AUTOMATIC RECOGNITION OF MICRO-EXPRESSIONS USING LOCAL BINARY PATTERNS ON THREE ORTHOGONAL PLANES AND EXTREME LEARNING MACHINE

Iyanu Pelumi ADEGUN

Student Number: 746542
Supervisor: Dr. Hima B. Vadapalli

School of Computer Science and Applied Mathematics

A Dissertation submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of *Master of Science.*

September 2017

Declaration of Authorship

I, Iyanu Pelumi Adegun declare that this dissertation is my own, unaided work. It is being submitted for the Degree of Master of Science at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.

Signature Date

Abstract

Recognition of micro-expressions is a growing research area as a result of its application in revealing subtle intention of humans especially under high stake situations. Owing to micro-expressions' short duration and low intensity, efforts to train humans in their recognition has resulted in very low performance. The use of temporal methods (on image sequences) and static methods (on apex frames) were explored for feature extraction. Supervised machine learning algorithms which include Support Vector Machines (SVM) and Extreme Learning Machines (ELM) were used for the purpose of classification. Extreme learning machines which has the ability to learn fast was compared with SVM which acted as the baseline model. For experimentation, samples from Chinese Academy of Micro-expressions (CASME II) database were used. Results revealed that use of temporal features outperformed the use of static features for micro-expression recognition on both SVM and ELM models. Static and temporal features gave an average testing accuracy of 94.08% and 97.57% respectively for five classes of micro-expressions using ELM model. Significance test carried out on these two average means suggested that temporal features outperformed static features using ELM. Comparison between SVM and ELM learning time also revealed that ELM learns faster than SVM. For the five selected micro-expression classes, an average training time of 0.3405 seconds was achieved for SVM while an average training time of 0.0409 seconds was achieved for ELM. Hence we can suggest that micro-expressions can be recognised successfully by using temporal features and a machine learning algorithm that has a fast learning speed.

Publication

Automatic Recognition Of Micro-Expressions Using Local Binary Patterns On Three Orthogonal Planes And Extreme Learning Machine [1].

Authors: Iyanu Pelumi Adegun and Hima Vadapalli.

Paper presented at Pattern Recognition Association of South Africa-Robotics and Mechatronics (PRASA-RobMech) International conference, IEEE, 2016.

Dedication

To the Almighty GOD who gives wisdom and out of whose mouth comes knowledge and understanding; Proverbs 2:6.

Acknowledgements

The successful completion of this dissertation was made possible with support from a number of people.

First, I would like to sincerely appreciate the contributions and support of my supervisor, Dr Hima Vadapalli. In addition to this, her patience, motivation and consistent guidance through out the period of carrying out this research contributed a great deal to completing the research successfully.

The support of the DST-NRF Centre of Excellence in Mathematical and Statistical Sciences (CoE-MaSS) towards this research is hereby acknowledged. Opinions expressed and all the conclusions arrived at, are mine and not attributed to the CoE. I would also like the acknowledge the support received from the Financial Aids and Scholarship Office towards a presentation of part of my research at the PRASA-RobMech International Conference in 2016.

I am very grateful to the Institute of Psychology, Chinese Academy of Sciences for making their database available for this research. The CASME II database was very helpful in achieving results.

Prof. Turgay Celic through whom I learnt the concept of machine learning on which this research is based is highly appreciated for his contributions.

Special thanks to my husband (Olumuyiwa) and son (Ayooluwa) for their support. The sacrifices they made when I had to stay extra hours on experiments and writing are highly appreciated. Also, I am grateful to my parents, parents-in-law and siblings who supported by constantly checking up on my research progress.

I would like to appreciate senior colleagues (Abejide Ade-Ibijola and Abiodun Modupe) and my colleague, Moses Rupenga for their willingness to always answer questions on my research.

Above all, I am grateful to God for the help I received from Him throughout the course of this research.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Expressions reveal what takes place in the human mind at a particular time. These are often displayed through speech, body gestures or facial expressions. Of the different modes of expression, facial expressions appear to be the most expressive means through which humans show their emotions [9], [10]. Facial expressions play a vital role in day to day life of humans. As a result, several researchers have developed automated systems for recognition and interpretation of facial expressions. For example, Vural et al. [11] proposed a system that identifies the level of drowsiness in drivers. Another example of facial expression recognition system is that of Whitehill et al. [12] which was used to get response/feedback from students while being taught.

Facial expressions fall under the category of 'macro-expressions'. These are the normal expressions that are seen daily in our interactions with people and last between 1/2 seconds and 4 seconds. However, in high stake situations, people often conceal or suppress their true emotions because of the fear of being caught [13]. These concealed emotions take place within the duration of 1/5 and 1/25 seconds and are known as 'micro-expressions'. These expressions were initially detected by Haggard and Isaacs [14] and at a later time by Ekman and Friesen [15]. Apart from the short duration of micro-expressions, they also possesses low intensity. These two unique features make the recognition of micro-expressions by humans more difficult even when they are trained to perform such tasks [15] [16]. Humans who are trained can recognise micro-expressions but with a very low accuracy due to the short duration and low intensity of these expressions[17]. Both macro-expressions and micro-expressions can be expressed in seven different forms which include sadness, happiness, fear, anger, surprise, disgust and contempt. Figure 1.1 shows the contrast between some macro- facial expression samples and micro-facial expression apex frame samples.

Owing to the limited ability of humans to recognise micro-expressions effectively, it becomes necessary to develop automated systems that have the ability to detect them automatically and effectively. Automatic micro-expression recognition involves the use of computer-based methods for recognition of micro-expressions.

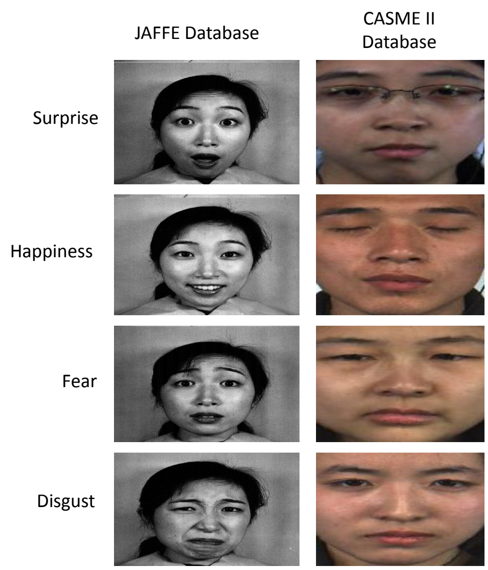

Figure 1.1: Samples of facial expression from JAFFE Database [2] and samples of micro-expression apex frames from CASME II database [3]

1.1 Potential Applications of Micro-expressions

Micro-expressions have been found to be relevant to many fields. They can be used by security officers or law enforcement agents to detect abnormal behaviours during interrogation of suspects. According to Ekman [15], it is believed that in high stake situations, people tend to suppress their true emotions which gives a clue to trained security/law enforcement agents about those with bad intentions. In the United States, the Transportation Security Administrators (TSA) at the airports use a technique known as Screening Passengers by Observation Technique (SPOT) to identify people who could be a threat to people in an aircraft. [18]. One major shortcoming of SPOT used by the TSA as identified by [19] is that the system was never subjected to controlled scientific tests.

Micro-expressions can also be used by health care practitioners for clinical diagnosis [20], [21]. This is useful when the health care practitioners need to re-assure themselves of a particular disease prediction made on a patient. In the earlier days of micro-expression detection, Paul Ekman [22] carried out an analysis on a patient who was believed to have depression based on an health practitioner's diagnosis. This patient made a suicidal attempt when he was informed about the diagnosis. The suicidal attempt was discovered by the health practitioner when the patient showed extreme sadness in a very short time but quickly suppressed it with smiles. An application in education would be when teachers can get an assessment of their teaching skills based on the response (emotions) exhibited by their students [23], [24], [25].

1.2 Research Motivation

The limitation of human ability (time and skills) in recognition of micro-expression has created a need to develop fully automatic systems for micro-expression recognition. Since the subject of automatic micro-expression recognition systems has come into existence, only a few studies have been carried out in this area. Some of these studies employed the use of static feature extraction technique which may not effectively extract features required to build micro-expression recognition systems. It is more effective to make use of temporal data for micro-expression expression recognition due to its short duration and low intensity. Results from past studies ([23], [26], [27]) have also shown that majority of the techniques used for automatic recognition of micro-expressions require longer training period which might amount to low or average performances.

In this study, we proposed the use of temporal feature extraction in conjunction with supervised machine learning algorithms for automatic recognition of micro-expressions. The combination of techniques mentioned above were compared with existing traditional techniques.

1.3 Research Hypothesis

Micro-expressions can be automatically recognised by combining temporal feature extraction technique and supervised machine learning techniques.

1.4 Research Questions

Main Research Question: Can effective models be built for automatic micro-expression recognition by using a combination of temporal feature extraction and supervised machine learning techniques?

