

**UNIVERSITY OF GHANA**

**COLLEGE OF BASIC AND APPLIED SCIENCES**



**DETECTING ANGER IN PERSUASIVE SPACES: AN EVALUATION OF FACIAL  
EXPRESSION ALGORITHMS**

**BY  
JACQUELINE ASOR KUMI  
(10485756)**

**THIS THESIS IS SUBMITTED TO THE UNIVERSITY OF GHANA, LEGON IN  
PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE AWARD OF  
MASTER OF PHILOSOPHY IN COMPUTER SCIENCE DEGREE**

**DEPARTMENT OF COMPUTER SCIENCE**

**JULY 2020**

## Declaration

I hereby declare that this dissertation is entirely my work unless otherwise indicated. No part of this dissertation work has been presented as the basis for the award of any degree to this or any other university.

.....  
DATE: ...October 5, 2020  
DR ISAAC WIAFE  
SUPERVISOR



.....  
DR EBENEZER OWUSU  
DATE: ...October 5, 2020  
CO-SUPERVISOR

.....  
DATE: ... October 5, 2020  
KUMI JACQUELINE ASOR  
STUDENT

## **Abstract**

Darwin's influential work on the recognition of the emotions in man and animals served as the starting point for emotion research. Based on his work, the basic emotion was theorized, from which several other emotions have been conceptualised. These emotions are recognised by both verbal and non-verbal form of communication. Facial expression is the significant and leading measure for recognition of emotions as 55% of what we communicate is expressed on our facial expressions. Therefore, facial expression has been applied in diverse fields to detect emotions such as lie detection and in the medical field for pain analysis; resulting in a plethora of algorithms or techniques.

Among the negative emotions, anger is said to be the most frequently experienced emotion yet the most unsatisfactorily handled emotion in both personal and social relations. It is also said to be the emotion that considerably affects the mental state of an individual. Further, its intensities like temper, hostility, annoyance, tantrum, agitation, and rage foster harm to the individual and the surrounding environment as well as have disruptive interpersonal and intrapersonal consequences.

Currently, anger recognition is performed as part of the multiclass emotion classification or done using physiological signals or speech data. To the best of our knowledge, there have not been any studies on how to detect only anger using facial expressions. And even with the existing approach, there have been some identified issues such as the overlapping of emotion: anger, fear, and disgust and the difficulty of some of the facial expressions algorithms in performing a multiclass recognition of emotions. For this reason, we argue that it is key that the recognition of the anger is done accurately. As, the detection of anger would be useful as it will provide useful information about peoples' intensity of anger to manage or control it as unregulated anger sometimes result in aggression or violence.

As such, we want to determine how these facial expression algorithms will perform when used for binary classification, specifically anger recognition. Therefore, we propose a framework of three models: SVM (machine learning algorithm), CNN (deep learning algorithm) and a novel ensemble learning algorithm and PCA as our dimensionality reduction function. The performance of our models was evaluated on JAFFE, CK+ and KDEF datasets. It was observed that our proposed models outperformed the state-of-the-art methods, with mention of our novel ensemble learning model which attained an accuracy of 100% on the JAFFE dataset. Thus we conclude that the proposed methods are effective for the recognition of anger using facial expressions and future work will look at evaluating the performance of these algorithms on a created database of Africans as well as employing these algorithms to detect anger in a persuasive space and persuade the individual from angry to another emotion for example happy.

Keywords: anger, persuasive spaces, anger recognition, facial expression, facial expression recognition, facial expression algorithms, deep learning, machine learning, ensemble learning algorithm, SVM, CNN

## **Dedication**

I dedicate this dissertation to my father and all who served a motivating and driving force for the successful completion of this work.

## **Acknowledgements**

First and foremost, I would like to express my gratitude to my father who consistently motivated me and made provisions available throughout this course. God bless you.

I am also grateful to all who assisted me in one way or the other to the fruitful completion of this course.

## Table of Contents

Declaration.....	I
Abstract.....	II
Dedication.....	IV
Acknowledgements.....	V
Table of Contents.....	VI
List of figures.....	XI
List of tables.....	XIII
List of abbreviations .....	XIV
Chapter 1.....	1
Introduction.....	1
1.1 Motivation and overview.....	1
1.2 Current approach .....	3
1.3 Challenges .....	4
1.4 Our approach and expected contribution.....	4
1.5 Aims and objectives.....	5
1.6 Structure of the thesis .....	5
Chapter 2.....	7
Literature review .....	7
2.1 Introduction .....	7
2.2 Psychological background of emotions.....	7
2.2.1 Models of emotions .....	7
2.3 Modalities of emotion recognition .....	8
2.4 Facial expression modality .....	10
2.6 Typical Facial Expression Recognition (FER) System .....	17
2.6.1 Machine and deep learning FER .....	20

2.6.2 Ensemble learning algorithms .....	26
2.7 Databases for facial expression recognition .....	33
2.8 Limitation of current work and contribution .....	34
2.9 Chapter summary.....	35
Chapter 3.....	37
Methodology .....	37
3.0 Introduction .....	37
3.1 Workflow.....	37
3.1.1 Image acquisition.....	37
3.1.1.1 JAFFE database .....	38
3.1.1.2 CK+ database.....	39
3.1.1.3 KDEF database .....	40
3.1.2 Pre-processing .....	40
3.1.2.1 Face detection.....	41
3.1.2.2 Image enhancement.....	43
3.1.2.2.1 Median blur.....	43
3.1.2.2.1 Histogram equalisation.....	44
3.1.3 Feature extraction .....	44
3.1.3.1 Local Binary Pattern.....	45
3.1.3.2 Histogram of Oriented Gradients .....	47
3.1.4 Feature selection.....	48
3.1.5 Classification models.....	49

3.1.5.1 Support Vector Machine.....	50
3.1.5.2 Convolutional Neural Network (CNN).....	51
3.1.5.3 Ensemble method .....	53
3.1.6 Elements involved .....	54
3.1.6.1 Programming language.....	54
3.1.6.2 Packages and development environments .....	55
3.1.6.2.1 Anaconda .....	56
3.1.6.2.2 Open Source Computer Vision Library (OpenCV).....	56
3.1.6.2.3 Tensorflow.....	57
3.1.6.2.4 Keras .....	57
3.1.6.2.5 Scikit-learn.....	57
3.1.7 Development.....	58
3.1.8 Chapter summary.....	60
Chapter 4.....	61
Experimental setup.....	61
4.1 Introduction .....	61
4.2 Hardware specification.....	61
4.3 Pre-processing stage .....	62
4.3.1 Database pre-processing.....	62
4.3.2 Image pre-processing.....	63
4.3.2.1 Grayscale and resizing of images.....	64
4.3.2.2 Face detection and cropping .....	64
4.3.2.3 Image enhancement .....	65

4.4 Feature extraction .....	66
4.4.1 LBP feature extraction.....	66
4.4.2 HOG feature extraction .....	66
4.4.3 Hybrid feature extraction.....	67
4.5 Feature selection .....	67
4.6 Classification .....	67
4.6.1 Model selection.....	67
4.6.2 Hyperparameters tuning .....	68
4.6.3 Settings and protocols.....	70
4.6.4 Data augmentation.....	70
4.7 Chapter summary.....	70
Chapter 5.....	72
Experimental results and discussion .....	72
5.1 Introduction .....	72
5.2 Evaluation metrics .....	72
5.3 Experiment I (training without data augmentation) .....	73
5.4 Experiment II.....	78
5.5 Discussion.....	86
5.5.1 Introduction .....	86
5.5.2 Performance of the models .....	86
5.5.3 Comparison of the state-of-the-art.....	87
5.6 Limitations.....	90

5.7 Chapter summary.....	91
Chapter 6.....	92
Conclusion .....	92
Bibliography .....	94
Appendix.....	118

**List of figures**

Figure 2.1: modalities of emotion recognition.....9

Figure 2.2 displays the basic emotions from the JAFFE database ..... 13

Figure 2.5: Measurements of facial expressions..... 13

Figure 2.3: processes of facial expression recognition .....20

Figure 3.1 shows sample images of the JAFFE database. ....39

Figure 3.2: displays a sample CK+ images.....40

Figure 3.3: shows KDEF images captured from different angles.....40

Figure 3.4 shows the rectangle features.....42

Figure 3.5 displays the calculation of area using integral image.....42

Figure 3.6: operation of the LBP operator. ....46

Figure 3.7: Three neighbour sets for different (P, R) used to construct a circularly symmetric LBP .....46

Figure 3.8: displays the operation of the LBP operator .....46

Figure 3.9 displays a HOG feature extraction process .....48

Figure 3.10: Steps involved in CNN FER .....53

Figure 3.11: the workflow of our research work .....59

Figure 4.1: Example of a gray scaled and resized CK+ angry image.....64

Figure 4.2: viola jones face detection on the left and a detected and cropped face on the right. ....65

Figure 4.3: a denoised JAFFE image using median blur. ....66

Figure 4.4 displays a CLAHE enhanced JAFFE image.....66

Figure 4.5: transformations applied to a JAFFE image .....71

Figure 5.1: Confusion matrices for the models: SVM, CNN and Ensemble learning for the JAFFE dataset – Experiment I .....76

Figure 5.2: Confusion matrices for the models: SVM, CNN and Ensemble learning for the CK+ dataset – Experiment I .....	77
Figure 5.3: Confusion matrices for the models: SVM, CNN and Ensemble learning for the KDEF dataset – Experiment I .....	78
Figure 5.4: Confusion matrices for the models: SVM, CNN and Ensemble learning for the JAFFE dataset – Experiment II.....	80
Figure 5.5: Confusion matrices for the models: SVM, CNN and Ensemble learning for the CK+ dataset – Experiment II .....	81
Figure 5.6: Confusion matrices for the models: SVM, CNN and Ensemble learning for the KDEF dataset – Experiment II.....	82
Figure 5.7: ROC curves on the JAFFE dataset.....	83
Figure 5.8: ROC curve on CK+ dataset.....	84
Figure 5.9: ROC curve on the KDEF dataset. ....	85

## List of tables

Table 2.1: Descriptions of Action Units, FACS description and their associated facial muscle .....	15
Table 2.2: summarises algorithms utilised by researchers for facial expression recognition..	27
Table 3.1: summary of the package development and environments.....	57
Table 4.1: a summary of the hardware specification. ....	62
Table 4.2: summary of the datasets.....	63
Table 4.3: summary of the datasets after data augmentation.....	71
Table 5.1 shows the performance (accuracy, recall, precision and F1-score) of the JAFFE dataset on the SVM, CNN and ensemble learning models.....	74
Table 5.2 displays the performance (accuracy, recall, precision and F1-score) of the CK+ dataset for the models: SVM, CNN and ensemble learning.....	75
Table 5.3: KDEF dataset performance (accuracy, precision, recall and f1-score) for CNN, SVM and ensemble learning models.....	75
Table 5.4: JAFFE dataset performance (accuracy) on the CNN, SVM and ensemble learning models.....	79
Table 5.5: Comparison of approaches on the JAFFE dataset .....	88
Table 5.6: Comparison of the proposed method with state-of-the-art.....	89
Table 5.7: Comparison of different techniques on JAFFE+ dataset.....	89

## **List of abbreviations**

**AAM --Active Appearance Models**

**Adadelta -- Adaptive Delta**

**Adagrad -- Adaptive Gradient**

**Adam -- Adaptive Momentum**

**ANN -- Artificial Neural Network**

**AUC -- Area Under ROC Curve**

**BU-3DFE -- Binghamton University 3D Facial Expression**

**BVP -- Blood Pressure Rate**

**CK -- Cohn Kanade**

**CK+ -- Extended Cohn Kanade**

**CLAHE -- Contrast Limited Adaptive Histogram Equalisation**

**CNN -- Convolutional Neural Network**

**CPU -- Central Processing Unit**

**DCNN -- Deep Convolutional Neural Network**

**DCT -- Discrete Cosine Transform**

**DGLTP -- Directional Gradient Local Ternary Pattern**

**DNN -- Deep Neural Network**

**ECG -- Electrocardiography**

**EEG -- Electroencephalography**

**ELM -- Extreme Learning Machine**

**EMG -- Electromyography**

**EOG -- Electrooculogram**

**FER -- Facial Expression Recognition**

**Fmris -- Functional Magnetic Resonance Imaging**

**Google Colab -- Google Colaboratory**

**HOG -- Histogram of Oriented Gradients**

**IFED -- Indian Facial Expression Image Database**

**JAFFE -- Japanese Female Facial Expression**

**KDEF -- Karolinska Directed Emotional Faces**

**KNN -- K-Nearest Neighbour**

**LBP -- Local Binary Pattern**

**LDA -- Linear Discriminant Analysis**

**LD-MGAD -- Local Descriptor with Modified Gray value Accumulation Value**

**LDN -- Local Directional Number**

**LDP -- Local Directional Pattern**

**LG -- Logistic Regression**

**LPQ -- Local Phase Quantization**

**LSTM -- Long Short-term Memory**

**MEG -- Magnetoencephalography**

**MLP -- Multilayer Perceptron**

**MOGA -- Multiobjective Genetic Algorithm**

**MRI -- Magnetic Resonance Imaging**

**MUFE -- Mevlana University Facial Expression**

**Nadam -- Nesterov Accelerated Gradient**

**NB -- Naïve Bayes**

**NIRS -- Near-Infrared Spectroscopy**

**NSGA -- Nondominated Sorting Genetic Algorithm**

**PCA -- Principal Component Analysis**

**PCA-LDA -- Principal Component Analysis - Linear Discriminant Analysis**

**PET -- Positron Emission Tomography**

**PSO-KNN -- Particle Swarm Optimization based K-Nearest Neighbour**

**RaFD -- Radboud Faces Database**

**RAM -- Random Access Memory**

**RBF -- Radial Basis Function**

**RF -- Random Forest**

**RMSprop -- Root Mean Square Propagation**

**RNN -- Recurrent Neural Network**

**ROC -- Receiver Operating Characteristic**

**SFEW -- Static Faces in the Wild**

**SFFS -- Sequential Feed Feature Selection**

**SIFT -- Scale Invariant Feature Transform**

**SOM -- Self-Organising Maps**

**SRC -- Sparse Representation based Classifier**

**SVM -- Support Vector Machine**

**TFEID -- Taiwanese Facial Expression Database**

**TPU -- Tensor Processing Unit**

## Chapter 1

### Introduction

#### 1.1 Motivation and overview

The interest in human emotions has spanned many centuries. Darwin's influential work on the expression of the emotions in man and animals served as the starting point for emotion research (Darwin, 1872; Petrushin, 2000). Subsequently, there has been a significant contribution from multidisciplinary fields such as psychology, computer science, medicine, sociology and so on (Mitsuyoshi & Ren, 2013).

Emotion is a complex feeling stimulated by internal and external stimuli which influences behaviour and mental processes resulting in the output of physical and physiological changes (Domínguez-Jiménez, Campo-Landines, Martínez-Santos, Delahoz, & Contreras-Ortiz, 2020). Variants of emotions have been proposed based on discrete theories by different psychologists, although they vary based on the number and type (Ortony & Turner, 1990). However, the most employed forms of emotion recognition are the basic emotions of: surprise, anger, happiness, fear, sadness and disgust, propounded by (Ekman, 1999). Emotions can be distinguished by modalities such as facial expressions, body postures or movement and physiological signals such as electroencephalography (EEG), electromyography (EMG), and electrocardiography (ECG) (Feidakis, 2016; Mitsuyoshi & Ren, 2013). Nevertheless, the facial expression is the most utilised measure as it is non-invasive, inexpensive due to the nonrequirement for hardware, gives easy and accurate detection of emotions as it serves as the medium through which 55% of human communication is displayed (Gonzalez-sanchez, Baydogan, Chavez-echeagaray, Robert, & Burleson, 2017; Mehrabian, 1968). Facial expression is a key factor in human communication revealing an individual's thoughts and emotions naturally during communication (Jameel, Singhal, & Bansal, 2016). Therefore, it can be concluded that the face

is an important feature of the body as it conveys an individual's personality, emotions, thoughts and ideas even before it has been verbalized; playing a significant role in human communication and social interaction (Dhall & Sethi, 2014; Mitsuyoshi & Ren, 2013). The components of the face that help in the expression of emotions include the eyes, eyebrows, mouth, forehead, lips, cheeks, chin, nose. For example, an angry face is characterized by a brow lowering, raising of the upper lid, and tightening of the lid and lip, respectively.

As such, the invention of human-centred interfaces by the next generation computing namely: persuasive computing has brought immense benefits which project the human user to the foreground. This next-generation computing readily responds to human communication, as these interfaces can perceive and understand human emotions and intentions which are communicated by the social and affective signals (Pantic, Pentland, Nijholt, & Hunag, 2007). These human-computer interaction interfaces seek to reshape the behaviour and intentions of individuals as well as seek to improve their health; hence the proposal of the construction of persuasive spaces or the use of persuasive technology to change an individual behaviour or emotion to a predetermined one (Stibe & Wiafe, 2018). Inspired by this vision, the field of human-computer interaction, computer vision and pattern recognition has witnessed colossal transformation including the automated analysis of facial expressions or facial behaviour with machine learning and deep learning algorithms.

Among the negative emotions, anger is said to be the most frequently experienced emotion yet the most unsatisfactorily handled emotion in both personal and social relations (Moritz, 2006). It is also said to be the emotion that considerably affects the mental state of an individual (Kudiri, Said, & Nayan, 2013). Anger is revealed to be a strong emotion influenced by several elements such as biological factors, psychological factors, physiological factors and environmental factors which include family, society and culture as well as also other emotions like fear, which serves as a springboard for anger (Shahsavarani, Noohi, Jafari, Kalkhoran, &

Hatefi, 2015; Zhan et al., 2018). Even though anger is noted solely as a negative emotion, it is a natural emotion which when expressed without restraint enhances an individual health. Anger can serve as a self-defence mechanism under stressful conditions as well as facilitates behaviour as it possesses motivating properties which drives an individual towards a goal-centred action (Moritz, 2006; Shahsavarani et al., 2015).

Despite, anger being a natural emotion and is expressed positively, its intensities like temper, hostility, annoyance, tantrum, agitation and rage can foster harm to the individual and the surrounding environment as well as bring about disruptive interpersonal and intrapersonal consequences (Kassinove, Sukhodolsky, Eckhardt, & Tsytsarev, 1997). On the other hand, when anger is repressed, it causes an individual to be meek and mild, lacking strength and initiative, passive in all circumstances and appearing lifeless always. For this reason, there is the need to recognise anger as it will provide useful information about peoples' intensity of anger so that there can be a regulation or management, as unregulated anger sometimes results in aggression or violence (Moritz, 2006; Shahsavarani et al., 2015).

## 1.2 Current approach

Presently, anger recognition is performed using the following approaches: based on the utilization of either physiological signals or audio or speech data (Chang, Lin, & Zheng, 2012; Chhabra, Vyas, Chatterjee, & Vob, 2017; Deng, Eyben, Schuller, & Burkhardt, 2018) or facial expressions for recognising general or a subset of emotions. To the best of our knowledge, there have not been any studies on how to detect only anger using facial expressions.

Therefore, our work has focused on recognising anger using facial expressions since it is the leading and significant measure among the modalities for anger recognition.

### 1.3 Challenges

Although the facial expression algorithms used for the recognition of the basic emotions or a subset of it has produced excellent results, some problems persist which need to be addressed.

1. Generally, there are some limitations with the multiclass classification of emotions, such as an overlap among the facial expressions namely disgust, anger, and fear. This is as a result of slight distinction among them which gives an untrue representation of the emotions when classified (Pell & Richards, 2011).
2. Further, it is postulated that a majority of the facial expression algorithms have difficulty in performing a multi-class classification, which is attested to with regards to the training time, computational time and the insufficient memory space (Kiran & Kushal, 2016; Shah, Sharif, Yasmin, & Fernandes, 2017).

### 1.4 Our approach and expected contribution

The current systems can detect facial expressions in general or a subset of emotions. To the best of our knowledge, there have not been any studies on how to detect only anger. We argue that anger detection needs to be done accurately, giving a true representation of the emotion. As the detection of anger could provide useful information about the intensity of anger and help to manage or control it, as unregulated anger sometimes results in aggression or violence (Moritz, 2006; Shamsavarani et al., 2015). To test and validate our arguments, our proposed framework will employ both machine learning and deep learning algorithm as well as a novel ensemble learning algorithm for our work (see chapters 3,4 and 5 for details). Our proposed models, in general, outperformed the state-of-the-art methods.

### 1.5 Aims and objectives

The aim of this study to conduct solely anger recognition using facial expression and compare the outcome to the state-of-the-art experiments to determine if the former would obtain higher accuracy.

Objectives:

1. To research on emotions and determine the significant measure for the measuring of emotions.
2. To research literature on facial expression algorithms.
3. To investigate and understand the algorithms used for facial expressions recognition (FER).
4. Integrate state-of-the-art machine learning and deep learning techniques into frameworks for anger recognition.
5. Discuss the performance of the various facial expression algorithms and databases utilised for anger recognition.
6. Compare our results with the outcome of the general FER experiments.

### 1.6 Structure of the thesis

The thesis is organized as follows:

Chapter 1 gives a succinct overview of the background of the study, the problem statement, aims and objectives and expected contribution.

Chapter 2 presents a review of the different algorithms at the three stages in facial expression recognition and provide insight into the background of emotion research as well as some psychology emotions theories.

In chapter 3 of this study, the experimental methodology is described. It involves the exploratory discussion of our different methodologies and datasets.

Chapter 4 details the implementation process. It presents details on the experimental setup, pre-processing, feature extraction and selection and classification.

Chapter 5 begins by investigating the results, evaluate the performance metrics used in adjudicating the performance of the algorithms; highlight what was done, what was achieved, or found, the implications of the results, strengths and limitations as well as recommendations for future research.

Our conclusions, as well as a summary of the work, are drawn in the final chapter of the thesis.

## Chapter 2

### Literature review

#### 2.1 Introduction

The expression of emotions in human and animals by Darwin (1872) in the nineteenth century, served as the premise for research on emotions. In his work, Darwin indicated that both humans and animals exhibit emotions of similar behaviour (Petrushin, 2000). Since then, there has been significant progress in the research on emotions and the past two decades has witnessed contributions from multidisciplinary fields such as psychology, medicine, sociology, neuroscience, endocrinology and computer science and with a colossal number of algorithms for automatic facial expression recognition being developed (Mitsuyoshi & Ren, 2013).

Therefore, in this chapter, we review and discuss relevant literature on the psychological background of emotion research, facial expression recognition and its resulting algorithms.

#### 2.2 Psychological background of emotions

##### 2.2.1 Models of emotions

Emotions can be described as things we feel which are caused by neurons that shoot electrons around the tiny pathways inside the amygdala, the emotion centre of the brain. Emotion can also be described as a complex experience involving related feelings which tends to move one out of his or her individuality (Moritz, 2006; Shamsavarani, Noohi, Jafari, Kalkhoran, & Hatefi, 2015). They come with physical and physiological changes which regulate our behaviour, due to reactions to internal and external stimuli (Domínguez-Jiménez et al., 2020). Emotion is a salient characteristic of humans. It plays a useful role in human communication as well as in the growth and regulation of interpersonal relationships (Ekman, 1999; Kirange & Deshmukh, 2012; Mangalagowri & Raj, 2017). It also affects thoughts, actions as well as the making of decisions (Izard, 2007).

### 2.3 Modalities of emotion recognition

In recognising emotions, several sources of emotional information have been proposed. These sources of emotion information serve as the primary data from which emotions can be inferred. They can be broadly classified into three groups namely biological indicators, behavioural indicators and physiological signals (see figure 1) (Feidakis, 2016; Mitsuyoshi & Ren, 2013). The biological indicators comprise facial expressions and body postures or gestures. The physiological signals are measurements based on electrical signals recording produced by the heart, skin, muscles, and brain. They include electroencephalography (EEG), electromyography (EMG), electrocardiography (ECG), respiration rate, skin conductance, electrooculogram (EOG), blood pressure rate, Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI), Magnetoencephalography (MEG), Functional Magnetic Resonance Imaging (fMRI), Near-Infrared Spectroscopy (NIRS). Further, speech signals and text represent the behavioural indicators for emotion recognition.

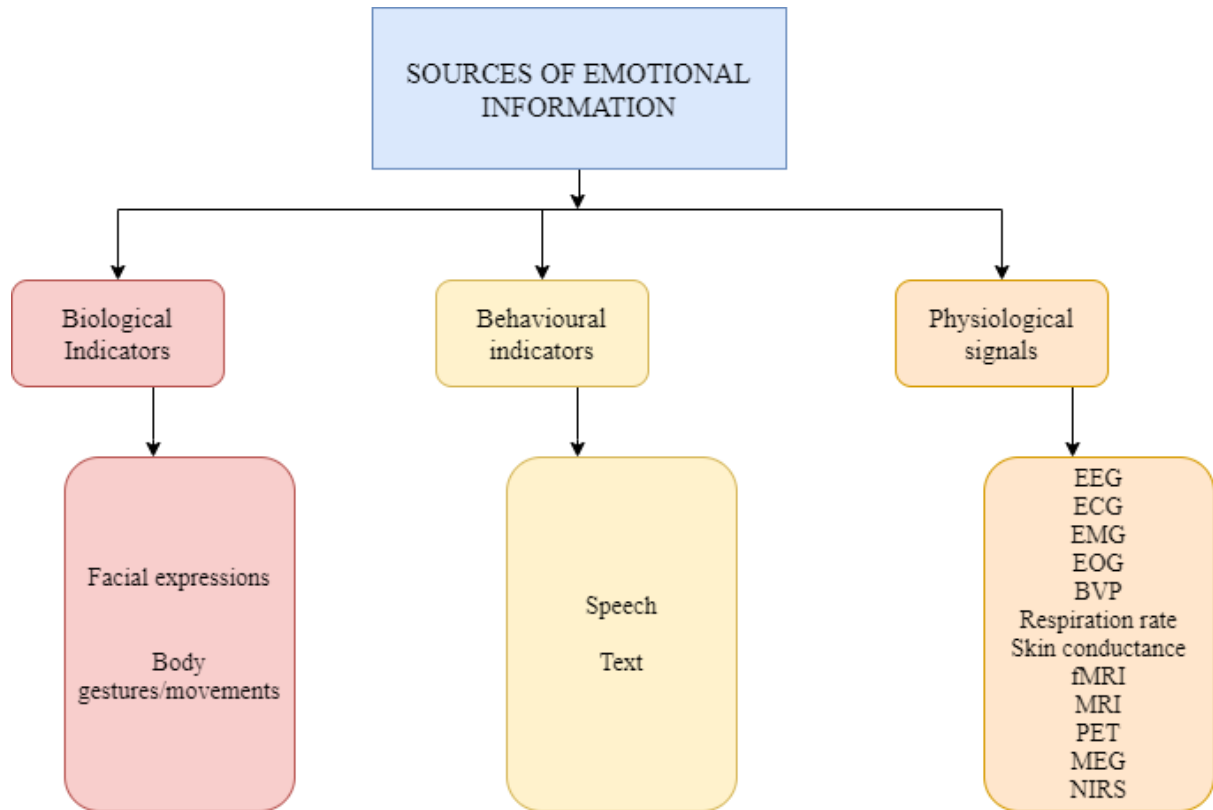


Figure 2.1: modalities of emotion recognition (Mitsuyoshi & Ren, 2013).

Therefore, when performing emotion recognition, either a unimodal or bimodal or multimodal source of the emotional information can be utilised. The unimodal approach involves the use of a technique whilst the bimodal or the multimodal involves the use of 2 or more sources of emotional information. However, from literature, bimodal emotion recognition can also be referred to as multimodal emotion recognition (Busso, Deng, Yildirim, & Bulut, 2004).

