DEGREE PROJECT IN TECHNOLOGY,
SECOND CYCLE, 30 CREDITS
STOCKHOLM, SWEDEN 2023

Analyzing How Blended Emotions are Expressed using Machine Learning Methods

Master Thesis

Disen Ling

KTH ROYAL INSTITUTE OF TECHNOLOGY
ELECTRICAL ENGINEERING AND COMPUTER SCIENCE

Authors

Disen Ling <disen@kth.se>
ICT Innovation: Data Science
KTH Royal Institute of Technology

Place for Project

Stockholm, Sweden
KTH

Examiner

Amir H. Payberah Stockholm
KTH Royal Institute of Technology

Supervisor

Magnus Boman Stockholm
KTH Royal Institute of Technology

Abstract

Blended emotion is a classification of emotional experiences that involve the combination of multiple emotions. Research on the expression of blended emotions allows researchers to understand how different emotions interact and coexist in an individual's emotional experience. Using machine learning to analyze mixed emotions may indeed bring new insights to the study of blended emotions. This thesis aims to explore blended emotion expression by testing machine learning models (SVM, Decision Tree, and Naive Bayes) trained on the single motion dataset on the blended emotion datasets and vice versa, to analyze the relationship between blended emotions and their constituent emotions. Furthermore, this thesis explores whether there is a dominant emotion in blended emotions and conducts an ablation study to investigate the importance of various facial features within each emotion. The results of testing models' generalization capabilities propose that blended emotion expressions are highly likely to result from the overlapping combinations of features from their constituent emotions or the combination of some features from one constituent emotion with some from another. Furthermore, based on the dataset used, this thesis also finds that happiness predominated in the blended emotion '*disgust & happiness*'. Additionally, an ablation study is conducted to identify the features that have the most significant impact on the accuracy and F1 score of single/pure emotion and blended emotion recognition across various recognition models.

Keywords

Blended emotion, Supervised learning, Model generalization capability, Ablation study

Abstract

”Blandade känslor” är en klassificering av känslomässiga upplevelser som innefattar en kombination av flera känslor. Forskning om uttryck av blandade känslor möjliggör för forskare att förstå hur olika känslor interagerar och samexisterar i en individs känslomässiga upplevelse. Användningen av maskininlärning för att analysera blandade känslor kan faktiskt ge nya insikter i studiet av blandade känslor. Denna avhandling syftar till att utforska uttryck av blandade känslor genom att testa maskininlärningsmodeller (SVM, beslutsträd och Naive Bayes) som är tränade på dataset med enskilda känslor på dataset med blandade känslor och vice versa, för att analysera sambandet mellan blandade känslor och deras beståndsdelar. Dessutom utforskar denna avhandling om det finns en dominerande känsla i blandade känslor och genomför en ablationsstudie för att undersöka betydelsen av olika ansiktsdrag inom varje känsla. Resultaten av testning av modellernas generaliseringsförmåga föreslår att uttryck av blandade känslor sannolikt härrör från överlappande kombinationer av drag från deras beståndsdelar eller en kombination av vissa drag från en beståndsdel med vissa från en annan. Vidare, baserat på det använda datasetet, finner denna avhandling också att glädje dominerar i den blandade känslan ’avsky och glädje’. Dessutom genomförs en ablationsstudie för att identifiera de drag som har störst påverkan på noggrannheten och F1-poängen för igenkänning av enskilda/rena känslor och blandade känslor över olika igenkänningsmodeller.

Nyckelord

Blandade känslor, Övervakad inlärning, Modellens generaliseringsförmåga, Ablationsstudie

Acknowledgements

Here, I would like to express my heartfelt gratitude to everyone who provided me with support and guidance throughout the process of writing this thesis.

First and foremost, I would like to extend my heartfelt gratitude to Professor Magnus Boman, my supervisor. He exhibited unparalleled patience and provided me with exceptional and insightful guidance throughout this project, enabling me to complete this master thesis project. I will forever remember his guidance and will carry his teachings with me in the days to come.

I would like to express my gratitude to Professor Petri Laukka at Stockholm University. His invaluable guidance and extensive knowledge of psychology provided me with crucial insights during this project. I consider myself fortunate and honored to have had the opportunity to receive his guidance.

I would also like to extend my gratitude to Jiachen Yao and Tim Lachmann. Through our exchanges and the insights gained from reading their research papers, I found inspiration and valuable assistance. I am truly thankful for the opportunity to engage with and learn from them.

Lastly, I want to express my heartfelt gratitude to my parents and friends for their unwavering emotional support and assistance throughout the project. They provided me with the strength to navigate through the challenging phases.

Thank you all.

Acronyms

ML - Machine Learning
SVM - Support Vector Machine
AU - Action Unit
LSTM - Long Short-Term Memory
VAE - Variational Autoencoder

Contents

Chapter 1

Introduction

1.1 Background

Psychological research on emotions has a rich history that has evolved over decades [1]. In the last thirty years, emotion and its neural substrates, activation, regulation, and functions have emerged as highly prominent and extensively researched subjects in various fields of psychology and related disciplines [2]. Researchers have focused on understanding how emotions are expressed, experienced, and regulated, as well as their influence on human behavior and mental health [3].

Over the years, a plethora of psychological and philosophical theories have been proposed to grasp the complexities of emotions [4]. Among them, there are two unique theories that are widely accepted: the basic emotion theory and the dimensional theory [5][6]. However, these two theories have contradicted one another. The disparity stems from whether emotions are defined as distinct entities or as a singular, independent dimension [7].

The basic emotion theory posits that emotions are composed of a limited number of basic emotions. These emotions are perceived to have evolved due to their adaptive significance in addressing fundamental life tasks. Each emotion exhibits distinct attributes, including signals, physiology, and antecedent events. Furthermore, commonalities exist among emotions, such as rapid onset, brief duration, involuntary emergence, automatic assessment, and coherence across responses. These shared and distinct characteristics stem from our evolutionary history and serve to differentiate emotions from other affective phenomena [8][9].

On the other hand, the dimensional theory of emotion posits that emotions can be understood and categorized along several continuous dimensions rather than being discrete and distinct entities [10]. This theory suggests that emotional experiences can be represented on various scales, such as valence (positive to negative), arousal (low to high intensity), and possibly other dimensions like dominance or approach-withdrawal [11][12]. In contrast to the basic emotion theory, which proposes a set of distinct and specific emotions, the dimensional theory provides a more continuous and nuanced framework for understanding the complexity of emotional experiences [13].

However, People's daily emotional experience is a complex construct that usually involves multiple emotions blended which is referred to as blended emotion [14]. Blended emotion is a broader classification of emotional experiences that involve the combination of multiple emotions compared with mixed emotion which necessitates the simultaneous experience of two emotions with opposing valences. However, in the case of blended emotion, there is no requirement for the emotions to possess opposite valences [15][16]. Investigating blended emotions allows researchers and psychologists to gain a more comprehensive understanding of how different emotions interact and coexist within an individual's emotional experiences [17]. Blended emotions are expressed has been an important and debated research topic in the field of emotion studies for the past 30 years. The central issue of the debate is whether multiple emotional states can truly coexist or if they are merely rapidly shifting between each other [18].

In recent years, the rapid strides in machine learning (ML) and information fusion have facilitated the capability to bestow machines/computers with the skill to understand, recognize, and analyze emotions [19]. The possible applications encompass automated driver assistance, healthcare, human-computer interaction, entertainment, marketing, education, and numerous other fields [20]. Currently, there have been numerous studies and significant achievements in utilizing machine learning for emotion recognition, resulting in very impressive outcomes [21][22][23][24]. However, among the multitude of studies, there is a lack of research that employs machine learning methods to investigate blended emotions. Therefore, this field still requires more exploration and research.

1.2 Problem

A recent study indicates that through dynamic facial/bodily, and vocal expressions, blended emotions, including combinations of both same-valence and other-valence emotions, can be accurately recognized [25]. Furthermore, facial and vocal expressions have been demonstrated in numerous studies to be effective for machine learning-based emotion recognition, yielding favorable recognition performance outcomes [26]. Therefore, understanding how blended emotions are expressed is essential for comprehending their nature and gaining valuable insights. Moreover, training machine learning models to recognize blended emotions could potentially lead to novel discoveries about how these complex emotional states are expressed. The research question can be developed based on the above statement:

How are blended emotions expressed?

1.3 Purpose

To comprehensively explore the relationship between blended emotions and their component emotions, two parallel tracks are pursued. First, machine learning models are trained on the single emotion dataset and then tested on the blended emotion dataset. Additionally, these models will be trained on the blended emotion dataset and tested on the single emotion dataset. This dual methodology aims to provide a comprehensive perspective on the interplay between blended and individual emotions.

1.4 Objective

The objectives of this master thesis are:

1) To explore whether blended emotion expressions result from the overlapping combination of features from both emotions.

2) To explore whether blended emotion expressions incorporate some features from one emotion alongside those from another.

3) To explore the presence of distinctive feature patterns in blended emotion expressions that are absent in individual emotions.

4) To explore whether a dominant emotion exists among the constituent emotions of blended emotion expressions.

5) To explore the pivotal features influencing emotional expression.

1.5 Outline

Chapter 2 furnishes a detailed technical background for the article, which includes introductions to SVM, Decision Tree, and Naive Bayes. Moreover, it presents some current research in the domains of blended emotions and emotion recognition, along with their respective outcomes.

Chapter 3 provides a detailed overview of the methodology employed in completing the thesis project. This includes the data preprocessing methods, the machine learning models utilized, the model evaluation criteria, as well as an explanation of the ablation study for feature importance.

Chapter 4 presents and explains the results obtained from various tests and analyses. This encompasses the Holdout set test results and Cross-validated aggregated results for each model across different datasets, as well as the analysis of the proportions of constituent emotions in blended emotions and the assessment of the importance of facial features in each emotion.

Chapter 5 conducts an analysis and discussion of the data results obtained from various experiments and analyses conducted in Chapter 4. Conclusions are drawn based on this analysis and discussion. This includes an analysis of blended emotion expressions, an assessment of the presence of dominant emotions in blended emotions, and an evaluation of the importance of various facial features across different emotions. Additionally, limitations within the project are analyzed, and future work is outlined and discussed.

Chapter 6 serves as a comprehensive summary of the entire project, featuring an analysis of the project's completion and offering certain prospects for the future.

Chapter 2

Background and Related Work

2.1 Machine Learning Models

2.1.1 Support Vector Machine

Support Vector Machine (SVM) is a supervised machine learning method applied to tasks like classification and regression. It has found extensive use in classification and the estimation of nonlinear functions [27]. The core concept of Support Vector Machine (SVM) is based on a linear classifier that seeks to maximize the margin within the feature space. SVM also incorporates the kernel trick, enabling it to effectively act as a nonlinear classifier [28]. SVM's learning approach is centered on maximizing this margin, which can be mathematically expressed as a convex quadratic programming problem. This can equivalently be minimized by the use of a regularized hinge loss function. The SVM learning process entails solving convex quadratic programming through optimization algorithms to attain optimal solutions [29].

2.1.2 Decision Tree

Decision Tree is a fundamental supervised machine learning algorithm used for classification and regression tasks. Decision tree models are tree-structured and, in classification problems, represent the process of classifying instances based on features [30]. The leaf nodes of a decision tree contain an output variable for prediction. Predictions can be made by traversing the tree's splits until a leaf node is reached and outputting that node's category value. Decision trees learn quickly and make predictions quickly. They can also solve a large number of problems and require no special preparation of data [31].

2.1.3 Naive Bayes

Naive Bayes constitutes a straightforward learning algorithm that combines Bayes' rule with a robust assumption of attributes being conditionally independent, given the class. Although this assumption of independence is frequently challenged in real-world scenarios, naive Bayes frequently achieves competitive classification accuracy. This, combined with its computational efficiency and numerous other advantageous traits, results in the widespread practical application of naive Bayes [32].

2.2 Related Work

2.2.1 Blended Emotion

Israelsson et al. conducted two studies on whether blended emotions can be recognized from dynamic facial, bodily, and vocal expressions by people [25]. Study 1 showed accurate recognition of emotion combinations from multi-modal (facial/bodily/vocal) expressions. Study 2 demonstrated emotion combinations can be recognized in uni-modal visual and auditory conditions. Both studies presented that blended emotions, including combinations of both same-valence and other-valence emotions, can be accurately recognized from dynamic facial/bodily and vocal expressions.