Sub-questions

1. Can temporal feature extraction techniques be used to effectively extract features for micro-expression recognition?
2. Can machine learning algorithms successfully classify micro-expressions?
3. Which of the machine learning algorithms for micro-expression recognition has a better performance?
4. Which of the machine learning algorithms has a better learning speed?
5. Aside from accuracy, can other measures be used to evaluate the performance of micro-expression recognition systems?

1.5 Research Objectives

Based on the questions stated above, this research seeks to achieve the following under-listed objectives.

1. To extract features from micro-expression samples using temporal feature extraction technique
2. To compare the performance and effectiveness of static feature extraction with temporal feature extraction techniques
3. To compare the performance of different machine learning algorithms for micro-expression recognition
4. To compare the learning speed of two machine learning algorithms
5. To evaluate the performance developed model using other measures aside from accuracy

1.6 Overview of Thesis

The thesis is structured as follows:

The first chapter is introductory.

In Chapter 2, literature on micro-expression recognition is reviewed. The review is based on different feature extraction and classification techniques used by various researchers.

Chapter 3 gives background information on the techniques used for the study and other related techniques. These include feature extraction techniques, classification methods and details on existing micro-expression databases.

Chapter 4 describes the methodology used in carrying out the research. It also provides the mathematical formulations of the methods used.

In Chapter 5, experimental results are presented, analysed and discussed in relation to some relevant literature. Chapter 6 concludes the dissertation. It presents an overall analysis of the research and areas for further study.

Chapter 2

Literature Review

2.1 Introduction

This chapter presents an in-depth review of related studies on automatic micro-expression recognition. The review was carried out based on existing micro-expression databases, type of feature extraction and classification techniques used for micro-expression recognition. In reviewing the relevant literature, details on the techniques used, results achieved and limitations/future works are presented.

2.2 Feature Extraction

Micro-expression recognition requires extraction of relevant features from the images which form feature vectors that are used as input for classification. Techniques for feature extraction are dependent on the form of images from which features are to be extracted. Extraction of features could be either from static (still) images without any form of variations [28] or features from temporal data in the form of image sequences or videos showing the continuous flow of facial movements.

2.2.1 Feature Extraction from Static Images

Some of the existing techniques for extracting features from static images include Gabor filters [2], Local Binary Patterns (LBP) [29] and Histogram of Gradients (HOG) [30]. Facial Expression Recognition (FER) systems and some micro-expression recognition systems have been successfully used to extract features through static feature extraction techniques on still images.

For instance, Wu et al. [26] carried out a study on the recognition of micro-expressions using single image frames. Gabor filters [2] which had 9 scales and 8 orientations was employed as the feature extraction technique. Gabor filters performed the function of filtering the images by alteration of all individual pixels in each image. Results showed a recognition accuracy of 85.42% which was an improvement over micro-expression recognition by trained human beings. However, Gabor filters can only be used to extract features in the spatial domain. They are also time consuming and memory intensive, as identified in [31].

Yan et al. [32] conducted a pilot study with the aim of analysing micro-expressions by quantifying and studying the pattern of facial movements. In their work, two feature extraction methods were applied. First, a Constraint Local Model (CLM) [33] which is an upgraded version of Active Appearance Model (AAM) [34] and Active Shape Model (ASM) [35]. Constraint Local Model was used to detect faces and to track feature points, thereafter, using feature points to create region of interests (ROI). The second feature extraction technique used was LBP which helped to extract features from the ROI and calculate difference between the frames. Result showed that the frame with the highest difference was marked as the apex frame. However, the study revealed that LBP does not have the ability to quantify dynamic information such as direction, velocity or subtle movements since they were meant for the purpose of describing textures of static images. Also, there was no criteria for determining onset and offset frames.

In recent times, a few studies ([36],[37]) have shown that micro-expression recognition systems can also use features extracted from static apex micro-expression frames and obtain results. In their work, they proved that apex frames are sufficient to recognise micro-expressions rather than using entire image sequences. Liong et al. [36] hypothesized that: "Obtaining a promising recognition accuracy is not dependent on the use of large number of frames". LBP and a newly proposed technique called Bi-Weighted Oriented Optical Flow (Bi-WOOF) were used for feature extraction after spotting apex frames. "Leave one subject out" cross validation was performed by using samples from one subject as testing data while the remaining samples were used as training data. This process was repeated for k times where k stands for the total number of subjects in the database. Final performance results were calculated by getting the average performance for each subject. SMIC and CASME II micro-expression databases were used to evaluate the system and they had F-Score performance of 61% and 62% respectively. Results showed that apex frames provided more information when compared with entire image sequences. However, the reason behind their findings was not fully discovered.

2.2.2 Feature Extraction from Image sequences (Temporal Data)

Recognition of micro-expressions require that features are extracted from dynamic facial images because of their short duration and low intensity on the facial muscles. Feature extraction from image sequences is relevant to both ordinary facial expression and micro-expression recognition. For facial expression recognition, general techniques include facial feature point tracking and optical flow [28] which have been used successfully in [38], [39], [40], [41] and [42]. For micro-expression recognition, some existing techniques meant for the purpose of extracting features from image sequences include variants of LBP, 3D Histogram of Gradients (HOG) and optical strain.

Zhao and Pietikainen [4] proposed an approach for recognition of dynamic textures by extending ordinary LBP to Volume LBP (VLBP) and LBP on Three Orthogonal Planes (LBP-TOP). These methods combine both static and temporal features during dynamic texture recognition. The effectiveness of these methods were evaluated using Cohn Kanade facial expression database [43]. Subjects from Cohn Kanade's database were separated randomly into n groups and 'leave one group out' cross validation was performed. Images were divided into 9×8 non-overlapping blocks and features were extracted from each block. These block features were concatenated to serve as the final feature vector. A description of this process is given in Figure 2.1. Both LBP-TOP and VLBP features were extracted using varying number of neighbouring points P and radii R_{XY}, R_{XT}, R_{YT}. For classification of the facial image samples, SVM was used. Results revealed that the highest average accuracy (96.26%) was achieved using LBP-TOP when the number of neighbouring points P was set to 8 and radius set to 3. For VLBP, the highest average accuracy (95.19%) was achieved when radii R_{XY}, R_{XT}, R_{YT} were set to 3, 2 and 3 respectively. Both LBP-TOP and VLBP were found to outperform the use of static textures.

Local Binary Patterns on Three Orthogonal Planes have also been widely used for micro-expression recognition, as seen in Yan et al. [3] while evaluating their spontaneous database (CASME II). Features were extracted from 5 x 5 blocks of facial images and five classes of micro-expressions (happiness, surprise, disgust, repression and others) were recognised. The radii in X and Y axes were varied between 1 and 4 while radius T was varied between 2 and 4. Performances at different radii values were also calculated and compared. Highest accuracy was achieved at values 1, 1 and 4 for XY, XT and YT planes respectively.

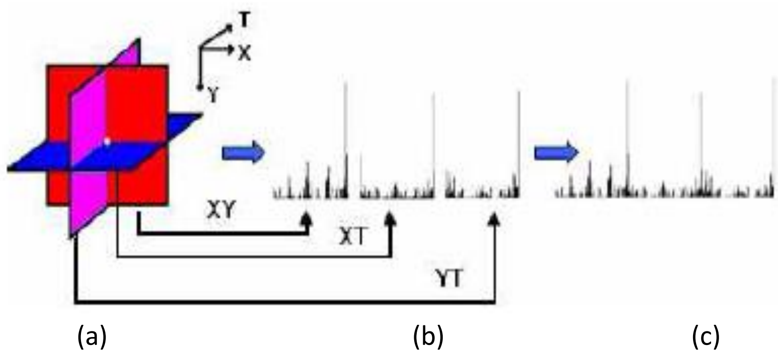

Figure 2.1: (a) A dynamic texture showing three planes, (b) feature histogram for each plane and (c) concatenated feature histograms [4]

Davison et al. [44] carried out an investigation to check whether micro-facial movement sequences can be distinguished from neutral face sequences. LBP-TOP and Gaussian Derivatives were used to detect the presence of both micro-facial movements and neutral expressions. SVM and random forest techniques were used for classification. The system was able to help humans to interpret micro-facial movements and what they mean in the context of a different situations.

Li et al. [6] and Pfister et al. [23] also used LBP-TOP feature extraction techniques while performing a baseline evaluation for the SMIC database. Three categories of SMIC dataset were used for their experiments. These categories were named based on the type of camera used for their acquisition. SMIC-HS means SMIC from High Speed camera, SMIC-VIS from visual camera while SMIC-NIR means SMIC from near infra-red camera. Local spatio-temporal features were extracted from pre-processed image sequences from the dataset while SVM was used for classification. The highest recognition accuracy of 49.30% was achieved on SMIC-HS dataset.