As such, drawing on the application of the unimodal and multimodal approach for emotion detection and classification by researchers, several studies for instance (Lalitha, Geyasruti, Narayanan, & Shravani, 2015; Rajvanshi, 2018; L. Sun, Zou, Fu, Chen, & Wang, 2019; Tzirakis, Trigeorgis, Nicolaou, Schuller, & Zafeiriou, 2017; S. Zhang, Zhang, Huang, & Gao, 2018) utilised speech signals for the classification of emotions. Also, researchers (Kaur, Singh, & Roy, 2018; Patil & Behele, 2018) employed Electroencephalography (EEG) whilst (Ahmed,

Bari, & Gavrilova, 2020) employed body gestures or movements and physiological signals deployed by (Chakladar & Chakraborty, 2018; H. Huang, Hu, Wang, & Wu, 2020). In addition, for multimodal emotion recognition, researchers (Chaparro et al., 2018; X. Huang et al., 2015; Yongrui Huang, Yang, Liao, & Pan, 2017; Matlovic, Gaspar, Moro, Simko, & Bielikova, 2016; T. Zhang, Zheng, Cui, Zong, & Li, 2019) adopted facial expressions and electroencephalogram signals. Also, Kudiri, Said, & Nayan (2016) used facial expressions and speech for emotion detection and classification. Keshari & Palaniswamy (2019) also utilised facial expression and body gestures and (Zheng, Liu, Lu, Lu, & Cichocki, 2019) performed speech emotion recognition fusing facial expressions and speech signals. Further, Nguyen, Nguyen, Sridharan, Dean, & Fookes (2018) performed a multimodal emotion recognition comprising of the measures: facial expression, pose, body movements, and voice.

#### 2.4 Facial expression modality

Research on facial expression dates back to the ancient times, making facial expressions a recognised and important modality among the non-verbal forms of communication as it can be inferred from the literature above that facial expressions is mostly combined with other modalities when performing emotion recognition (Zheng et al., 2019). Darwin's work on the universality of facial expressions of emotions across different cultures and different tribes served as a foundation for the empirical study on facial expressions (Darwin, 1872). Thus, making facial expressions the only measure with developed frameworks as it has been researched thoroughly in the past few decades (Keshari & Palaniswamy, 2019). Additionally, among the indicators for emotion recognition, facial expressions are argued to be a significant measure as well as the leading measure for emotion recognition as facial expressions convey 55% of what humans communicate and 7% and 38% through language and speech respectively (Mehrabian, 1968; Pantic et al., 2007). Emotions can be easily and accurately detected from the face (Abhang, Gawali, & Mehrotra, 2016; Happy, Patnaik, Routray, & Guha, 2017).

Furthermore, the use of the facial expression for emotion recognition has several advantages such as its non-invasiveness and relatively cheapness, as it does not involve any physical contact with the user employing sensors in comparison to the case of collecting EEG signals or has any requirement for expensive hardware (Gonzalez-sanchez et al., 2017). Facial expressions are useful in deciphering an individual's thoughts or state of mind during a conversation (Jameel et al., 2016). It also serves as the most real indicators that lend information on age, truthfulness, temperament, personality and the emotional state of a person (Apte, Basavaraj, & Nithin, 2016; Pantic & Bartlett, 2007). Hence, it can be concluded that the face is an important feature of the body, as it conveys an individual's personality, emotions, thoughts and ideas even before it has been verbalized; playing a significant role in human communication and social interaction (Dhall & Sethi, 2014; Mitsuyoshi & Ren, 2013).

Further, Darwin's research established the foundation for the conceptualisation of emotions and thus it received attention among various psychologists. Ekman (1970) validated Darwin theory on the universality of emotions irrespective of the tribes and cultures when he proposed the discrete theory of emotion namely the basic emotion. From that, several psychologists have theorised variants of emotions based on the basic theories, for example, Plutchik and Russell models (Ortony & Turner, 1990; Plutchik, 1987; Russell & Pratt, 1980). These conceptualised emotions vary according to the type and number, even though they are all borne out of Darwin and Ekman's universality of emotions. Nonetheless, the most employed emotions for emotion research based on these discrete theories are the basic emotions which are modelled by six classes: happiness, disgust, fear, surprise, sadness and anger (Ekman, 1999). The basic emotions are considered to be universal across different cultures and different people and are used in describing the affective states of individuals (Ekman, 1970; Haq & Jackson, 2010). Each basic emotion is characterised by a unique facial expression (Ekman, 1977) (refer to section 2.6 for details).

Conventionally, emotions are usually classified based on positivity and negativity (An, Ji, Marks, & Zhang, 2017). However, there are other classifications, such as the two-dimensional (2D) model proposed by Russell & Pratt (1980) and the eight primary emotions by (Plutchik, 1987). The 2D model is based on valence and arousal, pleasantness, and unpleasantness, on the other hand, the eight primary emotions represent the positivity and negativity of emotions as having been grouped in pairs such as joy and sadness.

Among the negative emotions, anger is revealed to be the most frequently experienced emotion yet the one most unsatisfactorily handled emotion in both personal and social relations (Moritz, 2006). It is also said to be the emotion that considerably affects the mental state of an individual (Kudiri et al., 2013).



Anger



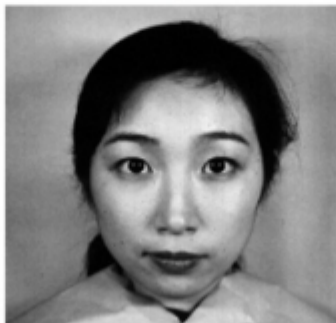
Disgust



Fear



Happiness



Neutral



Sadness

Figure 2.2 displays the basic emotions from the JAFFE database (Lyons, Akamatsu, Kamachi, & Gyoba, 1998). Left-to-right from top row: anger, disgust, fear, happiness, neutral and sadness

## 2.5 Measurements of facial expressions

The facial expression serves as the representation of signals which send forth messages (emotions) such as disgust, anger, happiness, surprise, fear, and sadness (Ekman, 1977). In detecting emotions from facial expressions, several methods have been proposed over the past two decades. Duchenne de Boulogne (1862) employed electric shocks to identify the various muscles motions. This approach helped identify the combination of the muscle motions that express an emotion. Moving on, photography was used to aid in deciphering emotions from facial expressions (Darwin, 1872). Ekman (1977) introduced the Facial Affect Scoring Technique (FAST), harnessing the discovery by Duchenne and Darwin. FAST's method of measuring emotions was not able to depict the different facial appearances for the basic emotions, nonetheless, it could correctly distinguish between pleasant and unpleasant facial expressions. Ekman (1977) further proposed the Facial Action Coding System (FACS) which is based on the study of the anatomical basis of facial expression. Thus coining a definition for facial expression, which is defined as the movement of one or more muscles of the face which conveys the emotions of an individual (Ekman, 1970). FACS could distinguish all visible facial behaviour; utilising a set of action units (AU) to describe every facial muscle activity. An AU represents a certain component of facial muscles movement (see Table 2.1) and emotions are described by a set of AU's. The facial muscles are displayed in Figure 2.2.

The components of the face that help in the expression of emotions include the eyes, eyebrows, mouth, forehead, lips, cheeks, chin, nose. For example, an angry face is characterized by a brow lowering, raising of the upper lid and tightening of the lid and lip respectively which corresponds to the following action units and muscles: AUs 4,5,7 and 23 and depressor

glabellae, depressor supercilli, corrugator supercilli, levator palpebrae superioris, orbicularis oculi and orbicularis orris respectively.

Yet, there were drawbacks with the use of FACS and this included:

1. early researchers who employed the use of Facial Action Coding System had to manually code the action units to unravel the basic emotion, making this process both labour and cost-intensive (Ekman, 1977). Also, learning to be proficient in coding in FACS could take 100 hours of training and an extra 2 hours for coding each image sequence (Littlewort, Bartlett, & Lee, 2007).
2. Irrelevant information is exposed by the FACS codes which setbacks data-driven facial expression recognition methods. For the sufficient number of facial expressions, it is not possible to have a training database which contains the existing number of about 7000 AU combinations, as such it leads to poor generalisation performance (Fasel, Monay, & Gatica-Perez, 2004).

Advances in technology have contributed immensely to the analysis of emotions, begetting automated facial expression recognition. This improvement in technology helps in extending the use of FACS beyond the behavioural research disciplines and helping in assessing the specific muscle movements associated with facial expressions at a faster and reliable way (Frank, 2001).

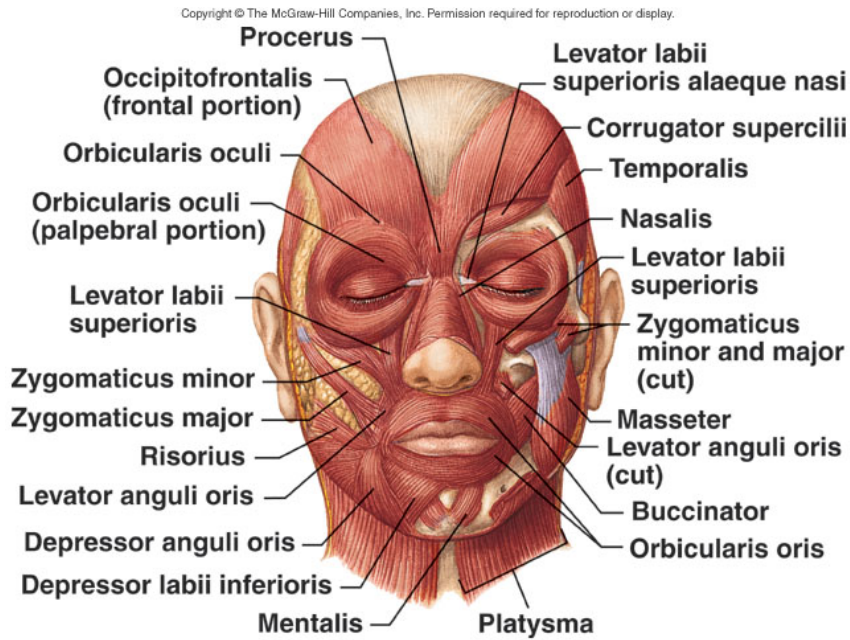


Figure 2.2: description of facial muscles.

Table 2.1: Descriptions of Action Units, FACS description and their associated facial muscle (Ekman & Friesen, 1976).

AU NUMBER	FACS DESCRIPTION	ASSOCIATED FACIAL MUSCLE
1	Inner brow raiser	Frontalis, Pars Medialis
2	Outer brow raiser	Frontalis, Pars Lateralis
4	Brow Lowering	Depressor Glabellae; Depressor Supercillii; Corrugator
5	Upper Lid Raiser	Levator Palpebras Superioris
6	Cheek Raiser	Orbicularis Oculi, Pars Orbitalis

7	Lid Tightener	Orbicularis Oculi, Orbicularis Palebralis
9	Nose Wrinkle	Levator Labii Superioris, Alaeque Nasi
10	Upper Lid Raiser	Levator Labii Superioris, Caput Infraorbitalis
11	Nasolabial Fold Deepener	Zygomatic Minor
12	Lip Corner Puller	Zygomatic Major
13	Cheek Puffer	Caninus
14	Dimpler	Buccinator
15	Lip Corner Depressor	Triangularis
16	Lower Lip Depressor	Depressor Labii
17	Chin Raiser	Mentalis
18	Lip Puckerer	Incisivii Labii Superioris. Incisive Labii Inferioris
20	Lip Stretcher	Risorius
22	Lip Funneler	Orbicularis Oris
23	Lip Tightener	Orbicularis Oris
24	Lip Pressor	Orbicularis Oris

25	Lips Part	Depressor Labii or Relaxation of Mentalis
26	Jaw Drop	Maseter, Temporary and Internal Pterygoid Relaxed
27	Mouth Stretch	Pterygoids; Digastric
28	Lip Suck	Orbicularis Oris

## 2.6 Typical Facial Expression Recognition (FER) System

The general framework of facial expression classification or recognition involves the following stages (refer figure 4): image acquisition, pre-processing, feature extraction, feature selection and classification (Valero, 2016). The stages will briefly be described below.

The first and foremost step in FER is the acquisition of either static or dynamic images. These images are either in the form of two-dimensional or three-dimensional spontaneous or posed images, captured either under controlled settings or “in the wild” conditions. The most utilised databases are the posed two-dimensional (2D) databases which includes Japanese Female Facial Expressions, Extended Cohn Kanade (CK+), Cohn Kanade (CK) (I . M Revina & Emmanuel, 2018) just to mention a few although these databases are still challenged with constraints of head pose and rotation variations (Pantic & Rothkrantz, 2000).

Pre-processing is performed to address the challenges associated with the 2D images as well as other issues such as image segmentation, deformation, and illumination variation as well as the removal of background noise from the images (Yunxin Huang, Chen, Lv, & Wang, 2019).

The next crucial stage in pre-processing is face detection. Face detection is the first step in any face image analysis or facial expression classification. It is useful in determining the existence of face in an image as well as aligning the face for the efficient extracting of the relevant features (Bhardwaj & Dixit, 2016; Martinez & Valster, 2016). Face detection can be grouped into four categories namely: knowledge-based, appearance-based, template-matching and feature invariant methods (M. H. Yang, Kriegman, & Ahuja, 2002). The knowledge-based method detects faces based on rules from human knowledge of facial images. It is mostly used for face localisation. Feature-invariant methods detect images based on learning from the structural features on the faces which does not change irrespective of illumination or pose variation. Furthermore, for template-matching face detection, manually pre-determined by experts patterns or models or templates of either whole or part of the facial image are stored and a correlation is found between the input images and the stored facial patterns whilst for appearance-based method detects face from the comparing facial image to a set of training image data. The knowledge-based method and feature-invariant method are mostly used for face localisation whilst the appearance-based method is mainly employed for both face detection (M. H. Yang et al., 2002).

After face detection, feature extraction is performed. Feature extraction is considered the most important step in facial expression classification. It helps in representing the facial image effectively by extracting the subtle changes of a facial image into a feature vector (Abouyahya, El Fkihi, Thami, & Aboutajdine, 2016; Bhardwaj & Dixit, 2016). Generally, feature extraction is categorised into geometric features, appearance features and hybrid feature method. Geometric features extract features using facial shape location. Appearance features extract features based on the pixel intensity information or texture (Yu & Liu, 2015). Hybrid feature method fuses geometric and appearance methods. Further, the hybrid method can be classified into decision-level and feature-level. The decision-level hybrid method utilises voting classifier

to ensemble the decision of all the feature sets whilst the feature-level method concatenate all the feature sets into one feature vector (X. Huang, 2014). Each type of feature extraction method has its associated advantages and limitations. Geometric methods have low computational cost; however, they are sensitive to noise and image spatial transform whilst the appearance feature methods are stable and accurate to image spatial transform, but they are computation expensive in comparison to geometric feature methods.

With the introduction of deep learning algorithms, these algorithms serve as their feature extractors or descriptors. Notable feature extraction methods include Histogram of Oriented Gradients (HOG), Gabor filter, Local Directional Pattern (LDP), Scale Invariant Feature Transform (SIFT), Linear Discriminant Analysis (LDA), Discrete Cosine Transform (DCT), Linear Binary Pattern (LBP) and Active Appearance Models (AAM).

Due to the high dimensionality of extracted features, feature selection is performed to discard irrelevant features, retaining the important feature vectors for accurate and acceptable classification. Feature selection is useful in reducing the computational cost and memory usage, improving data quality hence the predictive accuracy and increasing the speed of the algorithm (Khalid, Khalil, & Nasreen, 2014; Ladha & Deepa, 2011). Feature selection techniques worthy of mentioning are Adaboost, Linear Discriminant Analysis, Independent Component Analysis, Whitened Principle Component Analysis Laplacian Eigenmaps, Local Linear Embedding and Principal Component Analysis (PCA).

Classification is the final stage when conducting a facial expression classification experiment. Classification is in two folds: directly classifying into the various affective states or classifying into affective states after the detection of a particular action unit (Rizwan, 2013). The employed classifier categorizes the facial expressions into emotions: sadness, anger, joy, fear, happiness, smile, disgust and so on (I. M. Revina & Emmanuel, 2018). The algorithms utilised for

recognition of facial expressions are grouped into machine learning and deep learning algorithms. K-Nearest Neighbours, Naïve Bayes, Random forest, Hidden Markov Model, Extreme Learning Machine (ELM), Self-organising Maps (SOM), Sparse Representation based Classifier (SRC), Recurrent Neural Network (RNN), Deep Neural Network (DNN), Long Short-term Memory (LSTM) and Convolution Neural Network (CNN) are some variants of classification algorithms.

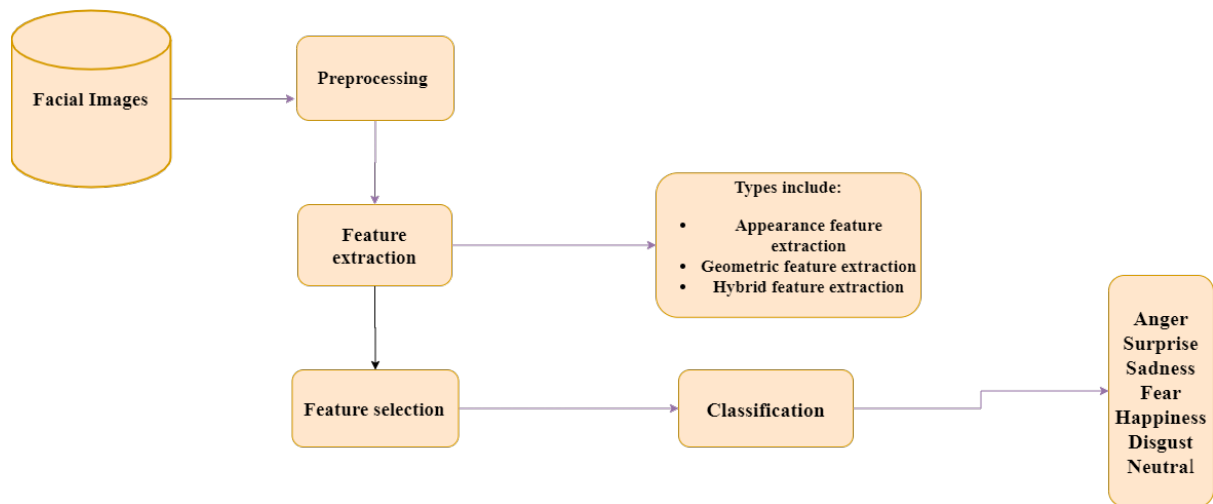


Figure 2.3: processes of facial expression recognition (FER).

Automatic facial expression recognition has become a hot research topic over the past two decades, with the utilisation of a plethora of algorithms. Thus, we seek to examine specifically the various algorithms utilized at the different phases of facial expression recognition experiments.

### 2.6.1 Machine and deep learning FER

In their work Zhong, Chen, & Liu (2014) proposed a novel method, Extended Nearest Neighbour for the classification of the facial expressions of the Japanese Female Facial expression databases. Gabor filter was deployed for the feature extraction process and Principal Component Analysis (PCA) for feature dimensionality reduction. The proposed method resulted in an accuracy of 93.01%, which in comparison to Gabor plus Support Vector Machine

(SVM) and Gabor with Neural Network attained accuracies of 91% and 90.01% respectively. Following, SVM and K-Nearest Neighbour (KNN) were utilized by the authors as classifiers to investigate on the performance of the feature extractors: Principal Component Analysis (PCA) and local binary pattern (LBP) on the JAFFE and Mevlana University Facial Expression (MUFE) database. SVM and JAFFE outperformed KNN and MUFE with regards to classifier and database respectively (Abdulrahman & Eleyan, 2015). Similarly, SVM outperformed KNN in a comparative analysis of their performance experimented on Extended Cohn Kanade (CK+) and Binghamton University 3D Facial Expression Database (BU-3DFE) databases. It was observed there was difficulty in classifying, as anger, fear and disgust produced similar results (Saeed, Al-Hamadi, Niese, & Elzobi, 2014). Likewise, Michel & El Kaliouby (2015) presented SVM for the classification of images of Cohn Kanade (CK) database. The presented algorithm attained an accuracy of 87.9%. Vo & Le (2016) proposed a fusion of CNN and SVM on Cohn Kanade (CK) database, achieving an accuracy of 96.04%. CNN served as the feature extraction method and SVM as the classification algorithm. Additionally, Mayya, Pai, & Manohara Pai (2016) performed facial expression recognition utilising Deep Convolution Neural Network (DCNN) and SVM for feature extraction and classification respectively on JAFFE and CK+ datasets. The proposed method attained an accuracy of 98.12%. In the work by Y. D. Zhang et al. (2016), biorthogonal wavelet entropy and fuzzy multiscale SVM were employed for facial expression classification to extract multiscale features as well as solve issues of noise and outliers. The proposed method achieved an accuracy of 96.77%. Furthermore, Kiran & Kushal (2016) presented Support Vector Machine for a multiclass facial expression classification on Japanese Female Facial Expression Database (JAFFE), Indian Facial Expression Image Database (IFED) and Taiwanese facial Expression Database (TFEID). The features were extracted using Bidirectional Local Binary Pattern resulting in recognition accuracies of 94.77%, 94.77% and 90.41% respectively. Also, the study indicated, anger, disgust and fear

had the same accuracy value of 88.88%. Additionally, Z. Wang, Jiang, Jiang, & Zhou (2016) employed SVM with a radial basis function kernel to distinguish the facial expressions of JAFFE database into 7 different facial expressions after learning the sparse representations of the facial images with K-SVD. More, a multi-class SVM with radial basis kernel was deployed in classifying facial expressions of CK+ after the pre-processing and extraction of the features with Viola Jones and Edge-Histogram Oriented Gradient (E-HOG) separately attaining accuracy of 96.4% (Candra, Yuwono, Chai, Nguyen, & Su, 2016). Furthermore, from a comparative study of their performance on the classification of facial expressions into the basic emotions, weighted feature gaussian kernel function SVM (WF-SVM) outperformed SVM with a gaussian kernel function with an average precision value of 93% to 83% (Wei & Jia, 2016). Besides, Borui, Liu, & Xie (2017) grouped the images of JAFFE database into the basic emotions using a multi-class SVM and Local Binary Pattern (LBP), Local Phase Quantization (LPQ) based on Gabor wavelet and Principal Component Analysis plus Linear Discriminant Analysis for extraction and selection of features respectively. The proposed method achieved an accuracy of 98.57 having anger and fear with the same accuracy value of 100. More, a comparative study of the performance of facial expression classification algorithms namely: Support Vector Machine (SVM), K-Nearest Neighbour (KNN) and Random Forests were conducted using extended Cohn-Kanade dataset (CK+). SVM surpassed KNN and Random Forest with accuracies of 80%, 75.15% and 76.97% respectively for considerably small amount of dataset. However, for large dataset, KNN and Random Forest outperformed SVM with accuracies of 98.85%, 98.85% and 90% respectively. Further, the results indicated some misclassification of anger and disgust facial expressions (Nugrahaeni & Mutijarsa, 2017). Further, Rashid et al. (2017) employed KNN and SVM with radial basis function to classify facial expressions of the JAFFE dataset, having Viola Jones algorithm and cross-correlation as the pre-processing and feature extraction method accordingly. The results indicated KNN

was an optimal algorithm for the experiment with an overall accuracy of 92.48%. Again, anger, disgust and fear had matrix values of 93.33, 93.10 and 93.75. Verma & Khunteta (2017) on the other hand, employed Gabor filter and ANN to categorize the facial expression of JAFFE database, achieving an accuracy of 85.7%. Likewise, Qayyum, Majid, Anwar, & Khan (2017) deployed Artificial Neural Network (ANN) to perform a facial expression classification of the databases: JAFFE, CK+ and MS-Kinect databases. Stationary wavelet transforms and Discrete Cosine Transform (DCT) were used for feature extraction and feature selection respectively. JAFFE surpassed the other databases with an accuracy of 98.83% from the outcome of the experiment. Breuer & Kimmel (2017) conducted facial expression classification and performance analysis on Extended Cohn Kanade, FER2013 and NovaEmotions databases using convolutional neural network (CNN). CNN on the Extended Cohn Kanade (CK+) database outperformed the other classifiers such as Gabor with SVM, LBPSVM with an accuracy of 98.62%, 89.8% and 95.1% respectively. Likewise, Lopes, de Aguiar, De Souza, & Oliveira-Santos (2017) compared the performance of six facial expressions to seven facial expressions using Extended Cohn Kanade (CK+), JAFFE and Binghamton University 3D Facial Expression (BU-3DFE) databases using CNN. The result indicated CNN performs best on CK+ when classifying six facial expressions with an accuracy of 98.92%. Also, Alizadeh & Fazel (2017) developed a convolutional neural network (CNN) for a facial expression recognition task and classified the facial expressions into anger, happiness, fear, neutral, surprise, sad and disgust using facial FER-2013 dataset. Likewise, Revina & Emmanuel (2018) classified facial expressions of the JAFFE database utilizing Particle Swarm Optimization based K-Nearest Neighbour (PSO-KNN). The features were extracted using Local Descriptor with Modified Gray value Accumulation Value (LD-MGAD). The proposed model attained an accuracy of 97.1%, with anger and fear obtaining similar values as anger and sad facial expressions misclassified as fear. M. I. Revina & Emmanuel (2018) presented SVM for the

classification of facial expressions of JAFFE and CK+ databases, achieving an accuracy of 88.63%. Local Directional Number (LDN) Pattern and Directional Gradient Local Ternary Pattern (DGLTP) were adopted for the feature extraction process. Further, Zarbakhsh & Demirel (2018) investigated the use of 3D images for facial expression detection to find an optimum low-dimensional feature sub-space for 3D facial expressional detection on Binghamton University known as BU- 3DFE dataset. Support Vector Machine (SVM) and Fuzzy SVM (FSVM) were utilized for the classification along with sequential feed feature selection (SFFS) and conventional t-test for the feature selection process. The results indicated an average accuracy of 87.67% for SFFS and FSVM and a matrix value of 85 for both anger and disgust. Likewise, SVM was adopted in classifying the facial expressions along with performing a comparative assessment of MMI, extended Cohn-Kanade (CK+) and static face in the wild (SFEW) databases. Weber local descriptor, a dual-fusion feature extraction method as well as discrete cosine transform (DCT) were utilised for both feature extraction and selection accordingly. The results confirmed CK+ as the excelling database from a comparative assessment of the databases. Also, it was difficult in classifying anger due to its misclassification as disgust or neutral. A deep convolutional neural network inspired by XCEPTION was proposed by Raksarikorn & Kangkachit (2018) to classify seven facial expressions using FER-2013 dataset. The suggested model outperformed the XCEPTION attaining an accuracy of 71.69%, 72.91% and 70% accordingly. In addition, Kumar, Kumar, & Sanyal (2018) deployed convolutional neural network (CNN) for training and classification of facial expressions of FER-2013 and Extended Cohn Kanade databases (CK+) into seven emotions namely anger, neutral, sad, happy, disgust, surprised and fear, achieving an accuracy of 90+%. Similarly, Mohammadpour, Khaliliardali, Hashemi, & Alyannezhadi (2018) adopted CNN to group facial expressions of CK+, JAFFE and BU-3DFE databases. The proposed method achieved an accuracy of 97.01% having CK+ as the excelling database. Li (2018) used

convolutional neural network (CNN) to classify facial expressions of JAFFE, CK+ and FER-2013 databases. JAFFE database achieved the topmost accuracy of 97.65. The results also showed anger and fear having closely related values. Also, CNN was utilized in classifying the images of the databases: JAFFE and CK+. The algorithm performed better on JAFFE than on CK+ database; having anger, fear and disgust have similar values (Farajzadeh & Hashemzadeh, 2018). Similarly, conditional neural network enhanced random forest (CoNERF) was utilized in classifying facial expressions of CK+, JAFFE, BU-3DFE and Labelled faces in the wild (LFW) databases. The proposed algorithm on JAFFE and CK+ obtained an accuracy of 99.02%, having anger and disgust with closely related values (Y. Liu et al., 2018). In addition, artificial neural network specifically multilayer perceptron with back propagation was utilized in categorization facial expressions of JAFFE, CK+ and Radboud Faces Database (RaFD) databases. Accuracies of 94.81%, 99.51% and 99.15% were obtained for the databases respectively (Islam, Mahmud, Hossain, Mia, & Goala, 2019). More, SVM excelled than KNN and multilayer perceptron (MLP) in a classification of facial expressions: normal, happy, angry, contempt, surprise, sad, fear and disgust of the Cohn Kanade database with accuracies of 93.53%, 82.97% and 79.79% separately. Histogram of oriented gradients was utilized for the feature extraction and PCA for the feature selection (Dino & Abdulrazzaq, 2019). Fan & Tjahjadi (2019) used combination of handcrafted and convolutional features with the classification algorithm SVM for facial expression recognition, achieving an accuracy of 92.15%. Also, Bellamkonda & Gopalan (2019) detected and classified facial expressions of the following databases: JAFFE, Cohn Kanade, MMI and Karolinska Directed Emotional Faces (KDEF) employing SVM with either local binary classifier or Gabor wavelet as the feature extraction algorithm. From the outcome, SVM plus Gabor wavelet gave a surpassing accuracy of 98.83% on the KDEF database.

