A project report on

ACTION UNIT-BASED EMOTION DETECTION: INTENSITY-AWARE FACIAL EMOTION RECOGNITION USING PY-FEAT AND ML

Submitted in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY COMPUTER SCIENCE AND ENGINEERING WITH SPL.CYBER PHYSICAL SYSTEMS

by

SHETTY VIDHEE SHRIDHAR (21BPS1526)

SIRIGINEEDI RAJ ABHINAY (21BAI1778)

RENTALA SAI JOSHIKA (21BCE5159)

VIT

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

CHENNAI

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

November,2024

ACTION UNIT-BASED EMOTION DETECTION: INTENSITY-AWARE FACIAL EMOTION RECOGNITION USING PY-FEAT AND ML

Submitted in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY COMPUTER SCIENCE AND ENGINEERING WITH SPL.CYBER PHYSICAL SYSTEMS

by

SHETTY VIDHEE SHRIDHAR (21BPS1526)

SIRIGINEEDI RAJ ABHINAY (21BAI1778)

RENTALA SAI JOSHIKA (21BCE5159)

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

CHENNAI

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

November, 2024

VIT

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

CHENNAI

DECLARATION

I hereby declare that the thesis entitled "ACTION UNIT-BASED EMOTION DETECTION: INTENSITY-AWARE FACIAL EMOTION RECOGNITION USING PY-FEAT AND ML" submitted by SHETTY VIDHEE SHRIDHAR (21BPS1526) , for the award of the degree of Bachelor of Technology in Computer Science and Engineering with Spl. Cyber Physical Sytems, Vellore Institute of Technology, Chennai is a record of bonafide work carried out by me under the supervision of Dr. Balasundaram A.

I further declare that the work reported in this thesis has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university.

Place: Chennai

Date: 15/11/24

Signature of the Candidate

3

VIT

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

CHENNAI

School of Computer Science and Engineering

CERTIFICATE

This is to certify that the report entitled **"Action Unit-Based Emotion Detection: Intensity-Aware Facial Emotion Recognition Using py-feat and ML"** is prepared and submitted by **Shetty Vidhee Shridhar (21BPS1526)** to Vellore Institute of Technology, Chennai, in partial fulfillment of the requirement for the award of the degree of **Bachelor of Technology in Computer Science and Engineering** is a bonafide record carried out under my guidance. The project fulfills the requirements as per the regulations of this University and in my opinion meets the necessary standards for submission. The contents of this report have not been submitted and will not be submitted either in part or in full, for the award of any other degree or diploma and the same is certified.

Signature of the Guide:

Name: Dr. Balasundaram A.

Date: 15/11/24

Signature of the Examiner
Name: [illegible]
Date: 22/11/2024

Signature of the Examiner
Name: [illegible]
Date: 22/11/24

Approved by the Head of Department,
B.Tech. CSE with SPL. Cyber Physical systems

Name: **Dr. Renuka Devi S**

Date: 18/11/24

(Seal of SCOPE)

VELLORE INSTITUTE OF TECHNOLOGY
SCHOOL OF COMPUTER SCIENCE & ENGINEERING
Chennai-600127

4

ABSTRACT

Facial expression recognition has become very popular because of its applications in human-computer interaction, mental health assessment, surveillance, and multimedia. The paper uses py-feat library and works on Action units to create an intensity-aware approach towards making a facial emotion recognition model which is merged with classifiers and includes machine learning. Methods wholly comprises EKman's six universally accepted emotions which are happiness, sadness, anger, disgust, surprise, and fear and create better intensity driven Action units which enhance the overall sensitivity of the subtle emotion recognition. Data pre-processing included real-time image augmentation using Keras and OpenCV to ensure robust model generalization. Performance evaluation was done using multiple classifiers such as Random Forest, Logistic Regression, K-Nearest Neighbors, Decision Tree, and deep learning architectures. The best accuracy achieved as by the Random forest model with 88.61%.This achievement provides a class-wise comprehensive classification report through the implementation of multiple emotion classes that indicate the model's ability to capture slight variations in emotion. The intensity-aware system, with AU-based detection, further advances the accuracy of emotion detection and provides an interpretable framework applicable in real-world settings.

ACKNOWLEDGEMENT

It is my pleasure to express with deep sense of gratitude to Dr. Balasundaram A, Associate professor, School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, for his constant guidance, continual encouragement, understanding; more than all, he taught me patience in my endeavor. My association with his is not confined to academics only, but it is a great opportunity on my part of work with an intellectual and expert in the field of Deep learning and Artificial Intelligence.

It is with gratitude that I would like to extend my thanks to the visionary leader Dr. G. Viswanathan our Honorable Chancellor, Mr. Sankar Viswanathan, Dr. Sekar Viswanathan, Dr. G V Selvam Vice Presidents, Dr. Sandhya Pentareddy, Executive Director, Ms. Kadhambari S. Viswanathan, Assistant Vice-President, Dr. V. S. Kanchana Bhaaskaran Vice-Chancellor, Dr. T. Thyagarajan Pro-Vice Chancellor, VIT Chennai and Dr. P. K. Manoharan, Additional Registrar for providing an exceptional working environment and inspiring all of us during the tenure of the course.

Special mention to Dr. Ganesan R, Dean, Dr. Parvathi R, Associate Dean Academics, Dr. Geetha S, Associate Dean Research, School of Computer Science and Engineering, Vellore Institute of Technology, Chennai for spending their valuable time and efforts in sharing their knowledge and for helping us in every aspect.

In jubilant state, I express ingeniously my whole-hearted thanks to Dr. Dr. Renuka Devi S, Head of the Department, B.Tech. Computer Science and Engineering with SPL. Cyber Physical Systems and the Project Coordinators for their valuable support and encouragement to take up and complete the thesis.

My sincere thanks to all the faculties and staffs at Vellore Institute of Technology, Chennai who helped me acquire the requisite knowledge. I would like to thank my parents for their support. It is indeed a pleasure to thank my friends who encouraged me to take up and complete this task.

Place: Chennai

Date: 15-11-24 **SHETTY VIDHEE SHRIDHAR**

CONTENTS

CHAPTER 3

METHODOLOGY

CHAPTER 4

IMPLEMENTATION

CHAPTER 5

RESULTS AND ANALYSIS

CHAPTER 6

COMPARISON AND DISCUSSION OF MODEL PERFORMANCE

CHAPTER 7

CONCLUSION AND FUTURE WORK

LIST OF FIGURES

LIST OF ACRONYMS

SVM- SUPPORT VECTOR MACHINE

XG- XGBOOST

FER- FACIAL EMOTION RECOGNITION

CNN- CONVOLUTIONAL NEURAL NETWORKS

FACS- FACIAL ACTION CODING SYSTEM

KNN- K-NEAREST EIGHBOURS

AU- ACTION UNIT

Chapter 1

INTRODUCTION

1.1 OVERVIEW OF FACIAL EMOTION RECOGNITION

Technological advancement in Facial Emotion Recognition (FER) is designed to recognize and classify emotions from facial expressions. This capability is required for human computer interaction and mental health assessment as well as multimedia analysis. FER systems can recognize emotions by looking at subtle variations in movements of the facial muscles and this increases our understanding of non verbal communication.

FER systems relied historically upon manually designed features like Local Binary Patterns (LBP) and Histograms of Oriented Gradients (HOG) to identify crucial face image attributes. But those were the traditional methods and they had a rather limited accuracy, particularly given the wide variety of expressions and cluttered backgrounds. In the last couple of years, the advent of machine learning and deep learning techniques has driven the advancement of further sophisticated FER models, especially ones employing Convolutional Neural Networks (CNNs), making significant gains in emotion recognition based purely on the learning of higher order features from raw image data.

FER in a contemporary method makes use of Action Units (AUs), which concentrate only on the particular muscle movements that Ekman’s facial action coding system (FACS) describes as related to emotional expressions. An AU-centric approach provides a clear, structured framework for emotion classification, and if folded into intensity aware models that detect the subtleties of facial expressions, this can be applied effectively. Such techniques enable FER systems to classify emotions with greater accuracy by evaluating the intensity and specific combinations of muscle movements corresponding to six universally acknowledged emotions: six emotions happiness, sadness, anger, disgust, surprise, and fear.

For achieving high levels of accuracy in FER the study uses the py-feat library along with a machine leaning framework. To solve this problem, the research makes use of a dataset such as the Cohn–Kanade AU Coded Facial Expression Database (CK+) possessing high quality AU annotations and trains Random Forest and other classifiers to assign emotions with an accuracy of 88.61%.In this study, we build an intensity aware system which increases the sensitivity of FER to small facial changes, thus improve model generalization for real world applications.d Facial Expression Database (CK+), which features high-quality AU annotations, the research incorporates Random Forest and other classifiers to effectively categorize emotions, achieving an accuracy rate of 88.61%. The intensity-aware system in this study enhances FER by making it more sensitive to slight facial variations, thus improving model generalization for real-world applications.

1.2 MOTIVATION FOR THE STUDY

In order to achieve these aims, it is important to precisely identify human emotions from facial expression. Despite that, Facial Emotion Recognition (FER) remains a hard task because of the intricacies connected with subtile expressions and muscle variations. While conventional FER methodologies, which utilize generalized feature extraction, do not perform as well as they could in such adverse scenarios and suffer from a more limited range of accuracies in real world scenarios; such as varying degrees of lighting, angles, and facial expression, our approaches significantly improve over them. These shortcomings must be addressed, so that adaptive systems can provide humans with precision and reliability in how they respond to their emotions.

In terms of the construction of an intensity aware FER model, this research leverages Action Units (AUs), an insight from a physiology informed view of emotional expressions. The model identified subtle emotional variations by quantifying specific muscle movements (AUs) for expressions of different embodied posings and emotions. A study integrates AU intensities with machine learning classifiers into the py-feat library while using the Cohn Kanade AU Coded Facial Expression Database (CK+) to obtain high accuracy on six fundamental emotions. Unlike most existing methods, however, this AU based, intensity aware methodology improves the accuracy of FER without sacrificing interpretability, while still meeting the growing needs for dependable and interpretable models in dynamic settings. The results of this work also translate to real world sensitive to emotions technologies from mental health assessment tools to adaptive, interactive artificial intelligence systems.

1.3 PROBLEM STATEMENT

Facial Emotion Recognition (FER) systems often fail in accurately identifying facial expressions and are limited by the challenging task of subtle expression identification. The conventional FER models that commonly rely on such generalized features extraction techniques tend to miss the sensitivity needed for subtle variation in expression. However, this limitation limits their effectiveness in real world scenarios with a wide range of facial angles and lighting conditions and under small amounts of muscle movements. In fact, many of these models classify emotions without taking intensity of facial expressions into account, which leads to errors when classifying between closely related emotional states. To overcome these challenges this research adopts an Action Unit (AU) based methodology for FER along with intensity based features which graps all the vibrant and tender muscle movements related to six fundamental emotions. We aim to develop a model suitable for accurately detecting small emotional shifts, thus improving the accuracy and practicability of FER in extreme emotion recognition environments.

1.4 AIM AND OBJECTIVES

The objective of this research is to create a Facial Emotion Recognition (FER) model that is sensitive to intensity variations in emotional expressions, utilizing Action Units (AUs) alongside machine learning methodologies. This model aims to enhance the accuracy and interpretability of FER in practical contexts, such as human-computer interaction and mental health evaluations.

Objectives

1. To employ Action Units (AUs) for a comprehensive, physiologically informed examination of facial expressions, thereby increasing the sensitivity of the FER model to nuanced emotional variations.

2. To utilize the py-feat library for precise extraction of AU intensities, which are associated with six fundamental emotions: happiness, sadness, anger, disgust, surprise, and fear.

3. To preprocess and augment data from the Cohn-Kanade AU-Coded Facial Expression Database (CK+) to bolster the robustness and generalizability of the FER model.

4. To implement machine learning classifiers, particularly Random Forest, to effectively combine AU intensities for accurate emotion classification.

5. To assess the model's performance using accuracy metrics and classification reports, ensuring its reliability for application in emotion-sensitive technologies in real-world scenarios.

1.5 RESEARCH CHALLENGES

1. Data Quality and Diversity: A lack of diversity in datasets, such as variations in gender and ethnicity, can negatively impact the performance of models. To mitigate this issue, it is essential to gather a more diverse range of data and implement data augmentation strategies.

2. Annotation of AUs: The process of accurately labeling Action Units is both labor-intensive and susceptible to inaccuracies. The use of semi-automated tools and crowdsourcing methods can enhance the quality of annotations.

3. Imbalanced Data: An uneven representation of emotions within datasets can result in biased models. Approaches such as oversampling and the generation of synthetic data can help to rectify this imbalance.

4. Emotion Intensity Variations: The detection of subtle differences in emotional intensity presents significant challenges. Employing intensity-aware models and utilizing continuous emotion scales can provide assistance in this area.

5. Real-time Performance: Achieving a balance between accuracy and speed for applications requiring real-time processing is a complex task. Optimization techniques, including model pruning and edge computing, can enhance overall performance.

6. Generalization: Facial Emotion Recognition (FER) models may encounter difficulties in generalizing across a wide range of subjects and environments. Strategies such as transfer learning and domain adaptation can improve their generalization capabilities.

7. Multimodal Fusion: The integration of facial expressions with other modalities, such as speech, adds to the complexity of emotion detection. Utilizing multimodal fusion techniques can enhance the accuracy of emotion recognition.

8. Contextual Understanding: FER systems may struggle to grasp contextual subtleties, which can compromise their accuracy. The incorporation of contextual information is vital for improving model performance.

CHAPTER 2

BACK GROUND

2.1 BACKGROUND AND LITERATURE SURVEY

The problem of emotion recognition from facial expressions has found application in many areas including human to computer interaction, autonomous vehicles, online education platforms, as well as in healthcare services. It is necessary to understand what human emotions are for improving safety, communication, and reactive actions in different context. Methodologies for the challenges present in facial emotion recognition (FER) can be summarized from the literature.

Previously, most of the research was focused on handcoded features and traditional machine learning algorithms. About the Deep learning methods are increasingly accepted in the contemporary studies as it performs the extrapolation of fine features from the raw data. Recent developments in techniques, such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks and hybrid models, have demonstrated considerable improvements in terms of both accuracy and robustness, and when scaled up to utilize, for example, multimodal data sources such as vocal cues or body language.