Guo et al. [45] also presented the use of LBP-TOP for feature extraction in a holistic manner. The term 'holistic' means that LBP-TOP features were extracted from entire image sequences rather than being extracted from blocks of an image. Their motivation for employing the holistic approach was to avoid information redundancy and to reduce the amount of calculations to be performed. Nearest Neighbour (NN) technique was used for classification. Results revealed a recognition accuracy of 65.83% achieved at LBP-TOP radii values of 1, 1 and 2 for X, Y and T planes respectively. However, the results showed that NN classification was not suitable for use when the number of samples increased.

Aside from variants of LBP, one other technique that has been used for extracting temporal features from image sequences is 3D Histogram of Oriented Gradients (HOG). 3D HOG was to serve as an improvement over the original HOG by Dalal and Triggs [30]. The initial version was meant for static 2D images while the 3D HOG was meant for the purpose of dynamic textures (videos).

Polikovsky et al. [27] and [46] explored the use of 3D HOG in their work. Their aim was to develop a descriptor that is specifically adapted for detecting facial movements by observing gradients histogram using a dataset comprising of 10 university students. Faces were divided into 12 regions and facial cubes were extracted from each of the regions. 3D HOG features were computed for each frame and the resulting histogram was concatenated and normalised, which formed the final feature vector. They were able to develop a micro-expression system capable of measuring the three phases of micro-expression. These three micro-expression phases included: Constrict (constriction of muscles), In Action (construction of muscles) and Release (release of muscles).

Optical strain which originated from optical flow is another dynamic feature extraction technique that has the ability to measure subtle changes that occur in the human face. A study on the use of optical strain to distinguish between macro- and micro-expressions were carried out in [47] and [48]. All the subjects' faces were divided into sub-regions and strain patterns were identified for each sub-region. Strain patterns for each sub-region were used to identify subtle changes before detecting micro-expressions. For macro-expressions, 85% spotting accuracy was achieved while 74% accuracy was achieved for micro-expression spotting.

Wang et al. [49] also carried out a study on the recognition of micro-expressions. Their intention was to reveal that colour may provide useful information for micro-expression recognition through dynamic textures on Tensor Independent Colour Space (TICS). A set of ROIs was defined based on FACS [50]. Later, dynamic texture histograms were calculated for each ROI. The experiments were conducted on CASME [7] and CASME II [3] micro-expression databases. Results showed that accuracy was 58.53% when CASME was used while accuracy was 61.85% on CASME II database.

2.3 Classification of Micro-expressions

Micro-expression recognition requires the use of machine learning algorithms for classification. Existing algorithms that have been used for this purpose include SVM, Nearest Neighbours, Random Forest, Decision Trees, Naive Bayes and ELM. Out of these, SVM - a traditional learning algorithm is commonly used in most studies on micro-expression recognition. Some of these studies are discussed below.

Wu et al. [26] proposed a means to recognise micro-expressions using a frame by frame approach. Gabor filter [2] were used for feature extraction from facial image frames while GentleSVM (a combination of Gentle boost and SVM algorithm) was used for classification. Their system recorded an accuracy of 95.83% for spotting and 85.42% for recognition. The limitation identified from the work is the need to increase the training set in order to achieve better results.

Pfister et al. [23] proposed a micro-expression detection system that uses SVM, Multiple Kernel Learning and Random Forest for classification. Two different corpora were used to evaluate the system (York Deception Detection Test (YorkDDT) corpus [51] and SMIC corpus [6]). On YorkDDT corpus, highest recognition rate of 83.0% was obtained from combination of Machine Kernel Learning (MKL) classifier with 10 frames of Temporal Interpolation Model (TIM 10). For SMIC corpus, highest recognition rate of 74.3% was obtained from combination of RF classifier with (TIM 10). SVM had a lower detection and recognition accuracy of 70.3% and 59.8% respectively.

Yan et al. [5] performed a baseline evaluation of CASME II database using LBP-TOP for feature extraction and SVM for classification. SVM classification results revealed that the highest average accuracy obtained was 63.41% at LBP-TOP radii values of 1, 1 and 4. Details on how SVM classification was performed was not stated in the literature.

Another study that used other classification algorithms apart from SVM include that of Shreve et al. [47] who developed a solution that has the ability to spot both macro- and micro-expressions without training any models. Magnitude of detected facial strain was used to recognise expressions. Their model was able to identify regular expressions, suppressed (rapid) micro-expressions and expressed (universal) macro-expressions separately. A spotting accuracy of 74% for micro-expressions and 85% for macro-expressions was achieved.

Wang et al. [52] employed the use of discriminant tensor subspace analysis (DTSA) to extract features and ELM was used for classification of micro-expressions. Extreme Learning Machine models were built using ten-fold cross validation to ensure that all test sets were independent. All samples were divided into ten subsets and at each time, one of the subsets were used as test set while the rest were used as training set. The number of hidden neurons were gradually increased from 1 to 5000 at an interval of 10. Their algorithm was tested on ORL [53], Yale [54] and YaleB [55] facial databases and CASME [7] databases. The highest accuracy was achieved with 2,751 hidden neurons using Yale database, 4,091 hidden neurons using ORL database and 661 hidden neurons using YaleB database. Their recognition accuracies were 91.56%, 99.78% and 99.21% respectively. On CASME database, accuracy of 46.90% was achieved. It was noted that the system had a higher recognition accuracy when facial expression databases were used and a lower accuracy when micro-expression database (CASME) was used for evaluation.

To improve the accuracy of micro-expression recognition, Guo et al. [56] combined centralised binary pattern on three orthogonal planes (CBP-TOP) with ELM as classifier. The underlying principle of CBP-TOP is similar to that of LBP-TOP. It compares the centre pixel with pairs of neighbouring points and calculate CBP codes in XY, XT and YT planes as opposed to LBP-TOP that compares the centre pixel with each neighbouring point. A recognition accuracy of 82.07% was obtained when evaluated using CASME database.

2.4 Micro-expression Databases

Micro-expression database samples can be categorised into acted or spontaneous samples. Acted micro-expression samples are those acquired by asking subjects to act out micro-expression after a careful explanation of what such expressions entails. For spontaneous samples, subjects' emotions were stimulated by real-time emotional occurrence. Micro-expression samples elicited spontaneously portray the actual micro-expression when compared with the acted samples as highlighted in [7]. Existing micro-expression database samples that were acted include USF-HD database [5] and Polikovsky's database [27] while spontaneous ones include Spontaneous Micro-expression database (SMIC) [6], Chine Academy of Sciences Micro-expression database (CASME) [7] and CASME II [3].

Shreve et al. [47] carried out a study on spotting and recogniton of macro- and micro-expressions. In their work, samples from three micro-expression datasets were used to test their algorithm. The datasets used were USF-HD dataset [5], Canal 9 Political debate dataset [57] and Ekman's videos [22] found on the Internet. Results from the experiments revealed that peak accuracy of 80% was achieved when samples from USF-HD dataset were used. However, USF-HD samples had an average duration of one minute which exceeds the normal duration of micro-expressions.

An extended study of Shreve et al. [47] was carried out by the same set of authors in [58]. In this study, samples from USF [5] and SMIC [6] micro-expression databases and Denver Intensity of Spontaneous Facial Action (DISFA) facial expression database [59] were used for testing their algorithm. SMIC samples are spontaneous when compared with USF samples which are acted. Performance of their algorithm was measured using Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve. Results revealed that SMIC was more promising when compared with USF micro-expression database and DISFA facial expression database.

Spontaneous micro-expression database samples were also used in [45], [60], [61] and [62] to test their micro-expression algorithms. The authors achieved better results when using these samples compared with acted samples. One major drawback of SMIC database is the lack of appropriate emotion labelling. Labelling was carried out based on subjects' self-reports only.

CASME database samples were used in other studies like [49], [52], [56]. These studies revealed that experiments carried out using CASME database had better performances when compared with earlier spontaneous micro-expression databases. Details on results from these studies were discussed in the classification section.

More recently, many of the studies on micro-expression analysis [32], [49], [60] and [62] have used CASME II database as a result of an increased sample size and camera resolution quality. More details on CASME II database is provided in background chapter.

Table 2.1 presents a summary of reviews carried out on micro-expression recognition with details on feature extraction and classification methods, databases used, results achieved and limitations of the each study.