Furthermore, Dubey & Dixit (2019) classified the facial expressions of JAFFE, CK+ and FER-2013 using CNN. The proposed algorithm achieved a classification accuracy of 97.90% on the images of CK+ and JAFFE into the various expressions. A deep convolutional neural network (DNN) was proposed to group the images of CK+ and JAFFE into the various expressions. Precision, recall, ROC and accuracy were utilized in evaluating the experiment; the algorithm performed excellently on the JAFFE database with an accuracy of 95.23% to 93.24% on CK+ dataset (D. K. Jain, Shamsolmoali, & Sehdev, 2019). In a quest to improve the performance of end-to-end frameworks for facial expression recognition for deep learning methods, Minaee & Abdolrashidi (2019) proposed attentional convolutional neural network using databases: JAFFE, FER-2013, CK+ and Facial Expression Research Group Database (FERG) and CK+. The proposed algorithm performs best on CK+ with an accuracy of 98%. Notwithstanding the advantages of the convolutional neural network, Sharma & Jain (2019) identified CNN to have a drawback of handling spatial information. Hence, the researchers proposed a bidirectional Long Short-Term Memory (LSTM) to categorize the facial expressions of the Cohn Kanade database as LSTM has an advantage of memory efficiency.

#### 2.6.2 Ensemble learning algorithms

T H H Zavaschi & Koerich (2011) utilised fused feature of Gabor and LBP and an ensemble of base classifiers SVM and used a multiobjective genetic algorithm (MOGA) as the pareto-optimal for selecting the best of classifiers. The proposed method achieved an accuracy of 96.2%. Also, Pons & Masip (2018) proposed an ensemble of CNN committee classifiers for facial expression recognition achieving an accuracy of 39.3%. More, D. H. Nguyen et al. (2019) performed FER using an ensemble of multilayer CNN obtaining an accuracy of 74.09%. Likewise, W. Sun, Zhao, & Jin (2019) proposed an ensemble of CNN for FER attained an accuracy of 96.15%. Xu, Pang, & Jiang (2019) conducted FER using a fusion of geometric features (HOG and DHOG) and achieved an accuracy of 96.4%.

Table 2.2: summarises algorithms utilised by researchers for facial expression recognition.

Work	Feature extraction	Feature selection	Algorithm	Database	Performance metric used
(Zhong et al., 2014)	Gabor	PCA	Extended Nearest Neighbour	JAFFE	accuracy
(Abdulrahman & Eleyan, 2015)	PCA/LBP		SVM /KNN	JAFFE/MUFE	Confusion matrix/accuracy (77% and 87%)
(Saeed et al., 2014)	Geometry-based		SVM/KNN	CK+	Confusion matrix
(Nugrahaeni & Mutijarsa, 2017)			SVM/KNN/ Random Forest	CK+	Accuracy
(Rashid et al., 2017)	cross-correlation/ MBWM		SVM/KNN	JAFFE	Confusion matrix

(Wang et al., 2016)	K-SVD		SVM with radial basis function kernel	JAFFE	Accuracy (97.138%)
(Candra et al., 2016)	E-HOG		SVM with radial basis function kernel	CK+	Accuracy (96.4%) /confusion matrix
(Wei & Jia, 2016)			Weighted feature gaussian kernel SVM	CK+	Precision (93%)
(Kiran & Kushal, 2016)	Bidirectional LBP		Multi-class SVM	JAFFE/TFEID/IFED	Accuracy (97.10%)
(Mahmood, Hussain, Iqbal, & Elkilani, 2019)	Weber local descriptor	DCT	SVM	MMI/CK+/SFEW	Confusion matrix/accuracy (98.62%)

(Qayyum et al., 2017)	Stationary wavelet transform	DCT	Artificial neural network	JAFFE/CK+/Ms-Kinect	Confusion matrix
(Dino & Abdulrazzaq, 2019)	Histogram oriented gradient	PCA	SVM/KNN/MLP	CK+	accuracy (93.53%)
(Islam et al., 2019)	Gabor	PCA	ANN (Multilayer Perceptron)	JAFFE/CK+/RaFD	Confusion matrix/accuracy (99.51%)
(Raksarikorn & Kangkachit, 2018)			CNN	FER-2013	Accuracy/precision/recall
(Kumar et al., 2018)	CNN		CNN	FER-2013/CK+	accuracy
(Minaee & Abdolrashidi, 2019)	Attentional CNN		Attentional CNN	JAFFE/CK+/FERG/FER-2013	accuracy
(Vo & Le, 2016)	CNN		SVM	CK	Accuracy (96.04%)

(Sharma & Jain, 2019)	LSTM		Bidirectional LSTM	CK	
(Breuer & Kimmel, 2017a)	CNN		CNN	CK+/FER-2013/NovaEmotions.	accuracy
(Lopes et al., 2017)	CNN		CNN	CK+/JAFFE/B U-3DFE	accuracy
(Dubey & Dixit, 2019)	CNN		CNN	JAFFE/CK+/FER-2013	accuracy
(Verma & Khunteta, 2017)	Gabor filter		ANN	JAFFE	accuracy
(Bellamkonda & Gopalan, 2019)	Gabor/LBP		SVM	JAFFE/MMI/K DEF	Accuracy (98.83)
(Alizadeh & Fazel, 2017)			CNN	FER-2013/Kaggle	

(Zarbakhs h & Demirel, 2018)		SFFS/conv ventional ttest	Fuzzy SVM	BU-3DFE	Confusion matrix  Accuracy (87.67%)
(Revina & Emmanuel , 2018)	LD-MGAD		PSO-KNN	JAFFE	Accuracy/confu sion matrix
(Mohamm adpour et al., 2018)	CNN		CNN	JAFFE/CK+/B U-3DFE	Accuracy/confu sion matrix
(Borui et al., 2017)	LBP+LQP+ Gabor		PCA-LDA	JAFFE	accuracy
(Z. Li, 2018)	CNN		CNN	JAFFE/CK+/F ER-2013	Accuracy/confu sion matrix
(Y. Liu et al., 2018)	CNN		CoNERF	JAFFE/CK+/L FW	Confusion matrix/accurac y
(Xu et al., 2019)	HOG +DHOG	PCA	SVM	JAFFE	Accuracy (96.4%)
(W. Sun et al., 2019)			Ensemble of CNN	CK+	Accuracy (96.15%)

(T H H Zavaschi & Koerich, 2011)	Gabor+LBP		Ensemble of SVM	CK+/JAFFE	Accuracy (96.2%)
(Farajzadeh & Hashemzadeh, 2018)	CNN		CNN	CK+/JAFFE	Confusion matrix/ ROC/ Precision
(M. I. Revina & Emmanuel, 2018)	LDN/DGL TP		SVM	JAFFE/CK+	Accuracy (88%)
(Michel & El Kaliouby, 2015)			SVM	CK	Accuracy
(D. K. Jain et al., 2019)	CNN		DNN	CK+/JAFFE	Accuracy/Precision/ Recall/ROC
(Abouyahya &			KNN+DTW	CK+	Precision/Recall

Fkihi, 2018)					
(Y. D. Zhang et al., 2016)	Biorthogonal Wavelet Entropy		Fuzzy Multiclass SVM	Captured images	Accuracy (96.77%)/ Confusion matrix
(Mayya et al., 2016)	DCNN		SVM	JAFFE/ CK+	Accuracy/ Confusion matrix (98.12%)
(Fan & Tjahjadi, 2019)	CNN features plus shape and appearance features		SVM	CK+	Accuracy (92.5%)

## 2.7 Databases for facial expression recognition

There exist standardised databases for facial expression recognition. These databases are employed for evaluating the performance of facial expression algorithms to make a meaningful comparison. These databases differ in terms of the following (Rizwan, 2013):

1. The uniqueness of subjects: characteristics such as face, shape, colour, number of subjects for capturing the images, age, ethnicity, skin colour distinguishes one database

from another. For instance, for some databases, there is an equal number of male and female subjects whilst others used only female subjects.

2. Posed versus spontaneous expressions: the databases are composed by asking the subjects to perform a series of expressions or are captured naturally. Most of the 2D databases are posed databases such as JAFFE with a few exceptions such as MMI which contains spontaneous smile facial expressions in addition to the posed facial expressions. Currently, FER is moving towards the use of spontaneous databases for facial expression analysis.
3. Face or head orientation: the databases capture their images from different angles, and this influence the performance of the facial expression algorithms. For example, the KDEF database has its expressions captured in five angles: full right profile, the full left profile, half right profile, half left profile and straight (Goeleven, De Raedt, Leyman, & Verschuere, 2008).

## 2.8 Limitation of current work and contribution

We have found shortcomings in the FER process. The current systems can detect facial expression in general or a subset of emotions. For instance, an algorithm categorizes the basic emotions namely disgust, anger, sadness, fear, surprise, and happiness plus neutral or a subset of it such as anger, disgust, and fear. To the best of our knowledge, there have not been any studies on how to detect anger only using facial expressions. More, with the multiclass classification of emotions, there are some drawbacks such as an overlap among the facial expressions namely disgust, anger and fear due to the slight distinction among them, giving an untrue representation of the emotions (Pell & Richards, 2011). Further, in confirmation to Pell's claim, concerns have been raised about the misclassification between angry and disgust facial expressions as observed from the results of experiments (Apte et al., 2016; Kwong,

Garcia, Abu, & Reyes, 2019; Nugrahaeni & Mutijarsa, 2017; Y. D. Zhang et al., 2016 ) (Kiran & Kushal, 2016; Talele, Shirsat, Uplenchwar, & Tuckley, 2017).

It is postulated a majority of the facial expression algorithms have difficulty in performing a multi-class classification, which is attested to with regards to the training time, computational time and the insufficient memory space (Kiran & Kushal, 2016; Shah et al., 2017). Also, despite the rapidly growing literature on anger detection, what is known is largely based on the utilization of either physiological signals or audio or speech data (Chang et al., 2012; Chhabra et al., 2017; Deng et al., 2018). Not much is known about the detecting of only anger using facial expressions. We argue that anger detection needs to be done accurately, giving a true representation of the emotion. As the detection of anger will provide useful information about peoples' intensity of anger to manage or control it, as unregulated anger sometimes results in aggression or violence (Moritz, 2006; Shahsavarani et al., 2015). Therefore, our contribution in this respect in these ways:

1. We have illustrated that only anger can be detected with facial expressions and facial expression algorithms (for reference read chapter 5).
2. We propose a novel ensemble learning algorithm for our anger recognition. For reference see chapter 3.
3. We have shown that our results can achieve high accuracies which exceed the state-of-the-art results. For reference read chapters 4 and 5.

## 2.9 Chapter summary

In this chapter, we have gained a comprehensive insight into our research, identified the limitations, and this helped us to effectively organise our work to achieve our objectives. The next chapter will detail how we plan to undertake our research work specifically explaining the methods we intend to deploy.



## Chapter 3

### Methodology

#### 3.0 Introduction

This chapter presents all the details related to the methods utilised during the different phases of facial expression recognition as well as the databases used during this research work. Also, it justifies the selected algorithms and databases.

#### 3.1 Workflow

This section details the methods deployed, which is depicted in figure 3.1.1 as well as the stages by which the research work was realised. The overall work was undertaken in the following phases: image acquisition, pre-processing, feature extraction, feature selection and classification.

##### 3.1.1 Image acquisition

In conducting research involving machine learning and deep learning algorithms, there is a dire need for datasets since these algorithms are driven by data. As such, careful diligence needs to be observed in the selection of the datasets for facial expression recognition, as an inappropriate selection such as datasets with noisy backgrounds might result in increasing the difficulty of the project as well as affect the overall recognition accuracy.

There are several facial expression databases; some can be easily downloaded whilst others require permission to be obtained before they can be downloaded. The permission is obtained by the mere filing of a form for formality sake and to prove the usage of the dataset strictly for academic purposes. The databases differ in terms of the following characteristics: dimensions namely two-dimensional and three-dimensional, image quality, posed and spontaneous, static and image sequences databases. The 2-dimensional databases are the most utilized due to their

availability publicly although there are challenged by with constraints of head pose and rotation variations (Pantic & Rothkrantz, 2000). On the other the 3-dimensional database is robust to these constraints due to the use of the 3D face scanner, are computationally expensive, requiring a lot of resources. Hence, these reasons account for the wide adoption of 2D facial expressions.

Therefore, for research work, after considerable weighing and evaluation of the pros and cons of both the two-dimensional and three-dimensional databases, the Japanese Female Facial Expression (JAFFE), Karolinska Directed Emotional Faces (KDEF) and Extended Cohn Kanade database (CK+) were selected. These 2D databases offer frontal, posed, noise-free background with labelled and validated images. A description of these databases are as follows. Nevertheless, much as we are aware these databases are not of African people, which is what we desire to make our work relevant to our people, they do serve a useful purpose to enable us quickly test our method against established methods. Therefore, our future work will focus on creating a database of African people.

#### 3.1.1.1 JAFFE database

The Japanese Female Facial Expression database contains 213 photographed images from ten Japanese females who displayed the basic emotion: anger, disgust, fear, surprise, sadness, happiness and neutral (Lyons et al., 1998). Each of the ten Japanese females posed 3 or 4 times per expression. The images are in grayscale and tiff format with a resolution of 256\*256 pixels. The JAFFE database is devoid of occlusion and illumination variation as the ten females were asked to tie their hair and show the real face. Also, adequate lightning was provided during capturing of the facial expression as the images were captured in a controlled environment.



Figure 3.1 shows sample images of the JAFFE database.

### 3.1.1.2 CK+ database

The Extended Cohn Kanade is an extension of the Cohn Kanade database (Kanade, Cohn, & Tian, 2000); an addition of spontaneous smile facial expressions whilst the remaining are posed facial expressions. It contains 529 image sequences from 123 subjects, within the ages of 18 to 50 years. Females form 69% and males 31%, having 81%, 13% and 6% as Euro-American, Afro-American and others accordingly. The image sequences differ in duration that is 6 to 10 seconds per frame and was videoed from the onset which is a neutral emotion to the formation of the peak emotion. The images sequences are well labelled and validated in png format, with a resolution of 640\*490 pixels (Lucey et al., 2010).



Figure 3.2: displays a sample CK+ images. Left-to-right from top row: anger, disgust, happiness, surprise, angry, contempt, sadness.

### 3.1.1.3 KDEF database

Karolinska Directed Emotional Faces (KDEF) database (Goeleven et al., 2008) is a posed facial expression database made up of 4900 images from 70 different subjects, 35 males and 35 females within the ages of 20 to 30 years. The 70 subjects displayed the 7 different emotional expressions, each expression photographed twice and shot at 5 different angles. The subjects are without visible make-up, spectacles, ornaments, beards and moustaches (Goeleven et al., 2008).



Figure 3.3: shows KDEF images captured from different angles

### 3.1.2 Pre-processing

Facial pre-processing is the first step in facial expression recognition. Pre-processing of facial expressions is performed to discard irrelevant information and improve the recognition accuracy of the important extracted features. It involves processes such as face detection and other image modification methods such as smoothening and normalisation.

### 3.1.2.1 Face detection

Face detection is the first step in any face image analysis or facial expression classification. It is useful in determining the existence of face in an image as well as aligning the face for the efficient extracting of the relevant features (Bhardwaj & Dixit, 2016; Martinez & Valster, 2016).

In this work, a face detection method proposed by Viola and Jones was employed. The Viola Jones face algorithm is a widely adopted face detection algorithm due to its robustness and computational simplicity; it works in the following phases: Haar-features, integral image, Adaboost and a cascade of classifiers (Viola & Jones, 2004).

The algorithm classifies images in the form of the value of simple features. The feature method is preferred over the pixel-feature since it operates in a much faster way. The features are classified into three groups namely: two-rectangle feature, three-rectangle feature, and four-rectangle feature. The two-rectangle feature is the value of the difference of the sum of pixels between the two rectangular regions. The three-rectangle feature quantifies the difference between the sum of pixels of the two exterior rectangles and the addition of pixels within the central rectangle and the four-feature rectangle computes the variation within the diagonal pairs of rectangles (Viola & Jones, 2004). The Haar-like features are displayed in figure 3.4.

However, the total number of features can be extremely huge, hence the proposal of an integral image for faster evaluation. The integral image is utilised as an image representation and it obtains the sum of values of a rectangular area. It is defined in equation 3.1:

$$k(x, y) = \sum_{x' \leq x, y' \leq y} b(x', y') \quad 3.1$$

where  $k(x, y)$  and  $b(x, y)$  is the integral image and actual image respectively (Viola & Jones, 2004).

The integral image sums the values of the rectangular feature from the origin to the point (x, y). Equation 3.2 is performed the calculation of the sum of values in a rectangular area bounded by  $(x_0, y_0)$  and  $(x_1, y_1)$  and  $Area_0 < Area_1$ . This is displayed in figure 3.5.

$$i(x, y) = i(x_1, y_1) - i(x_0, y_1) - i(x_1, y_0) + i(x_0, y_0) \quad 3.2$$

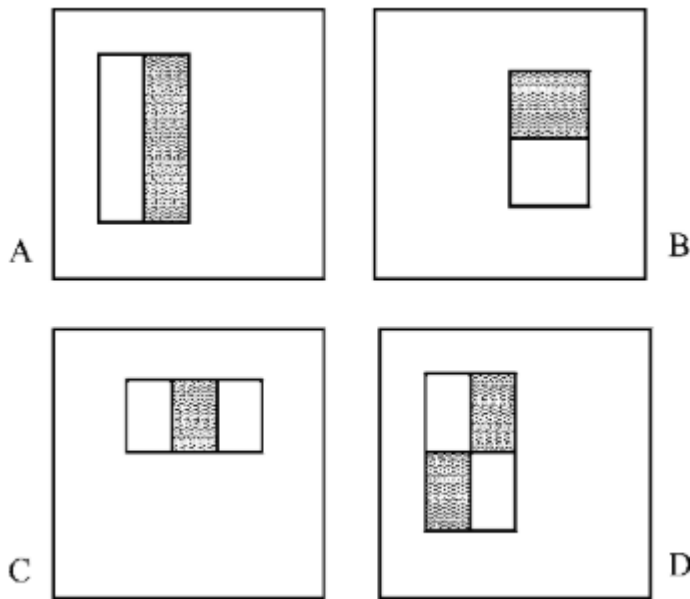


Figure 3.4 shows the rectangle features. A and B are the two-rectangular features and C and D are the three-rectangular features and four-rectangular features respectively (Viola & Jones, 2004).

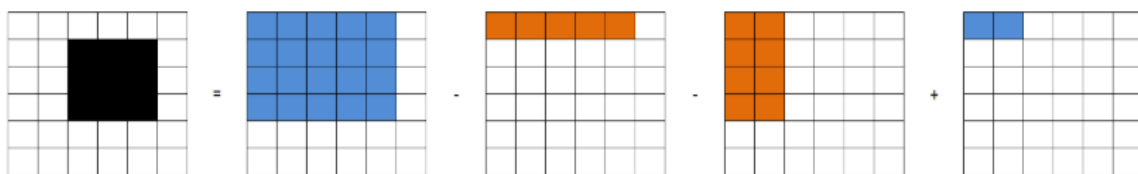


Figure 3.5 displays the calculation of area using integral image (Valero, 2016).

Although the calculation of the values in the rectangular area has been simplified, there is still a requirement for an amount of processing power to work on the whole image. Therefore,

Adaboost was introduced to select the relevant features from the face. Then it is passed to a complex cascade of classifiers for faster face detection.

In summary, the proposed method works in a form of haar-like features which will extract important characteristics from the face which we intend to use to train a machine learning algorithm which will detect faces in real-time.

### 3.1.2.2 Image enhancement

Image enhancement is an important factor in pre-processing of images as all images marginally contain some level of noise. Image enhancement is useful in extracting features accurately and efficiently as it helps in determining the accurate features to be extracted for classification which in turn improves the learning ability of classifiers (Tan & Jiang, 2019). The images were enhanced using the following methods: Contrast Limited Adaptive Histogram Equalisation and median blur. The main aim of image enhancement is to remove noise in images, thereby improving their quality.

#### 3.1.2.2.1 Median blur

Median blur is a nonlinear technique used to remove noise from images. It is characterised by its low computational speed and simplicity. Median blur is useful in preserving edges of an image. It operates as a filter sliding pixel by pixel over an image, exchanging the centre pixel with the median of gray levels in a window (Niu, Zhao, & Ni, 2017)

A two-convolutional median filter is symbolised as:

$$\hat{I}_{jk} = \text{median}\{I_{j+w,k+v}\} \quad 3.3$$

Given a H\*W image with  $1 \leq j \in \{1, 2, \dots, H\} \times \{1, 2, \dots, W\}$  and having  $w, v \in (-(w-1)/2, \dots, (w-1)/2)$

### 3.1.2.2.1 Histogram equalisation

Histogram equalisation is a pre-processing technique performed to enhance the contrast of an image after the removal of noise to extract features accurately and clearly. It works by transforming the intensity values of an image and can be described by the equation:

$$I_j = T(r_j) = \sum_{k=1}^j P_r(r_k) = \sum_{k=1}^j \frac{n_k}{n} \quad 3.4$$

Where  $I_j$  is the intensity value in the processed image which corresponds to  $r_j$  in the input image having input image's intensity values as  $P_r(r_k) = 1, 2, 3, \dots, L$ .

However, histogram equalisation has a drawback, it sometimes contrasts images and has the pixels fall within the same range. This intensifies the noise level of the image since it works on the whole face.

Therefore, in this work, we adopted contrast limited adaptive histogram equalisation (CLAHE), an extension of histogram equalisation for our image enhancement. CLAHE highlights and evaluates the facial regions. Then it divides the facial regions into tiles and applies the histogram equalisation on each tile (Zuiderveld, 1994).

### 3.1.3 Feature extraction

After the pre-processing stage, the next step is to extract the features. Feature extraction is mostly considered the most important step in facial expression classification, as the selection of the features is an important task. It helps in representing the facial image effectively by extracting the subtle changes of a facial image into a feature vector (Abouyahya et al., 2016; Bhardwaj & Dixit, 2016).

Therefore, in this work, two types of feature extraction methods are utilised: appearance and hybrid feature extraction method; LBP and fusion of HOG and LBP, respectively. These feature extraction methods resolve the two-dimensional database issue of illumination invariant as they are robust to it.

### 3.1.3.1 Local Binary Pattern

The Local Binary Pattern (LBP) is a commonly used appearance feature extraction method due to its numerous advantages. LBP has been widely adopted because it is easy to implement, invariant to rotations, robust to grayscale transformations caused by illuminations variations, tolerance to illumination, overcomes the problems of disequilibrium displacement, possess discriminative power, uses a modest amount of data and can save computational resource whilst retaining facial information (Ojala, Pietikäinen, & Mäenpää, 2002). LBP originally proposed by Ojala, Pietikäinen, & Harwood (1996), is a 2D texture algorithm distinct from the other traditional statistical and structural models for texture analysis. It comprises of two components: pattern and contrast, where contrast is the amount of texture. LBP operator thresholds between the centre pixel producing features with binary codes: 0 or 1 with a 3\*3 locality pixel (see figure 3.6).

Nonetheless, the 3\*3 neighbourhood performs poorly in encoding textures with large appearance changes and it is also unable to capture nonlocal macrotecture (L. Liu, Fieguth, Guo, Wang, & Pietikäinen, 2017). Therefore, the LBP operator has been extended with diverse neighbourhood sizes to handle the different scales. The extended LBP uses circular neighbourhoods and bilinearly interpolating the pixel values, which allows any number of points  $P$  and radius  $R$  in the neighbourhood (Ojala et al., 2002). Figure 3.7 displays the various neighbourhood sizes.

According to equation 3.5, a grayscale image  $I(x, y)$  is considered and let  $k_C$  symbolise the intensity of an arbitrary pixel  $(x, y)$ . Then,  $k_m$  represent the gray value of a sampling point in an evenly spaced circular neighbourhood of  $P$  sampling points and radius  $R$  around the point  $(x, y)$  (X. Huang, 2014). The LBP operator is defined as:

$$h_{P,R}(x_C, y_C) = \sum_{m=0}^{P-1} L((k_C - k_m) \geq 0) 2^m \quad 3.5$$

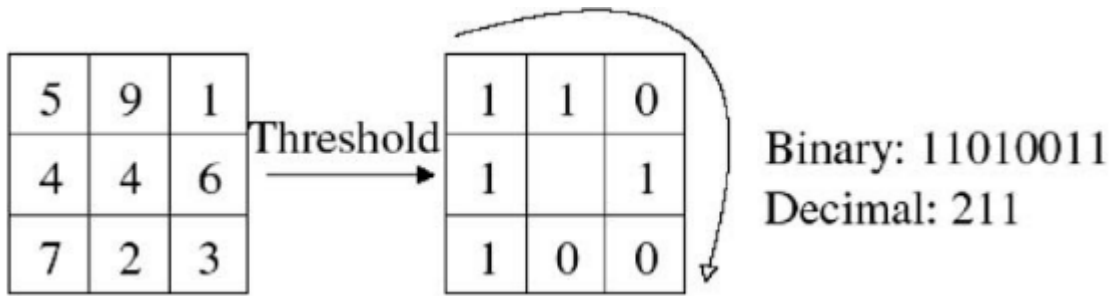


Figure 3.6: operation of the LBP operator.

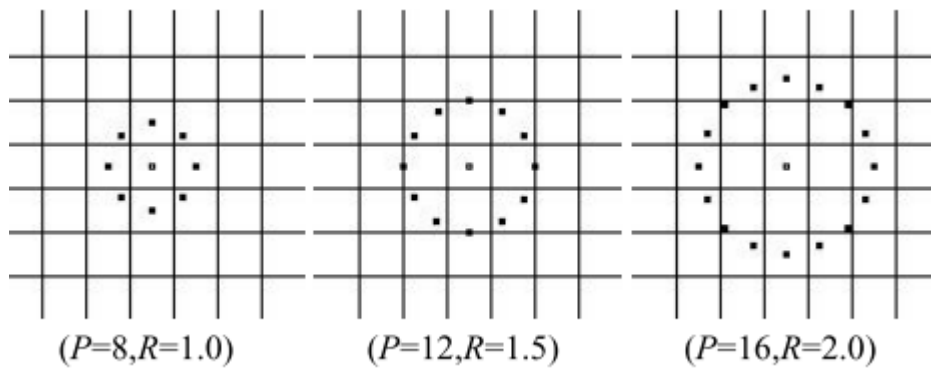


Figure 3.7: Three neighbour sets for different (P, R) used to construct a circularly symmetric LBP (Ojala et al., 2002).