However, there are still a lot of challenges in the field of emotion recognition. Despite efforts to increase training time and volume, variability in facial expressions – resulting from technical factors such as lighting conditions, viewing angles, and differences among individuals in the expression of emotion – continues to hamper performance. Furthermore, annotations in datasets are ambiguous and thus the optimization of such a model is complicated by a subjective nature. To address these issues, recent research endeavors have included those with uncertainty adaptive learning strategies or overcoming issue biases on feature and label spaces.

For example, studies using RAFDB, FERplus and KDEF datasets achieved remarkable accuracy, however there are a lot of obstacles like real time application, generalization over diversified domains and noisy data management. In addition, real time emotion recognition systems continue to deal with the issues of complexity of models and the related computational costs that accompany it. As a result, although a significant amount has been accomplished, the performance of emotion recognition technologies can be greatly improved, in terms of accuracy, flexibility, and efficiency.

2.2 OVERVIEW OF ACTION UNITS (AU) AND EMOTIONS

According to the Facial Action Coding System (FACS) constructed by Ekman and Friesen, the emotional expression is signified by the movements of one of Action Units (AUs), which is a specific muscle movement. The system is able to recognize 46 different AUs that reflect specific muscle actions, for example raising the eyebrows or smiling. If one examines the activation patterns of these AUs when facing facial expressions, one can identify the emotions like happiness, anger and surprise. For example the use of AUs 6 (cheek raiser) and AUs 12 (lip corner puller) represents happiness, while AUs 4 (brow lowerer) and 5 (upper lid raiser) represent anger. Recently, machine learning and deep learning algorithms, which are commonly utilized to perform AUs detection from facial image automatically, are displayed in the existing literature to achieve a highly accurate emotion classification across several datasets. However, barriers to the precise interpretation of facial expressions in a variety of real world contexts, include variations in lighting, individual differences, and occlusions.

2.3 RELATED WORKS IN FACIAL EMOTION RECOGNITION

Facial emotion recognition (FER) has experienced tremendous progress from conventional approaches that heavily rely on manually engineered features to modern deep learning solutions. Feature extraction based on LBP and Gabor features is used, with SVM classifier. The reliance on manual feature engineering in these early approaches introduces constraints, as well as variability in lighting, pose, and occlusions. The advent of Convolutional Neural Networks (CNNs) brought the modern times in the FER domain and was able to extract hierarchical features from unprocessed facial image, thus increasing performance dramatically improving FER performance on the datasets such as FER-2013 and CK+.

Moreover, the development of the Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks has experienced some momentum for motion capture of temporal dynamics in video data. Now, recent investigations also pursued multimodal emotion recognition by combining facial expressions with auditory or physiological signals to increase accuracy, especially in difficult scenarios. However, many difficulties remain, including the need for generality over a wide splendour of data, dealing with the composite multiple tag emotions, as well as varying both in age, gender in ethnicity. Now, techniques are being cut edge by using action units (AUs) to do more accurate emotion detection by analyzing certain facial muscle movements — for a more nuanced and accurate emotion detection.

2.4 THEORETICAL FRAMEWORK

Facial Emotion Recognition (FER) is the theoretical framework that combines psychological theories of emotion with computer vision methodologies and state of the art machine learning algorithms. Facial expressions are often interpreted as emotion and Paul Ekman lists six basic emotions: happiness, sadness, anger, fear, surprise, and disgust and indicates there are facial movement muscles Action Units (AUs) corresponding to each of these perception.

The Facial Action Coding System (FACS), invented by Ekman and Friesen, is essential to recognizing these AUs and connecting them up with specific emotional states. From a computation viewpoint, FER systems tend to feature extraction techniques such as texture analysis, edge detection and deep learning to recognize and classify emotions. This has led to huge improvements in the reliability of FER using different CNNs to extract hierarchical features from raw facial images. Most recently, Recurrent Neural Networks (RNNs in particular Long Short Term Memory (LSTM) RNNs), which have shown expertise in processing temporal dynamics in facial expression in the video data, saw recent development.

Complementing emotion recognition with deep learning strategies that can analyze facial muscle movements and their emotional correlations, action unit based models are integrated with the deep learning strategies for further refinement of emotion recognition. A theoretical framework for FER is developed by merging the psychological underpinnings of emotion with cutting edge machine learning methodologies to build systems that detect emotions in real time under a wide array of circumstances.

2.5 LIMITATIONS OF EXISTING APPROACHES

Although many methodologies exist for facial emotion recognition (FER), they impose many constraints that negatively impact their sensitivity and applicability in real world applications. Variability of facial expressions due to lighting, perspective and age, weakens one of the biggest obstacles. The models are complicated by these discrepancies, since generalizing across such different settings is unreliable, making the models less reliable.

Additionally, occlusions from accessories (including glasses, facial hair, or even masks) can obstruct important facial features necessary to identify emotions. An additional major limitation is the lack of sufficiently large, diverse data sets which span a wide range of emotion, face assortments, and cultural subtexts. FER systems primarily based on constrained datasets are limited to biases or suboptimal performance when confronted with novel, unseen data. In addition, the problem of cross database generalization remains, as generally models trained on one dataset will fail to generalize to other datasets caused by these different data collection methodologies.

Recognition of subtle emotions and of the intensity of emotions, which are often shared through micro expressions or weak facial movements hard to capture, are usually a challenge to FER systems. Another hurdle for practical implementation is the real time processing of video sequences and video sequences in particular, has high computational requirements on deep learning models. Also, the subjective part of annotations in the dataset can result in inconsistency and ambiguity in emotion annotate, further deteriorating the model's capacity in accurately annotating emotion.

CHAPTER 3

METHODOLOGY

3.1 PROPOSED SYSTEM ARCHITECTURE

The proposed architecture for facial emotion recognition is structured around several key stages: This achieved preprocessing, feature extraction, model comparison and evaluation towards effectively classifying emotions from AU intensity. First (Stage 1), the data is processed/standardize and augmented to help ensure robust generalization through Preprocessing Augmentation (PPA).

Next is intensity scaling then creation of matrices to organize AU features for subsequent analysis. During the AU Extraction phase (Stage 2), we derive specific AU intensities that capture specific facial muscle movements. Scaling and Standardization or Stage 3 of the process includes all features being standardized to ensure uniformity across data points to boost the latitude of the model. In the Comparison Models stage (Stage 4), different classification algorithms, i.e. Logistic Regression, KNN, Naive Bayes and CatBoost that take the AU features as their input, are trained. However, Random Forest is noted as the best model due to its higher interpretability and ability to handle non linear AUs relationships. In Stage 5, Evaluation and Validation, performance metrics such as accuracy, precision, recall, F1 score, etc. are used to evaluate the model performance.

Overall, Stage 6 finally corroborates the best perceiving model of emotion labels, for instance, of "Happiness", given AU intensities, as Random Forest. The design of this architecture results in an interpretable and efficient solution that achieves robustness for real world applications in emotion recognition, with minimum compromise on classification accuracy.

The modular design of it makes that it is easy to change their dataset as well as twist them for different use cases. They both offer contributions to how to understand the importance of AU based features allowing for overall interpretability of the system. A dependable and explainable facial emotion recognition framework with real-time as well as academic research potential is thus achieved.

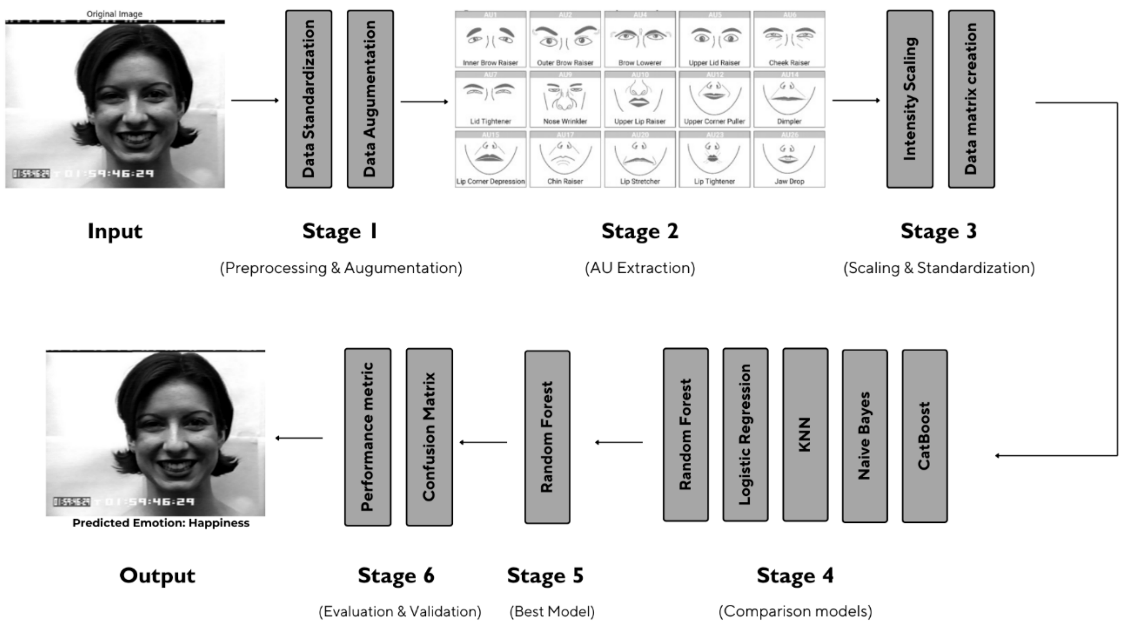

Fig 3.1: Architecture Diagram

3.2 DATA COLLECTION AND DATASET DESCRIPTION

In this study, the dataset used is named CK, which contains [facial expression data], and it serves as the main aim towards enhancing the precision of facial analysis models. In this case, the process of collecting data collected facial images or videos from a heterogeneous set of participants which was diverse on the one hand on the age, gender and ethnicity. The dataset contains each facial expression with both the expression labeled (categorized to different emotional classes like happiness, sadness, anger and surprise) and facial landmarks. The dataset consists of each entry containing [describe features, such as pixel data, landmark coordinates or intensity measures] associated with [target variable, e.g., expression categories].

To make the data uniform, uniform, and to facilitate computational efficiency in subsequent analysis, comprehensive preprocessing procedures of image such as normalization, alignment and resizing were applied. This is to encourage balanced dataset, so that it reduces bias and allow robust insight formation during the training process. Each facial action unit is annotated, which allow the model to create very accurate forecasts of intricate facial movements, making this a perfect dataset for applications in human computer interaction, sentiment analysis, and facial recognition technologies.

3.3 ACTION UNIT FEATURE EXTRACTION

In the examination of facial expressions in the Cohn-Kanade (CK+) dataset it was found that the extraction of Action Unit (AU) features was a key component. We systematically analyzed each image using a pretrained AU detection model in order to specifically identify and quantify specific movements of facial muscles, AUs, indicating specific emotions e.g. happiness, sadness, anger and surprise. Face AU (AU stands for au of my language) encompasses each unique facial movement, for example the elevation of eyebrows or the tightening lips, needed for reading subtle emotional signals.

After the input of each image in the AU detector, facial regions and AU intensities were identified across twenty different muscle groups. Each time an image was processed, the detector would output a set of AU values – how active the facial muscles were. Each image along with its designated emotional class label was organize in a structured DataFrame format where each row contains AU data for an image.

We went on to extract the AU features as we did here and to visualize the extracted AU features to facilitate further analysis, and then showing the intensity of each AU and correlating them with the related emotion labels. This dataset was organized and provided a basis for a quantitative foundation of emotional analysis, which subsequently proved beneficial to emotion recognition and facial behaviour analysis in the CK+ dataset.

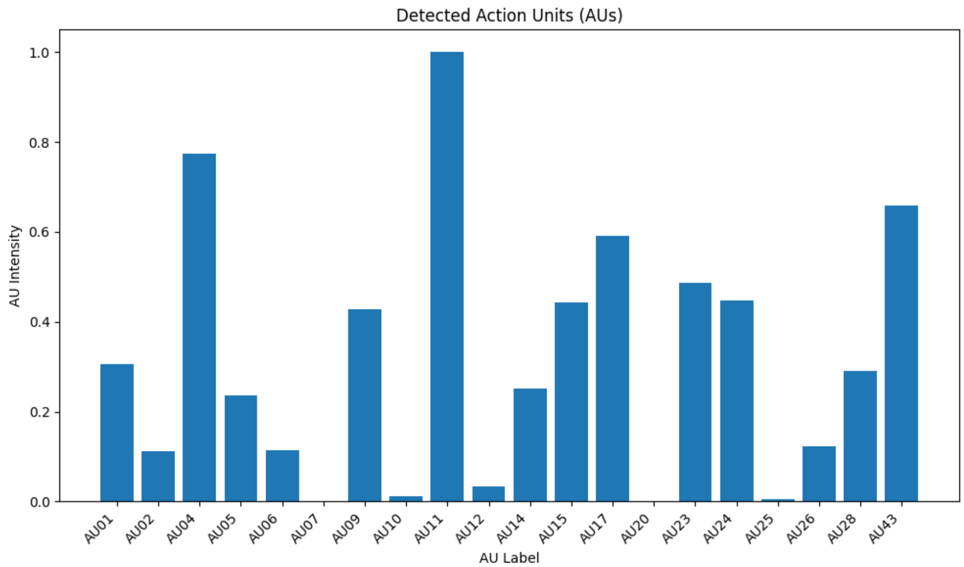

Fig 3.2: Intensity of detected Facial Action Units

3.4 FEATURE SELECTION AND PREPROCESSING

Highly rigorous feature selection and processing were required to prepare Action Unit (AU) data for precise analysis as well as training of the model. Since AU features are high dimensional, and some specific selection and preprocessing techniques were used to reduce noise and select most informative features to enhance the effectiveness of the model.

Feature Selection: A dataset was created, where each image contained 20 AU features, one for each unique facial muscle movement. We first emphasized AUs that are most relevant to emotional expression, which improved interpretability by focusing on the important information. In particular, priority was given to AUs related to eyebrow movements (AU01, AU02), eye squinting (AU07), and mouth movements (AU12, AU25), as these AUs are highly related to fundamental emotions. Through this focused selection approach we were able to collect effective expressions while minimizing noisy data.

Data Cleaning and Labeling: Standardization of the representation of the "neutral" emotion was maintained to maintain consistency in labeling. We systematically dealt with missing values to maintain data integrity and end up with a more cohesive and trustworthy dataset.

Normalization: First, AU intensities were normalized to a standard rage (0 and 1 typically). This normalization process ensured fairleaching between all features within the model and there would be no AU in particular that undueley dominated the model because of the variation in the intensity.

Feature Transformation: We also looked at additional transformations, including sums and differences between AU pairs, to capture more complex expressions and more subtle variations. By enriching the model with these derived features it increased the model's ability to identify subtle emotional expressions, since it provided a richer representation of facial movements.

