed approach. First and foremost, it attempts to address the risk of being sleepy or drowsy while operating a motor vehicle, as these circumstances can greatly increase the likelihood of an accident. Furthermore, the system acknowledges the possible risks associated with mentally unstable drivers, highlighting the significance of reducing risks for both the driver and other road users. The technology also attempts reduce the risks that come with driving carelessly, which can result in car crashes and other incidents. Additionally, the system stands out because it takes into account the combination of tiredness and facial emotion at the same time. This is a new method that sets it apart from previous research that does not fully handle this combination. All things considered, the suggested solution aims to improve traffic safety by tackling these complex issues in a thorough way.

1.4 Research Gap:

This section highlights a number of issues with emotion identification technology as it exists today. First of all, it can be said that it has restricted accessibility, implying that the general population does not have widespread access to the technology. Second, it draws attention to the lack of advanced algorithms, suggesting that the emotion identification systems in use today may not be equipped with powerful computational techniques. Thirdly, the difficulties of fully accounting for all emotions can be mentioned, suggesting that the technology may find it challenging to adequately analyze and categorize emotions. It can also be added that the majority of studies do not address the simultaneous assessment of tiredness and facial emotions and outlining possible directions for future research in this area.

1.5 Research Question:

The research questions are given below:

Q1. What is the relationship between a driver's mental state and a car accident?
Q2. Why is it important to detect the drowsiness?
Q3. Can the facial expressions of the drivers be recognized?
Q4. How can we use automated systems to lower the risk of traffic accidents?

This study explores the complex relationship that exists between the mental health of drivers and the risk of car crashes. It highlights how important it is to recognize tiredness in drivers and looks into whether facial expressions can be used as markers. The conversation also touches on the possibility of using automated systems as a preventative step to lower the likelihood of collisions, providing information about how technology and road safety interact.

1.6 Research Scope:

The objectives of this research are giveb below:

I. Mark down the drowsy drivers.
II. Improve the drivers driving efficiency.
III. Eliminate reckless driving.
IV. Increase drivers ability to focus.
V. Help to reduce road accidents.

The statement provides an in depth method to dealing with many road safety-related issues. The main objectives are to detect and deal with sleepy drivers, increase overall driving effectiveness, stop irresponsible driving, sharpen drivers' concentration, and eventually reduce the number of traffic accidents. The focus is on an integrated strategy to address particular driver-related issues in order to produce safer road conditions.
So, my main focus is to mark down the drowsy drivers, increase drivers ability to focus, help to reduce road accidents.

Additionally, we will also recognize some of the facial emotions for example: angry, yawn, happy, neutral etc (Khaireddin & Chen, 2021), to detect the mental condition of the driver, as too much joy can raise the excitement, on the other hand too much anger can manipulate the brain.

So, in this study we will use the combination of drowsiness and emotional situation to detect open or close eyes, also whether the driver is mentally stable to drive or not.

Chapter 2

Literature Review

2.0 Introduction

Research papers require a literature review, which is an extensive analysis of previous academic publications and studies that are relevant to a certain topic. By providing an overview of the past and present states of knowledge, it places the research problem in context, points out any gaps or disagreements in the literature, and suggests for the importance of the idea. Furthermore, by including relevant theories and concepts, the literature review informs methodological decisions, prevents duplication by highlighting previously studied topics, and develops a theoretical framework. In the end, it validates the claims and interpretations made in the research report, indicating analytical ability by demonstrating familiarity with important projects and discussions. All things considered, a literature review is crucial for ensuring that research is based in current knowledge, demonstrating its relevance, and guiding academic study.The literature review for this particular topic are given below

2.1 Literature Review

In order to classify face expressions, this research (Wawage & Deshpande, 2022) presents a Convolutional Neural Network-based real-time emotion recognition system for drivers. Through online surveys and in-person driving tests, the Pune, India-based study demonstrates the relationship between a driver's mood and driving behavior. Emotional experiences are recorded, features are extracted using AlexNet and VGG16, emotions are recognized using GGDA, and Multi-Class Log-Loss is reduced using a top-model (MLP). Some of the limitations are the requirement for a more extensive and varied dataset, the possibility of improving VGG-16, and the possibility of using speech analysis to determine mood. Important conclusions include the development of the system, new understandings of the emotional aspects of driving, and the significant influence of a driver's mood on their driving behaviors.

This research (Chand & Karthikeyan, 2022) introduces a revolutionary method that integrates emotion recognition and driver fatigue detection to reduce reckless driving. It achieves an impressive 93% accuracy in both areas. With the use of the Extended Cohn-Kanade (CK+) and Driver Drowsiness Detection (DDD) datasets, the study applies Convolutional Neural Networks (CNN) and OpenCV to efficiently develop the model. The work is divided into five sections that cover the suggested system, prior research, experimental findings, and conclusion. The overview by H. Varun Chand and J. Karthikeyan (2022) emphasizes the main causes of traffic accidents, emphasizing the importance of driver fatigue and attitude in particular. The suggested approach recognizes the necessity to take into account extra driver states or emotions and attempts to address these problems by using CNN-based sleepiness detection and emotion analysis.

To learn how to recognize group emotions in faces, the research (Kaviya & Arumugaprakash, 2020) presents a deep learning architecture that achieves 65% accuracy for FER-2013 and 60%

accuracy for bespoke datasets. The methodology consists of pre-processing, feature extraction, and group emotion prediction with voice synthesis using CNN and Haar filter. CNN is rather good at learning the features of facial expressions, and it does a great job of anticipating pleasant emotions. Concerns about model accuracy, possible misclassification, and the requirement for additional study are some of the limitations. FER-2013 (35,000 photos, seven emotions) and a bespoke dataset (25,000 images, five emotions) are the datasets that were used. All things considered, the work sheds light on difficult applications of group facial expression identification in a variety of disciplines.

In driving event identification, Support Vector Machine outperformed Convolutional Neural Networks and Single Shot identification. According to the study (Anigbogu et al., 2022), speeding is a major contributing factor to road crashes in Nigeria, highlighting the urgent need for initiatives to encourage safer and more environmentally friendly driving practices. Due to limitations such as a dataset that is geographically limited and a restricted focus on particular driving occurrences, more research is necessary to have a full picture of driving behavior. Without offering any concrete remedies, the report highlights the high number of traffic accidents in Africa, especially in Nigeria, and highlights the need for more study and initiatives to improve road safety across the continent. 3423 driving event photos unique to Nigeria make up the dataset, which is split between training (70%) and testing (30%) sets. Using a primary dataset, the methodology compares image processing and machine learning techniques.

The technique for identifying tiredness in drivers presented in this research (Vanjani & Varyani, 2019) uses CNNs and LSTMs to distinguish between alert and sleepy drivers with good accuracy. Limitations include the early stage of PERCLOS prediction and difficulties distinguishing between blinking and falling asleep from a single frame. After retraining the Inception-v3 model, the manually generated dataset for feature extraction obtained 97.5% accuracy on the validation set. The process includes creating an image dataset, using transfer learning, extracting features using CNN, and detecting tiredness using an LSTM across ten epochs. In conclusion, the study presents a reliable technique for identifying driver tiredness that makes use of CNNs and LSTMs and produces consistent classification outcomes.

For widespread accessibility, this research (Monish, 2023) provides a robust three-stage sleepiness detection method that is deployed on Android smartphones. It focuses on cutting-edge methods for improving feature maps so that drowsiness may be recognized efficiently. Compilation difficulties and computational complexity are among the limitations. The extensive analysis is enhanced by datasets like Southeast and expanded datasets. Convolutional neural networks are used in the technique, which emphasizes global average pooling for effective feature extraction.

With a focus on Android implementation, Kumar, Sai, and Kumar present a dependable three-stage smartphone-based drowsiness detection system (Jaiswal & Nandi, 2020). Popular convolutional neural networks are used in the methodology to extract features, and a three-stage detection system that makes use of the voiced to unvoiced ratio and PERCLOS is employed. The benefits of convolutional neural networks, such as high-level feature map extraction, are highlighted in the paper. Southeast and expanded databases for the categorization of distracted driving were among the datasets used. Compilation difficulties and computational complexity are among the limitations. The three-stage verification procedure, Android accessibility, and creative feature extractor design of the suggested framework are highly praised for their ability to effectively identify tiredness.

The authors provide a real-time emotion classifier that reduces compute complexity while outperforming state-of-the-art results, obtaining 74% accuracy (Sukhavasi et al., 2022). The model tackles issues including real-time accuracy, parameter needs, emotion integration, and differentiating between mixed emotions. It has been verified on eight different datasets. The study uses a convolutional neural network and covers relevant literature, network construction techniques, dataset characteristics, computational details, and outcomes. A custom lab dataset, JAFFE, FEI, IMFDB, TFEID, Chicago Face Database, CK, CK+, and Fer2013 are among the datasets. The model's importance in human-robot interaction is emphasized in the summary by Shruti Jaiswal and G C Nandi (2019), which highlights applications in child treatment, counseling, and geriatric care. The model's possible influence on real-time systems and customized robot development is covered in the paper's conclusion.

FER 2013, CK+, KDEF, and KMU FED are just a few of the datasets on which the study's hybrid network architecture for driver emotion prediction achieves high accuracy (Singh et al., 2023). Limitations include difficulty in identifying strong emotions under difficult driving situations, a lack of masked face images, and unresolved problems such as occlusions and illumination fluctuations. The approach classifies emotions with a support vector machine and convolutional neural network by combining Gabor and LBP features, and it performs exceptionally well on a variety of datasets. The research illustrates the improved accuracy of the innovative hybrid network under various settings and emphasizes the importance of constant driver emotion monitoring in the automobile sector.

The paper (Kartali et al., 2018) focuses on leveraging image processing, specifically Eye Aspect Ratio (EAR), to achieve 80% accuracy in drowsiness detection. The study's technique includes real-time AI algorithms, facial and eye movement tracking, driving pattern analysis, and physiological data analysis. A dataset is used to derive EAR threshold values for detecting driver sleepiness. The research highlights the importance of real-time detection in averting mishaps and emphasizes the necessity of an alert system. Notably, an 80% accuracy rate was achieved using the study's testing criteria. In order to account for the individual heterogeneity in EAR, the system is built to autonomously ascertain the appropriate detection threshold.

Five techniques for real-time emotion recognition from facial photos are compared in the article (Dua et al., 2021), with Affdex CNN demonstrating the highest accuracy at 85.05%, followed by AlexNet at 76.64%. Nevertheless, drawbacks include the inability to generalize to previously unseen photos or outdoor environments, dependence on a tiny sample size of 8 participants, and possible observer bias in the evaluation of facial photographs. The research uses a variety of methodologies, including CNN-based techniques, traditional HOG feature classification, the

OpenFace toolbox for extracting facial landmarks, and the Affdex SDK for emotion recognition. Notably, the study includes a scant discussion of possible avenues for future research.

In order to solve the common problem of accidents brought on by sleepy driving, the article (Chowdary et al., 2021) suggests a deep learning-based system that can identify driver drowsiness with 85% accuracy. The research recognizes its limitations, including the invasiveness of physiological tests and the underutilized application of deep learning in this field, and makes use of the NTH Drowsy Driver Detection (NTHU-DDD) video collection, which includes a variety of participants. Four deep learning models—AlexNet, VGG-FaceNet, FlowImageNet, and ResNet—are used in the suggested methodology. These models are categorized into four features and merged using an ensemble algorithm to achieve robust performance. The efficacy of the suggested system in identifying driver drowsiness is summarized by Mohit Dua, Shakshi, Ritu Singla, Saumya Raj, and Arti Jangra (2020). It achieves an accuracy of 85% in a variety of scenarios and acquires features from an extensive dataset.

The study (Hassouneh et al., 2020) uses pretrained convolutional neural networks to successfully demonstrate transfer learning in facial emotion recognition. With the best accuracy of 98.5%, Inception V3 has potential uses for speech and EEG signal emotion identification in the future. Constraints include the need for more research in order to extend networks to include speech and EEG inputs, dependence in traditional methods, and deep learning performance's dependency on data size. Emotion detection tests are conducted utilizing the CK+ database. The methodology includes performance evaluations using a variety of metrics, implementation parameter debates, CK+ database experiments, transfer learning, and the presentation of confusion matrices. Transfer learning for facial emotion identification is covered in the summary by M. Kalpana Chowdary, Tu N. Nguyen, and D. Jude Hemanth. Accuracy with pretrained models ranges from 94.2% to 98.5%, and the paper offers thorough insights into affective computing and model architectures.

The study (Jabbar et al., 2020) presents a real-time CNN-based emotion identification system that achieves 87.25% accuracy for EEG signals and 99.81% accuracy for facial landmarks. The algorithm's efficacy in recognizing emotions through the use of virtual markers and an optical flow algorithm is demonstrated by the results. Sensitivity to changes in the environment, the difficulty of integrating signals, and the emphasis on real-time applications are some of the limitations. The dataset includes 55 undergraduate students' face expression data, and the methodology includes 3-fold cross-validation, virtual marker placement, and Haar-like characteristics. The system focuses on emotionally categorizing expressions in children with autism and physically challenged people, highlighting areas for improvement through increased data collecting and feature extraction methods.

The paper (Liliana, 2019) presents a low-power solution for sleepiness detection, with an average accuracy of over 83% with a small 75KB model size. The creation of low computing capacity models appropriate for embedded systems, which satisfies the need for effective sleepiness detection, is the primary contribution. The study, which makes use of the NTHU Dataset, highlights the need for enhanced performance in difficult lighting circumstances by acknowledging limits associated with illuminance and face feature occlusion. The methodology uses an Android application for real-time sleepiness detection and a five-layer CNN that has been trained for satisfactory accuracy. The study emphasizes how important it is to prevent traffic accidents and praises industry collaboration in developing new technologies. In embedded systems and Android devices, the suggested lightweight CNN model provides an efficient real-time drowsiness detection method.

With an impressive average accuracy rate of 92.81%, this research (Jain et al., 2019) presents a revolutionary deep Convolutional Neural Network (CNN) technique for facial emotion identification. The paper addresses the complexity of facial expression identification using the enlarged Cohn-Kanade (CK+) dataset, and based on misclassification results, it suggests improvements to the system architecture. Using a deep CNN architecture with filter layers and a classification layer, the study technique tests various quantities of training and testing data sets from the CK+ database. Interestingly, as the amount of training data increases, the mean square error reduces. The work emphasizes the necessity of further investigation into alternate designs for improved outcomes, taking into account the diversity of human face expressions. According to D.Y. Liliana's (2019) summary, the study advances automated facial expression detection by using CNN to accurately classify eight fundamental emotion classes and by suggesting a novel method for automatic feature extraction.

The main conclusions of the research (Deng & Wu, 2019) show that the suggested DNN model outperforms current state-of-the-art techniques in face emotion recognition. The effectiveness of the model is validated by the large improvement in overall performance that results from integrating FCN and residual blocks. The model outperforms earlier techniques on the JAFFE and CK+ datasets, achieving superior classification across six facial emotion classes using a single DNN with convolution layers and deep residual blocks. Large head positions, non-frontal faces, and action similarity are among the difficulties, underscoring the continuous difficulty of emotion recognition in computer vision. Using datasets from Japanese Female Facial Expression (JAFFE) and Extended Cohn-Kanade (CK+), the methodology uses the DNN for feature extraction and training. The JAFFE dataset has 8363 photos, of which 8200 are used for training, and the CK+ dataset consists of 486 video sequences from 97 posers.

The study's (Ali et al., 2020) noteworthy discoveries include the development of a brand-new face-tracking algorithm (MC-KCF) and a non-contact technique for identifying driver weariness called DriCare. With the help of CelebA and YawDD datasets and video footage from ten participants in both drowsy and clear driving conditions, the technique makes use of MC-KCF for improved face tracking and a CNN for eye state assessment and yawning detection. Wanghua Deng and Ruoxue Wu's (2019) summary highlights the non-contact method that uses a camera mounted on the vehicle to assess tiredness by looking at things like blinking frequency, eye closure duration, and yawning. It draws attention to potential limits in specific situations, noting a drop in accuracy when taking environmental factors like darkness and the use of glasses into account.