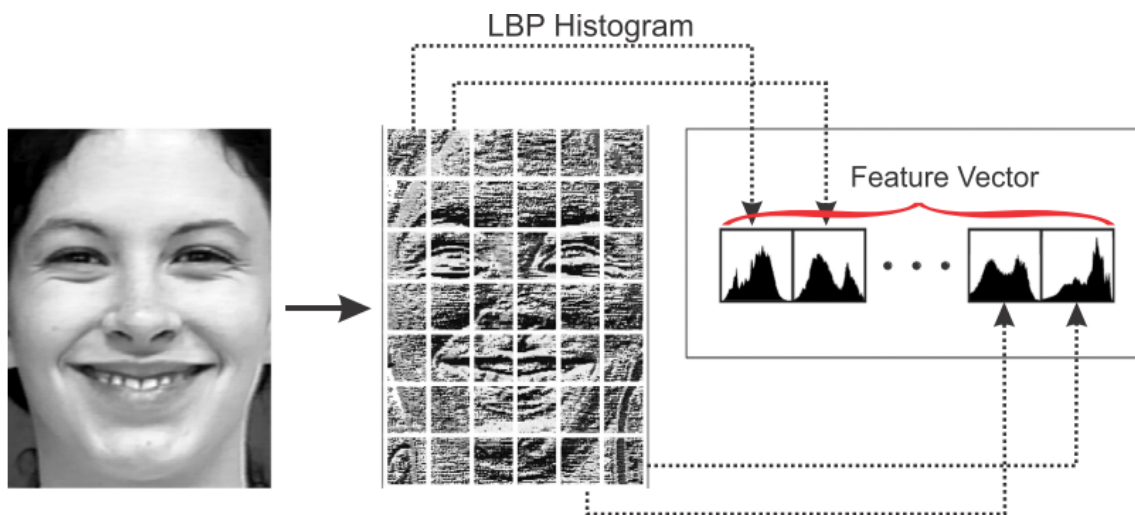


Figure 3.8: displays the operation of the LBP operator (Farajzadeh & Hashemzadeh, 2018).

### 3.1.3.2 Histogram of Oriented Gradients

HOG was first proposed by (Dalal & Triggs, 2005) for target detection. Due to its impressive performance, it has been widely adopted by researchers for facial expression recognition (Xu et al., 2019). HOG employs local gradient to define the shape of an image (Farajzadeh & Hashemzadeh, 2018).

The HOG operator works by dividing the images into small connected regions labelled cell and computes the gradient direction histogram for every single cell. Then, a HOG descriptor is formed from the combination of the gradient direction histogram (Dalal & Triggs, 2005).

HOG is characterised by two parameters: the size of cell and number of bins orientation. The size of the cell which represents the size per column and size per row is utilised in computing the histogram whilst the number of bins orientation is for the construction of the distance of the angles of the gradient (Nassih, Amine, Ngadi, & Hmina, 2019).

Mathematically, the HOG feature extraction is defined as follows:(Xu et al., 2019)

1. Using a one-dimensional differential template  $[-1,0,1]$ , the gradient value and gradient direction at a pixel  $((a, b))$  is calculated as follows:

$$R_a(a, b) = I(a + 1, b) - I(a - 1, b) \quad 3.6$$

$$R_b(a, b) = I(a, b + 1) - I(a, b - 1) \quad 3.7$$

Where the gradient of the vertical direction of the pixel and the gradient of the horizontal direction of the pixel is represented by  $R_a(a, b)$  and  $R_b(a, b)$  respectively.

2. The gradient amplitude and direction values are calculated as follows:

$$A(a, b) = \sqrt{R_a(a, b)^2 + R_b(a, b)^2}$$

$$\theta(a, b) = \arctan \theta \left( \frac{R_b(a, b)}{R_a(a, b)} \right) \quad 3.8$$

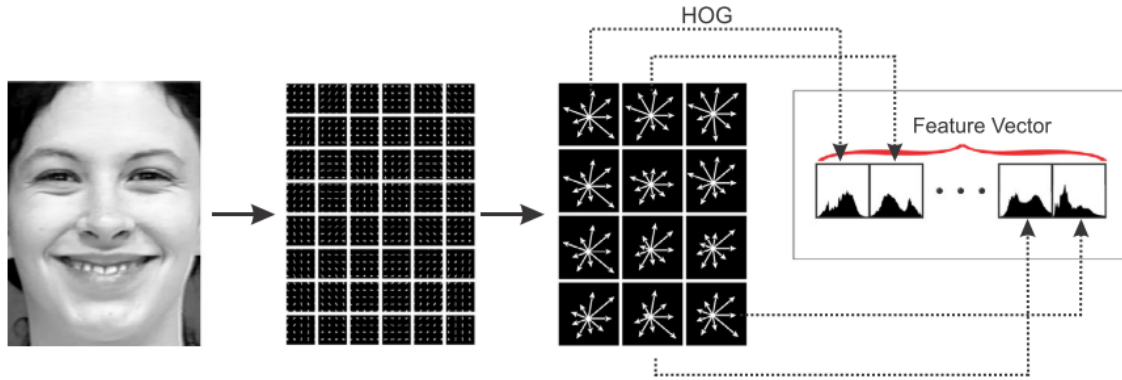


Figure 3.9 displays a HOG feature extraction process (Farajzadeh & Hashemzadeh, 2018).

#### 3.1.4 Feature selection

Due to the high dimensionality of extracted features, feature selection is performed to discard the irrelevant features retaining the important feature vectors for accurate and acceptable classification. Feature selection is useful in reducing the computational cost and memory usage, improve data quality hence the predictive accuracy and increase the speed of the algorithm (Khalid et al., 2014; Ladha & Deepa, 2011).

Notable techniques for feature selection are Adaboost, Linear Discriminant Analysis, Independent Component Analysis, Whitened Principle Component Analysis Laplacian Eigenmaps, Local Linear Embedding and Principal Component Analysis (PCA). Nonetheless, in comparison to the other methods, PCA seems to have received dominant research attention over the years.

PCA is a multivariate statistical technique with applicability to the fields of image compression and FER. It selects a small set of important inter-correlated features based on patterns from the high dimension data after analysis based on the similarities and differences of these feature vectors (Abdi & Williams, 2010; Turk & Pentland, 1991). So, in compressing the data as well

as providing a description for it, these selected features are represented as a set of new orthogonal variables called principal components, which are helpful in recognizing facial expressions effectively or excellently.

PCA produces excellent recognition rates due to its reduction in the sensitivity to noise as the redundant features are discarded. Further, it is comparatively invariant to changes in facial expression, has a low demand for memory and storage as well as low computational cost due to the reduced complexity of the features (Calder, Burton, Miller, Young, & Akamatsu, 2001; Karamizadeh, Abdullah, Manaf, Zamani, & Hooman, 2013).

In computing PCA, a series of steps are required to get the final set of dimensions. To start, a data set of  $q$  observations of  $m - dimensional vector, y = [y_1, y_2, y_3, \dots, y_n]^T$ . The vector is supposed to have a mean value of zero and this is attained by summing the values obtained from the subtraction of the mean  $\mu_x$  from each observation  $q$  of dataset  $y$ . Afterwards, the covariance matrix  $s_x$  is calculated to get the eigenvectors and the eigenvalues from the data. The dimensions containing the highest deviation within a dataset is the eigenvector. As such, some of these eigenvectors are selected as new features for our feature selection (Jan, 2017).

The functions are defined as follows:

$$U_x = \frac{1}{m} \sum_{i=1}^m y_i \quad 3.9$$

$$s_x = \sum_{i=1}^n (y_i - \mu_x)(y_i - \mu_x)^T \quad 3.10$$

### 3.1.5 Classification models

There are numerous machine learning and deep learning techniques employed for various applications. For our experiment, both machine learning and deep learning techniques were selected based on their performance and popularity. A brief description of the classification algorithms are as follows:

### 3.1.5.1 Support Vector Machine

It is the most exploitable machine learning algorithm for facial expression recognition due to its good classification accuracy; as it may even gain a better classification accuracy than the neural networks (Bhardwaj & Dixit, 2016). SVM has good generalisation ability especially when the labels are properly defined, efficiently process high-dimensional feature data and highly flexible to data size; making it a dynamic and interactive algorithm for facial expression recognition (Ekundayo & Viriri, 2019; Jakkula, 2011; Michel & El Kaliouby, 2015).

SVM belongs to a family of linear classifiers and used for regression and classification purposes. It forms a decision function or a hyperplane from given input vectors, maximizing the margin between the inputs. SVM views the classification problem as a quadratic optimization problem as it classifies the data with the set of support vectors; reducing the structural risk and the average error between the input and their target vectors (Vapnik, Golowich, & Smola, 1997). One against one and one against all are the two approaches available for SVM classification as well as kernel functions such as Radial Basis Function (RBF), linear, sigmoid, and polynomial.

The optimisation function for SVM, given a training-set of instance-label pairs  $(x_i, y_i), i = 1, 2, \dots, l$  where  $x_i \in R^n$  and  $y \in \{1 - 1\}^l$  is defined as:

$$\min_{w,b,\varepsilon} \frac{1}{2} \omega^T w + c \sum_{i=1}^l \varepsilon_i \quad 3.11$$

$$\text{subject to } y_i(\omega^T \phi(x_i) + w) \geq 1 - \varepsilon_i, \varepsilon_i \geq 0$$

According to equation 3.11, there is a mapping of the input vectors  $x_i$  to a higher or infinite-dimensional space defined by the function  $\phi$  and the error term is defined by  $C > 0$ . More, the dimensional space is formed within the linear separating hyperplane (Nugrahaeni & Mutijarsa, 2017).

The functions of the kernel are defined as follows:

$$\text{Linear kernel: } K(x_i, x_j) = x_i^T x_j. \quad 3.12$$

$$\text{Sigmoid kernel: } K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r). \quad 3.13$$

$$\text{Polynomial kernel: } K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0. \quad 3.14$$

Radial basis function (RBF):

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0. \quad 3.15$$

### 3.1.5.2 Convolutional Neural Network (CNN)

CNN originally proposed by Lecun, Bottou, Bengio, & Ha (1998), is an ‘end-to-end’ multi-layered algorithm; an advancement of artificial neural network (ANN) (Yunxin Huang et al., 2019). It is popularly employed for image recognition purposes as well as other computer vision tasks as it requires little or no data engineering.

The CNN model is a three-dimensional space measured by height, width and depth, It consists of a feature detection layer, a feature pooling layer and a classification layer, although the detection and pooling layer may consist of one or more than one step. The convolutional or feature detection layer with the help of a learnable kernel is used to compute the convolution of a set of neurons. Feature maps are produced from this layer and passed to the next layer for further computation. These feature maps are produced from the dot product of the kernel and input neurons. Furthermore, the convolutional layer is characterised by local connectivity which learns the interconnection among bordering pixels, shift-invariance to the location of the object and weight sharing in the same feature map (S. Li & Deng, 2018). The convolution operator is defined as follows:

$$Y = \sum_n^{N-1} W_n * X_n + b \quad 3.16$$

Where  $X \in R^{N \times W \times H}$  is the input with  $N$  channels and grids width  $W$  and height  $H$  pixels.  $W \in R^{N \times K \times K}$  is the convolution filter with a kernel size of  $K * K$ .  $W_n$  is the convolution kernel and the feature map is  $Y$ . Therefore, a convolution over a two-dimensional image  $M$  with kernel  $K$  is defined as (Breuer & Kimmel, 2017):

$$y(j, k) = (M * K)(j, k) = \sum_m \sum_n M(j - m)(k - n)K(m, n) \quad 3.17$$

There are two types of the pooling layer: average and max pooling. The pooling layer is useful in making the network cost-effective as well as reducing the spatial size of the feature maps. Further, all the neurons are converted to one dimension spatial size feature maps from the previous two-dimensional activated activation maps from the preceding layer in the fully connected layer for further feature representation and classification (S. Li & Deng, 2018).

$$y_i = f\left(\sum_{i=1}^m z_i \cdot \omega_{i,j} + b_j\right) \quad 3.18$$

The fully connected layer is represented in equation 3.18 (Jan, 2017). It computes the dot product between the kernel and the neuron from the input data, where  $y_i$  is the output neuron by taking the function  $f(z)$  of the given input  $z_i$  from the previous layer. Then, the sum of all inputs  $z_i$  with the dot product of each weight  $j$  plus bias  $b_j$  is calculated in the fully connected layer. CNN is a good option for classification since its inception due to its embedded feature extraction and selection and its computational efficiency. The neural network-based classifier CNN gives better accuracy than the other neural network-based classifiers (Revina & Emmanuel, 2018). Therefore, CNN was employed for our research. Our architecture is made up of three convolution layers and 2 dense layers.

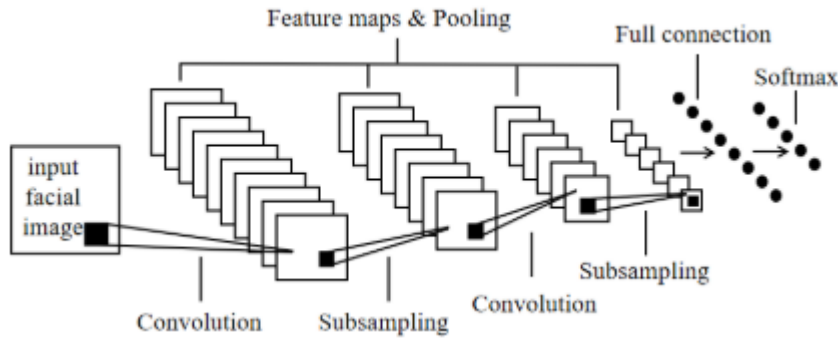


Figure 3.10: Steps involved in CNN FER (Yunxin Huang et al., 2019)

### 3.1.5.3 Ensemble method

In predicting an outcome, the ensemble method builds numerous models by employing either multiple different algorithms or training datasets (Kotu & Deshpande, 2015). Then the independent base models are combined using a technique such as averaging to produce a result. Ensemble methods are normally employed in supervised machine learning tasks. Ensemble methods usually produce optimal models from the combination of multiple base models since the generalisation error is minimised, as there is the likelihood that the error of the single model will be balanced by the other base models. Also, the averaging of the multiple different models with minimum bias leads to a higher prediction performance of the model as compared to a single predicted model (Valero, 2016). Ensemble methods in machine learning are good techniques as they solve issues of overfitting and have proven to be a computational cost-effective method (Dev & Eden, 2019; Sagi & Rokach, 2018).

The general framework for building an ensemble model is as follows. Given datasets  $i$  and  $k$  features  $K = \{(j_i, m_i)\}$ , ( $|K| = i, j_i \in R^k, m_i \in R$ ), the ensemble learning model  $\varphi$  uses an aggregation function  $F$ , combines  $B$  base models,  $\{g_1, g_2, \dots, g_b\}$  towards attaining an output defined as:  $\hat{m}_i = \varphi(j_i) = (g_1, g_2, \dots, g_b)$ , having  $\hat{m}_i \in Z$  for classification problems and  $\hat{m}_i \in R$  for regression problems (Dev & Eden, 2019; Sagi & Rokach, 2018).

For our work, the following base models: SVM, KNN, Naïve Bayes, Logistic Regression and Random Forest will be used. The obtained individual results will be combined and averaged to produce a better model for the recognition of anger. These base models were selected because of the following reasons. SVM performs excellently on binary classification (J. & Watkins, 1999). Naïve Bayes converges better with small datasets as compared to other models especially when its conditional independence assumption holds (Rish, Hellerstein, & Jayram, 2001). Random forest is non-parametric, efficient, has high prediction accuracy for many types of data, works perfectly with small sample size, has high dimensional space and complex data structures. Additionally, the random forest is less prone to overfitting as compared to many classifiers and works on many datasets (Agarwal, Baechle, Behara, & Rao, 2016; Breiman, 2001). KNN is easy to implement and similar to SVM as it finds the relationship between data points (Winters-Miner et al., 2015) and LG has the advantages of memory saving and better performance than other 2D approaches in two-class problem (Yun, Kim, Chi, & Yoon, 2007).

### 3.1.6 Elements involved

The elements utilised in implementing the three models: the SVM model, the CNN model and ensemble model are described below:

#### 3.1.6.1 Programming language

Choosing a machine learning programming language for this research work was a difficult task as there exists a number of these languages. Therefore, we evaluated the pros and cons of each of the most widely used machine learning programming languages and selected one of them. These popular machine learning programming languages include Python, C/C++, Go, R and Matrix Laboratory (MATLAB) (Gao et al., 2020).

MATLAB is an interactive, easy-to-use, fast programming language that is used for scientific computing. It can be employed for tasks such as data analysis, development of algorithms,

matrix manipulations, problem-solving and so on. MATLAB has good performance, provides an easy-to-use graphics, concise syntax and allows for easy language extension, however, a licence is required to use the product and some of its libraries. Go is an open-source programming language developed by Google with syntax similarity to C used for building simple, efficient, and reliable software. Its syntax is concise, expressive and enables the flexible segmented construction of programs. When implementing machine learning algorithms, libraries written in Go are used other than using other libraries in different languages. However, these machine learning libraries in Go are not numerous. R is an open-source programming language for statistical computing. R is highly graphical, producing high-quality images. Yet it is characterised by a steep learning curve and limited when analysing big data as it stores its data in the system memory (RAM). C/C++ is a powerful and efficient general-purpose programming languages used across multiple platforms. However, it is difficult developing and implementing machine learning algorithms using C/C++ as it is challenging learning C/C++ (Gao et al., 2020; Valero, 2016).

After careful analysis and evaluation, python was selected as the programming language for this research as it is easy to use, requires no licence, has easy portability, well-defined error model based on exception, good performance, simple and easy to learn, efficient, availability of a documentation and community support to resolve problems within the shortest possible time (Gao et al., 2020; Oliphant, 2007; Pérez, Granger, & Hunter, 2011).

### 3.1.6.2 Packages and development environments

The following packages and development environment were employed for this research, following the selection of python as the programming language. Familiarity, flexibility, simplicity, the availability of detailed documentation for easy implementation and community support were the reasons for the selection of these packages and development. As the success

of machine learning or deep learning project is dependent on the frameworks and libraries available to developers (Valero, 2016). A brief description of them is as follows.

#### 3.1.6.2.1 Anaconda

Anaconda is a complete, open-source package manager, environment manager, Python and R programming languages distribution for scientific computing and data science. It is easy to download and install and functions on cross-platforms. It simplifies package management as well as deployment. Anaconda provides a graphical user interface (GUI) which includes a link to all the applications which can be installed with just a mouse click. The applications included in the Anaconda package include JupyterLab, JupyterNotebook, Spyder, Orange, Glue, Visual Studio Code and RStudio. It simplifies installing of libraries and dependencies as it comes with over 250 automatically installed packages and over 7500 open-source libraries which can be installed using either pip or conda. In addition, multiple virtual environments can be created using Anaconda. For example, a Python 2.7 can be installed instead of the default python. Also, Anaconda provides detailed documentation as well as community support for additional help (Watkins, 2018).

#### 3.1.6.2.2 Open Source Computer Vision Library (OpenCV)

OpenCV is an open-source python library built for image and video analysis such as face detection and recognition, identifying objects, classification of objects in video etc with more than 2500 optimised computer vision and machine learning algorithms. It has the interfaces: Python, C++, MATLAB and Java interfaces and functions on a cross-platform. We utilised OpenCV mainly for our pre-processing stage as it contains functions such as the Viola Jones algorithm for face detection and image smoothening functions like median blur () and clahe () for histogram equalisation. This greatly reduces the efforts required for the pre-processing stage (Culjak, Abram, Pribanic, Dzapo, & Cifrek, 2012).

### 3.1.6.2.3 Tensorflow

Tensorflow is a free, open-source library used for numeric computation. Tensorflow operates using dataflow graphs and provides an end-to-end implementation and training of machine learning models particularly neural networks. Tensorflow allows for its deployment across diverse platforms such as GPU, CPU and TPU. The higher layers provide an application programming interface (API), commonly used in deep learning models (Culjak et al., 2012).

### 3.1.6.2.4 Keras

Keras is a deep learning API written in Python for developing and training of deep learning models. It is integrated into Tensorflow and was developed for faster experimentation of deep learning models. Keras has user-friendly, highly productive interface and modifiable and composable models. For this research, Keras was utilised for our data augmentation as well as the training of our deep learning models.

### 3.1.6.2.5 Scikit-learn

It is an open-source python library developed for the training of machine learning models, dimensionality reduction, model selection and feature extraction and normalisation. It is useful for the implementation of our SVM and ensemble model and evaluating our models using performance measures such as confusion matrix.

Table 3.1: summary of the package development and environments.

Operating system	Windows
Language	Python 3.7.4
Editor	Spyder (Anaconda)
Environments	OpenCV, Keras, Tensorflow, Scikitlearn

### 3.1.7 Development

For the development of our research work, a modular design method was used. The modular design allows the development of our anger recognition to be performed in modules or components, having each module performing a specific function. This allows changes to be made easily without or minimally affecting the other components (Valero, 2016). Thus, our work was divided into the pre-processing, feature extraction and classification stages. The pre-processing stage involves the grayscaling and resizing the image, face detection, image smoothing and normalisation. The feature extraction module involves the extracting of the features and labels. Then the classification stage receives the features and labels and creates the ensemble, CNN and SVM models, the classification of the facial expressions and the likelihood of each expression.

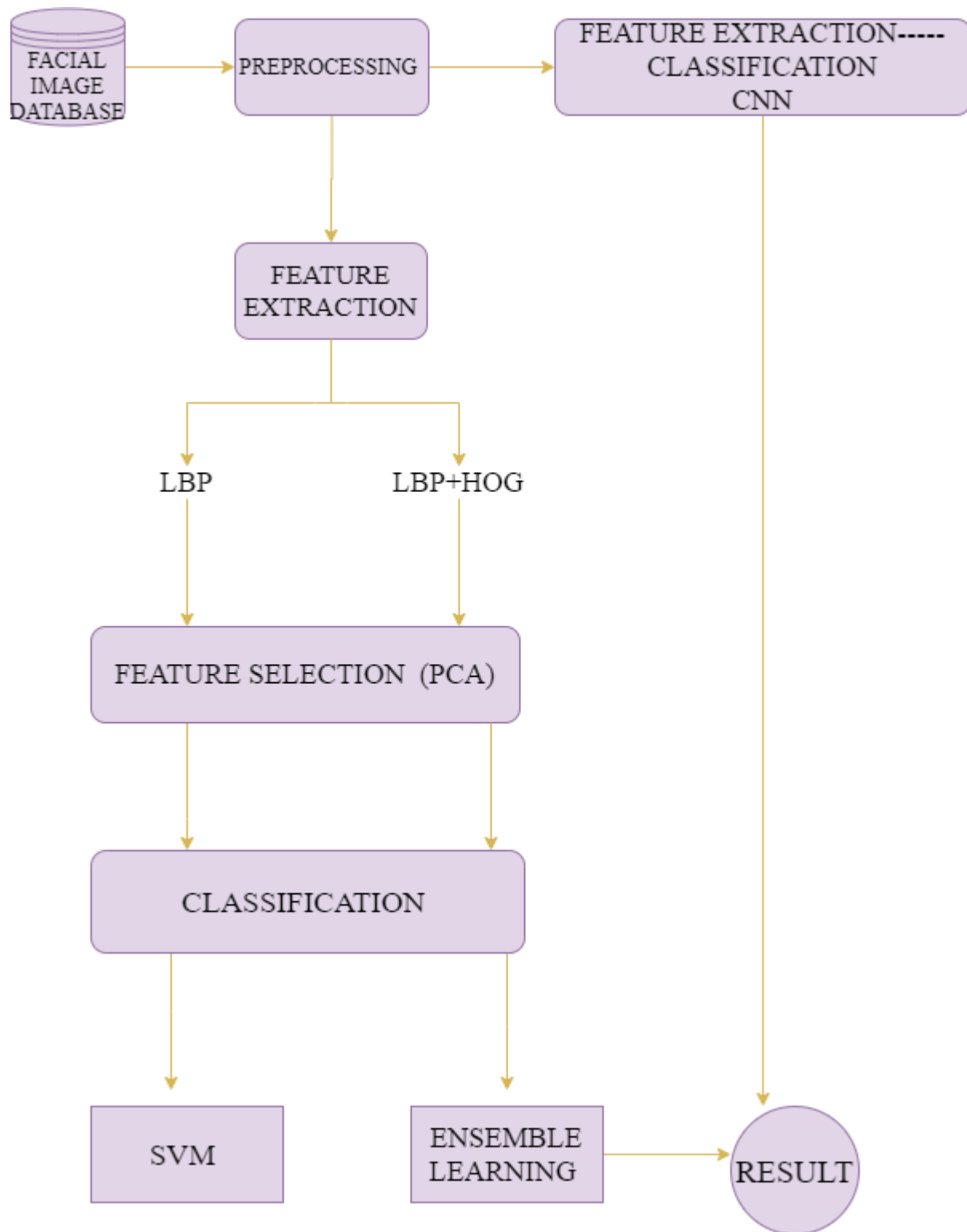


Figure 3.11: the workflow of our research work (Nassih et al., 2019).

### 3.1.8 Chapter summary

This chapter detailed how our research work is going to be undertaken particularly giving background and justification for the chosen methods. Moving on, we will discuss the implementation of these methods in the next chapter.

## Chapter 4

### Experimental setup

#### 4.1 Introduction

This chapter details the implementation of our research work. We describe how the features are extracted, fused, and reduced using feature dimensionality reduction methods, and classified using machine learning and deep learning algorithms.

Before proceeding to the details of the proposed experiment, a summary of the experiment to be conducted is described as follows. The datasets JAFFE, CK+ and KDEF will be utilised for the experiment. Two experiments will be conducted, outputting 3 models. Experiment I involve using the original datasets whilst experiment II involve the application of data augmentation technique to balance our dataset.

#### 4.2 Hardware specification

In performing a machine learning or deep learning project, selecting the hardware components is a key factor as the project is highly dependent on this component. The CPU hardware specification employed in experimenting is summarized in table 4.1. More for the TPU hardware specification, we employed Google Colaboratory (Google Colab). Google Colab is a free cloud service with colab notebooks which has been built on top of Jupyter notebooks and Ubuntu 18.04. Colab notebooks leverage on the power of Google's hardware, executing codes in the cloud. It allows for 'end-to-end' processes involved in facial expression recognition, from pre-processing to evaluating of models. It is particularly helpful in training deep neural networks as it provides Tensor Processing Unit (TPU), an integrated circuit particularly for neural network learning developed by Google (Feldman, 2018); since training neural networks

can be a protracted process depending on the model complexity and the resources available.

TPU provides access to Ram of 12.72GB and hard disk of 107.77GB (Valero, 2016).

Table 4.1: a summary of the hardware specification.

System model	HP Pavilion x360 m3 Convertible
Processor	Intel(R) Core (TM) i3-7100U CPU @ 2.40GHz, 2400 MHz, 2 Core(s), 4 Logical Processor(s)
Memory (RAM)	6.00 GB
System type	x64-based PC

### 4.3 Pre-processing stage

It outlines the steps taken in transforming our datasets due to the different features such as colour, size, number of emotions and resolution to get a unified input for the next stages.