Table 2.1: Summary of literature review carried out on micro-expression recognition

Author(s)	**Objective**	**Feature Extraction**	**Classification**	**Database**	**Results / Contributions**	**Limitations/ Future research**
Wu et al. [26]	To build a fully automatic micro-expression recognition system	Gabor Filters	Gentle SVM (Combination of Gentle Boost algorithm and SVM	Cohn & Kanade's dataset	Spotting Accuracy: 95.83%, Recognition Accuracy: 85.42%	Evaluation of algorithm performed on facial expression dataset and not micro-expression dataset
Yan et al. [32]	To quantify dynamic movements and spot apex frames for micro-expression analysis	CLM and LBP	-	CASME II	Apex frames were detected from samples. This reduced the amount of manual coding work during micro-expression analysis.	Ordinary LBP cannot quantify dynamic information such as direction and subtle movements
Pfister et al. [23]	To develop a micro-expression recognition system that performs better than human recognition	LBP-TOP	SVM, MKL & random forest classifiers	YorkDDT, SMIC	SMIC accuracy: 74.3%, YorkDDT accuracy: 83.0%	Dataset used had limited samples

Shreve et al. [58]	To propose a solution to macro- and micro-expression spotting in video sequences	Optical strain	-	USF, DISFA and SMIC datasets	Micro-expression spotting accuracy: 74%, Macro-expression spotting accuracy: 85%	Dataset used does not have ground truth labelling.
Yan et al. [3]	Baseline evaluation of CASME II database	LBP-TOP	SVM	CASME II	Recognition accuracy: 63.41%	There is a need to elicit samples from more natural communication and interactions in future
Wang et al. [52]	To use a novel algorithm for micro-expression recognition to improve accuracy	Discriminant Tensor Sub-space Algorithm (DTSA)	ELM	ORL, Yale and YaleB facial expression databases and CASME database	Recognition accuracy (CASME): 46.90%	Low accuracy on CASME database as a result of limited number of training samples
Guo et al. [45]	To perform micro-expression recognition using a novel approach	LBP-TOP	Nearest Neighbour	SMIC database	Accuracy: 65.83%	Little above average accuracy. Database samples were few

Guo et al. [56]	To propose a novel feature extraction technique for recognition of micro-expressions	CBP-TOP	ELM	CASME	Highest recognition accuracy: 82.07%	CBP-TOP was found to be slower than LBP-TOP because only the centre pixel was considered.
Wang et al. [63]	Extraction of subtle motion information from micro-expressions	Robust Principal Component Analysis (RPCA) and Local Spatio-temporal Descriptors (LSTD)	-	SMIC and CASME II	Accuracy (SMIC): 71.34%, Accuracy (CASMEII): 65.45%	Future work identified was to use proposed method for motion detection in surveillance

2.5 Motivation for the current work based on literature review

Owing to the critical role that feature extraction and classification plays in micro-expression recognition, varying combination of techniques have been used in the past. This was discussed within sections of this chapter. Studies like [26], [32], [36] and [37] performed feature extraction on static frames. As identified in [3], [23] and [45], temporal feature extraction is more relevant to micro-expression recognition compared with static feature extraction. The reason for this is the short duration of micro-expressions and their low intensity [63]. However, Liong et al. [36], [37] carried out a study to show that static feature extraction using apex frames from micro-expression samples are sufficient for recognition of micro-expressions.

In this study, LBP-TOP feature extraction technique was selected because of its ability to extract both spatial and temporal features [4]. These features were extracted using a holistic approach whereby spatio-temporal features are extracted from whole facial images instead of blocks of facial regions. This approach was selected so as to reduce information redundancy and avoid complexity during feature extraction process as seen in [45].

In micro-expression recognition, classification is often performed using machine learning models. Some of the existing models include: SVM [64], Random Forest [65] and Nearest Neighbour [66]. Of these models (mentioned above), SVM appears to be the commonly used model because of its good generalisation performance irrespective of bias in training sample [67]. However, SVMs have some major drawbacks which include their complexity and slow learning speed [68]. On the other hand, ELM [69] is a more recent machine learning algorithm that provides a simpler and faster means of developing classification models [52], [56].

For the purpose of this study, we propose a combination of temporal feature extraction technique (LBP-TOP) with ELM for classification of micro-expressions. Comparison was made between performance of using LBP on static images and that of using LBP-TOP for feature extraction from image sequences. Comparison was also made between the performance of traditional SVM and ELM models for micro-expression recognition.

2.6 Summary

This chapter presents a review of related literature. Studies dealing with recognition of micro-expressions using both spontaneous and non-spontaneous databases, static and dynamic feature extraction techniques and machine learning algorithms were dealt with. This chapter also presents a motivation for techniques used in the study. Motivation was based on some limitations identified in the previous work done in this field.

Chapter 3

Background

3.1 Introduction

This chapter gives a description of databases and models used in this study. Details on existing micro-expression databases, LBP and LBP-TOP feature extraction techniques, architecture of SVM and ELM with their mathematical formulations are presented.

3.2 Micro-expression Databases

For micro-expression recognition systems to perform well, it is essential to have a good set of data samples. There are two categories of micro-expression databases that exists. These include: Acted/posed micro-expression dataset and spontaneous micro-expression dataset. Further details on these existing micro-expression dataset are presented in the next section.

3.2.1 Acted USF-HD Database

USF-HD database [5] consists of 100 micro-expression samples and 181 macro-expression samples recorded at about 30 frames per second (fps). To obtain the micro-expression samples, subjects were asked to watch video clips that contained micro-expressions and thereafter repeat whatever had been watched. This type of database might not be effective for micro-expression recognition because a subject might forget whatever he/she had watched in the video.

3.2.2 Acted Polikovsky's Database

Polikovsky's database [27] is an acted micro-expression database that contains a total of 42 samples elicited from 10 subjects. These 10 subjects includess 5 Asians, 4 Caucasians and 1 Indian students from a university recorded at 200fps. These subjects were instructed to perform 7 basic emotions with low facial muscles intensity and to go back to the neutral face expression immediately, thus simulating the micro-expression motion. One of the major limitations of Polikovsky's database is its limited number of samples and the fact that the samples are non-spontaneous. Some subjects from the database are shown in Figure 3.1.

Figure 3.1: Subjects from Polikovsky's database [5]

3.2.3 Spontaneous SMIC Database

SMIC database [6] is made up of 164 spontaneous micro-expressions clips retrieved from 16 participants and recorded with a 100fps (frames per second) camera. Depicted emotions were labelled as positive (happy), negative (sad, anger, fear and disgust) and surprise. An example of a sequence from SMIC database is shown in Figure 3.2. This database was created because of the need to increase the sample size of micro-expression databases. Samples were acquired by showing emotional video clips to participants in an interrogation room. This targeted creating a representation of a high stake situation. Participants were also motivated to suppress their facial expressions. Self-report questionnaires that included questions on the type of emotions felt while watching the video clips were given to the participants. Labelling was carried out based on results from each participant's self-report. The experiments were conducted using three separate datasets based on the type of camera used for capturing the videos (HS- High Speed camera, VIS- Visual camera and NIR- Near Infrared camera). The samples resulting from these set of experiments contain sufficient data needed for automatic recogniton of micro-expressions. However, the need to extend the database such that micro-expressions can be spotted from live stream videos was identified.

Figure 3.2: Surprise sequence from SMIC Database [6]

3.2.4 Spontaneous CASME Database

Chinese Academy of Sciences Micro-expression database [7] consists of 195 spontaneous and dynamic facial micro-expressions recorded with two separate 60 fps cameras. When compared with previous micro-expression databases, CASME is spontaneous and not acted and also has a higher number of samples.The samples were elicited from about 1500 facial expressions with an onset duration that is less than 250 ms (within the normal duration of micro-expressions). These samples were coded with the onset, apex and offset frames. Onset frame is the first frame where changes from the neutral expression occurs. Apex frame is the frame where the highest intensity of the expression is reached while offset frame is the last frame before facial expression changes to neutral. The captured emotions were labelled and classified into seven categories: happiness, surprise, disgust, fear, sadness, repression and tense. A sequence of images from one of the subjects is shown in Figure 3.3.

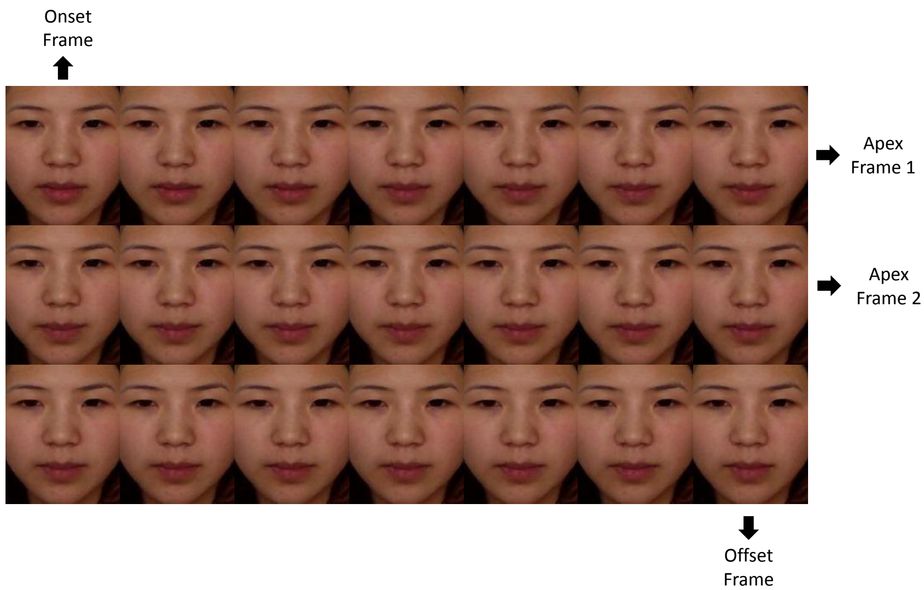

Figure 3.3: Repression micro-expression image sequence from one of the subjects in CASME Database [7] showing onset, apex and offset frames.