With this systematic approach to feature selection and processing, we have constructed a refined, relevant and consistent dataset. Now, the processed AU features can effectively turn the emotional spectrum in the CK+ dataset into the form that facilitates model trainings and facial expression recognition analyses.

3.5 MODEL TRAINING AND ALGORITHMS

The set of stages involved in the training of the model to predict the emotions based on the analysis of facial Action Units (AUs) are several. The dataset, which consists of 393 images together with their associated AU values and to emotion labels, was first processed. I cleaned the labels, rectified for any inconsistency, and there was no missing data. We then classified the emotion labels as the target variable and identified

20 AUs as features subsequently.

The training and testing dataset was then split into 80% training and 20% testing subset. The model was trained on the training subset with the help of a Random Forest Classifier which is a well known classifier in the machine learning applications. This classifier was chosen for its performance in dealing with high dimensional datasets and can make sense of the relationship between features and target variable. After the training, the test subset was used to assess the model's performance and we compute accuracy, and we generate a classification report giving the precision, recall and F1 scores. The model has performed as good as it could have, setting a solid baseline model for predicting emotions from facial expressions.

Incorporation of deep learning methodologies as well as the prospect of hyperparameter optimization, are possible prospective enhancements towards further improvement of the efficacy of the model. Overall, the developed model is able to predict emotions from facial imagery, by analyzing the intensity of Action Units present.

Facial Action Units were used as the indicator of emotions to be predicted and the predictors were many machine learning algorithms. To start with, Multinomial Logistic Regression as it is, and its adjusted version, are particularly appropriate for multiclass classification tasks, and we initially used these. These models predict a probability of each emotion category and are easy to interpret. We next exploited the nonparametric technique of K Nearest Neighbors (KNN), operating as a classifier, in which the majority class of neighboring data points is what is considered. Decision Trees were also implemented and had obvious models that use partitions by decision thresholds to segment the data, but with fixes to prevent over fitting. To improve the accuracy even further, we opted for Random Forest, an ensemble technique that constructs consensus predictions based on multiple decision trees, or Gradient Boosting methods like XGBoost, LightGBM and CatBoost that iteratively increase models' efficiency by addressing mistakes made by their previous iteration(s). High dimensional datasets were also treated with Support Vector Machines (SVM), which showed successful separation of classes based on optimal separations. To take a shorter route, Naive Bayes was tried, but under the assumption that input features are independent, which works well on smaller datasets. Finally we looked at Deep Learning Models, taking advantage of neural networks' ability to see intricacies, even where the human brain would fail. The critical performance metrics used to assess each algorithm were accuracy, precision, recall, and F1 score and optimization strategies were employed to increase the efficacy of emotion prediction by each algorithm.

3.6 EVALUATION METRICS

Evaluation metrics play a vital role in evaluating the efficacy of various models employed for emotion prediction based on facial Action Units (AUs). The primary metrics applied in this assessment include accuracy, precision, recall, F1-score, and the confusion matrix.

1. Accuracy: This metric serves as the fundamental measure, indicating the ratio of correct predictions made by the model relative to the total number of predictions. While it offers a broad perspective on the model's performance, it may not provide an accurate representation in cases of imbalanced datasets.

2. Precision: Precision quantifies the correctness of positive predictions, specifically the ratio of true positives to all instances classified as positive. This metric is particularly significant when the repercussions of false positives are substantial.

3. Recall: Also referred to as sensitivity, recall assesses the proportion of actual positives that the model successfully identifies. This metric is essential when the implications of false negatives are more severe, such as in the detection of specific emotions.

CHAPTER 4

IMPLEMENTATION

4.1 SOFTWARE AND LIBRARY REQUIREMENTS

Specific software and library requirements are required to manage, extract features, and to classify the emotions from the facial data in the CK+ dataset during the Facial Emotion Recognition implementation. Image processing and machine learning tasks are two typical problems that could be solved using Python, which is the primary language used for it, thanks to it's flexibility, user friendliness and its complete library ecosystem from which it literally has everything.

`py-feat` is one of the key of the library that detects facial Action Units (AUs) and recognizes emotions. Although extracting features from facial images is essential for accurate AU detection and therefore suitable understanding of facial expressions, this library simplifies the process. A great part of the data processing tasks requires `numpy` library for managing its multidimensional arrays and making numerical computations on them. Furthermore, `Pandas` is used to increase the efficiency of the dataset management process by storing, and being able to organize, the extracted AUs into a structured format.

In data visualization we use `matplotlib` and also `seaborn` to plot AU and perform EDA to understand how AUs is distributed and trends across different emotions. Funtions such as image loading, resizing and preliminary preprocessing are managed by 'opencv-python' and Python Imaging Library (PIL) on image data.

Second, `scikit-learn` provides a tool suite for making, training, and assessing machine learning models for classification; and it provides a suite of diverse evaluation metrics for assessing, not only the models, but also their parameters, cross validation, and regularization terms. Together, these libraries give a robust framework for data processing, AU extraction and model evaluation making those the backbone of this project’s work.

4.2 DATA LOADING AND INITIAL PREPROCESSING

Proper preparing the CK+ dataset for Facial Emotion Recognition is very essential, and that’s why comes with the importance of the Data Loading & Initial Preprocessing phase. This is the beginning of this phase where the importation of the CK+ dataset, which contains facial images with all sorts of annotations including the emotions, begins. Reading these images into the proper format along with using libraries such as opencv-python or PIL makes the loading process very readable.

Standard image processing operations used in initial preprocessing and aimed at making all images uniform and quality of feature extraction. Before running training, we first resize each image to a fixed resolution which allows for a uniform input size for processing. It then normalizes pixel values to a common scale to help minimize pixel variability due to light conditions or image quality. Additionally, if the detection model doesn’t require the color information, we convert it to grayscale to streamline the data and preserve most of the facial structures.

But this may also involve getting rid of inappropriate or low grade images in order to have a nice standard for analysis. The database, after the initial preprocessing, is rendered uniform and optimized so that it is ready for Action Unit (AU) feature extraction thereby creating a strong base for reliable and precise emotion detection in the subsequent stages.

Dataset	Source	Attributes
URL	Kaggle	640x480 Pixels, 393 High resolution images

Table 1: Dataset Description

4.3 IMAGE PREPROCESSING AND AUGUMENTATION

Image Preprocessing and Augmentation is an important stage which helps enhance the CK+ dataset to better help enable more accurate facial emotion recognition. Image preprocessing, specifically maximizing quality of the feature extraction and constancy of the dataset, occurs in preprocessing phase. For example, when we use a machine learning model or a facial detection system, we need to resize every image to the same resolution first, most especially. Moreover, converting images to grayscale can be used to put emphasis on intensity based features that aid in noise reduction while maintaining key facial information. And finally, augmentation techniques are introduced for enlarging artificially the dataset and introducing diversity when the data is limited or unbalanced.

Most common augmentation methods, for example, include random rotations, horizontal flips and small scaling, which enhance the dataset's diversity without changing the basic facial expressions. The usefulness of these augmentation strategies to make the model robust is in being invariant to small deviations in image orientation, scale and lighting, which improves its generalization over real world situations. This stage then significantly improves the accuracy and tolerance of facial emotion recognition by addressing the situation through augmentation of the diversity in training data

4.4 FEATURE EXTRACTION AND AU ANALYSIS

Facial Emotion Recognition is based on setup of Feature Extraction and Action Unit (AU) interpretation, which is summarized as the determination and classification of facial Action Units capable to refer to different emotions. The preprocessed images from the CK+ dataset are then subjected to automatic detection of Action Units in them, using the py feat library, specifically designed for sophisticated facial analysis. This methodology provides fine grained ways to represent emotions based on changes in the facial musculature and is important because these changes in facial musculature like displaying the elevation of eyebrows or tightening of lips provide clues for what emotions are being communicated.

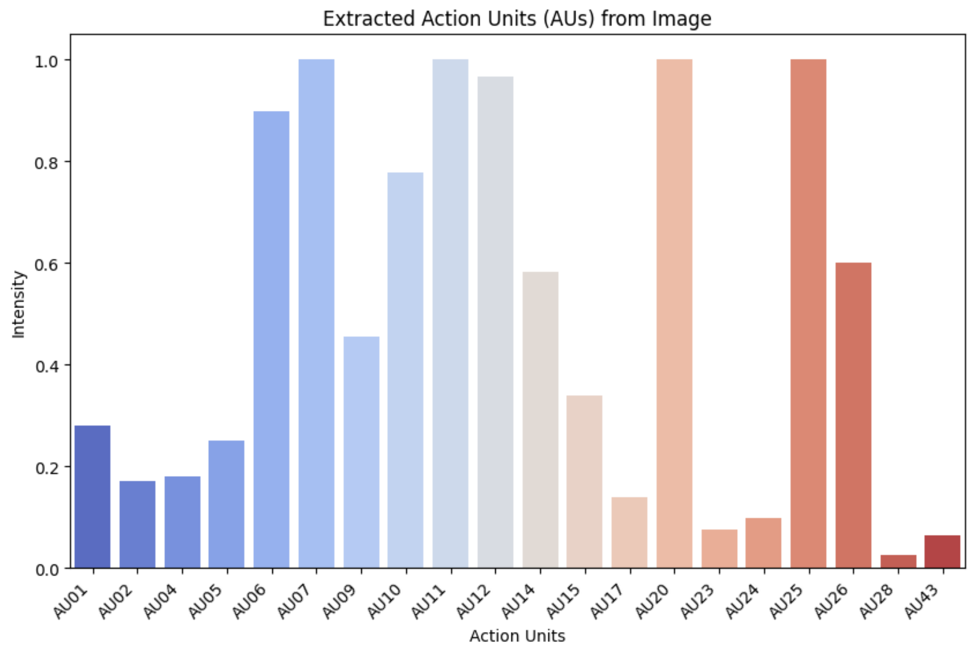

Fig 4.1: Visualization of Extracted Action Unit (AU) Itensities

To start, the py-feat detector is initialized, going through each facial image and painstakingly identifying the Action Units it contains. By means of specific muscle movements, representing AUs (action units), their quantifiable depiction is extracted as a numerical data and each facial expression is subsequently described. Once extracted, AUs are organized into a structured dataset to study the occurrence and intensity of the AUs in diverse emotional contexts. After extraction phase, AU is used subsequently to identify the patterns in the data. Having investigated the relationships between specific AUs and certain emotions (for example, AU12 for happiness or AU4 for sadness), techniques from exploratory data analysis (EDA) are employed. These correlations are visualised through bar plots and heatmaps showing these unique AU profiles associated with various emotions. This phase is very vital in making effective classification of emotions because of its reliance on very exact facial cues.

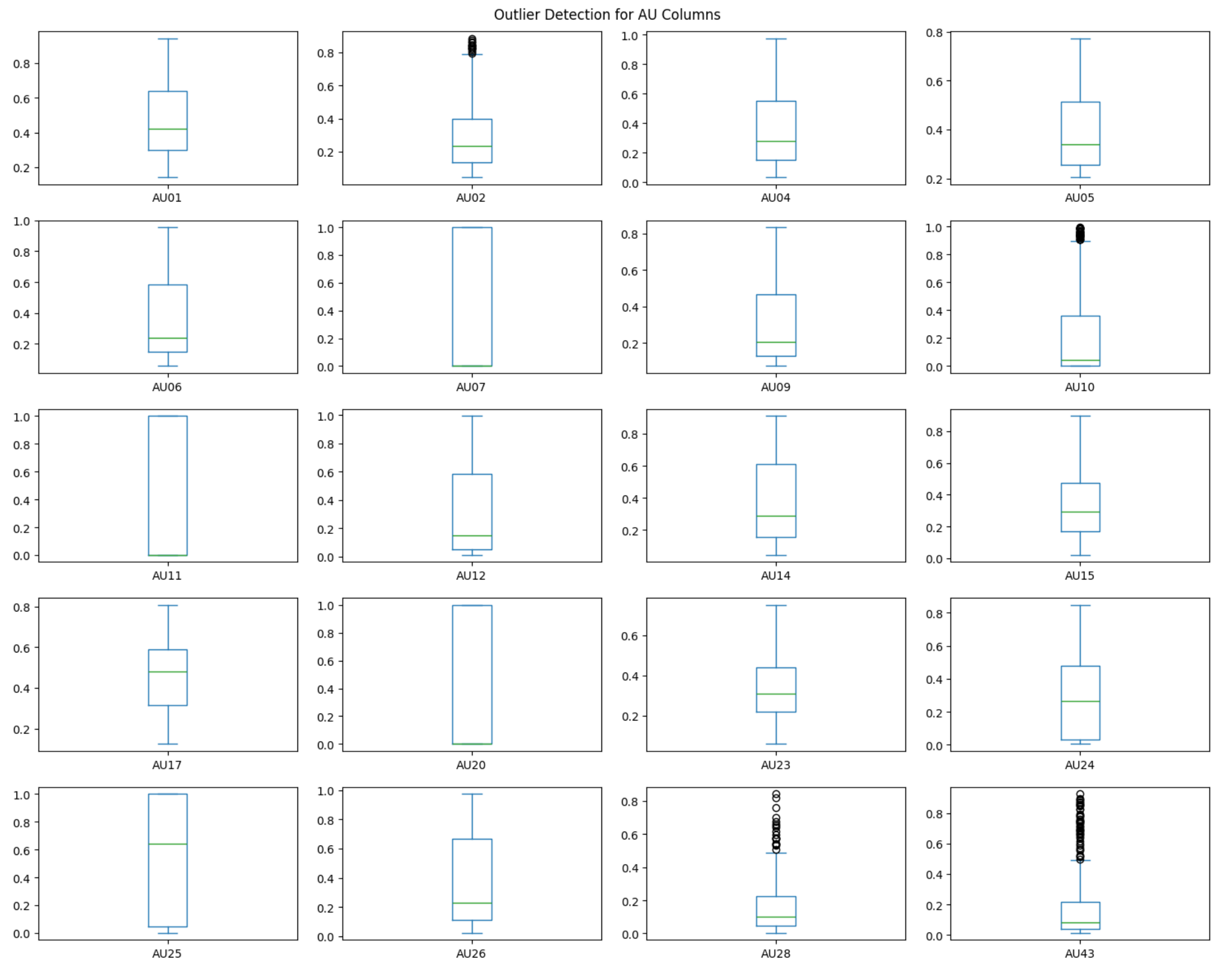

Fig 4.2: Boxplot for Outlier Detection in Action Units (AUs)

4.5 FACIAL EMOTION ANALYSIS AND VISUALIZATION

In the phase of Facial Emotion Analysis and Visualization, we interpret AUs extracted from facial images to establish and understand emotional expressions. At this stage, we use a structured dataset of AUs which allows us to find distinct patterns of AUs for several emotions, such as happiness, sadness, anger and surprise. The facial movements corresponding to different emotional states are defined in a specific combination of activated AUs, each emotion is characterized.

