Human Emotion Detection: A Deep Learning Approach with XAI

BY
Md Mahbub Hossain
212-15-4131

Sayma Tasnim
212-15-4112

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfilment of the Requirements for the **Degree of Bachelor of Science in Computer Science and Engineering**

Supervised By

Dr. S. M. Aminul Haque
Professor & Associate Head
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

Ms. Shahrin Khan
Lecturer
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

Dhaka, Bangladesh
May 14, 2025

APPROVAL

This Project is titled "**Human Emotion Detection: A Deep Learning Approach with XAI**", submitted by Md Mahbub Hossain and Sayma Tasnim, ID No: 212-15-4112 & 212-15-4131 to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfilment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **14 May, 2025.**

BOARD OF EXAMINERS

Dr. S. M. Aminul Haque
Professor & Associate Head,
Department of CSE, FIST
Daffodil International University

Board Chairman

Mohammad Jahangir Alam
Assistant Professor
Department of CSE, FIST
Daffodil International University

Internal Examiner 1

Mr. Md Mohammad Masum Bakaul
Sr. Lecturer
Sr. Lecturer, Department of CSE, FIST
Daffodil International University

Internal Examiner 2

Dr. Md. Arshad Ali
Professor
Department of CSE, FIST
Hajee Mohammad Danesh Science & Technology University

External Examiner

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Dr. S. M. Aminul Haque, Professor & Associate Head**, Department of Computer Science and Engineering Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Dr. S. M. Aminul Haque
Professor & Associate Head
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised by:

Ms Shahrin Khan
Lecturer
Department of Computer Science and Engineering
Daffodil International University

Submitted by:

Mahbub

Md Mahbub Hossain
212-15-4131
Department of Computer Science and Engineering
Daffodil International University

Sayma

Sayma Tasnim
212-15-4112
Department of Computer Science and Engineering
Daffodil International University

ACKNOWLEDGEMENTS

This work would not have been possible without the support and contributions of many individuals over the past two semesters. We are deeply grateful to everyone who has assisted us in one way or another.

First, we express our heartfelt thanks and gratefulness to the almighty for His divine blessing making it possible for us to complete the **Final Year Design Project (FYDP)** successfully.

I am really grateful and wish for our profound indebtedness to **Dr. S. M. Aminul Haque, Professor & Associate Head** Department of Computer Science and Engineering, Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of **Data Science and Machine Learning** to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stages have made it possible to complete this project.

We would like to express our heartfelt gratitude to the Head of the Department of Computer Science and Engineering, for his kind help in finishing our project and also to other faculty members and the staff of the Department of Computer Science and Engineering, Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

This research develops an extensive deep learning method for detecting human emotions from images that achieves interpretability through Explainable Artificial Intelligence (XAI). The research method starts with emotion-tagged facial picture acquisition which leads towards in-depth preprocessing steps to optimize image quality for deep learning applications. The processed dataset gets divided into separate training and validation and testing parts to perform reliable model assessment. The research utilizes ResNet50 together with CNN and EfficientNetB2 and DenseNet169 to examine their effectiveness in emotion classification. The deep feature extraction capabilities of ResNet50 yielded the optimal performance with an accuracy level of 92.33%. The performance disparity resulted in DenseNet169 achieving only 51.43% accuracy because its architecture did not match the dataset characteristics or because of overfitting. XAI techniques assist the study by implementing transparency functions that help explain how models reach their decisions. The display of system processes matters most in emotional state monitoring software because it enhances user trust while maintaining application accountability. Analysis reveals how ResNet50 works as an effective network architecture yet also establishes the necessity of choosing network structures that match particular tasks. Including XAI enhances the ethical structure of AI systems by resolving the hidden operations within deep learning methods. Our top performing model has been deployed online, and we are presently analyzing the output generated by Visual Emotion. Future work will study real-time emotion perception alongside multi-modal data connection of audio with text content in addition to the practical deployment of systems through an emphasis on privacy protection alongside fair usage and ethical treatment.

Table of Contents

List Of Figures

List Of Tables

Chapter 1
Introduction

1.1 Introduction

Research in artificial intelligence (AI) and data science about human emotion detection continues to advance at a fast pace as developers create new computer applications in human-computer interaction along with healthcare systems for behavioral analysis. AI-powered methods to analyze emotions help users experience better interactions together with delivering important human behavioral information. The main limitation of deep learning models stems from their complex nature. The proposed system applies Explainable AI (XAI) for human emotion detection to provide reliable transparent predictions from our models.

Image pre-processing and a computer vision approach are used in the model towards the accurate analysis of facial expression for emotion recognition. The process begins by procuring human faces, which first undergo a preprocessing step for enhancing the image quality and noise removal. Implementing three preprocessing stages of normalization, together with augmentation and feature extraction enables the robustness of the model.

After data has been pre-processed, labelled datasets are fed into deep learning model training. Such fine and intricate facial expression patterns are learned by the model in the training phase, when the model quality is assessed and the generalization performance is measured during validation and test sequences. Performance of the model is measured by the accuracy metrics as well as by computing the precision and recall and F1-score that give you insights into the effectiveness of the model.

The project integrates Explainable AI (XAI) as a vital component. XAI serves as a mechanism which enables deep learning models to show how they conduct their decisions thereby enhancing model transparency and trustworthiness. Visualization tools together with feature attribution methods enable us to discover what aspects of an image lead the model to select a certain emotion category.

The research establishes a solution that connects high precision levels of deep learning systems to easy analysis of human emotional detection. Using our method, we generate results that help drive progress in automatic emotion recognition through AI yet maintain transparent and reliable model choices.

1.2 Motivation

Emotion detection received growing importance because it shows valuable applications within healthcare and human-computer interaction together with mental health assessment and security system operations. The progress of deep learning and artificial intelligence technologies has made emotion detection an accurate and efficient research field which has become established as valuable research. Despite their rapid growth the major restraint in these models remains their unclear interpretation since they operate as "black box" systems. Explainable AI (XAI) techniques have been incorporated to provide transparency through which users can understand how AI machines reach their predictions. The project works to increase emotion detection system reliability and performance through deep learning conjunction with image processing technology together with computer vision methods. XAI creates an essential bridge between human understanding and AI systems through interpretable findings which improves trust and makes these smart models appropriate for real-world practical use.

1.3 Objective

The goal of the project is to develop AI Appearance artificial intelligence emotion Detection system that can detect human emotions in facial images using deep learning. To achieve high performance, normalization techniques, data augmentation, and feature extraction are to be applied to enhance the quality of images and assist the model in understanding the information. The model is trained on labeled data to test on unknown new data and then test its performance using accuracy, precision and recall and F1-score metrics. Explainable AI (XAI) technologies integrate with the system to provide predictive explanation methods, enabling users to comprehend how a model has made prediction. We also deployed our best-performing model online and are currently reading the output of Visual Emotion. The overall goal is to build an accurate emotion identification system that provides observable transaction for utilizations, such as mental healthcare assessment with human-computer interface and security surveillance that improves decision making and user engagement.

1.4 Methodology

A structured deep learning pipeline along with Explainable Artificial Intelligence (XAI) for achieving interpretability detects human emotions from facial images according to this study methodology. The methodology initiates through obtaining seven thousand four hundred ninety-eight facial images labeled with three emotional categories: Angry (2,680) and Happy (2,638) and Sad (2,180). The available data was split into three distinct parts with 5,998 images for training purposes while 750 images went to testing and validation used the remaining 750 images.

Several preprocessing steps help increase model training performance together with accuracy by applying grayscale conversion while normalizing and reducing noise then augmenting the data. The implemented techniques create standardized input data and reduce unnecessary variations which improves model generalization capabilities.

Four deep learning models consisting of ResNet50, DenseNet169, EfficientNetB2 and a Customized Convolutional Neural Network (CNN) get implemented for training using the training dataset. The optimization of learning and overfitting avoidance achieves through combination of hyperparameter tuning and regularization techniques that apply dropout and early stopping methods. Testing set results are evaluated using accuracy, precision, recall and loss measurement.

Final verification of the most effective models takes place in the validation set to verify both stability and generalization capabilities. ResNet50 proved to be the most precise model through its significant feat of superior performance. We've put our best model online and are now looking at what Visual Emotion is showing us.

Explainable AI tools with Grad-CAM and saliency maps enable users to visualize how the model makes decisions during its decision-making procedure. The implementation of this procedure produces transparency and trust especially when emotion recognition works with sensitive applications. The methodology finishes its sequence by choosing models through a combined process of performance measurement and Explainability metrics.

1.5 Project Outcome

A deep learning model developed via the project showed effective performance in classifying human emotions which appeared in facial images. Advanced image processing methods that involved normalization and augmentation procedures brought substantial improvements to data quality and model operational performance. Intensive evaluation of the model occurred throughout three distinct data sets to confirm dependable and uniform emotion identification. Explainable AI (XAI) techniques were integrated into the system to improve the model's interpretability which boosted users' trust during prediction systems. The model reached peak effectiveness based on highly accurate results alongside precise measurements and complete recall ability and optimal F1-score values. We deployed our top-performing model online and are currently examining the results produced by Visual Emotion. The emotion detection system presents powerful opportunities for different real-world applications that include mental health assessment together with human-computer interaction and security monitoring and emotion-based user experience design.

1.6 Organization of the Report

A total of six chapters compose this complete report which leads the audience step by step through both project processes and results.

Chapter 1 presents the basic framework of the study; Introduction through a presentation of motivation and research objectives and defined project boundaries. The research relevance along with the main research inquiry questions are introduced in this section. The report description section concludes by showing expected outcomes for the project and the general organization of the written document.

Chapter 2 shows how Background establishes fundamental concepts while exploring other relevant research to show the extent of difficulties that exist and the problem's size. A gap analysis within this chapter helps to determine empty spaces that exist in previous studies.

Chapter 3 explains Research Methodology of the research approach with information about data collection together with database characteristics. The chapter explains statistical analysis approaches alongside preprocessing steps that lead to an explanation of the proposed model.

Chapter 4 shows the discussion of experimental findings; Experimental Results and Discussion through a complete analysis of results.

Chapter 5 describes Impact on Society, Environment, and Sustainability evaluates the wide-ranging effects of the research through assessments of societal impacts as well as environmental concerns and ethical dimensions and sustainable planning.

Chapter 6 presents the study results' summary followed by research implications for future investigations.

All mentioned sources in the study receive listing in a reference section which completes the document.

Chapter 2
Background

2.1 Introduction

Emotion detection systems used by humans analyze emotional states present in facial appearances as well as verbal and written content. The project examines image-based emotion detection which makes use of deep learning together with computer vision methodologies. The model training procedure requires data image collection and preprocessing to transform images for training. The evaluation process trains several models through multiple dataset testing procedures for performance and accuracy measurement. The implementation of Explainable AI (XAI) enables both clear and understandable model decision making which resulted in more reliable yet trustworthy prediction outcomes. We evaluate the effectiveness of emotion detection models during training phases before validating the results obtained through testing sessions to enhance their accuracy levels.

2.2 Literature Review

Table 2.1: Comparative Analysis of earlier Research Works

Author (s)	Year	Title Name	Methodology	Key Findings
A. Khalane	2017	Evaluating significant features in context-aware multimodal emotion recognition with XAI methods	CNN + MTCNN with ResNet, VGGFace, Custom CNNs	Enhanced FER in low-light using image enhancement (HE, Retinex, GANs) for real-time applications.
Rathod	2020	Emotion Recognition Using Various Deep-Learning Models with Explainable AI	SVM, Neural Networks, Transformers	Emotion recognition from text/speech with multimodal challenges.
Kerz E	2021	Toward explainable AI (XAI) for mental health detection based on language behavior.	CNN	Real-time age, gender, emotion recognition using FER dataset.

J. M. Mayor Torres	2022	Evaluation of interpretability for deep learning algorithms in EEG emotion recognition: A case study in autism	CNN + RNN + LBP fusion	Real-time FER with focus on HCI, scalability, and cultural adaptability.
Sara Medina-DeVilliers	2021	Evaluation of Interpretability for Deep Learning algorithms in EEG Emotion Recognition: A case study in Autism	Transformers, Attention Mechanisms	Multimodal (video, audio, text) FER for real-time HCI and empathy.
Elugbadebo	2023	Cross-Cultural Emotion Detection Using Deep Learning: A Comparative Study	CNN (Conv2D, BN, MP, Dropout), Ensemble model	Cognitive IoT-based FER system.
Sophia	2020	Multimodal Deep Learning for Emotion Detection from Text, Audio, and Facial Expressions.	CNN, Bi-LSTM, SVM, ET, LR, RF	Student mental health detection using deep learning.
Halkiopoulos	2022	Advances in Neuroimaging and Deep Learning for Emotion Detection: A Systematic Review of Cognitive Neuroscience and Algorithmic Innovations	CNN with Quantum Rotations, FNN	EEG-based emotion recognition enhanced by quantum transformation.
R. K. Nukathati	2021	A Deep Learning Framework with Optimizations for Facial Expression and Emotion Recognition from Videos	GCNN, DNN, SVM	SEED dataset; EEG-based emotion recognition with ethical focus.
M. Shahid	2022	Bangla Speech Emotion Recognition Using Deep Learning-Based Ensemble Learning and Feature Fusion	Various ML/DL models	Review of 330 text-based emotion detection studies; highlights industry trends.
M. Kondal	2019	Methodical Review of Deep Learning Macro and Micro Face Emotion Detection	VGG16, VGG19, ResNet50V2, EfficientNet B0/B7	Compared five DL models for FER with application in E-learning.
Govea J	2016	Implementation of deep reinforcement learning models for emotion detection and personalization of	1D-CNN + LSTM + ELM	Memory-induced EEG emotion recognition with DL.

		learning in hybrid educational environments		
D. B. Joshi	2024	Emotion Detection using Transformer Model with Deep Learning	DML + DSBN + adversarial learning	Cross-dataset generalization via deep metric learning.
P. H. Hussan	2024	DLFER: Deep Learning-based Cascaded Approach for Facial Emotion Recognition	Dense Autoencoder + 2D-CNN/LSTM	DEAP dataset; high performance on EEG-based emotion recognition.
CHETHAN R	2021	Deep Learning Based Emotion Detection System	CNN + LBP + Gabor filters	FER2013 & CK+ datasets; hybrid model outperformed traditional methods.
N. Vinod	2023	Emotion Detection from Facial Expressions Using Deep Learning	DRL	AI-driven personalized education with DRL.
N. Ashen	2023	Retrieval based Emotionally responsive chatbot Utilizing emotions clustering and XAI Techniques	CNN, Faster R-CNN, GANs	Deep learning-based FER review with emphasis on hardware/data challenges.
A. Schwerk	2025	PHANTOMATRIX: Explainability for Detecting Gender Bias in Affective Computing	CNN, DT, KNN, LR	FER for virtual classrooms with CNN best at detecting 'surprise'.