#### 4.3.1 Database pre-processing

Three widely utilised databases: JAFFE, CK+ and KDEF were utilised to test the performance of our proposed models. To have a uniform standard input image, the databases were modified a bit to select only frontal, posed images for our experiment. The CK+ database has its images captured in the form of image sequences, from the neutral emotion transitioned to the peak emotion (the desired expression at the end of the image sequence). Also, the KDEF database has its images photographed from five angles: half left, half right, frontal, full left and full right (Goeleven et al., 2008). That being so, we selected only peak images and frontal images from the CK+ and the KDEF database respectively for this experiment. Subsequently, both the KDEF and JAFFE databases containing images expressing 7 different emotions whereas CK+ has 8 emotions. Therefore, to have a unified emotion type for our experiment, the contempt

emotion in the CK+ database was excluded. So, overall, 213, 329 and 980 images for JAFFE, CK+ and KDEF datasets respectively were selected and utilised for this experiment.

The datasets were manually prepared after the modification. They were categorised into emotions: “angry” and “not-angry” with emotion labels: 1 and 2 accordingly. The “not-angry” images consist of the other emotions apart from angry that is the combination of happy, sadness, fear, neutral, disgust and surprise. A summary of the databases is listed in table 4.2.

Table 4.2: summary of the datasets.

<b>DATABASE</b>	<b>TOTAL NUMBER</b>	<b>SIZE (PIXELS)</b>	<b>ANGRY</b>	<b>NOT-ANGRY</b>
<b>JAFFE</b>	213	256*256	30	183
<b>CK+</b>	327	640*490	45	282
<b>KDEF</b>	980	562*762	140	840

#### 4.3.2 Image pre-processing

The following procedures were implemented: grayscaling and resizing, face detection and cropping and image enhancement (Dagher, Dahdah, & Al Shakik, 2019).

#### 4.3.2.1 Grayscale and resizing of images

The database comes in different colour format. The JAFFE images are originally grayscale. The CK+ images are either coloured or grayscale. Likewise, the KDEF images are coloured. Therefore, the images were all first tested to know their colour format whether grayscale or coloured and then, the images were all grayscale to obtain a unified input of images (Dagher et al., 2019).

Follow-up was to ensure the images were all the same sizes. Hence to provide uniform image sizes, all the images were resized to from their original size to a size of 128\*128. The reduction of the sizes of images helps boost training time (Goeleven et al., 2008). It was observed there were no changes in the resolution of the images after the grayscale and resizing (Dagher et al., 2019).



Figure 4.1: Example of a gray scaled and resized CK+ angry image (from a size of 640\*480 to 128\*128).

#### 4.3.2.2 Face detection and cropping

After the grayscale and resizing of the images, the detection of faces from the images is performed. The viola jones algorithm implemented in the OpenCV library, which makes the face detection easier is used for this task. The viola jones algorithm is employed because it detects faces smoothly and at a faster rate. The face detection algorithm detects the face in the

images. It operates by looping through the images one after the other, finding shapes that resemble a face and scans a sub-window around it (Viola & Jones, 2004). The viola jones algorithm aligned and created a bounding box around the detected face in the image. Then, the rectangular area of box with the detected faces is cropped and saved for further pre-processing steps (Rani & Garg, 2014).

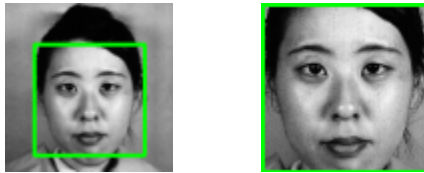


Figure 4.2: viola jones face detection on the left and a detected and cropped face on the right.

#### 4.3.2.3 Image enhancement

In this phase, the median filter was employed to denoise the images. The median filter is a non-linear filter which removes noise from images such as salt and pepper, impulse noise and so on. However, it preserves the edges of the images after denoising. It works as a sliding window and changes the gray level of each pixel with the median gray level in a locality of pixels (Rani & Garg, 2014; Tan & Jiang, 2019). Figure 5.3 displays a denoised JAFFE image.

Next, Contrast Limited Adaptive Histogram Equalisation (CLAHE) was utilised in improving the pixel intensities in an image, enhancing the facial features for easy extracting. CLAHE is performed after denoising in order not to intensify the noise in the images. It operates by dividing an image into non-overlapping contextual tiles and performs a histogram equalisation on each tile. Then, the adjoining tiles are combined using bilinear interpolation (Zhao, Georganas, & Petriu, 2010). In implementing the median filter and CLAHE for our experiment, the functions `median Blur ()` and `cv. createCLAHE ()` in OpenCV were utilised.



Figure 4.3: a denoised JAFFE image using median blur.



Figure 4.4 displays a CLAHE enhanced JAFFE image.

#### 4.4 Feature extraction

##### 4.4.1 LBP feature extraction

The LBP method is utilised for our feature extraction. As LBP makes use of neighbourhood radius and points, a series of experiments were performed to determine the optimal values for radius and points, respectively. Therefore, for our experiment, 24 and 8 were selected as the number of neighbourhood points and radius respectively which forms a circularly symmetric neighbour set. Then, the LBP operator is performed on all the images. Its computed histogram is texture-based image descriptor with a total feature vector of dimension 26 (Hossain, 2018).

##### 4.4.2 HOG feature extraction

The HOG method was applied to all the images. The optimal parameters found after a series of test were a cell size of 16 and 8 number of bins orientations. Thus, in calculating the feature size given an image size of 128\*128 pixels, the total feature size obtained was:

$$\frac{128 \times 128}{16 \times 16} \times 8 = 512$$

#### 4.4.3 Hybrid feature extraction

We utilised the hybrid feature extraction method specifically feature-level fusion method for the ensemble learning method feature extraction. Feature-level fusion locally concatenated the feature vectors of HOG and LBP.

For our CNN algorithm, feature extraction was performed in the convolution and the pooling layer.

#### 4.5 Feature selection

PCA was used for dimensionality reduction both for the LBP feature vector and the fused feature vector. This is important because the face contains some similarities in each facial expression and would create features for all. As such, these correlated irrelevant features need to be discarded retaining only the relevant ones. Therefore, PCA was set to maintain 98% of the bias between features, discarding noisy and irrelevant features (Jan, 2017).

#### 4.6 Classification

This stage centres on training the various models by varying and setting hyperparameters, applying data augmentation techniques to improve the models, using different datasets, and studying how the models perform by experimenting with it.

##### 4.6.1 Model selection

It is a difficult task as it involves selecting the architecture of the algorithms and its hyperparameters. Model selection involves selecting the architectures of the various algorithms and its hyperparameters. It is a difficult task as it involves tuning the hyperparameters and utilising the different datasets to get the best performing models.

The following subsections describe the elements that were considered during the training and testing stage as learning time and classification accuracy is dependent on the hyperparameters and the model architecture (Valero, 2016).

#### 4.6.2 Hyperparameters tuning

In training and getting an accurate and precise model, hyperparameter tuning needs to be performed. It helps in determining the best model and architecture for an algorithm as well as the right balance between bias and variance. Hyperparameters are the parameters which define a model architecture. Hyperparameter tuning is the process of selecting the optimal parameters for a model. Hyperparameter tuning can be defined as the process of adjusting the parameters or set of an algorithm to improve performance. A description of the selected hyperparameters which were hyper tuned are as follows:

1. The number of convolutional layers: the selection of convolution layers in building a CNN model is a key factor to prevent problems such as overfitting and vanishing and exploding gradient.
2. The number of hidden layers: the number of hidden layers chosen depends on the amount of data size used for training. This should be chosen carefully to precisely find the difference between bias and variance.
3. Activations functions: there are 4 activation functions used for CNN. This includes ReLu, sigmoid, Tanh and LeakyReLu, Sigmoid and Tahn are used for shallow networks. It is an important hyperparameter as it controls the firing of neurons.
4. Learning rate: it is a key factor in determining convergence to of an algorithm to a satisfactory solution as it determines the number of iterations. It should be tried in powers of 10 to determine the optimal one (Ramesh, 2018).
5. Dropout: it is a regularising technique to overfitting by finding an optimum bias-variance spot.
6. Number of Epoch: it determines the number of iteration and the time a training process will last and how well a model will fit on the train data as well as improve the generalisation error.

7. Optimiser: optimiser minimises the error during training. The optimiser speeds the convergence rate as well as optimise the internal parameters. The commonly used optimisers include Adaptive Momentum (Adam), Root Mean Square Propagation(RMSprop), Adaptive Gradient (Adagrad), Adaptive Delta (Adadelata) and Nesterov Accelerated Gradient (Nadam) (Prilianti, Brotosudarmo, Anam, & Suryanto, 2019).
8. Number of fully connected layer controls the quality and activation maps
9. Kernel: the kernels for SVM are sigmoid, polynomial, sigmoid, radial basis function (RBF) and linear. However, the most used kernel is RBF. The preference kernel determines the performance of an algorithm as well as the distinguishing of classes for classification purposes.
10. The number of trees ( $n\_estimators$ ): random forest can be defined as the grouping of trees. Therefore, we need to decide the number of trees to use for our computation as the computation efficiency is dependent on the number of trees utilised.
11. The number of features considered for splitting a node ( $max\_features$ ): it is the maximum number of features available to each tree in a random forest.
12. Distance metric: It helps to find the closest or similar learning points.
13. Number of neighbours ( $k$ ): is an important factor in determining the prediction model

To simplify the process of hyperparameter tuning and obtaining the optimal models at the same time, we utilised the hyperas module in Keras for the CNN models and grid search using cross-validation (GridSearchCv) module in scikit-learn library for both the ensemble and SVM models.

#### 4.6.3 Settings and protocols

The splitting of the dataset was done using scikit-learn `train_test_split`. The datasets were split into 80-20 ratio, 80% for the training of the model and 20% for testing. The images were randomly chosen, thus the procedure was repeated 20 times and the average accuracies were calculated (V. Jain, Lamba, Singh, Namboothiri, & Dhall, 2019). The models were trained one after the other on each dataset, using the same architectural model and hyperparameters (Minaee & Abdolrashidi, 2019).

For the SVM, the One Vs One approach was used where the classification would be done between the two labels. More, the decision-level fusion was utilised for combining the accuracies of our ensemble base models.

#### 4.6.4 Data augmentation

Data augmentation is a useful technique used for increasing the number of images. It helps resolves issues of overfitting (B. Yang, Cao, Ni, & Zhang, 2017).

This method was applied to our original datasets with the possibility of increasing the dataset (see table 4.3). We used the method: rotation, flipping and zooming. Rotation turned the image uniformly from the central point to 20 degrees clockwise; the image was flipped horizontally and vertically to create other samples; zooming was the result of randomly extracting sections of images and increasing their size (see figure 5.5).

#### 4.7 Chapter summary

We provided details on how our work was implemented. This is important for reproduction or future advancement of our work. Next chapter will discuss the outcome of our experiment.

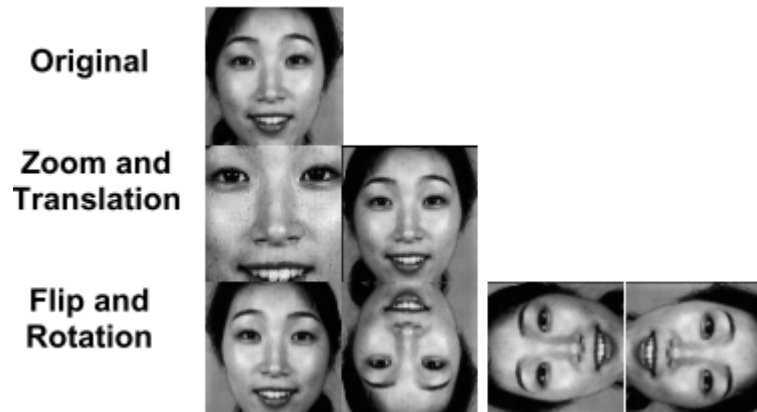


Figure 4.5:transformations applied to a JAFFE image (Valero, 2016).

Table 4.3: summary of the datasets after data augmentation.

<b>DATABASE</b>	<b>TOTAL NUMBER</b>	<b>SIZE (PIXELS)</b>	<b>ANGRY</b>	<b>NOT-ANGRY</b>
<b>JAFFE</b>	366	256*256	183	183
<b>CK+</b>	564	640*490	282	282
<b>KDEF</b>	1680	562*762	840	840

## Chapter 5

### Experimental results and discussion

#### 5.1 Introduction

In this section, the performance of the proposed methods and the obtained results are discussed below. Here, labels are represented as 1 for Angry and 2 for Not Angry.

#### 5.2 Evaluation metrics

The datasets: CK+, JAFFE and KDEF were used to evaluate the proposed methods. Also, the evaluation metrics confusion matrix, precision, recall, f1-score, and ROC curve were utilised for our experiment (Nisbet, Miner, & Yale, 2018).

A confusion matrix is a table that shows the association between the predicted label and the true label in a classification problem, thus describing the performance of a classifier.

The following error metrics are used in calculating other metrics are:

True positive (TP): number of correctly predicted labels. That is a positive label that is correctly predicted as positive.

True negative (TN): number of correctly predicted labels. That is a negative label that is correctly predicted as negative.

False positives (FP): number of labels falsely predicted as positive. That is when a label is predicted as positive when it is negative.

False negatives (FN): number of labels falsely predicted as negative. That is when a label is predicted as negative when it is positive.

Metrics such as accuracy, recall, precision and f1-score are computed from rates of a binary confusion matrix (Nisbet et al., 2018). Thus, the metrics are defined as:

$$Accuracy = \frac{TP+TN}{N} \quad 5.1$$

$$Precision(P) = \frac{TP}{TP+FP} \quad 5.2$$

$$Recall(R) = \frac{TP}{TP+FN} \quad 5.3$$

$$F1-score = 2 \times [(P \times R) / (P + R)] \quad 5.4$$

ROC curve (Shu et al., 2018):

$$\text{False-positive-rate (FPR)} = \frac{FP}{TN+FP} \quad 5.5$$

$$\text{True-positive-rate (TPR)} = \frac{TP}{TP+FN} \quad 5.6$$

Due to the severe imbalance nature of the datasets (see table 4.2), the performance metrics: accuracy, precision, recall, F1-Score and confusion matrix were used in evaluating the model, as using only accuracy would give a false representation of the model (Koehrsen Will, 2018). However, for the second experiment, accuracy, confusion matrix and ROC curves were employed as the datasets were balanced using the data augmentation technique. The macro average value is used as the precision value. Macro average is the sum of the two precisions of both classes divided by 2.

### 5.3 Experiment I (training without data augmentation)

The experiment results of the various classifiers are summarised in tables 5.1 to 5.3. It is interesting to note that generally the recognition accuracies and the precision and recall values varies based on the combination of descriptor and classifier, although some patterns are observed. The precision and recall values obtained were generally low and can be attributed to the imbalance nature of the datasets, as these two measures only give output on the relevant cases in a dataset.

In Table 5.1 below, we see the results of SVM, CNN and Ensemble learning models on the JAFFE dataset. In terms of accuracy, our ensemble learning model achieved 98% whilst SVM

and CNN attained 90% and 88% respectively. It also performed better in terms of F1-score, precision, recall and accuracy. The performance of the models can be attributed to the fusion of the descriptors and classifiers for the ensemble model, SVM’s ability to train on modest training data and CNN’s low performance justified by its requirement for huge training data. Further, the ensemble learning predictor ( a combination of KNN, SVM, Naïve Bayes (NB), Random forest (RF) and Logistic Regression (LB)) with HOG and LBP descriptors), achieved the greater accuracy of 98%; thus 8% and 10% greater than SVM and CNN accordingly.

Table 5.1 shows the performance (accuracy, recall, precision, and F1-score) of the JAFFE dataset on the SVM, CNN and ensemble learning models.

	Accuracy	Precision	Recall	F1-Score
SVM	0.9	0.9	0.89	0.89
CNN	0.88	0.7	0.59	0.61
ENSEMBLE	0.98	0.99	0.92	0.95

The overall value of the various performance metrics (accuracy, precision, recall and F1-score) on the CK+ dataset (Table 5.2) is significantly higher than the other datasets. SVM attained a higher accuracy of 97%, having CNN and Ensemble learning models with 94% and 93% respectively. SVM is thriving to achieve a good accuracy over the two tables probably due to its flexibility to perform on any amount of training data.

Table 5.2 displays the performance (accuracy, recall, precision, and F1-score) of the CK+ dataset for the models: SVM, CNN and ensemble learning.

	Accuracy	Precision	Recall	F1-Score
SVM	0.97	0.82	0.70	0.74
CNN	0.94	0.81	0.61	0.65
ENSEMBLE	0.93	0.46	0.50	0.48

According to Table 5.3, CNN achieves an accuracy of 93%, Ensemble learning had 92% and SVM with an accuracy of 89%. CNN model attained the leading accuracy of 93%. It is not surprising CNN attains the highest accuracy on the KDEF dataset as among the three utilised datasets for our experiment, KDEF contains the highest number of images (980 images). And as CNN requires a large amount of training data; thus, explains the accuracy achieved on the CNN model. Generally, the precision and recall values were low as they give the correct value of the relevant data points that are angry.

Table 5.3: KDEF dataset performance (accuracy, precision, recall and f1-score) for CNN, SVM and ensemble learning models.

	Accuracy	Precision	Recall	F1-Score
SVM	0.89	0.45	0.50	0.47
CNN	0.93	0.86	0.85	0.85
ENSEMBLE	0.92	0.90	0.76	0.81

The confusion matrices were utilised in further analysing the performance of our models (see figures 5.1 to 5.3). The confusion matrix is visualised by having the true label on the vertical axis and the expression recognised by the classifiers on the horizontal axis. The intensity of confusion of each expression with its counterparts is indicated in each row of the matrices (Abd El Meguid & Levine, 2014). Also, the grayscale levels across the figure present the inter-expression similarity across the two expressions (Ali, Iqbal, & Choi, 2016).

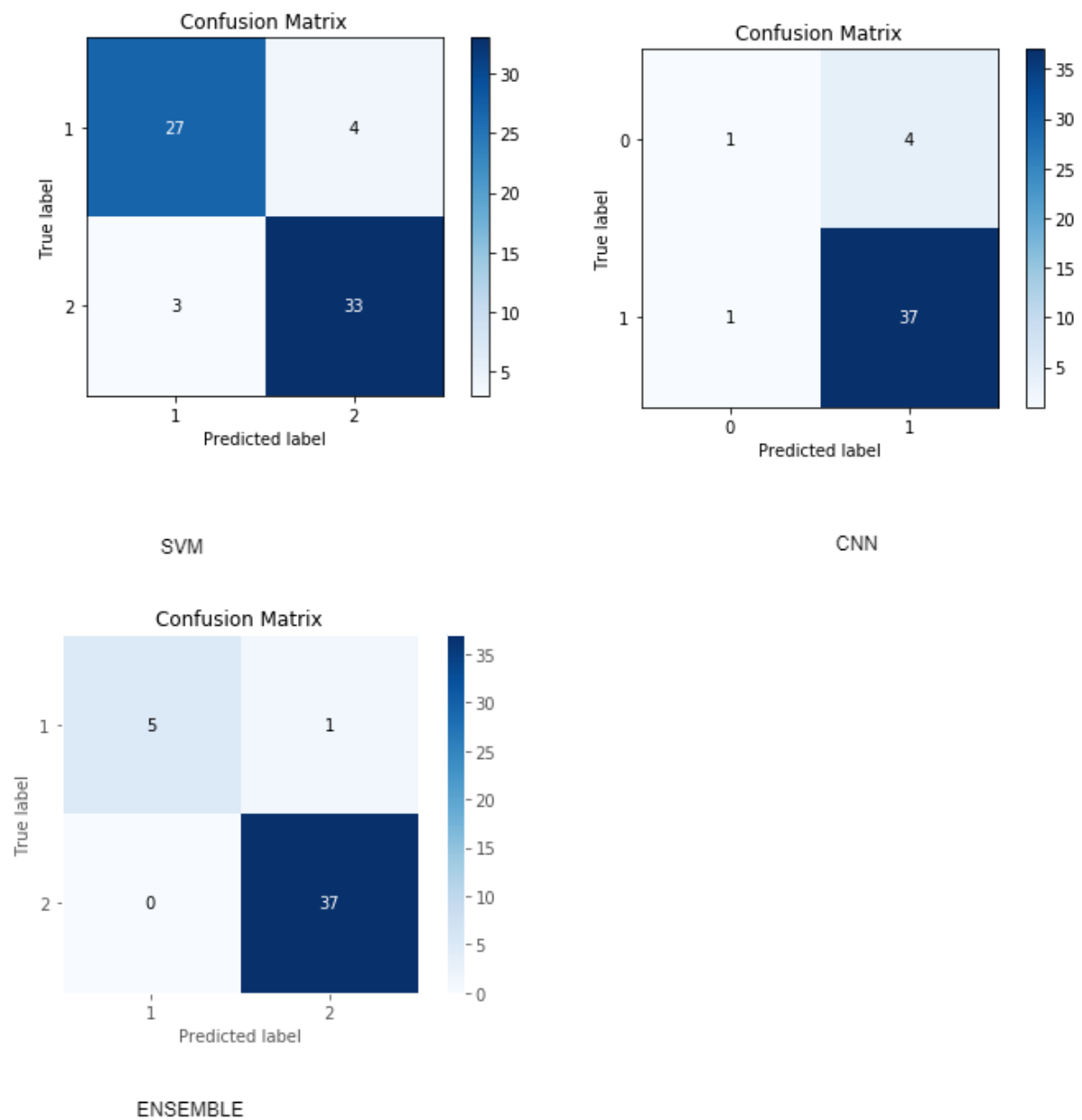


Figure 5.1: Confusion matrices for the models: SVM, CNN and Ensemble learning for the JAFFE dataset – Experiment I.

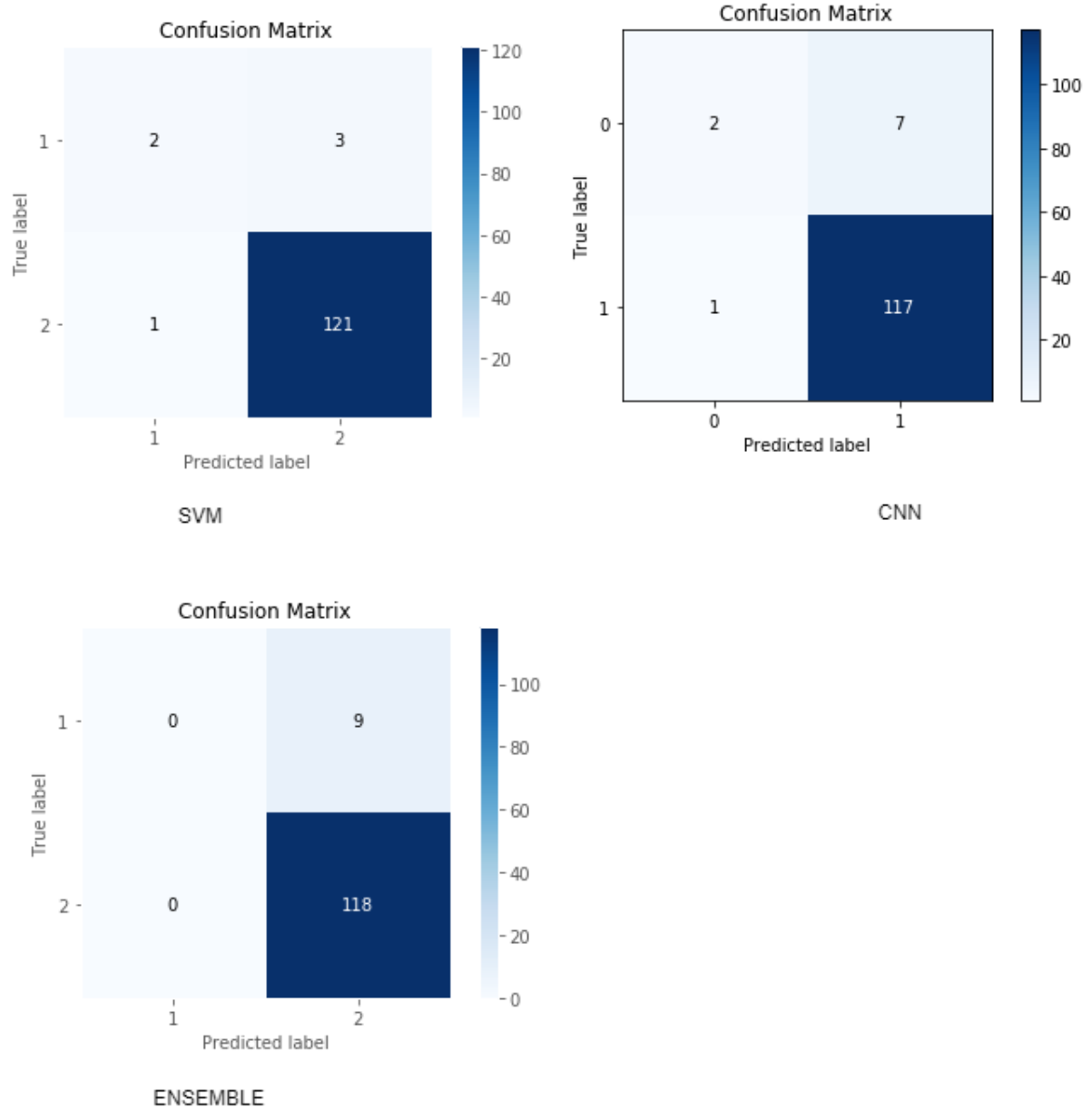


Figure 5.2: Confusion matrices for the models: SVM, CNN and Ensemble learning for the CK+ dataset – Experiment I.

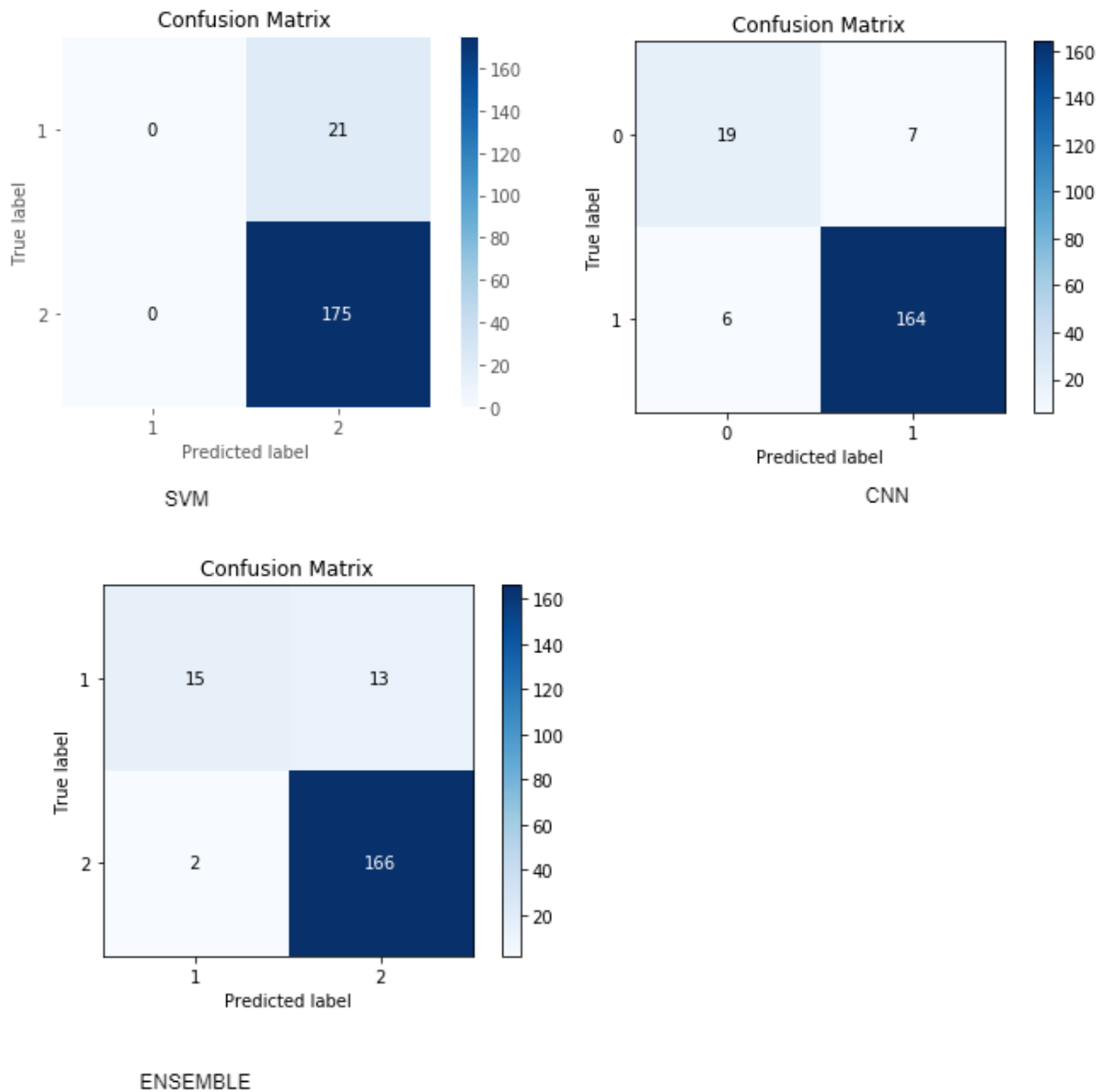


Figure 5.3: Confusion matrices for the models: SVM, CNN and Ensemble learning for the KDEP dataset – Experiment I.

