FACIAL EXPRESSION RECOGNITION OF AUTISM INDIVIDUALS FOR ANALYZING EMOTIONS USING MACHINE LEARNING

LEE QING REN

UNIVERSITI TEKNOLOGI MALAYSIA

PSZ 19:16 (Pind. 1/13)

UNIVERSITI TEKNOLOGI MALAYSIA

DECLARATION OF THESIS / UNDERGRADUATE PROJECT REPORT AND COPYRIGHT

Author's full name : Lee Qing Ren

Date of Birth : 25/07/2000

Title : Facial Expression Recognition Of Autism Individuals For Analyzing Emotions Using Machine Learning

Academic Session : 2023/20242

I declare that this thesis is classified as:

☐	**CONFIDENTIAL**	(Contains confidential information under the Official Secret Act 1972)*
☐	**RESTRICTED**	(Contains restricted information as specified by the organization where research was done)*
✓	**OPEN ACCESS**	I agree that my thesis to be published as online open access (full text)

1. I acknowledged that Universiti Teknologi Malaysia reserves the right as follows:
2. The thesis is the property of Universiti Teknologi Malaysia
3. The Library of Universiti Teknologi Malaysia has the right to make copies for the purpose of research only.
4. The Library has the right to make copies of the thesis for academic exchange.

Certified by:

SIGNATURE OF STUDENT	**SIGNATURE OF SUPERVISOR**
A20EC0065	DR. PANG YEE YONG
MATRIX NUMBER	**NAME OF SUPERVISOR**

NOTES : If the thesis is CONFIDENTIAL or RESTRICTED, please attach with the letter from the organization with period and reasons for confidentiality or restriction

“I hereby declare that we have read this thesis and in my
opinion this thesis is suffcient in term of scope and quality for the
award of the degree of Bachelor of Computer Science (Computer Graphics and
Multimedia Software)”

Signature	:	
Name of Supervisor	:	DR PANG YEE YONG
Date	:	4 JULY 2024

FACIAL EXPRESSION RECOGNITION OF AUTISM INDIVIDUALS FOR ANALYZING EMOTIONS USING MACHINE LEARNING

LEE QING REN

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Bachelor of Computer Science (Computer Graphics and Multimedia Software)

Faculty of Computing
Universiti Teknologi Malaysia

JULY 2024

DECLARATION

I declare that this thesis entitled *"Facial Expression Recognition Of Autism Individuals For Analyzing Emotions Using Machine Learning"* is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature :

Name : LEE QING REN

Date : 4 JULY 2024

DEDICATION

This thesis is dedicated to my father, who taught me that the best kind of knowledge to have is that which is learned for its own sake. It is also dedicated to my mother, who taught me that even the largest task can be accomplished if it is done one step at a time.

ACKNOWLEDGEMENT

In preparing this thesis, I was in contact with many people, researchers, academicians, and practitioners. They have contributed towards my understanding and thoughts. In particular, I wish to express my sincere appreciation to my main thesis supervisor, Dr Pang Yee Yong, for encouragement, guidance, critics and friendship. I am also very thankful to Dr Ajune Wanis binti Ismail, Dr Goh Eg Su and Dr Sim Hiew Moi for their guidance, advices and motivation. Without their continued support and interest, this thesis would not have been the same as presented here. I am grateful to all my family member.

ABSTRACT

Facial expression recognition in individuals with Autism Spectrum Disorder (ASD) remain challenges due to atypical expression patterns that could be hard to interpret by others comparing to neurotypical people. Moreover, finding a solution to reduce the computational time in training models while using large dataset is the current trends. Therefore, this research aims to investigate the state of the art of facial expression recognition using machine learning approaches, to develop a facial expression recognition model to classify emotions for individuals with ASD with less computational time and to evaluate the reliability and the accuracy of the proposed model for individuals with ASD. Two datasets of autistic facial emotions were collected, combined and preprocessed such as converting to grayscale, resizing, data augmentation and normalization. The emotion classes such as fear and surprise were dropped as the difference of the number of images was significant compared to the others. Feature extraction approaches such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), PCA+LDA and facial landmarks localization were carried out to find out which approach consumed the least computational time. The proposed model and the other classifiers utilized the least computational time feature extraction approach found to classify the autistic emotions. The proposed model was designated based on a referenced model. It was CNN based where it applied the fully connected layer for the autistic emotion classification. The performance was evaluated by using classification accuracy and confusion matrix and compared the result with the other classifiers. Based on the results, the proposed model had the highest accuracy (64.30%) among the classifiers. All the classifiers, including the proposed model learned and predicted well for the emotion joy while predictions of the other emotion classes such as angry, natural and sad were not reliable as the recall values were low.

ABSTRAK

Pengenalan ekspresi wajah pada individu dengan Gangguan Spektrum Autisme (ASD) masih menghadapi cabaran disebabkan oleh corak ekspresi atipikal yang sukar untuk ditafsirkan oleh orang lain berbanding dengan individu neurotipikal. Selain itu, mencari penyelesaian untuk mengurangkan masa pengiraan dalam melatih model sambil menggunakan set data besar adalah trend semasa. Oleh itu, penyelidikan ini bertujuan untuk menyiasat keadaan seni pengenalan ekspresi wajah menggunakan pendekatan pembelajaran mesin, untuk membangunkan model pengenalan ekspresi wajah untuk mengklasifikasikan emosi bagi individu dengan ASD dengan masa pengiraan yang kurang dan untuk menilai kebolehpercayaan dan ketepatan model yang dicadangkan untuk individu dengan ASD. Dua set data emosi wajah autistik telah dikumpulkan, digabungkan dan diproses semula seperti menukar kepada skala kelabu, menukar saiz, augmentasi data dan penormalan. Kelas emosi seperti takut dan terkejut digugurkan kerana perbezaan bilangan imej adalah ketara berbanding yang lain. Pendekatan pengekstrakan ciri seperti Analisis Komponen Utama (PCA), Analisis Diskriminan Linear (LDA), PCA+LDA dan penempatan tanda muka telah dijalankan untuk mengetahui pendekatan mana yang menggunakan masa pengiraan paling sedikit. Model yang dicadangkan dan pengelas lain menggunakan pendekatan pengekstrakan ciri masa pengiraan paling sedikit untuk mengklasifikasikan emosi autistik. Model yang dicadangkan telah ditetapkan berdasarkan model yang dirujuk. Ia berasaskan CNN di mana ia menggunakan lapisan sepenuhnya bersambung untuk pengelasan emosi autistik. Prestasi dinilai menggunakan ketepatan klasifikasi dan matriks kekeliruan dan membandingkan hasil dengan pengelas lain. Berdasarkan keputusan, model yang dicadangkan mempunyai ketepatan tertinggi (64.30%) di kalangan pengelas. Semua pengelas, termasuk model yang dicadangkan belajar dan meramalkan dengan baik untuk emosi kegembiraan manakala ramalan kelas emosi lain seperti marah, neutral dan sedih tidak boleh dipercayai kerana nilai panggil semula adalah rendah.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ASD	-	Autism Spectrum Disorder
ANN	-	Artificial Neural Network
CNN	-	Convolutional Neural Network
DCNN	-	Deep Convolutional Neural Network
ECG	-	Electrocardiogram
FER	-	Facial Expression Recognition
GAN	-	Generative Adversarial Network
KNN	-	K-Nearest Neighbors
LDA	-	Linear Discriminant Analysis
MLP	-	Multilayer Perceptron
PCA	-	Principal Component Analysis
RNN	-	Recurrent Neural Network
SVM	-	Support Vector Machine
VGG	-	Visual Geometry Group

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder, such that the developing brain is affected by both genetic and environmental factors, leading to deficits in social communication, the presence of restricted interests and repetitive behaviors (Black & MPH, 2014; Hodges et al., 2020; Sharma et al., 2018). In the book “DSM-5 Guidebook”, it stated that people with ASD have to meet the criteria such as deficiency in nonverbal communication, abnormal social approach, repetitive stereotyped behaviors, failure of reciprocal conversation, reduced sharing of feelings and emotions and failure to initiate or respond in social interactions (Black & MPH, 2014; Hodges et al., 2020). Children who were born by mothers who were exposed to psychotropic medication have a higher risk of having ASD according to the research (Gardener et al., 2009; Sharma et al., 2018). Therefore, potential patients often are diagnosed in childhood.

This research focuses on recognizing facial expressions made by individuals with ASD using machine learning approaches. This is because facial expression plays an important role in nonverbal communication other than body language or behavior as it encourages empathizing and understanding between each other. However, autistic individuals express their emotions through facial expression differently compared to norms based on the research findings. As a result, people find it is challenging to interpret the emotions by recognizing the facial expressions of the autistic people, leading to poor social communication between the autistic people and the others. It is necessary to have a solution to identify the emotion of the people with ASD accurately for clinical diagnosis, medical treatment and parental care. Fortunately, facial expression recognition for the people with ASD becomes a hot topic in society when the AI technology becomes more advanced. There are a lot of research carried out to propose solutions for classifying the emotions based on the facial expressions made by the people with ASD.

This project focuses on the facial expression recognition (FER) of autistic individuals to analyze emotions by using machine learning. This research aims to investigate the state of the art of facial expression recognition on ASD individuals using machine learning approaches, to propose a facial expression recognition model to classify emotions for individuals with ASD with using image inputs and to evaluate the reliability and the accuracy of the developed facial expression recognition prototype for individuals with ASD.

1.2 Problem Background

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder that may behave differently from other people and can be diagnosed early in childhood (Black & MPH, 2014, Basics About Autism Spectrum Disorder (ASD) | NCBDDD | CDC, 2022). Individuals with ASD have difficulties in social interaction, especially in expressing their emotions (Talaat, 2023). This is because the facial expressions made by people with ASD are atypical and recognized poorly by others compared to the normal people (Brewer et al., 2016). Therefore, it is quite challenging for healthcare professionals and caregivers to understand and address the emotional needs of individuals with ASD compared to normal people without identifying their current emotions accurately.

Nowadays, when the machine learning becomes more advanced, and it leads to the trends of efficiency model training (García-Martín et al., 2019). It is costly to train a model with a large dataset as it requires many computational resources (Guimarães et al. 2023). This research is inspired by the challenge of recognizing emotions of autistic people and plans to propose a model which uses less computational time to reach optimized result.

Hence, it is required to conduct a study to investigate the state of the art of facial expression recognition on ASD individuals using machine learning approaches. It is also important to propose a facial expression recognition model to classify emotions for individuals with ASD with less training using image inputs. Finally, it is necessary to evaluate the reliability and the accuracy of the developed facial expression recognition prototype for individuals with ASD.

1.3 Research Aim

Research aims to evaluate the proposed model in terms of accuracy and reliability by carrying out a comparison with the other approaches.

1.4 Research Question

The questions of the research are:

(a) How can machine learning approaches be effectively utilized for facial expression recognition in individuals with ASD?

(b) How to develop a model to classify emotions for individuals with ASD using less training time?

(c) How does the proposed prototype perform differently compared to the existing models in terms of accuracy?

1.5 Research Objectives

The objectives of the research are:

(a) To investigate the state of the art of facial expression recognition on ASD individuals using machine learning approaches.

(b) To propose a facial expression recognition model to classify emotions for individuals with ASD with less training time using image inputs.

(c) To evaluate the reliability and the accuracy of the developed facial expression recognition prototype for individuals with ASD by carrying out a comparison with the other approaches.

1.6 Research Scope

The scopes of the research are:

(a) The target user is autism individuals in ASD level 1.

(b) The prototype recognizes the autistic facial expressions and identify the emotion on a single target at a time.

(c) The camera is placed right in front of the target user.

(d) Emotions to be recognized are happiness, sadness, anger and neutral.

(e) Training dataset is the facial expressions of autistic individuals, in image file type and resized into same resolution.

1.7 Research Contribution

Autistic people have difficulties in socializing as it is hard for them to express emotions (Talaat, 2023). Many may interpret their emotions wrongly when their emotions show up atypically through their facial expressions (Brewer et al., 2016). It is a challenging task for healthcare professionals and caregivers to understand and address the emotional needs of individuals with ASD compared to neurotypical people. The future work of García-Martín et al., (2019) is looking into a way to reduce the computational cost required to train a model. Meanwhile, Guimarães et al. (2023) contributes into proposing an approach to predict the model training time. Therefore, inspired by the work above, this research aims to investigate the state of the art of facial expression recognition on the individuals with ASD using machine learning. It is to find out whether there are applications or systems developed to recognize the emotions of ASD individuals accurately so that the healthcare professionals and caregivers can plan strategies according to their emotions. Secondly, it aims to develop a facial expression recognition model to classify emotions for individuals with ASD with less computational time using image inputs. This is to reduce the training time required for the model so that the accuracy reaches the optimum with lesser computational time. Finally, it is to evaluate the reliability and the accuracy of the proposed prototype by comparing the obtained results with the other approaches. This is to find out which approach is suitable for the emotion classification task among the autistic people.

1.8 Report Organization

This thesis covers the entire chapter, from the beginning to the end of the research process. Introduction, Literature Review, Research Methodology, Research Design and Implementation, Results, Analysis and Discussion and Conclusion are the six chapters of this thesis.

Chapter 1 introduces this thesis to give an abstract. It consists of the introduction, problem background, research aim, research questions, research objectives, research scope and research contribution. The background and the goals of this thesis are covered in this chapter.

Chapter 2 focuses on the literature review. It covers the previous efforts that have been carried out by other researchers and developers. A lot of reviews would be carried out in this chapter to learn about the progress of the related research. It consists of the comparative analysis of algorithms and the system proposed in facial expression recognition among autistic individuals. In this chapter, the research domain of this thesis and the proposed solutions would be clarified.

Chapter 3 is about research methodology. It mainly discusses the workflow of the research and the justification. The performance measurement of the proposed solutions would also be discussed in this chapter.

Chapter 4 is about research design. It is about how the experiments are carried out and how the classifiers are evaluated. The experimental setup is discussed further in this chapter, such as the flow of the experiment, the feature extraction methods applied in the experiment, the architecture of the proposed model and the evaluation of the classifiers.

Chapter 5 is about the result and discussion. The results obtained from the research are analyzed and discussed in detail. The performance metrics such as computational time, dimensionality shape, classification accuracy and confusion matrix, are evaluated and explained with the analysis.

Finally, Chapter 6 is the conclusion. It is the last chapter which summarizes the whole thesis with a conclusion. It discusses important information such as how the objectives were achieved, the research contribution and the improvements and works to be performed in the future.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

In this chapter, it discusses the topics which are relatable with the research to have better understanding on ASD, facial expression recognition (FER), emotions, machine learning methods and deep learning neural networks. ASD is further discussed to investigate the diagnostic criteria, especially in emotion expressions. It looks further into the machine learning methods such as supervised learning and unsupervised learning along with the explanation. Moreover, it introduces some neural networks which can be implemented in FER for individuals with ASD. Each subchapter has the related works discussed to understand the state of the art. In simple words, it makes a review on the current state of the art techniques and algorithms used for facial expression recognition for autistic people.

2.2 Autism Spectrum Disorder (ASD)

The term “autism” represents the increased self-focus and often describes the aloof and unaware behaviors of individuals to the surroundings (Kana, 2022). A lot of studies stated that people with ASD would have to meet the criteria such as deficiency in nonverbal communication, abnormal social approach, repetitive stereotyped behaviors, failure of reciprocal conversation, reduced sharing of feelings and emotions and failure to initiate or respond in social interactions (Black & MPH, 2014; Carter, 2014; Hodges et al., 2020). This is because the brain is affected by both genetic and neurological factors associated with its development (Black & MPH, 2014; Hodges et al., 2020; Sharma et al., 2018). Children who were born by mothers that were exposed to psychotropic medication would have a higher risk of having ASD (Gardener et al., 2009; Sharma et al., 2018). This shows that ASD is a highly heritable disorder, and potential patients are often diagnosed in childhood as the gene in DNA has been changed.

As one of the criteria, autism individuals have difficulties in expressing their emotions in social interaction (Black & MPH, 2014; Carter, 2014; Hodges et al., 2020). Based on the critical review of Simmons et al. (2009) on ASD, there are a lot of studies showing that autistic individuals have deficiency in expressing emotional expressions. Therefore, facial expressions of ASD individuals are hard to be recognized compared to normal people as the accuracy of recognizing the autistic facial expressions is lower than that of neurotypical groups (Brewer et al., 2016). In the research of Brewer et al. (2016), the autistic expressions were recognized poorly and hard to interpret, indicating that individuals with ASD have atypical internal representations of emotions.

Based on DSM-V Autism Spectrum Disorder (American Psychiatric Association, 2013), it is classified to 3 severity levels, which are requiring support, requiring substantial support, and requiring very substantial support (Ousley & Cermak, 2014). Table 2.1 and Figure 2.1 shows the severity level of ASD with the description in social communication and repetitive stereotyped behaviors. Level 1 refers to the people with ASD requiring support, level 2 refers to the people with ASD requiring substantial support while level 3 refers to the people with ASD requiring very substantial support. It is challenging to interact with the higher severity level of ASD people. It is also hard to interpret the thoughts of the people with higher severity level of ASD as the facial expressions become lesser.

Table 2.1 Severity Level of ASD (American Psychiatric Association, 2013)

Severity Level	Social Communication	Repetitive Stereotyped Behaviors
Requiring support	Deficits in social communication with noticeable impairments without support. Deficits in social communication with clear atypical response. Possibility of decreased interest in social interactions.	Inflexibility of behavior. Deficits in multitasking. Difficulty in organizing and planning.
Requiring substantial support	Apparent deficits in verbal and nonverbal social communication. Apparent social impairments with supports. Limited initiation of social interactions Reduced or abnormal responses in communication.	Inflexibility of behavior. Difficulty in adapting changes Distress in changing tasks.
Requiring very substantial support	Severe deficits in verbal and nonverbal social communication skills cause severe impairments in functioning. Very limited initiation of social interactions, and minimal response to social overtures from others.	Inflexibility of behavior. Extreme distress in coping with change. Extreme distress in changing tasks.