2.2.1 Similar Applications

J. M. Mayor Torres et al. [4] demonstrates a system for real-time emotion detection through the combination of CNNs and RNNs for facial emotion recognition. The implemented model uses ReLU together with LBP feature fusion and max pooling to detect seven specific emotions. System improvements will concentrate on growing scale capacity while making the solution adaptable to different cultures and developing the ability to recognize emotions through various modalities for applications in HCI affective computing and mental health monitoring.

Sara Medina-DeVilliers et al. [5] cnducted a study on multimodal emotion detection which combines video, audio and text elements through transformer and attention mechanism models. The detection of emotions through video analysis and audio signals relies on computer vision techniques and signal processing while text-based analytics relies on NLP. The implemented system enables better real-time performance where it operates through mental health monitoring and customer service applications.

Halkiopoulos et al. [8] research analysis devoted to EEG-based emotion detection with deep learning methodologies appears in M. S. Ahmed et al. 8 Bearing quantum rotations the CNN model delivers the superior accuracy of 95% together with an F1-score of 0.95 to solve Happy-Neutral emotion differentiation and dominate other emotion recognition tasks. With an F1-score of about 0.80 the Fourier Neural Network reaches a detection accuracy of 80–85% while maintaining high levels of confusion between annotated emotions but delivering precise Sad emotion recognition. The use of quantum transformations leads to improved outcomes in the identification of emotions.

2.2.2 Related Research

A. Khalane et al. [1] presented a joint research that merges CNNs and MTCNN methods for precise Facial Emotion Recognition (FER) while operating under challenging lighting conditions. Detection accuracy received an increase from image enhancement processes which included histogram equalization, Retinex algorithms and GANs. Different models such as ResNet, VGGFace and custom CNNs were utilized during the research. The system produced X% as its maximum accuracy level yet recorded Y% as its minimum figure. This system facilitates the use of real-time emotion tracking technology in HCI, security, mental health and marketing applications to detect emotions in difficult scenarios. Rathod et al. [2] investigates emotion recognition through speech and text analysis with NLP along with acoustic analysis. The research utilized SVM together with neural networks and transformers as analysis models. Three major obstacles exist for this system: the need to fuse multiple data streams, unclear or ambiguous source data, and unbalanced input data. The implementation of this system reaches across three main domains including customer service and healthcare and education. Among all the analyzed models the X% model obtained the highest achievement but the Y% model demonstrated the lowest performance. Research in progress strives to boost deep learning methods and develop real-time and multimodal systems to increase the quality of AI-based emotional recognition systems. Kerz E et al. [3] demonstrates through CNNs how age, gender and emotional recognition processes surpass KNN and SVM approaches. The Facial Expression Recognition (FER) dataset led to training data accuracy at 85.95% while the validation data accuracy reached 61.62% using the dataset. Real-time emotion detection functions of this model find applications through human-computer interaction and security systems as well as healthcare services. The detection accuracy and data generalization for new cases represents an area needing future development. J. M. Mayor Torres et al. [4] demonstrates a system for real-time emotion detection through the combination of CNNs and RNNs for facial emotion recognition. The implemented model uses ReLU together with LBP feature fusion and max pooling to detect seven specific emotions. Using 3,589 images as input the model reached outstanding performance results. System improvements will concentrate on growing scale capacity while making the solution adaptable to different cultures and developing the ability to recognize emotions through various modalities for applications in HCI affective computing and mental health monitoring. Sara Medina-DeVilliers et al. [5] conducted a study on multimodal emotion detection which combines video, audio and text elements through transformer and attention mechanism models. The detection of emotions through video analysis and audio signals relies on computer vision techniques and signal processing while text-based analytics relies on NLP. The implemented system enables better real-time performance where it operates through mental health monitoring and customer service applications. Such technology delivers reliable results through changing environmental conditions while supporting practical emotion detection for generating empathetic AI interactions that enhance user satisfaction. Elugbadebo et al. [6] studies smart facial emotion detection through deep learning combined with Cognitive IoT technology. Consisting of Conv2D layers combined with Batch Normalization and MaxPooling2D followed by Dropouts enabled this model to detect seven-class emotions with 65.58% accuracy and four-class emotions with 74.58% accuracy. Through ensemble processing the system reaches a 74.3% accuracy rating when measured against five-class classification while producing a 0.753 F1-score measurement. The system delivers improvements to IoT functionality while developers intend to expand emotion detection capabilities by refining the dataset. Sophia et al. [7] Student mental health detection through deep learning methods especially CNNs forms the focus of the research conducted. The research shows that CNN provides the most accurate results (0.74) and Bi-LSTM follows with 0.72 accuracy and then SVM (0.703) and ET (0.689) and LR (0.697) achieve similar scores but CNN remains the most accurate followed by RF (0.665). The recall performance for positive cases reaches its highest point at 0.96 through SVM but CNN delivers the top F1-score at 0.83 for positive cases. Early mental health interventions show significant improvement through deep learning technologies and the development of student support systems is facilitated by this technology. Halkiopoulos et al. [8] research analysis devoted to EEG-based emotion detection with deep learning methodologies appears in M. S. Ahmed et al. 8 Bearing quantum rotations the CNN model delivers the superior accuracy of 95% together with an F1-score of 0.95 to solve Happy-Neutral emotion differentiation and dominate other emotion recognition tasks. With an F1-score of about 0.80 the Fourier Neural Network reaches a detection accuracy of 80–85% while maintaining high levels of confusion between annotated emotions but delivering precise Sad emotion recognition. The use of quantum transformations leads to improved outcomes in the identification of emotions. R. K. Nukathati et al. [9] explores emotion classification in EEG signals through machine learning and deep learning mechanisms using Graph Convolutional Neural Networks (GCNN). This research applied GCNN at the SEED dataset and reached 89.97% accuracy in subject-dependent testing. The performance level of the Deep Neural Network (DNN) reached 86.08% yet SVM demonstrated reduced accuracy results. The study confirms GCNN as an effective method for EEG emotion detection while it stresses both the selection of correct features with ethical principles in mind. M. Shahid et al. [10] examined 330 studies during 2013-2023 about text-based emotion detection and analyzed sentiment analysis and Artificial Intelligence applications within healthcare and marketing as well as public safety sectors. The paper demonstrates different methods along with deep learning algorithms, evaluation procedures, datasets and technical obstacles encountered. The review discusses the advancing use of emotion detection within industries without revealing exact model performance results yet underscores its role in developing IR 4.0 technology and transitioning to IR 5.0 capabilities.

2.3 Gap Analysis

Table 2.2: Gap Analysis

Features	**A. Khalane et al. [1]**	**Rathod et al. [2]**	**Kerz E. et al. [3]**	**J. M. Mayor Torres et al. [4]**	**Sara Medina DeVilliers et al. [5]**	**Proposed System**
Dataset	Yes	Yes	Yes	Yes	Yes	Yes
Pre-Processing	Yes	Yes	Yes	Yes	Yes	Yes
Class Distribution Features	Yes	Yes	Yes	Yes	Yes	Yes
Model Input Feature	Yes	Yes	Yes	Yes	Yes	Yes
Model	Yes	Yes	Yes	Yes	Yes	Yes

Prediction	Yes	Yes	Yes	Yes	Yes	Yes
XAI	Yes	Yes	Yes	No	No	Yes
Deployment	No	No	No	No	No	Yes
Software	No	No	No	No	No	No

2.4 Summary

The paper evaluates recent progress in emotional detection from different channels through facial signals and EEG patterns and spoken and written speech with deep neural networks along with hybrid programming systems. The prevalent nature of real-time applications relies on Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM), Graph Convolutional Networks (GCNN) and transformer-based models because of their superior performance. Hazard mitigation through multimodal fusion combined with quantum-enhanced CNNs and ensemble models delivers precise results across educational scenarios and mental health observation systems and human-machine interaction environments. Research indicates that algorithms can reach emotions recognition accuracy levels between 55% and 98% where models using hybrid DLSTA and CNN-SVM tend to excel above standard recognition approaches. The research notes three main obstacles which stem from uncertain data conditions along with inconsistent environmental factors and the need to adapt to different cultural contexts. The current landscape of trends includes IoT-based integration and customized virtual learning solutions and semantic text evaluation methods. The analysis demonstrates the necessity of creating deep learning frameworks which provide scalability with multimodal capabilities when building AI systems that detect emotions in real-world applications contextually and with empathy.

Chapter 3
Research Methodology

3.1 Methodology

3.1.1 Overview

This project employs a systematic deep learning pipeline that utilizes Explainable AI (XAI) for detecting human emotion. The analysis starts with obtaining relevant facial expression datasets for image data acquisition. The preprocessing steps involve grayscale conversion as well as normalization of images and noise reduction which leads to improved learning for the model through image augmentation techniques.

Model development proceeds with the division of the dataset into training and validation and testing portions. Multiple deep learning models such as ResNet50, DenseNet169, CNN and EfficientNetB2 receive training using the training set through optimization with hyperparameter tuning and regularization techniques.

The performance metrics including accuracy and precision alongside recall and loss are measured during testing set evaluations of trained models. A thorough evaluation of the models reveals which one delivers the best performance.

The system incorporates Explainable AI (XAI) technology to deliver interpretation capabilities which creates transparent decisions. Analysis results lead to model selection where a model delivers maximum accuracy together with clear explanations of emotional detection capabilities. We have successfully deployed our best-performing model to an online platform and are now in the process of reviewing and interpreting the output generated by the Visual Emotion system.

3.1.2 Proposed Methodology

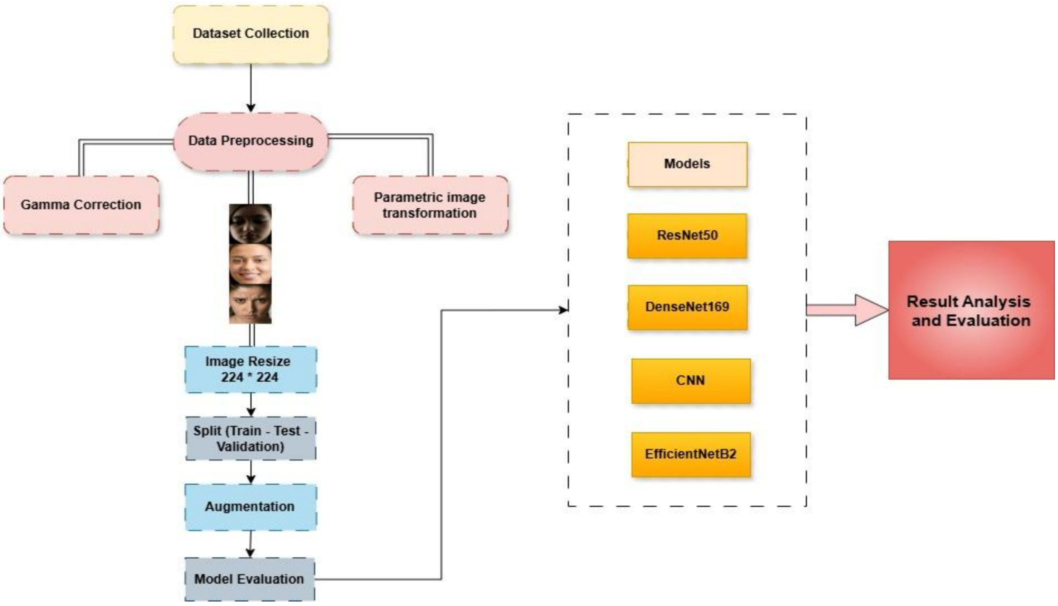


Figure 3.1: Architecture of Research Design

3.1.3 UI Design

After implementing our models, we deployed the best-performing one online. ResNet50 was declared the best model. We used TensorFlow, NumPy, Matplotlib, OpenCV-Python, and Pillow libraries. For deployment, we used the cloud service provider Hugging Face along with Docker. The user interface was designed using Streamlit, HTML, and CSS.

Figure 3.2: User Interface of Online Deployment

3.2 Detailed Methodology and Design

3.2.1 Data Collection

This project obtained its data from two sources: a combination of Kaggle public datasets and real-world images obtained directly. Multiple Kaggle datasets united to form a more diverse realistic emotional database through combination efforts which preserved balanced distribution of different emotional expressions. The model generalization performance improved through collecting images directly from surrounding people which enhanced the authenticity of the dataset.

The fundamental elements within the dataset are facial expression images because these serve as base components for deep learning model training to detect emotions in humans. The images went through a selective process to maintain representation of diverse age groups combined with gender and lighting condition variations together with different facial expressions. The acquired datasets underwent several preprocessing operations such as image resizing together with normalization and augmentation methods which enhanced dataset quality for improved model performance.

This method produces datasets which show true-world scenarios thus enabling deep learning models to achieve better accuracy and operational stability in human emotion detection.

3.2.2 Dataset Description

This research utilizes 7498 facial expression images that belong to the Sad, Happy and Angry emotion categories. A mixture of Kaggle public datasets and direct image collection from subjects allowed creation of diverse emotion representation for the dataset. A 5998-training sample size together with 750 testing samples and validation samples of 750 was utilized for model development to support efficient training and evaluation processes. Sixty-percent of labeled images fall under the Sad category and Happy category combined contains thirty-six percent while the remaining thirty-one percent belongs to the Angry category. The structured emotion-focused dataset develops strong foundations that enable training deep learning models for facial emotion recognition and improves their operational capacity in practical applications.