3.2.5 Spontaneous CASME II Database

Chinese Academy of Sciences Micro-expression II database [3] is an improved version of CASME database. It consists of 247 facial micro-expression samples retrieved from 26 participants. These samples were retrieved from 18 participants who were made to watch highly emotional video clips. While watching the videos, a screen was placed before each of the participants and a high resolution camera was used to record their emotions. After watching the clips, participants were told to rate the intensity of the video clips using a 7-point Lickert scale with 0 as the lowest and 6 as the highest.

Video recordings from each participant were divided into frames and pre-processed by removing facial and body movements that were not regular. Final micro-expression samples were selected based on recordings that had a total duration of less than 500 milliseconds or an onset duration of less than 250 milliseconds. These samples were labelled using FACS [70] but their labelling criteria were not the same with that of ordinary facial expressions. The micro-expression samples consist of seven (7) classes which includes happiness, surprise, repression and others. A sample image sequence from CASME II database is described in Figure 3.4. When compared with earlier versions of micro-expression databases, some of its outstanding features identified by Wang et al. [52] are listed as follows:

- Increase in sample size (247 samples)
- Video recordings with higher resolution of 200fps and proper illumination
- Spontaneous and dynamic micro-expression samples rather than acted or posed micro-expression samples
- Properly labelled emotions with coded onset and offset frames.

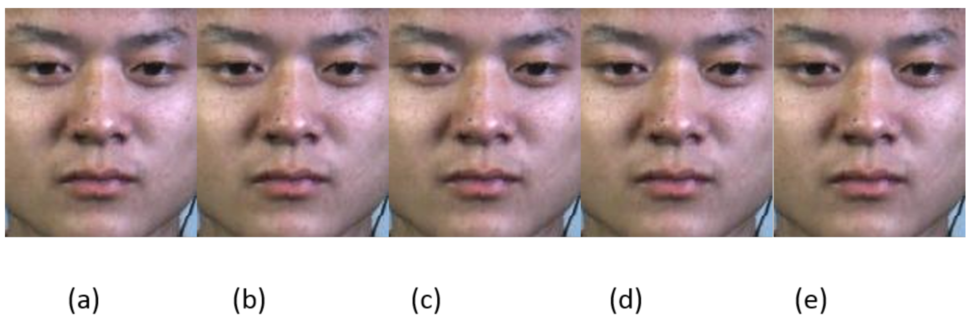

Figure 3.4: CASME II Happiness frame sequence [3] showing onset, offset and apex frames: (a) onset frame; (b) random frame between onset and apex frame; (c) apex frame; (d) random frame between apex and offset frame and (e) offset frame

3.3 Facial Action Coding System (FACS)

FACS was developed by Ekman and Friesen [71]. It was described as a method for measuring facial expressions based on the activities taking place in the underlying facial muscles. These facial movements are further classified into action units (AUs). Ekman and Friesen defined about 46 action units each of which corresponds to a particular activity in an underlying muscle or group of muscles. FACS has been used as a tool in classifying facial expressions into six universal emotions (disgust, anger, fear, sadness, happiness and surprise). However, labelling of spontaneous micro-expressions is different from that of facial expressions and non-spontaneous micro-expression databases. Owing to the fact that spontaneous micro-expressions are generated based on strong stimuli, it might be improper to classify micro-expressions into the six existing emotions as stated by Yan et al. in [5]. Therefore, the most recent micro-expression databases (CASME and CASME 2) labelled their emotions partly based on AUs and also on participants' self-report and content of video episodes [3]. Table 3.1 shows the difference between FACS coding for Cohn Kanade's facial expression database and CASME II micro-expression database.

Table 3.1: Differences between emotion labelling for facial expressions and micro-expressions database

Emotion	**Cohn Kanade [43]**	**CASME II [5]**
Surprise	AU1 + AU2 or AU5	AU1 + AU2 or AU25
Happiness	AU12	AU6 or AU12
Disgust	AU9, AU10 or AU4 + AU7	AU4 + AU7 or
Sadness	AU1 + AU4 + AU15 or AU11	AU1 + AU4
Repression	-	AU15 or AU17 or AU15 + AU17
Others	-	Other related facial movements

3.4 Feature Extraction Techniques

The process of extracting relevant features from data is critical to micro-expression recognition. In order to recognise micro-expressions accurately and effectively, we need to perform feature extraction. The next subsections present details on all feature extraction techniques used in this study with their mathematical formulations.

3.4.1 Local Binary Patterns

LBP was proposed by Ojala et al. [29] and their intention was to use it as a means of describing 2D textures of static images. The main idea of LBP is to compare the value of the centre pixel C of an image with the value of its neighbouring pixel P. If the centre pixel value is greater than the neighbouring pixel value, then, 0 is assigned, otherwise, 1 is assigned as described in Figure 3.5. This results into an 8 digit binary number comprising of 0s and 1s which is converted to decimal number and serves as the LBP value of the centre pixel. The mathematical description is described below:

Given a pixel p with intensity value v_p, radius r and N neighbouring pixels, a binary label is assigned to each of the neighbouring pixels. If the intensity value of

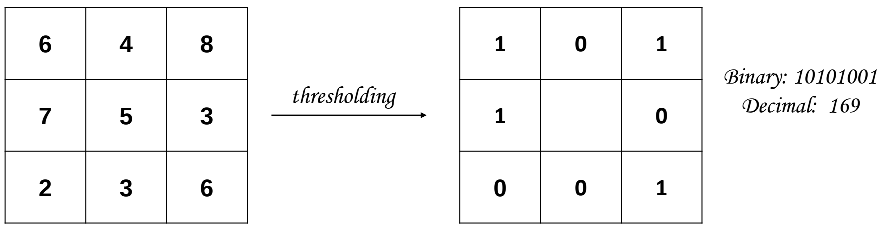

Figure 3.5: Thresholding on Local Binary Pattern

the given pixel is greater than that of the centre pixel, then 1 is assigned, otherwise, 0 is assigned to the pixel. LBP value is then given as:

$$LBP_{x_p,y_p} = \sum_{i=0}^{N-1} f(g_i - g_p)2^i \tag{3.1}$$

where x_p, y_p represents the co-ordinates of the centre pixel, g_p denotes the intensity value of the centre pixel and g_i is the intensity value of the i^{th} neighbouring pixel. 2^i is the weight that corresponds to the neighbouring pixel locations and $f(x)$ is a sign function defined as $f(x) = \{^{1,\ x\geq 0}_{0,\ x<0}\}$. The feature vector can then be derived by calculating the histogram of all the LBPs as described in (3.2).

$$H_i = \sum_{x=0} \sum_{y=0} I\{LBP(x,y) = i\}, (i = 0, ..., n-1) \tag{3.2}$$

where n represents the total number of labels produced and

$$I(X) = \{^{1,\ ifXistrue;}_{0,\ ifXisfalse.}$$

3.4.2 Local Binary Patterns on Three Orthogonal Planes

Local Binary Patterns on Three Orthogonal Planes was selected as a technique for feature extraction in this study because of its ability to extract temporal features from micro-expression samples. Ordinary LBP features were also extracted from apex and onset micro-expression frames and its performance compared with that of LBP-TOP.

LBP-TOP [4] is one of the spatio-temporal descriptors for dynamic textures (textures in motion) which was created in order to overcome the drawbacks of ordinary LBP. The major drawback of LBP is that it can only extract features from still images. LBP-TOP cannot be described without making reference to LBP and is described below.
LBP-TOP is one of the variants of ordinary LBPs and was proposed by Zhao et al. [4]. LBP-TOP was proposed as a result of the need to analyse textures that are time-dependent (i.e. videos).

For a given video with time length T, LBP value is calculated in three planes; XY, XT and YT, where XY provides spatial information while XT and YT supplies both spatial and temporal information about space-time transitions [4]. The LBP value for each of the planes is calculated using equation (3.1) and is concatenated into a single histogram which serves as the final feature vector. The corresponding LBP-TOP feature is given as

$$LBP - TOP_{P_{XY},P_{XT},P_{YT},R_{XY},R_{XT},R_{YT}}$$

while the histogram is described in equation (3.3), where P represents the number of neighbouring points and R represents the radius.

$$H(i) = \sum I\{f_j(x,y,t) = i\}, i = 0, ..., n_j - 1; j = 0, 1, 2. \tag{3.3}$$

where n_j represents the number of interest patterns in the jth plane and $f_j(x, y, t)$ represents the LBP code at pixel positions (x, y, t) along the jth plane. The final histogram is normalised using equation (3.4).

$$N_{i,j} = \frac{H_{i,j}}{\sum_{n_j-1}^{k=0} H_{k,j}} \tag{3.4}$$

Application of Uniform Local Binary Pattern

To reduce the dimensionality (feature vector length), uniform binning was introduced during the feature extraction process. This means that only important textures were put into consideration. The underlying principle behind uniform LBP is that, if a binary pattern contains at most two bitwise transitions (0-1 or 1-0), then that binary pattern is referred to as uniform, otherwise, if the binary pattern contains more than two bitwise transitions, then it is non-uniform [72]. For example, 01111100 has two (2) transitions from 0-1 and 1-0, and is therefore uniform, while 11010101 has six (6) transitions and is non-uniform pattern.