In order to successfully recognize seven emotions using machine learning and facial image analysis, the study (Zadeh et al., 2019) suggests a cascade of CNN, binary SVM, and multi-class support vector machine (SVM) methods. These deep learning approaches work well together to reliably identify emotions in real-world photographs. The intricacy of facial expressions, the difficulty of retraining the system, and the scarcity of data for thorough training are still obstacles, though. The dataset used in the study is 48x48 pixel grayscale photos with emotional classifications (happy, sad, angry, surprised, neutral, disgusted, and frightened) from kaggle.com. The methodology integrates CNN, Viola-Jones algorithm, SVM, K-means clustering, decision tree, and real-world and internet media data collecting. It uses a training dataset of 34,488 photos and a testing dataset of 1,250 images. Md. Forhad Ali et al.'s summary emphasizes the significance of facial expressions in nonverbal communication, addresses difficulties in identifying emotions, and looks at the field's historical basis and current approaches. Along with highlighting the possible uses in social media, entertainment, criminal justice, content analysis, and healthcare, it also emphasizes how important it is for machines to be able to identify and react to human emotions.

In order to increase CNN's recognition accuracy and training time, the study (Mehendale, 2020) presents a deep learning architecture for human emotion recognition. The experimental findings show an astounding 97% accuracy, which is higher than traditional CNN techniques. Some of the limitations are the need for additional optimization to improve time efficiency, uncertainties over the face database's dependability for practical uses, and the possibility of capturing conflicting emotions. The dataset is the JAFFE database, which contains 213 photos of Japanese female models in seven different emotional states. The methodology uses CNN for classification and Gabor filters for feature extraction. Gabor-filtered images are used as neural network inputs. Citing related studies in the field, the study emphasizes the importance of emotions in human-computer interaction and argues that the suggested deep learning framework is effective for facial emotion recognition when used with CNN and Gabor filters.

Using VGGNet and SGD with Nesterov momentum, Yousif Khaireddin and Zhuofa Chen (2021) achieve a state-of-the-art 73.28% testing accuracy in facial emotion detection on FER2013. With a validation accuracy of 73.59%, the Plateau scheduler's Reducing Learning Rate performs better than expected. Restrictions include the need for better face feature identification and difficulties with naturalistic emotion recognition. The study (Khaireddin & Chen, 2021) uses the FER2013 dataset as its focal point and uses an implementation of the VGG network with hyperparameter tuning, optimization algorithm trials, and learning rate scheduler exploration. The study's relevance, rigorous methodology, and state-of-the-art classification accuracy achievement are highlighted in the summary.

The research (Niu et al., 2021) emphasizes the superiority of LBP + ORB features over LBP or ORB alone in trials, highlighting the efficacy and accuracy of a revolutionary facial expression identification framework. One of the study's limitations is that it only looks at static photographs. To address these issues, it is suggested that future studies examine face expressions in video sequences. The work uses databases from Cohn-Kanade (CK+), JAFFE, and MMI to extract characteristics from 208 MMI films, 593 CK+ sequences, and 213 JAFFE images. The suggested approach uses a face detection and region division algorithm to improve computational efficiency, combines ORB and LBP features, and uses SVM for classification. In conclusion, the work presents a fast and accurate face emotion detection algorithm and suggests directions for further study to increase speed and accuracy.

This paper (Saini et al., 2021) explains the way TensorFlow is implemented in machine learning to extract human emotion from photos and classify them into different emotional states. The approach uses a pre-trained TensorFlow model to predict scores, log loss, and human sentiments using a dataset of human photos annotated with sentiments. Results are displayed using tables and graphs. The study emphasizes how this method of forecasting human attitude has the potential to be widely used for a variety of platforms. It does, however, acknowledge shortcomings, pointing out decreased predictive accuracy for classes like "Contempt" and "Surprise" in the model's output.

This research (Mellouk & Handouzi, 2020) examines the latest developments in deep learning-based automatic facial emotion recognition (FER), highlighting the possibility of improved machine interpretation accuracy. It draws attention to certain drawbacks, such as issues with head posture changes, brightness, and the existing systems' narrow concentration on just six fundamental emotions. The paper urges the investigation of broader emotion ranges and multimodal recognition systems as a means of addressing these limitations in future research. From a methodological perspective, it thoroughly examines recent research, talking about CNN and CNN-LSTM architectures, numerous databases that are essential for precise emotion recognition, and preprocessing methods. In addition to stressing the need for more robust deep learning models and bigger databases, the research also emphasizes the value of multimodal techniques like physiological and audio-visual fusion for the best emotion identification results.

2.2 Summary

The literature review concludes by providing insight into the state of research on drivers drowsiness and mental health detection and offering an in-depth understanding of the main theories, approaches, and conclusions. This review highlights the importance of the current study by identifying significant gaps and challenges in the body of literature through a summary of many sources. This summary produced a theoretical framework that guides the investigation , providing a solid basis for the study. This literature review helps us this study by providing direction on the research design and method as well as presenting the work in the larger context of the academic discussion. The objective of our research is to make a significant contribution to the field and enhance the understanding of drivers drowsiness and mental health detection by filling in the identified gaps and improving on the knowledge gained from earlier investigations.

Chapter 3

Methodology

3.0 Introduction

From data collection and analysis to conclusion interpretation, research procedures comprise a wide range of systematic approaches and frameworks designed to guide the entire inquiry process. Here in this methodology I have included the process from data collection all the way to the model evaluation. Methodologies come in a variety of forms, but they can be broadly divided into two categories: quantitative and qualitative.

3.0.1 Quantitative Method:

A systematic and practical approach to studying observed events is known as quantitative methodology. It makes use of computational, statistical, or mathematical approaches to quantify data and draw conclusions that may be applied to a wider range of situations. This research methodology is represented by the gathering of numerical data and the objective analysis of patterns and relationships using statistical techniques. Quantitative research has a strong focus on objectivity and aims to reduce researcher bias. To improve the generalization of findings, bigger sample sizes are frequently used. To collect data systematically surveys, questionnaires, and experiments are examples of structured instruments that are frequently used. Strict statistical methods are used during the analysis phase to find patterns or trends in the numerical data.

3.0.2 Qualitative Method:

Investigative methods known as qualitative methodology take an exploratory approach to understanding the importance and contextual aspects of social occurrences, human experiences, and behavior. Unlike quantitative methodologies, which rely on numerical data, qualitative research focuses sensitive insights gained through in-depth analysis. Focusing on non-numerical data collection and interpretation, recognizing and embracing the subjectivity of human experiences, favoring smaller, purposeful sample sizes to enable in-depth investigation, having a flexible research design that changes as the study progresses, and leaning toward inductive analysis to find emerging themes and patterns are some of its significant features. Methods used in qualitative research include ethnography, case studies, focus groups, interviews, and content analysis. This approach offers depth and insight that may be difficult to obtain from numerical data alone, giving a comprehensive and contextual understanding of complicated situations.

In this research, I am using qualitative method. The data that I have used in this research are non-numeric data. The data that are used in this topic includes images of persons eye and faces. The eye dataset contains eye open and close images. And the face dataset contains emotions of different types.

In this research, I am using deep learning architectures to detect driver's drowsiness and mental health. In the last several years, machine learning has increased in research and been used in many different applications, such as multimedia concept retrieval, text mining, spam detection, video recommendation, and image classification (Alzubaidi et al., 2021). A effective classification method with considerable performance across a wide range of application fields is deep learning (Peng et al., 2017). Deep learning is a branch of machine learning that specializes on multi-layered neural networks, or "deep neural networks." The word "deep" describes the network's depth, indicating that there are multiple hidden layers between the input and output layers. These networks are able to recognize and convey intricate data hierarchies and patterns. There are many types of deep learning algorithms. The most well-known and often used algorithm is CNN (Alzubaidi et al., 2021). Similar to traditional neural networks, the structure of CNNs was modeled by neurons found in the brains of humans and other animals (Alzubaidi et al., 2021). Convolutional layers, pooling layers, fully connected layers, and activation functions are the four main parts of CNN. These four components are described below:

- **Convolutional Layer:** The convolutional layer is the most important part of the CNN design. It is comprised of a variety of convolutional filters, also referred to as kernels (Alzubaidi et al., 2021). To extract local patterns and characteristics from the input data, these filters are applied.
- **Pooling Layers:** By downsampling, pooling layers minimize the spatial dimensions of the incoming data. Max pooling, which keeps the maximum value in a limited area, and average pooling, which determines the average value, are common pooling processes.
- **Activation Functions:** In a neural network, activation functions are mathematical operations performed on a neuron's output. They give the network non-linearity, which enables it to learn and approximate intricate, non-linear correlations found in the data.
- **Fully Connected Layer:** At the end of every CNN design comes the Fully Connected Layer. Every neuron in this layer is coupled to every other neuron in the layer before it. It serves as the classifier for CNN. Being a feed-forward ANN, it operates on the same principles as a traditional multiple-layer perceptron neural network. The final pooling or convolutional layer provides the FC layer's input. Following flattening, the feature maps are converted into a vector, which is this input. The final CNN output is represented by the FC layer's output (Alzubaidi et al., 2021).

Here, I will run four models for emotion and four same models for drowsiness detection separately. After data collection, I will perform data preprocessing then train the model and lastly evaluate the models with the test data.The workflow diagram is given below:

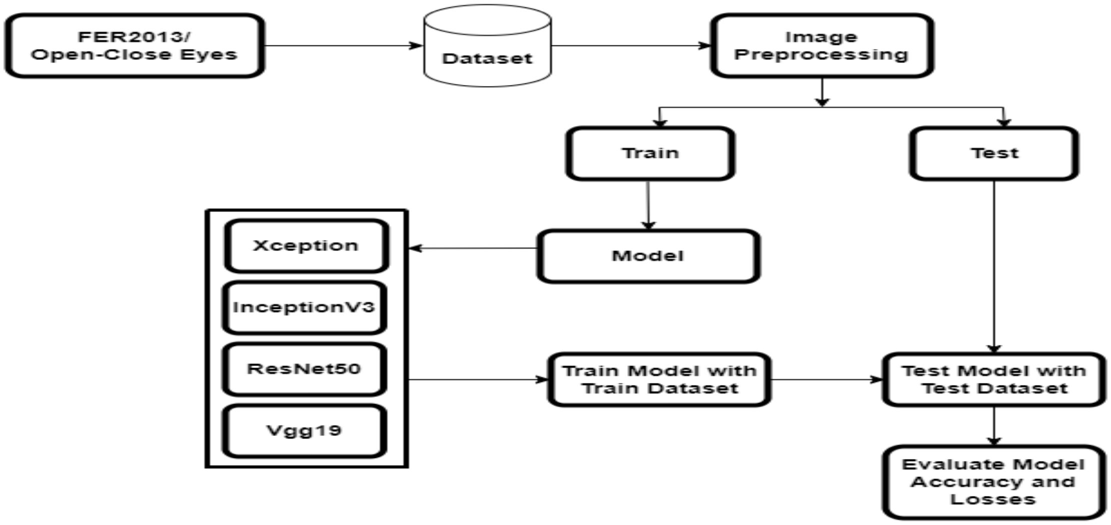


Figure 3. 1: Workflow diagram

3.2 Data Collection

For this research, I have collected data from two different datasets. One is FER-2013 and another one is drowsiness dataset for open eyes and closed eyes. Only static images are used to train all the models. The images of the emotion dataset contains only the face which are enough cropped and the drowsiness dataset contains only the eyes. I have used angry, fear, happy, neutral, sad and yawning for emotion classifications. The surprised faces that are similar to yawning has been used as yawning face. The the number of images per classification, also classification types along with sample images are given below:

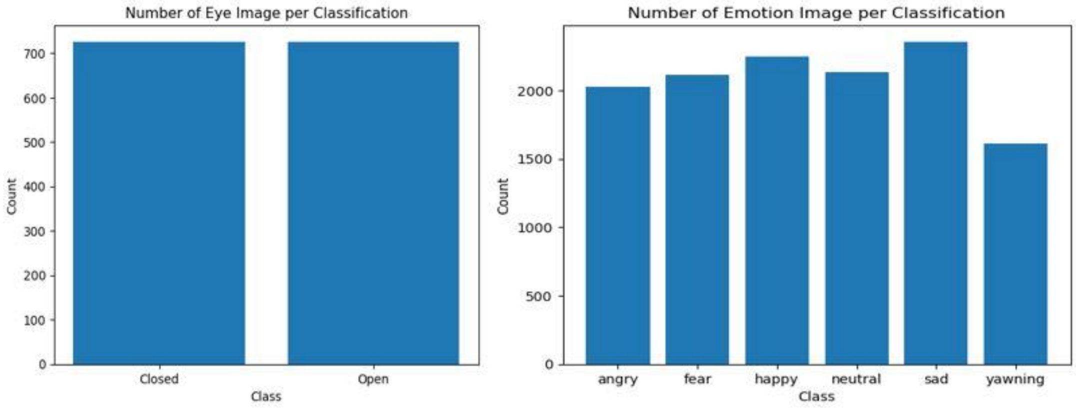


Figure 3. 2: Number of images per classification

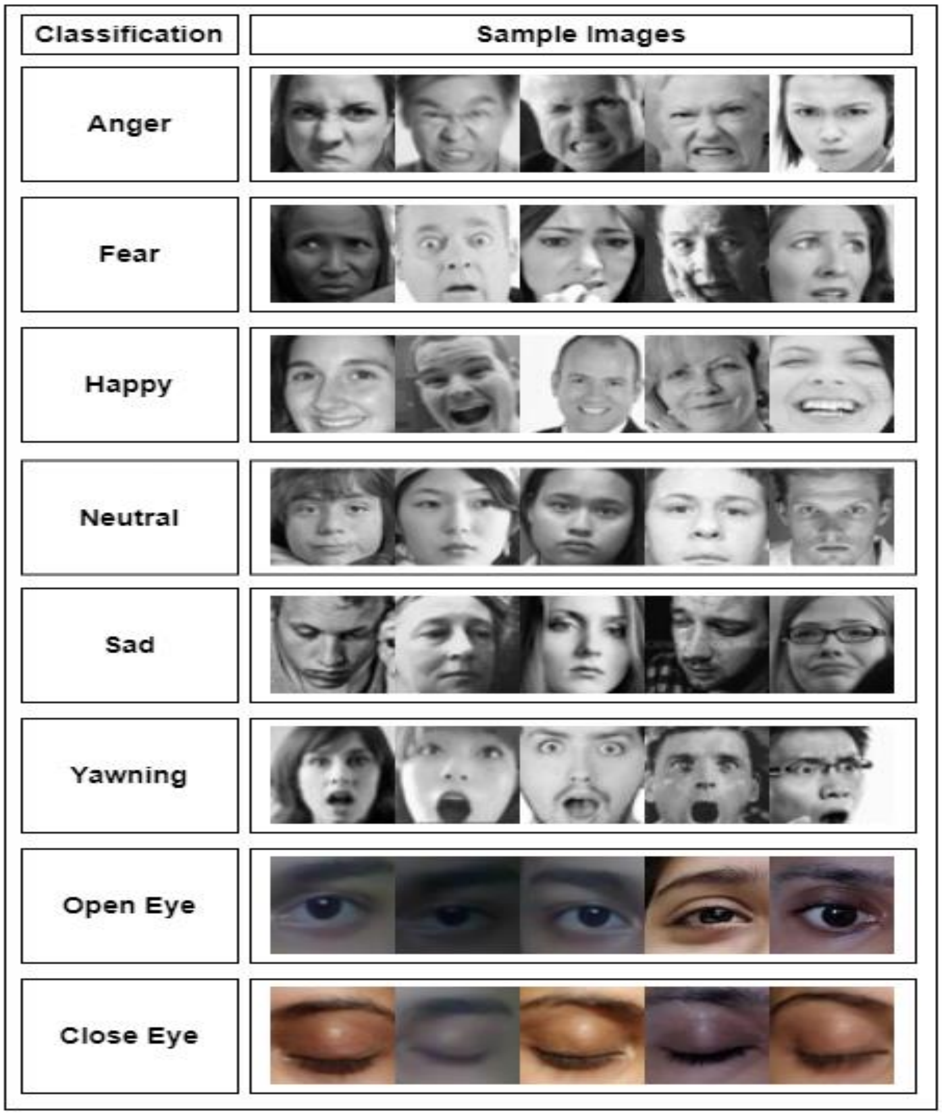


Figure 3. 3: Sample Images from Dataset

3.3 Data Preprocessing

In deep learning, data preprocessing is an essential step, especially for classification tasks. To prepare the raw data for model training, it must be cleaned and prepared. An effective machine learning model's performance and capacity for generalization can be greatly impacted by proper data preprocessing. Here, the TensorFlow Keras library's ImageDataGenerator class is used for data preprocessing. The data preprocessing are briefly given below:

Rescaling: By dividing each pixel value by 255, the image pixel values are normalized to the range [0, 1].