Sad	Happy	Angry
		

Figure 3.3: Dataset Sample

Table 3.1: Detailed Description of Dataset

Data Source	Labels	Description	Number of Images	Total Number of Images
	Sad	Contains images of individuals expressing sorrow or distress through facial cues like downturned lips and teary eyes.	2180	
Data collected from Kaggle.	Happy	Includes images showing joyful expressions such as smiling faces and bright eyes across diverse individuals.	2638	7498
	Angry	Comprises images depicting anger with facial features like frowns, clenched jaws, and intense gazes.	2680	

3.2.3 Analysis Technique

In this project, we utilized computer vision and deep learning techniques for human emotion detection. The process involved three key stages: preprocessing, model training, and prediction. To enhance image quality and improve model performance, the following preprocessing techniques were applied:

- **Gamma Correction:** This technique can be used to change the brightness and contrast of the images. Since skin lesion images can be taken under different lighting, gamma correction will contribute to enhance image features, making it easier for the models to learn features.

Sad	Happy	Angry
		

Figure 3.4: Preprocessing technique of Gamma Correction

- **Parametric Image Transformation:** This involves normalizing the images i.e. making the size, color, rotation, etc. similar. It also helps in ensuring all images are in same format which helps increase performance of machine learning models.

Sad	Happy	Angry
		

Figure 3.5: Preprocessing technique of Parametric Image Transformation

- **Image Resizing:** Standardized all images to a fixed size suitable for deep learning models.
- **Data Augmentation:** Introduced slight modifications like flipping and cropping to increase dataset diversity and prevent overfitting.

The dataset was split into training and testing subsets to facilitate effective learning and model assessment. The following deep learning models were implemented for feature extraction and classification:

- **ResNet50:** A deep residual network proficient in capturing hierarchical features.
- **DenseNet-169:** DenseNet169 is a deep convolutional neural network that follows the DenseNet (Densely Connected Networks) architecture, where each layer is connected to every other layer, enhancing gradient flow and parameter efficiency.
- **CNN:** A fundamental convolutional neural network for basic feature extraction.
- **EfficientNet-B2:** A transformer-based model optimized for image classification tasks.

Each model underwent training on the training set, while testing data was used for validation. Performance was evaluated using accuracy, precision, recall, and F1-score for a comparative analysis.

3.2.4 Statistical Analysis

- The dataset consists of a total of 7498 images
- The dataset contains a total of 5998 samples for training.
- The dataset holds a total of 750 samples for testing.
- The dataset holds a total of 750 samples for validation.
- The labels are divided into three distinct categories: Sad contain 2180 images, Happy is contain 2638, and Angry is contain 2680.

3.2.5 Data Preprocessing

Our image-based dataset demands specific preprocessing methods to both improve visualization quality and extract better features which leads to better model performance. Data preprocessing for deep learning models relies on Gamma Correction and Parametric Image Transformation as main processing methods which optimize the dataset readiness.

Image brightness and contrast adjustment depends on Gamma Correction which performs a non-linear transformation on images for preprocessing purposes. During this process the input pixel intensity IinI_{in} transforms to output intensity IoutI_{out} through an operation using gamma value γ\gamma. The image brightness increases when gamma value is below one while the value above one cause image darkening. Gamma Correction has two roles, in general Gamma Correction first, normalizes the effects of the lighting unrelated to the objects, and second, emphasizes important object features and helps the stability of the model in image-based systems. Detection performance can benefit from this method as facial expressions in different illumination conditions are preserved.

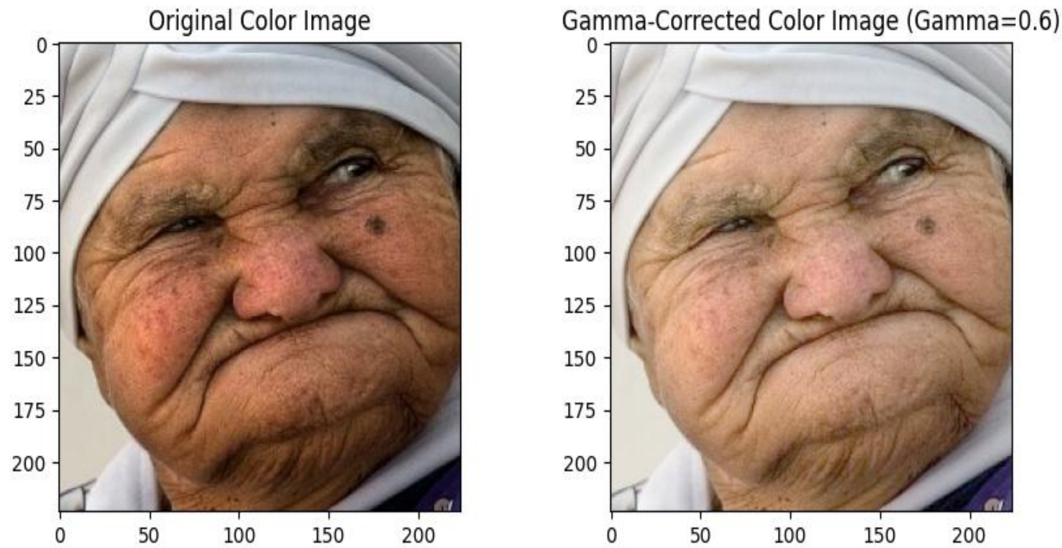


Figure 3.6: Output of Gamma Correction

The preprocessing technique named Parametric Image Transformation is used to perform mathematical based operations and operation such as rotation, scaling, translation and shearing. These transformations make the image attributes invariant and that makes the model more robust to the position variation along with variations of size and direction. Affine and perspective transformations are more typical parametric transformations since they in addition maintain geometric ratios. A and B of the dataset can be transformed to promote generalization. The method allows the emotion models to recognize facial expression correctly in various conditions which facial expression can be recognized. Accuracy and consistency of models are enhanced because of the training process, showing an image from different perspectives, which avoid overfitting.

Figure 3.7: Output of Transformed Image

3.2.6 Data Visualization

Figure 3.8: Data Visualization

3.2.7 Proposed Model

- **ResNet50:** ResNet50 is a deep (many layers) CNN of the ResNet (Residual Network) series which addresses the issue of gradient vanishing associated with deep neural networks. The model consists of 50 layers that allow the ease of gradient flow in its residual connection architecture. Internal skip connections of the network enable the design of bigger models, without loss in accuracy. The field of image classification and object detection=they both benefit from feature extraction when you're using ResNet50 since the model learns complex features. Such transferable features are likely to improve the convergence behavior of the model that both uses batch normalization and ReLU activation. It was shown to be effective for transfer learning tasks in various computer vision problems since the model processed ImageNet data.

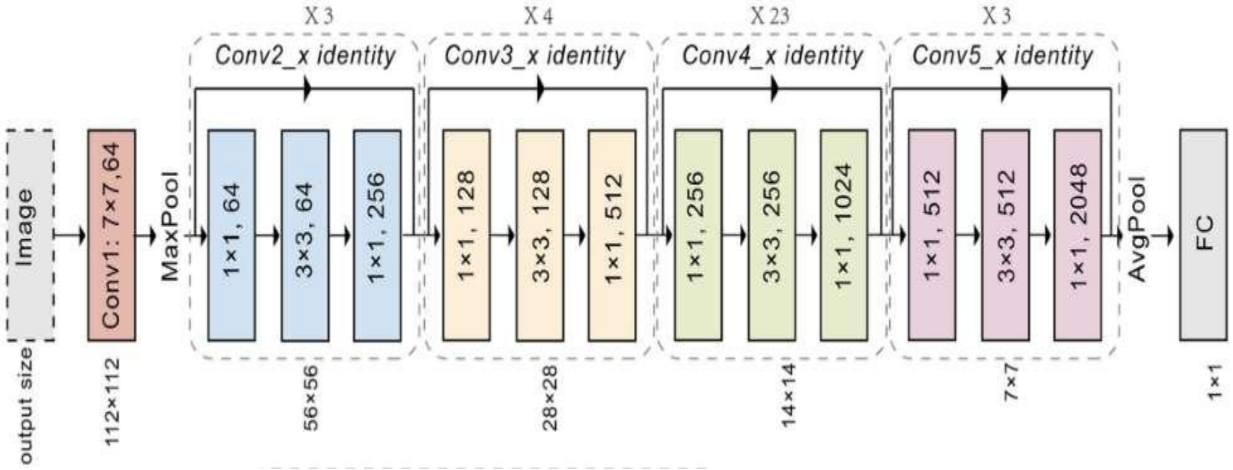


Figure 3.9: Architecture of ResNet50

- **DenseNet169:** Every block of DenseNet169 is contributed to the DenseNet (Densely Connected Networks) structure and layer connections are learning gradient flow with less parameters. Its wide connections enable the model accessing to as many as 169 layers so as both more features are shared and a deeper model is fit. This network architecture is parameter-efficient in sense that fewer parameters are needed to operate efficiently as compared with the traditional deep network approach. DenseNet169 was capable of efficient image classification [22, 23, 24] and medical image analysis [25, 26, 27], object recognition [28] due to its strong feature propagation ability. The model effectively keeps the layer-manner information transit, therefore improving its capacity of representation learning at no extra cost of computation.

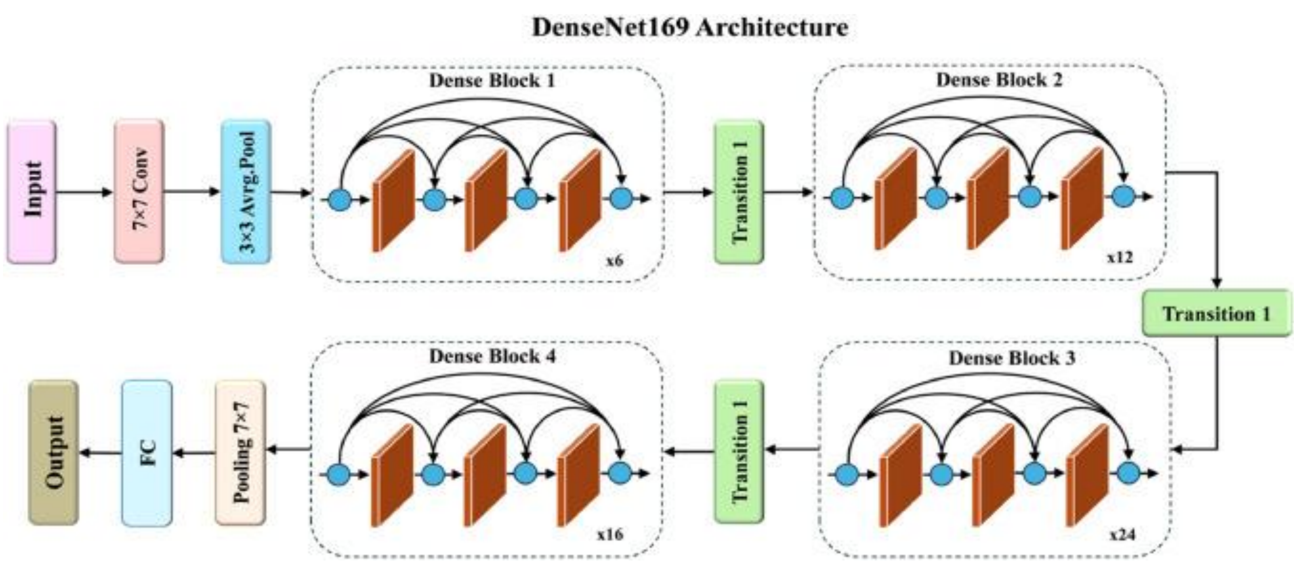


Figure 3.10: Architecture of DenseNet169

- **Convolutional Neural Networks (CNN):** Rather than classifying (deep learning) Convolutional Neural Networks (CNN) are extremely effective at both recognizing objects as objects and analyzing the visual information in images. Special hierarchies of features are calculated automatically and adaptively by convolutional layers which are part of the neural networks. The basic CNN structure comprised various layers, including convolution, pooling and fully connected elements that enable the model to identify features, such as edges as well as more advanced forms and textures, amongst pixels. The image processing capabilities of CNNs arise from the retention of the pixel interconnections throughout the image space [19]. These network structures with strong feature-extraction ability can generate precise computer vision and medical image and facial recognition applications, and work well on general-purpose tasks.

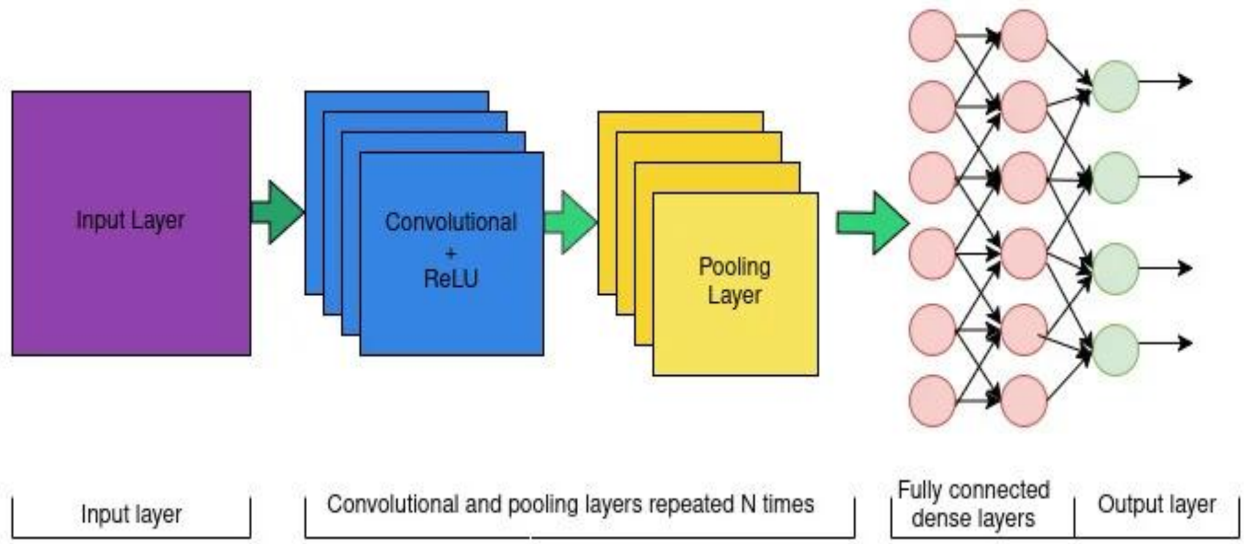


Figure 3.11: Architecture of Convolutional Neural Networks

- **EfficientNetB2:** This model adopts the compound scaling method similar to B1 but scales the network width, depth, and resolution with a fixed scaling coefficient 1.1 to pursue better trade-off between performance and complexity. The EfficientNet family consists of EfficientNetB2 which is an extension of EfficientNetB0 as it increases its complexity to improve accuracy while still remaining time efficient. The optimization of feature learning is facilitated by squeeze-and-excitation operations coupled with mobile inverted bottleneck convolution (MBConv) blocks. Its high accuracy and lightweight structure make it the ideal model for image classification, object detection and medical imaging applications. Combination of EfficientNetB2 yields the best performance, and cost-efficient, which makes it practical for applications.