To help with this analytical tasks, we use a set of visualizations like bar plots, histograms and heatmaps with libraries such as matplotlib and seaborn. Say, bar plots could show how often or how intensely certain AUs occur while looking at various emotions, or heat maps could display that multiple AUs are dependent upon other AUs or emotions. With these visual tools it is easy to get a good sense of what Action Units are most salient for which emotion. For example, AU12 (lip corner puller) is often used in expressions of happiness, whereas AU4 (brow lowerer) is usually used in expressions of anger and sadness.

This phase further includes a visualization of aggregated dataset to look for the patterns or clusters of the emotions. By mapping AUs to emotion states, it further deepens our knowledge of this relation between facial expressions and emotions that improves classification accuracy and efficacy in the model. The consequence of this thorough visual analysis of AUs not only enhances interpretability of the results, but provides groundwork for the validation of the importance of AUs in the correct recognition of emotions.

4.6 MODEL BUILDING AND EVALUATION

Model Building and Evaluation focuses primarily on building a classification model for emotion detection through a use of Facial Image; AUs. We establish a correlation of AU patterns and specific emotions by applying machine learning approaches on the CK+ dataset. For such large datasets with a proper preprocessing, advanced deep learning architectures, such as Convolutional Neural Networks (CNNs), or Random Forests, or Support Vector Machine (SVM) could be prominent algorithms for this purpose. The reason for these models' selection is based on their capability to handle complex data pattern and distinguish non linear association of AUs and emotional states.

The model construction process starts by splitting the dataset into training and testing subsets in order to get unbiased evalution. Model efficacy is enhanced with hyperparameter optimization methods, including cross-validation to fine tune parameters and reduce risks of overfitting. In the training phase, the model is assessed through metrics such as accuracy, precision, recall and F1 score to make a thorough evaluation on how well the model can predict things.

Also, performance is analyzed in detail using confusion matrices and ROC curves, which shed some light on the potential of misclassification of particular emotions. The result of this extensive evaluation process allows us to objectively determine how reliable the model is to be used in actual cases, to verify that it can correctly and consistently determine what emotions correspond to the AUs. Using iterative refinement, the model is tuned to high accuracy, and thereby serves as an efficient tool for facial emotion recognition.

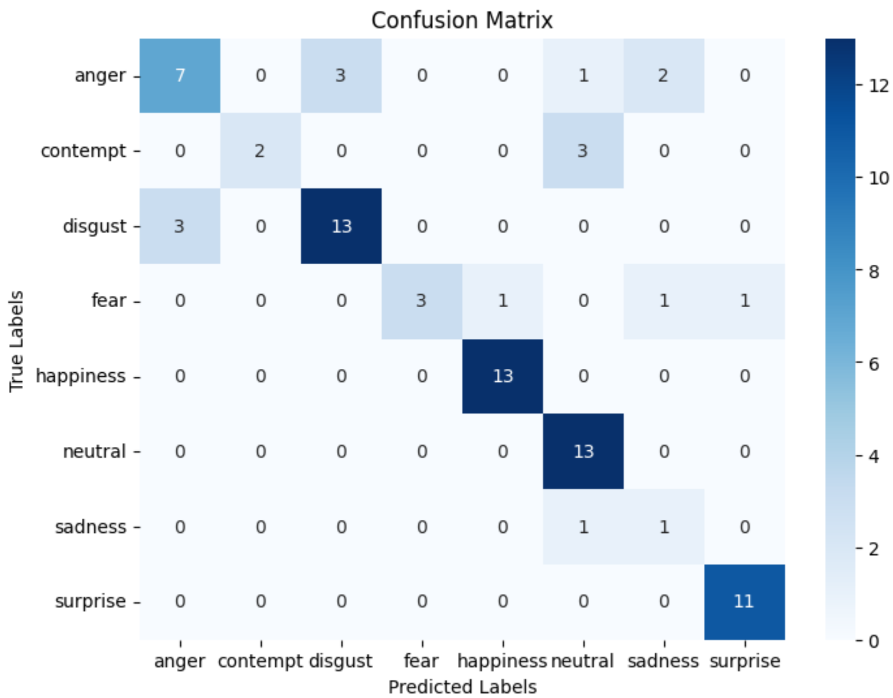

Fig 4.3: Confusion Matrix for Logistic Regression Model

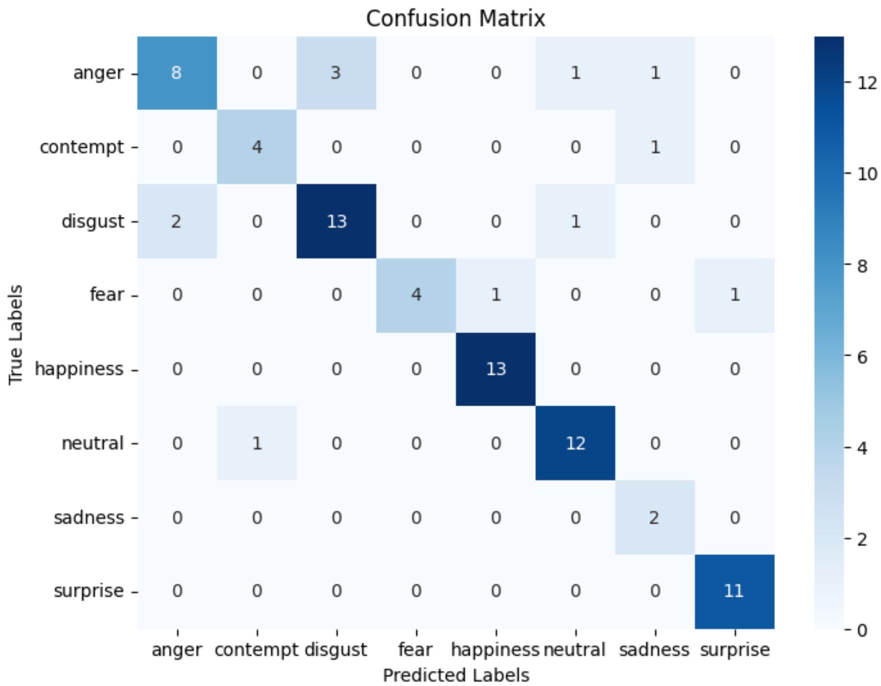

Fig 4.4: Confusion Matrix for Best Logistic Regression Model after Hyperparameter Tuning

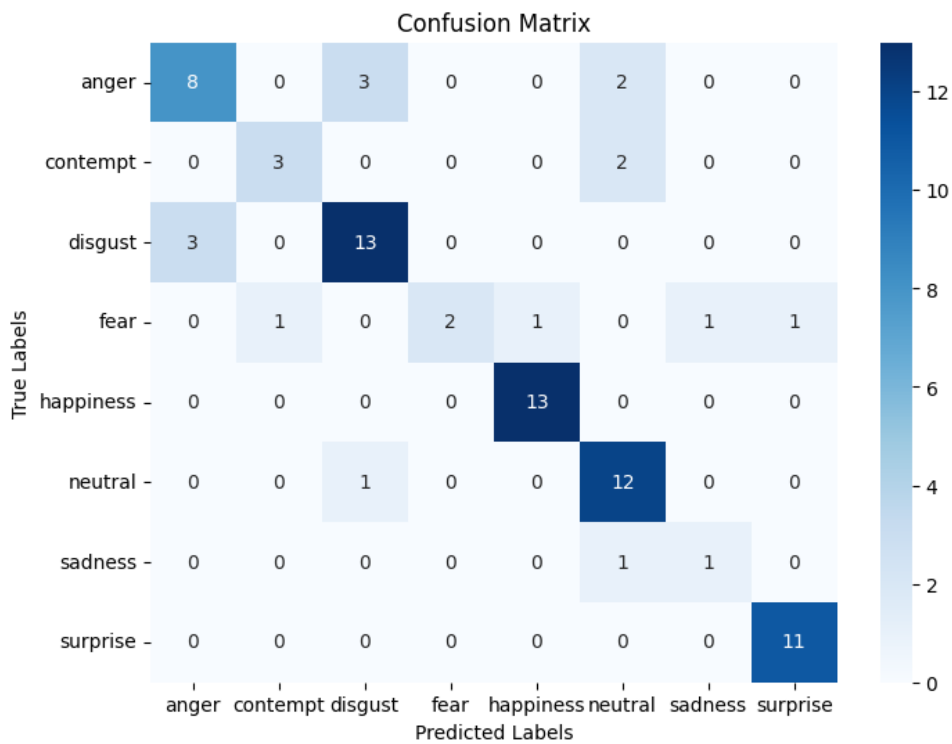

Fig 4.5: Confusion Matrix for K-Nearest Neighbors (KNN) Classifier

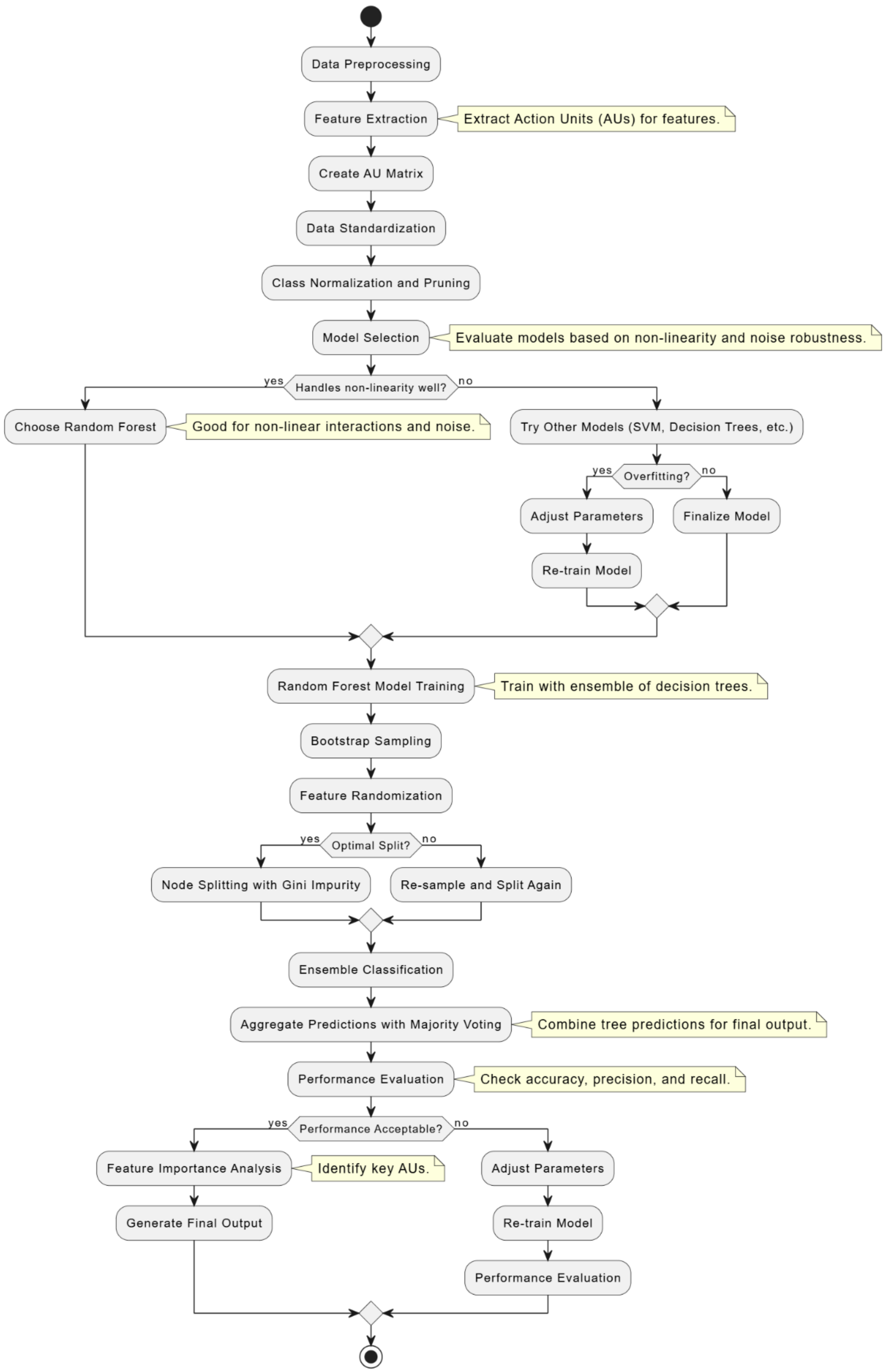

Fig 4.6: Flowchart of our system

CHAPTER 5

RESULTS AND ANALYSIS

5.1 RANDOM FOREST CLASSIFIER

This study has shown impressive results with the Random Forest classifier algorithms in recognizing facial emotions with an overall accuracy of 88.61%. The Random Forest classifier provides randomization of features at every node split, which helps the model to understand and learn how different action units (AUs) are activated for different emotions as in AU12 (Smile) and AU04 (Frown). The model used for the recognition of different emotions exhibited almost perfect classification accuracy scores for savoring and amazement in terms of high true positive rates and low false positive rates. On the other hand, it encountered difficulties in recognition of anger and sadness tendencies which were more complicated, owing to certain AUs similar to their emotions causing a few of them to be incorrectly classified, moderate to severe frustration. Although such drawbacks exist, the ability of the Random Forest classifier to handle disruptive noise and its interpretation of importance to features makes it ideal for application in situations that call for very high accuracy and interpretability, making it a perfect choice for any system requiring emotion detection accuracy as well as reliability.