Rotation Range: The photos can be rotated randomly by up to 20 degrees.

Shear Range: Applying shear transformations with a maximum shear intensity of 0.2

Zoom Range: Up to a maximum of 0.2, the photos are randomly zoomed into.

Horizontal Flip: Flips images horizontally at random.

Image Resizing: Images are resized to 224 by 224 pixels.

Batch Size: During training, 32 samples are used for each gradient update.

Color Mode: RGB color mode is used to read images.

Shuffle: The test generator's images aren't shuffled to keep the order.

Class Mode: As the research requires multi-class categorization, the class mode for training generators is set to 'categorical'. Since no labels are given during testing, class_mode in the testing generator is set to None.

3.4 Model:

After data preprocessing I have train the models with the training dataset. The deep learning architectures that I have been used here in this research are Xception, InceptionV3, Resnet50 and VGG19. These neural network models are used for feature extraction and classification tasks. I have used pre-trained weights of the ImageNet dataset. Also, I have set up the models to be trained with the Adam optimizer at a learning rate of 0.0001 for eye open and close detection models and 0.001 for facial emotion detection models, with accuracy serving as the evaluation measure and categorical crossentropy as the loss function. Also, I have been monitoring on the validation loss using the ReduceLROnPlateau callback. The learning rate will be lowered by a factor of 0.2 if the validation loss does not improve after 10 consecutive epochs. The learning rate is kept from getting too low by setting the minimum learning rate to 0.0001. Additionally, I have implemented an EarlyStopping callback in my code and configured it to observe validity loss over a period of 10 epochs for eye open and close detection models and 30 epochs for facial emotion detection models. The model architectures are given below:

3.4.1 Xception: The developer of the Keras deep learning package, François Chollet, introduced the deep convolutional neural network design known as Xception (Extreme Inception). One of the main parts of the design, depthwise separable convolutions, are an area of expertise for Xception. In comparison to conventional convolutional neural networks, the architecture is intended to deliver higher performance and efficiency. The diagram and the description of this model are given below:

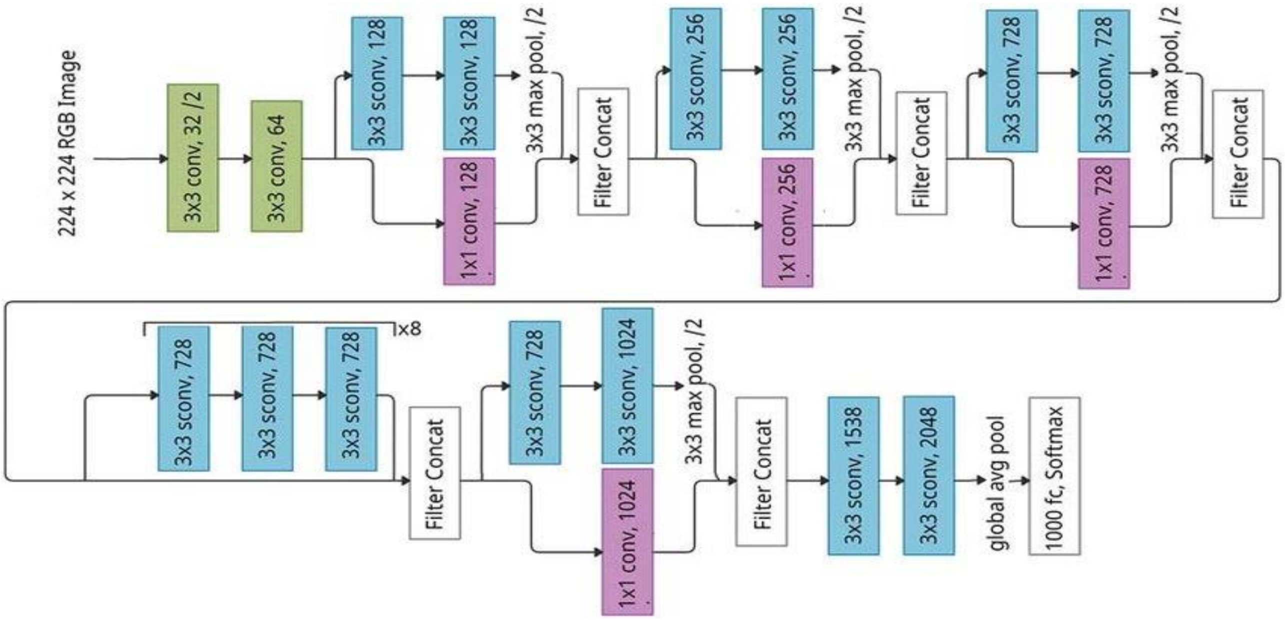


Figure 3. 4: Understanding of the Xception architecture. [Source: (Srinivasan et al., 2021)]

Entry Flow:

Input: A 299x299 RGB picture serves as the Xception model's input.

First Convolutional Block: Depthwise separable convolution with 32 filters, 3x3 kernel size, and 2 stride. Normalization of batches. ReLU activation.

Convolution Block 2: 3x3 kernel size, 64-filter, depth-wise separable convolution. Normalization of batches. ReLU activation.

Convolution Block 3: Identical as Block 2, but with an input-derived residual connection (shortcut connection).

Middle Flow: There are several identical blocks that make up the middle flow. Three depth-wise separable convolutional layers with batch normalization and ReLU activation are present in each block. In order to avoid the blocks and maintain information flow, residual connections are used. The intended network depth determines how many blocks are in the intermediate flow.

Exit Flow:

Block 13: 3x3 kernel size, 728 filters, depth-wise separable convolution. Normalization of batches. ReLU activation.

Block 14: 3x3 kernel size, depth-wise separable convolution with 1024 filters.Normalization of batches. ReLU activation.

Global Average Pooling: The Global Average Pooling layer is utilised to reduce spatial dimensions.

Fully connected Layer: Softmax activation for categorization in a fully connected layer.

The use of depthwise separable convolutions, in which conventional convolutions are split into depthwise and pointwise convolutions, is the main innovation of Xception. The goal of separating the spatial and channel-wise data is to more effectively capture dependencies. I have modified the input size as 224x224. Also, before the final classification layer, I have added a dropout layer. The Dropout layer is a regularization approach commonly used in neural networks to prevent overfitting. I have used 0.5 dropout rate, which means that at each training update, 50% of the input units will be randomly set to zero. Then for classification as I don't have 1000 classes for this system, for the final layer I have use the softmax activation function and the neurons will be 6 for facial emotions (angry, fear, happy, sad, neutral, yawning) and 2 for open and closed eyes.

3.4.2 InceptionV3: Convolutional neural network (CNN) architecture InceptionV3 is a member of the Inception model family. It is an advancement on the initial Inception (GoogLeNet) model and was unveiled by Google in 2015. The usage of inception modules, which are blocks of layers with varied filter sizes to collect characteristics at different scales, is what makes InceptionV3 so well-known. The diagram and the description of this model are given below:

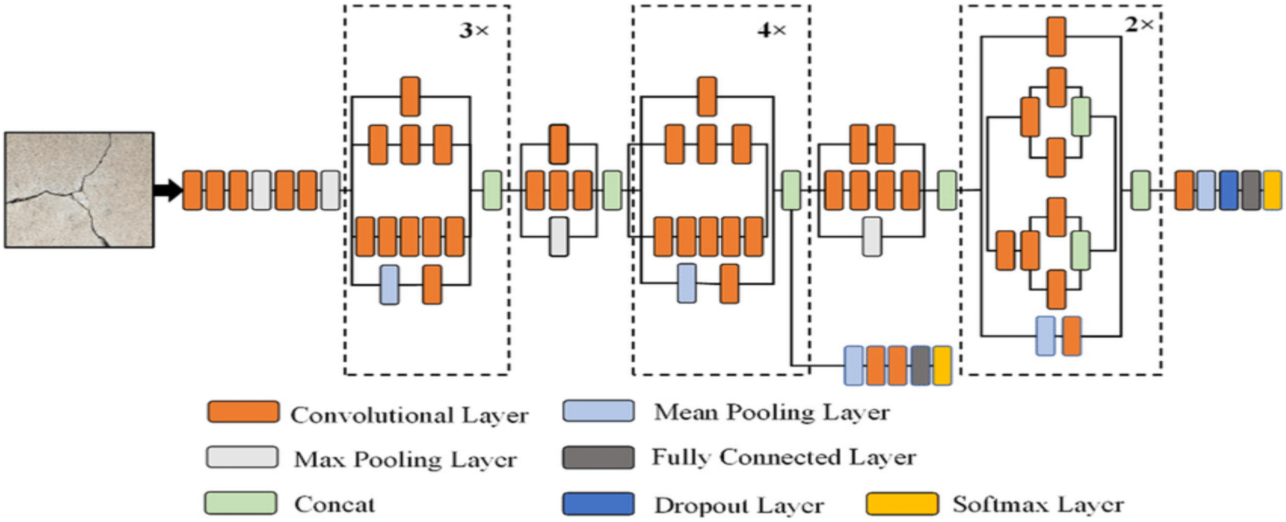


Figure 3. 5: Understanding of the InceptionV3 architecture. [Source: (Ali et al., 2021)]

Input Layer: The InceptionV3 model takes a 299x299 RGB picture as input.

Stem:

> First convolutional layer: 32 filters with a 3x3 kernel size and a 2 stride. Normalization of batches and ReLU activation.
>
> Convolutional Layer 2: 32 filters with a 3x3 kernel size make up Convolutional Layer 2. Normalization of batches and ReLU activation.
>
> Convolutional Layer 3: 64 filters with a 3x3 kernel. Normalization of batches and ReLU activation.
>
> Max Pooling Layer: 3 x 3 maximum pooling with a stride of 2.

Inception Blocks: InceptionV3 is made up of several repeating blocks, each of which is made up of different inception modules. A max pooling operation and 1x1, 3x3, and 5x5 convolutions are commonly included in each inception module. After every convolution, batch normalization and ReLU activation are implemented.

Optional Auxiliary Classifier: Two auxiliary classifiers are incorporated into InceptionV3 at various network depths to aid with gradient propagation during training. Global average pooling, a fully connected layer, batch normalization, and a softmax activation make up each auxiliary classifier. These classifiers help with training but are not utilized during inference.

Global Average Pooling: To decrease spatial dimensions, a global average pooling layer is used.

Fully Connected Layer: The last layer is a fully connected layer that uses classification-focused softmax activation.

The ImageNet dataset is used in InceptionV3. I have modified the input size as 224x224. Also, before the final classification layer, I have added a dropout layer. The Dropout layer is a regularization approach commonly used in neural networks to prevent overfitting. I have used 0.5 dropout rate, which means that at each training update, 50% of the input units will be randomly set to zero. Then for classification as I don’t have 1000 classes for this system, for the final layer I have use the softmax activation function and the neurons will be 6 for facial emotions (angry, fear, happy, sad, neutral, yawning) and 2 for open and closed eyes.

3.4.3 ResNet50: ResNet-50 (Residual Network with 50 layers) is a deep convolutional neural network architecture that introduced the concept of residual learning. It was designed to address the problem of vanishing gradients in very deep networks by using residual blocks. The diagram and the description of this model are given below:

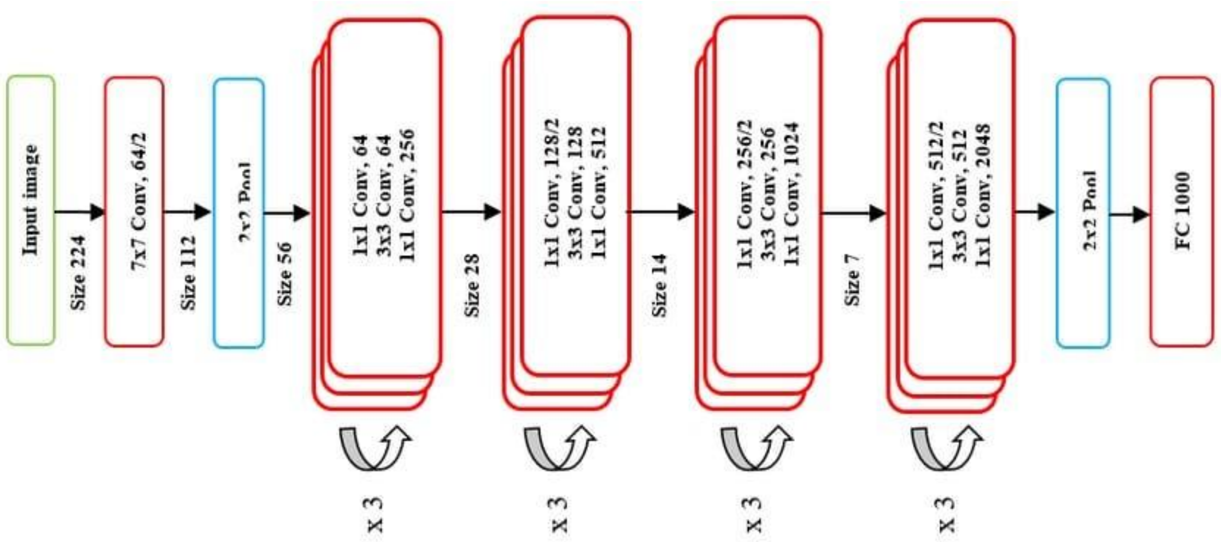


Figure 3. 6: Understanding of the ResNet50 architecture. [Source: towardsdatascience]

Input Layer: ResNet50 takes input images with a size of 224x224 pixels.

Convolutional and Pooling Layers:_The initial layers include a convolutional layer with a large kernel (7x7) followed by max-pooling.

Residual Blocks: The core of ResNet50 is the use of residual blocks. A residual block consists of two 3x3 convolutional layers surrounded by identity skip connections. The skip connection allows the input to bypass the convolutional layers, and the output is the sum of the original input and the result of the convolutional layers. ResNet-50 has several stacked residual blocks with varying numbers of filters (64, 128, 256, and 512). It uses rectified linear unit (ReLU) activation functions.

Global Average Pooling: After the stack of residual blocks, there is a global average pooling layer. This layer computes the average value of each feature map in the last convolutional layer, resulting in a fixed-size vector.

Fully Connected Layer: The global average pooling layer is followed by a fully connected layer with 1000 neurons, corresponding to the 1000 classes in the ImageNet dataset. The final layer uses softmax activation for classification.

This is the architecture of the ResNet50 model. ResNet-50 is characterized by its deep architecture (50 layers) and the use of residual blocks. I have added one dense layer before the final output layer with 128 neurons and added a 0.5 dropout. Also, for classification as I don't have 1000 classes for this system, for the output layer I have use the softmax activation function and the neurons will be 6 for facial emotions (angry, fear, happy, sad, neutral, yawning) and 2 for open and closed eyes.

3.4.4 VGG19: The University of Oxford's Visual Geometry Group introduced the VGG19, or Visual Geometry Group 19-layer model, a convolutional neural network (CNN) architecture. It is a development on the VGG16 model, except it has 19 layers as opposed to 16. The depth of the network is the main focus of VGG models, which have a straightforward and consistent architecture. The diagram and the description of this model are given below:

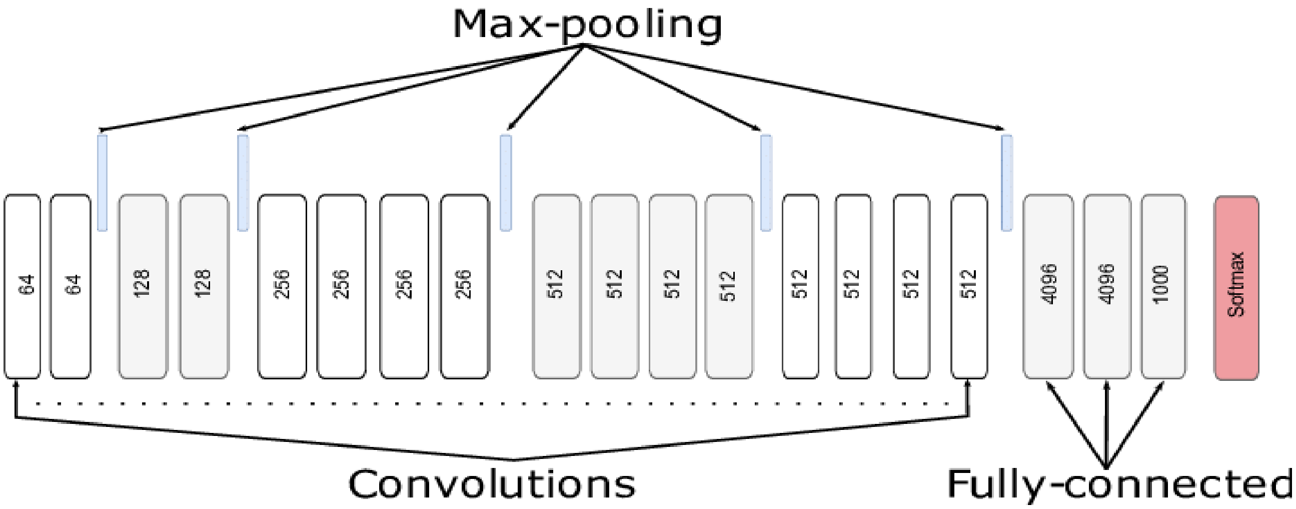


Figure 3. 7: Understanding of the VGG19 architecture. [Source: (Lagunas & Garces, 2018)]

Input Layer: 224x224 RGB image is used as the VGG19 model's input.