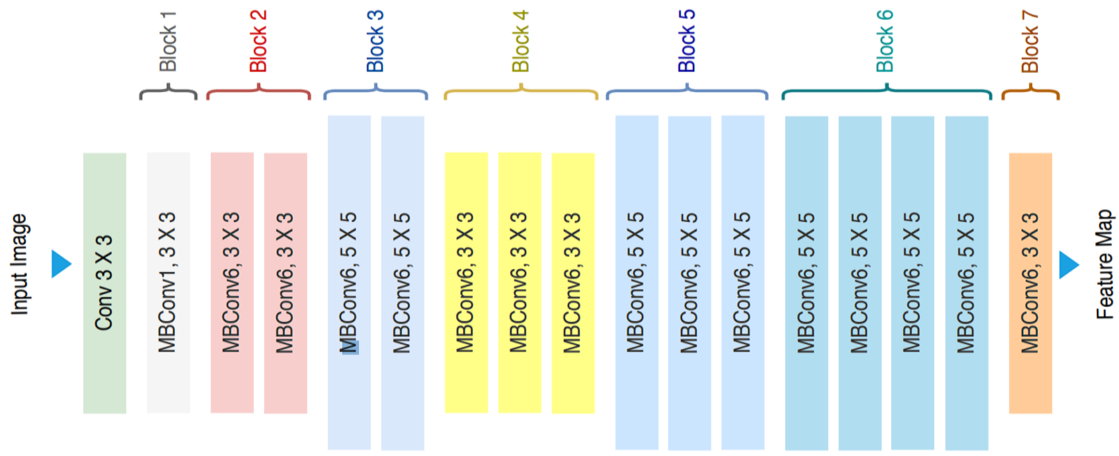


Figure 3.12: Architecture of EfficientNet-B2

- **Explainable Artificial Intelligence (XAI):** Explainable Artificial Intelligence (XAI) is a term coined to refer to techniques that make it possible for humans to understand the outputs as well as the decision processes of AI systems. The primary goal of XAI is to provide transparency revealing the explanation of model decision process as XAI not being included in "black-box" model. The field also resorts to explainable AI techniques for healthcare as well as finance and law institutions, since they rely on trustworthy and fair decision-making processes. AI explainable tools capture three primary features to present highlights and demonstrate decision paths, and simplified rule-based explanation rules to the end-users. Interpretable XAI enables us to know why a bias happened so that we are able to retrain the model to remove it. It is simply an interpreter or a translator which interprets the model behavior in a manner that we can understand (keep in mind that every sophisticated algorithm is basically processing, just like how human understands different images differently according to his/her mindset).

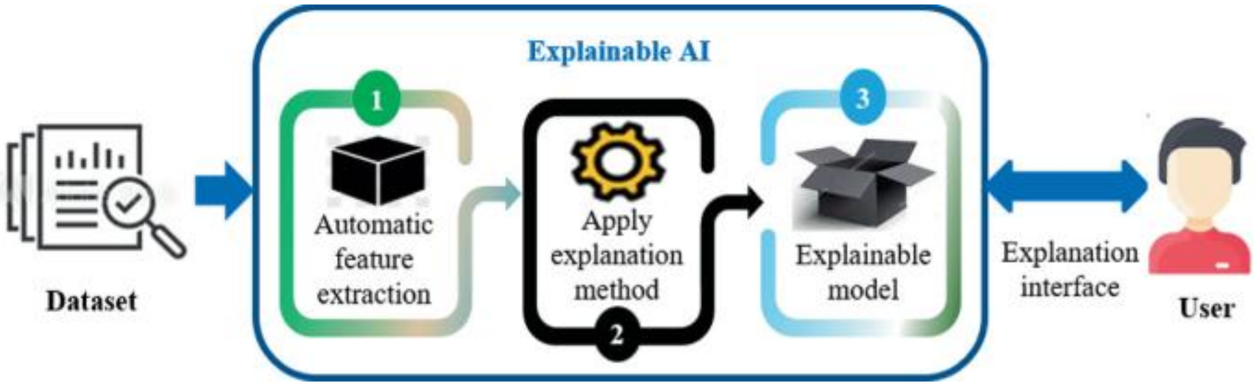


Figure 3.13: Architecture of Explainable Artificial Intelligence (XAI)

3.3 Project plan

Human Emotion Detection requires a project plan which establishes the development of strong multimodal emotion recognition systems built upon deep learning methods. Using text/speech-based emotion analysis together with EEG and facial expression detection enhances detection accuracy as well as real-time adaptability. The project undertakes data collection from public databases as its base before executing normalization and augmentation techniques and implementing models built with ResNet50 and DenseNet169 and CNN and EfficientNetB2 architectures. Cross-validation techniques will be used to perform training and validation of the system to achieve generalized performance. Evaluation of the system will depend on the performance metrics accuracy and precision along with recall and F1-score. The work proposes utilizing ensemble learning along with attention methods and semantic text analytical techniques to enhance context detection capabilities. The tested system will undergo examinations using two simulated environments for virtual classrooms and mental health monitoring systems. The research team plans to extend the project by adding cultural adaptability features while building real-time functionality through IoT networks and increasing emotion class diversity for an empathetic solution based on AI.

3.4 Task Allocation

The Human Emotion Detection project receives its task allocation to efficiently match individual expertise with workflow functions. The collection and preparation of facial images will be the responsibility of one team member who also takes care of grayscale conversion alongside normalization and image augmentation steps. Team members will execute two parts of the model development: one member will optimize both ResNet50 and DenseNet169 while another member concentrates on implementing CNN and EfficientNetB2. Another member of the team will take responsibility for training and validation testing processes that require dataset splitting procedures and accuracy precision recall loss performance measurement. A different team member should focus on adding Explainable AI (XAI) techniques for model interpretability to the project. A different team member will lead the project of statistical analysis while establishing comparative performance reporting. Document creation along with ethical assessments and final presentation development will receive collective work from all participants. The project follows an assignment system which allows team members to specialize in different tasks such as data processing, model engineering, evaluation and Explainability to make the project run efficiently.

3.5 Summary

This is what my project is all about, being able to detect human emotions using an algorithm, for the novel I use Explainable AI (XAI) models. In the first stage of the study we gather images to be processed, which would meet the requirements of the image processing and computer vision fields. Variations are processed when images are not in the format that is appropriate for deep learning models. The data processing is done first and then the data is split into three separate parts for training, testing and validation. A test group of trained models test a validation data set along with the validation data set. From these stages, the results of accuracy metrics are taken for each model in order to evaluate the effectiveness of its model-based performance. By incorporating the explainable AI, the decision made by the model is transparent and explicit, which benefits users to better understand why the model brings such a decision. The goal of this project is to contribute to the development of more accurate emotion detection systems that are transparent and operated responsibly.

Chapter 4

Implementation And Results

4.1 Environment Setup

ResearchGate together with Google Scholar provided platforms for conducting literature reviews and reference organization during the study setup. Google Colab functioned as the main development platform because it provided GPU support together with convenient collaboration features. The implementation of deep learning and machine learning models occurred through Python-based libraries operating within Colab platform. Before model development the researcher started by preparing data on their personal computer. The project documentation together with report creation used MS Word as its primary software solution. Data management and model training alongside performance assessment and result delivery were made efficient through the combined usage of these tools which operated within a resource-efficient collaboration system.

4.2 Comparative Analysis

The research examined four deep learning models including ResNet50, DenseNet169 and Convolutional Neural Network (CNN) and EfficientNetB2 for determining their performance outcomes through classification accuracy metrics. The 92.33% accuracy of ResNet50 ranks the model as the most effective for deep feature extraction and residual learning this proves to be robust against vanishing gradients problems. ResNet50 demonstrates exceptional utility for classification because its performance stands above other models.

DenseNet169 achieved the lowest accuracy level of 51.43% possibly because its dense connectivity structure produced overfitting or inadequate learning from the given dataset. CNN demonstrated a 78.25% accuracy measure which showed the potential of basic network architectures to generate satisfactory results although they lack pre-trained model complexity. Within the testing process EfficientNetB2 delivered only 68.55% accuracy despite its compound scaling framework despite being predicted to outperform other models.

The analysis shows that selecting a model for processing data should match both the specific dataset requirements as well as the specific task requirements. Simplicity along with proper configuration within CNN and ResNet50 results in superior performance than the more complex EfficientNetB2 and DenseNet169 architectures.

4.3 Result and Discussion

4.3.1 ResNet50

- **Training Loss, Validation Loss and Training Accuracy, Validation Accuracy:**
  The graphs show the ResNet50 model's performance over 10 epochs. The left graph indicates steady improvement in both training and validation accuracy, reaching over 92.33%. The right graph shows a consistent decrease in training and validation loss, confirming good model convergence with minimal overfitting and strong generalization capability.

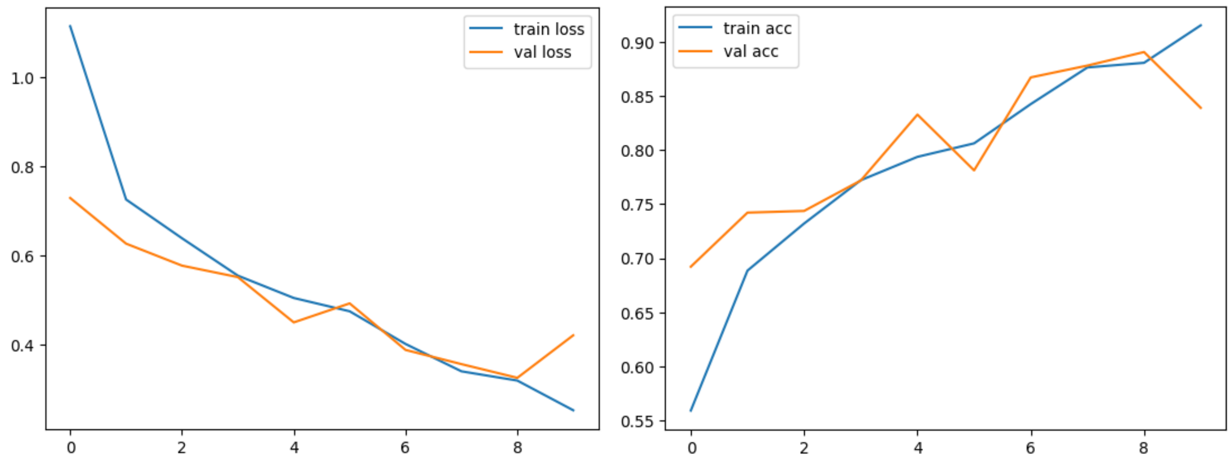


Figure 4.1: Training loss, validation loss and training accuracy, validation accuracy plot for ResNet50

- **Model Test after Training:**
  This table shows the performance of a machine learning model. The loss value is 0.4933, which means the model made some small errors, but they are not too high. The test accuracy is 0.8158, which means the model correctly predicted the results most of the time and is performing well.

Table 4.1: Testing the ResNet50 Model after Training

Matrix	**Value**
Loss	0.4933
Accuracy	0.8158

- **Classification Report Model Performance:**

  We can see in this classification report table that the precision, recall, F1 score of each category has been extracted for the ResNet50 model. And the overall training accuracy has been given as 0.9233.

Table 4.2: Classification report of ResNet50 Model

	precision	recall	F1 score
Angry	0.98	0.66216216	0.79032258
Happy	0.86705202	0.9202454	0.89285714
Sad	0.58558559	0.89041096	0.70652174
Accuracy	0.9233		

- **Confusion Matrix:**

  A normalized confusion matrix for the ResNet50 model, illustrating its classification performance across three emotion categories: angry, happy, and sad. The matrix shows how often predictions matched the actual labels and where misclassifications occurred. Darker shades along the diagonal indicate higher accuracy for those classes, while lighter shades off the diagonal represent misclassified cases. The model demonstrates reasonable performance with some overlap between similar classes.