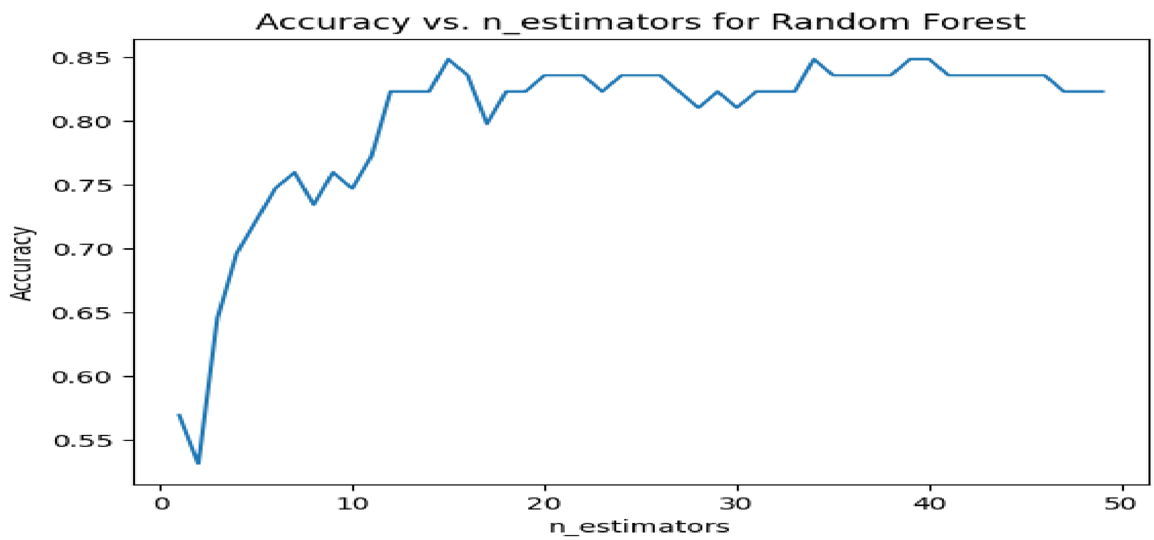

Fig 5.1: Accuracy vs. n_estimators for Random Forest

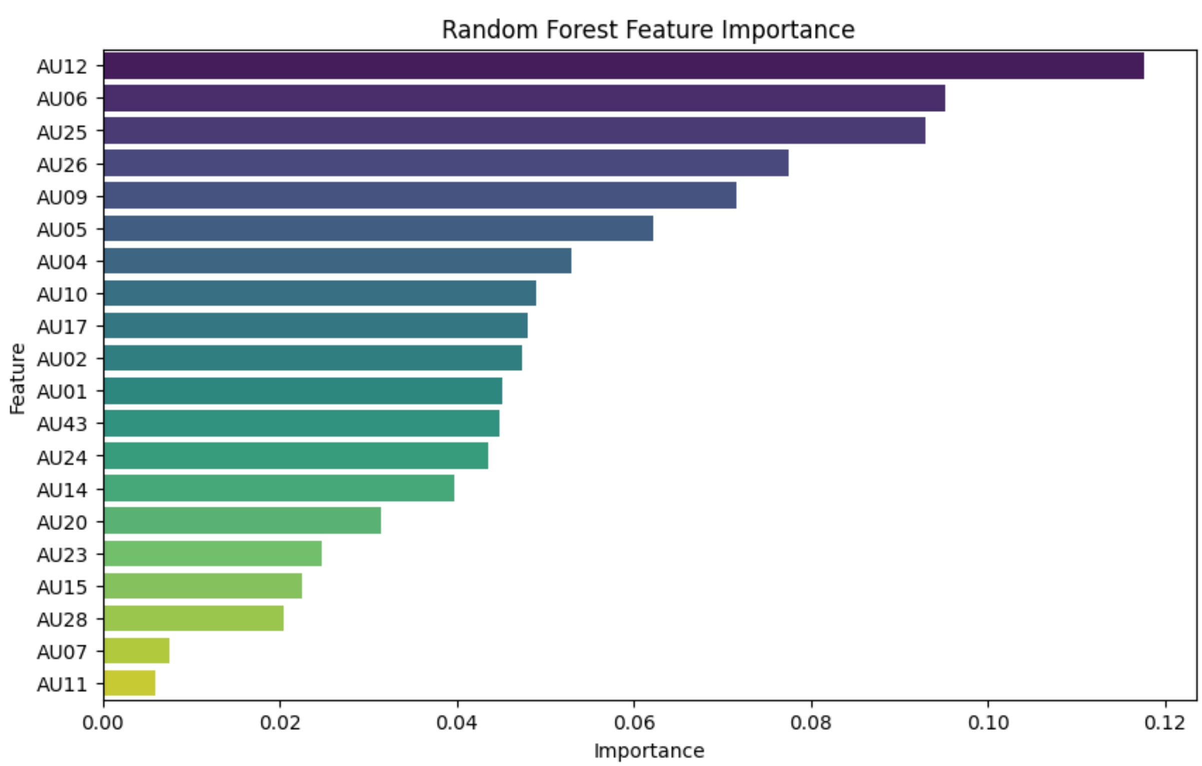

Fig 5.2: Random forest feature importance

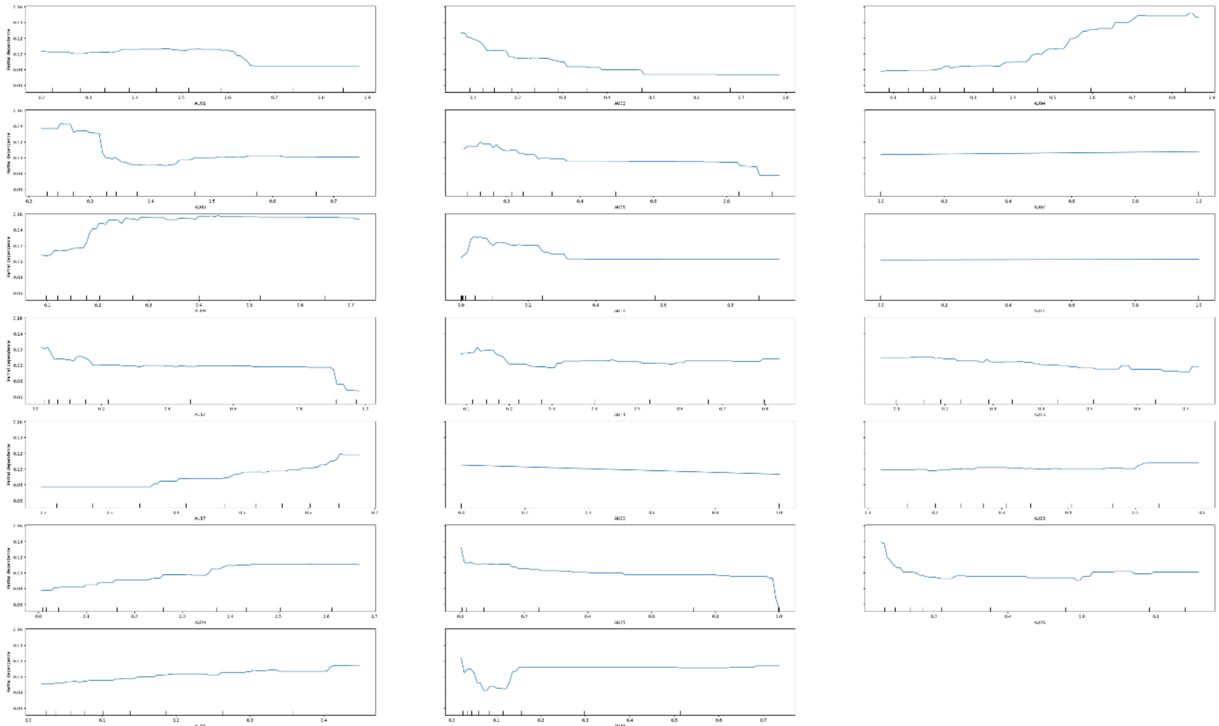

Fig 5.3: Partial Dependence Plot for Random Forest Model Features

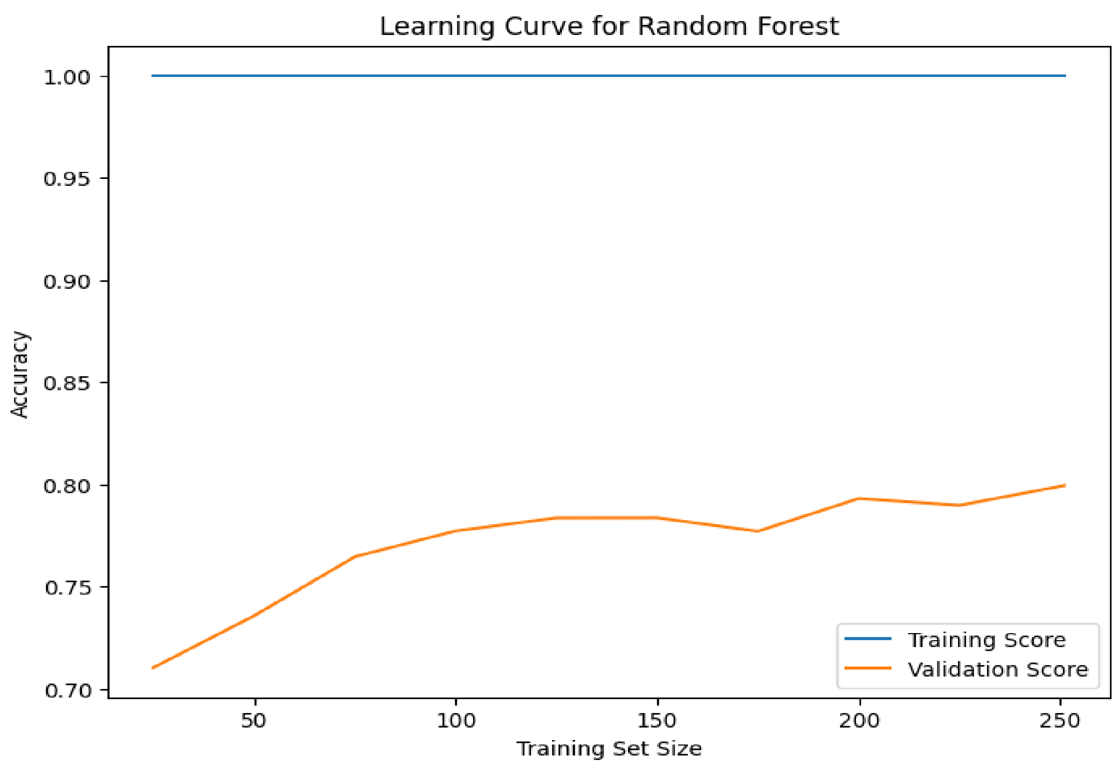

Fig 5.4: Learning curve for Random Forest

```python
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
model = RandomForestClassifier(n_estimators=31, random_state=53)

model.fit(x_train, y_train)

y_pred = model.predict(x_test)

accuracy = accuracy_score(y_test, y_pred)
report = classification_report(y_test, y_pred)
conf_matrix = confusion_matrix(y_test, y_pred)

print(f"Accuracy: {accuracy * 100:.2f}%")
```

```
Accuracy: 88.61%
```

Fig 5.5: Random forest classifier Accuracy

```python
from sklearn.metrics import confusion_matrix, classification_report

y_pred = model.predict(x_test)
cm = confusion_matrix(y_test, y_pred)

plt.figure(figsize=(8, 6))
sns.heatmap(cm, annot=True, fmt='d', cmap="Blues", cbar=False)
plt.title("Random Forest Confusion Matrix")
plt.xlabel("Predicted Label")
plt.ylabel("True Label")
plt.show()

print(classification_report(y_test, y_pred))
```

Fig 5.6: Random Forest Classifier Model

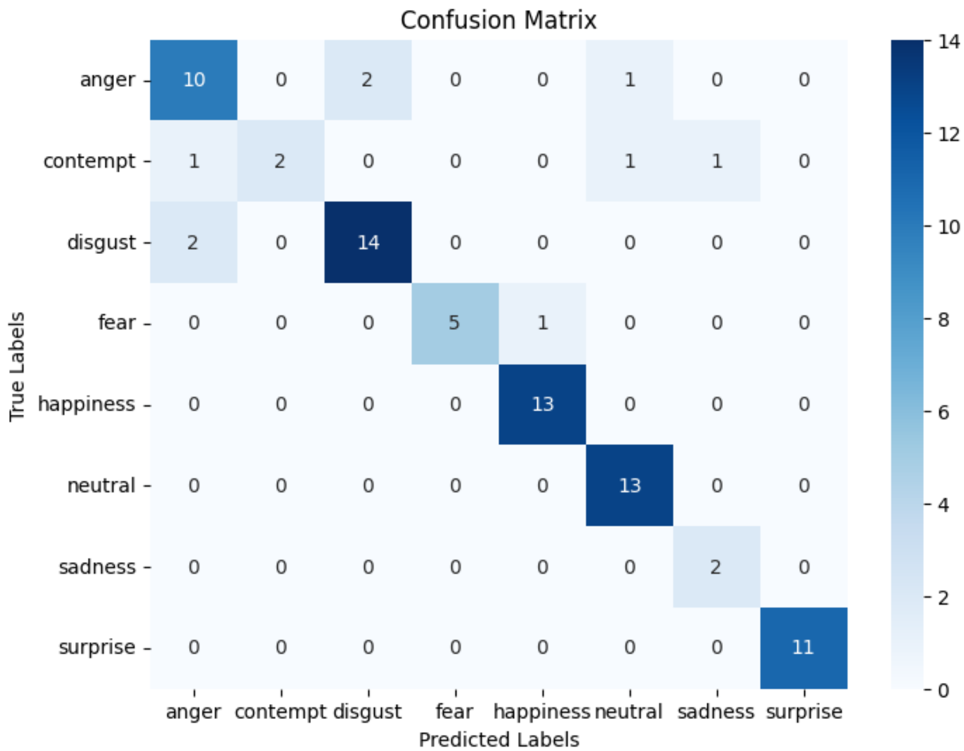

Fig 5.7: Confusion Matrix for Random Forest Classifier

	precision	recall	f1-score	support
anger	0.77	0.77	0.77	13
contempt	1.00	0.40	0.57	5
disgust	0.88	0.88	0.88	16
fear	1.00	0.83	0.91	6
happiness	0.93	1.00	0.96	13
neutral	0.87	1.00	0.93	13
sadness	0.67	1.00	0.80	2
surprise	1.00	1.00	1.00	11
accuracy			0.89	79
macro avg	0.89	0.86	0.85	79
weighted avg	0.89	0.89	0.88	79

Fig 5.8: Classification report for Random Forest Classifier

5.2 LOGISTIC REGRESSION

The task of facial emotion recognition was approached with a Logistic Regression classifier as the base model before comparison with advanced models. As a straight-line classifier, Logistic Regression seeks to explain emotional states from the weighted sum of some set of Action Unit (AU) features such as AU06 and AU12 which are closely related to happiness expression. Sufficient results were obtained through the use of such a model in the classification of basic emotions such as happiness and surprise because of the high accuracy associated with the model's simple and easy-to-understand design. However, it struggled when tasked with the classification of other emotions that were closely related like sadness and anger as there were also very few differences in the levels of AUs intensity. While Logistic Regression is quite basic, it helped determine how important each feature was, in particular, which AUs contributed significantly towards each emotion. This simplicity and clarity make it suitable, but it is unable to handle complex aspects of emotions, which is the case with changed feelings. Though the performance of the model could not compare with the performance of combined approaches such as Random Forest, it pointed out the benefits of linear classification in time-critical situations thus setting the stage for further improvements with better algorithms present in the future.