Convolutional Layers: There are 16 convolutional layers with small 3x3 filters in the architecture.To add non-linearity, a rectified linear unit (ReLU) activation function comes after each convolutional layer.

Max Pooling Layers: There is a max pooling layer with a 2x2 filter and a 2x2 stride after every set of convolutional layers.To minimize spatial dimensions and manage overfitting, max pooling is employed.

Fully Connected Layers: Three fully connected layers come after the convolutional layers.Each of the initial two completely connected layers contains 4,096 neurons.One thousand neurons make up the output layer, the third fully connected layer, which represents the number of classes in the ImageNet dataset.

Activation Function: After every convolutional and fully connected layer, ReLU activation functions are applied.

Softmax Layer: The output of the network is transformed into probability scores for each class by the softmax layer, which is the final layer.

This is the architecture of the VGG19 model. As I don't have 1000 classes for this system, I will use 128 and 64 neurons in the first two fully connected layer along with ReLU activation function. For classification, I have use the softmax activation function and the neurons will be 6 for facial emotions (angry, fear, happy, sad, neutral, yawning) and 2 for open and closed eyes in the output layer.

3.5 Model Evaluation

3.5.1 Accuracy:

Accuracy is one of the most popular metrics in multi-class classification (Grandini et al., 2020). The ratio of accurately predicted observations to total observations is known as accuracy. It offers an overall assessment of the performance of the model.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad \text{(Grandini et al., 2020)}$$

3.5.2 Precision:

The ratio of accurately predicted positive observations to the total number of predicted positives is known as precision (Grandini et al., 2020). It evaluates the positive predictions accuracy.

$$\text{Precision} = \frac{TP}{TP + FP} \quad \text{(Grandini et al., 2020)}$$

3.5.3 Recall:

The percentage of accurately predicted positive observations to all actual positive observations is known as recall (Grandini et al., 2020). It assesses how well the model can capture all pertinent cases.

$$\text{Recall} = \frac{\text{T P}}{\text{T P} + \text{F N}} \quad \text{(Grandini et al., 2020)}$$

3.5.4 F1-Score:

The harmonic mean of recall and precision is known as the F1-score (Hicks et al., 2022). It offers a well-rounded measurement that incorporates recall and precision.

$$\text{F1} - \text{Score} = \frac{2 \text{ x T P}}{2 \text{ x T P} + \text{FP} + \text{ F N}} \quad \text{(Hicks et al., 2022)}$$

3.5.5 Confusion Matrix:

A table that displays the counts of true positive, true negative, false positive, and false negative predictions can be used to create a confusion matrix, which provides an overview of the model's performance (Grandini et al., 2020).

$$M = \begin{pmatrix} TP & FN \\ FP & TN \end{pmatrix} \quad \text{(Hicks et al., 2022)}$$

3.5.6 Macro Average:

In a classification report, the term "macro average" refers to the process of determining performance metrics by computing the unweighted average of metrics (such precision, recall, and F1-score) over all classes.

3.5.7 Weighted Average:

A weighted average is a form of average in which various variables are assigned various weights. The weighted average is used to account for the imbalance in the number of examples across different classes when computing metrics like precision, recall, and F1-score in the context of a classification report. When dealing with multi-class classification difficulties, where certain classes may have more instances than others, this is especially helpful.

3.6 Summary

This study's approach consists of gathering data from selected sources and then going through an in-depth data preprocessing stage that involves data augmentation and other processes. Four deep learning architectures—Xception, InceptionV3, ResNet50, and VGG19—were chosen for developing models based on how well they fit the objectives of the study. After then, these models were trained using optimization methods and a predetermined set of hyperparameters. After training the models, each model's performance and capacity for generalization were evaluated using a different test dataset. The evaluation metrices that are used are accuracy, precision, recall, f1 score, confusion matrix, macro average, weighted average. For the particular goals of the study, this methodology provides an effective and accurate approach to the building and evaluation of deep learning models.

Chapter 4

Result and Discussion

4.0 Introduction

A research study's results chapter is important because it's the middle piece between the procedures and discussion/conclusion chapters. This section is essential to the clear structure and presentation of the research findings. The results chapter, which focuses on summarizing the data analysis findings, aims to respond to the research questions of the study by providing insights obtained from the carried out investigation. This chapter adds to the general comprehension of the study's findings by providing a well-organized and cohesive presentation of the results. It also lays the groundwork for the interpretation and discussion of the findings that will take place in other sections of the research document.

4.1 Previous Results

The results from previous study's described in literature review are given below: Pawan Wawage and Yogesh Deshpande uses AlexNet and VGG16 to predict car driver's emotions using facial expression, they acquired 87% accuracy (Wawage & Deshpande, 2022).

H. Varun Chand and J. Karthikeyan uses CNN as feature extractor and uses KNN and SVM as classifier to detect drivers drowsiness, they acquired 93% accuracy (Chand & Karthikeyan, 2022).

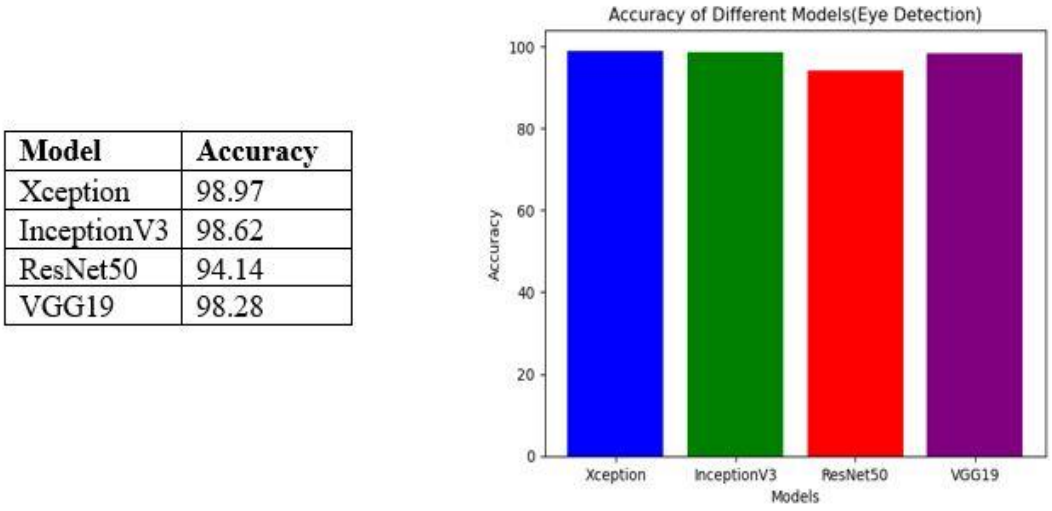
Kaviya and Arumugaprakash uses CNN to detect group facial emotions, and get an accuracy of 65% on FER2013 dataset (Kaviya & Arumugaprakash, 2020).

Kalpana Chowdary and others detect facial emotion for human–computer interaction applications using Resnet50, VGG19, Inception V3, MobileNet, where their experiment achieved an average accuracy of 96% (Chowdary et al., 2021).

Rateb Jabbar and others build Driver Drowsiness Detection Model Using Convolutional Neural Networks Techniques for Android Application using D2CNN-FLD, where they acquired an average of 83.33% of accuracy for all categories.

4.2 Result

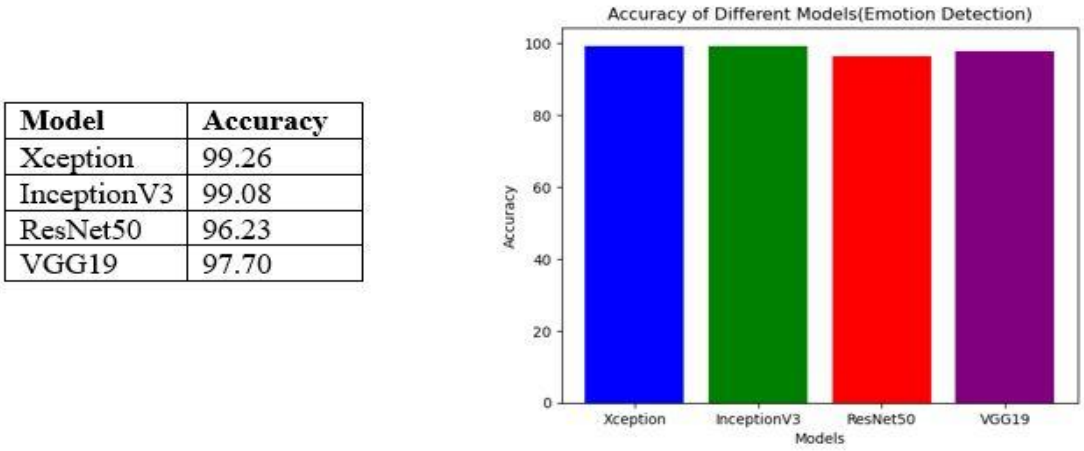
This chapter explains the performance results of the developed models. The model performance was evaluated based on the accuracy, recall, precision etc. In every model I have gone through 20 epochs for eye open close detection models and 100 epochs for emotion detection models. In eye open close detection models I have used 2 neuron based dense layer and for emotion detection models I have used 6 neuron based dense layer for classification as I have a total of 8 categories. The models accuracy are given below:



Model	Accuracy
Xception	98.97
InceptionV3	98.62
ResNet50	94.14
VGG19	98.28

Figure 4.1.1: Model Accuracy on Eye dataset

Here , we can see that the Xception model acquired the highest accuracy for eye open close detection, which is 98.97%. The other model InceptionV3, ResNet50 and VGG19 also acquired high accuracy but a slightly low accuracy than Xception. A bar chart is also given to visualize the accuracy for each model.



Model	Accuracy
Xception	99.26
InceptionV3	99.08
ResNet50	96.23
VGG19	97.70

Figure 4.1. 2: Model Accuracy on Emotion Dataset

For emotion detection models , we can see that the Xception model also acquired the highest accuracy here, which is 99.26%. The other model InceptionV3, ResNet50 and VGG19 also acquired high accuracy but a slightly low accuracy than Xception. A bar chart is also given to visualize the accuracy for each model. Now the difference graph of accuracy and loss during training and validation, also the confusion matrix and classification report of each model are given below:

Eye Open Close Detection

Xception:

The graph of accuracy and loss during training and validation are given below:

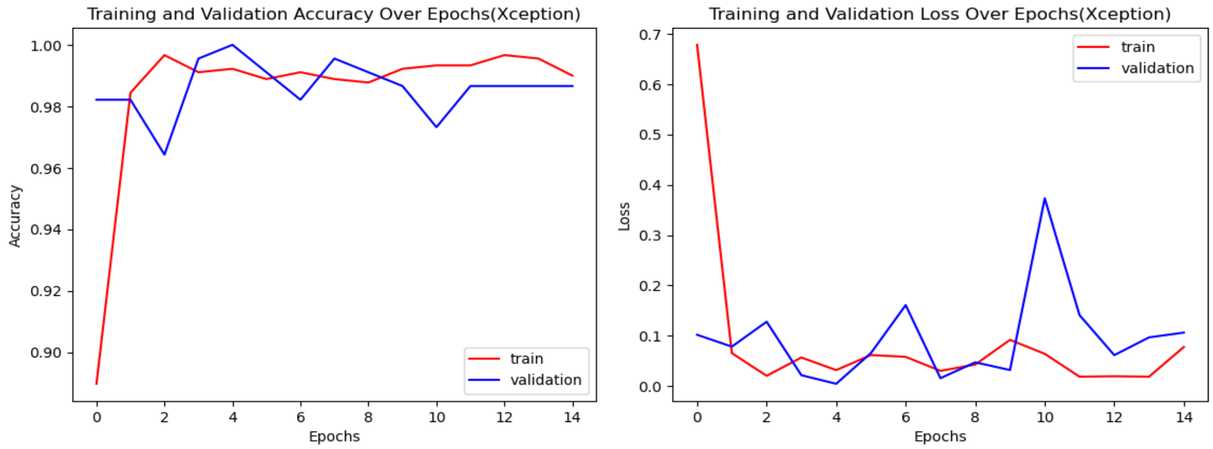


Figure 4.1. 3: Accuracy and Loss During Training and Validation(Xception)

As shown in the graphs, we can see that the accuracy is increasing with epochs and the loss is decreasing.

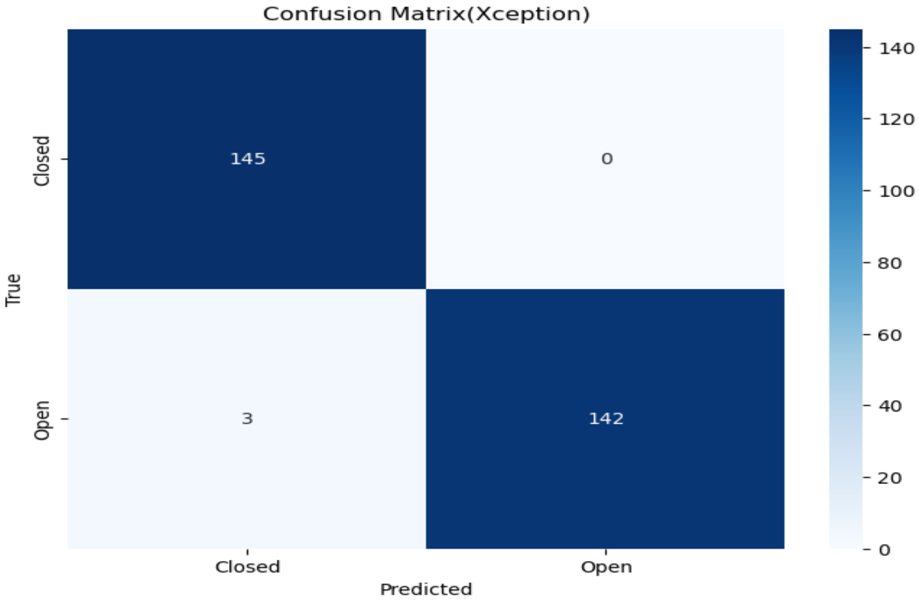
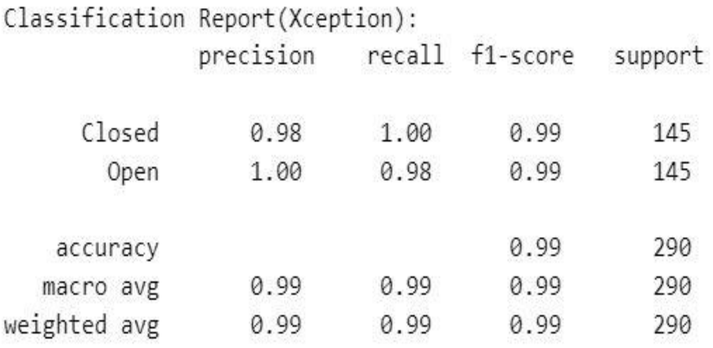


Figure 4.1. 4: Confusion Matrix (Xception)

From the confusion matrix, we can find that the model only misclassify 3 open eyes as closed eyes. The model perfectly predicted 145 closed eyes as closed eyes and 142 open eyes as open eyes.



```
Classification Report(Xception):
              precision    recall  f1-score   support

      Closed       0.98      1.00      0.99       145
        Open       1.00      0.98      0.99       145

    accuracy                           0.99       290
   macro avg       0.99      0.99      0.99       290
weighted avg       0.99      0.99      0.99       290
```

Figure 4.1. 5: Classification Report(Xception)

The Xception model's classification report shows excellent performance in dividing images into "Closed" and "Open" categories. With precision, recall, and F1-score metrics all averaging around 0.99, predictions are reliable and precise. The model's performance in identifying between the two groups is demonstrated by its 99% overall accuracy on a dataset including 290 samples.