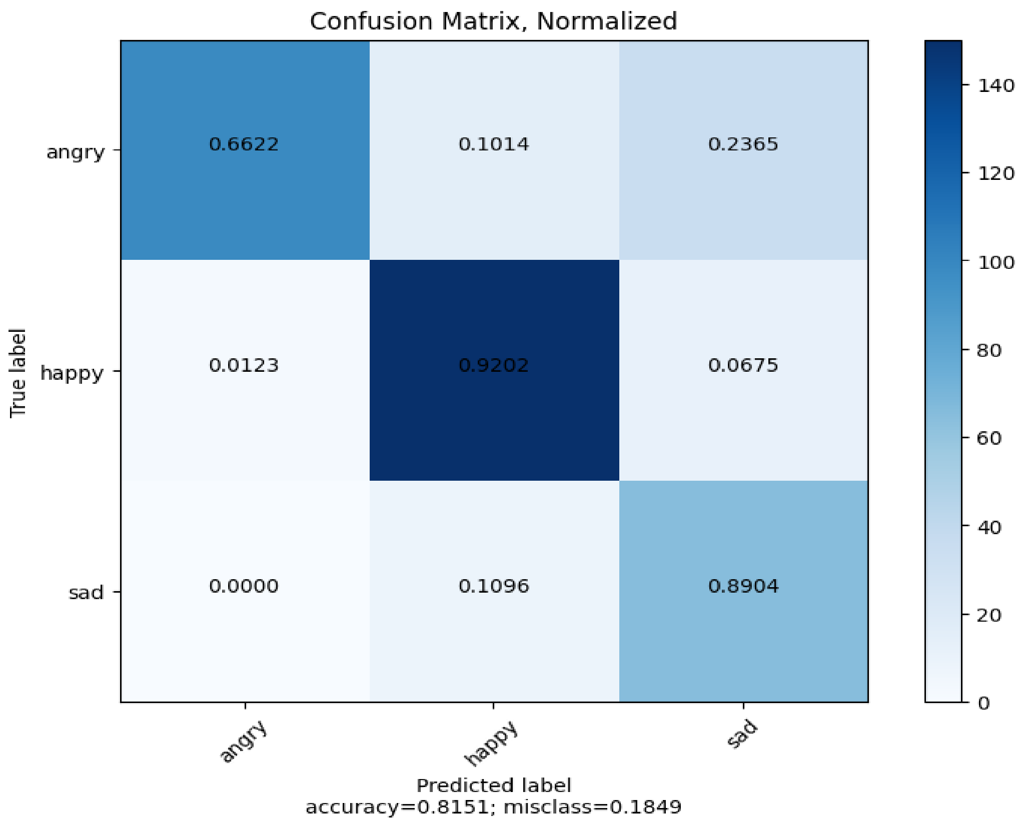


Figure 4.2: Confusion matrix for ResNet50

- **Prediction:**

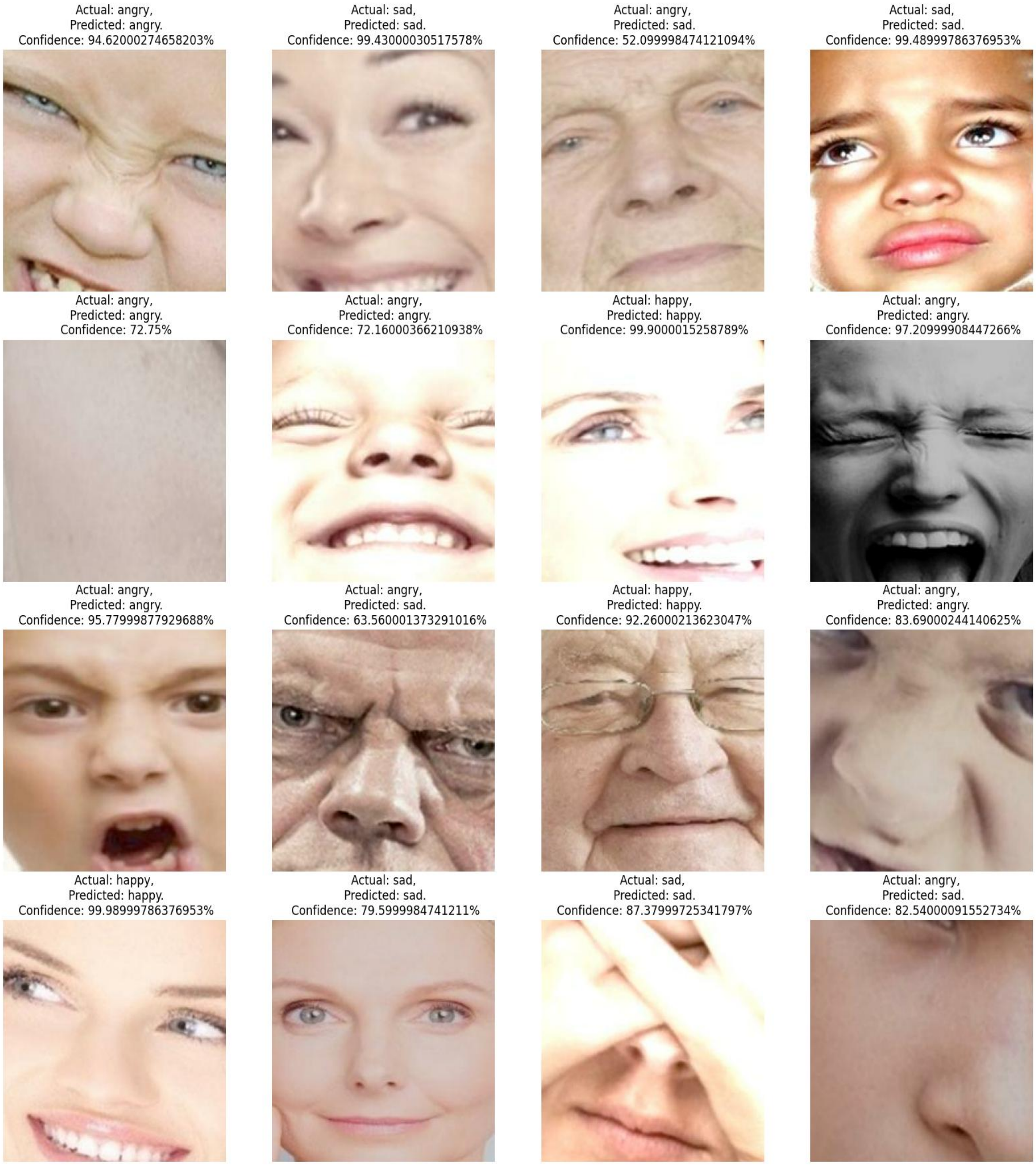


Figure 4.3: Prediction for ResNet50

4.3.2 DenseNet169

- **Training Loss, Validation Loss and Training Accuracy, Validation Accuracy:**

  The graphs show the DenseNet169 model's performance over 10 epochs. The left graph indicates steady improvement in both training and validation accuracy, reaching over 51.43%. The right graph shows a consistent decrease in training and validation loss, confirming good model convergence with minimal overfitting and strong generalization capability.

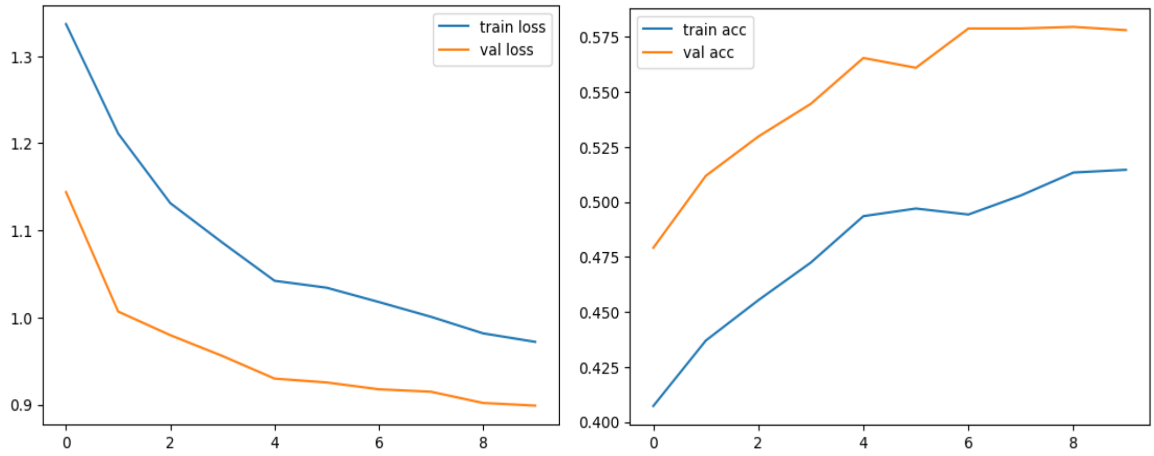


Figure 4.4: Training loss, validation loss and training accuracy, validation accuracy plot for DenseNet169

- **Model Test after Training:**

  This table shows the performance of a machine learning model. The loss value is 0.9168, which means the model made some small errors, but they are not too high. The test accuracy is 0.5485, which means the model correctly predicted the results most of the time and is performing well.

Table 4.3: Testing the DenseNet169 Model after Training

Matrix	Value
Loss	0.9168
Accuracy	0.5485

- **Classification Report Model Performance:**

  We can see in this classification report table that the precision, recall, F1 score of each category has been extracted for the DenseNet169 model. And the overall training accuracy has been given as 0.5143.

Table 4.4: Classification report of DenseNet169 Model

	precision	recall	F1 score
Angry	0.59288538	0.51724138	0.55248619
Happy	0.50126582	0.76153846	0.60458015
Sad	0.51785714	0.18831169	0.27619048
Accuracy	0.5143		

- **Confusion Matrix:**

  A normalized confusion matrix for the DenseNet169 model, illustrating its classification performance across three emotion categories: angry, happy, and sad. The matrix shows how often predictions matched the actual labels and where misclassifications occurred. Darker shades along the diagonal indicate higher accuracy for those classes, while lighter shades off the diagonal represent misclassified cases. The model demonstrates reasonable performance with some overlap between similar classes.