Li et al. introduced RAF-ML, an innovative multi-label facial expression database, along with a new deep learning algorithm aimed at addressing the challenge of limited blended emotion datasets [33]. A crowd-sourced annotation approach is utilized, involving 1.2 million labels from 315 participants, to identify multi-label expressions obtained from social networks. Subsequently, they developed an EM algorithm to filter out unreliable labels. Notably, RAF-ML stands as the first database in real-world settings to offer crowd-sourced recognition for multi-label expressions.

Watson et al. explored the implications of a hierarchical structure [34] which consists of (a) the higher order dimensions of nonspecific Positive Activation and Negative Activation and (b) multiple specific negative emotions (e.g., fear, sadness, and anger) and positive emotions (e.g., joviality, self-assurance, and attentiveness) at the lower level. The frequency of pure emotional states, same-valence emotional blends, and cross-valence mixed emotions is examined in a sizable momentary mood sample. In some cases, co-occurrence patterns can be accurately predicted from correlational data.

2.2.2 Emotion Recognition

Domínguez-Jiménez et al. introduced a model to detect three emotions—amusement, sadness, and neutrality—using physiological signals [35]. The aim was to create a dependable method for emotion recognition through wearable devices. The study employed video clips to elicit emotions in participants, monitoring heart rate and skin response. Extracted features were employed with a support vector machine for classifying emotions. Notably, using galvanic skin response features alone, accurate identification of amusement, sadness, and neutrality was achieved, with up to 100% accuracy on the test dataset.

Bazgir et al. introduced an emotion recognition system based on the valence/arousal model using electroencephalography (EEG) signals [36]. Specifically, EEG signals were split into gamma, beta, alpha, and theta bands via DWT (discrete wavelet transform), and spectral features were extracted. PCA (Principle component analysis) is applied to reduce dimensionality while maintaining independence. SVM, KNN, and ANN are applied to classify emotions. Results showed that cross-validated SVM (RBF kernel) using 10 EEG channels achieves 91.3% arousal and 91.1% valence accuracy in the beta band, outperforming existing algorithms applied to the DEAP dataset.

Minaee et al. presented a deep learning approach based on an attentional convolutional network that is able to focus on important parts of the face [37]. This approach outperforms previous models, which perform reasonably well on datasets of images captured in a controlled condition but fail to perform as well on more challenging datasets with more image variation and partial faces, on multiple datasets. Also, a visualization technique is used to find important facial regions to detect different emotions based on the classifier's output and experimental results showed that different emotions are sensitive to different parts of the face.

Chapter 3

Methods

3.1 Methodological Approach

This master's thesis project centers around the utilization of data-driven strategies to explore how blended emotions are conveyed using machine learning techniques. Consequently, the research methodology chosen for this project is primarily quantitative[38].

To collect quantitative data, facial features need to be extracted from the original video data. Subsequently, in order to ensure a sufficiently high data quality for machine learning training, data selection methods are applied to the extracted features from the videos. Following this, features are selected. Then, the features for each set of data corresponding to each video are averaged to represent that specific video. Lastly, the obtained data is subjected to z-score normalization to yield the final dataset which can be used for training and testing.

To obtain a more precise assessment of the machine learning models' performance, the cross-validation method is implemented. The average recognition accuracy obtained from cross-validation is then considered the final accuracy measure. Subsequently, to test the models' generalizability, the models trained on the single/pure emotion dataset are tested on the blended emotion dataset. Similarly, the models trained on the blended emotion dataset are tested on the single/pure emotion dataset to assess their performance. Based on the testing outcomes, analysis and inference of the expression of blended emotions are conducted. Additionally, the results will also be utilized to analyze whether a dominant emotion exists among the component emotions of blended emotional expressions.

In order to investigate the importance of various facial features in the expression of emotions, also known as analyzing feature importance, each feature in the feature list will be systematically removed, and the models will be retrained and tested accordingly. The testing results are compared with the training results obtained without removing any feature, allowing for analysis to determine the importance of the removed features [39].

3.2 Data

3.2.1 Data Source

The raw data utilized in this project consists of videos and can be divided into two parts: blended emotion videos and single/pure emotion videos. All of the videos is provided by Professor Petri Laukka and Alexandra Israelsson from the Department of Psychology at Stockholm University. The videos used in this project are part of a large-scale project on dynamic multi-modal emotional expression, in which, actors conveyed various emotions through facial gestures, body movements, and vocalizations. In the single/pure emotion videos, actors were tasked with individually performing the five emotions: anger, disgust, fear, happiness, and sadness. And in the blended emotion videos, actors were guided to convey complex emotions formed by combining each pair of anger, disgust, fear, happiness, and sadness (resulting in 10 emotional combinations). The goal was to ensure that both emotions were equally distinct in the resulting expressions [25].

The single/pure emotion videos were created with the participation of 18 actors, resulting in 18 videos for each single/pure emotion. However, due to the absence of a video depicting the emotion of sadness from actor A327, there are only 17 videos available for sadness. The blended emotion videos were created with the participation of 36 actors, resulting in 36 videos for each blended emotion. And each actor recorded videos featuring different emotional blend ratios for the same blended emotion. These blend ratios were set at 30:70, 50:50, and 70:30. In this project, videos for all three blend ratios were utilized as the raw data for blended emotions.

Each actor recorded videos showcasing various levels of emotional intensity for the same single or blended emotion. These intensity levels were categorized into four grades: low, medium, high, and very high, and they were numbered 1 to 4, respectively. In this project, only videos with a moderate intensity level were selected and used as the raw data. Therefore, in this project, a total of 89 single/pure emotion videos and 1080 blended emotion videos were used as the raw dataset.

3.2.2 Feature Extraction

To extract the facial features of actors conveying emotions in the original video data, OpenFace 2.2.0, an open-source facial behavior analysis toolkit, was utilized. OpenFace is a toolkit that possesses the ability to detect facial landmarks, estimate head pose, estimate eye gaze, and recognize facial action units [40].

Regarding facial landmark detection and tracking [41][42], this pertains to the procedure of recognizing crucial points or distinct features on an individual's face and subsequently ensuring continuous monitoring of their locations while the face is in motion within a sequence of images or videos. The OpenFace toolkit can detect and track 2D and 3D facial landmarks. For 2D facial landmarks, the output consists of 68 coordinates x_0, x_1, ... x_66, x_67, y_0, ... y_67. Meanwhile, the 3D facial landmarks are represented by their coordinates in 3D space, denoted as X_0, ... X_67, Y_0, ... Y_67, Z_0, ... Z_67. Figure 3.2.1 shows the 68 Facial Landmarks

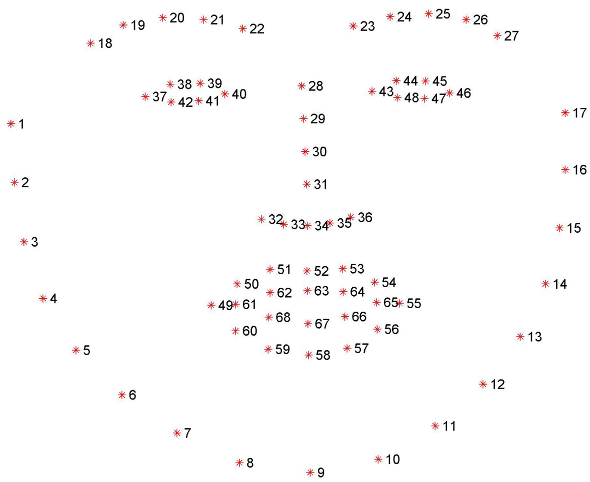

Figure 3.2.1: 68 Facial Landmarks are specific points on a person's face that are used for various computer vision and facial recognition tasks. These landmarks help in identifying and analyzing different facial characteristics [43].

Head pose estimation is regarded as the process of determining the orientation or position of a person's head in three-dimensional space relative to a reference point or coordinate system. The OpenFace toolkit can track the head pose of a subject in videos and provide corresponding data. In this context, pose_Tx, pose_Ty, and pose_Tz represent the position of the head relative to the camera, measured in millimeters. Notably, pose_Tz signifies the distance of the head from the camera, with positive values indicating a greater distance from the camera along the positive Z-axis. On the other hand, pose_Rx, pose_Ry, and pose_Rz denote the radians of rotation around the X, Y, and Z axes, respectively. The convention here is to use a left-hand coordinate system, where positive values signify counterclockwise rotation. These rotations can be understood as pitch (Rx), yaw (Ry), and roll (Rz). The rotations are represented in a world coordinate system with the camera as the origin.

Concerning the tracking of eye gaze [44], this involves the procedure of observing and analyzing the orientation in which an individual's eyes are directed. OpenFace toolkit can track the gaze direction vectors of human eyes in videos and provide corresponding vector and angle data. The gaze direction vector for the left eye is represented by gaze_0_x, gaze_0_y, gaze_0_z, while the right eye's gaze direction vector is represented by gaze_1_x, gaze_1_y, gaze_1_z. Additionally, gaze_angle_x and gaze_angle_y represent the gaze direction of the eyes in radians, averaged across both eyes and transformed into a more user-friendly format for representation in world coordinates.

Facial action unit detection [45], is based on the Facial Action Coding System (FACS) [46], a methodology for categorizing human facial movements based on their visual manifestations on the face. FACS encodes the actions of individual facial muscles through subtle changes in facial appearance that occur in an instant. Through FACS, it becomes feasible to code practically any anatomically plausible facial expression by breaking it down into the distinct AUs that contributed to its formation. The descriptions for each AU are as indicated in Table 3.2.1.

OpenFace toolkit can detect the intensity of 17 Action Units (AUs) on a scale from 0 to 5: AU01_r, AU02_r, AU04_r, AU05_r, AU06_r, AU07_r, AU09_r, AU10_r, AU12_r, AU14_r, AU15_r, AU17_r, AU20_r, AU23_r, AU25_r, AU26_r, and AU45_r. Additionally, it can also determine the presence of 18 AUs (with 0 indicating absence and 1 indicating presence): AU01_c, AU02_c, AU04_c, AU05_c, AU06_c, AU07_c, AU09_c, AU10_c, AU12_c, AU14_c, AU15_c, AU17_c, AU20_c, AU23_c, AU25_c, AU26_c, AU28_c, and AU45_c. Figure 3.2.2 shows some AUs can be recognized through OpenFace.

Action Unit	Description
1	Inner Brow Raiser
2	Outer Brow Raiser (unilateral, right side)
4	Brow Lowerer
5	Upper Lid Raiser
6	Cheek Raiser
7	Lid Tightener
9	Nose Wrinkler
10	Upper Lip Raiser
11	Nasolabial Deepener
12	Lip Corner Puller
13	Cheek Puffer
14	Dimpler
15	Lip Corner Depressor
16	Lower Lip Depressor
17	Chin Raiser
18	Lip Puckerer
20	Lip stretcher
22	Lip Funneler
23	Lip Tightener
24	Lip Pressor
25	Lips part
26	Jaw Drop
27	Mouth Stretch
28	Lip Suck
41	Lid droop
42	Slit
43	Eyes Closed
44	Squint
45	Blink
46	Wink

Table 3.2.1: Descriptions for each Action Units[47].

3.2.3 Data Filtering

Through OpenFace, facial data from each frame of the original video data is extracted. OpenFace evaluates the Confidence (how confident the tracker is in the current landmark detection image) and Success (whether the tracking was successful, if a face was present in the frame, or if the tracking is considered good) for the data extracted from each frame. In this project, if the confidence of more than 15% of the frames in a video is below 98%, or if Success is 0, then the video is considered low-quality and will be filtered out. This approach is taken to ensure that only high-quality data is utilized for training machine learning models. Comparison of the number of single/pure emotion videos and blended emotion videos before and after filtering are presented

Figure 3.2.2: A portion of the AUs recognizable by OpenFace[48]

in Table 3.2.2 and Table 3.2.3. Therefore, based on this data filtering approach, 12.3% of single/pure emotion videos and 23.5% of mixed emotion videos have been filtered out.

Emotion	Before Filtering	After Filtering
Anger	18	16
Disgust	18	16
Fear	18	17
Happiness	18	16
Sadness	17	13

Table 3.2.2: Comparison of the number of single/pure emotion videos before and after filtering.