#### 5.4 Experiment II

There was an improvement in the results of the evaluation metrics for the various models, after the application of the data augmentation technique as illustrated in tables 5.4 since the balancing of the two classes of the datasets.

Table 5.4 contains the results of the second experiment on the JAFFE, CK+ and KDEP datasets. Generally, there has been an improvement in the performance measures used in evaluating our models. The ensemble learning model is the best performing algorithm on the JAFFE dataset.

The best performing model increased by 2%, having the best performing SVM and CNN models with accuracies 97% and 97% on JAFFE and CK+ datasets, respectively. More, the SVM and CNN models attained the same accuracy on the KDEF dataset.

Table 5.4: JAFFE dataset performance (accuracy) on the CNN, SVM and ensemble learning models.

ACCURACY (%)			
	JAFFE	CK+	KDEF
<b>SVM</b>	97	95	95
<b>CNN</b>	96	97	95
<b>ENSEMBLE</b>	100	96	97

Overall, the best performing model was the ensemble learning model on the JAFFE dataset (see figure 5.4). All the models generally had some confusions except for the ensemble model in figure 5.4. The highest confusion occurred on (figure 5.3). From experiment I, the results indicate our true positives (angry) were just a considerable small amount whilst the true negatives (not angry) were a considerably large amount. This can be attributable to the imbalance nature of the dataset as with the JAFFE dataset, the angry images are 45 whilst the not angry images are 168 images.

It was noted that the best performing model on JAFFE dataset was the ensemble learning model as it had no misclassifications (figure 5.4).

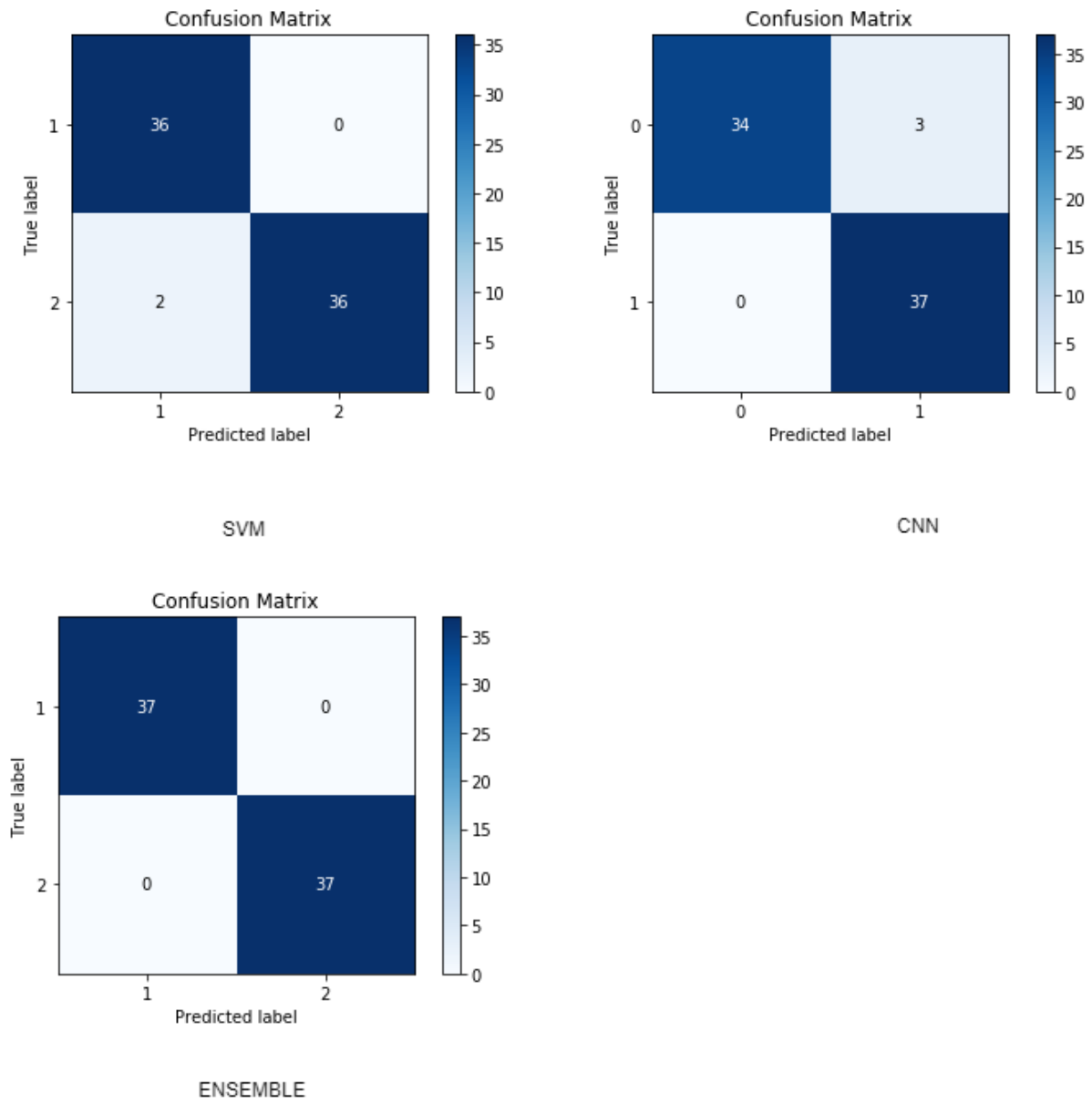


Figure 5.4: Confusion matrices for the models: SVM, CNN and Ensemble learning for the JAFFE dataset – Experiment II.

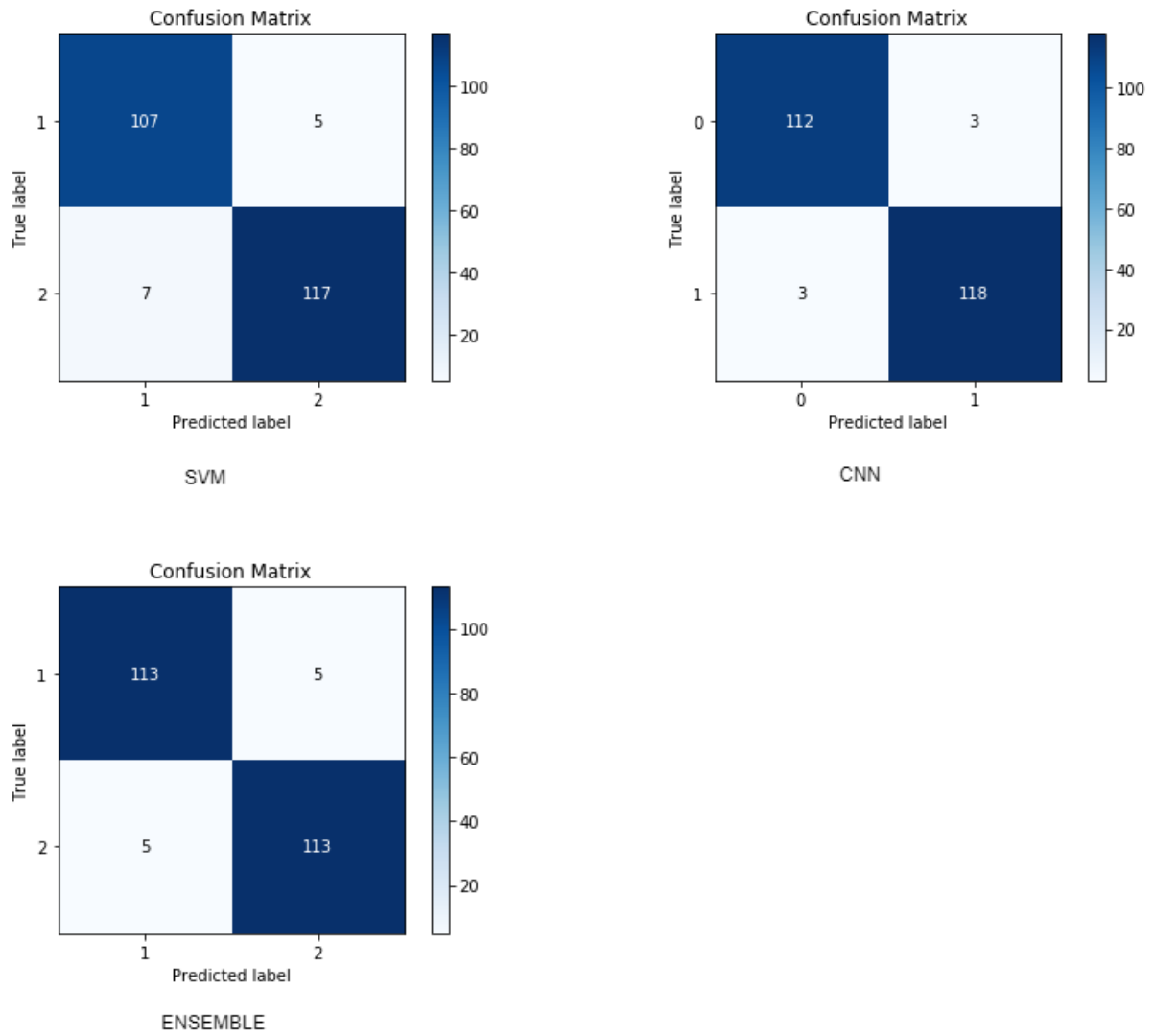


Figure 5.5: Confusion matrices for the models: SVM, CNN and Ensemble learning for the CK+ dataset – Experiment II.

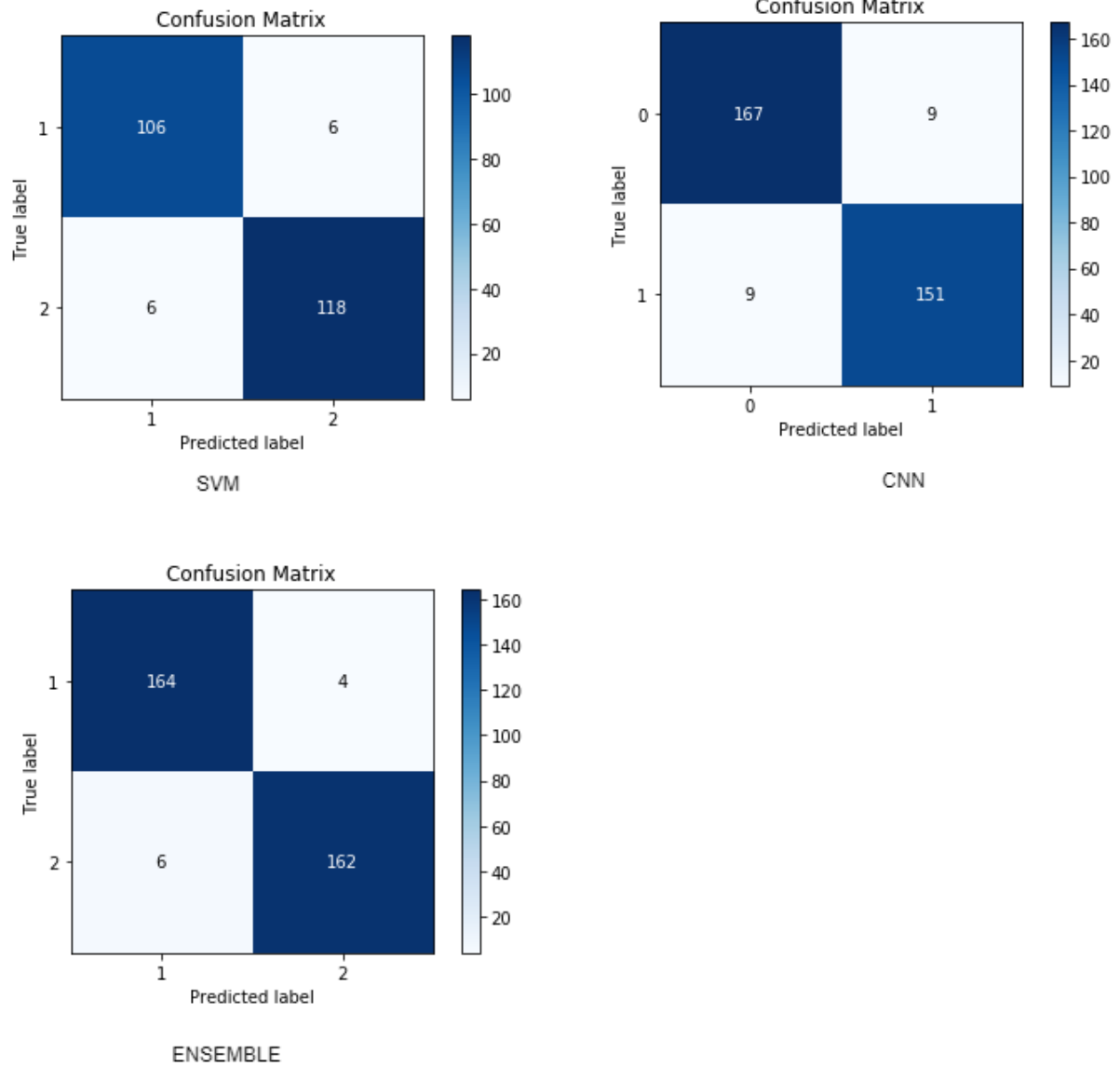


Figure 5.6: Confusion matrices for the models: SVM, CNN and Ensemble learning for the KDEF dataset – Experiment II.

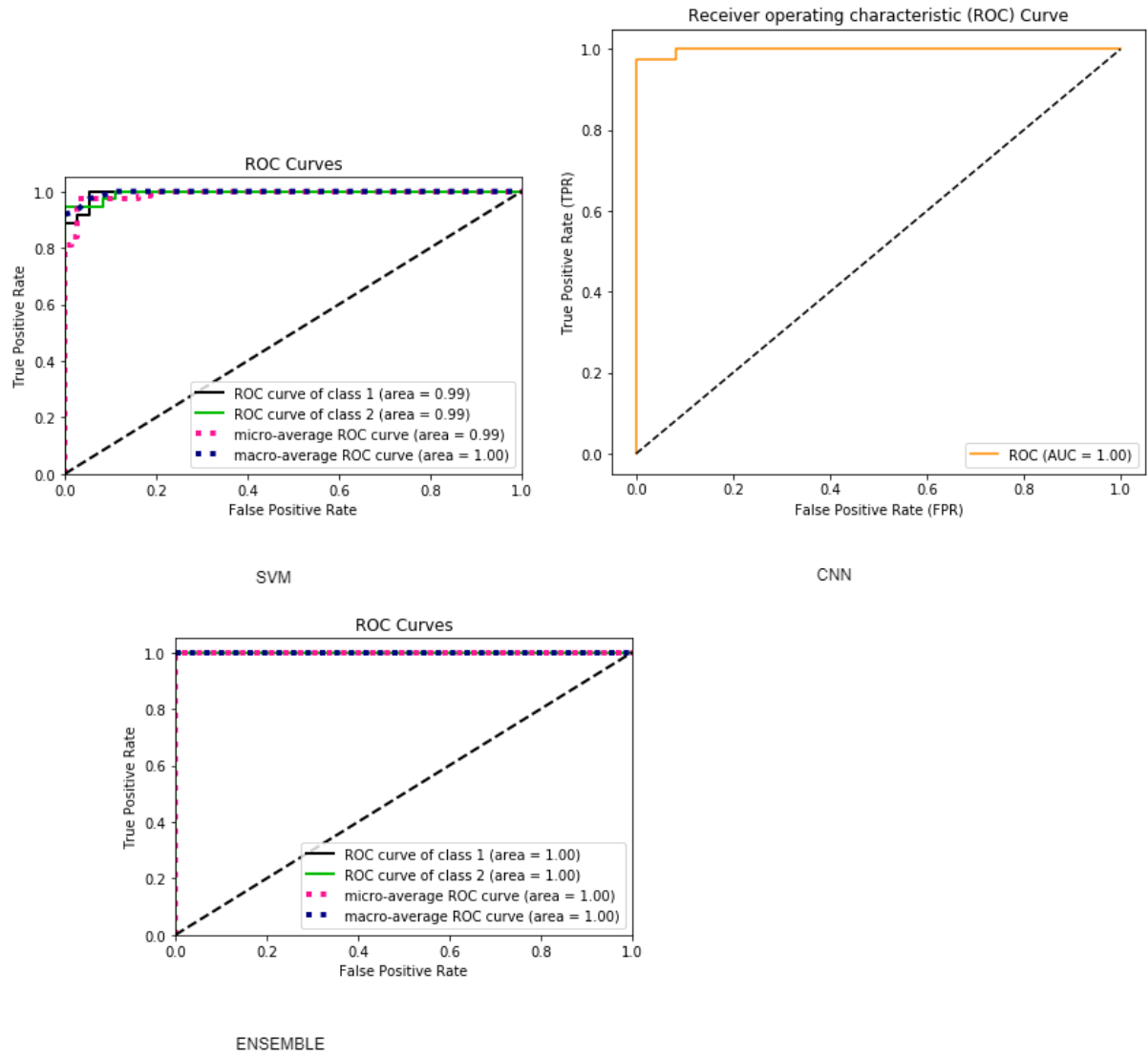


Figure 5.7: ROC curves on the JAFFE dataset.

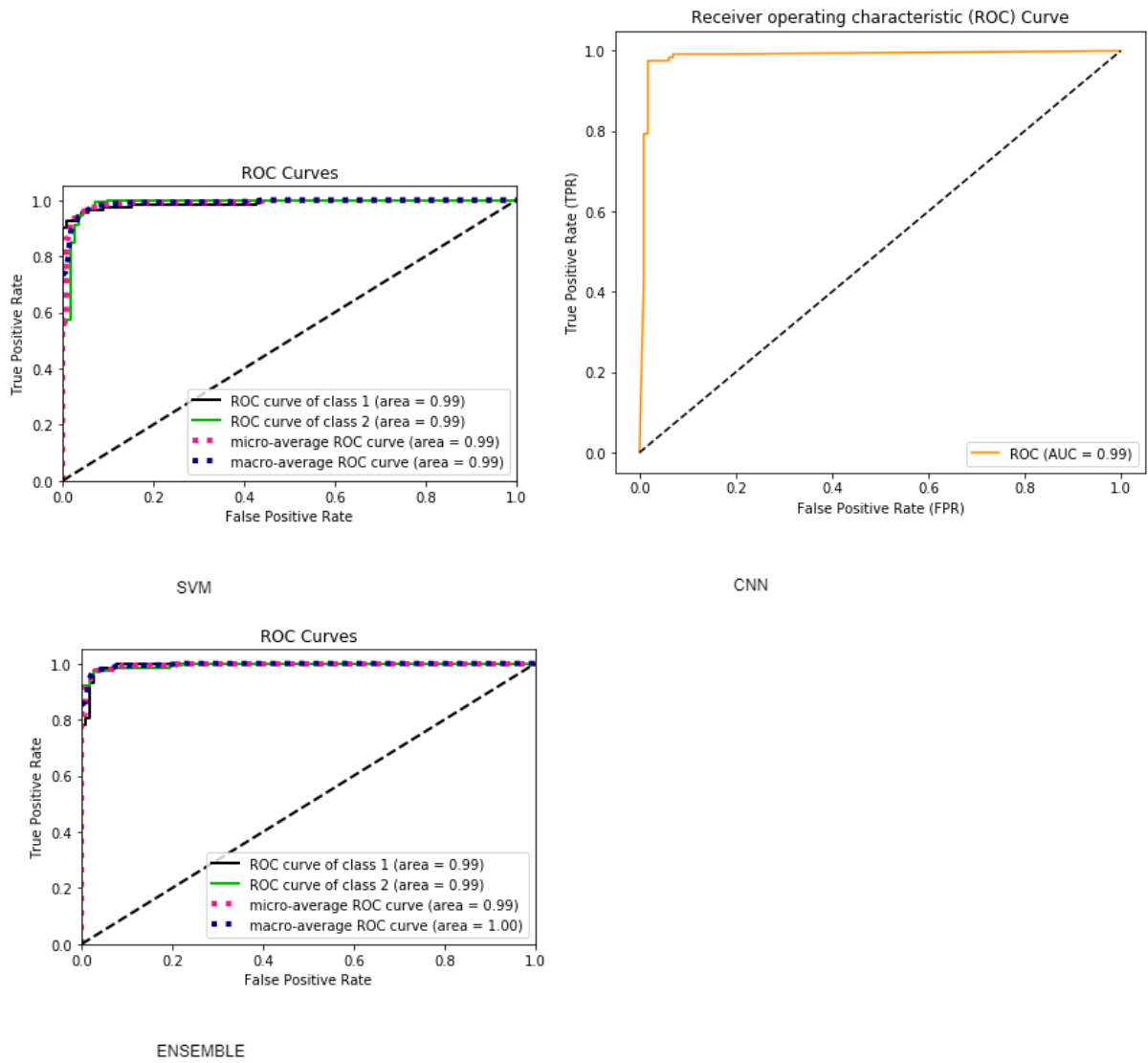


Figure 5.8: ROC curve on CK+ dataset.

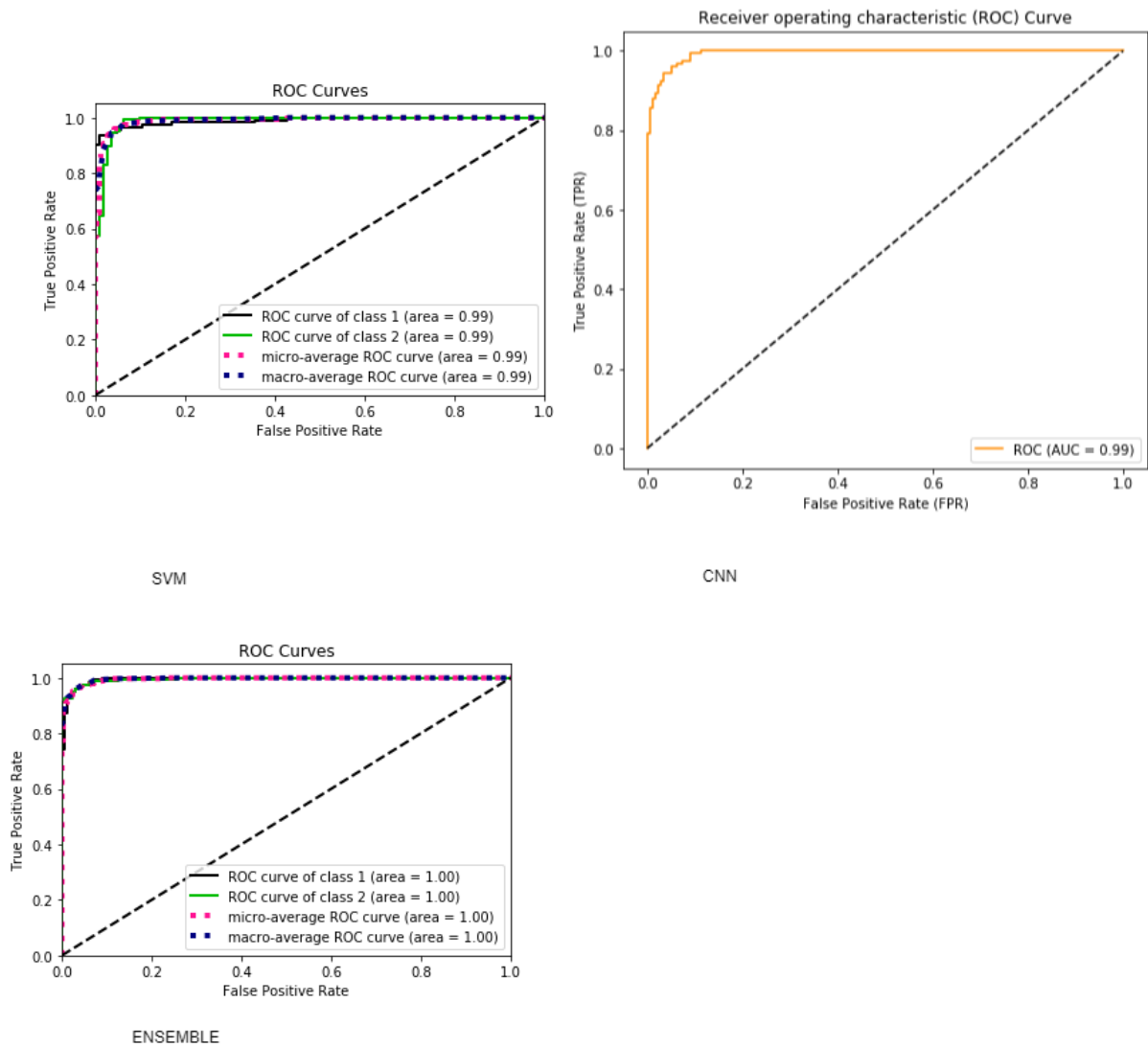


Figure 5.9: ROC curve on the KDEF dataset.

Figures 5.7, 5.8 and 5.9 display the Receiver Operating Characteristic (ROC) curves, comparing the outcome of the models. The displayed ROC curves were obtained after the application of the data augmentation technique.

As shown in figures 5.7, 5.8 and 5.9, the overall performance of ensemble learning model is higher than of SVM and CNN whilst SVM attains the overall lowest-performing model. Notwithstanding, all the classifiers produce excellent models, as there are all situated in the upper left corner of the ROC, the values are all above 0.5 (Shu et al., 2018).

## 5.5 Discussion

### 5.5.1 Introduction

The findings from this study indicate that the recognition of the emotions anger and not-angry where not-angry is the combination of the remaining emotions using facial expression as well as facial expression algorithms, results in higher accuracy.

This study revealed that, among the basic emotions introduced by Ekman, anger is the most frequently experienced yet poorly handled. And among the modes of expressing emotions, the facial expression is the most salient. However, the study conducted on the recognition of anger utilised speech data, physiological signals or general multiclassification of facial expressions that is all the facial expressions are been classified. However, we want to investigate how anger versus the other facial expressions will produce a higher accuracy. Therefore, facial expressions algorithms are evaluated, and the most utilised machine learning and deep learning algorithms are employed for our study. Also, we proposed an ensemble learning method for this study.