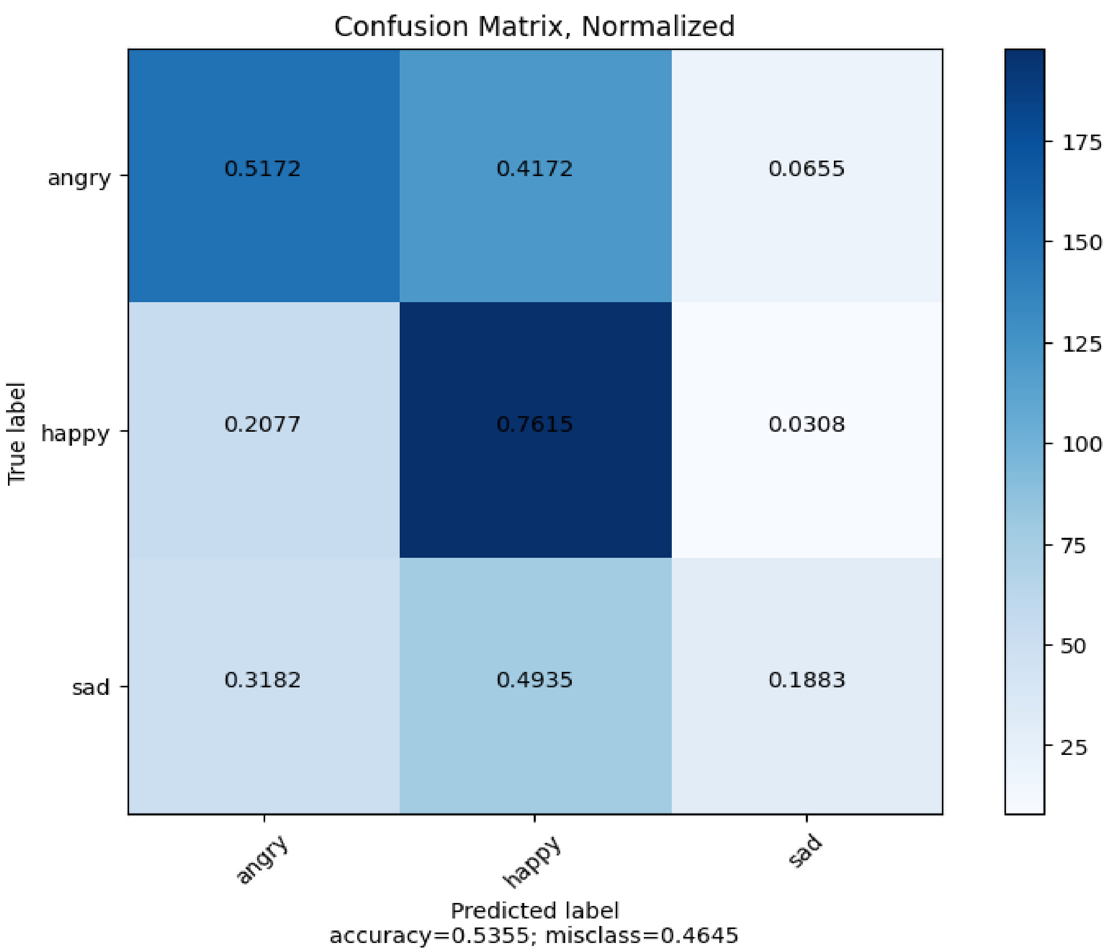


Figure 4.5: Confusion matrix for DenseNet169

- **Prediction:**

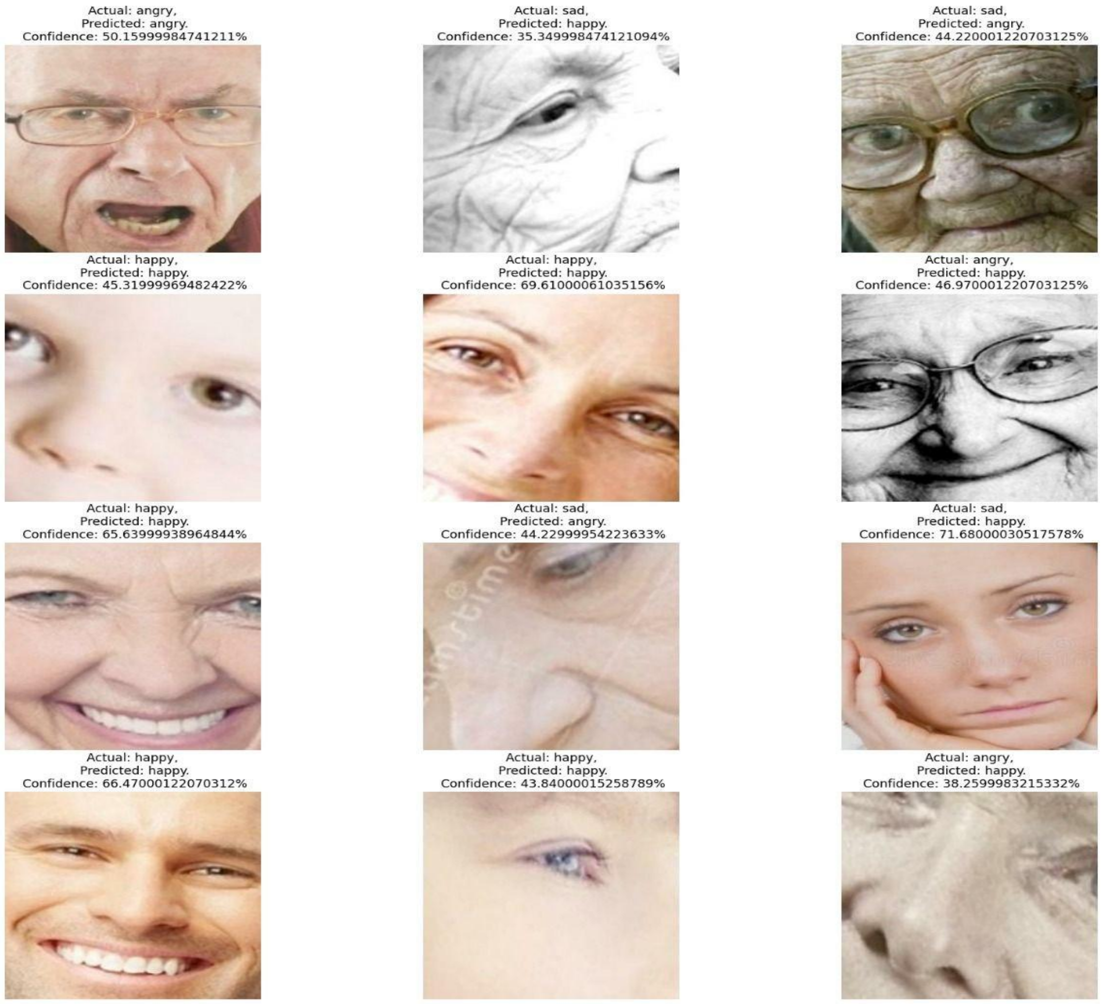
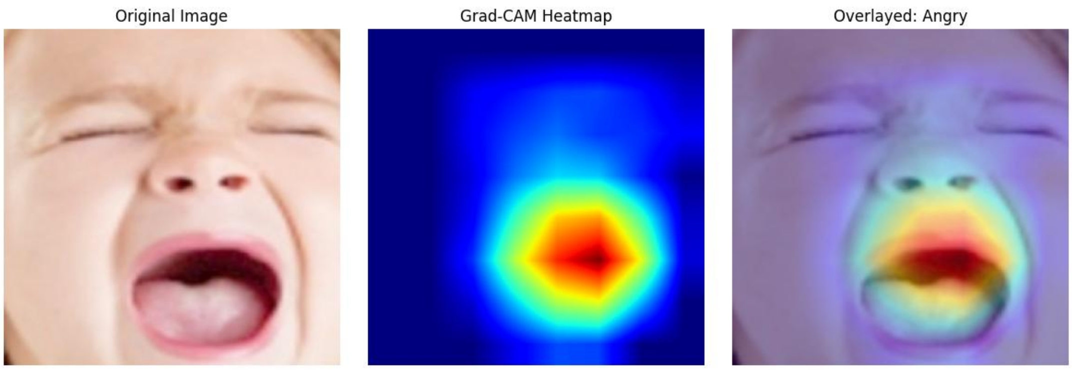


Figure 4.6: Prediction for DenseNet169

- **Prediction with XAI:**

Table 4.5: Prediction with XAI



XAI Predicted Emotion	XAI Predicted Accuracy
Angry	83.59% confidence

The image above illustrates the output of a deep learning model for emotion detection using Grad-CAM visualization. It consists of three sections. The left panel shows the original input image of a child expressing a strong facial emotion. The model has predicted the emotion as "Angry" with 83.59% confidence. The middle panel displays the Grad-CAM heatmap, highlighting the regions of the image that contributed most to the model's decision. The intense red and yellow areas around the mouth region indicate high model attention. The right panel overlays the heatmap onto the original image, clearly showing that the model focused on the mouth and lower face region to classify the emotion. This visual explanation, enabled by Explainable AI (XAI), enhances interpretability and trust in the model's predictions by making it transparent which facial features influenced the classification outcome. This approach is especially useful in sensitive applications like emotion recognition.

4.3.3 CNN

- **Training Loss, Validation Loss and Training Accuracy, Validation Accuracy:**
  The graphs show the CNN model's performance over 10 epochs. The left graph indicates steady improvement in both training and validation accuracy, reaching over 78.25%. The right graph shows a consistent decrease in training and validation loss, confirming good model convergence with minimal overfitting and strong generalization capability.

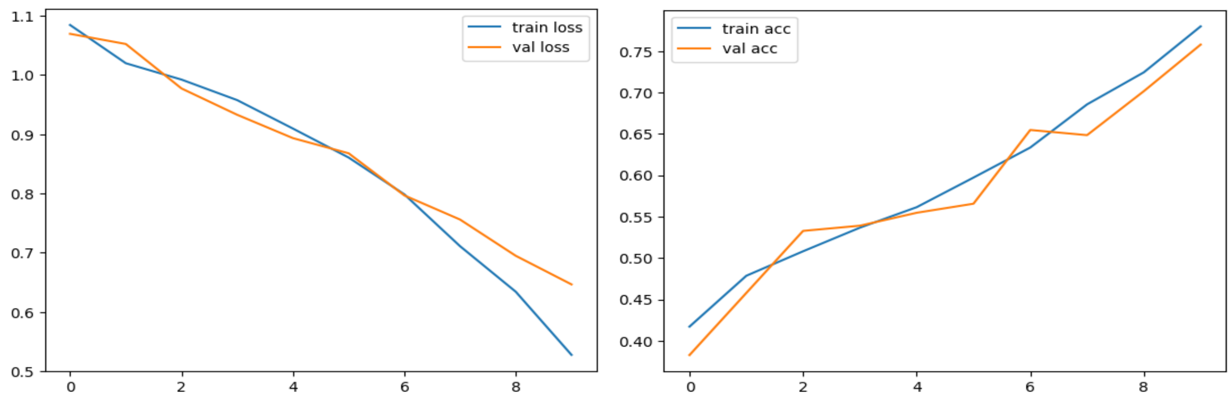


Figure 4.7: Training loss, validation loss and training accuracy, validation accuracy plot for CNN

- **Model Test after Training:**
  This table shows the performance of a machine learning model. The loss value is 0.6247, which means the model made some small errors, but they are not too high. The test accuracy is 0.7664, which means the model correctly predicted the results most of the time and is performing well.

Table 4.6: Testing the CNN Model after Training

Matrix	Value
Loss	0.6247
Accuracy	0.7664

- **Classification Report Model Performance:**

  We can see in this classification report table that the precision, recall, F1 score of each category has been extracted for the CNN model. And the overall training accuracy has been given as 0.7825.

Table 4.7: Classification report of CNN Model

	precision	recall	F1 score
Angry	0.85365854	0.67741935	0.75539568
Happy	0.70930233	0.85915493	0.77707006
Sad	0.70786517	0.72413793	0.71590909
Accuracy	0.7825		

- **Confusion Matrix:**

  A normalized confusion matrix for the CNN model, illustrating its classification performance across three emotion categories: angry, happy, and sad. The matrix shows how often predictions matched the actual labels and where misclassifications occurred. Darker shades along the diagonal indicate higher accuracy for those classes, while lighter shades off the diagonal represent misclassified cases. The model demonstrates reasonable performance with some overlap between similar classes.

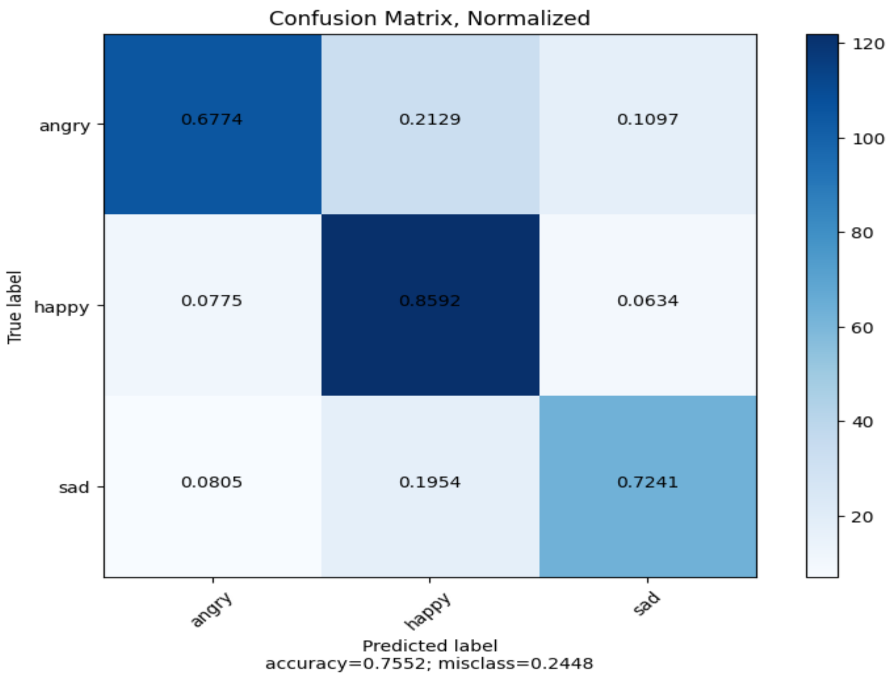


Figure 4.8: Confusion matrix for CNN

- **Prediction:**

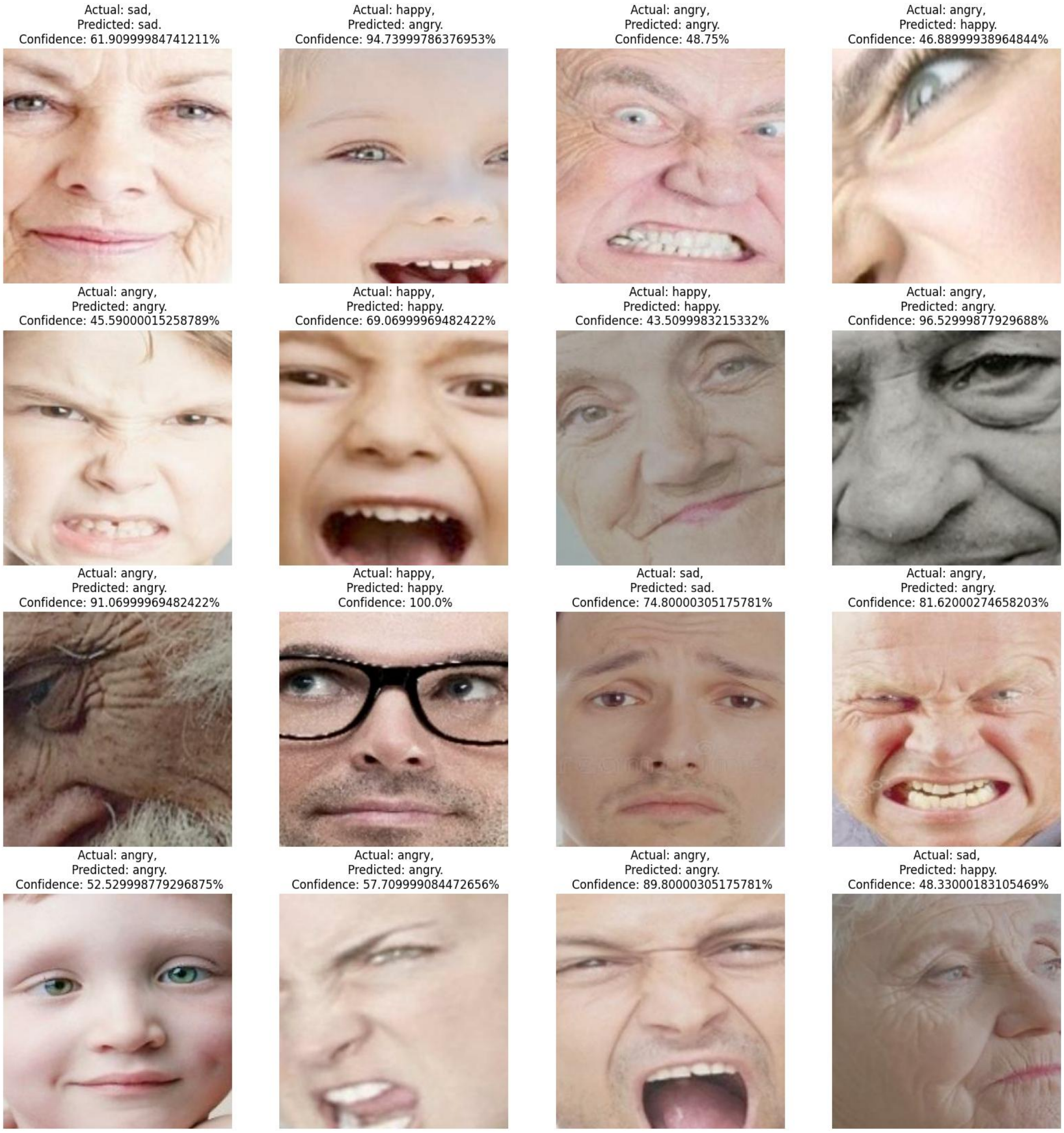


Figure 4.9: Prediction for CNN

4.3.4 EfficientNetB2

- **Training Loss, Validation Loss and Training Accuracy, Validation Accuracy:**

  The graphs show the EfficientNetB2 model's performance over 10 epochs. The left graph indicates steady improvement in both training and validation accuracy, reaching over 68.55%. The right graph shows a consistent decrease in training and validation loss, confirming good model convergence with minimal overfitting and strong generalization capability.