3.2.4 Feature Selection

OpenFace extracts a substantial number of facial features, yet not all of these are essential for training machine learning models. Therefore, feature selection is necessary to pick the relevant features needed for the models. After conducting a literature review, it was discovered that Tim Lachmann, a member of Laukka's research group, using the same database, found that for single modality models, employing Action Units (AUs) with intensity yielded the best recognition results in his project [49]. Consequently, the facial features chosen for this project are AU01_r, AU02_r, AU04_r, AU05_r, AU06_r, AU07_r, AU09_r, AU10_r, AU12_r, AU14_r, AU15_r, AU17_r, AU20_r, AU23_r, AU25_r, AU26_r, and AU45_r.

Emotion	Before Filtering	After Filtering
Anger&Disgust	108	82
Anger&Fear	108	82
Anger&Happiness	108	78
Anger&Sadness	108	80
Disgust&Fear	108	78
Disgust&Happiness	108	78
Disgust&Sadness	108	83
Fear&Happiness	108	88
Fear&Sadness	108	86
Happiness&Sadness	108	91

Table 3.2.3: CComparison of the number of blended emotion videos before and after filtering.

3.2.5 Linear Interpolation

After the data filtering process, the data of the expected quality is obtained. However, in some videos, there are still frames where the confidence is below 98% or success is 0. As a result, the data from those frames will not be used, and a linear interpolation method will be employed to fill in the facial features' values of those frames.

Linear interpolation is a mathematical technique used to estimate a value within a range based on two known values. It assumes a linear relationship between the known points and calculates an intermediate value based on their positions. In other words, it "connects the dots" between two data points and provides an approximation for the value at a given point between them. The interpolation is calculated based on the following formula [50]:

$$f(x_{missing}) = f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0} * (x_{missing} - x_0)$$

in which $f(x)$ means the values of facial features of the frame number x and $x_0 < x_{missing} < x_1$.

After filtering out data from frames with lower confidence scores, this method allows for the filling of missing values caused by data filtering, thereby supplementing the dataset.

3.2.6 Z-Score Normalization

Even though all the actors were tasked to express emotions in medium intensity, it is important to recognize that the definition of ”medium” varies among individual actors. In other words, the emotional intensity baselines differ among actors, leading to variations in the intensity of Action Units when they are acting even though they are expressing the same emotion at nominally the same intensity.

A method to account for variations in baseline among actors is to individually normalize all features within each actor and subsequently conduct analyses on the normalized features. Therefore, the Z-Score normalization method is applied separately to each feature of each actor. The following formula is used to perform a z-score normalization on each facial feature of each actor separately:

$$x^{'} = \frac{x-\mu}{\sigma}$$

where x is the original value, μ is the mean value of the target feature, and σ is the standard deviation of the target feature.

This method preserves variability between different emotions but eliminates baseline differences because all actors will have the same mean value = 0 and standard deviation = 1.

3.2.7 Data Preparation

The facial feature data extracted is organized in a time sequence. To convert this data into scalar values, the solution involves calculating the mean value of the time sequence of facial features extracted from each video. In this project, the mean value of the facial feature time sequence data from all videos is adopted to represent the facial features within each video which throws a lot of information away that can be preserved if the data points are presented on a time series.

3.3 Machine Learning Model Selection

According to the data section provided above, there are a total of 78 sets of data for all of the 5 single/pure emotions and a total of 826 sets of data for all of the 10 blended emotions. It’s clear that the dataset for training machine learning models is rather limited. With the available data being scarce, the choice of machine learning models must lean towards those capable of delivering satisfactory results with only a small amount of data. As a result, the necessary models should be simple, easy to train, and capable of performing well on classification tasks. Therefore, the final selection of models includes SVM, Decision Tree, and Naive Bayes.

3.4 Evaluation Metrics

Evaluation metrics serve as a method to measure the quality of the statistical or machine learning model. The confusion matrix is a very popular measure used while solving classification problems. It can be applied to binary classification as well as to multi-class classification problems [51]. A confusion matrix for binary classification is shown in table 3.4.1:

Table 3.4.1: Confusion matrix for binary classification.

		Predicted	
		Negative	Positive
Actual	Negative	TP	FN
	Positive	FP	TN

The label ”TN” corresponds to True Negative, representing the count of accurately classified negative examples. Likewise, ”TP” corresponds to True Positive, indicating the count of accurately classified positive examples. ”FP” refers to False Positive, representing the count of actual negative examples erroneously classified as positive. On the other hand, ”FN” stands for False Negative, indicating the count of actual positive examples mistakenly classified as negative [51].

Accuracy measures the ratio of correct predictions to the total predictions made and can be computed based on the confusion matrix [52]:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

Precision refers to the proportion of positive values among the total predicted positive instances, and it is defined by the following formula [52]:

$$Precision = \frac{TP}{TP + FP}$$

Recall refers to the proportion of negative values among the total predicted negative instances, and it is defined by the following formula [52]:

$$Recall = \frac{TN}{TN + FN}$$

The F1 score represents the harmonic mean of precision and recall, attributing significance to both factors:

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

In this project, accuracy and F1 score have been utilized to measure the classification capability of the machine learning models.

3.5 Ablation Study for Feature Importance

In ML, ablation refers to the process of eliminating a component from an ML system. An ablation study is conducted to assess the performance of a ML system by systematically removing specific components in order to comprehend the extent of contribution each component makes to the overall functionality of the system [53].

In this project, an ablation study is conducted to assess the importance of various facial features in emotion recognition [54][39]. Specifically, machine learning models were retrained and evaluated after removing a specific facial feature, aiming to investigate the importance of that feature in recognizing some given emotions.

Chapter 4

Results

4.1 Data and Data Processing Analysis

In this section, the computed results of Pearson correlation coefficients between AUs on the single/pure emotion dataset and blended emotion dataset will be presented.

4.1.1 Action Units

In order to explore the correlation coefficients between different Action Units, Pearson correlation coefficient matrices were separately calculated for both the single/pure emotion dataset and the blended emotion dataset. By comparing these two correlation coefficient matrices, it was found that the correlation coefficients between the two Action Units calculated from different datasets are similar. The Pearson correlation coefficient matrices for AUs in the single emotion dataset and the blended emotion dataset are shown in figure 4.1.1 and figure 4.1.2 respectively.

4.2 Machine Learning

This section will present the training and testing results of machine learning models on different datasets, and it will also present the generalizability testing results of the trained models. As mentioned in the previous section, Accuracy is used to evaluate the model's classification performance in multi-class scenarios, while F1 Score assesses the model's classification ability in binary classification tasks. All accuracy and F1

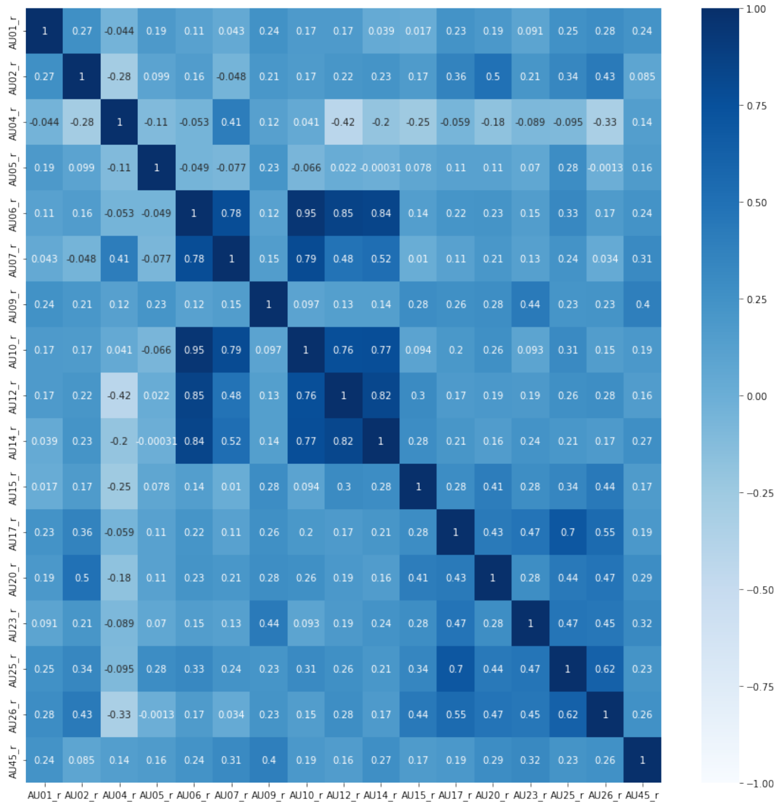

Figure 4.1.1: Pearson correlation coefficient matrix of action units of single/pure emotion dataset

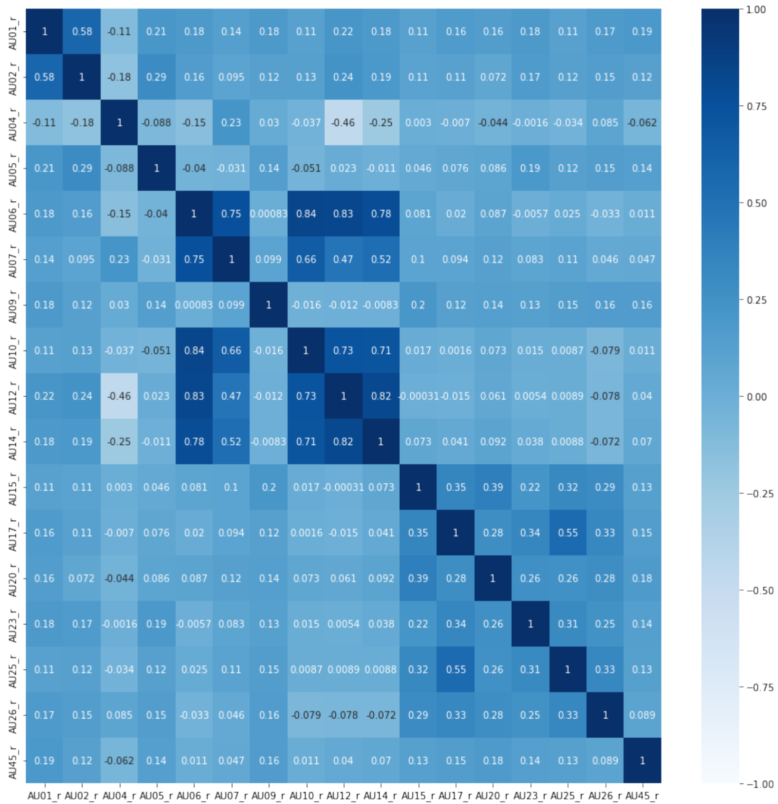

Figure 4.1.2: Pearson correlation coefficient matrix of action units blended emotion dataset

score values are calculated by averaging the results from cross-validation. However, to assess the models' generalization, models trained and tested on specific training and testing sets are also required. The test results of these models will also be shown in this section.

In this project, emotion recognition will be divided into two scenarios: multi-class classification and binary classification. In the case of multi-class classification, data will be directly labeled with their corresponding emotions. However, in the context of binary classification, the target emotion data will be labeled as 1, while the remaining data will be labeled as 0.

It's worth mentioning that all the results in this section are based on data that has undergone Z-score normalization.

4.2.1 Single/Pure Emotion Recognition

Table 4.2.1 and table 4.2.2 respectively represent the "cross-validated average results" and "holdout set test results" for single emotion recognition.

Emotion	SVM		Decision Tree		Naive Bayes	
	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score
Multi-Emotion	0.54		0.44		0.44	
Anger	0.79	0.00	0.70	0.16	0.60	0.36
Disgust	0.86	0.54	0.81	0.54	0.84	0.51
Fear	0.78	0.03	0.66	0.15	0.71	0.45
Happiness	0.99	0.97	0.95	0.89	0.95	0.90
Sadness	0.84	0.00	0.69	0.19	0.65	0.37

Table 4.2.1: Cross-validated aggregated results for single/pure emotion recognition

Emotion	SVM		Decision Tree		Naive Bayes	
	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score
Multi-Emotion	0.68		0.58		0.68	
Anger	0.79	0.00	0.79	0.33	0.68	0.40
Disgust	0.95	0.86	0.84	0.73	0.95	0.86
Fear	0.79	0.00	0.74	0.55	0.79	0.50
Happiness	1.00	1.00	0.95	0.89	1.00	1.00
Sadness	0.84	0.00	0.63	0.22	0.74	0.29

Table 4.2.2: Holdout set test results for single/pure emotion recognition

Among them, the models in Table 4.2.2 exhibit emotion recognition capabilities that surpass the cross-validated average. They will be utilized for recognizing blended emotions.