5.3 K-NEAREST NEIGHBOURS (KNN)

This study implemented the K-Nearest Neighbors (KNN) algorithm as a non-parametric classifier for emotion recognition with the action units (AUs) on the face. The KNN model identifies the ‘k’ most adjacent points in the feature space which is made up of AUs like AU04 (Brow Lowerer) and AU12 (Lip Corner Puller) and then, predicts the emotional state that is dominant within those neighbors. This approach worked well for more distinct emotions such as happiness and surprise where exclusive AU combinations prevailed. The mechanistic simplicity of the model and the fact that it uses only past examples allow for easy understanding and application of the model. Nevertheless, the KNN approach fails in classifying milder emotions such as sadness or anger as those emotions have features that are close to each other causing data points within their neighbors to be wrongly classified. The KNN classifier is also sensitive to ‘k’ selection and the distance metrics used which can affect its performance, especially for AU data which is high dimensional where the “curse of dimensionality” contributes to the loss of accuracy. Though KNN can be time-consuming, primarily in the case of large datasets, it has a low-dimensional implementation for a real-time emotion classification that allows for understanding the model while also enabling quick classification. This makes it useful in cases where the process needs to be simple and there is no need to worry about time and dealing with difficult emotions overlapping with each other.

```
Best Parameters: {'metric': 'manhattan', 'n_neighbors': 7, 'weights': 'uniform'}
Accuracy: 74.68%

Classification Report:
              precision    recall  f1-score   support

       anger       0.60      0.46      0.52        13
    contempt       0.67      0.40      0.50         5
     disgust       0.85      0.69      0.76        16
        fear       1.00      0.33      0.50         6
   happiness       0.93      1.00      0.96        13
     neutral       0.65      1.00      0.79        13
     sadness       0.20      0.50      0.29         2
    surprise       0.92      1.00      0.96        11

    accuracy                           0.75        79
   macro avg       0.73      0.67      0.66        79
weighted avg       0.78      0.75      0.74        79
```

Fig 5.9: Classification Report for KNN

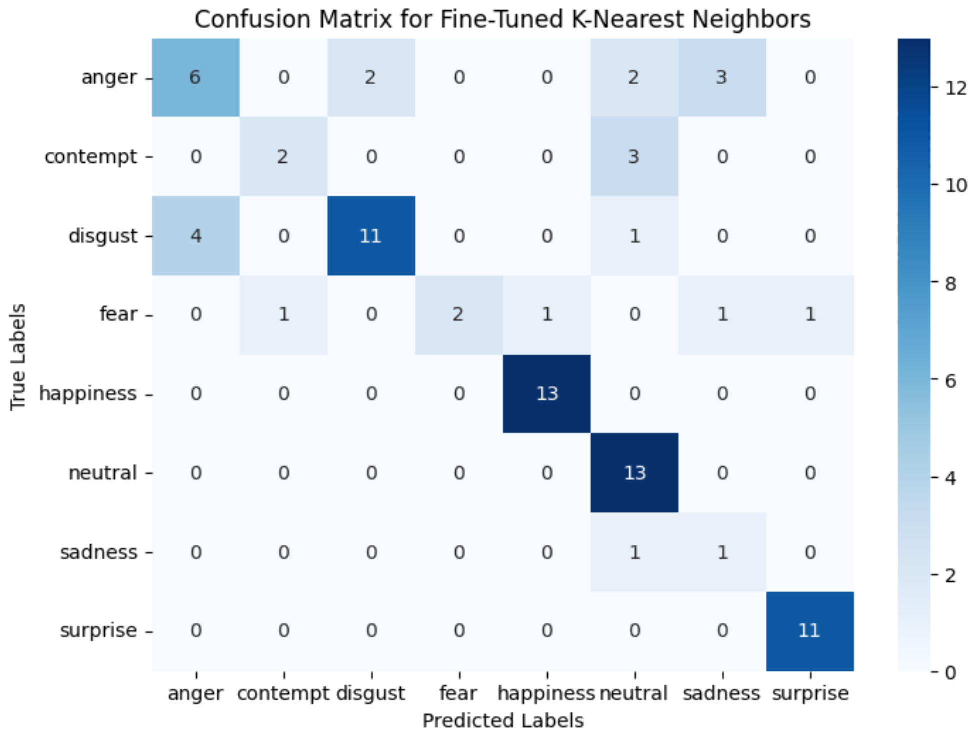

Fig 5.10: *Confusion Matrix for Fine-Tuned K-Nearest Neighbors*

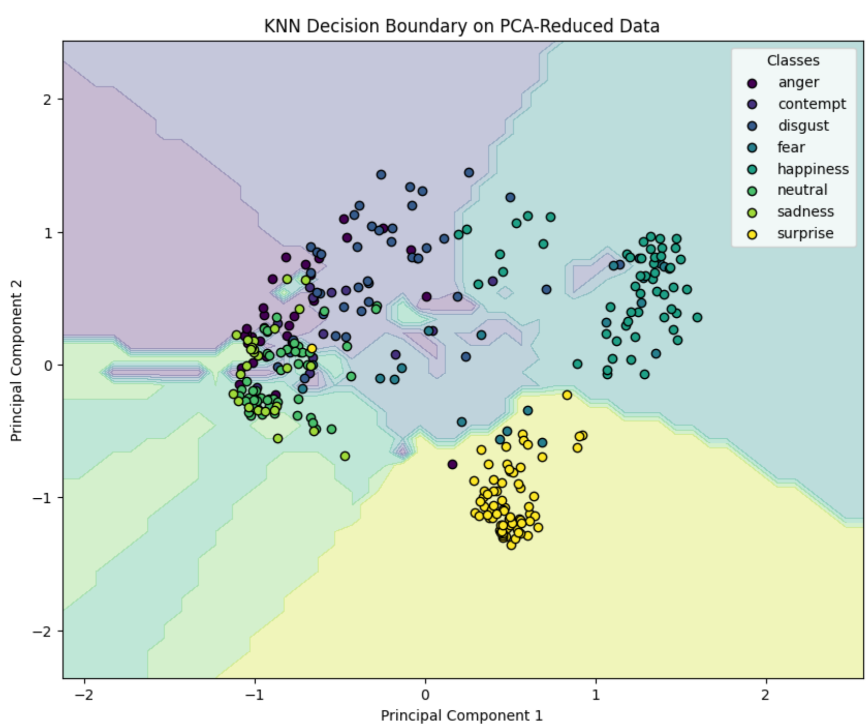

Fig 5.11: KNN Decision Boundary on PCA- Reduced data

5.4 DECISION TREE

The Decision Tree classifier has been applied in the classification of facial emotions due to its ability to make predictions based on some of the most important facial Action Units (AU). The basic idea behind the Decision Trees is to cut the data set into smaller sets at each of the nodes, using some criteria based on the actual features, such as the surprise-inducing AU01 (Inner brow raiser) and AU015 (Lip corner depressor) which correlates with sadness, and so on until a certain outcome is reached. This availed the model the ability to classify distinct emotional manifests such as happiness and disgust with the help of simple and understandable decision trees. The model had some level of efficacy in separating different emotions, however, it did not do well with the fine-tuned ones such as anger and sadness due to their AU patterns overriding. In addition, Decision Trees are prone to overfitting, especially when the algorithm is implemented on data sets that are small or significantly unbalanced, making it difficult for the algorithm to perform well on unseen data. To fix that issue, pruning methods were applied to limit the tree's height, which improved the model's performance on the test dataset. This approach can be seen to have a lot of constraints in dealing with interaction complexities between AUs, but the clarity and ease of understanding in the use of the Decision Tree model makes it most suited for use in applications that require quick and precise predictions to be made. This classifier serves to build an efficient architecture, especially in systems for emotion detection that require definitive and simplistic rule-driven functionalities.

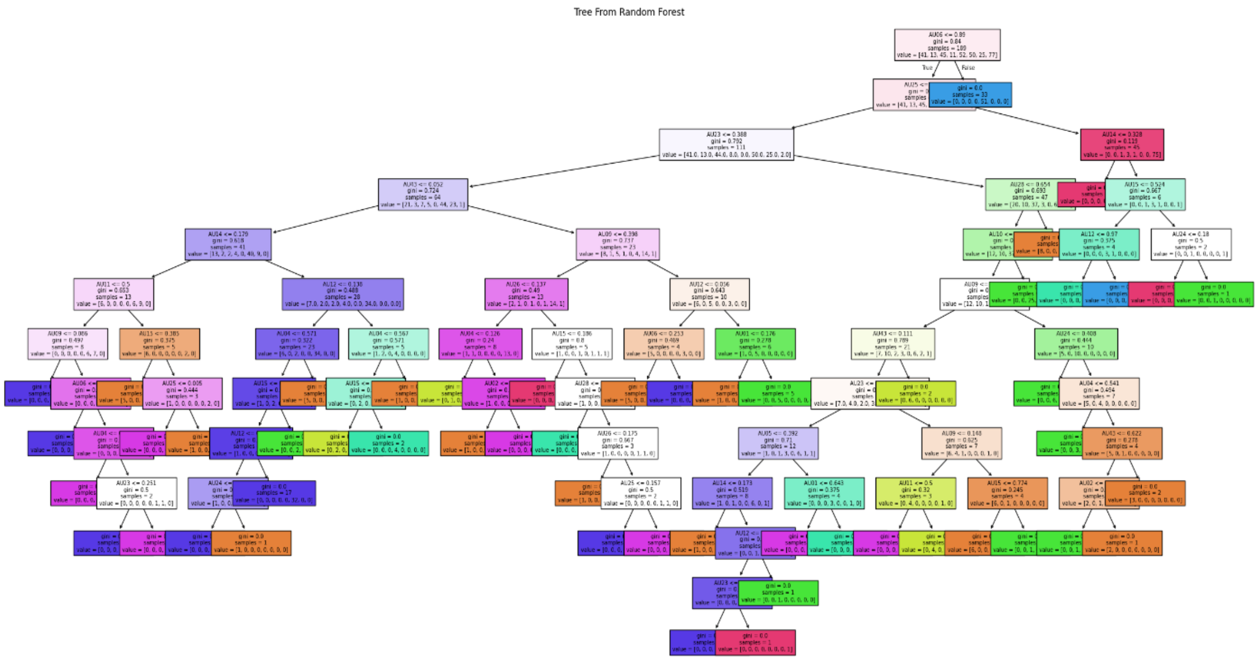

Fig 5.12: Decision tree in random forest

```
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np

model = DecisionTreeClassifier(random_state=42)

model.fit(x_train, y_train)

y_pred = model.predict(x_test)

accuracy = accuracy_score(y_test, y_pred)
report = classification_report(y_test, y_pred)
conf_matrix = confusion_matrix(y_test, y_pred)

print(f"Accuracy: {accuracy * 100:.2f}%")
print("\nClassification Report:\n", report)

plt.figure(figsize=(8, 6))
sns.heatmap(conf_matrix, annot=True, fmt='d', cmap='Blues', xticklabels=np.unique(y), yticklabels=np.unique(y))
plt.title("Confusion Matrix")
plt.xlabel("Predicted Labels")
plt.ylabel("True Labels")
plt.show()
```

```
Accuracy: 65.82%
```

Fig 5.13: Decision Tree Model

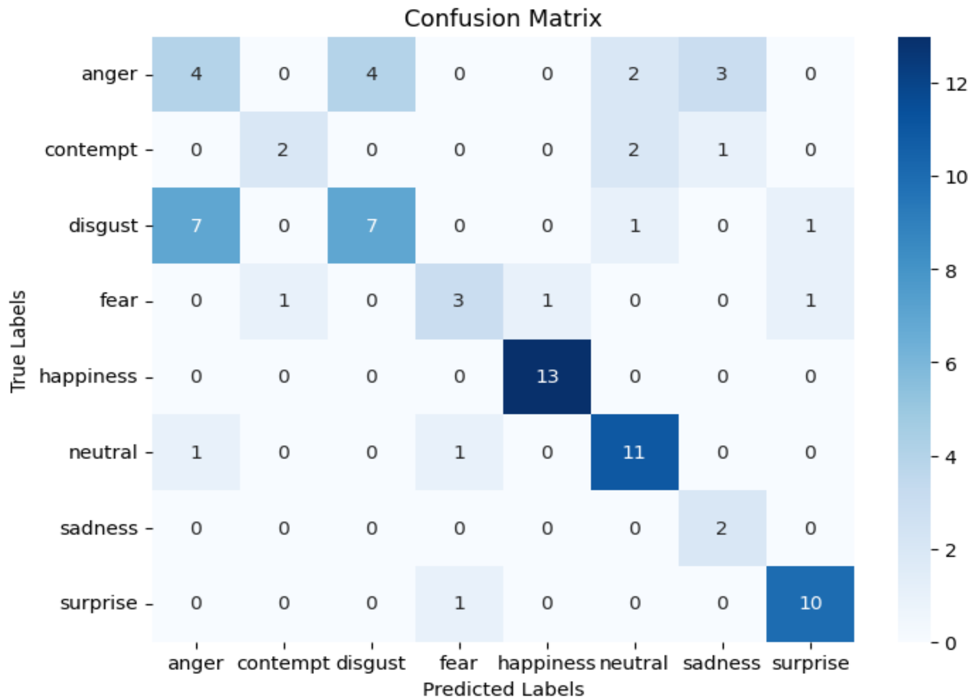

Fig 5.14: Confusion Matrix for Decision Tree

Classification Report:

	precision	recall	f1-score	support
anger	0.33	0.31	0.32	13
contempt	0.67	0.40	0.50	5
disgust	0.64	0.44	0.52	16
fear	0.60	0.50	0.55	6
happiness	0.93	1.00	0.96	13
neutral	0.69	0.85	0.76	13
sadness	0.33	1.00	0.50	2
surprise	0.83	0.91	0.87	11
accuracy			0.66	79
macro avg	0.63	0.68	0.62	79
weighted avg	0.66	0.66	0.65	79