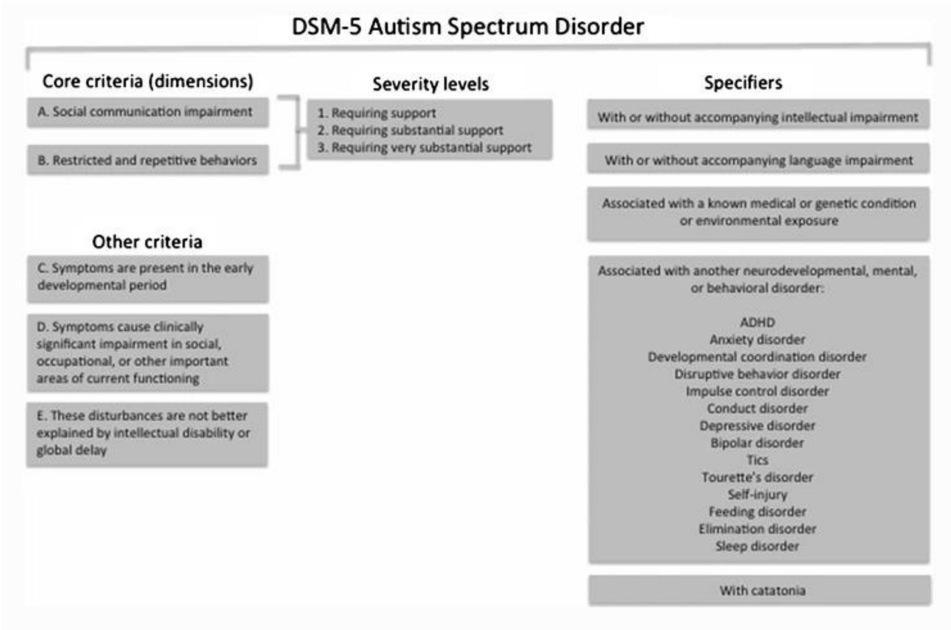

Figure 2.1 Severity Level of ASD (Ousley & Cermak, 2014)

2.3 Machine Learning

Machine learning is defined as the ability of a system to further learn or integrate new knowledge instead of applying the programmed knowledge only (Woolf, 2009). It impersonates humans in learning to carry out tasks with its algorithms by using the data collected (*What Is Machine Learning? | IBM*, n.d.). It has the flexibility to learn from experience and perform tasks based on the inputs from the situation. The learning system of machine learning is broken into 3 main components, which are decision process, error function and optimization process (Uadmin, 2022). Classification of input data is carried out based on the machine learning algorithms implemented in decision process (Uadmin, 2022). Meanwhile, the model accuracy is measured in error function and the is optimized by reducing the miss rate in optimization process (Uadmin, 2022).

There are 3 main types of machine learning techniques, which are supervised learning and unsupervised learning and reinforcement learning. Each type of the techniques has their specific learning methodology, such that supervised learning needs labeled data and teaching from the humans and unsupervised learning needs unlabeled data and teaching from the humans. Meanwhile, reinforcement learning is like supervised learning, but it does not require sample data (Wichert & Sa-Couto, 2021). In contrast, it learns by trial and error to find the best solution to the goal (Wichert & Sa-Couto, 2021).

In the study of Alvari et al. (2021), they combined the machine learning models with computer vision to analyze the micro-expressions of infants with ASD via videos using Openface, which is an AI-based software that integrated different tasks performed by different machine learning algorithms. The analysis showed that the intensity and the frequency of smiling expression of infants with ASD was lower than well-developed infants (Alvari et al., 2021). As a result, they lost the opportunity to experience effective interactions in the period of development (Alvari et al., 2021). Unfortunately, there was insufficient data to conclude the specific mechanism behind the atypical social cognitive development of infants with ASD (Alvari et al., 2021).

2.3.1 Supervised Learning

Supervised learning is one of the machine learning types such that model is trained using a labeled dataset, which means that each data has an expected output, under the supervision of the humans (GeeksforGeeks, 2023). The dataset is often split into 2 subsets, which are training set and test set to train model and evaluate model respectively (Liu, 2011). In simple words, models are trained using the split training set given by the humans and evaluated using the remaining set as test set in terms of accuracy. Figure 2.2 from Salian (2019) demonstrates how models learn in supervised learning. The labeled observations refer to the data that are split into two categories, which are training set and test set. The model is trained by the training set first and is evaluated using the testing set to generate the performance result of the model.

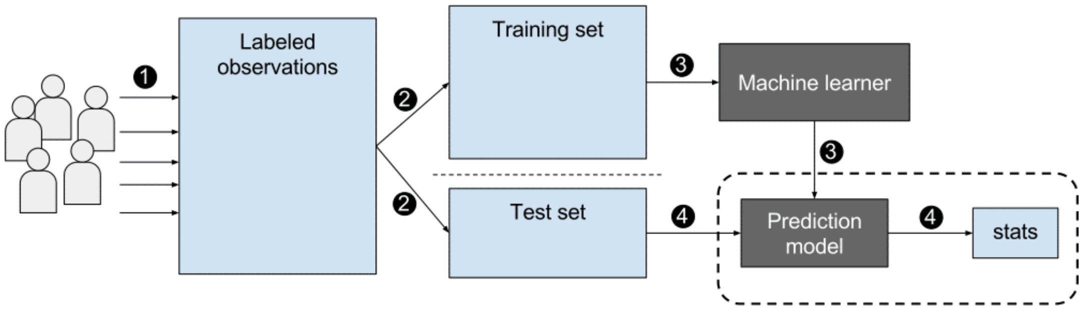

Figure 2.2 Classification in Supervised Learning (Salian, 2019)

The important part for supervised learning is the quality of the dataset. Feature selection performance could be dragged down by the missing values in the dataset (Muhammad & Yan, 2015). Hence, it is necessary to carry out data pre-processing such as outlier detection, which is noise removal for image data (Muhammad & Yan, 2015).

2.3.1.1 Support Vector Machines (SVM)

SVM is one of the machine learning algorithms that implements supervised learning. It performs well in text classification and other classification such as image classification (Liu, 2011).

A real time facial expression recognition system using SVM was proposed by Michel and Kaliouby (2003). The dataset was gathered via the face feature tracker to train the model to recognize unseen emotions (Michel & Kaliouby, 2003). It was assumed that the data gathered was always the front view of the posers and classified into 6 basic emotions (Michel & Kaliouby, 2003). Figure 2.3 demonstrates how the SVM model was trained to classify unseen expressions. Based on the findings, it was found that the recognition average accuracy of still image data was 86% while that of video was 71.8% (Michel & Kaliouby, 2003). The accuracy of different emotions classified were particularly different, some were higher, and some were lower (Michel & Kaliouby, 2003). Finally, the problem encountered was the accuracy affected by the head motion (Michel & Kaliouby, 2003).

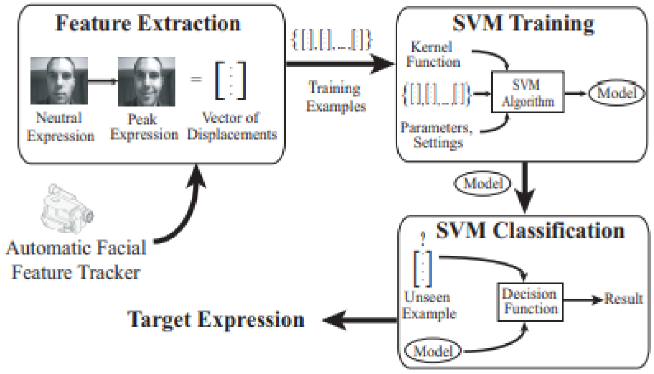

Figure 2.3 The Stages of SVM-Based Automated Expression Recognition (Michel & Kaliouby, 2003)

There is another study about a comparison between Support Vector Machine (SVM) and Artificial Neural Network (ANN) in the emotion classification of the autistic children and the result showed that SVM (90%) was more accurately than ANN (70%) (Rani 2019). Besides, there was a proposed emotion recognition system by Sivasangari et al. (2019) for autistic individuals. The authors used SVM as the classifier to recognize and predict emotions. Feature selection of the face and noise removal were carried out to improve the classification accuracy (Sivasangari et al., 2019). Face detection was performed using Haar feature-based cascade classifier, which is a machine learning approach in object detection (Sivasangari et al., 2019). Based on the findings, the proposed SVM had the highest accuracy compared to the other algorithms (Sivasangari et al., 2019).

2.3.1.2 Decision Tree

Decision Tree is a supervised learning algorithm that can be used for both classification and regression tasks. It generates a tree with a number of decisions by partitioning the data into subsets based on the input features. There are 3 nodes, such as root node, internal node and leaf node. Root node refers to the data inputs, internal node refers to the feature decisions and lead node represents the output of classification or regression. The root node is divided into sub-nodes to predict the final output using the decision rules in each internal node.

A hierarchical binary decision tree approach was suggested by Lee et al. (2011) to forecast the emotions from the spoken phrase input. The layers in the tree were designated to handle the most straightforward emotion predictions first, so that the error propagation could be minimized (Lee et al., 2011).

2.3.1.3 Random Forest

Random Forest is an ensemble learning model which it can be used in both classification and regression tasks. It is based on the mechanics of decision trees such that a random subset of the data set is used to measure a random subset of characteristics in each partition throughout the construction of each tree (GeeksforGeeks, 2024). It provides randomness during training and thus, it improves the accuracy of classification or regression tasks by reducing the overfitting problem.

Wang et al. (2019) suggested a hybrid strategy that integrated CNN and Random Forest to address the computational problem and the limitations of handcrafted feature extraction, such as lack ability of extracting and abstracting important features. The authors applied CNN to extract the feature and put the proposed random forest with 4.5 classifier at the last pooling layers of CNN to predict the class based on the probability. The research was carried out using 4 different facial emotion datasets from normal people, which were JAFFE, CK+, FER2013 and RAF-DB. This research showed that the proposed method is better than the other models in terms of accuracy.

2.3.1.4 AdaBoost

AdaBoost is the shortform of Adaptive Boosting. It is an ensemble learning algorithm that combines weak classifiers for boosting purpose to build a strong classifier and this improves the performance of weak classifiers. It starts with assigning equal weights to all samples and training the weak classifier with the weighted data. Then it calculates the error rate of the weak classifier and computes the alpha value. This alpha value represents the classifier’s accuracy. After that, it starts to update the sample weights such that it increases the weights for misclassified samples and vice versa. Lastly, the weak classifiers are combined by summing up the weights.

Owusu et al. (2014) studied on how to improve the facial emotion classification accuracy and execution time by proposing a neural-based AdaBoost based facial emotion recognition system. The proposed model achieved 96.83% and 92.22% using JAFFE and YALE dataset respectively. The execution time was 14.5ms for 100 × 100 pixel size. Moreover, neutral emotion class had the poorest classification precision compared to happy, surprise and disgust.

2.3.1.5 K Nearest Neighbors (KNN)

KNN is a supervised learning algorithm, such that it finds the nearest neighbors and assign the unclassified datapoint to the cluster. There are different ways to calculate the distance between the nearest neighbors and the unclassified datapoint, such as Euclidean Distance, Manhattan Distance and Minkowski Distance. Each method will have different calculations which result in different clustering of the neighbors.

Dino and Abdulrazzaq (2020) made a comparison among MLP, Naïve Bayes, decision tree and KNN to find out which classifier achieved the best accuracy with the smallest number of feature inputs. It applied 3 different feature selection methods to select 30 features, such as correlation, info gain and gain ration. It found out that KNN could achieve 91.01%, 90.17% and 89.45% accuracy while using correlation, info gain and gain ration respectively, with only 30 features used.

2.3.2 Unsupervised Learning

Unsupervised learning refers to the machine that learns to cluster the unlabeled data by the similarities and differences without any guidance. It is totally different from supervised learning. It is fast in clustering. However, there is a possibility for unsupervised learning model to categorize anomalies to one cluster that they are not supposed to, which means overestimation could happen (Pratt, 2020). As a result, the accuracy of the outputs using unsupervised learning is uncertain and it is difficult to validate the results as there are no labeled data (Pratt, 2020). Therefore, it requires experts to label the results (Pratt, 2020). Figure 2.4 shows how unsupervised learning carries out clustering.

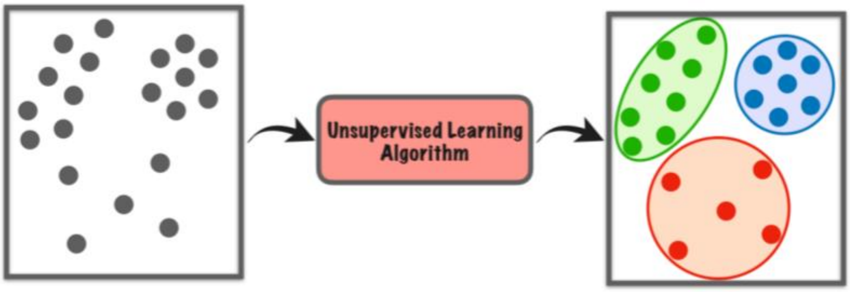

Figure 2.4 Clustering in Unsupervised Learning (Jeffares, 2020)

Kushki et al. (2015) proposed an unsupervised algorithm, which was based on Kalman filter, to be implemented in the real-time arousal detection of anxiety for children with ASD by physiological activities. The data was collected through electrocardiogram (ECG). Based on the results, it showed that the accuracy of the proposed algorithms was high with some potential improvements (Kushki et al., 2015). It would be great to have the data collected naturally, instead of under controlled laboratory settings (Kushki et al., 2015). Moreover, environmental noise was necessary to be considered to increase the system performance (Kushki et al., 2015).

From the study of Sivasangari et al. (2019), a comparison was made between the proposed SVM, decision tree, fuzzy logic and K-means. The accuracy of K-means almost had a similar accuracy result with the proposed SVM, yet it still was below the proposed SVM accuracy (Sivasangari et al., 2019).

There was a study that explored an unsupervised approach to classify the emotions of children with ASD (Kurian & Tripathi, 2022). The experiment was carried out using Generative Adversarial Network (GAN), which involved competing between 2 neural networks (Kurian & Tripathi, 2022). The result was the same for the datasets FER-2013 and CK+, which were 71.533% (Kurian & Tripathi, 2022). Although the accuracy was not as high as the previous work, the authors stated that it could be improved by feeding more data to the model (Kurian & Tripathi, 2022).

2.3.3 Comparison Between Supervised Learning and Unsupervised Learning

Table 2.2 summarizes the attributes of supervised and unsupervised learning with some examples of algorithms. It is suitable to implement supervised learning when the outputs are known. In contrast, unsupervised learning is implemented when the outputs remain unknown. It shows that it is suitable to use supervised learning in image recognition.

Table 2.2 Comparison between Supervised Learning and Unsupervised Learning

Aspect	Supervised Learning	Unsupervised Learning
Type of problems	Classification Regression	Association Clustering
Data Type	Labeled data	Unlabeled data
Supervision	External supervision	No supervision
Approach	Map the label inputs to the known outputs	Recognize the similarities and predict the outputs
Algorithms	SVM Decision trees Random forest	K-means PCA Gaussian Mixture Models
Field	Image recognition Spam filters Price prediction	Anomaly detection Recommendation system

2.4 Deep Learning

Deep learning is a subset of machine learning demonstrated in Figure 2.5 from Li et al. (2021). Machine learning and deep learning are similar in imitating humans to learn. Machine learning is a more generalized learning method while deep learning is specific in learning and predicting through the neural network of a machine. The concept of the neural network of a machine is based on the structure of the human brain neural network (Gupta, 2018). In simple words, deep learning is built of multiple layers of neural network, which consist of an input layer, one or multiple hidden layers and an output layer, to process data and make accurate prediction (Awan, 2022). It performs feature extraction based on the data input to extract the similar features of the same label before classifying the labels using decision boundaries (Awan, 2022).

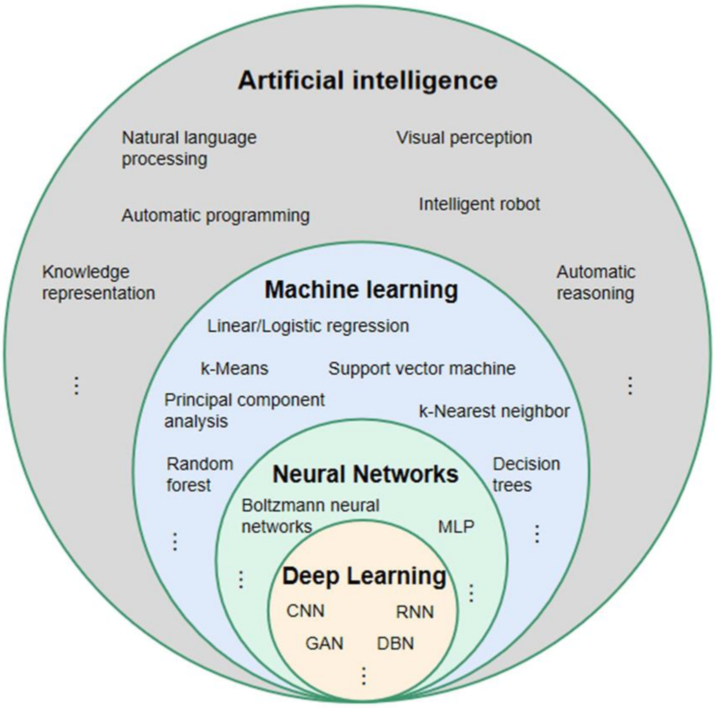

Figure 2.5 Relationship Between Artificial Intelligence, Machine Learning, Neural Networks and Deep Learning (Li et al., 2021)

There are 2 methods to train neural networks, which are forward propagation and backpropagation. Forward propagation, as indicated in its name, carries out the calculations and stores the intermediate values, traversing from the input layer to the hidden layers and finally to the output layers. In contrast, backpropagation traverses from the output layer to the input layer. The intermediate values stored in forward propagation are then used to compute the model gradients to find the error during learning. Since calculating and storing the intermediate variables are important for both propagations, it is necessary to have sufficient large memory to do the computation.