4.2.2 Blended Emotion Recognition

Table 4.2.3 and Table 4.2.4 respectively represent the "cross-validated average results" and "holdout set test results" for blended emotion recognition. Among them, the models in Table 4.2.4 exhibit emotion recognition capabilities that basically equal the cross-validated average. They will be utilized for recognizing blended emotions.

Emotion	SVM		Decision Tree		Naive Bayes	
	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score
Multi-Emotion	0.35		0.22		0.28	
Anger & Disgust	0.90	0.00	0.84	0.25	0.80	0.19
Anger & Fear	0.90	0.01	0.84	0.29	0.72	0.32
Anger & Happiness	0.91	0.00	0.83	0.11	0.81	0.22
Anger & Sadness	0.90	0.00	0.83	0.13	0.82	0.25
Disgust & Fear	0.91	0.00	0.85	0.22	0.71	0.24
Disgust & Happiness	0.92	0.32	0.88	0.33	0.83	0.42
Disgust & Sadness	0.90	0.00	0.83	0.19	0.77	0.27
Fear & Happiness	0.89	0.00	0.84	0.29	0.83	0.33
Fear & Sadness	0.90	0.00	0.82	0.16	0.69	0.29
Happiness & Sadness	0.90	0.19	0.85	0.35	0.83	0.39

Table 4.2.3: Cross-validated aggregated results for blended emotion recognition

Emotion	SVM		Decision Tree		Naïve Bayes	
	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score
Multi-Emotion	0.39		0.26		0.30	
Anger & Disgust	0.90	0.00	0.83	0.12	0.84	0.36
Anger & Fear	0.90	0.00	0.82	0.17	0.69	0.33
Anger & Happiness	0.91	0.00	0.84	0.12	0.78	0.18
Anger & Sadness	0.91	0.00	0.77	0.13	0.76	0.09
Disgust & Fear	0.91	0.00	0.91	0.43	0.74	0.31
Disgust & Happiness	0.93	0.40	0.87	0.21	0.82	0.46
Disgust & Sadness	0.90	0.00	0.84	0.18	0.82	0.29
Fear & Happiness	0.89	0.00	0.84	0.31	0.81	0.27
Fear & Sadness	0.89	0.00	0.79	0.05	0.69	0.31
Happiness & Sadness	0.88	0.00	0.86	0.33	0.85	0.48

Table 4.2.4: Holdout set test results for blended emotion recognition

4.2.3 Model Generalizability Results

When testing a single/pure emotion recognition model on a blended emotion dataset, for the multi-class classification scenario, accurate recognition of any component emotion of a blended emotion is considered accurate recognition, whereas failure to do so is considered inaccurate recognition. For instance, if models recognize the blended emotion ”anger & disgust” as either ”anger” or ”disgust,” it is considered accurate recognition. However, if it is recognized as any other emotion, it would be considered inaccurate recognition.

For the binary classification scenario, within the blended emotion dataset, instances, where the component emotions contain the target emotion, will be labeled as 1, while instances where the target emotion is not present will be labeled as 0. Table 4.2.5 represents the generalizability results of single/pure emotion recognition models.

Emotion	SVM		Decision Tree		Naïve Bayes	
	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score
Multi-Emotion	0.64		0.60		0.59	
Anger	0.61	0.00	0.59	0.23	0.59	0.44
Disgust	0.64	0.16	0.60	0.50	0.65	0.38
Fear	0.79	0.00	0.62	0.36	0.65	0.46
Happiness	0.76	0.60	0.79	0.67	0.77	0.65
Sadness	0.59	0.00	0.57	0.34	0.57	0.39

Table 4.2.5: Generalizability results of single/pure emotion recognition models

When testing blended emotion recognition models on the single emotion dataset, correctly recognizing a single emotion data instance as a blended emotion containing that emotion is considered accurate recognition for the multi-class classification scenario. If it is recognized as any other blended emotion, it would be considered inaccurate recognition. For example, accurately recognizing an ”anger” data instance as ”anger & disgust,” ”anger & fear,” ”anger & happiness,” or ”anger & sadness” would be considered accurate recognition, while recognizing it as any other blended emotion would be considered inaccurate recognition.

For the binary classification scenario, within the single-emotion dataset, instances that are component emotions of the target blended emotion will be labeled as 1, while other instances will be labeled as 0. Table 4.2.6 represents the generalizability results of blended emotion recognition models.

4.3 Z-Score Normalization

To examine the impact of applying Z-Score normalization to the data on the machine learning model’s recognition performance, a comparison is made between models

Emotion	SVM		Decision Tree		Naive Bayes	
	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score
Multi-Emotion	0.79		0.62		0.65	
Anger & Disgust	0.59	0.00	0.62	0.29	0.62	0.35
Anger & Fear	0.58	0.00	0.72	0.52	0.74	0.69
Anger & Happiness	0.59	0.00	0.60	0.16	0.77	0.61
Anger & Sadness	0.63	0.00	0.59	0.11	0.67	0.28
Disgust & Fear	0.58	0.00	0.54	0.18	0.45	0.30
Disgust & Happiness	0.69	0.40	0.74	0.55	0.83	0.75
Disgust & Sadness	0.63	0.00	0.68	0.39	0.69	0.40
Fear & Happiness	0.58	0.00	0.63	0.36	0.71	0.47
Fear & Sadness	0.62	0.00	0.67	0.41	0.69	0.62
Happiness & Sadness	0.63	0.00	0.67	0.32	0.85	0.48

Table 4.2.6: Generalizability results of blended emotion recognition models

trained using the original data and the data normalized with Z-Score. In this project, accuracy is used to evaluate the models' classification ability in multi-class problems, while both accuracy and F1 score are employed to measure the models' classification capability in binary classification problems. The accuracy and F1 score results for single/pure emotion recognition and blended emotion recognition with data without implementing Z-Score normalization are shown in Table 4.3.1 and Table 4.3.2 respectively and obtained by averaging across cross-validation which is chosen to mitigate issues like overfitting and selection bias.

Emotion	SVM		Decision Tree		Naive Bayes	
	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score
Multi-Emotion	0.53		0.40		0.45	
Anger	0.79	0.00	0.64	0.12	0.56	0.34
Disgust	0.83	0.27	0.77	0.43	0.74	0.19
Fear	0.79	0.00	0.74	0.35	0.80	0.64
Happiness	0.98	0.97	0.97	0.93	0.98	0.97
Sadness	0.84	0.00	0.71	0.19	0.55	0.26

Table 4.3.1: Cross-validated aggregated results for single/pure emotion recognition with data before implementing Z-Score normalization

Table 4.3.3 and Table 4.3.4 show the "holdout set test results" of single/pure emotion recognition and blended emotion recognition with data without implementing Z-Score normalization.

Emotion	SVM		Decision Tree		Naive Bayes	
	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score
Multi-Emotion	0.30		0.22		0.25	
Anger & Disgust	0.90	0.00	0.84	0.24	0.68	0.37
Anger & Fear	0.90	0.00	0.84	0.16	0.69	0.29
Anger & Happiness	0.91	0.00	0.84	0.18	0.81	0.17
Anger & Sadness	0.90	0.00	0.82	0.17	0.82	0.14
Disgust & Fear	0.91	0.00	0.85	0.22	0.81	0.11
Disgust & Happiness	0.91	0.00	0.87	0.28	0.82	0.41
Disgust & Sadness	0.90	0.00	0.83	0.18	0.73	0.20
Fear & Happiness	0.89	0.00	0.83	0.21	0.82	0.32
Fear & Sadness	0.90	0.00	0.82	0.19	0.67	0.24
Happiness & Sadness	0.89	0.03	0.84	0.30	0.83	0.34

Table 4.3.2: Cross-validated aggregated results for blended emotion recognition with data before implementing Z-Score normalization

Emotion	SVM		Decision Tree		Naive Bayes	
	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score
Multi-Emotion	0.58		0.42		0.37	
Anger	0.79	0.00	0.79	0.50	0.68	0.40
Disgust	0.79	0.00	0.84	0.40	0.84	0.57
Fear	0.79	0.00	0.79	0.33	0.63	0.22
Happiness	1.00	1.00	1.00	1.00	1.00	1.00
Sadness	0.84	0.00	0.58	0.00	0.47	0.17

Table 4.3.3: Holdout set test results for single/pure emotion recognition with data before implementing Z-Score normalization

4.4 Dominant Emotion Exploration

To explore whether a dominant emotion exists among the component emotions of blended emotion, which implies that there is a predominant emotion present during the expression of a blended emotion, an analysis was conducted on the recognition results of the single/pure emotion recognition models on the blended emotion dataset. Specifically, for each blended emotion, the analysis will involve analyzing the distribution of recognition results for instances that are accurately recognized. For example, in the case of ”anger & disgust”, the percentages of instances accurately identified as ’anger’ and ’disgust’ will be calculated. Through analyzing the proportions of the different component emotions, exploration will be conducted to determine whether a dominant emotion exists. Table 4.4.1, Table 4.4.2 and Table 4.4.3 respectively present the distribution of component emotions in the recognition results obtained by employing a single/pure emotion recognition SVM model on blended

Emotion	SVM		Decision Tree		Naive Bayes	
	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score
Multi-Emotion	0.25		0.24		0.22	
Anger & Disgust	0.90	0.00	0.83	0.26	0.75	0.30
Anger & Fear	0.90	0.00	0.85	0.13	0.68	0.32
Anger & Happiness	0.91	0.00	0.83	0.17	0.77	0.13
Anger & Sadness	0.91	0.00	0.81	0.11	0.81	0.11
Disgust & Fear	0.91	0.00	0.85	0.29	0.86	0.20
Disgust & Happiness	0.91	0.00	0.87	0.48	0.79	0.41
Disgust & Sadness	0.90	0.00	0.84	0.07	0.78	0.14
Fear & Happiness	0.89	0.00	0.85	0.24	0.84	0.37
Fear & Sadness	0.89	0.00	0.84	0.12	0.71	0.27
Happiness & Sadness	0.89	0.00	0.82	0.06	0.81	0.23

Table 4.3.4: Holdout set test results for blended emotion recognition with data before implementing Z-Score normalization

emotion datasets under emotion blending ratios of 50:50, 30:70, and 70:30." Similarly, Table 4.4.4, Table 4.4.5, and Table 4.4.6, as well as Table 4.4.7, Table 4.4.8, and Table 4.4.9, respectively demonstrate the distribution of recognition results obtained by the Decision Tree and Naive Bayes models under these three scenarios.

4.5 Ablation Study for Feature Importance

An ablation study is conducted to assess the importance of each facial feature in emotion recognition. Specifically, 17 action units will be successively removed, and then reinserted, with the data modified after each removal used for training a machine learning model. The test results of the trained models will be compared to those of models trained without the removal of any action units. This comparison aims to assess the importance of the action unit that has been removed.