InceptionV3:

The graph of accuracy and loss during training and validation are given below:

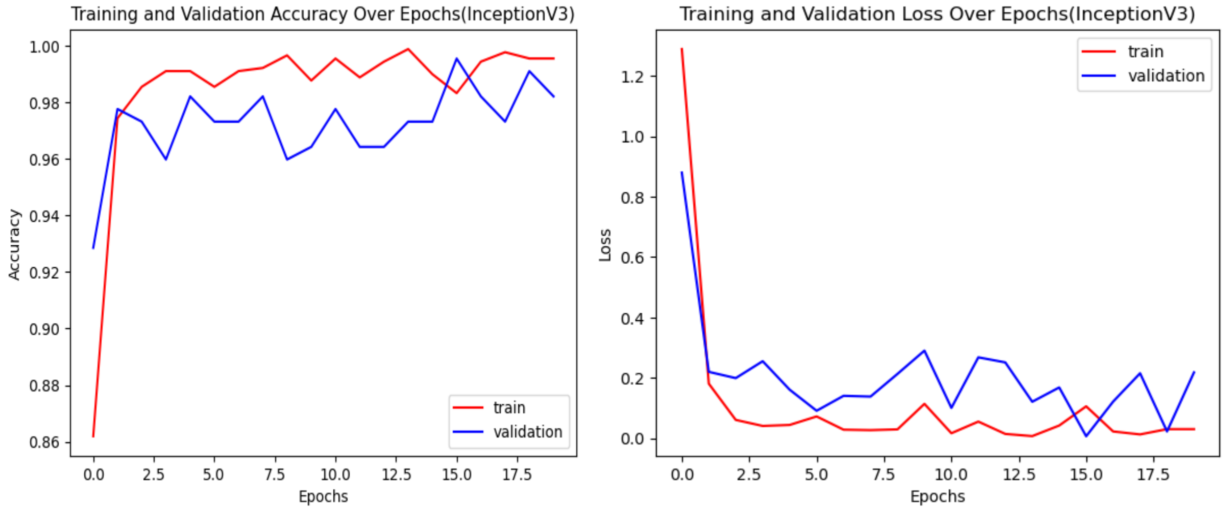


Figure 4.1. 6: Accuracy and Loss During Training and Validation(InceptionV3)

The graphs shows how the accuracy increases and the loss decreases with epochs.

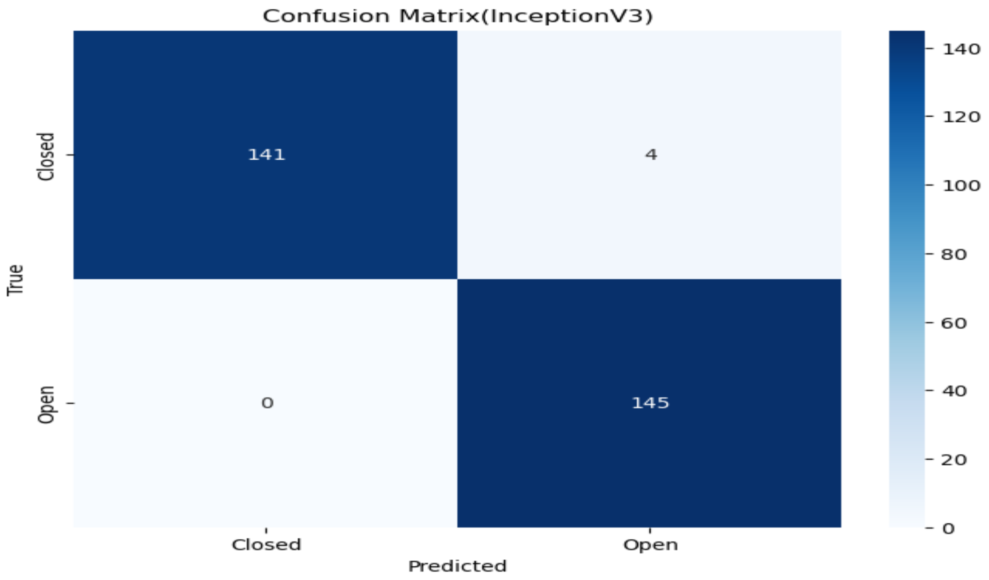
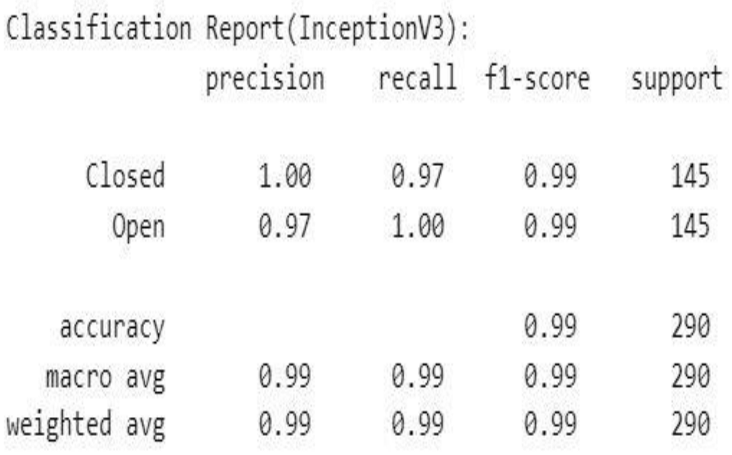


Figure 4.1. 7: Confusion Matrix(InceptionV3)

From the confusion matrix, we can find that the model only misclassify 4 closed eyes as open eyes. The model perfectly predicted 145 open eyes as open eyes and 141 closed eyes as closed eyes.



```
Classification Report(InceptionV3):
              precision    recall  f1-score   support

      Closed       1.00      0.97      0.99       145
        Open       0.97      1.00      0.99       145

    accuracy                           0.99       290
   macro avg       0.99      0.99      0.99       290
weighted avg       0.99      0.99      0.99       290
```

Figure 4.1. 8: Classification Report(InceptionV3)

With a 99% accuracy rate, the InceptionV3 model's categorization report shows outstanding accuracy. It shows a great capacity to correctly categorize cases with excellent precision, recall, and F1-score for both "Closed" and "Open" classes. The model's general performance across classes is further supported by the weighted averages and macros.

ResNet50:

The graph of accuracy and loss during training and validation are given below:

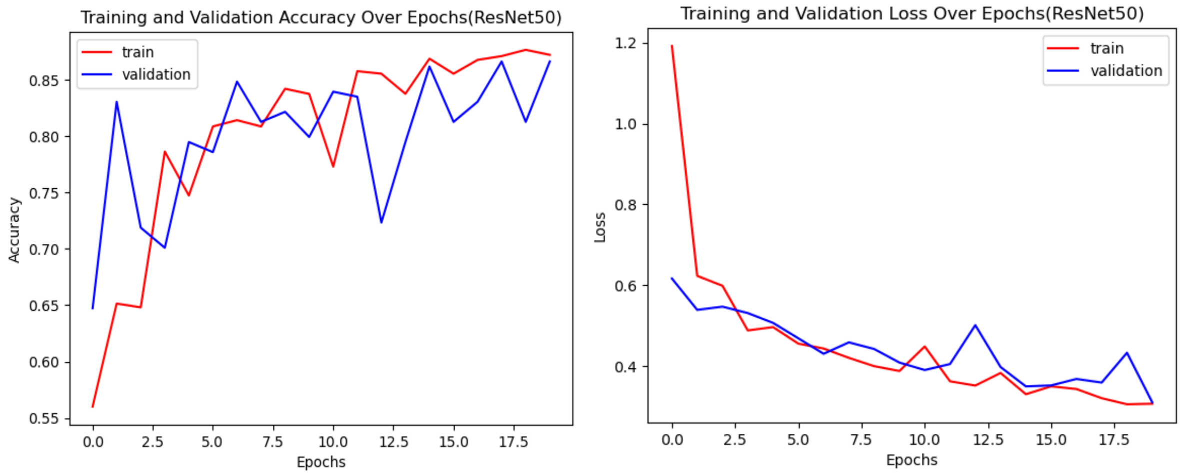


Figure 4.1. 9: Accuracy and Loss During Training and Validation(ResNet50)

The graphs shows how the accuracy increases and the loss decreases with epochs.

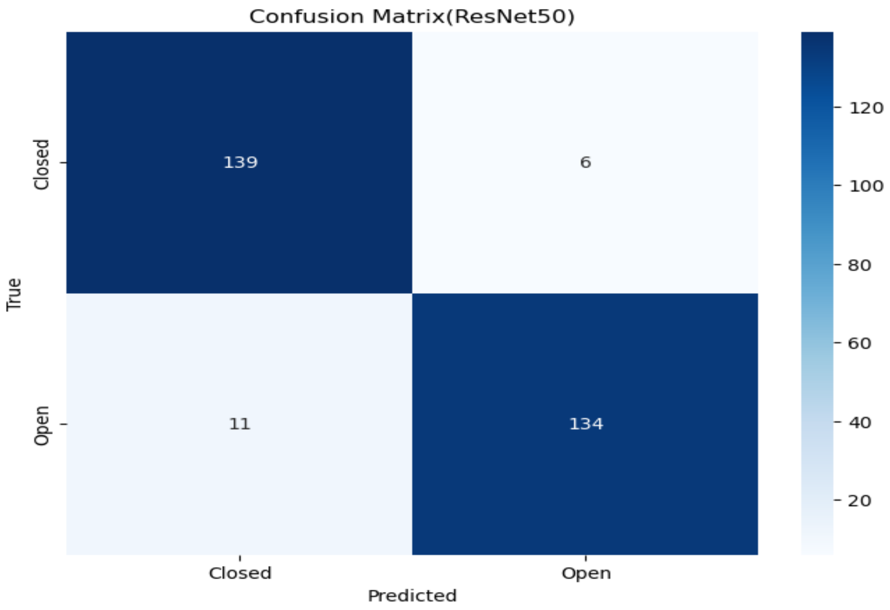
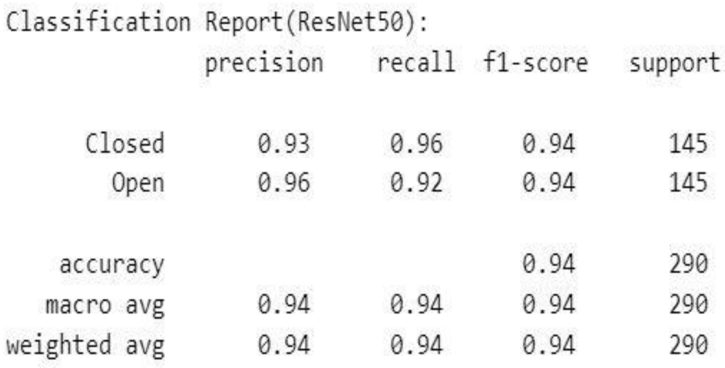


Figure 4.1. 10: Confusion Matrix(ResNet50)

The confusion matrix shows that the model misclassify 11 open eyes as closed eyes and 6 closed eyes as open eyes. The model perfectly predicted 134 open eyes as open eyes and 139 closed eyes as closed eyes.



```
Classification Report(ResNet50):
              precision    recall  f1-score   support

      Closed       0.93      0.96      0.94       145
        Open       0.96      0.92      0.94       145

    accuracy                           0.94       290
   macro avg       0.94      0.94      0.94       290
weighted avg       0.94      0.94      0.94       290
```

Figure 4.1. 11: Classification Report(ResNet50)

The ResNet50 model's performance served as the basis for the classification report. For the "Closed" and "Open" classes, it displays recall, precision, and F1-score. Using 290 samples in the dataset, the model's overall accuracy was 94%. High performance is shown in the report, with precision, recall, and F1-score values averaging 0.94, showing a fair and accurate classification for both classes.

VGG19:

The graph of accuracy and loss during training and validation are given below:

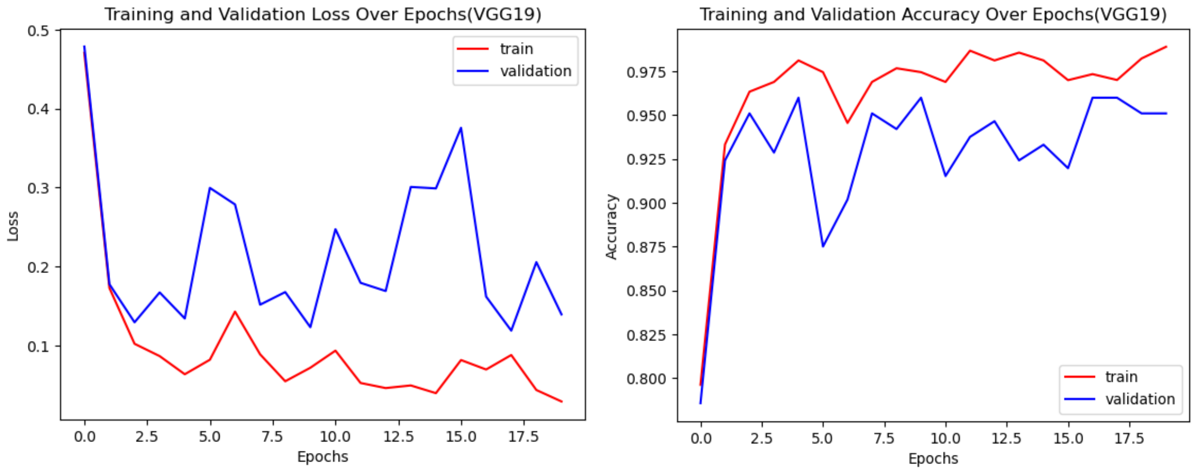


Figure 4.1. 12: Accuracy and Loss During Training and Validation(VGG19)

The graphs indicates how the accuracy increases and the loss decreases with epochs.

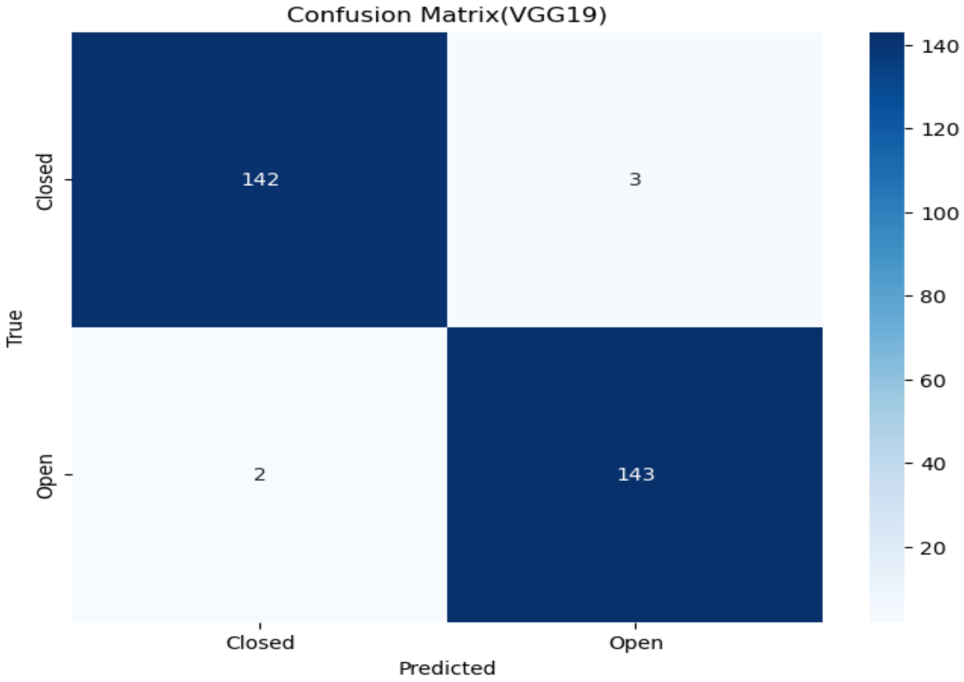
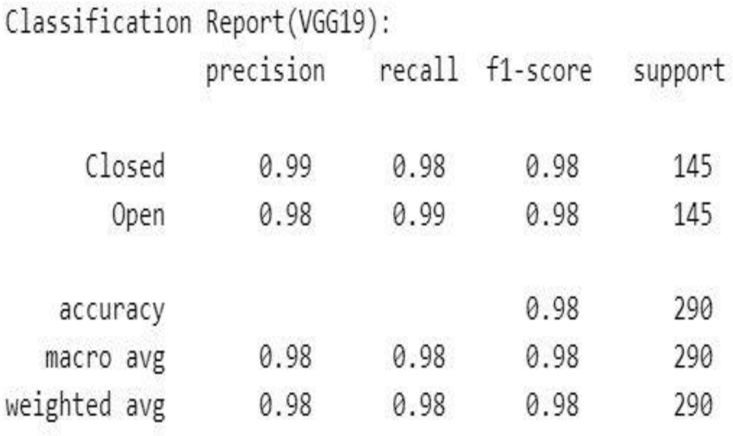


Figure 4.1. 13: Confusion Matrix(VGG19)

The confusion matrix shows that the model misclassify 2 open eyes as closed eyes and 3 closed eyes as open eyes. The model perfectly predicted 143 open eyes as open eyes and 142 closed eyes as closed eyes.



```
Classification Report(VGG19):
              precision    recall  f1-score   support

      Closed       0.99      0.98      0.98       145
        Open       0.98      0.99      0.98       145

    accuracy                           0.98       290
   macro avg       0.98      0.98      0.98       290
weighted avg       0.98      0.98      0.98       290
```

Figure 4.1. 14: Classification Report(VGG19)

The VGG19 model's classification report provides an overview of its performance on a dataset that is divided into two classes: "Closed" and "Open." For every class, the F1-score, accuracy, and recall metrics are given. With 290 samples in total, the model's accuracy was 98% overall. Additionally shown are the weighted averages and macros for precision, recall, and F1-score, all of which show good overall performance.