2.4.1 Artificial Neural Network (ANN)

Artificial neural network (ANN), which is also known as Multilayer Perceptron (MLP), is a simple and classic neural network, which is inspired by the human brain (Gupta, 2018). Figure 2.6 illustrates the architecture of ANN by Bartzatt (2014). The data with weightage is collected by the perceptron and transmitted to the hidden layers by firing activation functions (Gupta, 2018; Sadiq et al., 2019). When the nodes in the hidden layers reached the threshold, they would traverse to the output layer (Sadiq et al., 2019). ANN is suitable for large number of datasets as human experts could not interfere with the quantitative evidence stored in the network (Sadiq et al., 2019).

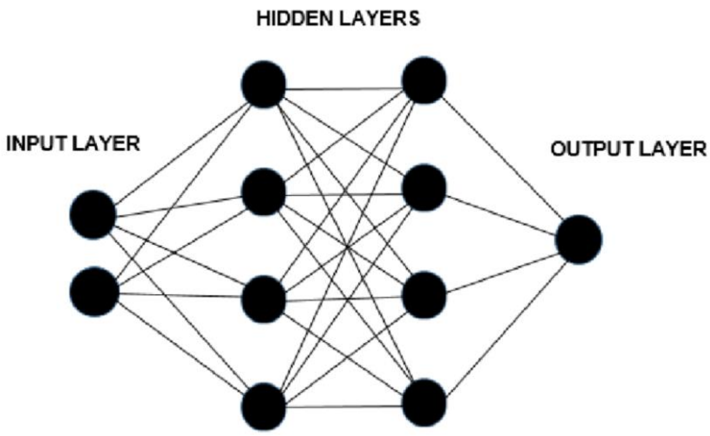

Figure 2.6 Architecture of ANN (Bartzatt, 2014)

There was a study about a comparison between Support Vector Machine (SVM) and Artificial Neural Network (ANN) in the emotion classification of the autistic children and the result showed that SVM (90%) was more accurately than ANN (70%) (Rani 2019). There is few research about facial emotion recognition using ANN for people with ASD.

2.4.2 Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) can deal with data in the form of multiple arrays, such as image data which carries the color and brightness intensity (LeCun et al., 2015; Mishra, 2021; Pai, 2023). It is suitable for various fields, such as image classification, object detection, segmentation and video processing as it can encapsulate the spatial features of an image. It is designated to learn hierarchical representations of data with the inspiration of the structure and function of the visual cortex in the brain (LeCun et al., 2015). Figure 2.7 illustrates the architecture of CNN with 3 layers, which are convolutional layer, pooling layer and a fully connected layer. Among the layers, the convolutional layer is the main building block as it is responsible for most of the computational load (Mishra, 2021). It takes a kernel to carry out dot product with the image (Mishra, 2021). In pooling layer, the output is split into partitions and the partitions are replaced by the value in the neighborhood, which is decided using statistical methods, to decrease the computational load (Mishra, 2021). In fully connected layer, it maps the input to the output by interconnecting the neurons (Mishra, 2021).

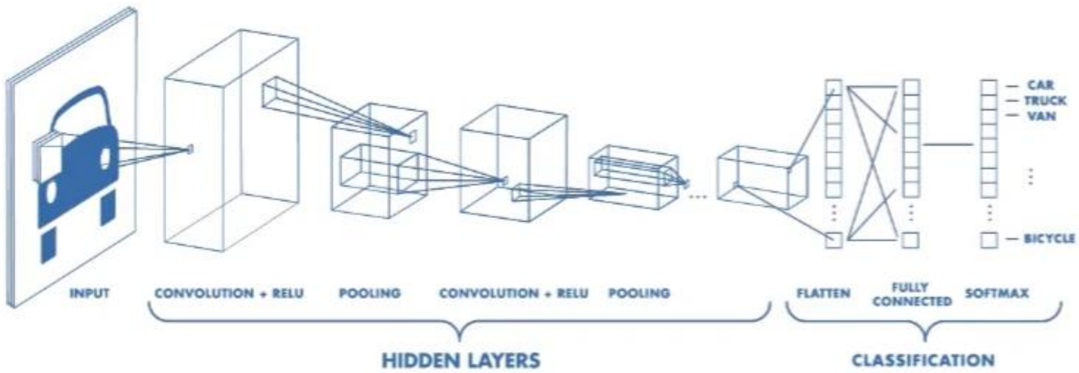

Figure 2.7 Architecture of CNN (*Introduction to Deep Learning: What Are Convolutional Neural Networks? Video*, n.d.)

Mollahosseini et al. (2016) presented a new deep neural network by adopting the main layers of CNN, which are convolution layer and max pooling layer, and adding 4 inception layers to recognize facial expressions. The number of neurons and layers were increased to improve the capability of the network in learning (Mollahosseini et al., 2016). Based on the experiments, the classification accuracy is increased with reduced number of iterations in training (Mollahosseini et al., 2016).

An approach using customized CNN to recognize the emotions of autistic individuals who might have a meltdown in real time with 80.6% accuracy using FER+ dataset (Silva et al., 2021).

Talaat (2023) proposed a real time facial emotion system among autistic children using Deep CNN architecture to recognize 6 basic emotions, such as joy, anger, sadness, surprise, fear and neutral. It did not require feature extraction and was better than former CNN based algorithms (Talaat, 2023).

2.4.3 Comparison Between Classifiers

Table 2.3 compares SVM, ANN, CNN, decision tree, random forest, KNN and AdaBoost in terms of data, data size, recurrent connections, parameter sharing, spatial relationship, complexity and field. CNN is powerful and popular for image data, and it needs a large data size to ensure the accuracy reaches the optimum. There is no recurrent connection, so it is not suitable for real time system development. Nevertheless, it supports parameter sharing and spatial relationships. Based on the attributes, it is suitable for image recognition and facial expression recognition.

ANN, which is also known as multilayer perceptron (MLP) is a simple and classic neural network which takes tabular and text data as input. It requires a large data size to train the neural network to ensure accuracy. There is no recurrent connection, parameter sharing, and spatial relationship supported. Hence, it is only suitable for predictive tasks, not suitable for handling image-based tasks as it does not handle image data and support spatial relationship.

SVM is a flexible algorithm that applies supervised learning. It receives the input data in numerical format and requires a small amount of data to train a model. If the data format is an image, it is required to change the image data into numerical format. It does not have recurrent connections and parameter sharing but supports spatial relationship based on the kernel, which enables it to act as a classifier to classify the images. It is suitable for image classification, text classification and prediction.

Decision tree prefers either numerical or categorical data type. The same goes for random forest, KNN and AdaBoost. The data size preferred is moderate. There is no recurrent connection and spatial relationship, but it has parameter sharing. This is because it involves splitting to determine the final output where it needs to use the previous parameters' weights. The flow from the root node to the output is sequential and not recursive. It depends on the weights in the tree instead of calculating distance or angle to classify the inputs. The complexity is from low to medium and it is suitable for the classification and regression tasks. At the same time, random forest, which is decision tree based, the aspects are similar to the decision tree. However, the only difference is that it provides randomness inside the tree. Thus, the complexity is from medium to high.

KNN prefers either numerical or categorical data type, which means it is suitable for both classification and regression tasks. The data size preferred is from small to medium. There are no recurrent connections and parameter sharing. However, it has spatial relationship as it classifies the data using the nearest neighbor calculated. The complexity is low as it does not involve complex computation.

AdaBoost prefers both numerical and categorical also. The size of data is from small to medium. Meanwhile, there are no recurrent connections, parameter sharing and spatial relationship. This is because it combines the weak classifiers to build a strong classifier by updating the weak classifiers' weights and hence, the complexity is medium. The task suitable for AdaBoost is classification.

Table 2.3 Comparison Among Classifiers

Classifier	SVM	ANN	CNN	Decision Tree	Random Forest	KNN	AdaBoost
Data	Numerical	Tabular and text data	Image data	Numerical and Categorical	Numerical and Categorical	Numerical and Categorical	Numerical and Categorical
Size of data	Small	Large	Large	Moderate	Large	Small to Medium	Small to Medium
Recurrent Connections	No	No	No	No	No	No	No
Parameter Sharing	No	No	Yes	Yes (trees)	Yes (trees)	No	No
Spatial relationship	Yes (Based on the kernel)	No	Yes	No	No	Yes (Distance)	No
Complexity	Flexible	Simple	Depends on architecture	Low to Medium	Medium to High	Low	Medium
Field	Image classification Text classification Prediction	Predictive analysis Complex problem solving	Image recognition Facial expression recognition	Classification and Regression	Classification and Regression	Classification and Regression	Classification

2.5 Computer Vision

Computer vision enables the computer to comprehend the environment by images and videos and is widely used in computer graphics, image processing and pattern recognition. (Shirai, 1987). It is because the computer learns by dealing with the image data in multiple arrays with the implementation of machine learning. Once the computer is well-trained, it is able perform specific tasks that it was trained to such as segmentation, face recognition, facial expression recognition, object detection etc. that related to image or video data.

Voulodimos et al. (2018) carried out a brief review on the computer vision using a deep learning approach. It was found that most of the applications utilized CNN architecture to enable computer vision. In the field of face recognition, feature extraction was needed to extract the features from the face to generate a low dimensional representation of data, enabling predictions by the classifier (Voulodimos et al., 2018). Some examples of face recognition systems that used

CNN architecture were Google's FaceNet and Facebook's DeepFace (Voulodimos et al., 2018).

2.5.1 Facial Expression Recognition (FER)

Facial Expression Recognition (FER) is a subcategory of computer vision and pattern recognition. It involves recognizing human emotions or facial expressions in static or real time with machine learning. FER has become more popular in recent years as people believe that it has potential in the area of psychological healthcare, human-computer interaction and security. It requires a model to be trained using a machine learning algorithm, such as CNN or another suitable algorithm, on a dataset of labeled or unlabeled facial expression images before making predictions.

Hassouneh et al. (2020) proposed a real time facial emotion recognition system using CNN and LSTM classifiers. There were 2 approaches implemented in this study. CNN classifier was used to classify and detect the emotions while LSTM was used to classify the detected ECG signals, which was prepared for validation purpose (Hassouneh et al., 2020). Based on the results obtained, the accuracy using facial landmarks was higher than that of using ECG signals was 87.25% (Hassouneh et al., 2020).

Abdullah and Al-Allaf (2021) proposed an optimized Visual Geometry Group (VGG) based on CNN architecture to recognize the facial expression of the autistic children while playing with computer or mobile. The dataset used in this study was CK+ 48 group and JAFFE group (Abdullah & Al-Allaf, 2021). The results obtained showed that the accuracy of recognizing emotions using CK+ 48 datasets was higher than that of using JAFFE datasets (Abdullah & Al-Allaf, 2021).

Talaat (2023) proposed an emotion detection framework to be implemented in IoTs using CNN for recognizing facial emotions of autistic children in real time. There were 3 layers for the framework, which were cloud layer, fog layer and IoT layer (Talaat, 2023). Cloud layer was responsible to collect the image data and save in cloud storage, fog layer mainly focused on classifying emotions based on the input and IoT layer was an API layer that sent and received the data (Talaat, 2023).

2.6 Feature Extraction

Feature extraction works perfectly for complex data. It aims to extract the attributes from the raw input data, such as images, audio and text to create a new set of features. For example, image data are extracted into an array of pixel values, which contain the color and brightness intensity information. The extracted features are fed to the training model or network to perform specific tasks, such as classification and recognition. Figure 2.8 shows how an image data is extracted into features. It helps to reduce the dimensionality of the data by focusing on the most informative features and reducing the noise and irrelevant information. This helps to improve the interpretability, accuracy and efficiency of machine learning models.

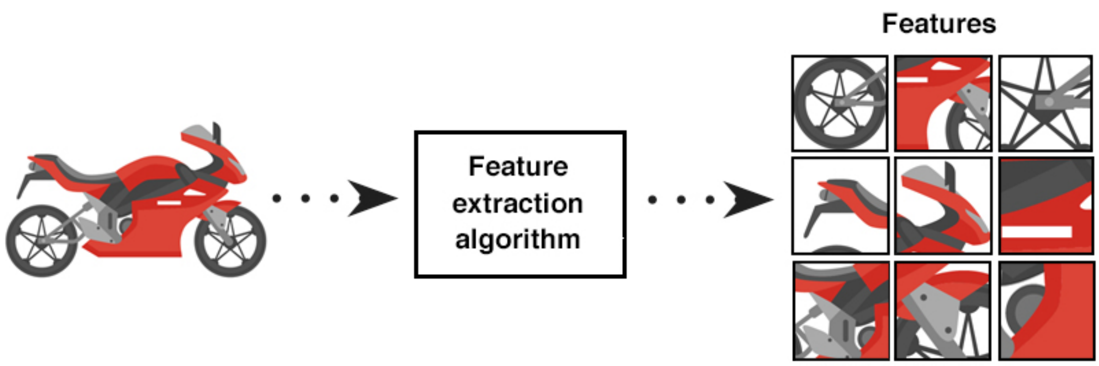

Figure 2.8 Feature Extraction of Image Data (Educative, n.d.)

2.6.1 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is an appearance based unsupervised algorithm that utilizes linear dimensionality reduction technique, such that it captures the important features in the data and discard the irrelevant or redundant information by maximizing the variances. It is unable to capture complex nonlinear relationships in the data. PCA requires the data to be standardized and the covariance to be computed based on the standardized data. Then the eigenvectors and eigenvalues are computed then to select the principal components based on the descending eigenvalues. The original data is projected onto the selected principal components to obtain the lower dimensional representation by taking the dot product of the standardized data and the principal components.

There are some studies that utilized PCA in feature extraction for facial expression recognition. Smitha and Vinod (2015) utilized PCA method to find the most important features that explain the data variance for autistic children to identify the distinctive features. The extracted features were then classified by the machine learning algorithms (Smitha & Vinod, 2015).

Sivasangari et al. (2019) implemented PCA method in the proposed emotion recognition system for people with ASD. The image data with RGB values were converted to grayscale and extracted into features, such as eyelid, nose tip, eyeball, mouth corner etc. using eigenvalue and eigenvector calculation (Sivasangari et al., 2019).

2.6.2 Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is an appearance based supervised algorithm in feature extraction. It can maximize the separation between different classes by creating a linear combination of the features and identify the features that differentiate the classes the best. It requires the labelled data to be pre-processed and standardized for the class mean computation. Then it computes the scatter matrices and solve the generalized eigenvalue problem to select the discriminant components before the projection. It is often used in the field of classification due to the discriminant functions that can separate the classes.

Marasamy and Sumathi (2012) proposed a method to resolve the drawbacks of LDA by applying wavelet transform. It was found that the performance of wavelet LDA was enhanced in face recognition tasks with the implementation of wavelet transform. However, PCA and LDA were able to be reached to the optimal accuracy faster than the wavelet LDA with the increasing number of data fit.

Sun et al. (2021) carried out a study to compare the feature extraction algorithms between LDA and facial landmark detection using the Cohn-Kanda+ dataset. It was found that the facial landmark detection performed better than LDA in emotion recognition as the accuracy was higher.

2.6.3 PCA + LDA

PCA+LDA starts with the dimensionality reduction first using PCA followed by LDA to extract the important features, which maximize the class separability. It can reduce the data dimensionality shape and extract the important features. PCA+LDA starts with the dimensionality reduction first using PCA followed by LDA to extract the important features, which maximize the class separability.

The combination can perform nicely on emotion classification task. Varma et al. (2019) carried out an analysis of PCA and LDA on its performance in facial expression recognition on 6 emotions using SVM and HMM classifiers. The performance metrics used were classification accuracy and confusion matrix. The result showed that the combination of PCA and LDA performed well with SVM instead of HMM.

2.6.4 Feature Point Tracking

Feature point tracking or facial landmark detection is a technique to track the features in a sequence of images or frames. It identifies the movement of selected feature points across the consecutive frames to analyze the motion, deformation, or the other changes over time. Firstly, the feature points of interest in the initial frame are manually specified or automatically detected using the algorithms such as corner detectors, scale-invariant feature transform (SIFT), or speeded-up robust features (SURF). Then, the initial positions of the feature points need to be estimated in the subsequent frames to establish the correspondence between frames. After that, the feature points are tracked across frames by estimating their new positions via the feature point tracking algorithms such as, Lucas-Kanade method, Kanade-Lucas-Tomasi (KLT) tracker, or pyramidal methods. Techniques such as Kalman filtering, robust estimators, or RANSAC (Random Sample Consensus) could be implemented to handle the occlusions and outliers found during the feature point tracking, improving the robustness.

Gondhi et al. (2017) combined Viola-Jones feature detection algorithm with Harris corner detection to reduce the complicated pre-processing in the facial landmarks detection system and make it easy to implement. It was found that the detection time was low, and the accuracy was high.

Song et al. (2018) proposed a CNN-based dense landmark detector called ConcatNet, that converted a given set of sparse landmarks into more precise and dense landmarks. Based on the findings, the accuracy of the landmark positions was increased with a small dataset and the number of the sparse landmarks was increased too. Sun et al. (2021) found that the facial landmark detection performed better than LDA in emotion recognition as the accuracy was higher.

Table 2.4 shows the comparison between PCA, LDA and feature point tracking in facial expression recognition. PCA is suitable for exploratory data analysis as it maximizes the variances to capture the important features and reduce the data dimensionality by projecting the data onto a lower dimensional space to generate the principal components. Meanwhile, LDA is suitable for classification task as it works with labelled data and maximizes the separation between the classes by creating a linear combination of the features. Feature point tracking is suitable for facial expression recognition as it focuses on the predefined geometric landmark position to extract the facial components to form a feature vector that represents the face geometry. In this case, emotion recognition could be carried out easily using the feature vectors extracted.