Based on the obtained comparison results, the 6 action units with the most significant impact on the test results for each model will be selected. These will include the top 3 action units that contribute to the highest increase in F1 score and the top 3 action units that lead to the greatest decrease in F1 score. However, if it's not possible to identify the top 3 action units that contribute to the highest increase in F1 score and the top three action units that lead to the greatest decrease in F1 score, then selections will be made based on what's available. Depending on the circumstances, this could involve choosing the top 2, 1, or even none. Through the aforementioned steps, the important action units for each model have been collected for the 5 single emotions

True Emotion	Pred_Emotion	Accuracy	Predicted Rate
Anger & Disgust	Anger Disgust	0.57	0.25 0.75
Anger & Fear	Anger Fear	0.81	0.19 0.81
Anger & Happiness	Anger Happiness	0.54	0.14 0.86
Anger & Sadness	Anger Sadness	0.19	0.4 0.6
Disgust & Fear	Disgust Fear	0.64	0.44 0.56
Disgust & Happiness	Disgust Happiness	0.86	0.32 0.68
Disgust & Sadness	Disgust Sadness	0.67	0.72 0.28
Fear & Happiness	Fear Happiness	0.54	0.4 0.6
Fear & Sadness	Fear Sadness	0.76	0.86 0.14
Happiness & Sadness	Happiness Sadness	0.59	0.94 0.06

Table 4.4.1: Blended emotion (blend ratio:5050) recognition result of SVM single/pure emotion recognition model

True Emotion	Pred_Emotion	Accuracy	Predicted Rate
Anger & Disgust	Anger Disgust	0.65	0.06 0.94
Anger & Fear	Anger Fear	0.86	0.21 0.79
Anger & Happiness	Anger Happiness	0.73	0.16 0.84
Anger & Sadness	Anger Sadness	0.36	0.7 0.3
Disgust & Fear	Disgust Fear	0.76	0.32 0.68
Disgust & Happiness	Disgust Happiness	1	0.2 0.8
Disgust & Sadness	Disgust Sadness	0.63	0.94 0.06
Fear & Happiness	Fear Happiness	0.83	0 1
Fear & Sadness	Fear Sadness	0.5	0.77 0.23
Happiness & Sadness	Happiness Sadness	0.31	0.8 0.2

Table 4.4.2: Blended emotion (blend ratio:3070) recognition result of SVM single/pure emotion recognition model

True Emotion	Pred_Emotion	Accuracy	Predicted Rate
Anger & Disgust	Anger Disgust	0.61	0.53 0.47
Anger & Fear	Anger Fear	0.89	0.28 0.72
Anger & Happiness	Anger Happiness	0.38	0.16 0.84
Anger & Sadness	Anger Sadness	0.23	0.67 0.33
Disgust & Fear	Disgust Fear	0.6	0.67 0.33
Disgust & Happiness	Disgust Happiness	0.96	0.65 0.35
Disgust & Sadness	Disgust Sadness	0.76	0.77 0.23
Fear & Happiness	Fear Happiness	0.52	0.81 0.19
Fear & Sadness	Fear Sadness	0.68	0.9 0.1
Happiness & Sadness	Happiness Sadness	0.87	0.96 0.04

Table 4.4.3: Blended emotion (blend ratio:7030) recognition result of SVM single/pure emotion recognition model

True Emotion	Pred_Emotion	Accuracy	Predicted Rate
Anger & Disgust	Anger Disgust	0.36	0.3 0.7
Anger & Fear	Anger Fear	0.62	0.44 0.56
Anger & Happiness	Anger Happiness	0.35	0.22 0.78
Anger & Sadness	Anger Sadness	0.38	0.3 0.7
Disgust & Fear	Disgust Fear	0.5	0.5 0.5
Disgust & Happiness	Disgust Happiness	0.93	0.44 0.56
Disgust & Sadness	Disgust Sadness	0.59	0.38 0.62
Fear & Happiness	Fear Happiness	0.21	0.17 0.83
Fear & Sadness	Fear Sadness	0.76	0.5 0.5
Happiness & Sadness	Happiness Sadness	0.72	0.76 0.24

Table 4.4.4: Blended emotion (blend ratio:5050) recognition result of Decision Tree single/pure emotion recognition model

True Emotion	Pred_Emotion	Accuracy	Predicted Rate
Anger & Disgust	Anger	0.58	0.4
	Disgust		0.6
Anger & Fear	Anger	0.57	0.38
	Fear		0.62
Anger & Happiness	Anger	0.62	0.12
	Happiness		0.88
Anger & Sadness	Anger	0.54	0.13
	Sadness		0.87
Disgust & Fear	Disgust	0.52	0.23
	Fear		0.77
Disgust & Happiness	Disgust	0.96	0.21
	Happiness		0.79
Disgust & Sadness	Disgust	0.67	0.61
	Sadness		0.39
Fear & Happiness	Fear	0.72	0
	Happiness		1
Fear & Sadness	Fear	0.69	0.39
	Sadness		0.61
Happiness & Sadness	Happiness	0.53	0.35
	Sadness		0.65

Table 4.4.5: Blended emotion (blend ratio:3070) recognition result of Decision Tree single/pure emotion recognition model

True Emotion	Pred_Emotion	Accuracy	Predicted Rate
Anger & Disgust	Anger	0.46	0.69
	Disgust		0.31
Anger & Fear	Anger	0.68	0.37
	Fear		0.63
Anger & Happiness	Anger	0.38	0.5
	Happiness		0.5
Anger & Sadness	Anger	0.62	0.44
	Sadness		0.56
Disgust & Fear	Disgust	0.52	0.31
	Fear		0.69
Disgust & Happiness	Disgust	0.96	0.65
	Happiness		0.35
Disgust & Sadness	Disgust	0.72	0.57
	Sadness		0.43
Fear & Happiness	Fear	0.32	0.6
	Happiness		0.4
Fear & Sadness	Fear	0.74	0.61
	Sadness		0.39
Happiness & Sadness	Happiness	0.87	0.92
	Sadness		0.08

Table 4.4.6: Blended emotion (blend ratio:7030) recognition result of Decision Tree single/pure emotion recognition model

True Emotion	Pred_Emotion	Accuracy	Predicted Rate
Anger & Disgust	Anger Disgust	0.39	0.45 0.55
Anger & Fear	Anger Fear	0.65	0.41 0.59
Anger & Happiness	Anger Happiness	0.42	0 1
Anger & Sadness	Anger Sadness	0.42	0.27 0.73
Disgust & Fear	Disgust Fear	0.36	0.9 0.1
Disgust & Happiness	Disgust Happiness	0.93	0.37 0.63
Disgust & Sadness	Disgust Sadness	0.7	0.63 0.37
Fear & Happiness	Fear Happiness	0.36	0.3 0.7
Fear & Sadness	Fear Sadness	0.76	0.77 0.23
Happiness & Sadness	Happiness Sadness	0.55	0.88 00.12

Table 4.4.7: Blended emotion (blend ratio:5050) recognition result of Naive Bayes single/pure emotion recognition model

True Emotion	Pred_Emotion	Accuracy	Predicted Rate
Anger & Disgust	Anger Disgust	0.5	0.23 0.77
Anger & Fear	Anger Fear	0.75	0.19 0.81
Anger & Happiness	Anger Happiness	0.65	0.06 0.94
Anger & Sadness	Anger Sadness	0.46	0.54 0.46
Disgust & Fear	Disgust Fear	0.68	0.41 0.59
Disgust & Happiness	Disgust Happiness	1	0.16 0.84
Disgust & Sadness	Disgust Sadness	0.81	0.68 0.32
Fear & Happiness	Fear Happiness	0.76	0 1
Fear & Sadness	Fear Sadness	0.35	0.44 0.56
Happiness & Sadness	Happiness Sadness	0.34	0.45 0.55

Table 4.4.8: Blended emotion (blend ratio:3070) recognition result of Naive Bayes single/pure emotion recognition model

True Emotion	Pred_Emotion	Accuracy	Predicted Rate
Anger & Disgust	Anger Disgust	0.43	0.5 0.5
Anger & Fear	Anger Fear	0.71	0.65 0.35
Anger & Happiness	Anger Happiness	0.31	0.38 0.62
Anger & Sadness	Anger Sadness	0.54	0.71 0.29
Disgust & Fear	Disgust Fear	0.4	0.8 0.2
Disgust & Happiness	Disgust Happiness	0.96	0.52 0.48
Disgust & Sadness	Disgust Sadness	0.86	0.52 0.48
Fear & Happiness	Fear Happiness	0.23	0.71 0.29
Fear & Sadness	Fear Sadness	0.61	0.58 0.42
Happiness & Sadness	Happiness Sadness	0.83	0.96 0.04

Table 4.4.9: Blended emotion (blend ratio:7030) recognition result of Naive Bayes single/pure emotion recognition model

and 10 blended emotions. Summarizing the obtained data in tables results in 15 tables, corresponding to 5 single/pure emotions and 10 blended emotions.

In the tables, the symbol ”O” indicates that the specific action unit is among the top 3 action units whose removal would result in the greatest increase in F1 score for that specific model. The symbol ”X” indicates that the specific action unit is among the top 3 action units whose removal would result in the greatest decrease in F1 score for that specific model. And symbol ”-” indicates the specific action unit is not among the 6 most important action units for that specific model. Those feature importance tables of single/pure emotions are shown as follows:

The feature importance tables for blended emotions are shown as follows:

Removal	SVM	Decision Tree	Naive Bayes
Remove AU_02	–	O	–
Remove AU_05	–	–	O
Remove AU_06	–	–	X
Remove AU_10	–	–	X
Remove AU_14	–	–	O
Remove AU_20	–	O	–
Remove AU_23	–	O	X
Remove AU_25	–	–	O
Remove AU_26	–	X	–

Table 4.5.1: Important action units of Anger

Removal	SVM	Decision Tree	Naive Bayes
Remove AU_01	–	–	O
Remove AU_04	X	O	–
Remove AU_05	–	–	O
Remove AU_07	O	X	–
Remove AU_09	–	X	–
Remove AU_10	X	X	X
Remove AU_14	X	O	–
Remove AU_17	O	–	O
Remove AU_20	–	–	X
Remove AU_25	–	O	X
Remove AU_45	O	–	–

Table 4.5.2: Important action units of Disgust

Removal	SVM	Decision Tree	Naive Bayes
Remove AU_01	–	–	O
Remove AU_04	O	X	–
Remove AU_05	–	–	O
Remove AU_06	O	X	–
Remove AU_07	–	O	–
Remove AU_09	–	O	–
Remove AU_10	–	–	X
Remove AU_12	O	X	X
Remove AU_14	X	O	X
Remove AU_25	–	–	O
Remove AU_26	X	–	–

Table 4.5.3: Important action units of Fear

Removal	SVM	Decision Tree	Naive Bayes
Remove AU_02	–	O	–
Remove AU_05	–	–	O
Remove AU_06	–	–	O
Remove AU_09	–	–	X
Remove AU_12	X	X	X
Remove AU_14	X	–	–
Remove AU_17	–	X	–
Remove AU_20	–	X	–
Remove AU_26	–	O	–
Remove AU_25	–	O	–
Remove AU_45	–	–	X

Table 4.5.4: Important action units of Happiness

Removal	SVM	Decision Tree	Naive Bayes
Remove AU_01	–	X	O
Remove AU_04	–	–	O
Remove AU_05	–	–	X
Remove AU_06	–	–	X
Remove AU_07	–	X	–
Remove AU_09	–	–	O
Remove AU_10	–	–	X
Remove AU_14	–	X	–
Remove AU_20	–	O	–
Remove AU_23	–	O	–
Remove AU_25	–	O	–

Table 4.5.5: Important action units of Sadness

Removal	SVM	Decision Tree	Naive Bayes
Remove AU_01	–	–	O
Remove AU_02	–	O	–
Remove AU_10	–	X	–
Remove AU_12	O	–	–
Remove AU_15	–	O	O
Remove AU_20	–	O	O
Remove AU_25	–	X	–
Remove AU_45	O	X	–

Table 4.5.6: Important action units of Anger & Disgust

Removal	SVM	Decision Tree	Naive Bayes
Remove AU_01	–	–	O
Remove AU_02	–	–	X
Remove AU_06	–	–	O
Remove AU_07	–	X	–
Remove AU_10	–	–	X
Remove AU_14	–	–	O
Remove AU_17	–	–	X
Remove AU_20	–	X	–
Remove AU_25	–	X	–

Table 4.5.7: Important action units of Anger & Fear

Removal	SVM	Decision Tree	Naive Bayes
Remove AU_01	–	–	X
Remove AU_02	–	O	–
Remove AU_04	–	X	–
Remove AU_05	–	X	–
Remove AU_06	–	–	X
Remove AU_12	–	–	X
Remove AU_14	–	X	–
Remove AU_17	–	O	–
Remove AU_26	–	O	–

Table 4.5.8: Important action units of Anger & Happiness

Removal	SVM	Decision Tree	Naive Bayes
Remove AU_01	–	–	X
Remove AU_04	–	O	–
Remove AU_07	–	–	X
Remove AU_09	–	X	–
Remove AU_10	–	O	–
Remove AU_12	–	–	X
Remove AU_15	–	O	–