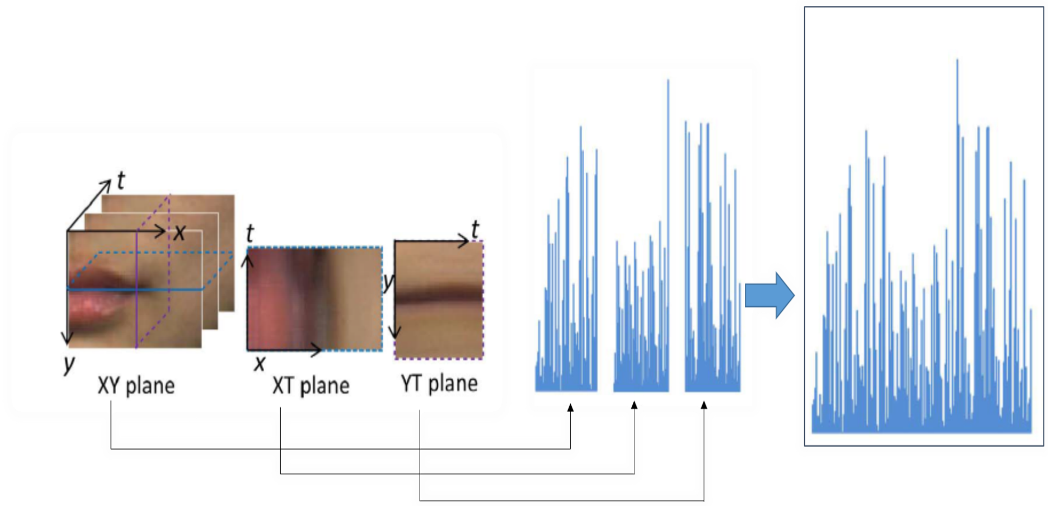

Figure 3.6: CASME II micro-expression sample showing XY, XT and YT planes with their corresponding histograms [3]

3.5 Machine Learning Algorithms

Machine learning is essential to automatic micro-expression recognition. It is categorised into supervised, un-supervised and semi-supervised. For supervised learning, there is always a set of training samples that are labelled with their corresponding target/output data. The algorithm learns from the dataset and is able to predict correctly based on the input that is given. For unsupervised learning, only the input data is provided while the target or output data is not given. The learning model tries to discover the patterns that are suitable for getting the corresponding class of the input data. Semi-supervised machine learning uses a few labelled data with many unlabelled data.

Some of the existing supervised learning algorithms for automatic micro-expression recognition include decision trees, neural networks, k nearest neighbours, SVM, ELM and random forest classifiers. For the purpose of this study, we employed the use of ELM and SVM motivated by their success from past studies. SVM has been widely used for both macro- and micro-expression recognition and for other classification tasks while ELM has not been used by many for the same purpose.

3.5.1 Support Vector Machines

Support Vector Machine is a supervised machine learning algorithm that is used for binary classification. It can also be used to perform regression tasks.

How does SVM work?

Considering that we have two a dataset with two separate classes that are linearly inseparable, SVM uses a hyperplane to separate the group of data into their appropriate classes. There could be more than a single hyperplane separating the classes and the one with the largest margin is chosen as the best/most correctly classified. Figure 3.7 gives a description of how classification is performed using SVM. The mathematical model is described as follows:

Consider a set of two-class problem given as $\{x_i, y_i\}_{i=1}^{N}$, where x_i and y_i represents input and output vectors respectively and $y_i \epsilon \{-1, +1\}$ is the class label of input x_i. SVM aims at finding an optimal decision boundary (hyperplane) that can classify all points correctly. The hyperplane is written as $w(x_i) + b = 0$ where w is the optimal set of weights and b represents the optimal bias. The hyperplane is optimised using minimising the optimal weight (3.4) subject to the constraints in equation (3.5).

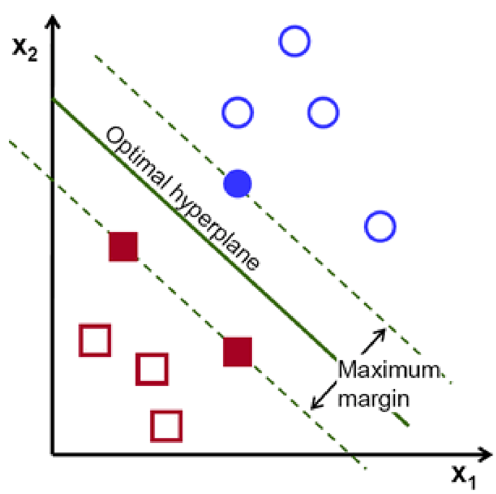

Figure 3.7: SVM Architecture [8]

$$Minimize\ \frac{1}{2}\|w\|^2 \tag{3.5}$$

$$Subject\ to\ y_i(w(x_i)+b) \geq 1 \qquad \forall i \tag{3.6}$$

Thus, Lagrangian multipliers α are introduced and result is shown in equation (3.7)

$$L_M \ = \ \frac{1}{2}\|w\|^2 \ - \ \sum_{i=1}^{n} \alpha_i y_i (x_i.w + b) + \sum_{i=1}^{n} \alpha_i \tag{3.7}$$

where $i = \{1, ..., n\}$. To minimise L_M with respect to w and b, its derivatives are completely taken away by setting the gradient of the Lagrangian multiplier to 0 as seen in equation (3.7)

$$\sum_{i=1}^{n} \alpha_i y_i = 0 \tag{3.8}$$

$$w = \sum_{i=1}^{n} \alpha_i y_i x_i \tag{3.9}$$

Thereafter, w is re-substituted onto equation (3.6) and results in:

$$L_P = \sum_{i} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i . x_j \tag{3.10}$$

To train the dataset, L_P is maximised using equations (3.7) and (3.8) as constraints. When there is a training point such that its Lagrangian multiplier $\alpha_i 0$, then such a point is called support vectors, hence the name support vector machine.

3.5.2 Extreme Learning Machine

Extreme Learning Machine is a learning algorithm for the single hidden layer feedforward neural networks (SLFN) proposed by Huang et al. [69]. This learning algorithm was proposed to overcome the drawbacks of traditional feed-forward neural networks. As stated by Huang et al. [69], one of the major drawbacks of traditional feed-forward neural network is their slow learning speed. Some of the advantages of ELM over other traditional learning algorithms of SLFN are stated in [52], [73].

- ELM does not require parameter tuning
- ELM has an extremely fast learning speed as compared with other learning algorithms such as back propagation (BP) algorithm

- ELM is very useful in training SLFNs with many non-differentiable activation functions.

According to Wang et al. [73], the most superior and impressive of these features is the fast training speed compared to other traditional learning algorithms. The mathematical model is described below:

For N distinct samples (x_i, t_i) where $x_i = (x_{i1}, x_{i2}, ..., x_{in})^T \in R^n$ and $t_i = (t_{i1}, t_{i2}, ..., t_{in})^T \in R^m$, SLFN can be modeled using equation (3.11)

$$\sum_{i=1}^{\tilde{N}} \beta_i g(w_i.x_j + b_i) = p_j, j = 1, ..., N \tag{3.11}$$

where $g(x)$ represents the activation function, $\tilde{N}$ represents the number of hidden neurons, the weight vector that connects the i^{th} hidden node and the input nodes is represented by $w_i = (w_{i1}, w_{i2}, ..., w_{im})$ and b_i represents the threshold of the hidden node. $\beta_i = (\beta_{i1}, \beta_{12}, ..., \beta_{im})^T$ is the weight vector connecting the i^{th} node and the output nodes while $w_i.x_j$ is defined as the inner product of w_i and x_j. Equation (3.10) can be re-written linearly as:

$$H\beta = T \tag{3.12}$$

where $H(w_1, ..., w_{\tilde{N}}, b_1, ..., b_{\tilde{N}}, x_1, ..., x_N) = \begin{bmatrix} g(w_1.x_1 + b_1) & \cdots & g(w_{\tilde{N}}.x_i + b_{\tilde{N}}) \\ \vdots & \ddots & \\ g(w_1.x_1 + b_1) & \cdots & g(w_{\tilde{N}}.x_N + b_{\tilde{N}}) \end{bmatrix}_{N\times\tilde{N}}$