Facial Emotion Detection

Xception:

Below are the graphs, that shows the accuracy and loss throughout training and validation:

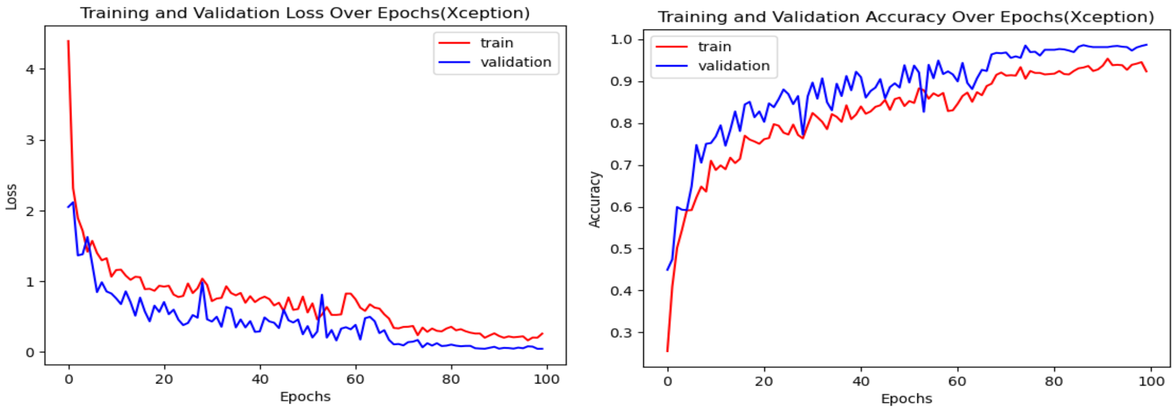


Figure 4.1. 15: Accuracy and Loss During Training and Validation(Xception)

The plots show how, with each epoch, accuracy rises and loss falls.

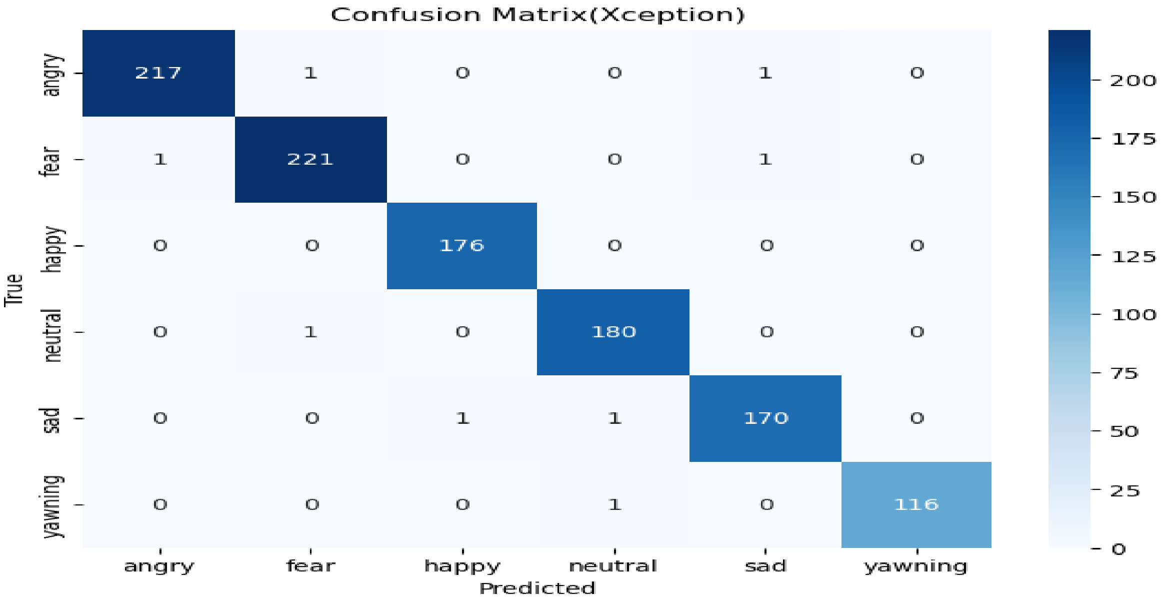
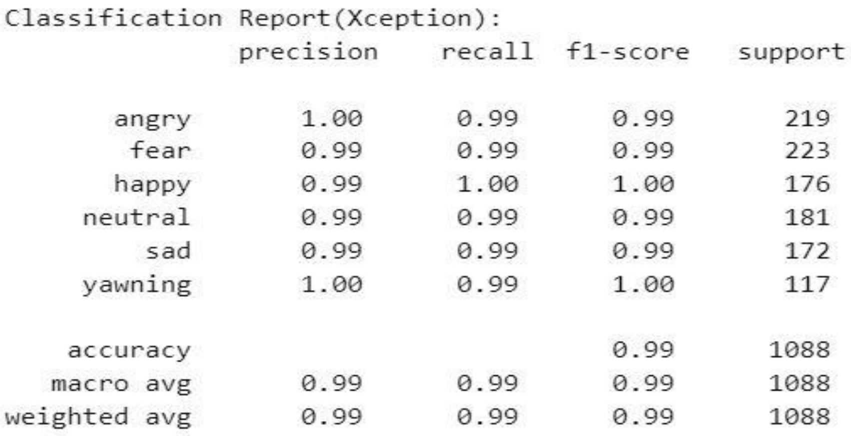


Figure 4.1. 16: Confusion Matrix(Xception)

The confusion matrix shows that the model misclassify some classes but overall it gives the highest accuracy.



```
Classification Report(Xception):
              precision    recall  f1-score   support

       angry       1.00      0.99      0.99       219
        fear       0.99      0.99      0.99       223
       happy       0.99      1.00      1.00       176
     neutral       0.99      0.99      0.99       181
         sad       0.99      0.99      0.99       172
     yawning       1.00      0.99      1.00       117

    accuracy                           0.99      1088
   macro avg       0.99      0.99      0.99      1088
weighted avg       0.99      0.99      0.99      1088
```

Figure 4.1. 17: Classification Report(Xception)

This classification report provides an overview of a model's (Xception) performance on a given dataset. With an overall accuracy of 99%, the model demonstrated strong precision, recall, and F1-score across multiple emotion classes (angry, fear, happy, neutral, sad, and yawning). Strong overall model performance is also indicated by the weighted averages and macros.

InceptionV3:

Below are the graphs, that shows the accuracy and loss throughout training and validation:

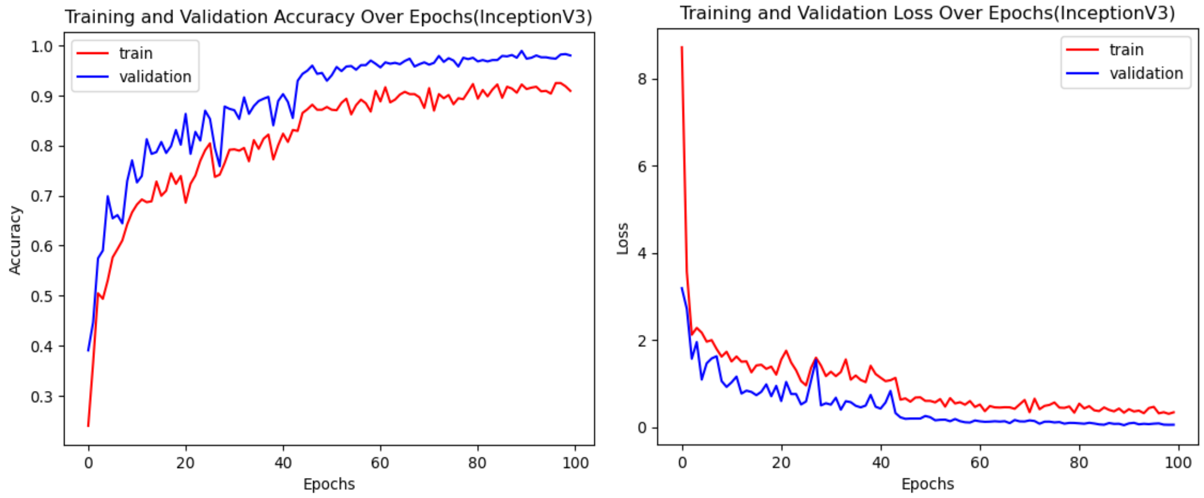


Figure 4.1. 18: Accuracy and Loss During Training and Validation(InceptionV3)

The plots shows how, with each epoch, accuracy rises and loss falls.

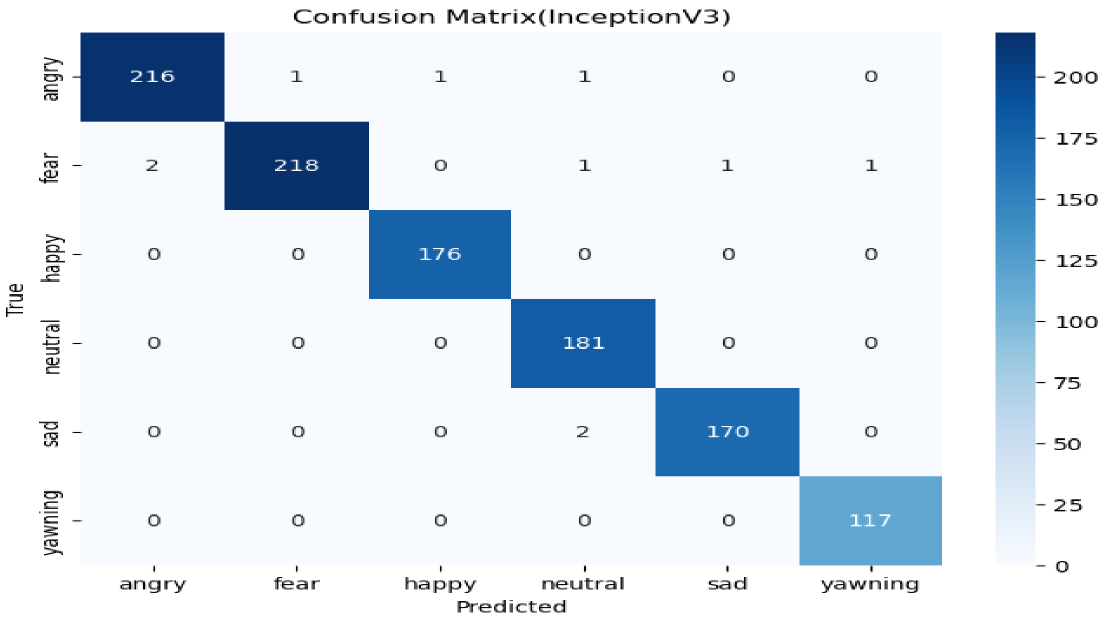
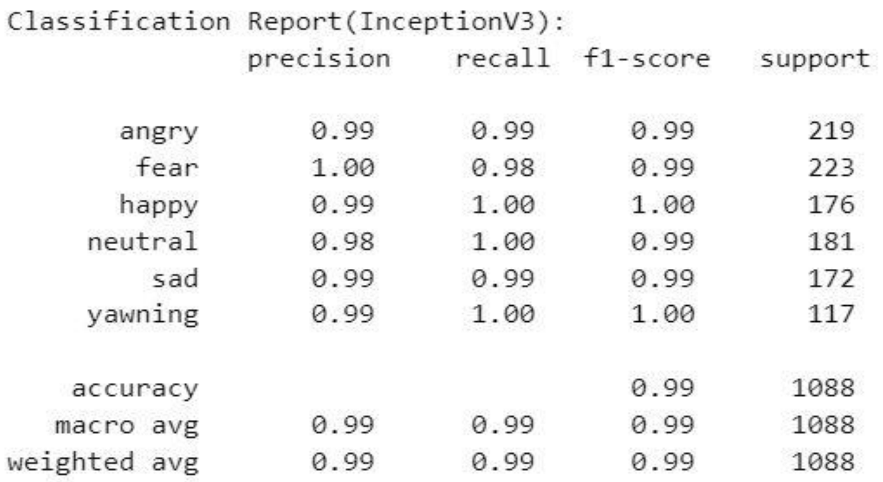


Figure 4.1. 19: Confusion Matrix(InceptionV3)

The confusion matrix shows that the model misclassify 1 or 2 classes but overall it gives a higher accuracy.



```
Classification Report(InceptionV3):
              precision    recall  f1-score   support

       angry       0.99      0.99      0.99       219
        fear       1.00      0.98      0.99       223
       happy       0.99      1.00      1.00       176
     neutral       0.98      1.00      0.99       181
         sad       0.99      0.99      0.99       172
     yawning       0.99      1.00      1.00       117

    accuracy                           0.99      1088
   macro avg       0.99      0.99      0.99      1088
weighted avg       0.99      0.99      0.99      1088
```

Figure 4.1. 20: Classification Report(InceptionV3)

This is the InceptionV3 model's classification report for a dataset that has six emotion classes. Metrics including recall, precision, and F1-score are included in the report for each class (angry, fear, happy, neutral, sad, and yawning). With a 99% overall accuracy, the model performs well. The precision, recall, and F1-score macro and weighted averages are all 99%, suggesting stable and balanced performance in every class.

ResNet50:

The graphs shows the accuracy and loss during training and validation:

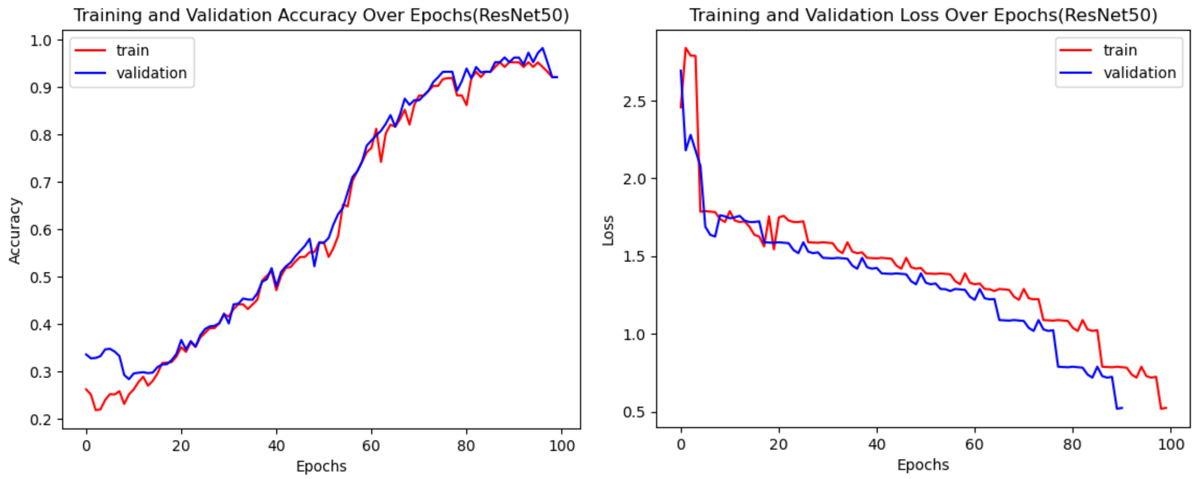


Figure 4.1. 21: Accuracy and Loss During Training and Validation(ResNet50)

The plots shows how, with each epoch, accuracy rises and loss falls.