Table 2.4 Comparison Between PCA, LDA and Feature Point Tracking (Rajan et al., 2019; Kumar, 2023)

Algorithm	PCA	LDA	PCA + LDA	Feature Point Tracking
Method	Appearance-based	Appearance-based	Appearance-based	Geometric-based
Objective	Capture the most variation in the data	Capture the most separation between the classes	Capture the most variation in the data and the most separation between the classes	Focus on predefined geometric landmark position
Supervision	Unsupervised	Supervised	Hybrid	Semi-supervised
Dimensionality Reduction	Projecting the data onto a lower dimensional space	Creating a linear combination of the features that maximizes the separation between the classes	Projecting the data onto a lower dimensional space and creating a linear combination of the features that maximizes the separation between the classes	Facial components are extracted to form a feature vector that represents the face geometry
Output	Principal components	Discriminant functions	Principal components and discriminant functions	Facial features
Interpretation	Exploratory data analysis	Classification	Exploratory data analysis and classification	Facial Expression Recognition

2.7 Emotions

Emotion is defined as a complex psychological state that involves a range of physiological and behavioral responses in a situation. The physiological state changes based on the current emotion, such as sweating for being nervous and increasing blood pressure for being angry. The intensity of emotion felt for the current time varies with the duration, let it be positive or negative. One's behavior and decision making are also affected or portrayed by the current emotion.

Ekman (2000) suggested that there are 6 basic emotions, such as happiness, sadness, anger, fear, disgust and surprise. Complex emotions are formed by the combination of multiple basic emotions (Ekman, 2000). Humans are built complex, in the way of thinking and feeling. Furthermore, he also discussed the physiological states contributed by emotions.

Brewer et al., (2016) showed the deficit in expressing typical facial expressions among autistic people as their expressions were recognized less than that of neurotypical individuals. It also stated the problems encountered in recognizing emotions based on the facial expressions shown as there were 2 different expressions, which were spontaneous expressions, which referred to the current genuine emotional state, and posed expressions, which were used in social interaction widely to express manner (Brewer et al., 2016). Individuals with ASD might be impaired in both expression types and their emotional expressions in real life interaction could be different significantly (Brewer et al., 2016).

Trevisan et al. (2018) carried out an analysis to investigate the facial expressions in individuals with ASD. It was found that the facial expressions expressed by autistic people were atypical in appearance and less frequent. In addition, the quality and accuracy of their facial expressions were perceived to be worse. Emotions expressed by them were also shorter in time compared to neurotypical people.

Table 2.5 shows the facial expression features of the corresponding emotions based on Darwin's facial description (1872). The facial features of the emotions are iconic. However, as discussed by Trevisan et al. (2018), the facial expressions of autistic people are atypical, less frequent and low quality, making the emotions hard to be recognized based on Darwin's facial description (1872). However, the facial features described in Table 2.5 are still applied to the autistic people. Figure 2.9 shows the autistic facial expressions from the dataset recompiled by (Talaat, 2023) for anger, joy, sadness and natural. Figure 2.9(a) shows the autistic anger, figure 2.9(b) shows the autistic joy, figure 2.9(c) shows the autistic natural and figure 2.9(d) shows the autistic sadness. It is difficult to distinguish the emotions of ASD people through their facial expressions as there are no significant changes between the emotion classes.

Table 2.5 Facial Expression Features of Emotions (Darwin, 1872; Turabzadeh et al., 2018)

Emotions	Facial Features
Happiness	Eyes sparkle Drawn back at mouth corners Wrinkled skin under eyes
Sadness	Depressed mouth corner Raised inner eyebrows
Anger	Wide open eyes Compressed mouth Raised nostrils

Figure 2.9 Autistic Facial Expressions (Talaat, 2023)

2.8 Dataset

A dataset is a collection of structured or unstructured data that has been arranged and categorised for a particular use. It can be perceived as a methodical portrayal of data or observations from the real world. Datasets are frequently used to train models, test hypotheses, or draw conclusions from the data in the field of research and machine learning. The data instance is represented as files or rows in a table along with the features. There are many types of data, such as image, audio, text and so on. The technique applied in machine learning depends on the data type to preform pre-processing and other specific tasks.

Table 2.6 shows the overview of facial expression datasets, which are CK+, FER2013, MMI and AffectNet. Each dataset has different ways of collecting data, properties and labelling method. CK+, FER2013 and AffectNet are image-based datasets while MMI are video-based datasets. CK+ is verified among the datasets as the data is collected in a controlled laboratory using standardized protocols. Meanwhile, FER2013, MMI and AffectNet collects the data from the internet. AffectNet is the largest size of dataset among the 4 datasets. FER2013 provides large-scale dataset with diverse facial expressions from real world scenarios. In other words, the accuracy of emotion recognition using FER2013 dataset would be higher as it includes real world circumstances although the accuracy is affected by the automatic annotation. MMI is suitable for real time application as it consists of videos that provide the combination of posed and spontaneous emotions.

Table 2.6 Overview of Facial Expression Datasets

Types of Datasets	CK+	FER2013	MMI	AffectNet
Data Collection	Collected in a controlled laboratory using standardized protocols	Collected from the internet	Webcam recordings and video interviews	Collected from the internet
Label	FACS	User tags	Dimensional models	Crowdsourcing
Size	593 image frames from 123 participants	Approximately 30000 images, divided into training, validation, and testing sets	2900 videos and high-resolution still images of 75 subjects	1M images
Resolution	640 x 490 pixels	48 x 48 pixels	Not specified	Not specified
Properties	Focuses on intensity variations within the basic facial expressions, providing nuanced data for expression analysis.	Large-scale dataset with diverse facial expressions from real-world scenarios but the accuracy depends on the automatic annotation.	Provides a combination of posed and spontaneous expressions. Captures the variability and complexity of human emotions.	Captures the variability and cultural differences in expressing emotions.

2.9 Performance Measure

A system, process, or intervention's efficacy, efficiency, and success are evaluated and assessed as part of the performance measurement process. It entails gathering and analyzing data and metrics to evaluate performance and track the development of objectives. In other words, it aims to evaluate the performance to identify the strengths and weaknesses so that improvements can be made.

There are many performance measures that could assess the performance of trained models, such as classification accuracy, confusion matrix, Area Under Curve (AUC), mean squared error, training time etc. Classification accuracy performs well when the number of samples for each class is equivalent. Meanwhile, confusion matrix demonstrates the predicted output in the form of matrix to describe the number of correct classified class and the number of incorrect classified class. AUC is suitable for binary classification problem as it depends on the probability in ranking the sample randomly. It is easy to compute the gradient of the squared of the difference between the original and predicted values so that the models can focus on the large errors.

In terms of training time, there are several measures such as training time per epoch, iterations per second and batch processing time. Training time per epoch refers to the time taken to complete the training process, which includes data loading, preprocessing, model initialization, forward and backward passes, parameter updates and so on. Iterations per second measures the number of training iterations that the model can perform in one second, which indicates the speed of the models in learning and updating the parameters. Meanwhile, batch processing time measures the time taken to process and train the model on a single batch of training data.

Sivasangari et al. (2019) used classification accuracy and confusion matrix to measure the proposed model and carry out the comparison between different classifiers. However, it did not show the training time needed for the proposed model. Silva et al. (2021) also used classification accuracy and confusion matrix to measure the proposed model. Based on the accuracy over epoch graph, it took approximately 70 epochs to converge and complete training. Rani (2019) did not demonstrate the training time but clarified that it took 24 minutes and 40 seconds for training.

2.10 Chapter Summary

ASD is categorized as 3 groups, which are requiring support (level 1), requiring substantial support (level 2) and requiring very substantial support (level 3). People with ASD have difficulties in social interactions, especially in emotion reciprocity in nonverbal communication. In other words, normal people find it is difficult to communicate with individuals with ASD as their emotional representation is atypical. There are a lot of studies in this area to prove the statement. The facial expressions of people with ASD are atypical and hard to be recognized compared to the normal people.

Supervised learning is the learning method of machine under the supervision of human experts using labeled dataset. There is a study that showed that the performance is affected by the missing values. It is required to carry out data pre-processing to improve the dataset quality, avoiding the impact of the dataset to the performance. There are 2 types of evaluations methods in supervised learning such as Hold-out validation and Cross-Validation (CV) to measure the accuracy of the trained model.

In contrast, unsupervised learning does not require human supervision and labeled data to learn and classify. In contrast, it can learn to cluster classes by identifying the similarities and differences of the unlabeled data. Nevertheless, mistaking an anomaly into a cluster is possible and this will affect the accuracy of the outputs. In addition, it is difficult to validate the results as there is no labeled data.

SVM applies supervised learning and performs well in classification. There were a lot of studies that proposed facial expression recognition system using SVM. It was found that the average accuracy of still image data was higher than video. Furthermore, it was found that the classification of the autistic emotions was more accurate than ANN using SVM classifier. It was also found that the emotion recognition of individuals with ASD using SVM classifier was better than other algorithms.

ANN, which is also known as multilayer perceptron, is a deep learning neural network, which often requires many datasets to achieve optimal accuracy. The limitation of ANN is that the quantitative evidence stored in the network during the computation cannot be interfered with by human experts. There was a study which showed that the accuracy of ANN is lower than SVM in emotion recognition of individuals with ASD. There was only a few research on facial emotion recognition using ANN for individuals with ASD.

K-nearest neighbors, which is KNN in short, is a supervised algorithm that classifies data by calculating the nearest neighbor. In other words, the unclassified data is predicted with the nearest distance between the neighbors and itself. There was research that found out KNN could achieve a nice accuracy with only 30 features used.

It was found out that KNN could be used to predict the emotions from the spoken phrase input by a hierarchical binary decision tree approach, which was suggested by Lee et al. (2011).

AdaBoost could be used to improve the facial emotion classification accuracy. In the research of Owusu et al. (2014), the proposed neural based achieved 96.83% and 92.22% using JAFFE and YALE dataset respectively.

To overcome the handcrafted feature extraction in extracting meaningful features and abstracting data, a hybrid strategy was proposed by Wang et al. (2019), such that CNN and Random Forest are combined. This research showed that the proposed method was better than the other models in terms of accuracy.

CNN is more powerful algorithms than ANN as it able to deal with the data with multiple arrays, especially image data. Therefore, it is suitable for various fields, especially in image recognition. There were a lot of studies on facial expression recognition for the individuals with ASD using CNN. A new deep neural network with 4 inception layers adding into CNN architecture was presented to recognize the emotions of autistic people and increase the classification accuracy. Besides, a customized CNN was proposed to recognize the emotions of autistic individuals in real time with 80.6% accuracy using FER+ dataset. A real time facial emotion system among autistic children using Deep CNN architecture was proposed without feature extraction and the result showed that Deep CNN was better than the former CNN based algorithms (Talaat, 2023).

A real time facial emotion recognition system using CNN and LSTM classifiers to classify the emotions and the ECG signals received was proposed and it was found that the accuracy using facial landmarks was higher than that of using ECG signals. An emotion detection framework in IoTs using CNN was also proposed for recognizing facial emotions of autistic children in real time. There were 3 layers for the framework, which were cloud layer, fog layer and IoT layer. Cloud layer was responsible to collect the image data and save in cloud storage, fog layer mainly focused on classifying emotions based on the input and IoT layer was an API layer that sends and receives the data.

There are various methods in feature extraction, such as PCA, LDA and feature point tracking. PCA is suitable for exploratory data, LDA is suitable is suitable for classification tasks while feature point tracking works specifically for facial expression recognition as it tracks the facial feature points. It depends on the circumstances to select which approach to be used. In this case, feature point tracking is preferrable as it was found that the facial landmark detection performed better in emotion recognition (Sun et al., 2021).

Emotion refers to the feeling associated with the physiological state. Based on the previous literature, there are 6 basic emotions, such as happiness, sadness, anger, fear, disgust and surprise while complex emotions are the combination of multiple basic emotions. There were a lot of studies that showed that the atypical emotions in autistic people were recognized poorly by others. Another study also found that the autistic people expressed the emotions atypically and less frequently, resulting that the quality and accuracy of their facial expressions were perceived poorly.

Datasets are necessary to train models or networks to perform specific tasks, such as classification, recognition and prediction. There are 4 common facial expression datasets such as CK+, FER2013, MMI, and AffectNet. It reminded that the methods used in the data collection for each dataset are different. The data in FER2013 is collected from the internet and labelled with the user tags in 48 x 48 resolution. It is a large-scale dataset with diverse facial expressions from real-world scenarios, but the accuracy depends on the automatic annotation.

Performance measurement is crucial to investigate whether the proposed model has achieved the standards and objectives. Classification accuracy and confusion matrix are generally good to demonstrate the overall performance of the proposed model as they depict the overall accuracy and the number of correct and incorrect predicted classes in the form matrix. There is a few research that showed the training time taken for their proposed models. A graph of accuracy or loss over the number of epochs or iterations can demonstrate the training time required for the proposed model.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

In this chapter, it focuses on the research framework and the methods or technologies that would be implemented in this research. The research framework consists of 2 phases, which are analysis phase and implementation phase. Both phases contain the flow of stages. In short, the analysis phase discusses the problem statement, literature review and data collection. Meanwhile, the implementation phase details out how the research is going to be carried out. It includes data preprocessing, feature extraction, prototype design and development, training and testing and result analysis. The methods used are further discussed in the subchapters created to clarify the technologies or methods that are going to apply, such as feature extraction and performance measurement. Besides, the software and hardware requirement are also mentioned in the subchapter

3.2 Research Framework

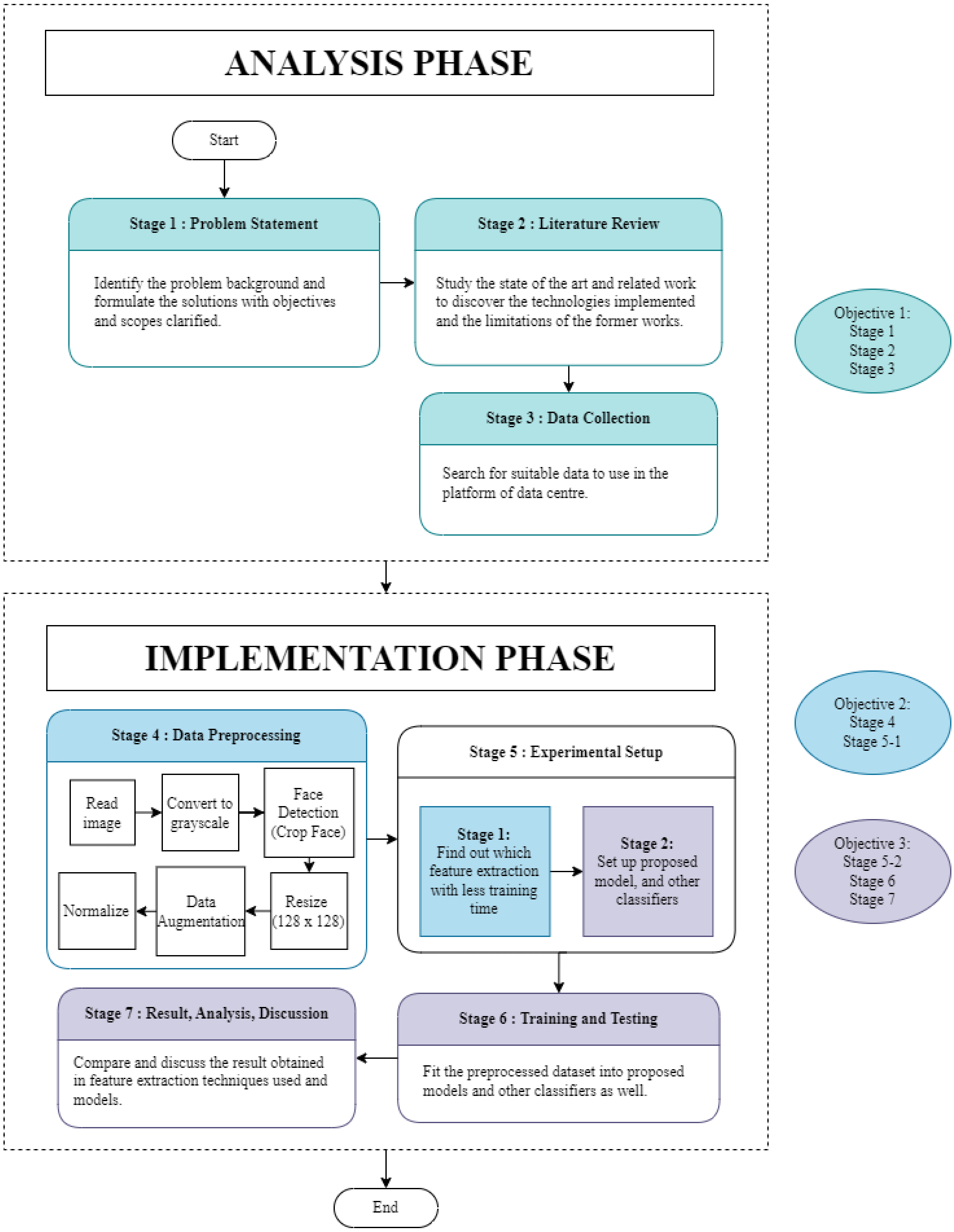

Figure 3.1 Overview of the Research Framework

3.2.1 Analysis Phase

In stage 1, the problem statement was identified along with the objectives and scopes. The solutions were proposed in this stage based on the problem background. In stage 2, literature review was performed to study the state of the art and the related works from the former research. People with ASD express emotions atypically through their faces and it is hard to make interpretation precisely to understand their emotional states. Many research works have proposed a model to classify the emotions of the autistic individuals by static images or real time videos. However, the computational time of training a model is costly (Guimarães et al. 2023) and it becomes a research trend to reduce the computational time required with a large data size. Therefore, it is crucial to reduce the computational time for the model. The first objective was accomplished in this stage whereas the state of the art has been explored. In stage 3, data collection was carried out. It needed to search for a suitable dataset for training through various platforms. Finally, the dataset decided to be used in this research was the raw facial expression images of autistic individuals found in Kaggle, provided by Talaat (2023) combined with another dataset, which was provided by Mashuque Alamgir et al. (2023).