,

$$\beta = \begin{bmatrix} {\beta_1}^T \\ \vdots \\ {\beta_{\tilde{N}}}^T \end{bmatrix}_{\tilde{N}\times M} \text{ and } T = \begin{bmatrix} {t_1}^T \\ \vdots \\ {t_{\tilde{N}}}^T \end{bmatrix}_{\tilde{N}\times M}$$

where H represents the hidden layer output matrix. According to the theorem of Huang et al. [74], it is assumed that since input weights w_i and hidden layer biases H are randomly generated, the output weight β can be determined by finding the minimum norm least square (LS) solution to the linear system $H\beta = T$. The LS solution is described in equation (3.13).

$$\hat{\beta} = H^{\dagger}T \tag{3.13}$$

where $H^\dagger$ is the Moore-Penrose generalised inverse of H. ELM algorithm can be summarised in three steps:

Given a training set $X = \{(x_i, t_i)|x_i \in R_n t_i \in R_m, i = 1, 2, ..., N\}$ with an activation function $g(x)$ and number of hidden neurons $\tilde{N}$

1. Randomly assign the input weights w_i and hidden layer bias $b_i = 1, ..., \tilde{N}$
2. Calculate the hidden layer output matrix H
3. Calculate the output weight

3.6 Multi-class Classification

Multi-class classification is defined as a combination of several binary classifiers. A binary classifier is such that contains instances with just two groups. It can be labelled either as positive or negative, yes or no, 1 or 0, 1 or -1. There are two major types of multi-class classification. This includes One Versus One (OVO) multi-class classification and One Versus All (OVA) multi-class classification. These methods are described in the next sections.

3.6.1 OVO Multi-class Classification

OVO multi-class classification deals with building $\frac{N(N-1)}{2}$ classifiers. For instance, given a 5 class problem, labelled as a, b, c, d and e, 10 pairs will derived as {a,b}, {a,c}, {a,d}, {a,e}, {b,c}, {b,d}, {b,e}, {c,d}, {c,e} and {d,e}. Each classifier is made to distinguish between each pair of classes resulting in a total of 10 classifiers.

3.6.2 OVA Multi-class Classification

In the case of OVA multi-class classification, for N classes we have N binary classifiers which compares a given class with the other $N - 1$ classes. The i^{th} binary classifier is trained with positive samples that belongs to class i and negative samples that belong to class $(i-1)$. To test a sample whose class is unknown, voting strategy is performed whereby the classifier that generates the highest output is said to be the winner and the class label is assigned to the given sample.

In this study, OVA multi-class classification was performed for all our experiments.

3.7 Measures for Performance Evaluation of Micro-expression Recognition

To measure the performance of the system, four measures were used which includes: accuracy, precision and recall, F-score measure and confusion matrix. Confusion matrix performs the function of describing the performance of the classifier based on the test sample data. Precision helps to measure how relevant the result of the classifier is while recall helps to measure the completeness of the classifier. Accuracy performs the function of measuring the correctness of the classifier. Other researchers who have worked on micro-expression recognition have often times used accuracy as the only means of measuring performance. The formulation for these measures are described with four parameters named as True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN). Table 3.2 gives the description of a confusion matrix for a binary classifier.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.14}$$

$$Precision = \frac{TP}{TP + FP} \tag{3.15}$$

$$Recall = \frac{TP}{TP + FN} \tag{3.16}$$

$$FScore = \frac{2 \times precision \times recall}{precision + recall} \tag{3.17}$$

Table 3.2: Confusion Matrix

	Positive	Negative
Positive	TP	FP
Negative	FN	TN

3.8 Summary

In this chapter, background details of computational methods used for this research were presented. Existing micro-expression databases were explained in details and a comparison was made between emotion labelling for both macro-expressions and micro-expressions. Mathematical description of feature extraction and classification methods used were also presented. Finally, measures used for evaluation of performance (precision, recall and F-score) were explained.

Chapter 4

Methodology

4.1 Introduction

This chapter presents details on all the methods used in this study. Methods are presented based on the hypothesis stated earlier. The micro-expression recognition process is made up of three main stages which includes data preparation, feature extraction and classification as illustrated in Figure 4.1. In the following sections, details on preparation of data samples used for experiments are provided. For feature extraction, LBP and LBP-TOP were used and classification was performed using SVM and ELM. Two data formats were used for the experiments conducted. The first data format involves the use of apex and onset frames for recognition while the second format involves the use of video sequences for micro-expression recognition.

4.2 Data Preparation

The data samples used for this study were acquired from CASME II database [3]. Micro-expression samples from CASME II were already pre-processed before being made available to the public. Some of the pre-processing carried out by the database owners includes face detection and registration, division of recorded videos into frames, labelling of samples with relevant action units and coding with onset, apex and offset frames. CASME II database has a total of 247 micro-expression samples with a sampling rate of 200fps. One of the samples from CASME II database is shown in Figure 4.2.

In this study, a total of 220 micro-expression samples were used for the first set of experiments (micro-expression recognition using LBP features extracted from apex and onset frames). For the second set of experiments (micro-expression recognition using LBP-TOP features extracted from image sequences of each sample), a total of 230 samples were used. The variation in the number of samples used for the two set of experiments is as a result of some samples whose coding did not include their correct apex, onset and offset labels. These samples without correct labels were left out of the total samples used for the first set of experiment, hence, we had 220 onset and apex frame samples. Since the second set of experiment used entire image sequences, a total of 230 samples were used. Details on the number of samples for each micro-expression class are presented in Table 4.1. Pre-processing involved conversion of frames from RGB into grey-scale images.

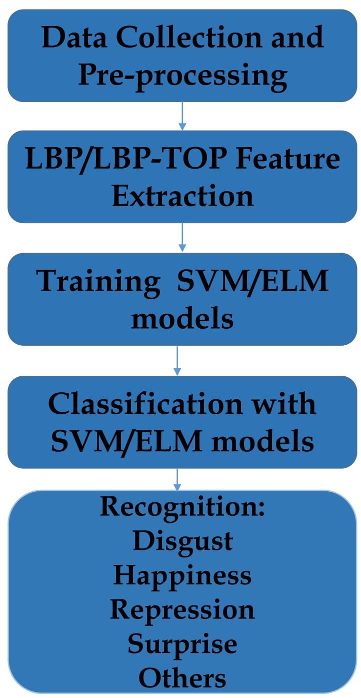

Figure 4.1: Architecture of Micro-expression Recognition System

Figure 4.2: Disgust sequence from CASME II [3] showing onset, offset and apex frames: (a) onset frame; (b-d) random frame between onset and apex frame; (e) apex frame; (f-g) random frame between apex and offset frame and (h) offset frame. (Size of each frame in the sequence = 144×176)

Table 4.1: Number of samples for each micro-expression class in CASME II and selected samples for each experiments

Class	**CASME II**	**No. of Selected Apex Frames**	**No. of Selected Image Sequences**
Disgust	60	57	59
Happiness	33	25	30
Repression	27	27	24
Surprise	25	23	25
Others	102	88	92
Total	**247**	**220**	**230**

4.3 Feature Extraction using Local Binary Pattern

The first set of experiments were performed using LBP on onset frames and apex frames. As presented in Table 4.1 above, a total of 220 samples that includes 57 disgust, 25 happiness, 27 repression, 25 surprise and 88 others samples were used for this experiment. Selection of these samples was done based on the labelling provided by the database owners. For each grey-scale image, LBP values were obtained by comparing the centre pixel values with all the neighbouring pixel values resulting in a 8-digit binary number converted into decimal. Thereafter, the histograms of these LBP values were calculated with intensity levels on y co-ordinates and pixel values (from 1 to 256) on x co-ordinates which resulted in a feature vector size of 1×256 for each sample. An illustration of LBP histogram for apex and onset frame samples from one of the subjects are shown in Figure 4.3 and Figure 4.4 respectively.