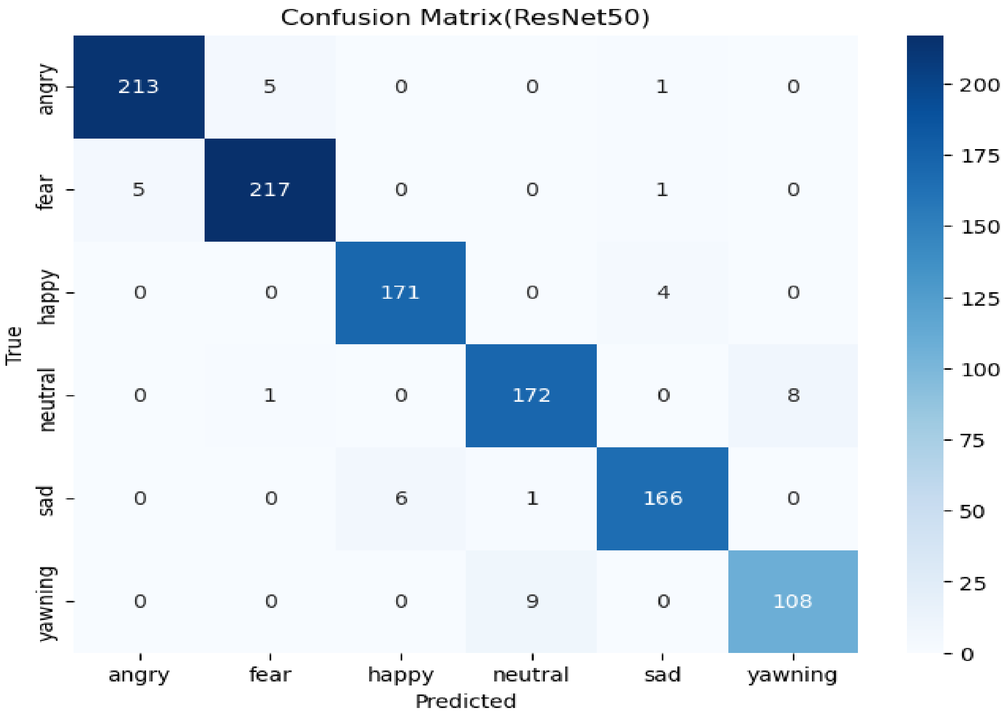
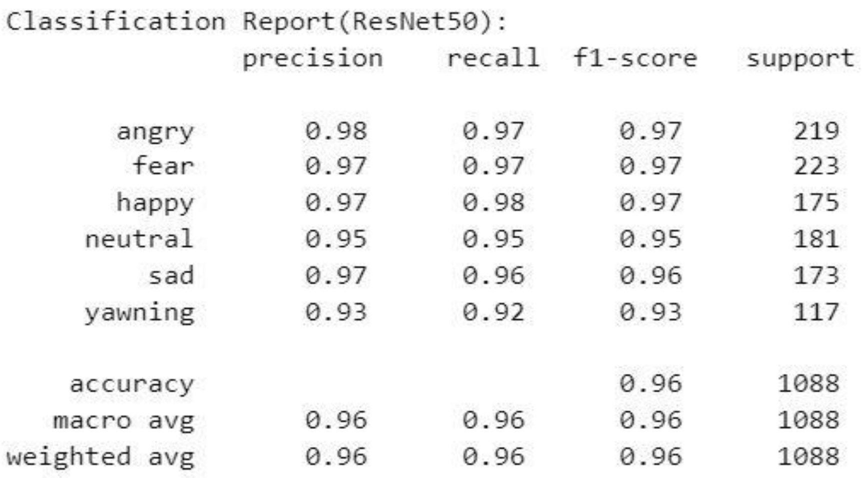


Figure 4.1. 22: Confusion Matrix(ResNet50)

The confusion matrix shows that the model misclassify some classes but overall it gives a good accuracy.



```
Classification Report(ResNet50):
              precision    recall  f1-score   support

       angry       0.98      0.97      0.97       219
        fear       0.97      0.97      0.97       223
       happy       0.97      0.98      0.97       175
     neutral       0.95      0.95      0.95       181
         sad       0.97      0.96      0.96       173
     yawning       0.93      0.92      0.93       117

    accuracy                           0.96      1088
   macro avg       0.96      0.96      0.96      1088
weighted avg       0.96      0.96      0.96      1088
```

Figure 4.1. 23: Classification Report(ResNet50)

The results of a ResNet50 model's performance metrics on a dataset of six emotion classes are displayed in this classification report. For every kind of emotion (anger, frightened, happy, neutral, sad, yawning), there is a precision, recall, and F1-score offered. Furthermore, the percentage of successfully identified instances is provided as 96%, indicating overall accuracy. Additionally included are the macro and weighted averages for precision, recall, and F1-score, which show how well the model performed overall across all classes. All things considered, the model performs successfully, consistently achieving excellent results in various emotional areas.

VGG19:

The graphs shows the accuracy and loss during training and validation:

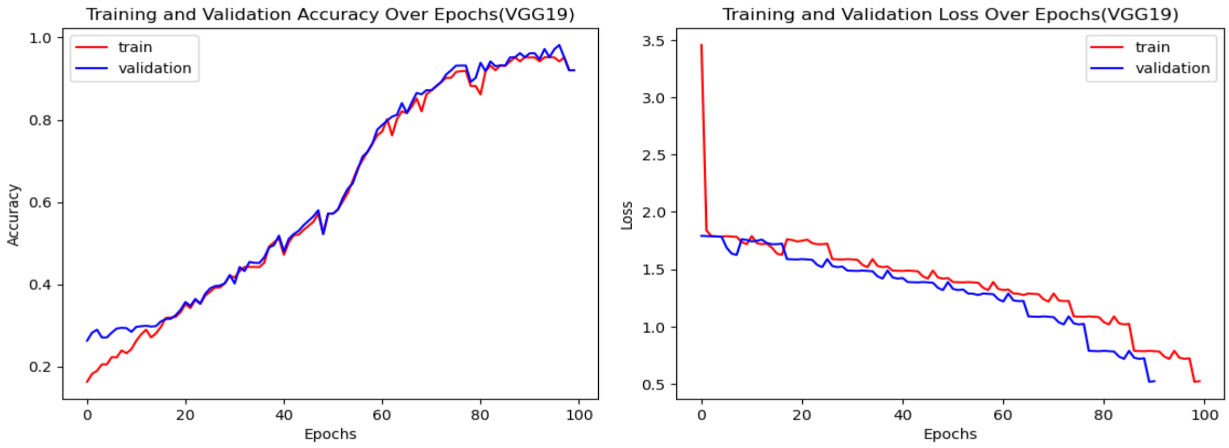


Figure 4.1. 24: Accuracy and Loss During Training and Validation(VGG19)

The graphs shows how, with each epoch, accuracy rises and loss falls.

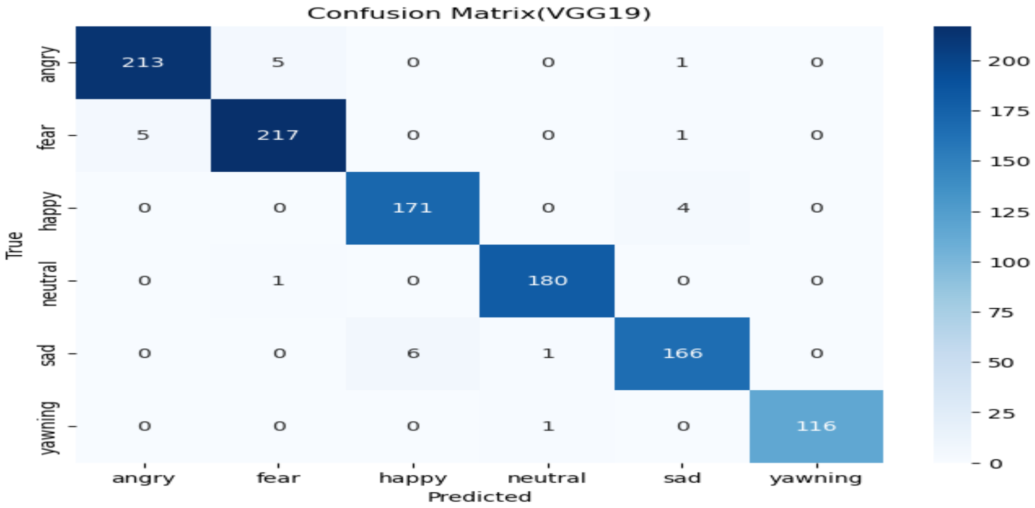
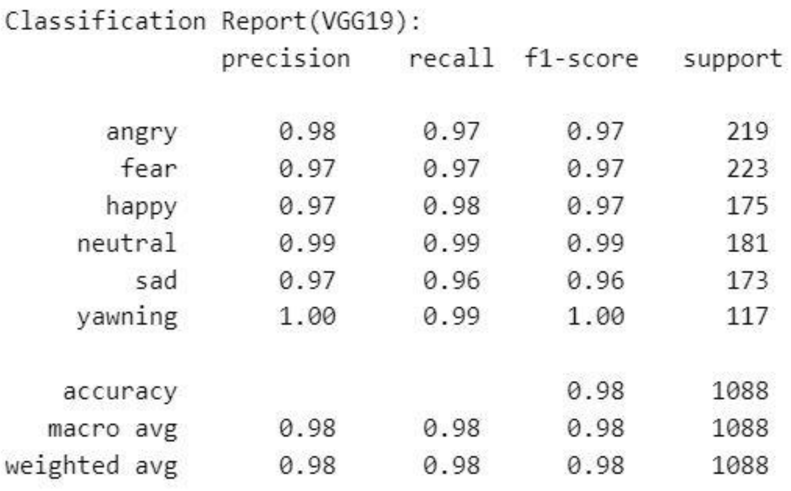


Figure 4.1. 25: Confusion Matrix(VGG19)

The confusion matrix shows that the model misclassify some classes but overall it gives a good accuracy.



```
Classification Report(VGG19):
              precision    recall  f1-score   support

       angry       0.98      0.97      0.97       219
        fear       0.97      0.97      0.97       223
       happy       0.97      0.98      0.97       175
     neutral       0.99      0.99      0.99       181
         sad       0.97      0.96      0.96       173
     yawning       1.00      0.99      1.00       117

    accuracy                           0.98      1088
   macro avg       0.98      0.98      0.98      1088
weighted avg       0.98      0.98      0.98      1088
```

Figure 4.1. 26: Classification Report(VGG19)

A VGG19 model's classification report on a dataset includes metrics for precision, recall, and F1-scores for several emotion classes. With a 98% total accuracy rate, the model performs well. It is particularly good at identifying neutral expressions (99% precision and recall) and yawning (100% precision and 99% recall). A balanced and accurate categorization across all classes is indicated by the weighted and macro averages of 98% for precision, recall, and F1-score.

4.3 Discussion

I have run the model for 20 epochs on eye open close dataset and 100 epochs on facial emotion dataset. The models performance was evaluated based on the accuracy, recall, precision etc. I have used 2 neuron-based dense layer for eye open-close detection models, and since I have eight categories in total, I have used 6 neuron-based dense layer for emotion detection models. In this research, I have used 4 models(Xception, InceptionV3, ResNet50 and VGG19). The four models accuracy are given in the result section. Xception model performs best for both eye and emotion dataset. The Xception model gives 98.97% accuracy for eye open and close detection dataset and 99.26% for facial emotion detection dataset. For eye dataset, the Xception model only misclassifies 3 open eyes as closed eyes. And for emotion dataset, the model misclassify some classes but overall it gives the highest accuracy. With exceptional precision, recall, and F1-score metrics average approximately 0.99, the Xception model performs exceptionally well when classifying images into "Closed" and "Open" categories. With a total accuracy of 99% across 290 samples in the dataset, it produces reliable and accurate predictions. Across multiple emotion classes (angry, fear, happy, neutral, sad, and yawning), the Xception model demonstrated excellent precision, recall, and F1-score, achieving an outstanding 99% overall accuracy on the dataset. Both weighted averages and macro average showed great performance, showing the overall efficiency of the methodology.

The idea of real-time implementation of this research is described through the following diagram:

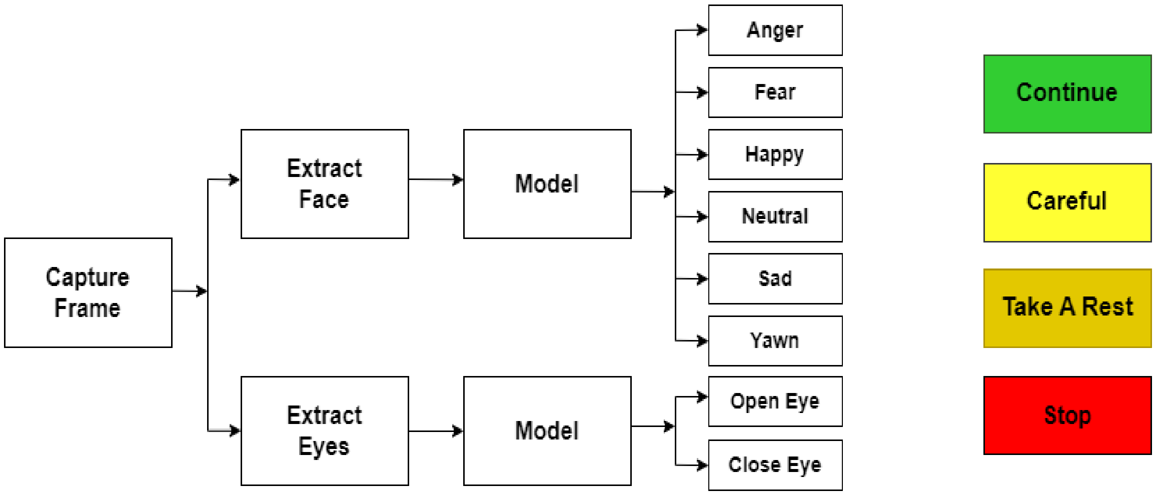


Figure 4.2. 1: The Real-Time Workflow

In real-time, the camera will capture the frames. And from those frames face and eyes will be extracted. Then these extracted face and eyes will be given to the models for detection. The facial emotion detection model will classify the face as angry, fear, happy, neutral, sad or yawning. And the eye open close detection model will classify eyes as open or closed.

Based on the combination of emotion and eye open close detection the system will show the following message(Continue, Careful, Take a rest or Stop). For example, for close eyes it need to show stop message, in this scenario the facial emotion might not be considered. For open eyes the other three messages will be shown based on the emotions.Such as for anger, fear and happy it will show careful. For neutral it will show continue. For sad and yawning it will show take a rest. This can be the real-time implementation for this research.

4.4 Summary

In summary, the result chapter includes all the model evaluation results for the 4 mentioned models. Along them Xception model works best for both tasks. For eye detection, it obtained 98.97% accuracy, and for face emotion detection, 99.26% accuracy. The camera records frames in real time, and will extract eyes and faces. These features are then incorporated into detection models. For example, the eye model classifies eyes as open or closed, while the facial expression model classifies faces as angry, fear, happy, neutral, sad, or yawning.