Datasets

The dataset was obtained from the website Kaggle with the URL: https://www.kaggle.com/datasets/fatmamtalaat/autistic-children-emotions-dr-fatma-m-talaat, which was collected by Gerry through internet searches and recompiled by Talaat (2023). It consisted of 1327 images of autistic children with 6 basic emotions such as 48 natural, 350 joy, 200 sadness, 67 anger, 30 fear and 63 surprise. The images were varied in size, fidelity and facial orientation.

The number of images in fear, surprise, anger and natural were too small and this leaded to imbalanced data and biased emotion classification. Besides, overfitting problems would also occur if the dataset was too small. Therefore, another dataset was requested and combined with this dataset to increase the dataset size. The dataset was collected and recompiled by Mashuque Alamgir et al. (2023) in their introduced novel of facial expressions database of autistic children. The emotions in this dataset were happy, sadness, neutral and angry. These classes were combined with the joy, sadness, natural and anger in the previous dataset.

The emotion classes like fear and surprise were dropped because there was a significant difference between the other classes in terms of the number of images. This was to prevent imbalanced data, biased classification and overfitting problems. Moreover, the scope of this research focused on classifying happy, sad, neutral and angry only. Hence, there were a total of 782 images in the current dataset for the 4 classes.

3.2.2 Implementation Phase

Preprocessing

The dataset underwent preprocessing to standardize the data. Firstly, it started with converting the data loaded class by class into grayscale. This was to reduce the channel from 3 to 1. Then face detection was carried out to crop the face detected for the image. The cropped face image was resized into (128, 128). Then it carried out data augmentation using the cropped face image. After that, it continued with normalization. However, different approaches would have different methods in normalizing the data. Feature extraction approaches such as LDA, PCA and PCA+LDA used Standard Scaler, which is a built-in function provided by sklearn, to carry out normalization. For the approach facial landmark localization, it used histogram equalizer to normalize the image to improve the contrast. Then the dataset was split into 80% training and 20% validation.

3.2.2.1 Experimental Setup

Stage 1: Feature Extraction

Experimental setup in stage 1 aimed to find out which feature extraction technique used the least training time. 4 approaches would be tested out in this stage, such as facial landmarks localization, LDA, PCA and LDA+PCA.

The performance metrics to measure the 3 methods were time consumed and the shape of the output, which would be used in the next stage later. The training time taken for the proposed model was measured by using the built-in library “time” in Python. The code snippets were added at the start and the end of the training loop to calculate the elapsed time. The shape of the extracted feature in an array was printed out to see the dimensionality reduction.

Stage 2: Proposed Model and Other Classifiers

Stage 2 of experimental setup focused on the proposed model and other classifiers. Various classifiers such as SVM, KNN, MLP, KNN, Random Forest, Decision Tree and AdaBoost were built along with the proposed model using the same preprocessing method and feature extraction approach (which consumed the least computational time in stage 1). In other words, the preprocessing and feature extraction approach were constant along the experiment to better investigate the results. The performance metrics used were classification accuracy and confusion matrix, which consisted of precision, recall, f1-score and a heatmap that showed the prediction results for each class. Graphs of loss and accuracy were plotted to investigate the learning curve of the proposed model.

3.2.3 Performance Measurement

The approach for the performance measurement was classification accuracy and confusion matrix. It aimed to evaluate the performance of a model. These two were used to compare the result of the proposed model with the other classifiers. The classification accuracy described the overall accuracy of the emotion classification prototype while the confusion matrix helped to visualize the performance of the model.

The classification accuracy is the ratio of the number of correct classified emotions to the total number of the classification. It is transformed into percentage by multiplying the ratio with 100%. The calculation is summarized as below:

$$Accuracy = \frac{(Number\ of\ Correct\ Classification)}{(Total\ Number\ of\ Classification\ Made)} \times 100\%$$

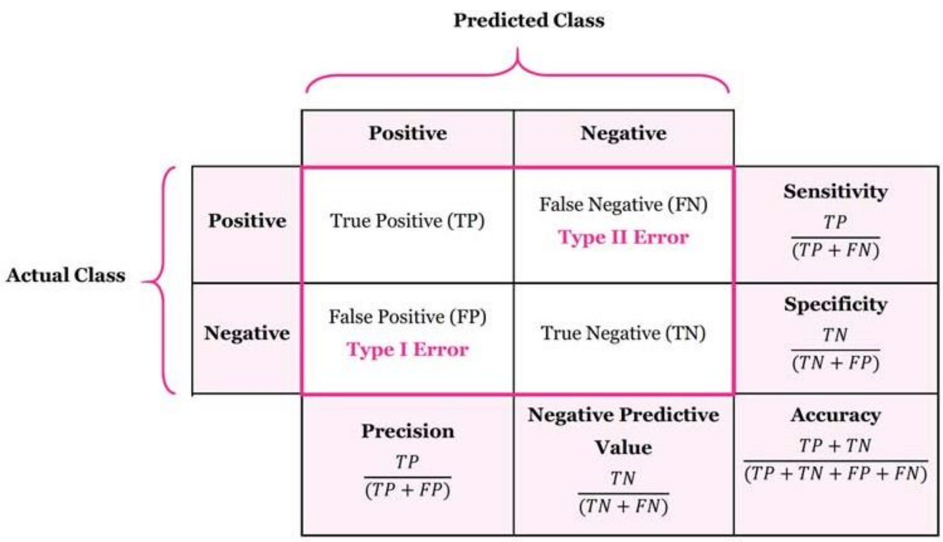

Figure 3.2 Confusion Matrix

Figure 3.2 shows an example of confusion matrix. It helps to depict the actual and predicted class in a heatmap. It visualizes the actual and predicted class. It can generate the other evaluation metrics, such as sensitivity, specificity, precision, recall, which is as known as sensitivity, and accuracy by carrying out arithmetic calculation. It has 4 categories, which are True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). TP refers to the number of the correct classified positive class, TN refers to the number of the correct classified negative class, FP refers to the number of the incorrect classified positive class while FN refers to the number of the incorrect classified negative class. Type I Error refers to the FP classification while Type II Error refers to the FN classification. Meanwhile, the training time of the proposed model is demonstrated using the accuracy and the number of iterations in a line graph.

3.2.4 Software and Hardware Requirement

The proposed method was executed on TensorFlow with Keras APIs, and built-in libraries imported from OpenCV. The research was working on Acer Nitro AN515-58, 12th Gen Intel(R) Core(TM) i7-12650H 2.70 GHz with 16GB installed RAM in x64 operating system.

3.3 Chapter Summary

In summary, the research started with the problem identification and the formulation of solutions. The objectives and scope of the research were clarified also in this stage to ensure the research was on track. Next, literature review was carried out to learn more about the state of the art and related works to identify the limitations and improvements to be made and the technologies implemented by the previous work to figure out what technologies were suitable for this research. Afterwards, it started to look for the data collection of the autistic individuals. The dataset was collected from the website Kaggle. Since the data was raw image of autistic people's emotions, the dataset required preprocessing and standardization which would be carried out in the next stage.

In the next stage, it started with data preprocessing. The image data collected from Kaggle were converted into grayscale and then cropped using the face detection approach, which refers to the Haar Cascade Classifier provided by OpenCV. After the face was cropped out, the image was resized into (128, 128). Then it underwent data augmentation followed by normalization. Normalization method for PCA, LDA and PCA+LDA was Standard Scaler while normalization method for facial landmark localization was histogram equalizer. Afterwards, it went to the experimental setup, which consisted of 2 stages. Stage 1 focused on finding out the least computational time consumed while stage 2 used the approach found in stage 1 to do feature extraction and design the proposed model and other classifiers. Then it started to train and test the proposed model and other classifiers. Then the result obtained in stage 1 and 2 would be compared and discussed using the performance metrics such as classification accuracy and confusion matrix.

CHAPTER 4

RESEARCH DESIGN

4.1 Introduction

This chapter discusses the proposed solutions and the experiment design, which is also known as the experimental setup. It mainly focuses on explaining the flow of the proposed method and the techniques used. Preprocessing carried out includes grayscale, face detection, resize, data augmentation and normalization. The concept of the experimental setup and the proposed model are explained in detail in this chapter.

4.2 Proposed Solution and Experimental Setup

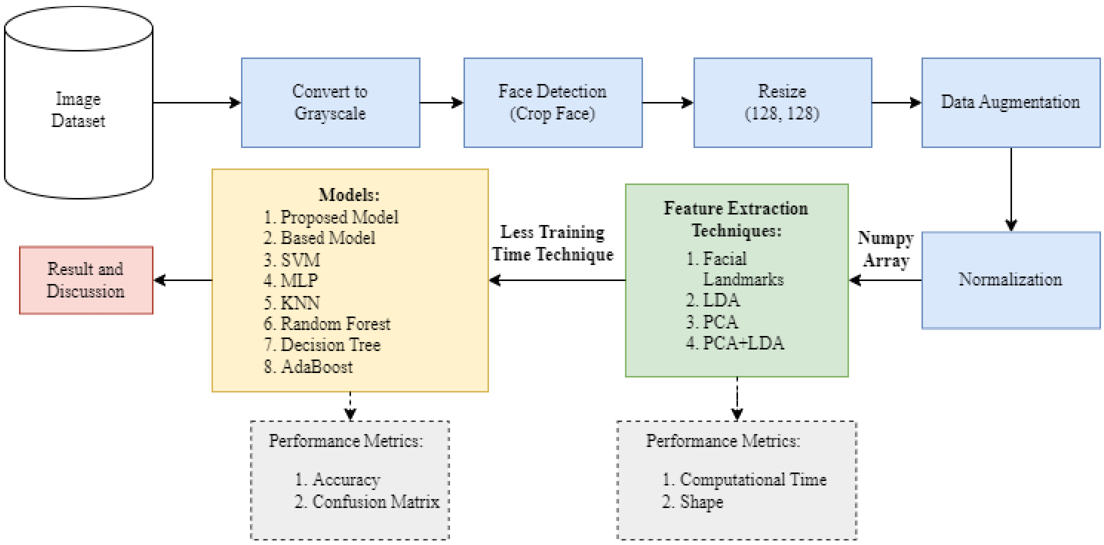

Figure 4.1 Flow of the Experimental Setup

Figure 4.1 shows the flow of the experimental setup. The image dataset is loaded and converted to grayscale. The grayscale image then undergoes face detection to crop the face out. The image then is resized into (128, 128) for standardization. After that, data augmentation is carried out followed by normalization. The normalized image data undergoes 4 feature extraction approaches, such as facial landmark localization, LDA, PCA and PCA+LDA. The performance metrics for these 4 approaches are computational time and dimensionality shape for the extracted data. Next, among these 4 approaches, the least computational time is chosen to implement in creating models. All models, including the proposed model, based model, SVM, MLP, KNN, Random Forest, Decision Tree and AdaBoost, use this feature extraction technique to learn and predict the emotion of autistic individuals. The performance metrics to evaluate the models are classification accuracy and confusion matrix, which consists of precision, recall and f1-score. Moreover, graphs of accuracy and loss are plotted to depict the learning performance in the proposed model. Finally, the results obtained in the feature extraction and classification experiments are analyzed and discussed.

4.2.1 Preprocessing

The dataset consists of different resolutions raw images. Therefore, preprocessing is important to standardize the data. The preprocessing sequence is converting into grayscale, face detection (where the face image is cropped), resizing, data augmentation and normalization. These preprocessing steps are carried in a for-loop to iterate the images for each emotion class. It joins the current directory with the subdirectory which consists of classes to form a file path and iterate the image inside. The image file path is stored temporarily to pass to the face detection function. This function carries out grayscale on the image first before starting to crop the face with the help of Haar Cascade Classifier, which is provided by OpenCV. In the end, it will return a (128, 128) grayscale cropped face image to the function call.

After that, the grayscale cropped face image undergoes data augmentation, which consists of rotation, flip, brightness and blurring in sequence. This is to increase the number of images and improve the dataset quality. The original cropped face image is stored in an empty list first along with its label in another list. The augmented image with its label comes next. After all the iterations are done, the feature list and the label list is converted into a NumPy array. Normalization is the next step. However, different feature extraction approaches use different methods of normalization. For LDA, PCA and PCA+LDA, the method used is Standard Scaler. The image is flattened to make it suitable to fit in LDA, PCA and PCA+LDA. Meanwhile, facial landmark localization uses histogram equalizer for normalization. The extracted data is in the form of vectorized landmarks, which consist of the coordinates and center of gravity of each landmark predicted. Last but not least, the preprocessed data is split into 80% training set and 20% validation set.

4.2.1.1 Feature Extraction

There are 4 feature extraction approaches needed to be tested out to find out the least computational time approach, such as facial landmark localization, LDA, PCA and PCA+LDA. Therefore, a timer is started before fitting the preprocessed images to the respective feature extraction function, and ended after the process is finished. The total time taken is printed in the unit of seconds. Similarly, the shape of the extracted features is printed for checking the shape reduced.

4.2.1.2 Facial Landmark Localization

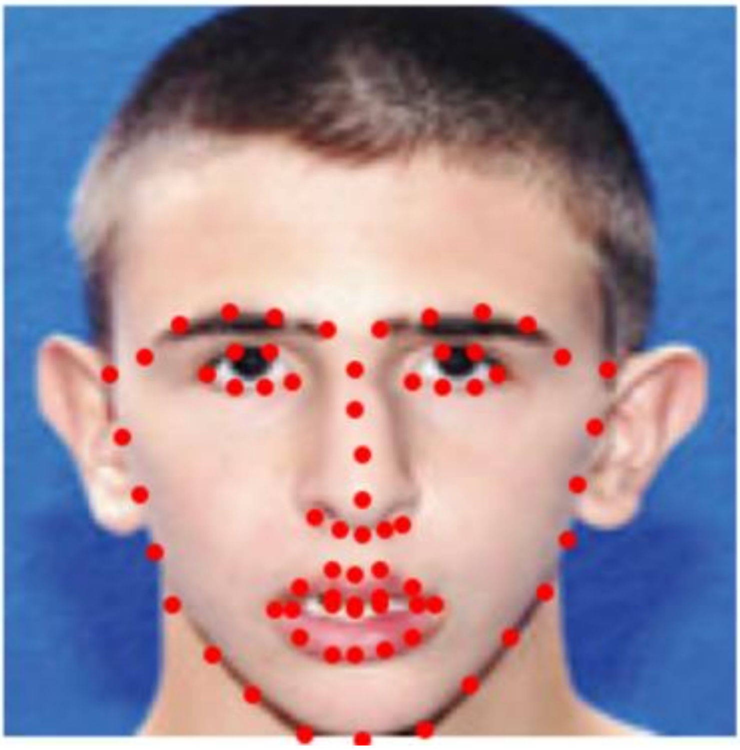

Figure 4.2 Example of 68 Facial Landmarks Predictor

Facial landmark localization is an approach that detects facial features. The library used here is dlib, which is a popular toolkit for machine learning in face recognition and facial landmark detection. The facial landmark predictor used is 68 face landmarks. Figure 4.2 shows an example of the 68 facial landmarks predicted on a human face. The markers plotted on the facial landmarks give a summarization of facial expressions.

The facial landmarks detected are stored in two lists, one is for x-coordinate list (x-list), and one is for y-coordinate list (y-list) (Ahafeez, n.d.). Then, it needs to find out the mean of x-list and y-list to find out the central deviance, such as x-central and y-central (Ahafeez, n.d.). These calculations are served constructing an array that containing of vectorized landmarks, which stand for x-list, y-list, distance and angle between the centre of gravity of each landmark (x-central, y-central), which refers to the (Ahafeez, n.d.). The extracted facial landmarks are stored in a NumPy array.

4.2.1.3 LDA

LDA is initialized with 3 n_components since the maximum number of n_component is the number of classes - 1. The split training set and validation set are fitted to the LDA classifier by using fit_transform() and transform() respectively. This is because fit_transform() will combine the data with the label while transform() will not do so. Using transform() will just simply transform the data in the validation set. A scatter plot will be plotted to show the performance of the extracted data.

4.2.1.4 PCA

A bit similar to LDA, it is required to determine the number of n_component. This n_component is the number of principal components that are kept. PCA is initialized with ‘randomized’ svd_solver and set whiten to true. Then the training set is fit to PCA by using fit_transform() while the validation set is fit to PCA by using transform() using the same reason as in LDA. A scatter plot will be plotted to show the performance of the extracted data.

4.2.1.5 PCA+LDA

PCA+LDA is the combination feature extraction approach. The method begins with PCA followed by LDA. It uses the same values in both PCA and LDA that was carried out before. Similarly, a scatter plot will be plotted to show the performance of the extracted data.

4.2.2 Proposed Model Architecture

The proposed model architecture continues with the feature extraction part. The least computational time approach will be applied to the proposed model and the other classifiers.