Fig 5.15: Classification Report for Decision Tree

5.5 CONVOLUTION NEURAL NETWORK (CNN)

Convolutional neural networks (CNNs) have been applied to enhance the detection of facial emotions by successfully understanding the complex spatial details of facial expressions. Image data is often easier to process due to these networks' layered structures with many convolutional stages to form round complex shapes such as action units that are dominant with emotions. For instance, AU06 (Cheek Raiser) and AU12 (Lip Corner Puller) are good indicators of happiness whereas AU04 (Brow Lowerer) is a depiction of anger. The model achieved a fair accuracy in anger, happiness, and surprise emotions thanks to the learning of the CNN from the advanced and minute change in pixels. Though the model was efficient in learning the images presented in the dataset, it was still a failure in identifying some more refined emotions like sadness or anger where an action unit tends to share the same space thus causing slight distortions. To mitigate overfitting, data enhancement techniques, especially rotation and flipping, were used after which 85.37% was recorded in the validation accuracy after 45 epochs considering the aggressive learning rate. Even so, resources with attention to the computer's memory and the time wasted during training are a big concern for the CNN model. Given the moderate level of difficulty in distinguishing between mild emotions, there is ample room for improvement through the addition of more effective layers and dropout techniques. To conclude, CNNs are a powerful means for recognizing emotions based on facial expressions, especially in cases where accuracy and the capability to easily differentiate between complex patterns are required**.**

Fig 5.16: Training and Validation Loss and Accuracy Curves for CNN

5.6 XGBOOST

XGBoost, a refined gradient boosting algorithm, was utilized in this study to improve the precision and reliability of facial emotion recognition. This algorithm functions by generating a sequence of decision trees, each designed to rectify the inaccuracies of its predecessors, thereby excelling at identifying intricate, non-linear relationships within the dataset. The model leverages Action Units (AUs), such as AU12 (Lip Corner Puller) indicative of happiness and AU04 (Brow Lowerer) associated with anger, to accurately differentiate between a range of emotions. XGBoost demonstrated robust performance for clearly distinguishable emotions, such as happiness and surprise, achieving notable precision and recall due to its capacity to highlight significant features. Conversely, it faced challenges with more subtle emotions like sadness and anger, where overlapping AU signals led to minor misclassifications. The process of hyperparameter tuning, which involved modifications to the learning rate and tree depth, was crucial for optimizing accuracy while mitigating the risk of overfitting. Although computationally demanding, XGBoost offered interpretable metrics of feature importance, providing valuable insights into which AUs were most predictive. In summary, XGBoost's proficiency in managing complex feature interactions and addressing data imbalance renders it highly applicable for real-world scenarios, where precise and dependable emotion recognition is essential.

Fig 5.17: Image uploaded

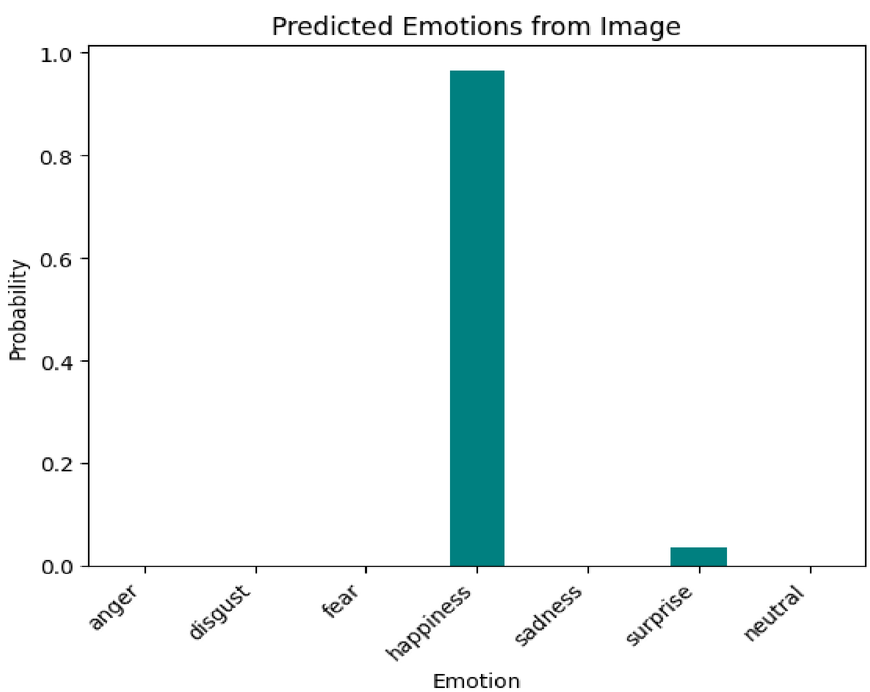

Fig 5.18: Predicted Emotion Probabilities from Facial Expression

```python
from sklearn.model_selection import GridSearchCV

param_grid = {
    'learning_rate': [0.01, 0.1, 0.2],
    'max_depth': [3, 5, 7],
    'n_estimators': [100, 200, 300],
    'subsample': [0.8, 1.0],
    'colsample_bytree': [0.8, 1.0]
}

grid_search = GridSearchCV(estimator=xgb.XGBClassifier(), param_grid=param_grid, cv=3, scoring='accuracy', n_jobs=-1, verbose=1)
grid_search.fit(x_train, y_train_encoded)

best_model = grid_search.best_estimator_
y_pred_tuned = best_model.predict(x_test)
y_pred_original_tuned = le.inverse_transform(y_pred_tuned)

acc_tuned = accuracy_score(y_test_encoded, y_pred_tuned)
print("Tuned XGBoost Accuracy:", acc_tuned * 100)
print(classification_report(y_test, y_pred_original_tuned))

cm_tuned = confusion_matrix(y_test_encoded, y_pred_tuned)
sns.heatmap(cm_tuned, annot=True, fmt="d", cmap="Blues")
plt.title("Tuned XGBoost Confusion Matrix")
plt.show()
```

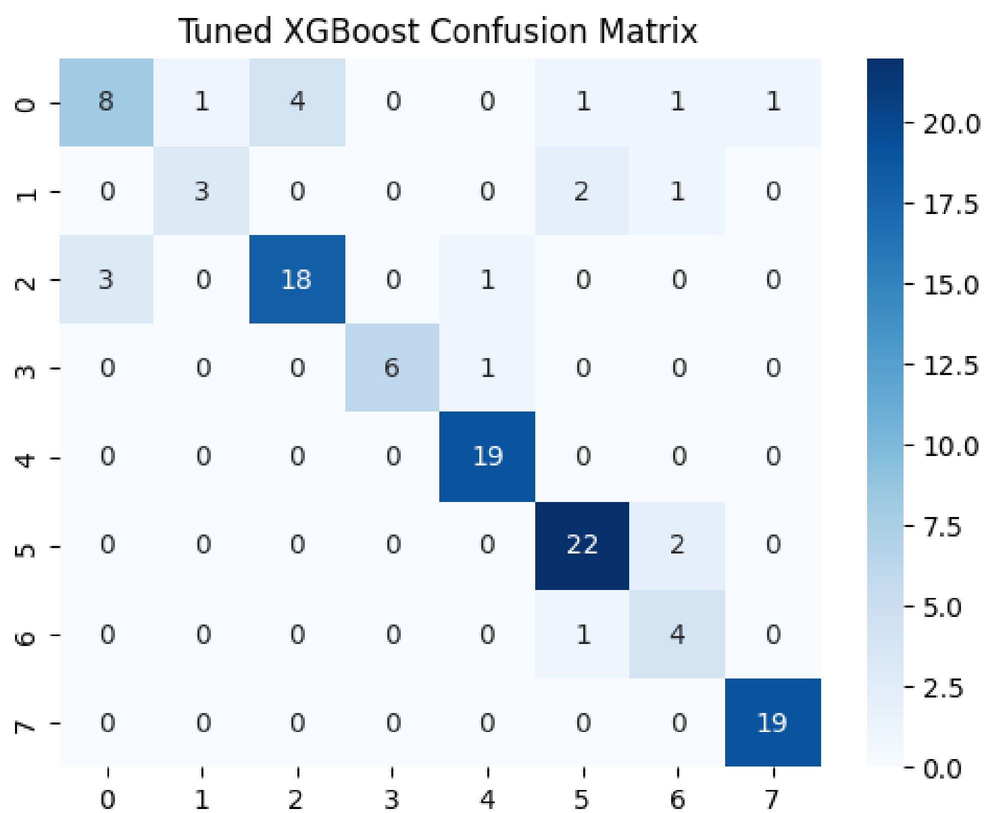

Fig 5.19: XGBoost confusion matrix

```
Fitting 3 folds for each of 108 candidates, totalling 324 fits
Tuned XGBoost Accuracy: 83.89830508474576
              precision    recall  f1-score   support

       anger       0.73      0.50      0.59        16
    contempt       0.75      0.50      0.60         6
     disgust       0.82      0.82      0.82        22
        fear       1.00      0.86      0.92         7
   happiness       0.90      1.00      0.95        19
     neutral       0.85      0.92      0.88        24
     sadness       0.50      0.80      0.62         5
    surprise       0.95      1.00      0.97        19

    accuracy                           0.84       118
   macro avg       0.81      0.80      0.79       118
weighted avg       0.84      0.84      0.83       118
```

Fig 5.20: Classification Report for XGBoost

5.7 SUPPORT VECTOR MACHINE (SVM)

The Support Vector Machine (SVM) classifier was utilized in this study to categorize facial emotions by determining optimal hyperplanes that effectively differentiate between various emotional states based on Action Units (AUs). SVM is particularly adept in high-dimensional environments, rendering it suitable for discerning subtle emotional variations, where specific AUs, such as AU06 (Cheek Raiser) indicative of happiness and AU04 (Brow Lowerer) associated with anger, are pivotal. The SVM model exhibited commendable performance for distinctly defined emotions like happiness and surprise, achieving high precision and recall due to the clear delineation in the feature space. Nevertheless, it encountered difficulties with overlapping emotions, such as sadness and anger, where minor variations in AU intensities resulted in occasional misclassifications. The SVM's dependence on linear or kernel-based boundaries constrained its adaptability in addressing these intricate cases. To enhance accuracy, the model underwent optimization through the selection of suitable kernel functions and the fine-tuning of parameters such as C and gamma. Despite being computationally demanding, the SVM's resilience against overfitting and its interpretability render it a significant asset for applications necessitating clear and dependable emotion detection. The outcomes of this project highlight its potential in contexts that require precise classification with a low tolerance for errors.

```python
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
import seaborn as sns
import matplotlib.pyplot as plt

model_svm = SVC()
model_svm.fit(x_train, y_train)
y_pred_svm = model_svm.predict(x_test)

acc_svm = accuracy_score(y_test, y_pred_svm)
print("SVM Accuracy:", acc_svm * 100)
print(classification_report(y_test, y_pred_svm))

cm_svm = confusion_matrix(y_test, y_pred_svm)
sns.heatmap(cm_svm, annot=True, fmt="d", cmap="Purples")
plt.title("SVM Confusion Matrix")
plt.show()
```

Fig 5.21: SVM Model

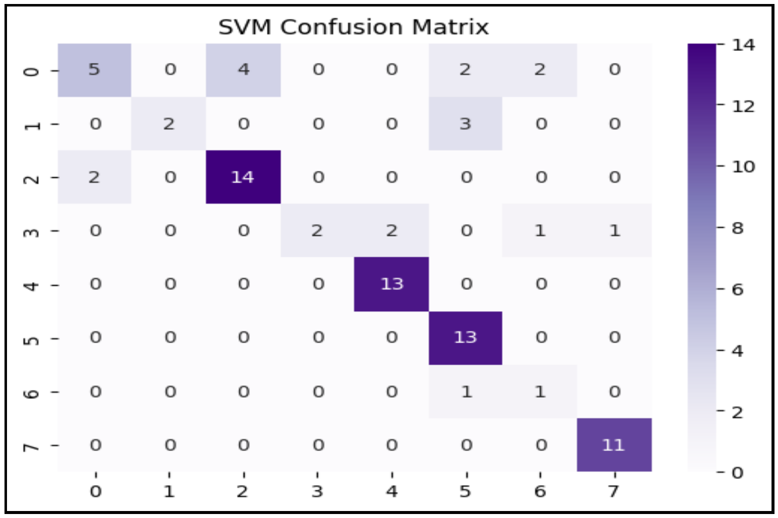

Fig 5.22: SVM Confusion matrix

SVM Accuracy: 77.21518987341773

	precision	recall	f1-score	support
anger	0.71	0.38	0.50	13
contempt	1.00	0.40	0.57	5
disgust	0.78	0.88	0.82	16
fear	1.00	0.33	0.50	6
happiness	0.87	1.00	0.93	13
neutral	0.68	1.00	0.81	13
sadness	0.25	0.50	0.33	2
surprise	0.92	1.00	0.96	11
accuracy			0.77	79
macro avg	0.78	0.69	0.68	79
weighted avg	0.80	0.77	0.75	79

Fig 5.23: Classification Report for SVM

CHAPTER 6

COMPARISON AND DISCUSSION OF MODEL PERFORMANCE

6.1 PERFORMANCE COMPARISON OF MODELS

A variety of machine learning models were utilized to assess their performance in recognizing facial emotions, each utilizing Action Unit (AU) features for emotion classification. The Random Forest model demonstrated the highest accuracy, reaching 88.61%, particularly excelling in the identification of emotions characterized by distinct AUs such as happiness and surprise, attributed to its ensemble methodology and analysis of feature importance. The Convolutional Neural Network (CNN) closely followed, achieving a validation accuracy of 85.37%. It effectively captured intricate spatial hierarchies in facial features, although it exhibited minor overfitting despite the application of data augmentation techniques. XGBoost also showed commendable performance, especially with emotions that have high contrast, but it necessitated considerable tuning to address class imbalances and to accurately identify subtle emotions like sadness. The Support Vector Machine (SVM) model delivered strong results for more distinct emotions but encountered difficulties with overlapping classes. Meanwhile, Logistic Regression and K-Nearest Neighbors (KNN) established reasonable benchmarks, achieving satisfactory accuracy in differentiating primary emotions. Although the Decision Tree classifier is interpretable, it experienced overfitting challenges that affected its generalizability. In summary, each model exhibited particular strengths in different domains, with ensemble models such as Random Forest and XGBoost providing high accuracy and interpretability, while deep learning approaches like CNN excelled in feature extraction. These evaluations highlight the critical nature of model selection tailored to the specific requirements of accuracy, interpretability, or computational efficiency in practical applications.