Table 4.5.9: Important action units of Anger & Sadness

Removal	SVM	Decision Tree	Naive Bayes
Remove AU_02	–	–	O
Remove AU_04	–	O	–
Remove AU_07	–	O	X
Remove AU_10	–	X	–
Remove AU_12	–	X	X
Remove AU_14	–	–	O
Remove AU_17	–	X	X
Remove AU_20	–	–	O
Remove AU_23	–	O	–

Table 4.5.10: Important action units of Disgust & Fear

Removal	SVM	Decision Tree	Naive Bayes
Remove AU_02	O	–	–
Remove AU_04	–	O	–
Remove AU_05	–	O	–
Remove AU_07	–	–	O
Remove AU_09	O	X	–
Remove AU_10	X	X	X
Remove AU_12	–	–	O
Remove AU_14	–	O	–
Remove AU_15	–	–	O
Remove AU_17	O	–	–
Remove AU_20	–	–	X
Remove AU_23	X	–	–
Remove AU_26	–	X	X
Remove AU_45	X	–	–

Table 4.5.11: Important action units of Disgust & Happiness

Removal	SVM	Decision Tree	Naive Bayes
Remove AU_01	–	X	O
Remove AU_02	–	–	X
Remove AU_04	–	–	X
Remove AU_05	–	–	O
Remove AU_06	–	O	–
Remove AU_09	–	X	O
Remove AU_10	–	O	–
Remove AU_20	–	X	–
Remove AU_23	–	O	–
Remove AU_25	–	–	X

Table 4.5.12: Important action units of Disgust & Sadness

Removal	SVM	Decision Tree	Naive Bayes
Remove AU_01	–	O	–
Remove AU_02	–	X	X
Remove AU_04	–	–	X
Remove AU_05	O	–	–
Remove AU_07	–	–	O
Remove AU_09	O	O	–
Remove AU_12	–	X	X
Remove AU_14	–	–	O
Remove AU_20	–	X	O
Remove AU_23	O	–	–
Remove AU_25	–	O	–

Table 4.5.13: Important action units of Fear & Happiness

Removal	SVM	Decision Tree	Naive Bayes
Remove AU_01	–	–	O
Remove AU_02	–	O	–
Remove AU_05	–	O	–
Remove AU_06	–	X	O
Remove AU_09	–	–	X
Remove AU_14	–	–	O
Remove AU_23	–	O	–
Remove AU_26	–	X	–
Remove AU_45	–	X	–

Table 4.5.14: Important action units of Fear & Sadness

Removal	SVM	Decision Tree	Naive Bayes
Remove AU_01	–	–	O
Remove AU_02	–	–	O
Remove AU_04	–	–	X
Remove AU_06	O	–	–
Remove AU_07	X	–	X
Remove AU_09	O	O	–
Remove AU_12	X	X	–
Remove AU_17	X	X	X
Remove AU_20	–	–	O
Remove AU_23	–	X	–
Remove AU_45	O	–	–

Table 4.5.15: Important action units of Happiness & Sadness

Chapter 5

Discussion

5.1 Major Findings

5.1.1 Emotion Recognition Ability

For single/pure emotion recognition, there are 5 distinct single/pure emotions, so the accuracy for random recognition is 0.2. Based on the aggregated cross-validation results for single/pure emotion recognition from Table 4.2.1 and the holdout set test results of Table 4.2.2, it can be observed that the accuracy of each model in multi-emotion recognition is higher than 0.2 which means the trained models have achieved test results worthy of further scrutiny. This implies that the trained machine learning models are capable of recognizing single/pure emotions with acceptable or at least near-acceptable accuracy.

For blended emotion recognition, there are 10 distinct emotions, so the accuracy for random recognition is 0.1. Based on the aggregated cross-validation results for blended emotion recognition from Table 4.2.3 and the holdout set test results of Table 4.2.4, it can be observed that the accuracy of each model in multi-emotion recognition is higher than 0.1 which means the trained models have achieved likewise noteworthy test results.

5.1.2 Z-Score Normalization

Comparing the holdout set test results between Table 4.3.3 and Table 4.2.2, in the context of single/pure emotion recognition, for the specific training set and test set pair, applying Z-score normalization to the data makes the data more suitable for ML models to classify in multi-emotion classification tasks. Additionally, in binary classification tasks, certain emotions become more easily classifiable, notably the ”Disgust” emotion data which exhibits a significant enhancement in classification capability. While the Decision Tree model achieved slightly better results in recognizing anger and happiness without applying Z-score normalization, the difference is not substantial. Moreover, this outcome could also be attributed to the stochastic parameters in the Decision Tree model from the scikit-learn package.

And comparing the holdout set test results between Table 4.3.4 and Table 4.2.4, in the context of blended emotion recognition, for the specific training set and test set pair, applying Z-score normalization makes the data more suitable for machine learning models to classify in multi-emotion classification tasks. However, in binary classification tasks, applying Z-score normalization only slightly improves the classifiability of certain blended emotions by the models. Meanwhile, it can lead to a decrease in classification capability for some blended emotions. Overall, based on the content of these two tables, applying Z-score normalization makes the data more amenable to classification.

To comprehensively investigate the impact of Z-score normalization on the data's classification capability, an analysis will be performed using the cross-validation aggregated results of model testing before and after applying Z-score normalization to the dataset.

Based on the aggregated cross-validation results for single/pure emotion recognition from Table 4.3.1 and Table 4.2.1, it can be observed that in multi-emotion classification tasks, Z-score normalization tends to slightly improve the data's classification capability by SVM and Decision Tree models. However, the recognition capability of data by the Naive Bayes model in multi-emotion classification tasks shows a minor decline. Due to the minimal nature of these improvements and declines, it can be inferred that applying Z-score normalization to the data has a generally small impact on the data's classification capability in multi-emotion classification tasks under single/pure emotion recognition scenarios. In binary classification tasks, the impact of applying Z-score normalization to the data is more pronounced. After applying Z-score normalization, the recognition capability of the ”Disgust” emotion by the model is significantly enhanced, which aligns with the observations from the specific training and test set pair. However, the recognition capability of the ”Fear” emotion by the model experiences a notable decline, which differs from the results observed on the specific training and test set pair. Apart from the mentioned two emotions, for other emotions, applying Z-score normalization to the data generally does not significantly affect their recognition capability by machine learning models.

Based on the aggregated cross-validation results for blended emotion recognition from Table 4.3.2 and Table 4.2.3, it can be observed that in multi-emotion classification tasks, the application of Z-score normalization enhances the data's classification capability by SVM and Naive Bayes models, while the classification capability by the Decision Tree model remains unchanged. In binary classification tasks, in most cases, applying Z-score normalization to the data usually enhances the data's recognition capability by the models. However, in certain instances, applying Z-score normalization to the data can lead to a decrease in the models' recognition capability for certain blended emotions. Specifically, for the Decision Tree model, applying Z-score normalization results in a decline in its recognition capability for three blended emotions: Anger & Happiness, Anger & Sadness, and Fear & Sadness. Nevertheless, due to the relatively small magnitude of the decline, this situation could also be attributed to the randomness in the Decision Tree model. Simultaneously, for the Naive Bayes model, applying Z-score normalization to the data leads to a more noticeable decrease in its recognition capability for the Anger & Disgust mixed emotion.

Overall, applying Z-score normalization to the data generally makes the data more amenable to machine learning classification in most cases. However, in certain instances, the opposite might occur, but these cases do not negate the overall improvement of the data.

5.1.3 Model Generalizability

According to our definition of accurate recognition of blended emotions by single emotion recognition models in Section 4.2.3, the probability of randomly recognizing a blended emotion as a specific single/pure emotion and accurately recognizing it is 0.4. According to Table 4.2.5, the test results of single emotion recognition models on blended emotion data show that in the case of multi-emotion classification, all models achieve emotion recognition accuracy exceeding 0.4 which means trained models can achieve higher accuracy than random recognition accuracy on blended emotion data for multi-emotion classification problems. In binary classification tasks, the Naive Bayes model also achieved an acceptable test result.

Based on the above analysis, it can be inferred that blended emotions' expressions at least possess some important characteristics of their constituent emotions' expressions. According to our definition of accurate recognition of single/pure emotions by blended emotion recognition models in Section 4.2.3, the probability of randomly recognizing a single/pure emotion as a specific blended emotion and accurately recognizing it is 0.4. According to Table 4.2.6, the test results of blended emotion recognition models on single/pure emotion data show that in the case of multi-emotion classification, all models achieve emotion recognition accuracy exceeding 0.4. SVM even achieves an impressive accuracy of 0.79 on single/pure emotion data. In binary classification tasks, both the Decision Tree model and the Naive Bayes model achieved good test results.

Based on the conclusion in Section 5.1.1, machine learning models only achieve acceptable test results in blended emotion recognition. However, those models can achieve high recognition accuracy on single emotion data. Therefore, based on the above analysis, it can be concluded that blended emotion expressions do not possess unique characteristic patterns that are absent in their constituent emotions.

Overall, based on the analysis and conclusions, blended emotion expressions are highly likely to result from the overlapping combinations of features from constituent emotions or the combination of certain features from one constituent emotion with certain features from another constituent emotion.

5.1.4 Dominant Emotion Exploration

Emotion Blend Ratio

Before investigating the presence of dominant emotions in blended emotions, it is important to validate whether the emotion blend ratio of emotions in the original blended emotion data is accurate. In theory, if the emotion blend ratio is changed, the proportion of a certain blended emotion being recognized as its constituent emotion should also change accordingly. For instance, consider the blended emotion ”anger & disgust.” Compared to the case where the emotion blend ratio is 50:50, when the emotion blend ratio becomes 30:70, in cases where the blended emotion ”anger & disgust” is accurately recognized, there should be a higher proportion of cases recognized as ”disgust” and a lower proportion of cases recognized as ”anger.” Similarly, when the emotion blend ratio changes to 30:70, the proportion of cases recognized as ”disgust” should decrease, while the proportion of cases recognized as ”anger” should increase.

For the SVM model, according to Table 4.4.1, Table 4.4.2 and Table 4.4.3, it can be observed that changing the emotion blend ratio of blended emotions results in corresponding changes in the proportions of blended emotions being recognized as their constituent emotions in the accurate recognition results in most cases. However, when the emotion blend ratio changes from 50:50 to 30:70, the changes in the proportions of ”anger & sadness” and ”disgust & sadness” being recognized as their constituent emotion are opposite to what would be expected. Based on the changes in the emotion blend ratio, in cases where these two blended emotions are successfully recognized, the proportion of cases recognized as ”sadness” should increase compared to when the emotion blend ratio is 50:50. However, the observed results are in the opposite direction. In reality, there is a higher proportion of cases recognized as ”anger” and ”disgust.”

For the Decision Tree model, according to Table 4.4.4, Table 4.4.5, Table 4.4.6, it can be observed that in most cases, if the emotion blend ratio changes, the proportion of a certain blended emotion being recognized as its constituent emotions also changes accordingly. However, there are also cases that are opposite to what would be expected. When the emotion blend ratio changes from 50:50 to 70:30, the proportion of cases where ”anger & fear” and ”disgust & fear” are recognized as ”fear” should decrease. However, in reality, the proportion of cases where ”anger & fear” and ”disgust & fear” are recognized as ”fear” has actually increased. When the emotion blend ratio changes from 50:50 to 30:70, the proportion of cases where ”anger & disgust” is recognized as ”disgust” and ”disgust & sadness” is recognized as ”sadness” should increase. But the proportion of cases where ”anger & disgust” is recognized as ”disgust” and ”disgust & sadness” is recognized as ”sadness” actually decreased.

For the Naive Bayes model, according to Table 4.4.7, Table 4.4.8, Table 4.4.9, it is evident that in the majority of cases, when there is a change in the emotion blend ratio, the proportion of a specific blended emotion being recognized as its constituent emotions also tends to change correspondingly. However, there are also some cases that are opposite to what would be expected. When the emotion blend ratio changes from 50:50 to 70:30, the proportion of cases where ”disgust & fear” is recognized as ”disgust”, ”disgust & sadness” is recognized as ”disgust”, and ”fear & sadness” is recognized as ”fear” should increase but the actual results show a decrease instead. When the emotion blend ratio changes from 50:50 to 30:70, the proportion of cases where ”anger & sadness” and ”disgust & sadness” are recognized as ”sadness” should increase but the results show a decrease instead.