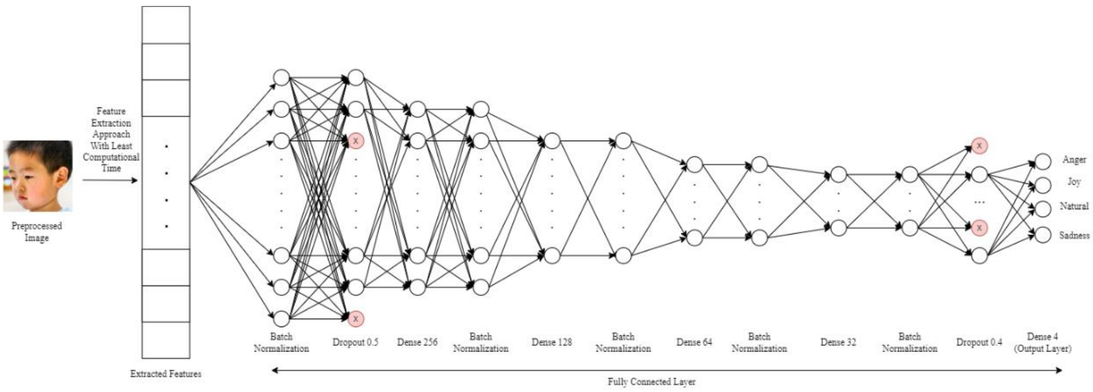

Figure 4.3 Feature Extraction Approach Continue with Fully Connected Layer

Figure 4.3 illustrates the architecture of the proposed model with CNN based fully connected layer. The convolutional and pooling layers are substituted with the least computational time feature extraction approach, leaving fully connected layer for the autistic facial emotion classification. This design is to reduce the computational load as CNN architecture mainly consumes a lot of computational time for the convolutional and pooling layers (Mishra, 2021). After the preprocessed image undergoes the feature extraction, the extracted features then enter the fully connected layer, where the machine learns the patterns to classify the autistic facial emotions.

The fully connected layer is designated as in Figure 4.4. The input shape of the model is the input shape of the extracted features. There are 5 Batch Normalization, 4 Dense with ‘ReLu’ activation function layers and 1 Dense with ‘softmax’ activation function for the output layer. In addition, there are 2 Dropout layers, which are added at the beginning and in the end before the output layer.

The Batch Normalization layer is added before every dense layer to ensure the inputs are normalized while transitioning to another layer. This improves the duration of the proposed model training time and makes the proposed model more stable. The proposed model is moderately complex with 12 layers, as it requires to predict the autistic facial emotions which are hard to recognize and identify the patterns. The dense layers are down sampling by using a divider of 2 to decrease uniformly the units in the dense layers while transitioning across the layers. This is to compress the learned features uniformly into more compact representations by reducing the number of neurons so that the proposed model can learn better the complex patterns of the autistic facial emotions.

The proposed model applies L2 regularizer in the first dense layer only with 0.0005 to prevent overfitting by adding a penalty to the loss function. This can generalize the proposed model so that it can predict well for the unseen data. Similarly, Dropout layers are added at the beginning and end of the layer architecture with 0.5 and 0.4 respectively to prevent overfitting. This is because the Dropout layer drops a percentage of neurons randomly, to force the proposed model to learn the redundant representations of the fitted data.

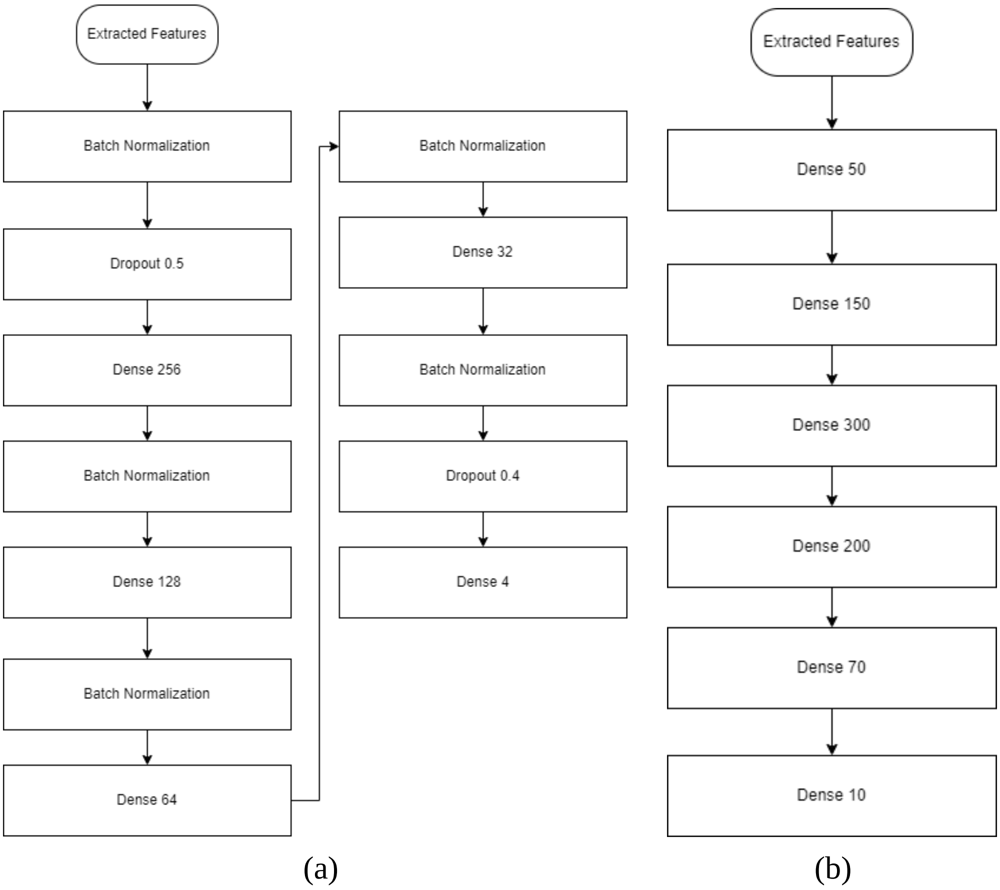

Figure 4.4 (a) Proposed Model Architecture Layout, (b) Based Model Architecture Layout

Figure 4.4 (a) shows the layout of the proposed model. This architecture takes the model architecture from the research project by DhruvDholakiaCE (n.d.) as reference to modify. The proposed model uses the same activation function for the dense layers and output layer as the referenced model. There are several modifications have been made. Firstly, 5 Batch Normalization layers are added before each dense layer. Secondly, there are 2 Dropout layers are added at the beginning and in the end of the layer architecture with value 0.4 and 0.5 respectively. Moreover, the proposed model uses L2 regularizer in the first dense layer while the referenced model does not do so. Last but not least, the units of the dense layers for the proposed model are decremented uniformly by a divisor of 2 and finally the last dense layer is set to the number of emotion classes to be classified. Meanwhile, the units in the dense layers for the referenced model are incremented and decremented not uniformly before the output layer.

Hyperparameters

Table 4.1 Hyperparameters of the Proposed Model

Hyperparameters	Value	Description
Number of Layers	12	5 Batch Normalization layers 2 Dropout layers 5 Dense layers
Learning Rate	0.0010	Default learning rate
Number of Epochs	50	More number of epochs to train the model to ensure it converges to the optimized accuracy
Batch Size	32	Default batch size
Shuffle	True	Shuffle the training data to prevent the model learns the sequence instead of the patterns
Callbacks	Early Stopping	Monitor = 'val_accuracy' patience = 10 restore_best_weights = True
	ReduceLROnPlateau	Monitor = 'val_accuracy' Factor = 0.5 Patience = 5 min_lr = 1e-6

Table 4.1 shows the hyperparameters used for the proposed model. The referenced model also uses the hyperparameters listed in Table 4.1 except for the number of layers as both model architectures are different. The number of layers for the referenced model is 6 while the number of layers for the proposed model is 12 based on Figure 4.4. The learning rate is set to default and the learning rate will be modified by the callback "ReduceLROnPlateau" whenever the validation accuracy is not improved for 5 epochs. The number of epochs is set to 50 to ensure that the accuracy has converged. The batch size is set to default, 32 which is a moderate quantitative value because larger batch size requires more computational time.

Shuffle is set to true to ensure the model learns the pattern representation instead of the sequence representation. Lastly, early stopping is implemented to monitor and stop the model learning whenever it starts to overfitting with 10 patience waits.

Other Classifiers

SVM, MLP, KNN, Random Forest, Decision Tree and AdaBoost have their own built-in classifier in sklearn. Therefore, it just needs to import the required libraries from sklearn to test the extracted features out. The classification accuracy and confusion matrix will be shown to see the performance.

4.2.2.1 Performance Metrics

The performance metrics here are classification accuracy and confusion matrix. Only the proposed model will have graphs of loss and accuracy. The classification accuracy will be printed with 2 decimal places while the confusion matrix will be shown in heatmap to visualize the actual and predicted labels. There are metrics in the confusion matrix such as precision, recall and f1-score for the analysis.

4.3 Chapter Summary

The experimental setup starts with preprocessing, which consists of converting into gray scale, face detection to crop out the face detected, resizing into 128x128 and normalization. There are 2 different ways to carry out normalization. One is using Standard Scaler which is a built-in library provided by sklearn for PCA, LDA and PCA+LDA and another is using histogram equalizer for facial landmarks localization. After that, the extracted features will be stored in a NumPy array. The same goes for the labels. Then, the extracted features with the labels will be split into training and validation sets in the ratio 80:20.

The first experiment stage begins with finding out the least computational time of feature extraction approach among LDA, PCA, PCA+LDA and facial landmarks localization. For LDA, PCA and PCA+LDA, the training set is fitted using their respective built-in functions to extract the data. Meanwhile, dlib library is used to predict facial landmarks of the training set to extract the data. The time function is used to record down the time taken for each approach to complete.

The second experiment stage continues with the previous stage. The feature extraction approach with the least computational time will be used for the models. The proposed model layout is designated based on the model architecture by DhruvDholakiaCE (n.d.), where the extracted features are connected with fully connected layers. This is because the feature extraction part has been done in the previous stage, so the convolutional and pooling layers are excluded. Batch normalization and dropout layers are used to prevent overfitting occurrences. The same reason goes to the kernel regularizer applied. Other classifiers such as MLP, SVM, KNN, Random Forest, Decision Tree and AdaBoost will use the same training and validation sets to analyze and compare the results. The graphs of loss and accuracy will be plotted for the proposed model to visualize the performance. In addition, the classification accuracy and confusion matrix, as the performance metrics, will be shown for all models.

CHAPTER 5

RESULT AND DISCUSSION

5.1 Introduction

This chapter consists of the results obtained using the research design. It will analyze the result and discuss the result in detail. It mainly focuses on the research design stage 1 and stage 2, which refer to feature extraction and model classification. Moreover, the problems encountered during the research will also be discussed in the chapter.

5.2 Feature Extraction Result

Table 5.1 Feature Extraction Result

Approach	Computational Time					Average	Dimensionality Shape
Trials	1	2	3	4	5		
Facial Landmark Localization	1.3358	1.1310	0.6775	0.6556	0.6813	0.8962	(625, 268)
LDA	1.1066	1.2037	1.1101	1.1050	1.1277	1.1306	(416, 3)
PCA	0.3018	0.3024	0.2877	0.2840	0.2898	0.2328	(416, 64)
PCA + LDA	0.5519	0.6179	0.3452	0.3450	0.3254	0.4371	（416, 3）

Table 5.1 shows the result of feature extraction approaches such as facial landmarks localization, LDA, PCA and PCA+LDA. The result obtained showed that PCA consumed the least amount of computational time to complete the feature extraction among the 4 approaches. The average time taken was 0.2328s, which

outranked the other approaches. However, the dimensionality shape reduced by PCA was only (416, 64), which was the second-best reduced dimensionality shape. In contrast, LDA took the longest time to finish feature extraction to reduce the dimensionality shape into (416, 3). PCA+LDA performed secondly with an average of 0.4371s and dimensionality shape of (416, 3). Unfortunately, facial landmarks localization took the third longest computational time to complete feature extraction. Moreover, the dimensionality shape of the features extracted by facial landmarks localization was (416, 268).

It was found that PCA was the least computational time among the approaches. The time consumption involved computing the covariance matrix of the data and eigenvalue decomposition. This explained why it was relatively fast as it involved matrix operations without iterative processes. Meanwhile, LDA required more computational time to compute the scatter matrices to solve a generalized eigenvalue problem. This computation was complex and high consumption of computational time. It was clear to see the difference between LDA and PCA+LDA in terms of computational time. Without using PCA beforehand, LDA required more computational time to extract features. In PCA+LDA, PCA was applied at first to reduce the dimensionality shape of data and therefore LDA only extracted the important features using a smaller dimensionality shape of data which indirectly reduced the computational time consumed. For facial landmarks localization, the computational time consumed was reasonable as there were many stages of detection for the dlib 68 landmarks predictor. It required scanning through the image, predicting the landmarks' locations and recording down the coordinates for each predicted landmark.

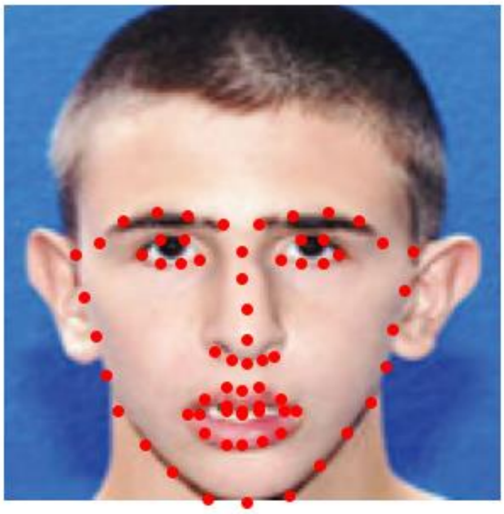

Figure 5.1 Sixty-Eight Facial Landmarks Predicted

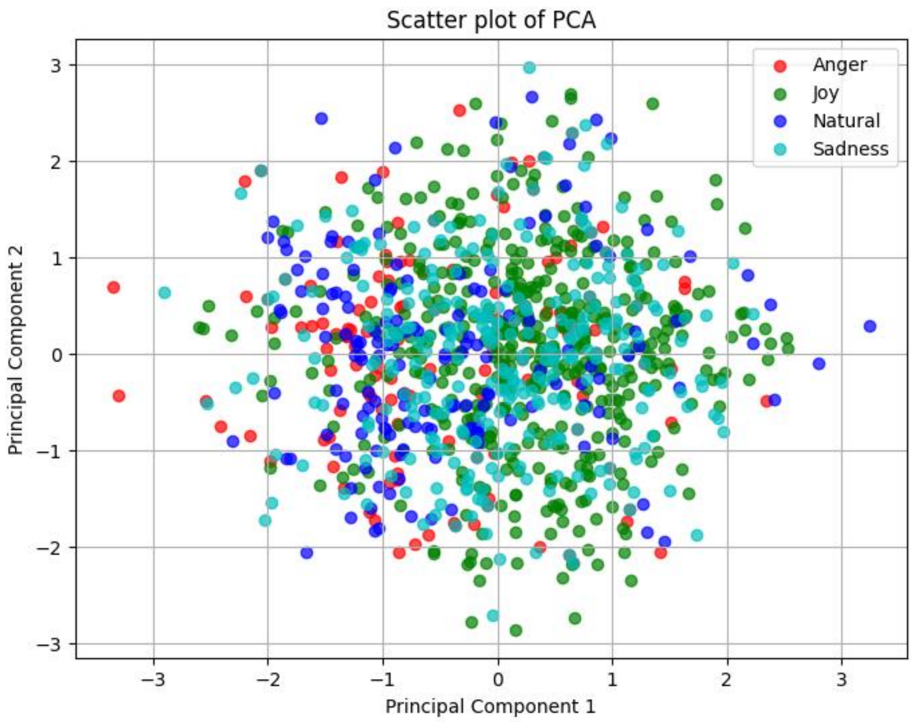

Figure 5.2 Scatter Plot of PCA

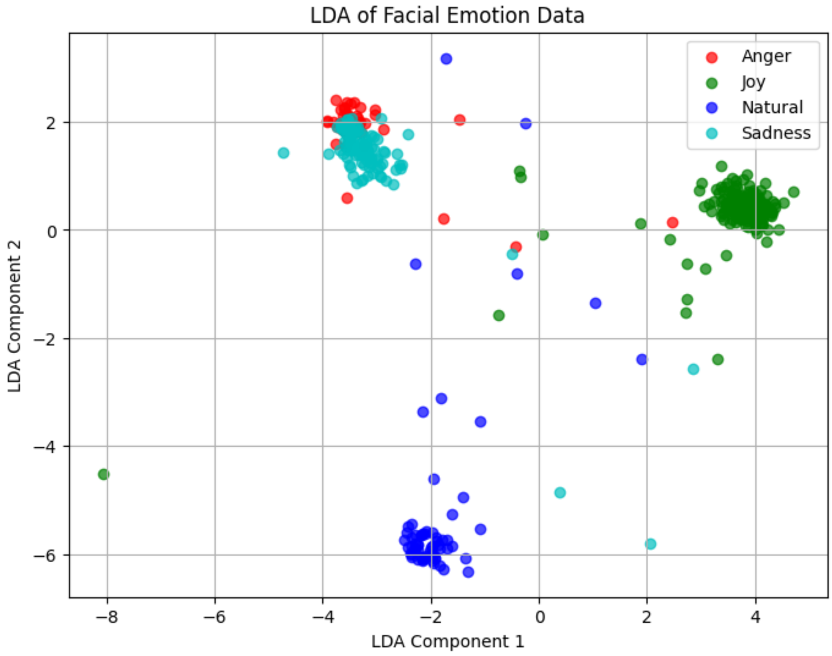

Figure 5.3 Scatter Plot of LDA

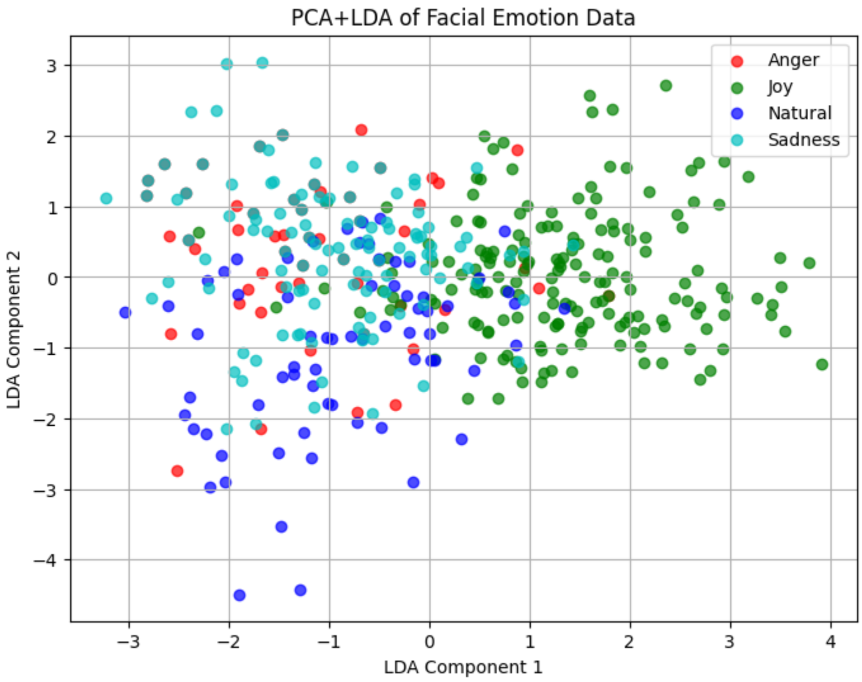

Figure 5.4 Scatter Plot of PCA+LDA

Figure 5.1 shows the 68 facial landmarks predicted using dlib library. The markers plotted on the image represented the facial landmarks predicted. This showed that the predictor from dlib library worked well. The locations of each landmark were used to train and validate the model at the next stage.