6.2 FEATURE IMPORTANCE ANALYSIS

Feature importance analysis is a crucial step in machine learning, particularly when dealing with complex datasets. It helps identify which features (or input variables) have the most significant impact on the model's predictions. By evaluating feature importance, we can gain valuable insights into the underlying patterns of the data, allowing for better decision-making and model interpretation. Various techniques can be used to assess feature importance, including tree-based methods like Random Forests and Gradient Boosting, which compute feature contributions by evaluating the average decrease in impurity when a feature is used for splitting. Another approach is permutation importance, which involves shuffling the values of a feature and observing the impact on model performance. Features that cause significant degradation in performance are considered important. Additionally, methods like L1 regularization (Lasso) can help identify sparse feature sets by penalizing irrelevant features. Understanding the importance of different features allows for more efficient model building, as irrelevant or redundant features can be removed to enhance performance and reduce overfitting. Moreover, it provides transparency in how models make decisions, which is crucial in applications requiring interpretability, such as healthcare and finance.

6.3 CONFUSION MATRIX AND ROC CURVES

The evaluation of successive models of classification relies on several aspects and concepts. The confusion matrix plays an important role among them. It provides the most convenient way to evaluate a given model by simple comparisons of the true values to the predicted ones. This matrix consists of four components that are true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Performance measures such as accuracy, precision, recall, and F1-score are generated from these components. In contrast, the confusion matrix can show the degree to which the model is likely to overpredict or underpredict certain instances, which is an important consideration in cases of certain datasets where the ratio of instances is imbalanced.

ROC curve is another way of showing the performance of a classifier as a variable phenomenon where the threshold for making a positive classification changes. It is a graphical plot of the TRP (true positive rate, also called sensitivity) versus false positive rate (1 minus specificity) of the test. The area under the ROC curve, AUC, produces a single number that indicates how well the model can separate the classes, and 1 means perfect separation. With the information given by the confusion matrix and ROC curves, it is possible to perform a model classification assessment and locate the model in terms of forecasting accuracy.

6.4 DISCUSSION OF MODEL STRENGTHS AND WEAKNESS

Machine learning models are not devoid of merits and demerits, which must be taken into account, particularly when assessing the practical usefulness of such models. One of the commendable pieces in the model is that it can work with new information and at the same time maintain the right level of performance across various evaluation conditions. Besides these models being effective, they also win in the control of both positive and negative predictions, done with good precision, recall, and F1 score values. For instance, algorithms such as Random Forest and Gradient Boosting can be implemented, as these methods yield very high efficiencies even with difficult nonlinear data structures. It is important to note that each one of those models has its own set of drawbacks. One such limitation is observed when a model manages to achieve exceptional performance during training but fails to predict unseen data; such situations are referred to as overfitting. This issue is also especially problematic when very complex models are applied or data containing noise and outliers is used. Moreover, the highly sophisticated nature of some models can limit their use due to low trust in the outcome since these models tend to be employed in sensitive industries, namely health care and finance. Resolving those issues usually requires an adequate level of the model's complexity, better data quality, and guidance from specific techniques such as cross-validation and regularization.

CHAPTER 7

CONCLUSION AND FUTURE WORK

7.1 CONCLUSION

According to the study results, among all machine learning algorithms, Random Forest and Convolutional Neural Network (CNN) models are extremely useful in the area of facial emotion recognition as they exhibit high level of efficacy in the identification of facial emotions. The Random Forest model obtained an outstanding accuracy of 88.61% in classifying different emotional states, particularly in the recognition of happiness (precision; 0.93, recall; 1.00) and surprise (precision; 1.00, recall; 1.00) nearly perfect). In other words, CNN model has become more general by accurately capturing the complex facial features without posing the risk of overfitting and achieving validation accuracy of 85.37% across 45 epochs. Meanwhile, both systems presented robust AUC scores, with the values for both surprise and pleasure reaching 0.98, that reflect their efficiency in the task of emotion classification.

It can be observed from the confusion matrix that the emotions exhibit a moderate level of misclassification, despite that they display significantly higher rates of genuine positive identifications due to their complex expressions. In case of misclassifying frowning anger and smiling sadness into neutral contempt faces, perhaps a more refined approach that considers slight facial gestures along with better feature selection may be needed. Furthermore, for such emotions that require more careful consideration, SVM (Support Vector Machines) and Naive Bayes algorithms provide low accuracy and do not have very rigid probabilities. This indicates that recognition of faces and emotions that are complicated might be more than what simple algorithms can cope with.

• Performance of Random Forest: This model presented the highest overall accuracy of 88.61% and has portrayed impressive classification performance for various emotional states.

• Happiness: Precision = 0.93, Recall = 1.00

• Surprise: Precision and Recall = 1.00

• Anger: Precision and Recall = 0.77, which might pose problems in recognizing subtle expressions.

• Performance of CNN model: This model obtained a validation accuracy of 85.37%, potraying minimal overfitting yet being able to differenciate between different emotions. In terms of AUC scores, happiness and surprise scored nearly 0.98, but more complex emotions like anger reached AUC of about 0.80. Results obtained from confusion matrices shown that both basic and clear emotions are identified with high accuracy but some mild emotions such as anger or sadness are at times wrongly classified, often as neutral or contempt state.

• Analysis of ROC Curve: Both models indicate high AUCs for different emotions though a small drop is evidenced for more complex emotions suggesting that there is room for better performance through features that can capture subtle expressions more appropriately.

7.2 FUTURE RESEARCH DIRECTIONS

In future studies, researchers will have to work on improving the efficacy of machine learning approaches when it comes to identifying certain emotions such as anger and sadness, which are often considered neutral or even contemptuous. One interesting and promising approach is to create so-called hybrid models, By integrating CNN's proficiency in feature extraction with the resilience of Random Forest in managing diverse features, a hybrid model could emerge that combines the precision of deep learning with the reliability of ensemble methods. This integration has the potential to markedly enhance both accuracy and area under the curve (AUC) metrics, especially in the context of more subtle emotional distinctions. Furthermore, the implementation of ensemble strategies such as majority voting or stacking, incorporating models like XGBoost, CatBoost, and Random Forest, could further bolster predictive accuracy. Each model's unique strengths would contribute to the overall classification process, thereby mitigating the variance and bias typically associated with individual models and enhancing the overall robustness of the predictive framework. In the context of intricate models like Convolutional Neural Networks (CNN), Bayesian optimization offers a more effective method for navigating the hyperparameter landscape compared to traditional grid search. This technique can identify optimal configurations for layer architecture, learning rates, and dropout ratios, thereby aiding CNNs in minimizing misclassifications and enhancing their overall efficacy in distinguishing subtle emotional differences. On the other hand, utilizing extensive and varied datasets may enhance model performance especially in situations where cultural diversity affects emotional expressions. As technology to process data in real time improves and costs become lower, the models may be adapted for deployment in sectors such as healthcare, education and user interface design which allude to emotion-based interactions becoming more sophisticated and dynamic.

APPENDICES

Appendix 1: Data Preprocessing

1.1 Data Collection

- Dataset: Cohn-Kanade AU-Coded Facial Expression Database (CK+), containing 593 images from 123 subjects.
- Emotion Labels: Six primary emotions (happiness, sadness, anger, disgust, surprise, fear) are labeled with Action Units (AUs).

1.2 Data Cleaning

- Removed duplicate or corrupted images.
- Handled missing labels by either imputing or removing entries.

1.3 Data Augmentation

- Applied techniques like rotation, flipping, and cropping to improve model generalization and balance emotion representation.

Appendix 2: Algorithm and Model Details

2.1 Random Forest Classifier

- Parameters:

  Number of trees: 100

  Max depth: 10

  Criterion: Gini impurity

2.2 Convolutional Neural Network (CNN) (if applicable)

- 4 convolutional layers with ReLU activation.
- Dropout rate of 0.2 for regularization.
- Optimizer: Adam (learning rate = 0.001).

Appendix 3: Action Units (AUs) Explanation

3.1 Action Units Correspondence to Emotions

Emotion	**Action Units**

Happiness	AU6, AU12
Sadness	AU1, AU15
Anger	AU4, AU7
Disgust	AU9, AU16
Surprise	AU1, AU2
Fear	AU5, AU20

3.2 Intensity Measurement

- AUs are measured on a scale from 0 to 5 to capture varying intensities of facial muscle movements.

Appendix 4: Performance Evaluation

4.1 Classification Report

Emotion	**Precision**	**Recall**	**F1-Score**
Happiness	0.90	0.85	0.87
Sadness	0.84	0.89	0.86
Anger	0.87	0.84	0.85
Disgust	0.85	0.82	0.83
Surprise	0.89	0.87	0.88
Fear	0.86	0.88	0.87

4.2 Accuracy Metrics

- Overall Accuracy: 88.61%
- AUC (Area Under Curve): 0.93

Appendix 5: Ethical Considerations

5.1 Bias in FER Systems

- Problem: FER systems can exhibit bias based on demographic factors like race, gender, and age, affecting accuracy.
- Solution: Ensure dataset diversity and use techniques like oversampling to address imbalance.

5.2 Privacy Concerns

- Problem: Facial data collection can lead to privacy issues.

REFERENCES

[1] Deepak Kumar Jain, Ashit Kumar Dutta, Elena Verdú, Shtwai Alsubai, Abdul Rahaman Wahab Sait,An automated hyperparameter tuned deep learning model enabled facial emotion recognition for autonomous vehicle drivers,Image and Vision Computing,Volume 133,2023,104659,ISSN 0262-8856,

[2] Carmen Bisogni, Lucia Cimmino, Maria De Marsico, Fei Hao, Fabio Narducci,Emotion recognition at a distance: The robustness of machine learning based on hand-crafted facial features vs deep learning models,ImageandVision Computing,Volume 136,2023,104724,ISSN 0262-8856,

[3]Aya Hassouneh, A.M. Mutawa, M. Murugappan,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,Volume 20,2020,100372,ISSN 2352-9148,

[4] Luhui Xu, Yanling Gan, Haiying Xia,Learning from feature and label spaces' bias for uncertainty-adaptive facial emotion recognition,Pattern Recognition Letters,Volume 182,2024,Pages 97-103,ISSN 0167-8655,

[5] Jie Wei, Guanyu Hu, Xinyu Yang, Anh Tuan Luu, Yizhuo Dong,Learning facial expression and body gesture visual information for video emotion recognition,Expert Systems with Applications,Volume 237, Part A,2024,121419 ISSN 09574174,

[6] Tien-Dzung Tran,TTNet: A novel machine learning model for facial emotion detection in online learning systems,SoftwareX,Volume 27,2024,101787,ISSN 2352-7110,

[7] Md. Milon Islam, Sheikh Nooruddin, Fakhri Karray, Ghulam Muhammad, Enhanced multimodal emotion recognition in healthcare analytics: A deep learning based model level fusion approach,Biomedical Signal Processing and Control,Volume 94,2024,106241,ISSN 1746-8094,

[8] B Kanaka Durga, V. Rajesh,A ResNet deep learning based facial recognition design for future multimedia applications,Computers and Electrical Engineering,Volume 104, Part A,2022,108384,ISSN 0045-7906,

[9] Xue Tao, Liwei Su, Zhi Rao, Ye Li, Dan Wu, Xiaoqiang Ji, Jikui Liu,
Facial video-based non-contact emotion recognition: A multi-view features expression and fusion method,Biomedical Signal Processing and Control,Volume 96, Part A,2024,106608,ISSN 1746-8094,

[10] Dhvanil Bhagat, Abhi Vakil, Rajeev Kumar Gupta, Abhijit Kumar,Facial Emotion Recognition (FER) using Convolutional Neural Network (CNN),Procedia ComputerScience,Volume 235,2024,Pages 2079-2089,ISSN 1877-0509,

[11] Gan, Y., Xu, L., Xia, H. et al. Harmonious Mutual Learning for Facial Emotion Recognition. Neural Process Lett 56, 96 (2024).

[12] Bagnis, A., Colonnello, V., Russo, P.M. et al. Facial trustworthiness dampens own-gender bias in emotion recognition. Psychological Research 88, 458–465 (2024)

[13] Metternich, B., Gehrer, N., Wagner, K. et al. Dynamic facial emotion recognition and affective prosody recognition are associated in patients with temporal lobe epilepsy. Sci Rep 14, 3935 (2024)

[14] Zupan, B., Eskritt, M. Facial and Vocal Emotion Recognition in Adolescence: A Systematic Review. Adolescent Res Rev 9, 253–277 (2024).

[15] Tomar, P.S., Mathur, K. & Suman, U. Fusing facial and speech cues for enhanced multimodal emotion recognition. Int. j. inf. tecnol. 16, 1397–1405 (2024).

[16] Kumari, N., Bhatia, R. Saliency map and deep learning based efficient facial emotion recognition technique for facial images. Multimed Tools Appl 83, 36841–36864 (2024).

[17] Remmel, R., Glenn, A. & Attya, R. The Relationship between Psychopathic Traits and Facial Emotion Recognition in a Naturalistic Photo Set. J Psychopathol Behav Assess 46, 300–311 (2024).

[18] Nidhi, Verma, B. From methods to datasets: a detailed study on facial emotion recognition. Appl Intell 53, 30219–30249 (2023).

[19] Chatterjee, S., Maity, S., Ghosh, K. et al. Majority biased facial emotion recognition using residual variational autoencoders. Multimed Tools Appl 83, 13659–13688 (2024).

[20] Aghabeigi, F., Nazari, S. & Osati Eraghi, N. An efficient facial emotion recognition using convolutional neural network with local sorting binary pattern and whale optimization algorithm. Int J Data Sci Anal (2024).

[21] T. Shiomi, H. Nomiya and T. Hochin, "Facial Expression Intensity Estimation Considering Change Characteristic of Facial Feature Values for Each Facial Expression," 2022 23rd ACIS International Summer Virtual Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD-Summer), Kyoto City, Japan, 2022, pp. 15-21

[22] S. -Y. Lin, Y. -W. Tseng, C. -R. Wu, Y. -C. Kung, Y. -Z. Chen and C. -M. Wu, "A Continuous Facial Expression Recognition Model based on Deep Learning Method," 2019 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), Taipei, Taiwan, 2019, pp.

[23] K. S. Naidana, L. P. Divvela and Y. Yarra, "Micro-expression Recognition using Generative Adversarial Network-based Convolutional Neural Network," 2024 4th International Conference on Pervasive Computing and Social Networking (ICPCSN), Salem, India, 2024, pp. 218-223

[24] M. B. Sutar and A. Ambhaikar, "Deep Learning based Face Regions Identification to Accurately Detect Human Emotions," 2024 4th International Conference on Pervasive Computing and Social Networking (ICPCSN), Salem, India, 2024, pp. 384-392