Based on the analysis above, across all three models, it has been observed that in most cases, when the emotion blend ratio changes, the proportion of a certain blended emotion being recognized as its constituent emotions also changes accordingly. Therefore, the annotated emotion blend ratios on the blended emotion dataset have been verified to be correct.

At the same time, it was also observed across all three models that when the emotion blend ratio shifts from 50:50 to 30:70, the proportion of instances that ”disgust & sadness” is recognized as ”sadness” decreased, contrary to the expected increase. This either implies that ”sadness” itself is easily misclassified as ”disgust,” or it raises doubts about the accuracy of the labeled emotion blend ratios of ”disgust & sadness” in the original data.

Dominant Emotion

In the context where the accuracy of emotion blend ratio annotations has been confirmed, the analysis will be centered on scenarios with emotion blend ratios of 50:50. This analysis aims to determine whether a dominant component emotion exists in blended emotions. To ensure the validity of the analysis and eliminate potential confounding factors, two premises need to be met. Firstly, the model’s recognition accuracy for blended emotions should be above a certain threshold that is defined as acceptable, set at 60%. Secondly, the model’s ability to recognize the component emotions of mixed emotions should be relatively consistent. In the multi-emotion recognition scenario, the accuracy of each model in recognizing each single/pure emotion is shown in the following table 5.1.1:

Based on Table 5.1.1, Table 4.4.1, Table 4.4.4 and Table 4.4.7, only ”disgust & happiness” satisfies the requirements. Among the three models, the instances of

Model	Anger	Disgust	Fear	Happiness	Sadness
SVM	0.25	1	0.75	1	0.33
Decision Tree	0.25	0.75	0.75	1	0
Naive Bayes	0.25	1	0.75	1	0.33

Table 5.1.1: Accuracy of each model in recognizing each single/pure emotion

blended emotion ”disgust & happiness” which are accurately recognized are more likely to be recognized as ”happiness,” with recognition probabilities of 0.68, 0.56, and 0.63 for the SVM, Decision Tree, and Naïve Bayes models, respectively. Therefore, based on the analysis results, it can be concluded that with the current dataset, in the blended emotion ”disgust & happiness,” happiness appears to be the dominant component single/pure emotion.

However, due to the limited size of the dataset and the only acceptable accuracy of the models, the conclusions drawn are likely to apply only to the dataset used. To address the question of whether a dominant emotion exists within blended emotions, further investigation will necessitate more extensive data and the utilization of more sophisticated models.

5.1.5 Ablation Study for Feature Importance

In Section 4.4, the results of the feature ablation study are presented, which investigates the importance of various features for different emotions under each model. In this section, the importance of various features for both single emotions and blended emotions will be analyzed separately. The analysis aims to identify features whose removal significantly affects the model’s recognition performance across the three models and these features are then considered important for emotion recognition.

Single/pure emotion

For ”disgust”, based on Table 4.5.2, it can be observed that removing the feature ”AU_10” from the data will decrease the recognition ability of all models for ”disgust”. On the other hand, removing the feature ”AU_17” will increase the recognition ability of SVM and Naive Bayes models for ”disgust”. This indicates that ”AU_10” is important for recognizing ”disgust”, while the presence of ”AU_17” can introduce interference in recognizing “disgust”.

For ”happiness”, based on Table 4.5.4, it can be observed that removing the feature ”AU_12” from the data will decrease the recognition ability of all models for ”happiness”. This indicates that ”AU_12” is crucial for recognizing ”happiness”.

For ”fear”, according to Table 4.5.3, it can be observed that removing the feature ”AU_12” from the data will decrease the recognition ability of the decision tree and Naive Bayes models for ”fear” but also will increase the recognition ability of the SVM model for ”fear”. Additionally, removing the feature ”AU_12” from the data will decrease the recognition ability of the SVM and Naive Bayes models for ”fear” but also will increase the recognition ability of the decision tree model for ”fear”. This indicates that ”AU_12” and ”AU_14” are important for recognizing ”fear,” but whether these two features enhance or diminish the models’ recognition ability is difficult to conclude based on the current result and analysis.

For ”anger” and ”sadness”, based on Table 4.5.1 and Table 4.5.5, it can be observed that there are no specific features of particular importance that, when removed, decrease or increase the recognition ability of multiple models for ”anger” or ”sadness”. The reason for such results could possibly be that all models exhibit poor recognition abilities for ”anger” and ”sadness.” As a result, even if there are important features for recognizing these two emotions, it is challenging to observe them through experiments.

Blended emotion

In the context of blended emotions, an analogous analysis was performed for each specific blended emotion, presenting the results without reiterating the detailed process.

For ”anger & disgust”, the presence of ”AU_15” and ”AU_20” can introduce interference in its recognition. For ”disgust & fear”, ”AU_12” and ”AU_17” are important for its recognition. For ”disgust & happiness”, ”AU_10” and ”AU_26” are important for its recognition. For ”fear & happiness”, ”AU_02” and ”AU_12” are important for its recognition. While the presence of ”AU_09” can introduce interference in its recognition. For ”happiness & sadness”, ”AU_07”, ”AU_12” and ”AU_17” are important for its recognition. While the presence of ”AU_09” can introduce interference in its recognition.

For other blended emotions, it can be observed that there are no specific features of particular importance that, when removed, decrease or increase the recognition ability of multiple models for them.

During the analysis of important features for blended emotions, it was noticed that certain features present in the important feature tables for blended emotions were not found in the important feature tables for single/pure emotions. In most cases, the results are acceptable, as removing a specific feature only results in a decrease or increase in the recognition capability of that blended emotion in one of the models. This situation is normal and also acceptable. For example, for ”fear & happiness”, removing ”AU_23” will decrease the recognition ability of the SVM model for it. And ”AU_23” is absent in the important feature tables of ”fear” and ”happiness”. However, ”AU_23” actually is not so important because it only leads to the recognition ability of one model decrease. Therefore it is acceptable for its absence in the important feature tables of ”fear” and ”happiness”.

In one case, the result is confusing. As mentioned above, the presence of ”AU_15” can introduce interference in recognizing ”anger & disgust”. However, ”AU_15” is absent in the important feature tables of ”anger” and ”disgust”. After checking the original tables, it can be observed that removing ”AU_15” will increase the F1 score from 0.19 to 0.67 in the Naive Bayes model which suggests that ”AU_15” is an interfering feature for recognizing ”disgust & happiness”.

Some assumptions can be proposed for this situation. First, ”AU_15” may be highly correlated with other features that provide similar information. In this case, the model suffers from multicollinearity while retaining this feature, and the improved F1 score after removal is due to the elimination of the multicollinearity, making it easier to generalize the model to new data. However, based on figure 4.1.2 and figure 4.1.1, ”AU_15” is not highly correlated with other action units. Therefore, that assumption should be incorrect.

Second, ”AU_15” may contain a lot of noise or irrelevant information that interferes with the model’s performance. When removed, the model can better focus on the truly useful features, resulting in improved accuracy. Third, ”AU_15” may lead to an improvement in model over-fitting. Over-fitting is when a model performs well on training data but poorly on test data because it is too complex and fits too much of the noise from the training data. Removing a feature may reduce the complexity of the model and reduce the likelihood of over-fitting, thus improving generalization.

However, verifying these assumptions is very difficult. Based on the available knowledge, it's challenging to provide a conclusive analysis and interpretation of this result. But, this direction is indeed a research avenue that holds significant promise and is worth further exploration in the future.

5.2 Limitations and Future Work

While this thesis has explored the field of blended emotion expression to some extent and presented some insights, there are still many limitations.

The first limitation is related to the quantity of data. Due to the small dataset, the trained model is prone to overfitting on the limited samples and underfitting the target task. As a result, the model trained in this manner may not necessarily yield good recognition results on other similar datasets, although this assertion still requires further validation.

The second limitation lies in the choice of models employed in this project, all of which are relatively simple ML models. However, multi-emotion classification is a complex process, and while simple models have the advantage of faster training, they often struggle to achieve the same level of performance as complex models when faced with such intricate classification problems.

The third limitation is constrained by the nature of the data. While facial expressions can to a considerable extent reflect human emotions, in real-life situations, humans may involuntarily or intentionally conceal their true emotions, a phenomenon known as social masking[19], making it challenging for the model to be applied effectively in practical life scenarios.

The last limitation stems from the absence of temporal analysis of video data in this project. Instead, the approach involved directly averaging the data to represent the entire video. This approach neglects the variations in facial features over time in the expression of emotions, focusing solely on the overall intensity of facial features in emotional expression. Jiachen, a master student from the Karolinska Institute was also involved in the study of blended emotion expression. While I used supervised learning, she employed unsupervised learning. She used the same dataset as this project, employed unsupervised machine learning algorithms for temporal analysis of the data, and discovered that the temporal changes in features are a significant factor in classifying different emotions[55]. Therefore, due to the absence of temporal analysis of the data in this project, there is a certain limitation associated with it.

In future research, it may be possible to delve deeper into the study of blended emotion expression from two directions. First, increasing the quantity of data could enable the application of more complex emotion recognition models and enhance the model's generalizability. Second, conducting temporal analysis of the data, such as utilizing LSTM, VAE, or other temporal analysis models, can provide a better understanding of the principles behind blended emotion expression. However, achieving this would also require a substantial amount of data for training neural networks. Alternatively, employing the sliding window method to process data can help avoid the substantial information loss that comes with directly averaging all the data points. Furthermore, this method is easier to implement compared to complex neural networks like LSTM.

In summary, while this project has some limitations, it still serves as a strong foundation and offers valuable insights for researching blended emotion expression through supervised machine learning models.

Chapter 6

Conclusions

The thesis analyzes the process of expressing blended emotions by investigating the relationship between blended emotions and their constituent emotions. Based on the analysis in Chapter 5, it is highly likely that blended emotion expressions involve the combination of some features from one emotion and some features from another emotion, or they are composed of overlapping combinations of both emotional states. This also suggests that in blended emotion expressions, there are no unique feature patterns that do not exist in single/pure emotions.

Additionally, about explorations whether a dominant emotion exists among the constituent emotions of blended emotion expressions. In the blended emotion ”disgust & happiness,” happiness appears to be the dominant component single/pure emotion. However, it is not sufficient to conclude that happiness dominates in ”disgust & happiness” solely based on this analysis. Nor does it prove the presence of a dominant emotion in equivalent or equally intense blended emotions. To establish these claims, further experimentation on a larger dataset is required, along with the use of more complex and precise models. Furthermore, about the feature importance list for each emotion, there are some action units that are found to be important for recognizing specific emotions, both in single/pure emotions and blended emotions.

However, taking the average of all data points directly results in a significant loss of information, which represents a limitation and weakness of this project. In the future, it will be essential to perform time series-based analysis on blended emotion data using either a simple sliding window method or complex neural network models like LSTM or VAE.

In conclusion, this thesis has provided valuable and meaningful insights into the study of blended emotion expression using supervised learning methods. It also offers data and conclusions that can serve as references for future research.

Reference

[1] Gendron, Maria and Feldman Barrett, Lisa. "Reconstructing the past: A century of ideas about emotion in psychology". In: *Emotion review* 1.4 (2009), pp. 316–339.

[2] Izard, Carroll E. "The many meanings/aspects of emotion: Definitions, functions, activation, and regulation". In: *Emotion Review* 2.4 (2010), pp. 363–370.

[3] Gross, James J and Muñoz, Ricardo F. "Emotion regulation and mental health." In: *Clinical psychology: Science and practice* 2.2 (1995), p. 151.

[4] Scherer, Klaus R et al. "Psychological models of emotion". In: *The neuropsychology of emotion* 137.3 (2000), pp. 137–162.

[5] PS, Sreeja and Mahalakshmi, G. "Emotion models: a review". In: *International Journal of Control Theory and Applications* 10.8 (2017), pp. 651–657.

[6] Lindquist, Kristen A, Siegel, Erika H, Quigley, Karen S, and Barrett, Lisa Feldman. "The hundred-year emotion war: are emotions natural kinds or psychological constructions? Comment on Lench, Flores, and Bench (2011)." In: *Psychological Bulletin* 139.1 (2013), pp. 255–263.