Figure 5.2 shows the scatter plot of PCA, depicting the clustering of each class, anger, joy, neutral and sadness. There was not a clear clustering for each class, as all classes were mixed. This was because PCA did not use class labels to fit transform the data. Moreover, the purpose of PCA was to maximize the variance instead of class separability.

Different from PCA, LDA's purpose was to maximize the class separation while minimizing the spread within each class. It was required to use class labels to transform the data. Figure 5.3 shows the scatter plot of LDA, showing the clustering of each class. The classes of anger and sadness were not clustered well. This was because the facial expressions of the autistic individuals were not expressed different significantly from each other.

Figure 5.4 shows the scatter plot of PCA+LDA, showing a clear direction of each class spreading. With the combination of PCA and LDA, it maximized the variance and class separability, which was why the clusters were spreading towards a specific direction.

5.3 Model Result

Referenced Model

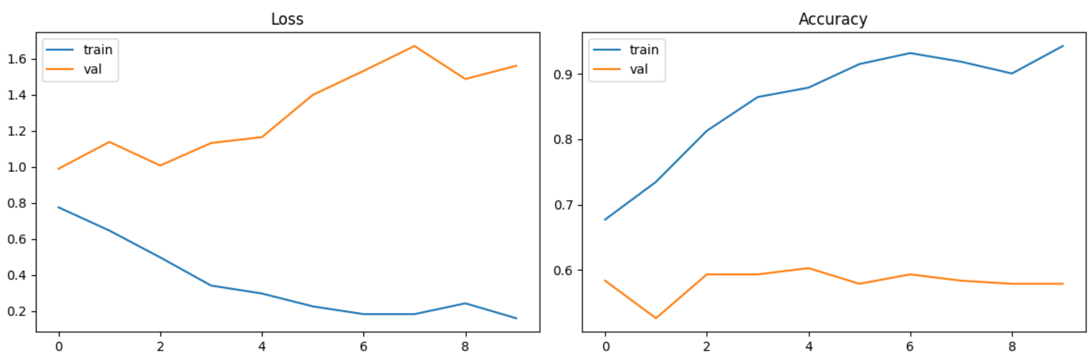

Figure 5.5 Referenced Model's Loss and Accuracy Graphs

Figure 5.5 shows the referenced model's loss and accuracy graphs. The left one represents the loss graph while the right one represents the accuracy graph. Based on the loss graph, it was clearly shown that the referenced model was overfitting starting from the beginning. The validation accuracy did not increase starting from epoch 0 while the accuracy could reach up to 0.9. It was stopped at epoch 10 due to early stopping callback. This is because of the small dataset size fitted although the referenced model's architecture is simple.

Proposed Model

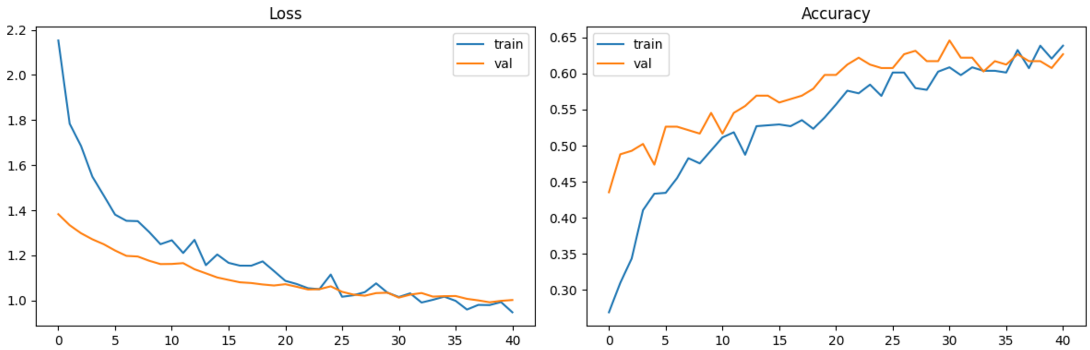

Figure 5.6 Proposed Model's Loss and Accuracy Graphs

Figure 5.6 shows the proposed model's loss and accuracy graphs. The left one represents the loss graph while the right one represents the accuracy graph. These two graphs visualized the learning performance of the proposed model. Based on the loss graph, the training loss starting from 2.2 decreased to 0.9 in a smooth curve with a little fluctuation. Meanwhile, the validation loss starting from 1.4 decreased to 1.0 in a smooth curve. The difference between the training loss and validation loss was not significant. This indicated that the proposed model was not overfit thanks to the early stopping in the callbacks.

In terms of accuracy learning curve, the training accuracy starting from 0.25 increased to 0.65 while the validation accuracy starting from 0.43 increased to 0.62. The validation accuracy became consistent starting from 30 to 40, indicating that the proposed model was not learning anymore due to lack of data. Therefore, it proved that the proposed model improved the performance and reliability of the referenced model.

Overall

Table 5.2 Model Classification Result

Model	Trials	Accuracy	Precision				Recall				F1-Score			
			A	J	N	S	A	J	N	S	A	J	N	S
Proposed Model	1	62.20%	0.27	0.78	0.47	0.50	0.09	0.88	0.28	0.70	0.14	0.82	0.35	0.58
	2	65.07%	0.55	0.78	0.58	0.49	0.18	0.91	0.52	0.57	0.27	0.84	0.55	0.53
	3	64.59%	0.38	0.76	0.58	0.53	0.15	0.89	0.48	0.62	0.22	0.82	0.53	0.57
	4	65.55%	0.60	0.80	0.57	0.50	0.18	0.88	0.45	0.68	0.28	0.84	0.50	0.58
	5	64.11%	0.78	0.76	0.48	0.49	0.21	0.90	0.48	0.55	0.33	0.82	0.48	0.52
	Average	64.30%	0.52	0.78	0.54	0.50	0.16	0.89	0.44	0.62	0.25	0.83	0.48	0.56
Referenced Model	1	46.41%	0.00	0.48	0.00	0.38	0.00	0.91	0.00	0.25	0.00	0.63	0.00	0.30
	2	46.89%	0.00	0.46	0.00	0.64	0.00	1.00	0.00	012	0.00	0.63	0.00	0.21
	3	53.11%	0.00	0.55	0.39	0.55	0.00	0.95	0.31	0.29	0.00	0.69	0.35	0.38
	4	55.50%	0.00	0.69	0.36	0.43	0.00	0.79	0.17	0.70	0.00	0.73	0.23	0.53
	5	58.37%	0.33	0.66	0.43	0.48	0.03	0.91	0.31	0.52	0.06	0.77	0.36	0.50
	Average	52.06%	0.07	0.57	0.24	0.50	0.07	0.91	0.16	0.38	0.01	0.69	0.18	0.38
SVM	1	56.46%	0.24	0.82	0.38	0.42	0.21	0.82	0.38	0.45	0.23	0.82	0.38	0.43
	2	56.46%	0.24	0.82	0.38	0.42	0.21	0.82	0.38	0.45	0.23	0.82	0.38	0.43
	3	56.46%	0.24	0.82	0.38	0.42	0.21	0.82	0.38	0.45	0.23	0.82	0.38	0.43
	4	56.46%	0.24	0.82	0.38	0.42	0.21	0.82	0.38	0.45	0.23	0.82	0.38	0.43
	5	56.46%	0.24	0.82	0.38	0.42	0.21	0.82	0.38	0.45	0.23	0.82	0.38	0.43
	Average	56.46%	0.24	0.82	0.38	0.42	0.21	0.82	0.38	0.45	0.23	0.82	0.38	0.43
MLP	1	64.11%	0.40	0.84	0.44	0.50	0.24	0.89	0.38	0.61	0.30	0.87	0.41	0.55
	2	63.16%	0.35	0.85	0.44	0.47	0.18	0.89	0.41	0.59	0.24	0.87	0.43	0.52
	3	61.72%	0.40	0.85	0.39	0.44	0.24	0.90	0.38	0.50	0.30	0.87	0.39	0.47
	4	63.64%	0.38	0.86	0.41	0.49	0.24	0.89	0.41	0.57	0.30	0.88	0.41	0.53
	5	65.55%	0.41	0.89	0.52	0.48	0.27	0.89	0.45	0.61	0.33	0.89	0.48	0.54
	Average	63.64%	0.41	0.86	0.44	0.48	0.23	0.89	0.41	0.58	0.29	0.88	0.42	0.52
KNN	1	50.72%	0.33	0.69	0.25	0.45	0.33	0.75	0.34	0.30	0.33	0.72	0.29	0.36
	2	50.72%	0.33	0.69	0.25	0.45	0.33	0.75	0.34	0.30	0.33	0.72	0.29	0.36
	3	50.72%	0.33	0.69	0.25	0.45	0.33	0.75	0.34	0.30	0.33	0.72	0.29	0.36

	4	50.72%	0.33	0.69	0.25	0.45	0.33	0.75	0.34	0.30	0.33	0.72	0.29	0.36
	5	50.72%	0.33	0.69	0.25	0.45	0.33	0.75	0.34	0.30	0.33	0.72	0.29	0.36
	Average	50.72%	0.33	0.69	0.25	0.45	0.33	0.75	0.34	0.30	0.33	0.72	0.29	0.36
Random Forest	1	58.37%	0.58	0.65	0.54	0.44	0.21	0.92	0.24	0.43	0.31	0.76	0.33	0.43
	2	58.85%	0.54	0.65	0.62	0.45	0.21	0.95	0.17	0.45	0.30	0.77	0.27	0.45
	3	60.77%	0.55	0.70	0.46	0.47	0.18	0.96	0.21	0.50	0.27	0.81	0.29	0.48
	4	59.81%	0.55	0.68	0.45	0.45	0.18	0.97	0.17	0.46	0.27	0.80	0.25	0.46
	5	60.77%	0.58	0.69	0.59	0.45	0.21	0.93	0.34	0.45	0.31	0.79	0.43	0.45
	Average	59.71%	0.56	0.67	0.53	0.45	0.20	0.95	0.23	0.46	0.29	0.79	0.31	0.45
Decision Tree	1	46.89%	0.23	0.67	0.39	0.35	0.24	0.64	0.41	0.36	0.24	0.66	0.40	0.35
	2	47.85%	0.21	0.70	0.35	0.37	0.21	0.66	0.38	0.39	0.21	0.68	0.37	0.38
	3	48.80%	0.16	0.73	0.29	0.41	0.15	0.71	0.34	0.39	0.16	0.72	0.31	0.40
	4	44.98%	0.14	0.67	0.31	0.36	0.12	0.64	0.38	0.38	0.13	0.66	0.34	0.37
	5	47.37%	0.19	0.71	0.33	0.35	0.18	0.69	0.38	0.34	0.18	0.70	0.35	0.34
	Average	47.18%	0.19	0.70	0.33	0.37	0.18	0.67	0.38	0.37	0.18	0.68	0.35	0.37
AdaBoost	1	52.63%	0.50	0.72	0.38	0.36	0.15	0.73	0.34	0.52	0.23	0.72	0.36	0.42
	2	52.63%	0.50	0.72	0.38	0.36	0.15	0.73	0.34	0.52	0.23	0.72	0.36	0.42
	3	52.63%	0.50	0.72	0.38	0.36	0.15	0.73	0.34	0.52	0.23	0.72	0.36	0.42
	4	52.63%	0.50	0.72	0.38	0.36	0.15	0.73	0.34	0.52	0.23	0.72	0.36	0.42
	5	52.63%	0.50	0.72	0.38	0.36	0.15	0.73	0.34	0.52	0.23	0.72	0.36	0.42
	Average	52.63%	0.50	0.72	0.38	0.36	0.15	0.73	0.34	0.52	0.23	0.72	0.36	0.42

Table 5.2 shows the classification result of all models, including the proposed model. The average of each performance metric was calculated. The proposed model had the highest accuracy (64.30%) among the classifiers. The referenced model had the accuracy of 52.06%, which was the second lowest accuracy. Meanwhile, MLP scored 63.64% accuracy, which was the second highest accuracy classifier. Random forest had 59.71% accuracy while decision tree had the lowest accuracy (47.18%). The performance metrics for SVM, KNN and AdaBoost were consistent, with 56.46%, 50.72% and 52.63% respectively.

For neural networks, such as MLP, the referenced model and the proposed model, the accuracy was varied based on what the machine learns. The deviation of the proposed model based on the 5 trials was from 62.20% to 65.55 with 3.35 deviation difference. Meanwhile, the deviation of the referenced model was from 46.41% to 58.37% with 11.96 deviation difference. The proposed model deviation difference was smaller than that of MLP. The deviation of MLP was from 61.72% to 65.55% with 3.83 deviation difference. The referenced model's deviation difference was greatest while the proposed model's deviation difference was the smallest. The proposed model deviation difference was smaller than that of MLP.

In terms of precision, recall and f1-score, it showed that the emotion joy was well learned by most of the classifiers as the precision for this class was always the lead. However, the other emotion classes still had rooms to be improved as the model did not learn well about the features for anger, natural and sadness to do emotion classification for the autistic individuals. The precision, recall and f1-score for 3 of these emotions had not reached the optimized value. The concern of the autistic individuals could not express their facial emotion well and the characteristics were hard to identify were a big challenge in emotion classification using image. Besides, the number of emotions such as anger, natural and sadness in the dataset was lesser than joy. This led to bias classification, which had been portrayed by the smaller recall values of the other emotion classes such as anger and natural, which indicated that the number of classes predicted correctly was smaller.

5.4 Chapter Summary

Based on the result, it was concluded that PCA was the feature extraction approach that consumed the least computational time compared to facial landmarks localization, LDA and PCA+LDA. The proposed model, with 64.30% accuracy, performed better than the other classifiers, such as SVM, MLP, KNN, Random Forest, Decision Tree and AdaBoost. Moreover, it was proved that the proposed model had successfully improved the referenced model in terms of accuracy and reliability such that the accuracy was improved from 52.06% to 64.30% accuracy and the proposed model was not overfit while the referenced model was overfit. The emotion class, joy, had a greater number of images than the others and hence, this led to joy being more likely to be predicted out of the other 3 classes. This was portrayed by the values of recall of each class.

CHAPTER 6

CONCLUSION

6.1 Project Achievements

The first objective of the project was to investigate the state of the art of facial expression recognition on ASD individuals using machine learning approaches. This objective was achieved in the analysis stage, which was clarified in the proposed methods. The literature reviews were carried out to achieve this objective so that it helped to further understand the current research progress and trends in this field. Furthermore, this objective was necessary to determine which approaches to be experimented with.

The second objective was to propose a FER model to classify emotions for individuals with ASD with less training time using image inputs. This objective was achieved in the implementation phase. There were 2 different stages, one undergoing the feature extraction to find out the feature extraction approach with the least computational time, while the other one focused on the classification performance of classifiers to determine which classifier was better. Therefore, the second objective was accomplished when PCA was proved to be the feature extraction method with the least computational time among the 4 approaches.

The third objective was to evaluate the reliability and the accuracy of the developed facial expression recognition prototype for individuals with ASD by comparing the results with the other approaches. This objective was achieved in the comparison between the proposed model and the other classifiers, taking the training results from the second stage of the experimental setup, which focused on the performance of the proposed model and the other classifiers. The results showed that the proposed model had the highest classification accuracy compared to the others. In addition, the proposed model successfully improved the accuracy of the referenced model. The reliability of the proposed model was portrayed in the loss graph plotted, which had shown that the loss and validation loss were decreasing to 1.0 and the difference between 2 curves was not significant. The proposed model’s loss and accuracy graphs showed it successfully improved the reliability of the referenced model.

6.2 Contribution

The contribution in this work was successfully compared LDA, PCA, PCA+LDA and facial landmarks localization and found out that PCA consumed the least amount of computational time among approaches. In addition, this work successfully proposed a facial emotion model for autistic people with the highest accuracy with 64.30% compared to the other classifiers while using the same preprocessing and feature extraction methods.

In terms of knowledge, this research identified PCA as the most computational efficient feature extraction approach among LDA, PCA+LDA and facial landmarks localization. This proved that the computation of the covariance matrix and eigenvalue decomposition in PCA consumed smaller portions of computational resources compared to the other approaches. This provided a clear guideline in selecting feature extraction approaches with lesser computational time for model training.

Moreover, this research found that LDA consumed more computational time compared to PCA+LDA. This was because PCA+LDA reduced the dimensionality shape of the data first with the help of PCA and LDA transformed the data with the dimensionality reduced data with lesser computation time consumed. This showed that the relationship between the dimensionality shape of the data and the computational time. Based on the experimental stage 1, it provided an insight in understanding the trade-offs between different feature extraction approaches and their impact on computational efficiency.