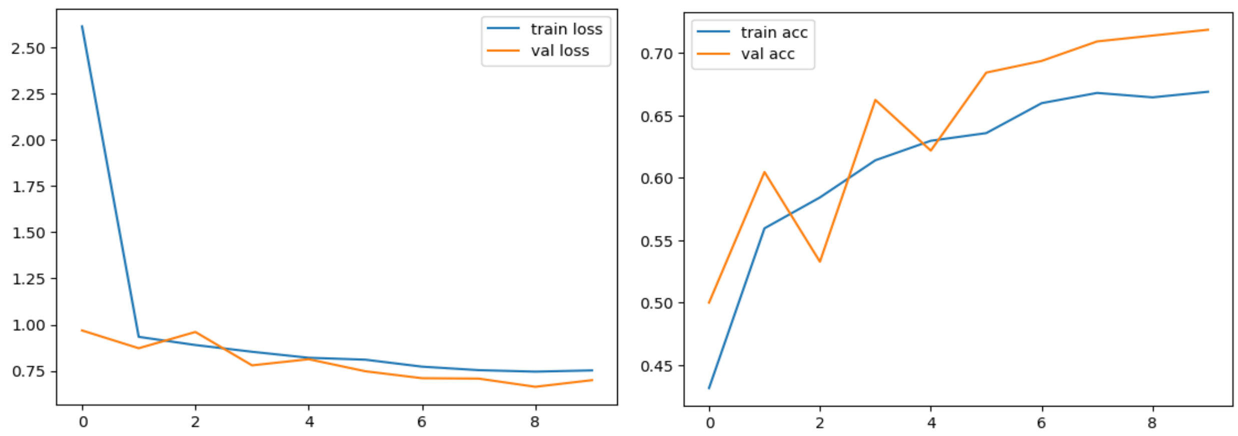


Figure 4.10: Training loss, validation loss and training accuracy, validation accuracy plot for EfficientNetB2

- **Model Test after Training:**

  This table shows the performance of a machine learning model. The loss value is 0.6911, which means the model made some small errors, but they are not too high. The test accuracy is 0.7226, which means the model correctly predicted the results most of the time and is performing well.

Table 4.8: Testing the EfficientNetB2 Model after Training

Matrix	**Value**
Loss	0.6911
Accuracy	0.7226

- **Classification Report Model Performance:**

  We can see in this classification report table that the precision, recall, F1 score of each category has been extracted for the EfficientNetB2 model. And the overall training accuracy has been given as 0.6855.

Table 4.9: Classification report of EfficientNetB2 Model

	precision	**recall**	**F1 score**
Angry	0.81746032	0.67763158	0.74100719
Happy	0.78807947	0.75796178	0.77272727
Sad	0.52336449	0.74666667	0.61538462
Accuracy	0.6855		

- **Confusion Matrix:**

  A normalized confusion matrix for the EfficientNetB2 model, illustrating its classification performance across three emotion categories: angry, happy, and sad. The matrix shows how often predictions matched the actual labels and where misclassifications occurred. Darker shades along the diagonal indicate higher accuracy for those classes, while lighter shades off the diagonal represent misclassified cases. The model demonstrates reasonable performance with some overlap between similar classes.

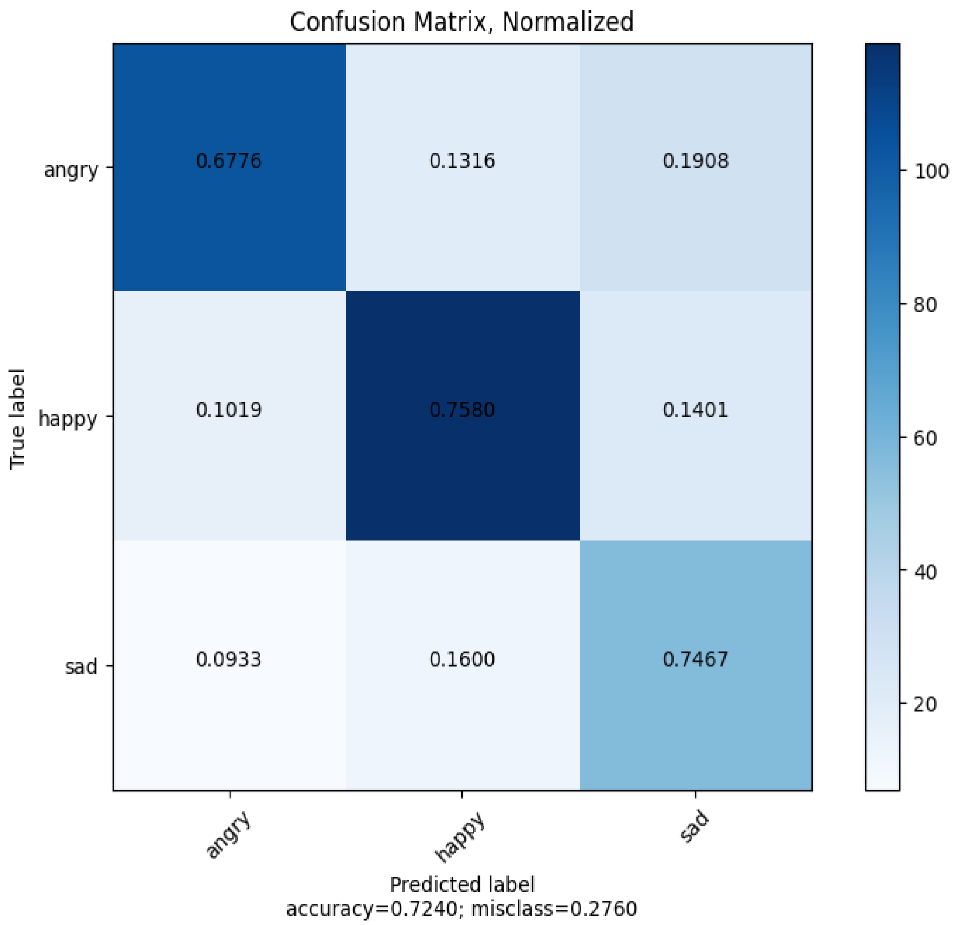


Figure 4.11: Confusion matrix for EfficientNetB2

• **Prediction:**

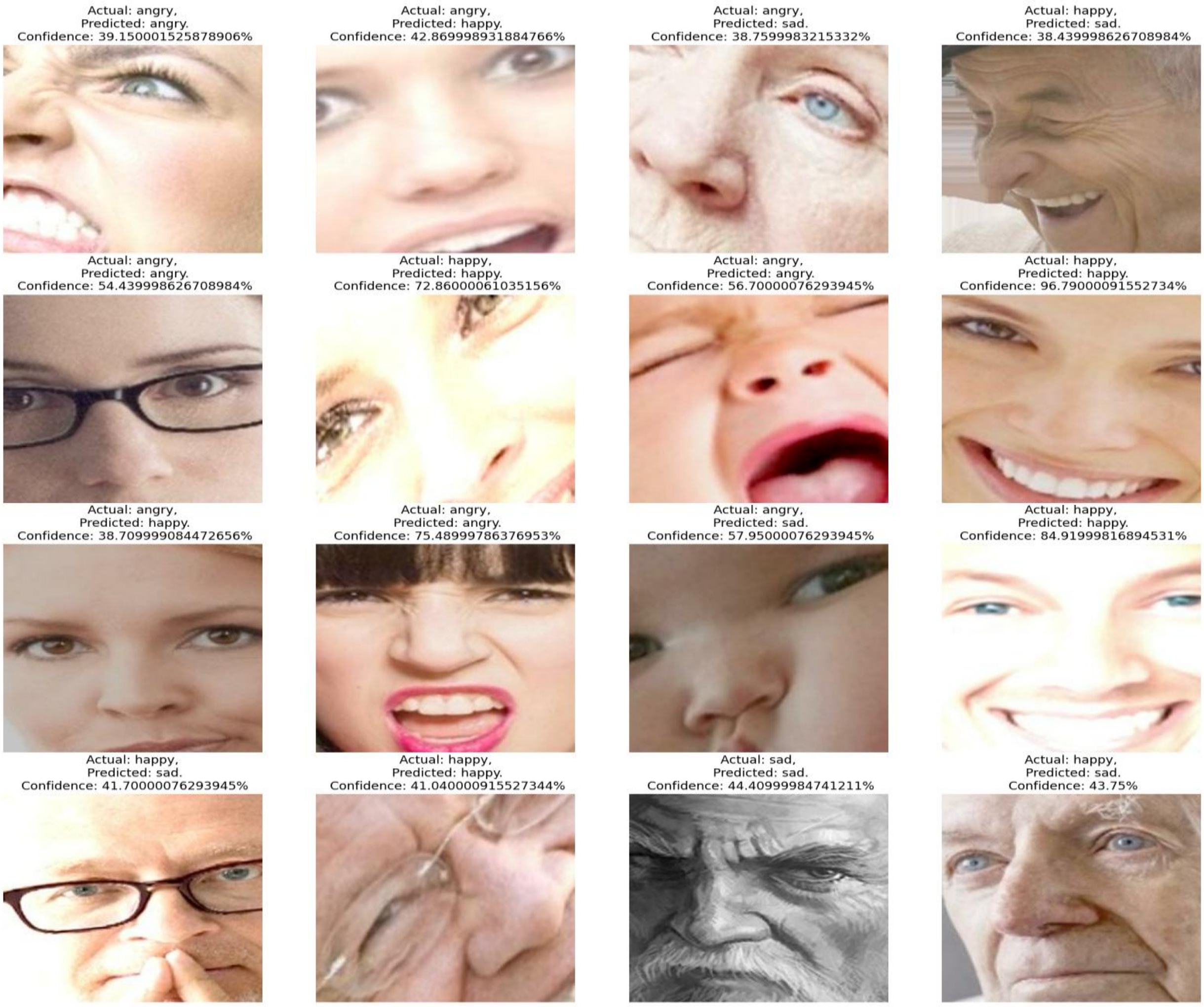


Figure 4.12: Prediction for EfficientNetB2

4.4 Summary

A XAI-based deep learning system operates within the "Human Emotion Detection: A Deep Learning Approach with XAI" project to discern human emotions utilizing photographic records. The first stage of the work involves collecting images which specifically serve computer vision and image processing functions. The images receive different processing operations which improve their quality before deep learning model usage. The data processing results in three separate data subdivisions for training, testing and validation purposes. A group of multiple models uses training data to conduct tests on testing data with validation data applied for model assessment. Each model undergoes evaluation through analysis of performance metrics and results produced during the testing stages. The implementation of Explainable AI ensures the model's decision-making becomes clear to end-users thus enabling better interpretation abilities.

Chapter 5

Engineering Standards And Design Challenges

5.1 Compliance with the Standards

5.1.1 Software Standard

Standard software requirements in the Human Emotion Detection project use established development platforms and tools to offer reliable standards in both efficiency and team collaboration. ResearchGate together with Google Scholar provide scientific literature which can be trusted. Google Colab functions as the main platform developers use for deep learning model creation and training through Python and TensorFlow and Keras programming frameworks. The development as well as the evaluation of machine learning models and deep learning models uses default programming tools. The project team uses their personal computer equipped with GPU components to run tests from the local work environment. The research reporting adheres to academic standards throughout the process whereas documentation uses Microsoft Word to deliver professional and clear findings.

5.1.2 Hardware Standard

- **MINIMUM:**
  - ❖ OS: Windows 7,8,8.1
  - ❖ Processor: I3 latest gen
  - ❖ Memory: 4 GB RAM
  - ❖ Graphics: 4GB GRAPHICS
  - ❖ DirectX: Version 11
  - ❖ Storage: 3 GB available space
- **RECOMMENDED:**
  - ❖ OS: Windows 10
  - ❖ Processor: I5
  - ❖ Memory: 8 GB RAM
  - ❖ Graphics: 6GB or above GRAPHICS
  - ❖ DirectX: Version 12
  - ❖ Storage: 3 GB available space

5.1.3 Communication Standard

The communication language utilized by Human Emotion Detection is a communication protocol that allows to establish a uniform communication to make good team work. Project status is tracked via Zoom and Google Meet scheduling that merges updates of task assignments with solutions of challenges. The project is under version control and information is available to everyone, as we share updates and documentations via Google Drive. 1. The WhatsApp and Email is for Quick Daily communication and for Current Affairs. All documentation, including reports, is to be written using structured word-processor files in MS Word. Using the feedback loops, the project maintains an open channel of communication that fosters transparence and timely revision processes and team-based decision-making processes to aid swift workflow from project start to project finish.

5.2 Impact on Society, Environment and Sustainability

5.2.1 Impact on Life

Explainable AI (XAI) In the Human Emotion Detection project, there's a clear opportunity for transformative, cross-cutting positive social good across sectors with better human-machine interaction and better processing of emotions by AI systems. Major societal benefits are- This system is a screening stage to catch early signs of emotional distress and assist health care professionals in identifying those who are at the risk for depression and anxiety for early intervention. The interactions of businesses with their customers can be enhanced when emotion detection technology is embedded into customer service and interactive systems; by responding adaptively and in an empathetic manner to customers' moods and emotions, they can provide personalized responses. Emotion detection can be used in educational institutions to track and observe the emotional states of students to help teacher in adjusting teaching strategy and support student's emotion needs. By implementing this system security facilities are enabled to detect threatening emotional states while the security value increases of public surveillance. The integration of XAI provides emotion detection systems with transparency in addition to trust resulting in the removal of public concerns on AI decision and privacy.

This promotes AI emotional comprehension for enhancing social empathy amend human society.

5.2.2 Impact on Society & Environment

The key focus of the Human Emotion Detection project is in human-AI interaction and AI knowledge transparent; however, its effective use of resource may bring environmental benefits. Significant environmental effects are-Accelerated methods employed for training of deep learning models reduce the computational resources that need to be used for training, which mitigates energy usage during long AI training tasks; Deploying models in the cloud achieves resource sharing that has lower local hardware requirements and is more efficient and environmentally friendly in terms of optimized power usage (data centers waste less power and environmental damage related to climate change is lower); Developers are building widely valuable tools like emotion recognition tools for AI in mental health support or customer service. This gives developers a sustainable advantage where they no longer need to rely on adding more human staff and office properties and pressure the physical environment to serve its services to their customers, but increasingly serve customers online with their virtual users and add less physical infrastructure and by that, become more environmentally friendly.

The project will deliver a double return to society when in operation to promote sustainable progress of technology.

5.2.3 Ethical Aspects

Several ethical factors affect the implementation and use of the Human Emotion Detection project to develop responsible AI methods. All data used for training as well as testing must receive informed consent and necessitates secure storage to safeguard the privacy of users' sensitive information (images of individuals). The training data must include diverse populations because biases in emotion detection should be eliminated between demographic groups such as age and gender and ethnicity. Measures should be developed to prevent discrimination together with unfair treatment practices. XAI technology ensures both transparency and accountability feature by making AI decision-making processes easy to understand so users can trust and monitor AI systems for their prediction outputs. Social protection measures need implementation to stop unethical situations where emotion detection technology gets misused via unauthorized surveillance or manipulation activities.