### 5.5.2 Performance of the models

Although different datasets were used, all the images were pre-processed to contain a unify input which strengthens the experimental results. The experimental results were further improved in experiment II with the application of the data augmentation technique to make our datasets balanced. Our experiments achieved remarkable results in comparison to the results of Tables 5.5-5.7. Generally, the best model for our experiment is the ensemble learning classifier. The high performance obtained can be attributed to the fusion of the descriptors HOG and LBP and the classifiers. The average accuracy obtained was 97% and the overall best classifier was the ensemble classifier with an accuracy of 100%. Further, it was observed the deep learning model performed well on the CK+ dataset and the machine learning algorithms on JAFFE dataset.

### 5.5.3 Comparison of the state-of-the-art

In this section, in evaluating the effectiveness of our model, we perform a comparison of our results with state-of-art results (Lai & Ko, 2014). It is not possible to straightforward compare the state-of-the-art experiments to our experiment because of the different experimental setting and protocols. Also, as per literature survey, existing work has not performed or reported results on any binary classification of facial expressions ( that is categorising of the 7 emotional states into angry and not-angry) on the contrary, on several multiclass classifications. Therefore, we focus our attention on the performance of these state-of-art facial expression algorithms in comparison to ours. The state-of-the-art experiments are discussed in this section and are indicated in Tables 5.5 - 5.7 (Alphonse & Dharma, 2017). Although the feature extractors differ, the classifiers are the same. It can be observed that our proposed methods obtained noticeable results, particularly our ensemble learning algorithm. We believe our ensemble learning model made a difference due to the fusion of HOG and LBP as well as SVM, KNN, RF, NB, and LG.

Our SVM classifier is outstandingly comparable to the expression classification accuracies in Table 5.5. The techniques used for building the SVM model is the same as the one in (Abdulrahman & Eleyan, 2015), however the results differ greatly. Also, diverse feature extraction techniques or descriptors were used such as LDA, LDN, HOG, DGLTP and Gabor, although the SVM classifier was utilised. Yet, the resulting expression recognition accuracies were not the best.

Follow on, the recognition accuracy of our CNN model is compared to the state-of-the-arts experiment in Table 5.6. It can be observed our proposed CNN model obtained a promising result in comparison to the others, even with the ensemble of CNN in (W. Sun et al., 2019).

Further, in table 5.7, our proposed novel ensemble model outperformed the state-of-the-arts ensemble methods.

Table 5.5: Comparison of approaches on the JAFFE dataset

Author	Technique	Classifier	Dataset	Accuracy (%)	Remarks
(Abdulrahman & Eleyan, 2015)	PCA/LBP	SVM	JAFFE	87	obtained an accuracy of 87% from an investigation of the performance of PCA+LBP and SVM on JAFFE.
(M. I. Revina & Emmanuel, 2018)	LDN/DGLTP	SVM	JAFFE	88.63	utilised SVM and LDN and DGLTP as classifier and feature extractors respectively on JAFFE dataset achieving an accuracy of 88.63%.
(Shah et al., 2017)	LDA	SVM	JAFFE	93.97	used threefold SVM and LDA to recognise the basic emotions on the JAFFE dataset attaining accuracy of 93.97%.

(Bellamkonda & Gopalan, 2019)	Kirsh+LBP	SVM	JAFFE	86	combined an edge detection algorithm called kirsh with LBP and SVM as the classifier. Their method obtained an accuracy of 86%.
<b>Proposed SVM model</b>	<b>LBP+PCA</b>	<b>SVM</b>	<b>JAFFE</b>	<b>97</b>	<b>Proposed LBP+PCA and SVM classifier on JAFFE, achieving an accuracy.</b>

Table 5.6: Comparison of the proposed method with state-of-the-art.

Author	Classifier	Dataset	Accuracy (%)
(Z. Li, 2018)	CNN	CK+	95.21
(W. Sun et al., 2019)	Ensemble of CNN	CK+	96.2
(Vo & Le, 2016)	CNN	CK+	92
<b>Ours</b>	<b>CNN</b>	<b>CK+</b>	<b>97</b>

Table 5.7: Comparison of different techniques on JAFFE+ dataset

Author	Technique	Classifier	Dataset	Accuracy (%)	Remarks
(T H H Zavaschi & Koerich, 2011)	Gabor+LBP	Ensemble of SVM	JAFFE/CK+	96.2	Used NSGA as a selection ranking method.
(Thiago H.H. Zavaschi, Britto, Oliveira, & Koerich, 2013)	LBP+Gabor	Ensemble of classifiers (MOGA)	JAFFE/CK+	96.2	Created a pool of classifiers and used MOGA to select the best ensemble model
<b>Our proposed ensemble learning model</b>	<b>LBP+HOG PCA (for dimension reduction)</b>	<b>An ensemble of KNN, SVM, NB, LG, and RF</b>	<b>JAFFE</b>	<b>100</b>	<b>Proposed a novel ensemble model for recognition of anger.</b>

### 5.6 Limitations

There are some limitations in terms of the number of angry facial expression images in the existing facial expression databases. Therefore, it makes it difficult in training an accurate model to recognise anger as observed in experiment I, where the recognition accuracies obtained were generally high whilst the precision and recall which gives the actual values, on the contrary, were low. This untrue accuracy representation can be attributed to the imbalance

nature of the datasets. As such, we resolved this issue by balancing the datasets using data augmentation (Bargshady et al., 2020).

Future work will look at creating a database of Africans, use these proposed algorithms to detect anger and compare their performance to the standard existing databases as well as employing these algorithms to detect anger in a persuasive space and persuade the individual from angry to another emotion for example happy.

### 5.7 Chapter summary

In this chapter, we looked at our experimental results. The experiment was conducted in two phases: with data augmentation and without data augmentation. It was observed that the accuracies were improved in the second experiment after the application of the data augmentation technique. Particularly, our proposed novel ensemble learning method outperformed the existing experiments in the literature. Thus, we conclude that the proposed methods are effective for the recognition of anger using facial expressions.

## Chapter 6

### Conclusion

Charles Darwin's influential work served as the premise for research in emotions. These emotions are recognised using indicators such as speech data, physiological signals and so on. Among the indicators for recognition of emotions, facial expression is a significant and leading measure as 55% of what we communicate is expressed in our facial expressions.

The current systems can detect facial expressions in general or a subset of emotions. To the best of our knowledge, there have not been any studies on how to detect only anger using facial expressions. Further, the multiclass classification of emotions has drawbacks such as the overlapping among the facial expressions which gives an untrue representation of the emotion when classified. We argued that anger detection needs to be done in an accurate way, giving a true representation of the emotion.

Therefore, in this research work, we propose a framework to perform binary classification of facial expressions for recognition of the emotions: angry and not angry and compare the outcome to the state-of-the-art experiments; having identified facial expression as the leading and significant measure for detecting emotions. We employed the most utilised algorithms from literature as well as propose a novel ensemble learning algorithm. The algorithms are SVM, CNN and a novel ensemble learning algorithm. The ensemble learning algorithm is a fusion of the feature sets HOG and LBP as well as a fusion of SVM, KNN, RF, NB, and LG. The experiment was conducted in two phases due to the imbalance nature of the dataset and the proposed methods were evaluated on JAFFE, KDEF and CK+ datasets. The SVM, CNN and ensemble models achieved accuracies of 97% on JAFFE dataset, 97% on CK+ dataset and 100% on JAFFE dataset, respectively. Our novel proposed an ensemble learning algorithm is the best performing model. Also, our models perform better than the state-of-art models.

Future work, we plan to create a database of Africans, use these proposed algorithms to detect anger and compare their performance to the standard existing databases as well as employing these algorithms to detect anger in a persuasive space and persuade the individual from angry to another emotion for example happy.

## Bibliography

- Abd El Meguid, M. K., & Levine, M. D. (2014). Fully automated recognition of spontaneous facial expressions in videos using random forest classifiers. *IEEE Transactions on Affective Computing*, 5(2), 141–154. <https://doi.org/10.1109/TAFFC.2014.2317711>
- Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4), 433–459. <https://doi.org/10.1002/wics.101>
- Abdulrahman, M., & Eleyan, A. (2015). Facial expression recognition using Support Vector Machines. *2015 23rd Signal Processing and Communications Applications Conference, SIU 2015 - Proceedings*, 276–279. <https://doi.org/10.1109/SIU.2015.7129813>
- Abhang, P. A., Gawali, B. W., & Mehrotra, S. C. (2016). Multimodal Emotion Recognition. *Introduction to EEG- and Speech-Based Emotion Recognition*, 113–125. <https://doi.org/10.1016/b978-0-12-804490-2.00006-3>
- Abouyahya, A., El Fkihi, S., Thami, R. O. H., & Aboutajdine, D. (2016). Features extraction for facial expressions recognition. *International Conference on Multimedia Computing and Systems -Proceedings*, 16, 46–49. <https://doi.org/10.1109/ICMCS.2016.7905642>
- Abouyahya, A., & Fkihi, S. El. (2018). An optimization of the k-nearest neighbor using dynamic time warping as a measurement similarity for facial expressions recognition. *ACM International Conference Proceeding Series*, 1–5. <https://doi.org/10.1145/3230905.3230921>
- Agarwal, A., Baechle, C., Behara, R. S., & Rao, V. (2016). Multi-method approach to wellness predictive modeling. *Journal of Big Data*, 3(1), 1–23. <https://doi.org/10.1186/s40537-016-0049-0>
- Ahmed, F., Bari, A. S. M. H., & Gavrilova, M. L. (2020). Emotion Recognition from Body Movement. *IEEE Access*, 8, 11761–11781.

<https://doi.org/10.1109/ACCESS.2019.2963113>

Ali, G., Iqbal, M. A., & Choi, T. S. (2016). Boosted NNE collections for multicultural facial expression recognition. *Pattern Recognition*, 55, 14–27.

<https://doi.org/10.1016/j.patcog.2016.01.032>

Alizadeh, S., & Fazel, A. (2017). Convolutional Neural Networks for Facial Expression Recognition. Retrieved from <http://arxiv.org/abs/1704.06756>

Alphonse, A. S., & Dharma, D. (2017). Enhanced Gabor (E-Gabor), Hypersphere-based normalization and Pearson General Kernel-based discriminant analysis for dimension reduction and classification of facial emotions. *Expert Systems with Applications*, 90, 127–145. <https://doi.org/10.1016/j.eswa.2017.08.013>

An, S., Ji, L. J., Marks, M., & Zhang, Z. (2017). Two sides of emotion: Exploring positivity and negativity in six basic emotions across cultures. *Frontiers in Psychology*, 8(APR), 1–14. <https://doi.org/10.3389/fpsyg.2017.00610>

Apte, A., Basavaraj, A., & Nithin, R. K. (2016). Efficient Facial Expression Ecognition and classification system based on morphological processing of frontal face images. *2015 IEEE 10th International Conference on Industrial and Information Systems, ICIIS 2015 - Conference Proceedings*, 366–371. <https://doi.org/10.1109/ICIINFS.2015.7399039>

Bargshady, G., Zhou, X., Deo, R. C., Soar, J., Whittaker, F., & Wang, H. (2020). Enhanced deep learning algorithm development to detect pain intensity from facial expression images. *Expert Systems with Applications*, 149, 113305. <https://doi.org/10.1016/j.eswa.2020.113305>

Bellamkonda, S., & Gopalan, N. P. (2019). Facial Expression Recognition Using Kirsch Edge Detection, LBP and Gabor Wavelets. *Proceedings of the 2nd International Conference on Intelligent Computing and Control Systems, ICICCS 2018, (Iciccs)*, 1457–1461.

<https://doi.org/10.1109/ICCONS.2018.8662971>

- Bhardwaj, N., & Dixit, M. (2016). A Review: Facial Expression Detection with its Techniques and Application. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 9(6), 149–158. <https://doi.org/10.14257/ijcip.2016.9.6.13>
- Borui, Z., Liu, G., & Xie, G. (2017). Facial expression recognition using LBP and LPQ based on Gabor wavelet transform. *2016 2nd IEEE International Conference on Computer and Communications, ICCC 2016 - Proceedings*, 365–369. <https://doi.org/10.1109/CompComm.2016.7924724>
- Breiman, L. E. O. (2001). Random Forest (LeoBreiman) .pdf, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Breuer, R., & Kimmel, R. (2017a). A Deep Learning Perspective on the Origin of Facial Expressions, 1–16. Retrieved from <http://arxiv.org/abs/1705.01842>
- Breuer, R., & Kimmel, R. (2017b). *A Deep Learning Perspective on the Origin of Facial Expressions*. Israel Institute of Technology. Retrieved from <http://arxiv.org/abs/1705.01842>
- Busso, C., Deng, Z., Yildirim, S., & Bulut, M. (2004). Analysis of Emotion Recognition using Facial Expressions, Speech and Multimodal Information. *Icmi*, 205–211. <https://doi.org/10.1145/1027933.1027968>
- Calder, A. J., Burton, A. M., Miller, P., Young, A. W., & Akamatsu, S. (2001). A principal component analysis of facial expressions. *Vision Research*, 41(9), 1179–1208. [https://doi.org/10.1016/S0042-6989\(01\)00002-5](https://doi.org/10.1016/S0042-6989(01)00002-5)
- Candra, H., Yuwono, M., Chai, R., Nguyen, H. T., & Su, S. (2016). Classification of facial-emotion expression in the application of psychotherapy using Viola-Jones and Edge-Histogram of Oriented Gradient. *Proceedings of the Annual International Conference of*

- the IEEE Engineering in Medicine and Biology Society, EMBS, 2016-Octob(Di)*, 423–426. <https://doi.org/10.1109/EMBC.2016.7590730>
- Chakladar, D. Das, & Chakraborty, S. (2018). EEG based emotion classification using “correlation Based Subset Selection.” *Biologically Inspired Cognitive Architectures*. <https://doi.org/10.1016/j.bica.2018.04.012>
- Chang, C. Y., Lin, Y. M., & Zheng, J. Y. (2012). Physiological angry emotion detection using support vector regression. *Proceedings of the 2012 15th International Conference on Network-Based Information Systems, NBIS 2012*, 592–596. <https://doi.org/10.1109/NBiS.2012.78>
- Chaparro, V., Gomez, A., Salgado, A., Quintero, O. L., Lopez, N., & Villa, L. F. (2018). Emotion Recognition from EEG and Facial Expressions: A Multimodal Approach. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2018-July*, 530–533. <https://doi.org/10.1109/EMBC.2018.8512407>
- Chhabra, P., Vyas, G., Chatterjee, J., & Vob, S. H. (2017). An automatic system for recognition and assessment of anger using adaptive boost. *Proceedings - 2016 International Conference on Micro-Electronics and Telecommunication Engineering, ICMETE 2016*, 151–154. <https://doi.org/10.1109/ICMETE.2016.89>
- Culjak, I., Abram, D., Pribanic, T., Dzapo, H., & Cifrek, M. (2012). A brief introduction to OpenCV. *MIPRO 2012 - 35th International Convention on Information and Communication Technology, Electronics and Microelectronics - Proceedings*, 1725–1730.
- Dagher, I., Dahdah, E., & Al Shakik, M. (2019). Facial expression recognition using three-stage support vector machines. *Visual Computing for Industry, Biomedicine, and Art*, 2(1),

0–8. <https://doi.org/10.1186/s42492-019-0034-5>

- Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, I*, 886–893. <https://doi.org/10.1109/CVPR.2005.177>
- Darwin, C. (1872). The expression of the emotions in man and animals. *The Expression of the Emotions in Man and Animals*. <https://doi.org/10.1037/10001-000>
- Deng, J., Eyben, F., Schuller, B., & Burkhardt, F. (2018). Deep neural networks for anger detection from real life speech data. *2017 7th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos, ACIIW 2017, 2018-Janua*, 1–6. <https://doi.org/10.1109/ACIIW.2017.8272614>
- Dev, V. A., & Eden, M. R. (2019). *Gradient Boosted Decision Trees for Lithology Classification. Computer Aided Chemical Engineering* (Vol. 47). Elsevier Masson SAS. <https://doi.org/10.1016/B978-0-12-818597-1.50019-9>
- Dhall, S., & Sethi, P. (2014). Geometric and Appearance Feature Analysis for Facial Expression Recognition. *International Journal of Advanced Engineering Technology*, 5(3), 1–11.
- Dino, H. I., & Abdulrazzaq, M. B. (2019). Facial Expression Classification Based on SVM, KNN and MLP Classifiers. *2019 International Conference on Advanced Science and Engineering, ICOASE 2019*, 70–75. <https://doi.org/10.1109/ICOASE.2019.8723728>
- Domínguez-Jiménez, J. A., Campo-Landines, K. C., Martínez-Santos, J. C., Delahoz, E. J., & Contreras-Ortiz, S. H. (2020). A machine learning model for emotion recognition from physiological signals. *Biomedical Signal Processing and Control*, 55, 101646. <https://doi.org/10.1016/j.bspc.2019.101646>
- Dubey, S., & Dixit, M. (2019). Facial expression recognition using deep convolutional neural

networks. *Computing Publications*, 8(1), 130–135.  
<https://doi.org/10.1109/KSE.2017.8119447>

Duchenne de Boulogne, G. . (1862). *The Mechanism of Human Facial Expression*. Cambridge University Press. <https://doi.org/10.1097/00006534-199203000-00032>

Ekman, P. (1970). *Universal-Facial-Expressions-of-Emotions*. California Mental health.

Ekman, P. (1977). *Facial Expression*, (1972), 97–116.

Ekman, P. (1999). *Basic Emotions*. *Encyclopedia of Personality and Individual Differences*.  
[https://doi.org/10.1007/978-3-319-28099-8\\_495-1](https://doi.org/10.1007/978-3-319-28099-8_495-1)

Ekman, P., & Friesen, W. (1976). *Mesasuring facial movement*.pdf.

Ekundayo, O., & Viriri, S. (2019). *Facial expression recognition: A review of methods, performances and limitations*. *2019 Conference on Information Communications Technology and Society, ICTAS 2019*. <https://doi.org/10.1109/ICTAS.2019.8703619>

Fan, X., & Tjahjadi, T. (2019). *Fusing dynamic deep learned features and handcrafted features for facial expression recognition*. *Journal of Visual Communication and Image Representation*, 65, 1–6. <https://doi.org/10.1016/j.jvcir.2019.102659>

Farajzadeh, N., & Hashemzadeh, M. (2018). *Exemplar-based facial expression recognition*. *Information Sciences*, 460–461, 318–330. <https://doi.org/10.1016/j.ins.2018.05.057>

Fasel, B., Monay, F., & Gatica-Perez, D. (2004). *Latent semantic analysis of facial action codes for automatic facial expression recognition*. *MIR'04 - Proceedings of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval*, 181–188.  
<https://doi.org/10.1145/1026711.1026742>

Feidakis, M. (2016). *A Review of Emotion-Aware Systems for e-Learning in Virtual Environments*. *Formative Assessment, Learning Data Analytics and Gamification: In ICT*

*Education*. Elsevier Inc. <https://doi.org/10.1016/B978-0-12-803637-2.00011-7>

Feldman, M. (2018). Google Offers Glimpse of Third-Generation TPU Processor. Retrieved from <https://www.top500.org/news/google-offers-glimpse-of-third-generation-tpu-processor/>

Frank, M. G. (2001). Facial Expression. *International Encyclopedia of the Social & Behavioral Sciences*, 5230–5234. <https://doi.org/10.1016/B0-08-043076-7/01713-7>

Gao, K., Mei, G., Piccialli, F., Cuomo, S., Tu, J., & Huo, Z. (2020). Julia language in machine learning: Algorithms, applications, and open issues. *Computer Science Review*, 37, 100254. <https://doi.org/10.1016/j.cosrev.2020.100254>

Goeleven, E., De Raedt, R., Leyman, L., & Verschuere, B. (2008). The Karolinska directed emotional faces: A validation study. *Cognition and Emotion*, 22(6), 1094–1118. <https://doi.org/10.1080/02699930701626582>

Gonzalez-sanchez, J., Baydogan, M., Chavez-echeagaray, M. E., Robert, K., & Burleson, W. (2017). *Affect Measurement : A Roadmap Through Approaches, Technologies and Data Analysis. Emotions and Affect in Human Factors and Human-Computer Interaction*. Elsevier Inc. <https://doi.org/10.1016/B978-0-12-801851-4/00011-2>

Happy, S. L., Patnaik, P., Routray, A., & Guha, R. (2017). The Indian Spontaneous Expression Database for Emotion Recognition. *IEEE Transactions on Affective Computing*, 8(1), 131–142. <https://doi.org/10.1109/TAFFC.2015.2498174>

Haq, S., & Jackson, P. (2010). Machine Audition: Principles, Algorithms and Systems, chapter 8. *Multimodal Emotion Recognition*, 398–423. Retrieved from <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Machine+Audition+:+Principles+,+Algorithms+and+Systems#1>

Hossain, M. M. (2018). *Facial Expression Recognition Based on LBP and CNN: A*

- Comparative Study Using SVM Classifier*. Rajshahi University of Engineering and Technology.
- Huang, H., Hu, Z., Wang, W., & Wu, M. (2020). Multimodal Emotion Recognition Based on Ensemble Convolutional Neural Network. *IEEE Access*, 8(2), 3265–3271. <https://doi.org/10.1109/ACCESS.2019.2962085>
- Huang, X. (2014). *Methods for Facial Expression Recognition With Applications in Challenging Situations*. *Emotion Review*. University of Oulu, Finland. Retrieved from <http://emr.sagepub.com/content/6/2/113.short>
- Huang, X., Kortelainen, J., Zhao, G., Li, X., Moilanen, A., Seppänen, T., & Pietikäinen, M. (2015). Multi-modal emotion analysis from facial expressions and electroencephalogram. *Computer Vision and Image Understanding*, 147, 114–124. <https://doi.org/10.1016/j.cviu.2015.09.015>
- Huang, Yongrui, Yang, J., Liao, P., & Pan, J. (2017). Fusion of Facial Expressions and EEG for Multimodal Emotion Recognition. *Computational Intelligence and Neuroscience*, 2017, 1–8. <https://doi.org/10.1155/2017/2107451>
- Huang, Yunxin, Chen, F., Lv, S., & Wang, X. (2019). Facial expression recognition: A survey. *Symmetry*, 11(10), 1–28. <https://doi.org/10.3390/sym11101189>
- Islam, B., Mahmud, F., Hossain, A., Mia, M. S., & Goala, P. B. (2019). Human facial expression recognition system using artificial neural network classification of gabor feature based facial expression information. *4th International Conference on Electrical Engineering and Information and Communication Technology, ICEEICT 2018*, 364–368. <https://doi.org/10.1109/CEEICT.2018.8628050>
- Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas. *Association for Psychological Science*, 2(3), 260–280.

- J., W., & Watkins, C. (1999). Support Vector Machines for Multi-Class Pattern Recognition. *ESANN*, 219–224.
- Jain, D. K., Shamsolmoali, P., & Sehdev, P. (2019). Extended deep neural network for facial emotion recognition. *Pattern Recognition Letters*, 120, 69–74. <https://doi.org/10.1016/j.patrec.2019.01.008>
- Jain, V., Lamba, P. S., Singh, B., Namboothiri, N., & Dhall, S. (2019). Facial expression recognition using feature level fusion. *Journal of Discrete Mathematical Sciences and Cryptography*, 22(2), 337–350. <https://doi.org/10.1080/09720529.2019.1582866>
- Jakkula, V. (2011). Tutorial on Support Vector Machine (SVM). *School of EECS, Washington State University*, 1–13. Retrieved from <http://www.ccs.neu.edu/course/cs5100f11/resources/jakkula.pdf>
- Jameel, R., Singhal, A., & Bansal, A. (2016). A comprehensive study on Facial Expressions Recognition Techniques. *Proceedings of the 2016 6th International Conference - Cloud System and Big Data Engineering, Confluence 2016*, 478–483. <https://doi.org/10.1109/CONFLUENCE.2016.7508167>
- Jan, A. (2017). *Deep Learning Based Facial Expression Recognition and Its Applications*. Brunel University, London.
- Kanade, T., Cohn, J. F., & Tian, Y. (2000). Comprehensive database for facial expression analysis. *Proceedings - 4th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2000*, 46–53. <https://doi.org/10.1109/AFGR.2000.840611>
- Karamizadeh, S., Abdullah, S. M., Manaf, A. A., Zamani, M., & Hooman, A. (2013). An Overview of Principal Component Analysis. *Journal of Signal and Information Processing*, 04(03), 173–175. <https://doi.org/10.4236/jsip.2013.43b031>
- Kassinove, H., Sukhodolsky, D. G., Eckhardt, C. I., & Tsytsarev, S. V. (1997). Development

- of a Russian State-Trait anger Expression Inventory. *Journal of Clinical Psychology*, 53(6), 543–557. [https://doi.org/10.1002/\(SICI\)1097-4679\(199710\)53:6<543::AID-JCLP3>3.0.CO;2-L](https://doi.org/10.1002/(SICI)1097-4679(199710)53:6<543::AID-JCLP3>3.0.CO;2-L)
- Kaur, B., Singh, D., & Roy, P. P. (2018). EEG Based Emotion Classification Mechanism in BCI. In *Procedia Computer Science* (pp. 752–758). <https://doi.org/10.1016/j.procs.2018.05.087>
- Keshari, T., & Palaniswamy, S. (2019). Emotion Recognition Using Feature-level Fusion of Facial Expressions and Body Gestures. *Proceedings of the 4th International Conference on Communication and Electronics Systems, ICCES 2019*, (Icces), 1184–1189. <https://doi.org/10.1109/ICCES45898.2019.9002175>
- Khalid, S., Khalil, T., & Nasreen, S. (2014). A survey of feature selection and feature extraction techniques in machine learning. *Proceedings of 2014 Science and Information Conference, SAI 2014*, 372–378. <https://doi.org/10.1109/SAI.2014.6918213>
- Kiran, T., & Kushal, T. (2016). Facial expression classification using Support Vector Machine based on bidirectional Local Binary Pattern Histogram feature descriptor. *2016 IEEE/ACIS 17th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, SNPD 2016*, 115–120. <https://doi.org/10.1109/SNPD.2016.7515888>
- Kirange, D. K., & Deshmukh, R. R. (2012). EMOTION CLASSIFICATION OF NEWS HEADLINES USING SVM, 5, 104–106.
- Koehrsen Will. (2018). Beyond Accuracy: Precision and Recall - Towards Data Science. *Media.Com*, 19, 1–4. Retrieved from <https://towardsdatascience.com/beyond-accuracy-precision-and-recall-3da06bea9f6c>
- Kotu, V., & Deshpande, B. (2015). Data Mining Process. *Predictive Analytics and Data*

*Mining*, (1), 17–36. <https://doi.org/10.1016/b978-0-12-801460-8.00002-1>

Kudiri, K. M., Said, A. M., & Nayan, M. Y. (2013). Emotion detection using relative grid based coefficients through human facial expressions. *International Conference on Research and Innovation in Information Systems, ICRIIS, 2013*, 45–48. <https://doi.org/10.1109/ICRIIS.2013.6716683>

Kudiri, K. M., Said, A. M., & Nayan, M. Y. (2016). Human emotion detection through speech and facial expressions. *2016 3rd International Conference on Computer and Information Sciences, ICCOINS 2016 - Proceedings*, 351–356. <https://doi.org/10.1109/ICCOINS.2016.7783240>

Kumar, G. A. R., Kumar, R. K., & Sanyal, G. (2018). Facial emotion analysis using deep convolution neural network. *Proceedings of IEEE International Conference on Signal Processing and Communication, ICSPC 2017, 2018-Janua(July)*, 369–374. <https://doi.org/10.1109/CSPC.2017.8305872>

Kwong, J. C. T., Garcia, F. C. C., Abu, P. A. R., & Reyes, R. S. J. (2019). Emotion Recognition via Facial Expression: Utilization of Numerous Feature Descriptors in Different Machine Learning Algorithms. *IEEE Region 10 Annual International Conference, Proceedings/TENCON, 2018-Octob(October)*, 2045–2049. <https://doi.org/10.1109/TENCON.2018.8650192>