Chapter 5

Conclusion

5.1 Conclusion

The key goal of this research is to combine the drivers drowsiness state with emotional state identified through their facial expressions, to achieve a system that can be used to reduce road accicents. Throughout the world, millions of people are dying for irresponsible driving. This study suggests an improved pretrained model based approach for detecting driver's inattention. In most of the drowsiness detection system the main focus is to detect yawning or close eyes but I have considered to add facial emotion as extra safety feature which will keep track if the driver is mentally stable to drive. The main objective of this extra feature is to detect driver's mental mood so that the system can alert drivers if they gets angry or too excited for speeding or other things. I have used four models for eye open close detection and the same models for facial emotion detection. The four models are Xception, InceptionV3, ResNet50 and VGG19. Images are fed into these models to classify eyes as open or close, also 6 emotions are given to detect the drivers emotional state. The tasks of feature extraction and classification are performed by these neural network models. For the models in this study, I used 20 epochs for eye open-close training and 100 epochs for facial expression. For the eye open closed dataset, the following models accuracy are: Xception 98.97%, InceptionV3 98.62%, ResNet50 94.14%, and VGG19 98.28%. Additionally, the models' accuracy for the face emotion recognition dataset is 99.26% for Xception, 99.08% for InceptionV3, 96.23% for ResNet50, and 97.70% for VGG19. The Xception outperformed other models in both tasks. For eye detection, it obtained 98.97% accuracy, and for face emotion detection, 99.26% accuracy. The Xception model showed remarkable precision, recall, and F1-score across several emotion classes (angry, fear, happy, neutral, sad, and yawning), with an impressive 99% overall accuracy on the dataset. The macro average and weighted averages both performed successfully, highlighting the methodology's general effectiveness. The above model will identify instable drivers when it detects yawning face, close eyes and inappropriate mentality using deep learning model. Although I have managed to get a higher result, there is still much space for development. More human photos that are diverse in kind are essential. In real-time the camera will record frames and will extract eyes and faces. These features are then put into detection models. For example, the eye model classifies eyes as open or closed, while the facial expression model classifies faces as angry, fear, happy, neutral, sad, or yawning. Based on the combination of these classifications, the system will notify the drivers about their mental conditions. The system might display one of the following messages (Continue, Careful, Take a rest or Stop) based on the combination of the emotion and eye open close detection. So, my main focus is to mark down the drowsy drivers, increase drivers ability to focus, help to reduce road accidents, ensure drivers and their passengers safety by successfully detecting if the drivers eyes are closed or open and also the drivers mental state through their facial expressions.

5.2 Limitation

The above models gave high accuracy but there are other deeper models that might outperform these models. The dataset size for eye open close detection was moderate and it gave a high accuracy for test sets, but in real time scenario it might need to be trained with more datasets. Also for emotion, as the dataset is limited because I have only used FER2013 dataset for training and testing. So, the model might not perform well for other datasets. Also, I have considered 5 facial emotions and 1 emotion for yawning, but in real scenario there might be other emotions or there might be some combination between multiple facial emotions. Also I did not consider night time scenarios.

5.3 Future work

In this research I have developed 4 models for eye open close detection and 4 models for drivers facial emotion detection dataset. Both these models identify separately. One model only identifies if the eyes are closed or open. And the other model identifies the drivers mental state from their facial expressions. So, there are scope to find if there are any possible way to combine two models into one model.Also, by combining night vision technology, the system's functionality can be further improved. By adding this, the system would be ensured to function properly in situations involving driving at night, improving overall performance and safety. By adding night vision capabilities, the system will be able to identify and recognize objects better and adapt to different driving circumstances more easily. This will increase the system's usefulness and dependability in a variety of situations. Also voice recognition technology can be used to analyze a driver's emotional state. It is possible to identify emotions that express information about the driver's present mood by paying attention to their vocal expressions . This method makes use of the natural connection between voice patterns and emotions to provide a deeper understanding of the driver's emotional-condition.

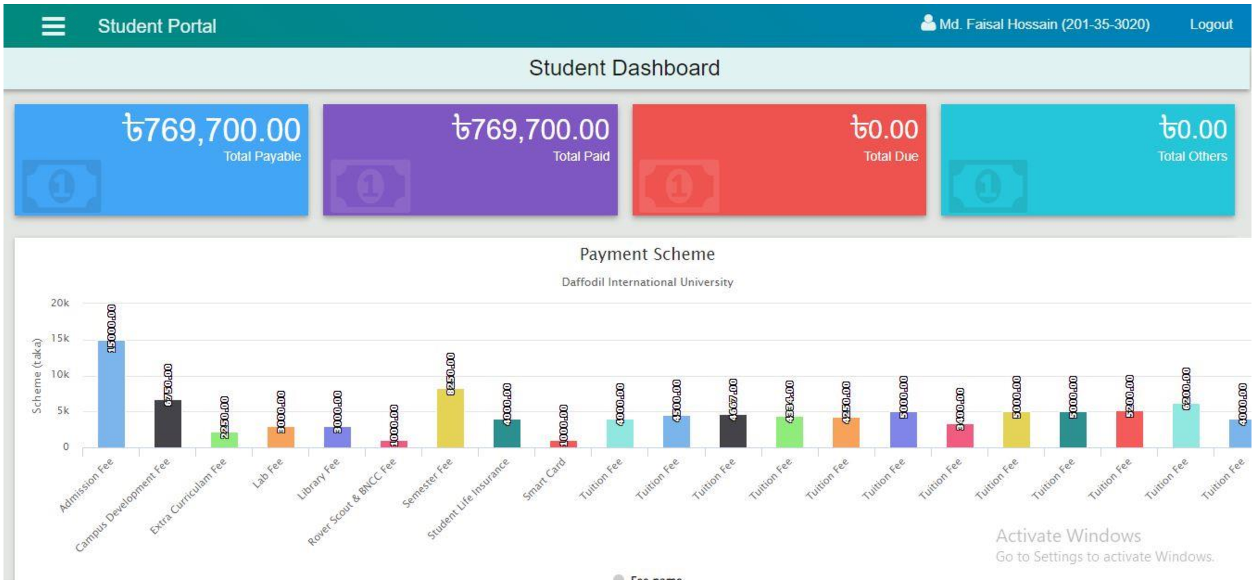
Additionally, driving behaviours such as braking, safedriving, speeding can also be considered to present a more improved system. The combination of these things can give a more better and upgrade system to mark down the drowsy drivers, increase drivers ability to focus, help to reduce road accidents.

References

- Wawage, P., & Deshpande, Y. (2022). Real-time prediction of car driver’s emotions using facial expression with a convolutional neural network-based intelligent system. International Journal of Performability Engineering, 18(11), 791.
    - Chand, H. V., & Karthikeyan, J. (2022). CNN Based Driver Drowsiness Detection System Using Emotion Analysis. Intelligent Automation & Soft Computing, 31(2).
    - Kaviya, P., & Arumugaprakash, T. (2020, June). Group facial emotion analysis system using convolutional neural network. In 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184) (pp. 643-647). IEEE.
    - Anigbogu, K. S., Inyiama, H. C., Onyenwe, I., & Anigbogu, S. O. (2022). Driver behavior model for healthy driving style using machine learning methods.
    - Vanjani, H. B., & Varyani, U. (2019). Identify dozyness of person using deep learning. International journal of applied engineering research, 14(4), 845-848.
    - Monish, M. DRIVER’S BEHAVIORAL RECOGNITION USING CNN.
    - Jaiswal, S., & Nandi, G. C. (2020). Robust real-time emotion detection system using CNN architecture. Neural Computing and Applications, 32(15), 11253-11262.
    - Sukhavasi, S. B., Sukhavasi, S. B., Elleithy, K., El-Sayed, A., & Elleithy, A. (2022). A hybrid model for driver emotion detection using feature fusion approach. International journal of environmental research and public health, 19(5), 3085.
    - Singh, J., Kanojia, R., Singh, R., Bansal, R., & Bansal, S. (2023). Driver Drowsiness Detection System: An Approach By Machine Learning Application. arXiv preprint arXiv:2303.06310.
    - Kartali, A., Roglić, M., Barjaktarović, M., Đurić-Jovičić, M., & Janković, M. M. (2018, November). Real-time algorithms for facial emotion recognition: A comparison of different approaches. In 2018 14th Symposium on Neural Networks and Applications (NEUREL) (pp. 1-4). IEEE.
    - Dua, M., Shakshi, Singla, R., Raj, S., & Jangra, A. (2021). Deep CNN models-based ensemble approach to driver drowsiness detection. Neural Computing and Applications, 33, 3155-3168.
    - Chowdary, M. K., Nguyen, T. N., & Hemanth, D. J. (2021). Deep learning-based facial emotion recognition for human–computer interaction applications. Neural Computing and Applications, 1-18.
    - Hassouneh, A., Mutawa, A. M., & Murugappan, M. (2020). Development of a real-time emotion recognition system using facial expressions and EEG based on machine learning and deep neural network methods. Informatics in Medicine Unlocked, 20, 100372.
    - Jabbar, R., Shinoy, M., Kharbeche, M., Al-Khalifa, K., Krichen, M., & Barkaoui, K. (2020, February). Driver drowsiness detection model using convolutional neural networks techniques for android application. In 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT) (pp. 237-242). IEEE.
    - Liliana, D. Y. (2019, April). Emotion recognition from facial expression using deep convolutional neural network. In Journal of physics: conference series (Vol. 1193, p. 012004). IOP Publishing.

- Jain, D. K., Shamsolmoali, P., & Sehdev, P. (2019). Extended deep neural network for facial emotion recognition. Pattern Recognition Letters, 120, 69-74.
- Deng, W., & Wu, R. (2019). Real-time driver-drowsiness detection system using facial features. Ieee Access, 7, 118727-118738.
- Ali, M. F., Khatun, M., & Turzo, N. A. (2020). Facial emotion detection using neural network. the international journal of scientific and engineering research.
- Zadeh, M. M. T., Imani, M., & Majidi, B. (2019, February). Fast facial emotion recognition using convolutional neural networks and Gabor filters. In 2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI) (pp. 577-581). IEEE.
- Mehendale, N. (2020). Facial emotion recognition using convolutional neural networks (FERC). SN Applied Sciences, 2(3), 446.
- Khaireddin, Y., & Chen, Z. (2021). Facial emotion recognition: State of the art performance on FER2013. arXiv preprint arXiv:2105.03588.
- Niu, B., Gao, Z., & Guo, B. (2021). Facial expression recognition with LBP and ORB features. Computational Intelligence and Neuroscience, 2021, 1-10.
- Saini, G. K., Chouhan, H., Kori, S., Gupta, A., Shabaz, M., Jagota, V., & Singh, B. K. (2021). Recognition of human sentiment from image using machine learning. Annals of the Romanian Society for Cell Biology, 1802-1808.
- Mellouk, W., & Handouzi, W. (2020). Facial emotion recognition using deep learning: review and insights. Procedia Computer Science, 175, 689-694.
- Peng, S., Jiang, H., Wang, H., Alwageed, H., & Yao, Y. D. (2017, April). Modulation classification using convolutional neural network based deep learning model. In *2017 26th Wireless and Optical Communication Conference (WOCC)* (pp. 1-5). IEEE.
- Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., ... & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of big Data*, *8*, 1-74.
- Lagunas, M., & Garces, E. (2018). Transfer learning for illustration classification. *arXiv preprint arXiv:1806.02682*.
- Srinivasan, K., Garg, L., Datta, D., Alaboudi, A. A., Jhanjhi, N. Z., Agarwal, R., & Thomas, A. G. (2021). Performance comparison of deep cnn models for detecting driver's distraction. *CMC-Computers, Materials & Continua*, *68*(3), 4109-4124.
- Ali, L., Alnajjar, F., Jassmi, H. A., Gocho, M., Khan, W., & Serhani, M. A. (2021). Performance evaluation of deep CNN-based crack detection and localization techniques for concrete structures. *Sensors*, *21*(5), 1688.
- Grandini, M., Bagli, E., & Visani, G. (2020). Metrics for multi-class classification: an overview. *arXiv preprint arXiv:2008.05756*.
- Hicks, S. A., Strümke, I., Thambawita, V., Hammou, M., Riegler, M. A., Halvorsen, P., & Parasa, S. (2022). On evaluation metrics for medical applications of artificial intelligence. *Scientific reports*, *12*(1), 5979.

ACCOUNT CLEARANCE

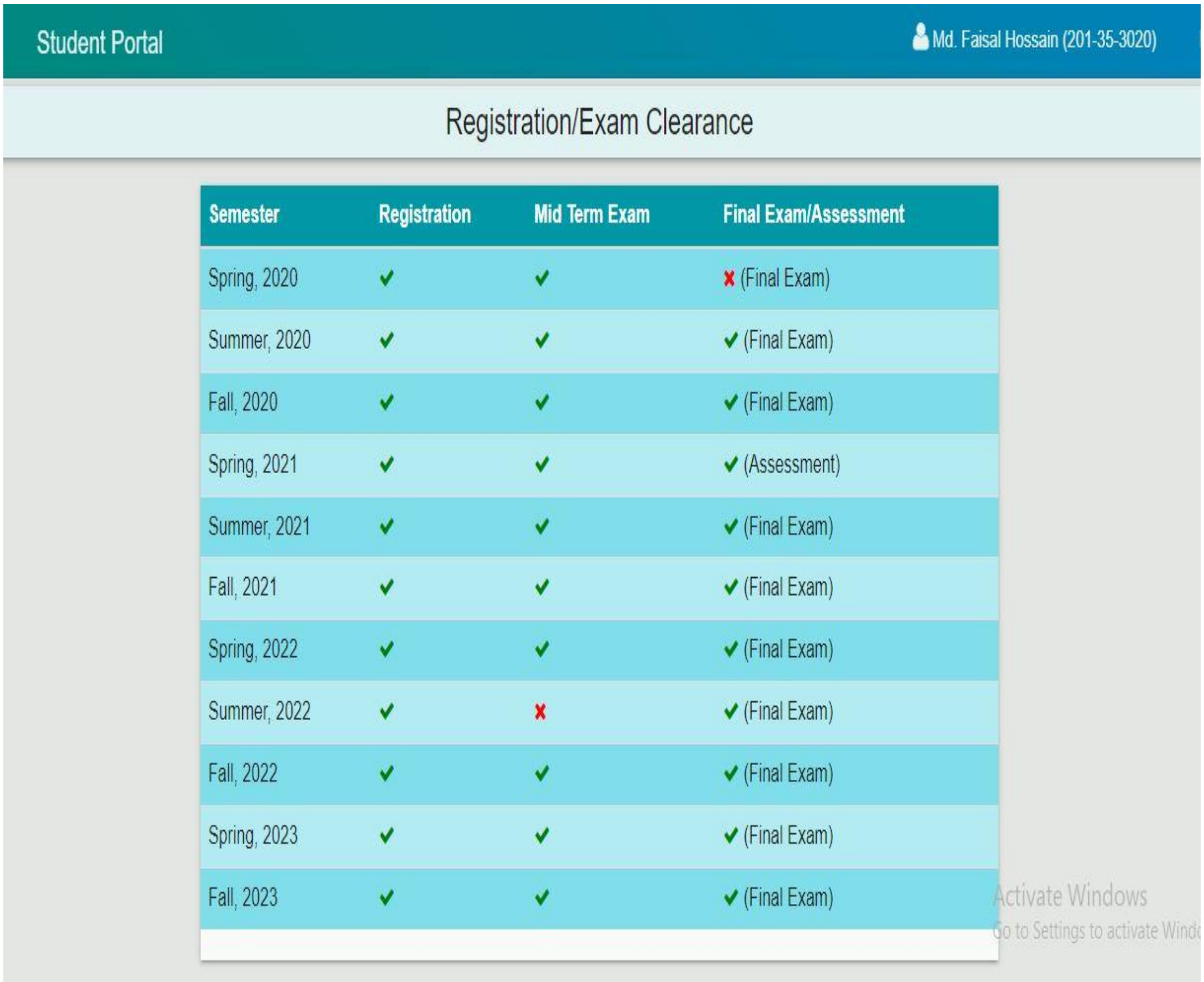


Student Dashboard

৳769,700.00	Total Payable
৳769,700.00	Total Paid
৳0.00	Total Due
৳0.00	Total Others

Payment Scheme

Daffodil International University


Student Portal
Md. Faisal Hossain (201-35-3020)
Activate Windows
Go to Settings to activate Windows

Registration/Exam Clearance

Semester	Registration	Mid Term Exam	Final Exam/Assessment
Spring, 2020	✓	✓	✗ (Final Exam)
Summer, 2020	✓	✓	✓ (Final Exam)
Fall, 2020	✓	✓	✓ (Final Exam)
Spring, 2021	✓	✓	✓ (Assessment)
Summer, 2021	✓	✓	✓ (Final Exam)
Fall, 2021	✓	✓	✓ (Final Exam)
Spring, 2022	✓	✓	✓ (Final Exam)
Summer, 2022	✓	✗	✓ (Final Exam)
Fall, 2022	✓	✓	✓ (Final Exam)
Spring, 2023	✓	✓	✓ (Final Exam)
Fall, 2023	✓	✓	✓ (Final Exam)