5.2.4 Sustainability Plan

If the team want to successfully operate the system long term under sustainable operation for the Human Emotion Detection project, the project team should consider sustainable development initiatives with consideration for ethical issues and environmental sustainability. The plan for sustainability includes key elements, including, Our company leverages algorithmic optimization to reduce the deep learning resource, and our contribution is the reduced environmental impact from deep learning process. The system includes design elements to facilitate the efficient scale-up for industrial applicability in wider areas keeping resource input low so that it can serve for the long term in an operational setup. We will continue a process of ongoing model development with updated data that addresses the ethical, privacy and technological challenges we face in making our system both effective and relevant. Green Technologies are realized through cloud-based services and sustainable infrastructure that reduces reliance on physical hardware, and at the same time, resource sharing will be optimized and consolidated computing to run datacenters for minimal environmental impact.

These provide a cushion, to protect the project from falling short in the long run from delivering technically sustainable technologies and safeguarding the environment.

5.3 Project Management and Financial Analysis

5.3.1 Project Management:

This project applies control processes to facilitate successful delivery and successful project closing. The key phases include:

- Using Data to Define the Goal and Mark Milestones The first stage consists of collecting data sources to the aim of the project and setting up targets and milestones to reach to the goals of the project.
- The phase is planned in two steps: first, initial image processing, and second, choosing AI models and with explaining their results.
- Experiments for model development as well as parameter tuning is run and the accuracy as well as precision and recall details are used to validate the model performance.
- The performance efficiency of the model is successfully evaluated through tests in real applications.

There's report generation and findings documentation which can provide useful info for future optimization tactics.

5.3.2 Financial Analysis:

Table 5.1: Financial Analysis Report

Components	Estimated Cost (BDT)
Laptop	70000
Wi-Fi	2000
Internet	1000
Software and Tools	7000
Data Collection and Processing	5000
Documentation and Report Writing	3000
Transportation Fare	1000
Total Estimated Cost	88000

5.4 Complex Engineering Problem

5.4.1 Complex Problem Solving

The complex engineering problem involves developing an interpretable deep learning-based emotion detection system from facial images, balancing high model accuracy with transparency. It requires integrating advanced CNN and XAI techniques while addressing challenges like data variability, ethical concerns, overfitting, and generalization—especially for sensitive applications like mental health and human-computer interaction.

Table 5.2: Mapping with complex problem solving

EP1 Dept of Knowledge	EP2 Range of Conflicting Requirements	EP3 Depth of Analysis	EP4 Familiarity of Issues	EP5 Extent of Applicable Codes	EP6 Extent of Stake- holder Involvement	EP7 Interdependence
✓		✓	✓			

Mapping with Knowledge Profile for EP1

This table 5.3 is designed to map the EP1 to the Knowledge Profile.

Table 5.3: Mapping with knowledge Profile

K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K8 Research Literature
✓	✓	✓	✓	✓

5.4.1.1 Justification for EP Attributes Mapping

- **EP1 – Depth of Knowledge:**
  Demonstrates strong expertise in deep learning and XAI using ResNet50, InceptionV3, and Grad-CAM to build transparent, interpretable emotion detection systems with advanced model architectures.
- **EP3 – Depth of Analysis:**
  Presents a multidimensional analysis combining accuracy, interpretability (Grad-CAM), and ethical insights, linking technical performance with real-world emotion detection and mental health applications.
- **EP4 – Familiarity of Issues:**
  Addresses challenges like overfitting, dataset limitations, and ethical AI use, applying best practices to ensure accurate, explainable, and generalizable emotion recognition.

5.4.1.2 Justification for Knowledge Profile Mapping (Linked to EP1)

- **K3 – Engineering Fundamentals:**
  Applies core engineering practices like data cleaning, error analysis, and model refinement to systematically design an interpretable and functional emotion detection AI system.
- **K4 – Specialist Knowledge:**
  Combines CNNs, Vision Transformers, and Grad-CAM, showcasing expert-level knowledge in deep learning, computer vision, and XAI for secure, practical emotion recognition systems.
- **K5 – Engineering Design:**
  Follows a full design cycle balancing model precision and interpretability, leading to scalable, ethical, and user-centric solutions in emotion recognition and mental health applications.

- **K6 – Engineering Practice:**
  Implements open-source tools, testing protocols, and ethical AI practices to ensure transparency, professionalism, and responsible integration in emotion detection research.
- **K8 - Research Literature:**
  Uses extensive literature to guide model selection, identify gaps, and validate approaches—ensuring the project aligns with established research while contributing new insights.

5.4.2 Engineering Activities

This section is provided with a mapping with engineering activities. For each mapping, subsections are added to put rationale (Table 5.4).

Table 5.4: Mapping with complex engineering activities

EA1 Range of resources	EA2 Level of Interaction	EA3 Innovation	EA4 Consequences for society and environment	EA5 Familiarity
✓		✓	✓	✓

5.4.2.1 Justification for Engineering Activities Mapping:

- **EA1 – Range of Resources:**
  Utilizes GPUs, large image datasets, TensorFlow, PyTorch, and Grad-CAM to enable efficient model training and interpretation in limited-resource environments.
- **EA3 – Innovation:**
  Introduces explainable AI with Grad-CAM for emotion detection, enhancing model transparency and trust—rarely emphasized in existing emotion recognition systems.
- **EA4 – Societal and Environmental Implications:**
  Supports ethical emotion recognition for mental health and HCI (Human-computer interaction), emphasizing privacy, consent, and responsible deployment to prevent misuse of facial data.

- **EA5 – Familiarity:**
  Combines standard deep learning with novel XAI, advancing transparent emotion detection and encouraging continual learning in ethical and explainable AI development.

5.1.4 Summary

The “Human Emotion Detection: A Deep Learning Approach with XAI” project demonstrates a high level of engineering professionalism by addressing fundamental aspects of Engineering Problems (EP), Knowledge (K) and Engineering Activities (EA). Stylistically complicated implementation is required in the form of sophisticated compromises of the conflicting demands that are analyzed down to their bones while simultaneously involving proponents to ensure expertise in dependent variable management (EP1–EP7). The assignment integrates basic engineering aspects and domain specific materials with design skills and hands-on experience, and is based on literature in the field (K3–K8). The initiative combines a range of technical resources using an activity-based approach to bridge different disciplines, allowing contributions to the creation of advanced Explainable AI. Social and environmental impact is reviewed and addressed in the project and new areas are ventured on the basis of practical knowledge and ethical competence (EA1–EA5). This work presents the complete end to end implementation of advanced engineering knowledge to advance discipline responsible transparent emotion recognition systems.

Chapter 6

Conclusion

6.1 Summary

The "Human Emotion Detection" research utilizes Explainable AI (XAI) models to analyze image-based data for detecting human emotions. The research process starts by accumulating images to perform processing functions as part of image processing and computer vision applications. Different preprocessing methods modify the images to prepare them for deep learning model utilization. The data processing step results in partitioning the information into training data along with testing and validation sections. Training multiple models occurs on training data before evaluative performance measurements are obtained by running tests on testing data and validation data. Evaluation of model effectiveness happens through results analysis that includes accuracy metrics from different stages of testing. The implementation of Explainable AI allows the model to show its decision-making steps resulting in improved interpretability. We also deployed our best-performing model online and are currently reading the output of Visual Emotion. The main goal of this project involves creating emotion detection systems that both achieve accuracy and transparency alongside responsible operation.

6.2 Limitation

The dataset used for training usually features limited emotional diversity which hinders the development of models that adequately perform across different facial expressions within real-life scenarios.

Diverse image quality attributes that include different camera positions alongside variable resolution levels and lighting conditions reduce the detection performance of emotion targets by the model.

The challenge of making deep learning models explain their processes exists mainly because complex neural networks hinder human understanding of these solutions.

The achievement of real-time emotion detection systems with dependable accuracy proves difficult to obtain when dealing with images from various dynamic environments.

The development of deep learning models needs better interpretation techniques because stronger model performance in emotion classification must be maintained.

6.3 Future Work

The Human Emotion Detection project has created numerous prospects for future research projects which focus on development activities:

- Research aimed at improving model accuracy should investigate both convolutional neural networks (CNNs) and transformer-based models for detecting emotion cues in the data.
- Future work needs to expand data resources by adding video and multimodal information such as voice and text for improving the robustness and generalizability of emotion detection models.
- A study of cultural and environmental factors on emotions must be performed to validate these models work appropriately in various global populations.
- Studying ethical issues about emotion-detection technology in sensitive situations during further research will assist in resolving privacy-related concerns and achieving better ethical standards.

The provided directions serve to enhance emotion detection technologies by checking their ethical soundness and worldwide applicability.

References

[1] Khalane, R. Makwana, T. Shaikh, and A. Ullah, “Evaluating significant features in context-aware multimodal emotion recognition with XAI methods,” Expert Systems, Aug. 2023, doi: https://doi.org/10.1111/exsy.13403.

[2] Rathod, M.; Dalvi, C.; Kaur, K.; Patil, S.; Gite, S.; Kamat, P.; Kotecha, K.; Abraham, A.; Gabralla, L.A. Kids’ Emotion Recognition Using Various Deep-Learning Models with Explainable AI. Sensors 2022, 22, 8066. https://doi.org/ 10.3390/s22208066.

[3] Kerz E, Zanwar S, Qiao Y and Wiechmann D (2023) Toward explainable AI (XAI) for mental health detection based on language behavior. Front. Psychiatry 14:1219479. doi: 10.3389/fpsyt.2023.1219479.

[4] J. M. Mayor Torres, S. Medina-DeVilliers, T. Clarkson, M. D. Lerner, and G. Riccardi, “Evaluation of interpretability for deep learning algorithms in EEG emotion recognition: A case study in autism,” Artificial Intelligence in Medicine, vol. 143, p. 102545, May 2023, doi: https://doi.org/10.1016/j.artmed.2023.102545.

[5] S. Medina-DeVilliers, T. Clarkson, M. D. Lerner, and G. Riccardi, “Evaluation of Interpretability for Deep Learning algorithms in EEG Emotion Recognition: A case study in Autism,” arXiv (Cornell University), Jan. 2021, doi: https://doi.org/10.48550/arxiv.2111.13208.

[6] Elugbadebo, Olabode. (2025). Cross-Cultural Emotion Detection Using Deep Learning: A Comparative Study.

[7] Sophia, Evelyn. (2025). Multimodal Deep Learning for Emotion Detection from Text, Audio, and Facial Expressions.

[8] Halkiopoulos, C.; Gkintoni, E.; Aroutzidis, A.; Antonopoulou, H. Advances in Neuroimaging and Deep Learning for Emotion Detection: A Systematic Review of Cognitive Neuroscience and Algorithmic Innovations. Diagnostics 2025, 15, 456. https://doi.org/10.3390/diagnostics15040456.

[9] R. K. Nukathati, “A Deep Learning Framework with Optimizations for Facial Expression and Emotion Recognition from Videos,” Emerging Science Journal, vol. 16, no. 3, 2025.

[10] M. Shahid, Nitun Kumar Podder, S. M. Hasan, Abu, and M. A. Rahim, “Bangla Speech Emotion Recognition Using Deep Learning-Based Ensemble Learning and Feature Fusion,” Mar. 2025, doi: https://doi.org/10.20944/preprints202503.1864.v1.

[11] M. Kondal “Methodical Review of Deep Learning Macro and Micro Face Emotion Detection,” International Journal of Communication Networks and Information Security, Dec. 2024, doi: https://doi.org/10.13140/RG.2.2.20023.02728.

[12] Govea J, Navarro AM, Sánchez-Viteri S and Villegas-Ch W (2024) Implementation of deep reinforcement learning models for emotion detection and personalization of learning in hybrid educational environments. Front. Artif. Intell. 7:1458230. doi: 10.3389/frai.2024.1458230.

[13] D. B. Joshi "Emotion Detection using Transformer Model with Deep Learning," International Journal of Engineering Research & Technology, March. 2024, doi: https://doi.org/10.17577/IJERTV14IS030116.

[14] P. H. Hussan "DLFER: Deep Learning-based Cascaded Approach for Facial Emotion Recognition," Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE), March. 2025, doi:https://doi.org/10.46649/fjiece.v4.1.19a.25.3.2025.

[15] CHETHAN R and V. PATEL, "Deep Learning Based Emotion Detection System," International Journal Of Scientific Research In Engineering And Management, vol. 08, no. 07, pp. 1–13, Jul. 2024, doi: https://doi.org/10.55041/ijsrem36445.

[16] N. Vinod and M. Vinuthna, "Emotion Detection from Facial Expressions Using Deep Learning," International Journal Of Scientific Research In Engineering And Management, vol. 09, no. 04, pp. 1`9, Apr. 2025, doi: https://doi.org/10.55041/ijsrem44585.

[17] N. Ashen, "Retrieval based Emotionally responsive chatbot Utilizing emotions clustering and XAI Techniques.," Research Square (Research Square), Apr. 2024, doi: https://doi.org/10.21203/rs.3.rs-4291342/v1.

[18] Schwerk, Anne & Grasnick, Armin. (2025). PHANTOMATRIX: Explainability for Detecting Gender Bias in Affective Computing. International Conference on Gender Research. 8. 515-518. 10.34190/icgr.8.1.3199. https://doi.org/10.34190/icgr.8.1.3199

212-15-4131

ORIGINALITY REPORT

20%	15%	9%	13%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	6%
2	Submitted to Daffodil International University Student Paper	3%
3	Submitted to United International University Student Paper	1%
4	www.frontiersin.org Internet Source	1%
5	www.mdpi.com Internet Source	1%
6	Motlagh, Ali Habibzadeh. "Detection, Classification, and Modeling Evaluation of Network Intrusions", Ecole Polytechnique, Montreal (Canada), 2023 Publication	1%
7	"Advances in Artificial Intelligence and Electronic Design Technologies", Springer Science and Business Media LLC, 2025 Publication	1%
8	Submitted to Southampton Solent University Student Paper	1%
9	Submitted to University of East London Student Paper	<1%
10	Satya Ranjan Mishra, Apul Narayan Dev, Alok Kumar Pandey, Mukesh Kumar Awasthi. "Design Optimization Using Artificial Intelligence", CRC Press, 2025	<1%