[7] Bestelmeyer, Patricia EG, Kotz, Sonja A, and Belin, Pascal. "Effects of emotional valence and arousal on the voice perception network". In: *Social cognitive and affective neuroscience* 12.8 (2017), pp. 1351–1358.

[8] Ekman, Paul. "An argument for basic emotions". In: *Cognition & emotion* 6.3-4 (1992), pp. 169–200.

[9] Ekman, Paul. "Are there basic emotions?" In: *Psychological Review* 99.3 (1992), pp. 550–553.

[10] Harmon-Jones, Eddie, Harmon-Jones, Cindy, and Summerell, Elizabeth. “On the importance of both dimensional and discrete models of emotion”. In: *Behavioral sciences* 7.4 (2017), p. 66.

[11] Barrett, Lisa Feldman and Russell, James A. “The structure of current affect: Controversies and emerging consensus”. In: *Current directions in psychological science* 8.1 (1999), pp. 10–14.

[12] Citron, Francesca MM, Gray, Marcus A, Critchley, Hugo D, Weekes, Brendan S, and Ferstl, Evelyn C. “Emotional valence and arousal affect reading in an interactive way: neuroimaging evidence for an approach-withdrawal framework”. In: *Neuropsychologia* 56 (2014), pp. 79–89.

[13] Posner, Jonathan, Russell, James A, and Peterson, Bradley S. “The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology”. In: *Development and psychopathology* 17.3 (2005), pp. 715–734.

[14] Berrios, Raul. “What is complex/emotional about emotional complexity?” In: *Frontiers in psychology* 10 (2019), p. 1606.

[15] Cross Francis, Dionne Indera, Hong, Ji, Liu, Jinqing, Eker, Ayfer, Lloyd, Kemol, Bharaj, Pavneet Kaur, and Jeon, MiHyun. “The dominance of blended emotions: A qualitative study of elementary teachers’ emotions related to mathematics teaching”. In: *Frontiers in psychology* 11 (2020), p. 1865.

[16] Berrios, Raul, Totterdell, Peter, and Kellett, Stephen. “Investigating goal conflict as a source of mixed emotions”. In: *Cognition and Emotion* 29.4 (2015), pp. 755–763.

[17] Braniecka, Anna, Trzebińska, Ewa, Dowgiert, Aneta, and Wytykowska, Agata. “Mixed emotions and coping: The benefits of secondary emotions”. In: *PloS one* 9.8 (2014), e103940.

[18] Fa-Da, PAN and Qi-Wei, LI. “Theoretical Arguments and Evidence about Simultaneous Experience and Processing of Mixed Emotions”. In: *Journal of East China Normal University (Educational Sciences)* 29.4 (2011), p. 40.

[19] Zhang, Jianhua, Yin, Zhong, Chen, Peng, and Nichele, Stefano. “Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review”. In: *Information Fusion* 59 (2020), pp. 103–126.

[20] Bota, Patricia J, Wang, Chen, Fred, Ana LN, and Da Silva, Hugo Plácido. "A review, current challenges, and future possibilities on emotion recognition using machine learning and physiological signals". In: *IEEE Access* 7 (2019), pp. 140990–141020.

[21] Jang, Eun-Hye, Park, Byoung-Jun, Kim, Sang-Hyeob, Chung, Myung-Ae, Park, Mi-Sook, and Sohn, Jin-Hun. "Emotion classification based on bio-signals emotion recognition using machine learning algorithms". In: *2014 International Conference on Information Science, Electronics and Electrical Engineering*. Vol. 3. IEEE. 2014, pp. 1373–1376.

[22] Tarunika, K, Pradeeba, RB, and Aruna, P. "Applying machine learning techniques for speech emotion recognition". In: *2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. IEEE. 2018, pp. 1–5.

[23] Singh, Shubham Kumar, Thakur, Revant Kumar, Kumar, Satish, and Anand, Rohit. "Deep learning and machine learning based facial emotion detection using CNN". In: *2022 9th International Conference on Computing for Sustainable Global Development (INDIACom)*. IEEE. 2022, pp. 530–535.

[24] Ahirwal, Mitul Kumar and Kose, Mangesh Ramaji. "Audio-visual stimulation based emotion classification by correlated EEG channels". In: *Health and Technology* 10.1 (2020), pp. 7–23.

[25] Israelsson, Alexandra, Seiger, Anja, and Laukka, Petri. "Blended Emotions can be Accurately Recognized from Dynamic Facial and Vocal Expressions". In: *Journal of Nonverbal Behavior* (2023), pp. 1–18.

[26] Khan, Amjad Rehman. "Facial emotion recognition using conventional machine learning and deep learning methods: current achievements, analysis and remaining challenges". In: *Information* 13.6 (2022), p. 268.

[27] Wang, Haifeng and Hu, Dejin. "Comparison of SVM and LS-SVM for regression". In: *2005 International conference on neural networks and brain*. Vol. 1. IEEE. 2005, pp. 279–283.

[28] Zoppis, Italo, Mauri, Giancarlo, and Dondi, Riccardo. "Kernel Methods: Support Vector Machines". In: *Encyclopedia of Bioinformatics and Computational Biology*. Ed. by Shoba Ranganathan, Michael Gribskov, Kenta Nakai, and

Christian Schönbach. Oxford: Academic Press, 2019, pp. 503–510. ISBN: 978-0-12-811432-2. DOI: https://doi.org/10.1016/B978-0-12-809633-8.20342-7. URL: https://www.sciencedirect.com/science/article/pii/B9780128096338203427.

[29] Jakkula, Vikramaditya. "Tutorial on support vector machine (svm)". In: *School of EECS, Washington State University* 37.2.5 (2006), p. 3.

[30] Ali, Jehad, Khan, Rehanullah, Ahmad, Nasir, and Maqsood, Imran. "Random forests and decision trees". In: *International Journal of Computer Science Issues (IJCSI)* 9.5 (2012), p. 272.

[31] Kotsiantis, Sotiris B, Zaharakis, Ioannis, Pintelas, P, et al. "Supervised machine learning: A review of classification techniques". In: *Emerging artificial intelligence applications in computer engineering* 160.1 (2007), pp. 3–24.

[32] Webb, Geoffrey I, Keogh, Eamonn, and Miikkulainen, Risto. "Naïve Bayes." In: *Encyclopedia of machine learning* 15.1 (2010), pp. 713–714.

[33] Li, Shan and Deng, Weihong. "Blended emotion in-the-wild: Multi-label facial expression recognition using crowdsourced annotations and deep locality feature learning". In: *International Journal of Computer Vision* 127.6-7 (2019), pp. 884–906.

[34] Watson, David and Stanton, Kasey. "Emotion Blends and Mixed Emotions in the Hierarchical Structure of Affect". In: *Emotion Review* 9.2 (2017), pp. 99–104. DOI: 10.1177/1754073916639659. eprint: https://doi.org/10.1177/1754073916639659. URL: https://doi.org/10.1177/1754073916639659.

[35] Domínguez-Jiménez, Juan Antonio, Campo-Landines, Kiara Coralia, Martínez-Santos, Juan C, Delahoz, Enrique J, and Contreras-Ortiz, Sonia H. "A machine learning model for emotion recognition from physiological signals". In: *Biomedical signal processing and control* 55 (2020), p. 101646.

[36] Cai, Jing, Xiao, Ruolan, Cui, Wenjie, Zhang, Shang, and Liu, Guangda. "Application of electroencephalography-based machine learning in emotion recognition: A review". In: *Frontiers in Systems Neuroscience* 15 (2021), p. 729707.

[37] Minaee, Shervin, Minaei, Mehdi, and Abdolrashidi, Amirali. "Deep-emotion: Facial expression recognition using attentional convolutional network". In: *Sensors* 21.9 (2021), p. 3046.

[38] Mackenzie, Noella and Knipe, Sally. "Research dilemmas: Paradigms, methods and methodology." In: *Issues in educational research* 16.2 (2006), pp. 193–205.

[39] Hooker, Sara, Erhan, Dumitru, Kindermans, Pieter-jan, and Kim, Been. "Evaluating Feature Importance Estimates". In: *arXiv* (2018). URL: https://arxiv.org/pdf/1806.10758.pdf.

[40] Baltrušaitis, Tadas, Robinson, Peter, and Morency, Louis-Philippe. "OpenFace: An open source facial behavior analysis toolkit". In: *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 2016, pp. 1–10. DOI: 10.1109/WACV.2016.7477553.

[41] Zadeh, Amir, Chong Lim, Yao, Baltrusaitis, Tadas, and Morency, Louis-Philippe. "Convolutional experts constrained local model for 3d facial landmark detection". In: *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2017, pp. 2519–2528.

[42] Baltrusaitis, Tadas, Robinson, Peter, and Morency, Louis-Philippe. "Constrained local neural fields for robust facial landmark detection in the wild". In: *Proceedings of the IEEE international conference on computer vision workshops*. 2013, pp. 354–361.

[43] Barra, Paola, Bisogni, Carmen, Nappi, Michele, and Ricciardi, Stefano. "Fast QuadTree-Based Pose Estimation for Security Applications Using Face Biometrics: 12th International Conference, NSS 2018, Hong Kong, China, August 27-29, 2018, Proceedings". In: Aug. 2018, pp. 160–173. ISBN: 978-3-030-02743-8. DOI: 10.1007/978-3-030-02744-5_12.

[44] Wood, Erroll, Baltrusaitis, Tadas, Zhang, Xucong, Sugano, Yusuke, Robinson, Peter, and Bulling, Andreas. "Rendering of eyes for eye-shape registration and gaze estimation". In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 3756–3764.

[45] Baltrušaitis, Tadas, Mahmoud, Marwa, and Robinson, Peter. "Cross-dataset learning and person-specific normalisation for automatic action unit detection". In: *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*. Vol. 6. IEEE. 2015, pp. 1–6.

[46] Ekman, Paul and Friesen, Wallace V. "Facial action coding system". In: *Environmental Psychology & Nonverbal Behavior* (1978).

[47] Cohn, Jeffrey F, Ambadar, Zara, and Ekman, Paul. “Observer-based measurement of facial expression with the Facial Action Coding System”. In: *The handbook of emotion elicitation and assessment* 1.3 (2007), pp. 203–221.

[48] Baltrusaitis, Tadas. *Action Units*. https://github.com/TadasBaltrusaitis/OpenFace/wiki/Action-Units. 2019.

[49] Lachmann, Tim. “Machine Learning for Emotion Recognition-Automated emotion recognition for 44 unique emotions”. In: (2022).

[50] Meijering, Erik. “A chronology of interpolation: from ancient astronomy to modern signal and image processing”. In: *Proceedings of the IEEE* 90.3 (2002), pp. 319–342.

[51] Kulkarni, Ajay, Chong, Deri, and Batarseh, Feras A. “5 - Foundations of data imbalance and solutions for a data democracy”. In: *Data Democracy*. Ed. by Feras A. Batarseh and Ruixin Yang. Academic Press, 2020, pp. 83–106. ISBN: 978-0-12-818366-3. DOI: https://doi.org/10.1016/B978-0-12-818366-3.00005-8. URL: https://www.sciencedirect.com/science/article/pii/B9780128183663000058.

[52] Sharma, Deepak Kumar, Chatterjee, Mayukh, Kaur, Gurmehak, and Vavilala, Suchitra. “3 - Deep learning applications for disease diagnosis”. In: *Deep Learning for Medical Applications with Unique Data*. Ed. by Deepak Gupta, Utku Kose, Ashish Khanna, and Valentina Emilia Balas. Academic Press, 2022, pp. 31–51. ISBN: 978-0-12-824145-5. DOI: https://doi.org/10.1016/B978-0-12-824145-5.00005-8. URL: https://www.sciencedirect.com/science/article/pii/B9780128241455000058.

[53] Wikipedia contributors. *Ablation (artificial intelligence) — Wikipedia, The Free Encyclopedia*. [Online; accessed 14-August-2023]. 2023. URL: https://en.wikipedia.org/w/index.php?title=Ablation_(artificial_intelligence)&oldid=1154046811.

[54] Merrick, Luke. “Randomized ablation feature importance”. In: *arXiv preprint arXiv:1910.00174* (2019).

[55] Yao, Jiachen. “Using unsupervised learning and feature analysis to investigate the composition of blended emotions”. In: (2023).