Ladha, L., & Deepa, T. (2011). Feature Selection Methods And Algorithms. *International Journal on Computer Science and Engineering*, 3(5), 1787–1797. Retrieved from <http://journals.indexcopernicus.com/abstract.php?icid=945099>

Lai, C. C., & Ko, C. H. (2014). Facial expression recognition based on two-stage features extraction. *Optik*, 125(22), 6678–6680. <https://doi.org/10.1016/j.ijleo.2014.08.052>

Lalitha, S., Geyasruti, D., Narayanan, R., & Shravani, M. (2015). Emotion Detection using

- MFCC and Cepstrum Features, *70*, 29–35. <https://doi.org/10.1016/j.procs.2015.10.020>
- Lecun, Y., Bottou, L., Bengio, Y., & Ha, P. (1998). Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, (November), 1–46. <https://doi.org/10.1109/5.726791>
- Li, S., & Deng, W. (2018). Deep Facial Expression Recognition: A Survey. *IEEE*, 1–25. Retrieved from <http://arxiv.org/abs/1804.08348>
- Li, Z. (2018). A discriminative learning convolutional neural network for facial expression recognition. *2017 3rd IEEE International Conference on Computer and Communications, ICC 2017, 2018-Janua*, 1641–1646. <https://doi.org/10.1109/CompComm.2017.8322818>
- Littlewort, G. C., Bartlett, M. S., & Lee, K. (2007). Faces of pain: Automated measurement of spontaneous facial expressions of genuine and posed pain. *Proceedings of the 9th International Conference on Multimodal Interfaces, ICMI'07*, (May 2014), 15–21. <https://doi.org/10.1145/1322192.1322198>
- Liu, L., Fieguth, P., Guo, Y., Wang, X., & Pietikäinen, M. (2017). Local binary features for texture classification: Taxonomy and experimental study. *Pattern Recognition*, *62*, 135–160. <https://doi.org/10.1016/j.patcog.2016.08.032>
- Liu, Y., Yuan, X., Gong, X., Xie, Z., Fang, F., & Luo, Z. (2018). Conditional convolution neural network enhanced random forest for facial expression recognition. *Pattern Recognition*. <https://doi.org/10.1016/j.patcog.2018.07.016>
- Lopes, A. T., de Aguiar, E., De Souza, A. F., & Oliveira-Santos, T. (2017). Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order. *Pattern Recognition*, *61*, 610–628. <https://doi.org/10.1016/j.patcog.2016.07.026>

- Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, CVPRW 2010*, (July), 94–101. <https://doi.org/10.1109/CVPRW.2010.5543262>
- Lyons, M., Akamatsu, S., Kamachi, M., & Gyoba, J. (1998). Coding facial expressions with Gabor wavelets. *Proceedings - 3rd IEEE International Conference on Automatic Face and Gesture Recognition, FG 1998*, 200–205. <https://doi.org/10.1109/AFGR.1998.670949>
- Mahmood, A., Hussain, S., Iqbal, K., & Elkilani, W. S. (2019). Recognition of Facial Expressions under Varying Conditions Using Dual-Feature Fusion. *Mathematical Problems in Engineering, 2019*, 1–12. <https://doi.org/10.1155/2019/9185481>
- Mangalagowri, S. G., & Raj, P. C. P. (2017). EEG feature extraction and classification using feed forward backpropagation algorithm for emotion detection. In *2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques, ICEECCOT 2016* (pp. 183–187). IEEE. <https://doi.org/10.1109/ICEECCOT.2016.7955211>
- Martinez, B., & Valster, M. (2016). *Advances in face detection and facial image analysis. Advances, Challenges, and Opportunities in Automatic Facial Expression Recognition*. <https://doi.org/10.1007/978-3-319-25958-1>
- Matlovic, T., Gaspar, P., Moro, R., Simko, J., & Bielikova, M. (2016). Emotions detection using facial expressions recognition and EEG. *Proceedings - 11th International Workshop on Semantic and Social Media Adaptation and Personalization, SMAP 2016*, 18–23. <https://doi.org/10.1109/SMAP.2016.7753378>

- Mayya, V., Pai, R. M., & Manohara Pai, M. M. (2016). Automatic Facial Expression Recognition Using DCNN. *Procedia Computer Science*, 93, 453–461. <https://doi.org/10.1016/j.procs.2016.07.233>
- Mehrabian, A. (1968). Communication Studies. *Institute of Judicial Studies*.
- Michel, P., & El Kaliouby, R. (2015). Facial expression recognition using Support Vector Machines. *2015 23rd Signal Processing and Communications Applications Conference, SIU 2015 - Proceedings*, 276–279. <https://doi.org/10.1109/SIU.2015.7129813>
- Minaee, S., & Abdolrashidi, A. (2019). Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network. Retrieved from <http://arxiv.org/abs/1902.01019>
- Mitsuyoshi, S., & Ren, F. (2013). Emotion Recognition. *The Journal of The Institute of Electrical Engineers of Japan*, 125(10), 641–644. <https://doi.org/10.1541/ieejjournal.125.641>
- Mohammadpour, M., Khaliliardali, H., Hashemi, S. M. R., & Alyannezhadi, M. M. (2018). Facial emotion recognition using deep convolutional networks. In *2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation, KBEI 2017* (Vol. 2018-Janua, pp. 0017–0021). <https://doi.org/10.1109/KBEI.2017.8324974>
- Moritz, D. A. (2006). Understanding Anger. *The American Journal of Nursing*, 78(1), 81. <https://doi.org/10.2307/3424476>
- Nassih, B., Amine, A., Ngadi, M., & Hmina, N. (2019). DCT and HOG Feature Sets Combined with BPNN for Efficient Face Classification. *Procedia Computer Science*, 148, 116–125. <https://doi.org/10.1016/j.procs.2019.01.015>
- Nguyen, D. H., Kim, S. H., Lee, G. S., Yang, H. J., Na, I. S., & Kim, S. H. (2019). Facial Expression Recognition Using a Temporal Ensemble of Multi-level Convolutional Neural Networks. *IEEE Transactions on Affective Computing*, 3045, 1–12.

<https://doi.org/10.1109/TAFFC.2019.2946540>

Nguyen, D., Nguyen, K., Sridharan, S., Dean, D., & Fookes, C. (2018). Deep spatio-temporal feature fusion with compact bilinear pooling for multimodal emotion recognition. *Computer Vision and Image Understanding*, 174(July), 33–42. <https://doi.org/10.1016/j.cviu.2018.06.005>

Nisbet, R., Miner, G., & Yale, K. (2018). Model Evaluation and Enhancement. *Handbook of Statistical Analysis and Data Mining Applications*, 215–233. <https://doi.org/10.1016/b978-0-12-416632-5.00011-6>

Niu, Y., Zhao, Y., & Ni, R. R. (2017). Robust median filtering detection based on local difference descriptor. *Signal Processing: Image Communication*, 53(November 2016), 65–72. <https://doi.org/10.1016/j.image.2017.01.008>

Nugrahaeni, R. A., & Mutijarsa, K. (2017). Comparative analysis of machine learning KNN, SVM, and random forests algorithm for facial expression classification. In *Proceedings - 2016 International Seminar on Application of Technology for Information and Communication, ISEMANTIC 2016* (pp. 163–168). IEEE. <https://doi.org/10.1109/ISEMANTIC.2016.7873831>

Ojala, T., Pietikäinen, M., & Harwood, D. (1996). A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, 29(1), 51–59. [https://doi.org/10.1016/0031-3203\(95\)00067-4](https://doi.org/10.1016/0031-3203(95)00067-4)

Ojala, T., Pietikäinen, M., & Mäenpää, T. (2002). Multiresolution gray scale and rotation invariant texture classification with local binary patterns. *Pattern Recognition*, 1842, 404–420. [https://doi.org/10.1007/3-540-45054-8\\_27](https://doi.org/10.1007/3-540-45054-8_27)

Oliphant, T. E. (2007). Python for scientific computing. *Computing in Science and Engineering*, 9(3), 10–20. <https://doi.org/10.1109/MCSE.2007.58>

- Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review*, 97(3), 315–331. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=mnh&AN=1669960&site=ehost-live>
- Pantic, M., & Bartlett, M. S. (2007). *Machine Analysis of Facial Expressions*. *Intech* (Vol. 5). <https://doi.org/http://dx.doi.org/10.5772/57353>
- Pantic, M., Pentland, A., Nijholt, A., & Hunag, T. S. (2007). Human Computing and Machine Understanding of Human Behaviour. *Springer-Verlag : Human COmputing*, 9359(4), 47–71. Retrieved from <https://ibug.doc.ic.ac.uk/media/uploads/documents/LNAI-PanticEtAl-CAMERA.pdf>
- Pantic, M., & Rothkrantz, L. Ü. M. (2000). Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12), 1424–1445. <https://doi.org/10.1109/34.895976>
- Patil, A., & Behele, K. (2018). Classification of Human Emotions Using Multiclass Support Vector Machine. *2017 International Conference on Computing, Communication, Control and Automation, ICCUBEA 2017*, 1–4. <https://doi.org/10.1109/ICCUBEA.2017.8463656>
- Pell, P. J., & Richards, A. (2011). Cross-emotion facial expression aftereffects. *Vision Research*, 51(17), 1889–1896. <https://doi.org/10.1016/j.visres.2011.06.017>
- Pérez, F., Granger, B. E., & Hunter, J. D. (2011). Python: An ecosystem for scientific computing. *Computing in Science and Engineering*, 13(2), 13–21. <https://doi.org/10.1109/MCSE.2010.119>
- Petrushin, V. A. (2000). Emotion recognition in speech signal: Experimental study, development, and application. *6th International Conference on Spoken Language Processing, ICSLP 2000*, (Icslp).

- Plutchik, R. (1987). The nature of emotions. *Philosophical Studies*, 52(3), 393–409.  
<https://doi.org/10.1007/BF00354055>
- Pons, G., & Masip, D. (2018). Supervised Committee of Convolutional Neural Networks in Automated Facial Expression Analysis. *IEEE Transactions on Affective Computing*, 9(3), 343–350. <https://doi.org/10.1109/TAFFC.2017.2753235>
- Prilianti, K. R., Brotosudarmo, T. H. P., Anam, S., & Suryanto, A. (2019). Performance comparison of the convolutional neural network optimizer for photosynthetic pigments prediction on plant digital image. *AIP Conference Proceedings*, 2084, 1–9.  
<https://doi.org/10.1063/1.5094284>
- Qayyum, H., Majid, M., Anwar, S. M., & Khan, B. (2017). Facial Expression Recognition Using Stationary Wavelet Transform Features. *Mathematical Problems in Engineering*, 2017(1). <https://doi.org/10.1155/2017/9854050>
- Rajvanshi, K. (2018). An Efficient Approach for Emotion Detection from Speech Using Neural Networks. *International Journal for Research in Applied Science and Engineering Technology*, 6(5), 1062–1065. <https://doi.org/10.22214/ijraset.2018.5170>
- Raksarikorn, T., & Kangkachit, T. (2018). Facial Expression Classification using Deep Extreme Inception Networks. *Proceeding of 2018 15th International Joint Conference on Computer Science and Software Engineering, JCSSE 2018*, 1–5.  
<https://doi.org/10.1109/JCSSE.2018.8457396>
- Ramesh, S. (2018). A guide to an efficient way to build neural network architectures- Part I: Hyper-parameter selection and tuning for Dense Networks using Hyperas on Fashion-MNIST. Retrieved June 14, 2020, from <https://towardsdatascience.com/a-guide-to-an-efficient-way-to-build-neural-network-architectures-part-i-hyper-parameter-8129009f131b>

- Rani, J., & Garg, K. (2014). An Interface for Extracting and Cropping of Face from Video Frames. *International Journal of Computer Science and Information Technologies*, 5(3), 4394–4397.
- Rashid, M. I., Hasan, M., Yeasmin, N., Shahnaz, C., Fattah, S. A., Zhu, W. P., & Ahmed, M. O. (2017). Emotion recognition based on vertical cross correlation sequence of facial expression images. *Midwest Symposium on Circuits and Systems, 2017-Augus*, 1344–1347. <https://doi.org/10.1109/MWSCAS.2017.8053180>
- Revina, I. Michael, & Emmanuel, W. R. S. (2018). Facial Expression Recognition via Modified GAD Features with PSO-KNN. *2018 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 4(Icssit), 145–149. <https://doi.org/10.1109/icssit.2018.8748697>
- Revina, I. M., & Emmanuel, W. R. . (2018). A Survey on Human Face Expression Recognition Techniques. *Journal of King Saud University - Computer and Information Sciences*, 1–10. <https://doi.org/10.1016/j.jksuci.2018.09.002>
- Revina, M. I., & Emmanuel, S. W. R. (2018). Face expression recognition using LDN and Dominant Gradient Local Ternary Pattern descriptors. *Journal of King Saud University - Computer and Information Sciences*, 1319–1578. <https://doi.org/10.1016/j.jksuci.2018.03.015>
- Rish, I., Hellerstein, J., & Jayram, T. (2001). An analysis of data characteristics that affect naive Bayes performance. *IBM TJ Watson Research Center*, 30, 1–8. <https://doi.org/10.1.1.138.672>
- Rizwan, A. K. (2013). *Detection of emotions from video in non-controlled environment*. Université Claude Bernard - Lyon I.
- Russell, J. A., & Pratt, G. (1980). A description of the affective quality attributed to

- environments. *Journal of Personality and Social Psychology*, 38(2), 311–322.  
<https://doi.org/10.1037//0022-3514.38.2.311>
- Saeed, A., Al-Hamadi, A., Niese, R., & Elzobi, M. (2014). Frame-based facial expression recognition using geometrical features. *Advances in Human-Computer Interaction*, 2014.  
<https://doi.org/10.1155/2014/408953>
- Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), 1–18. <https://doi.org/10.1002/widm.1249>
- Shah, J. H., Sharif, M., Yasmin, M., & Fernandes, S. L. (2017). Facial expressions classification and false label reduction using LDA and threefold SVM. *Pattern Recognition Letters*, 0, 1–8. <https://doi.org/10.1016/j.patrec.2017.06.021>
- Shahsavarani, A. M., Noohi, S., Jafari, S., Kalkhoran, M. H., & Hatefi, S. (2015). Assessment & Measurement of Anger in Behavioral and Social Sciences: A Systematic Review of Literature. *International Journal of Medical Reviews*, 2(3), 279–286.
- Sharma, N., & Jain, C. (2019). Characterization of Facial Expression using Deep Neural Networks. *2019 5th International Conference on Advanced Computing and Communication Systems, ICACCS 2019*, 492–495.  
<https://doi.org/10.1109/ICACCS.2019.8728386>
- Shu, L., Xie, J., Yang, M., Li, Z., Li, Z., Liao, D., ... Yang, X. (2018). A review of emotion recognition using physiological signals. *Sensors (Switzerland)*, 18(7), 1–41.  
<https://doi.org/10.3390/s18072074>
- Stibe, A., & Wiafe, I. (2018). Beyond Persuasive Cities: Spaces that Transform Human Behavior and Attitude. *Persuasive Technology*, 1–2.
- Sun, L., Zou, B., Fu, S., Chen, J., & Wang, F. (2019). Speech emotion recognition based on DNN-decision tree SVM model, 115(April), 29–37.

<https://doi.org/10.1016/j.specom.2019.10.004>

- Sun, W., Zhao, H., & Jin, Z. (2019). A facial expression recognition method based on ensemble of 3D convolutional neural networks. *Neural Computing and Applications*, 31(7), 2795–2812. <https://doi.org/10.1007/s00521-017-3230-2>
- Talele, K., Shirsat, A., Uplenchwar, T., & Tuckley, K. (2017). Facial expression recognition using general regression neural network. *IEEE Bombay Section Symposium 2016: Frontiers of Technology: Fuelling Prosperity of Planet and People, IBSS 2016*, 1–6. <https://doi.org/10.1109/IBSS.2016.7940203>
- Tan, L., & Jiang, J. (2019). *Image Processing Basics. Digital Signal Processing*. <https://doi.org/10.1016/b978-0-12-815071-9.00013-0>
- Turk, M., & Pentland, A. (1991). Eigenfaces for Face Detection / Recognition. *Journal of Cognitive Neuroscience*, 3(1), 1–11. <https://doi.org/10.1162/jocn.1991.3.1.71>
- Tzirakis, P., Trigeorgis, G., Nicolaou, M. A., Schuller, B. W., & Zafeiriou, S. (2017). End-to-End Multimodal Emotion Recognition Using Deep Neural Networks. *IEEE Journal on Selected Topics in Signal Processing*, 11(8), 1301–1309. <https://doi.org/10.1109/JSTSP.2017.2764438>
- Valero, H. G. (2016). *Automatic Facial Expression Recognition*. University of Manchester. Retrieved from [http://studentnet.cs.manchester.ac.uk/resources/library/thesis\\_abstracts/MSc16/FullText/GamboaValero-Hugo-diss.pdf](http://studentnet.cs.manchester.ac.uk/resources/library/thesis_abstracts/MSc16/FullText/GamboaValero-Hugo-diss.pdf)
- Vapnik, V., Golowich, S. E., & Smola, A. (1997). Support vector method for function approximation, regression estimation, and signal processing. *Advances in Neural Information Processing Systems*, 281–287.
- Verma, K., & Khunteta, A. (2017). Facial expression recognition using Gabor filter and multi-

- layer artificial neural network. *IEEE International Conference on Information, Communication, Instrumentation and Control, ICICIC 2017*, 1–5.  
<https://doi.org/10.1109/ICOMICON.2017.8279123>
- Viola, P., & Jones, M. (2004). Robust Real-Time Face Detection Intro to Face Detection. *International Journal of Computer Vision*, 57(2), 137–154.
- Vo, D. M., & Le, T. H. (2016). Deep generic features and SVM for facial expression recognition. *NICS 2016 - Proceedings of 2016 3rd National Foundation for Science and Technology Development Conference on Information and Computer Science*, 80–84.  
<https://doi.org/10.1109/NICS.2016.7725672>
- Wang, Z., Jiang, R., Jiang, X., & Zhou, T. (2016). Learning sparse representations by K-SVD for facial expression classification. *Proceedings of 2015 4th International Conference on Computer Science and Network Technology, ICCSNT 2015*, 01(Iccsnt), 772–775.  
<https://doi.org/10.1109/ICCSNT.2015.7490856>
- Watkins, D. (2018). Getting started with Anaconda Python for data science | Opensource.com. Retrieved June 12, 2020, from <https://opensource.com/article/18/4/getting-started-anaconda-python>
- Wei, W., & Jia, Q. (2016). Weighted Feature Gaussian Kernel SVM for Emotion Recognition. *Computational Intelligence and Neuroscience*, 2016.  
<https://doi.org/10.1155/2016/7696035>
- Winters-Miner, L. ., Bolding, P. ., Hilbe, J. ., Goldstein, M., Hill, T., Nisbet, R., ... Miner, G. . (2015). Chapter 15. Prediction in Medicine – The Mining Algorithms of Predictive Analytics. In *Practical Predictive Analytics and Decisioning Systems for Medicine* (pp. 1–21).
- Xu, Y., Pang, Y., & Jiang, X. (2019). A Facial Expression Recognition Method Based on

- Improved HOG Features and Geometric Features. *Proceedings of 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference, IAEAC 2019*, (Iaeac), 1118–1122. <https://doi.org/10.1109/IAEAC47372.2019.8997772>
- Yang, B., Cao, J., Ni, R., & Zhang, Y. (2017). Facial Expression Recognition Using Weighted Mixture Deep Neural Network Based on Double-Channel Facial Images. *IEEE Access*, 6, 4630–4640. <https://doi.org/10.1109/ACCESS.2017.2784096>
- Yang, M. H., Kriegman, D. J., & Ahuja, N. (2002). Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1), 34–58. <https://doi.org/10.1109/34.982883>
- Yu, H., & Liu, H. (2015). Combining appearance and geometric features for facial expression recognition. *Sixth International Conference on Graphic and Image Processing (ICGIP 2014)*, 9443, 944308. <https://doi.org/10.1117/12.2179066>
- Yun, W. H., Kim, D. H., Chi, S. Y., & Yoon, H. S. (2007). Two-dimensional logistic regression. *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*, 2(1), 349–353. <https://doi.org/10.1109/ICTAI.2007.48>
- Zarbakshsh, P., & Demirel, H. (2018). Fuzzy SVM for 3D facial expression classification using sequential forward feature selection. *Proceedings - 9th International Conference on Computational Intelligence and Communication Networks, CICN 2017, 2018-Janua*, 131–134. <https://doi.org/10.1109/CICN.2017.8319371>
- Zavaschi, T H H, & Koerich, A. L. (2011). FACIAL EXPRESSION RECOGNITION USING ENSEMBLE OF CLASSIFIERS Pontifical Catholic University of Paran ´ a Department of Computer Science Curitiba , PR , Brazil Federal University of Paran ´ a Department of Computer Science, 1489–1492.
- Zavaschi, Thiago H.H., Britto, A. S., Oliveira, L. E. S., & Koerich, A. L. (2013). Fusion of

- feature sets and classifiers for facial expression recognition. *Expert Systems with Applications*, 40(2), 646–655. <https://doi.org/10.1016/j.eswa.2012.07.074>
- Zhan, J., Ren, J., Sun, P., Fan, J., Liu, C., & Luo, J. (2018). The neural basis of fear promotes anger and sadness counteracts anger. *Neural Plasticity*, 2018(Figure 1). <https://doi.org/10.1155/2018/3479059>
- Zhang, S., Zhang, S., Huang, T., & Gao, W. (2018). Speech Emotion Recognition Using Deep Convolutional Neural Network and Discriminant Temporal Pyramid Matching. *IEEE Transactions on Multimedia*, 20(6), 1576–1590. <https://doi.org/10.1109/TMM.2017.2766843>
- Zhang, T., Zheng, W., Cui, Z., Zong, Y., & Li, Y. (2019). Spatial-Temporal Recurrent Neural Network for Emotion Recognition. *IEEE Transactions on Cybernetics*, 49(3), 939–947. <https://doi.org/10.1109/TCYB.2017.2788081>
- Zhang, Y. D., Yang, Z. J., Lu, H. M., Zhou, X. X., Phillips, P., Liu, Q. M., & Wang, S. H. (2016). Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation. *IEEE Access*, 4, 8375–8385. <https://doi.org/10.1109/ACCESS.2016.2628407>
- Zhao, Y., Georganas, N. D., & Petriu, E. M. (2010). Applying Contrast-limited Adaptive Histogram Equalization and integral projection for facial feature enhancement and detection. *2010 IEEE International Instrumentation and Measurement Technology Conference, I2MTC 2010 - Proceedings*, 861–866. <https://doi.org/10.1109/IMTC.2010.5488048>
- Zheng, W. L., Liu, W., Lu, Y., Lu, B. L., & Cichocki, A. (2019). EmotionMeter: A Multimodal Framework for Recognizing Human Emotions. *IEEE Transactions on Cybernetics*, 49(3), 1110–1122. <https://doi.org/10.1109/TCYB.2018.2797176>

- Zhong, S., Chen, Y., & Liu, S. (2014). Facial expression recognition using local feature selection and the extended nearest neighbor algorithm. *Proceedings - 2014 7th International Symposium on Computational Intelligence and Design, ISCID 2014, 1*, 328–331. <https://doi.org/10.1109/ISCID.2014.108>
- Zuiderveld, K. (1994). Contrast Limited Adaptive Histogram Equalization. In *Graphics Gems IV* (pp. 474–485). USA: Academic Press Professional, Inc.

## Appendix

Source code for SVM on JAFFE dataset.

```
#OpenCV module

import cv2

import os

import numpy as np

from imutils import paths

from PIL import Image

from pyimagesearch.localbinarypatterns import LocalBinaryPatterns

from sklearn import svm

from sklearn.decomposition import PCA

from sklearn.metrics import classification_report, confusion_matrix

from sklearn.metrics import accuracy_score, precision_score, recall_score

from sklearn.model_selection import train_test_split, learning_curve, cross_val_score

from sklearn.model_selection import GridSearchCV

from sklearn.svm import SVC

from sklearn.svm import LinearSVC

from sklearn.preprocessing import StandardScaler
```

```
import matplotlib.pyplot as plt

#function to detect face using OpenCV

def detect_face(img):

    gray = cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)

    face_cascade = cv2.CascadeClassifier('./haarcascade_frontalface_default.xml')

    faces = face_cascade.detectMultiScale(gray, scaleFactor=1.2, minNeighbors=5);

    if (len(faces) == 0):

        return None, None

    (x, y, w, h) = faces[0]

    return gray[y:y+w, x:x+h]

print("=====
=====")

print("          PROCESSING")

print("=====
=====\n")

emotion = {'angry' : 1, 'not-angry' : 2}

def face_det_crop_resize(img_path):

    img = cv2.imread(img_path)

    gray = cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)

    cv2.imwrite(img_path, gray)
```

```
face_cascade = cv2.CascadeClassifier('./haarcascade_frontalface_default.xml')

img = cv2.imread(img_path)

gray = cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)

faces = face_cascade.detectMultiScale(gray, 1.3, 5)

for (x,y,w,h) in faces:

    face_clip = img[y:y+h, x:x+w]

    cv2.imwrite(img_path, cv2.resize(face_clip, (128, 128)))

def preprocessing(data_folder_path):

    dirs = os.listdir(data_folder_path)

    for dir_name in dirs:

        subject_dir_path = data_folder_path + "/" + dir_name

        subject_images_names = os.listdir(subject_dir_path)

        for image_name in subject_images_names:

            image_path = subject_dir_path + "/" + image_name

            print("Detecting, cropping, resizing, and saving : ", image_name)

            if face_det_crop_resize(image_path):

                print(image_path)

                print(image_name)

def prepare_training_data(data_folder_path):

    dirs = os.listdir(data_folder_path)
```

```
faces = []

labels = []

data = []

desc = LocalBinaryPatterns(24,8)

for dir_name in dirs:

    subject_dir_path = data_folder_path + "/" + dir_name

    subject_images_names = os.listdir(subject_dir_path)

    for image_name in subject_images_names:

        if image_name.startswith("."):

            continue;

        image_path = subject_dir_path + "/" + image_name

        print(image_name)

        print(image_path)

        image = cv2.imread(image_path)

        print(image)

        gray = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)

        for x, y in emotion.items():

            if dir_name == x:

                labels.append(y)

        #face = detect_face(image)
```

```
#cv2.resize(face,(128,128))

#faces.append(face)

#smoothened_faces = image_smoothing(gray,image_path)

hist = desc.describe(gray)

#clahe = cv2.createCLAHE(clipLimit=2.0, tileGridSize=(8,8))

#c11 = clahe.apply(smoothened_faces)

#hist = desc.describe(c11)

#labels.append(label)

data.append(hist)

#new_data = np.array(data)

cv2.destroyAllWindows()

cv2.waitKey(1)

cv2.destroyAllWindows()

return data, labels

def image_smoothing(image_name, image_path):

    #print("Smoothing face on {}".format(image_path))

    blur = cv2.GaussianBlur(image_name,(5,5),0)

    medianBlur = cv2.medianBlur(blur,5)

    print("Smoothened face on {}".format(image_path))

    return medianBlur
```

```
print("Pre-processing.....")

preprocessing("data")

print("Preparing training data.....")

data, labels = prepare_training_data("data")

print('\n')

#data = np.array(data)

print("Data prepared")

#print total faces and labels

#print("Total faces: ", len(faces))

print("Total faces detected in the data: ", len(data))

print("Total labels detected in the data: ", len(labels))

print('\n')

#new_data = image_smoothing(data)

new_data = np.array(data)

scaler = StandardScaler()

scaler.fit(new_data)

scaled_data = scaler.transform(new_data)

pca = PCA(n_components=26, whiten=False)

pca.fit(scaled_data)
```