Based on the comparative performance analysis of the classification models, the proposed model successfully improved the referenced model in terms of accuracy and reliability. Moreover, this work contributed valuable insights into the strengths and weaknesses of each classifier in the autistic facial emotion recognition. The proposed model with a customized fully connected layer had the highest accuracy among the classifiers. This showed the feasibility of utilizing a customized fully connected layer after a feature extraction approach. This also helped in understanding which classifier was weak in the autistic facial emotion classification.

6.3 Suggestions for Improvement and Future Works

The improvement that could be made is to collect more images of the facial expressions of the individuals with ASD. The current dataset used is considered small and it is not enough to reach the optimal accuracy as the model will overfit immediately. Nevertheless, finding an alternative way to increase the model accuracy in the autistic individuals' emotions classification using small dataset can be the future improvement to be made.

Therefore, solving the fundamental problem first, the future work is to increase the dataset size of the facial emotions of autistic individuals. Moreover, finding an alternative way to increase the model classification accuracy using small dataset.

6.4 Conclusion

In conclusion, the research's objectives were successfully achieved. The first objective was achieved by looking into the literature reviews of the facial emotion recognition of autistic people.

The second objective was achieved in the experimental setup stage 2, which focused on finding out the feature extraction method with the least computation time, and it was found that PCA took the least time consumption among the 4 approaches.

The third objective was achieved in the experimental setup stage 2, which focused on the performance of the classifiers, and the result and discussion stage in the implementation phase. A comparison between the proposed model and the other classifiers as well as the referenced model was made, and the comparison result showed that the proposed model had the highest classification accuracy. Moreover, the loss graph showed that the model was not overfitting, as the difference between the 2 curves was not significant. This showed that the proposed model was reliable and improved the reliability of the referenced model.

Lastly, there are some improvements to be made, such as collecting more facial expressions of autistic people to increase the dataset size and finding an alternative solution to increase the classification accuracy while using small dataset.

REFERENCES

Abdullah, N. E., & Al-Allaf, A. F. (2021). Facial Expression Recognition (FER) of Autism Children using Deep Neural Networks. In *2021 4th International Iraqi Conference on Engineering Technology and Their Applications (IICETA)*. https://doi.org/10.1109/iiceta51758.2021.9717550

Ahafeez. (n.d.). *GitHub - ahafeez8/Emotion-Recognition-OpenCV-Dlib*. GitHub. https://github.com/ahafeez8/Emotion-Recognition-OpenCV-Dlib/tree/master

Alvari, G., Furlanello, C., & Esposito, G. (2021). Is Smiling the Key? Machine Learning Analytics Detect Subtle Patterns in Micro-Expressions of Infants with ASD. *Journal of Clinical Medicine*, *10*(8), 1776. https://doi.org/10.3390/jcm10081776

Awan, A. A. (2022). *Deep Learning - A Tutorial for Data Scientists*. https://www.datacamp.com/tutorial/tutorial-deep-learning-tutorial

Bartzatt, R. (2014). Determination of Dermal Permeability Coefficient (Kp) by Utilizing Multiple Descriptors in Artificial Neural Network Analysis and Multiple Regression Analysis. *Journal of Scientific Research and Reports*, *3*(22), 2884–2899. https://doi.org/10.9734/jsrr/2014/13125

Basics About Autism Spectrum Disorder (ASD) | NCBDDD | CDC. (2022, December 9). Centers for Disease Control and Prevention. https://www.cdc.gov/ncbddd/autism/facts.html

Black, D. W., MD, & MPH, J. E. G. M., JD. (2014). *DSM-5 Guidebook* (5th ed., Vol. 3).

Brewer, R., Biotti, F., Catmur, C., Press, C., Happé, F., Cook, R. J., & Bird, G. (2016). Can Neurotypical Individuals Read Autistic Facial Expressions?

Atypical Production of Emotional Facial Expressions in Autism Spectrum Disorders. *Autism Research*, *9*(2), 262–271. https://doi.org/10.1002/aur.1508

Carter, M. J. (2014). Diagnostic and Statistical Manual of Mental Disorders, 5th ed. *Therapeutic Recreation Journal*, *48*(3).

Darwin, C. (1872). The expression of the emotions in man and animals. In *John Murray eBooks*. https://doi.org/10.1037/10001-000

DhruvDholakiaCE. (n.d.). *Supervised-PCA-vs-LDA-as-Pre-processing-for-Image-Classification/ghodsi_supervised_pca.ipynb at main · DhruvDholakiaCE/Supervised-PCA-vs-LDA-as-Pre-processing-for-Image-Classification*. GitHub. https://github.com/DhruvDholakiaCE/Supervised-PCA-vs-LDA-as-Pre-processing-for-Image-Classification/blob/main/ghodsi_supervised_pca.ipynb

Dino, H. I., & Abdulrazzaq, M. B. (2020). Comparison of four classification algorithms for facial expression recognition. *PolyTechnic/PolyTechnic (Erbil. Online)*, *10*(1), 74–80. https://doi.org/10.25156/ptj.v10n1y2020.pp74-80

Educative. (n.d.). *What is feature extraction?* Educative: Interactive Courses for Software Developers. https://www.educative.io/answers/what-is-feature-extraction

Ekman, P. (2000). Basic Emotions. In T. Dalgleish & M. Power (Eds.), *Handbook of Cognition and Emotion* (pp. 45–60). John Wiley & Sons.

García-Martín, E., Rodrigues, C. F., Riley, G., & Grahn, H. (2019). Estimation of energy consumption in machine learning. *Journal of Parallel and Distributed Computing*, *134*, 75–88. https://doi.org/10.1016/j.jpdc.2019.07.007

Gardener, H., Spiegelman, D., & Buka, S. L. (2009). Prenatal risk factors for autism: comprehensive meta-analysis. *British Journal of Psychiatry*, *195*(1), 7–14. https://doi.org/10.1192/bjp.bp.108.051672

GeeksforGeeks. (2023). Supervised and Unsupervised learning. *GeeksforGeeks*. https://www.geeksforgeeks.org/supervised-unsupervised-learning/

GeeksforGeeks. (2024, February 22). *Random Forest algorithm in machine learning*. GeeksforGeeks. https://www.geeksforgeeks.org/random-forest-algorithm-in-machine-learning/

Gondhi, N. K., Kour, N., Effendi, S., & Kaushik, K. S. (2017). An efficient algorithm for facial landmark detection using haar-like features coupled with corner detection following anthropometric constraints. In *2017 2nd International Conference on Telecommunication and Networks (TEL-NET)*. https://doi.org/10.1109/tel-net.2017.8343517

Guimarães, M., Carneiro, D., Palumbo, G., Oliveira, F., Oliveira, Ó., Alves, V., & Novais, P. (2023). Predicting model training time to optimize distributed machine learning applications. *Electronics*, *12*(4), 871. https://doi.org/10.3390/electronics12040871

Gupta, A. (2018, June 8). *Introduction to Deep Learning: Part 1*. AIChE. https://www.aiche.org/resources/publications/cep/2018/june/introduction-deep-learning-part-1

Hassouneh, A., Mutawa, A. M., & Murugappan, M. (2020). Development of a Real-Time Emotion Recognition System Using Facial Expressions and EEG based on machine learning and deep neural network methods. *Informatics in Medicine Unlocked*, *20*, 100372. https://doi.org/10.1016/j.imu.2020.100372

Hodges, H., Fealko, C., & Soares, N. (2020). Autism spectrum disorder: definition, epidemiology, causes, and clinical evaluation. *Translational Pediatrics*, *9*(S1), S55–S65. https://doi.org/10.21037/tp.2019.09.09

Introduction to Deep Learning: What Are Convolutional Neural Networks? Video. (n.d.). https://www.mathworks.com/videos/introduction-to-deep-learning-what-are-convolutional-neural-networks--1489512765771.html

Jeffares, A. (2020, September 15). Supervised vs Unsupervised Learning in 3 Minutes - Towards Data Science. *Medium*. https://towardsdatascience.com/supervised-vs-unsupervised-learning-in-2-minutes-72dad148f242

Kana, R. K. (2022). *The Neuroscience of Autism*. Academic Press.

Kartali, A., Roglic, M., Barjaktarovic, M., Duric-Jovicic, M., & Jankovic, M. M. (2018). Real-time Algorithms for Facial Emotion Recognition: A Comparison of Different Approaches. In *2018 14th Symposium on Neural Networks and Applications (NEUREL)*. https://doi.org/10.1109/neurel.2018.8587011

Kumar, A. (2023, April 13). *PCA vs LDA Differences, Plots, Examples - Data Analytics*. Data Analytics. https://vitalflux.com/pca-vs-lda-differences-plots-examples/

Kurian, A., & Tripathi, S. (2022). Unsupervised Domain Adaptation for Facial Emotion Recognition in Autistic Children. In *Ambient intelligence and smart environments*. IOS Press. https://doi.org/10.3233/aise220022

Kushki, A., Khan, A., Brian, J., & Anagnostou, E. (2015). A Kalman Filtering Framework for Physiological Detection of Anxiety-Related Arousal in Children With Autism Spectrum Disorder. *IEEE Transactions on Biomedical Engineering*, *62*(3), 990–1000. https://doi.org/10.1109/tbme.2014.2377555

LeCun, Y., Bengio, Y., & Hinton, G. E. (2015). Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lee, C., Mower, E., Busso, C., Lee, S., & Narayanan, S. (2011). Emotion recognition using a hierarchical binary decision tree approach. *Speech Communication*, *53*(9–10), 1162–1171. https://doi.org/10.1016/j.specom.2011.06.004

Leung, F., Sin, J., Dawson, C., Ong, J. H., Zhao, C., Veić, A., & Liu, F. (2022). Emotion recognition across visual and auditory modalities in autism spectrum disorder: A systematic review and meta-analysis. *Developmental Review*, *63*, 101000. https://doi.org/10.1016/j.dr.2021.101000

Li, S., Deng, Y., Zhu, Z., Hua, H., & Tao, Z. (2021). A Comprehensive Review on Radiomics and Deep Learning for Nasopharyngeal Carcinoma Imaging. *Diagnostics*, *11*(9), 1523. https://doi.org/10.3390/diagnostics11091523

Liu, B. (2011). *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*. Springer.

Marasamy, P., & Sumathi, S. (2012). *Automatic recognition and analysis of human faces and facial expression by LDA using wavelet transform*. https://doi.org/10.1109/iccci.2012.6158798

Mashuque Alamgir, F., Mostasim Hasnain Saif, S., Md. Hossain, S., Al Hadi, A., & Shafiul Alam, Md. (2023). Facial Expression Database of Autism Spectrum Disorder Children. *European Chemical Bulletin*. https://doi.org/10.48047/ecb

Michel, P., & Kaliouby, R. E. (2003). *Real time facial expression recognition in video using support vector machines*. https://doi.org/10.1145/958432.958479

Mishra, A. (2021, December 29). Metrics to Evaluate your Machine Learning Algorithm. *Medium*. https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234

Mishra, M. (2021, December 15). Convolutional Neural Networks, Explained - Towards Data Science. *Medium*. https://towardsdatascience.com/convolutional-neural-networks-explained-9cc5188c4939

Mollahosseini, A., Chan, D., & Mahoor, M. H. (2016). *Going deeper in facial expression recognition using deep neural networks*. https://doi.org/10.1109/wacv.2016.7477450

Muhammad, I., & Yan, Z. (2015). SUPERVISED MACHINE LEARNING APPROACHES: A SURVEY. *ICTACT Journal on Soft Computing*, *05*(03), 946–952. https://doi.org/10.21917/ijsc.2015.0133

Ousley, O. Y., & Cermak, T. (2014). Autism Spectrum Disorder: Defining Dimensions and Subgroups. *Current Developmental Disorders Reports*, *1*(1), 20–28. https://doi.org/10.1007/s40474-013-0003-1

Owusu, E., Zhan, Y., & Mao, Q. R. (2014). A neural-AdaBoost based facial expression recognition system. *Expert Systems With Applications*, *41*(7), 3383–3390. https://doi.org/10.1016/j.eswa.2013.11.041

Pai, A. (2023). CNN vs. RNN vs. ANN – Analyzing 3 Types of Neural Networks in Deep Learning. *Analytics Vidhya*. https://www.analyticsvidhya.com/blog/2020/02/cnn-vs-rnn-vs-mlp-analyzing-3-types-of-neural-networks-in-deep-learning/

Pratt, M. K. (2020). unsupervised learning. *Enterprise AI*. https://www.techtarget.com/searchenterpriseai/definition/unsupervised-learning

Rajan, S., Chenniappan, P., Devaraj, S., & Madian, N. (2019). Facial expression recognition techniques: a comprehensive survey. *Iet Image Processing*, *13*(7), 1031–1040. https://doi.org/10.1049/iet-ipr.2018.6647

Rani, P. (2019). Emotion Detection of Autistic Children Using Image Processing. In *IEEE International Conference on Image Information Processing*. https://doi.org/10.1109/iciip47207.2019.8985706

Sadiq, R., Rodriguez, M. J., & Mian, H. R. (2019). Empirical Models to Predict Disinfection By-Products (DBPs) in Drinking Water: An Updated Review. In *Elsevier eBooks* (pp. 324–338). https://doi.org/10.1016/b978-0-12-409548-9.11193-5

Sakib, S., Ahmed, N., Kabir, A. I., & Ahmed, H. (2018). An Overview of Convolutional Neural Network: Its Architecture and Applications. *ResearchGate*. https://doi.org/10.20944/preprints201811.0546.v1

Salian, I. (2019, August 20). *Difference Between Supervised, Unsupervised, & Reinforcement Learning | NVIDIA Blog*. NVIDIA Blog. https://blogs.nvidia.com/blog/2018/08/02/supervised-unsupervised-learning/

Sevilla, J. (2022, January 20). Estimating Training Compute of Deep Learning Models. *Epoch*. https://epochai.org/blog/estimating-training-compute

Sharma, S. R., Gonda, X., & Tarazi, F. I. (2018). Autism Spectrum Disorder: Classification, diagnosis and therapy. *Pharmacology & Therapeutics*, *190*, 91–104. https://doi.org/10.1016/j.pharmthera.2018.05.007

Shirai, Y. (1987). Three-Dimensional Computer Vision. In *Springer eBooks*. https://doi.org/10.1007/978-3-642-82429-6

Silva, J. O., Oliveira, I. L., Topolniak, Z. H., & Alvarez, A. G. (2021). *A CNN Approach Implemented to Emotional Facial Expression Recognition for the*

Prevention of Autistic Meltdowns. https://doi.org/10.1109/scla53004.2021.9540183

Simmons, D., Robertson, A. E., McKay, L. S., Toal, E. R., McAleer, P., & Pollick, F. E. (2009). Vision in autism spectrum disorders. *Vision Research*, *49*(22), 2705–2739. https://doi.org/10.1016/j.visres.2009.08.005

Sivasangari, A., Ajitha, P., Rajkumar, I., & Poonguzhali, S. (2019). Emotion recognition system for autism disordered people. *Journal of Ambient Intelligence and Humanized Computing*. https://doi.org/10.1007/s12652-019-01492-y

Smitha, K. G., & Vinod, A. P. (2015). Facial emotion recognition system for autistic children: a feasible study based on FPGA implementation. *Medical & Biological Engineering & Computing*, *53*(11), 1221–1229. https://doi.org/10.1007/s11517-015-1346-z

Song, H., Kang, J., & Lee, S. (2018). *ConcatNet: A Deep Architecture of Concatenation-Assisted Network for Dense Facial Landmark Alignment*. https://doi.org/10.1109/icip.2018.8451375

Sun, L., Dai, J., & Shen, X. (2021). *Facial emotion recognition based on LDA and Facial Landmark Detection*. https://doi.org/10.1109/icaie53562.2021.00020

Takahashi, Y., Murata, S., Idei, H., Tomita, H., & Yamashita, Y. (2021). Neural network modeling of altered facial expression recognition in autism spectrum disorders based on predictive processing framework. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-94067-x

Talaat, F. M. (2023). Real-time facial emotion recognition system among children with autism based on deep learning and IoT. *Neural Computing and Applications*. https://doi.org/10.1007/s00521-023-08372-9

Trevisan, D. A., Hoskyn, M., & Birmingham, E. (2018). Facial Expression Production in Autism: A Meta-Analysis. *Autism Research, 11*(12), 1586–1601. https://doi.org/10.1002/aur.2037

Turabzadeh, S., Meng, H., Swash, R. M., Pleva, M., & Juhar, J. (2018). Facial Expression Emotion Detection for Real-Time Embedded Systems. *Technologies (Basel)*, *6*(1), 17. https://doi.org/10.3390/technologies6010017

Uadmin. (2022, April 19). *What Is Machine Learning (ML)? - I School Online*. UCB-UMT. https://ischoolonline.berkeley.edu/blog/what-is-machine-learning/

Varma, S., Shinde, M., & Chavan, S. S. (2019). Analysis of PCA and LDA features for facial expression recognition using SVM and HMM classifiers. In *Springer eBooks* (pp. 109–119). https://doi.org/10.1007/978-3-030-16848-3_11

Voulodimos, A., Doulamis, A., & Protopapadakis, E. (2018). Deep Learning for Computer Vision: A Brief Review. *Computational Intelligence and Neuroscience*, *2018*, 1–13. https://doi.org/10.1155/2018/7068349

Wang, Y., Li, Y., Song, Y., & Rong, X. (2019). Facial expression recognition based on random forest and convolutional neural network. *Information, 10*(12), 375. https://doi.org/10.3390/info10120375

Wichert, A., & Sa-Couto, L. (2021). *Machine Learning: A Journey to Deep Learning with Exercises and Answers*. World Scientific Publishing Company.

Woolf, B. P. (2009). Machine Learning. In *Elsevier eBooks* (pp. 221–297). https://doi.org/10.1016/b978-0-12-373594-2.00007-1

Appendix A Gantt Chart