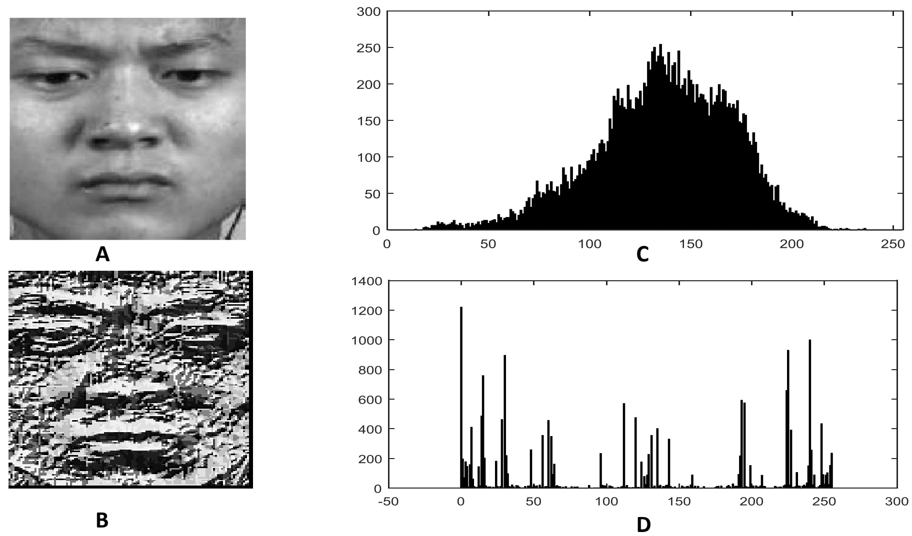

Figure 4.3: Apex frame for a disgust sample from CASME II [3] showing **A**: Gray-scale Image, **B**: Histogram of Grayscale Image, **C**: Local Binary Pattern and **D**: Histogram of Local Binary Pattern

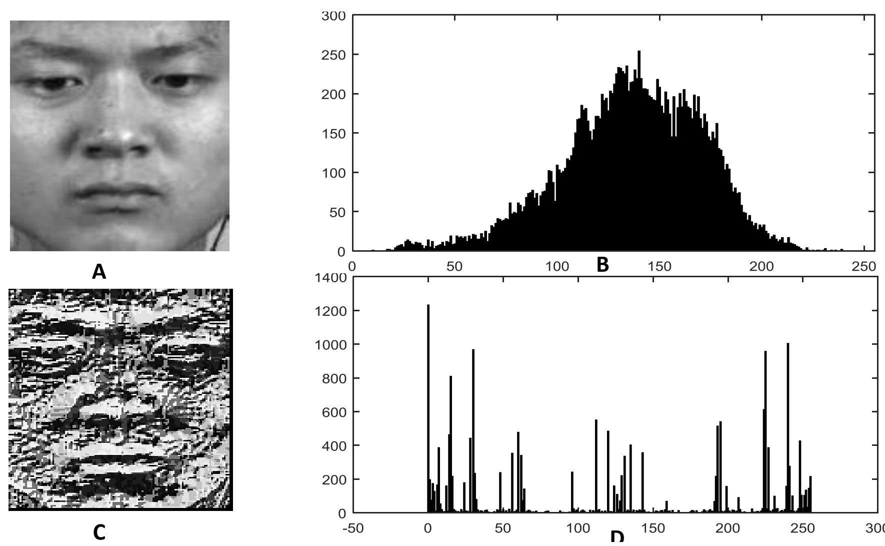

Figure 4.4: Onset frame for a disgust sample from CASME II [3] showing **A**: Gray-scale Image, **B**: Histogram of Grayscale Image, **C**: Local Binary Pattern and **D**: Histogram of Local Binary Pattern

4.4 Feature Extraction using Local Binary Pattern on Three Orthogonal Planes

The second experiment involved extraction of LBP-TOP features from entire image sequences. To extract LBP-TOP features, grey-scale image sequences were first read (i.e., micro-expression videos readily converted into frames of varying lengths). Thereafter, LBP-TOP features were extracted. In this experiment, radii values at x, y planes were varied between 1 and 4 while for t plane radii values were varied between 2 and 4. Number of neighbouring points for XY, XT and YT planes were set to 8 which gave a total of 2^8 patterns (i.e. 256 patterns) for each of the samples.

For each image sequence, LBP values were calculated in XY, XT and YT planes along with their respective histograms. The histogram derived for each of the three planes were concatenated and the result was a 3×256 feature vector for each image sequence. To reduce the length of the feature vector, uniform binning was applied to extract only uniform patterns from the 256 histogram bins. Application of uniform patterns reduced the length of the feature vector from 256 to 59. This resulted in a 3×59 feature vector for each image sequence which was converted into a single row vector of size 1×177. For all 230 samples used for these experiments, 230×177 features were used.

4.5 Classification Models

One Versus All classification was performed on both experiments using SVM and ELM models. Five micro-expression classes were used for all experiments (disgust, happiness, repression, surprise and others). In other words, we had a total of five classifiers. Training was performed using all samples belonging to each class as positive samples labelled as 1 while samples from the remaining four (4) classes were used as negative samples labelled as 0. The description of the labels are given in Table 4.2. In Section 4.5.1, training and optimisation of SVM is presented while in Section 4.5.2, training of ELM for recognition of micro-expressions is presented.

Table 4.2: Table showing positive and negative samples along with their given labels in parenthesis

Positive Samples	**Negative Samples**
Disgust (1)	Happiness (0) Repression (0) Surprise (0) Others (0)
Happiness (1)	Disgust (0) Repression (0) Surprise (0) Others (0)
Repression (1)	Happiness (0) Disgust (0) Surprise (0) Others (0)
Surprise (1)	Happiness (0) Disgust (0) Repression (0) Others (0)
Others (1)	Happiness (0) Disgust (0) Repression (0) Surprise (0)

4.5.1 Support Vector Machine based Model

Support vector machine based model was built for the purpose of classification. Selection of SVM as baseline model was motivated by its success in past micro-expression studies like [3], [23] and [45]. Five-fold cross validation was used to divide the samples into five subsets. Four training subsets and one test subset were generated at random five times for each of the five classifiers. SVM models were built using the two data formats that we had (apex and onset frames and image sequences). A total of 220 samples were used for apex frame experiment while 230 samples were used for experiments performed using image sequences. Details on training that was performed and how parameters were optimised are also presented in the next sections.

SVM Training on Apex and Onset Frames using LBP Features

Training SVM on apex and onset frames was performed by loading 1×177 feature vector for 220 apex and onset samples acquired after extraction of LBP features. Five fold cross validation was then performed to divide the samples into five independent subsets. Linear SVM kernel was used for training. An average training accuracy of 96.15% was achieved using LBP features from apex and onset frames as presented in Table 4.3.

SVM training on Image Sequences using LBP-TOP Features

To perform SVM training using image sequences, a similar procedure used for apex frame experiments was followed. Feature vector of size 1×177 for 230 image sequences acquired from LBP-TOP feature extraction were loaded into each classifier. Thereafter, five-fold cross validation was performed to divide the samples into five independent subsets. Optimisation of training model was performed by recording training accuracy at varying LBP-TOP radii values in x, y and t planes.

Table 4.3: Training accuracy of SVM using apex and onset frames on LBP features using linear kernel

Class	Apex Frames	Onset Frames	Average
Disgust	94.99	95.11	95.05
Happiness	97.62	97.62	97.62
Repression	97.62	97.61	97.62
Surprise	98.12	97.96	98.04
Others	93.16	91.71	92.44
Average	**96.30**	**96.00**	**96.15**

These values were varied between 1 and 4 for x and y planes while variation for t plane was between 2 and 4. Training results showed that the highest average training accuracy (94.00%) was achieved at $R_x = 1$, $R_y = 1$ and $R_t = 3$ as presented in Table 4.4. Training results as presented in the Table 4.4 revealed that there was no more increase in average training accuracy as from radii values $R_x = 3$, $R_y = 3$ and $R_t = 2$ to $R_x = 4$, $R_y = 4$ and $R_t = 4$.

Table 4.4: Optimisation of LBP-TOP radii values using SVM. Training accuracy was recorded in percentage (%)

R_x, R_y, R_t	**Surp.**	**Disg.**	**Happ.**	**Others**	**Repr.**	**Average**
1, 1, 2	95.22	86.96	87.83	91.30	97.39	91.74
1, 1, 3	**89.15**	**91.95**	**93.80**	**97.61**	**97.50**	**94.00**
2, 2, 2	88.70	93.04	95.65	93.04	88.70	91.82
2, 2, 3	88.26	89.57	96.52	93.92	95.65	92.78
2, 2, 4	88.70	89.57	97.83	93.48	95.65	93.05
3, 3, 2	86.09	88.70	97.39	85.65	89.13	89.39
3, 3, 3	86.09	88.69	97.39	85.22	90.44	89.57
3, 3, 4	86.52	89.13	96.96	88.70	87.82	89.83
4, 4, 2	85.22	87.83	96.96	86.52	87.83	88.87
4, 4, 3	83.04	83.05	85.22	92.61	94.78	87.74
4, 4, 4	83.91	85.22	85.65	94.78	94.78	88.86

4.5.2 Extreme Learning Machine Model

Extreme Learning Machine (ELM) models using LBP and LBP-TOP were built by using five-fold cross validation to partition the data samples into five subsets. ELM model was selected as our second model because of its learning speed and more effective training ability [69]. Training this model was performed using four subsets out of the five subsets while the remaining one subset was reserved for testing purpose. This process was repeated five times and the average training accuracy was calculated. Training results in terms of training time and training accuracy were compared with training results for both SVM (baseline model) and ELM. Results from Training time are presented in Chapter 5.

ELM Training on Apex and Onset Frames using LBP Features

Training accuracy was recorded for varying number of hidden neurons (between 50 and 400) at an interval of 50 using LBP features as shown in Table 4.5. It was discovered that training accuracy increased with increasing number of hidden neurons. There was a constant 100% training accuracy from 200 hidden neurons for the five classifiers as shown in Figure 4.5. This informed the decision to choose 300 as number of hidden neurons during validation of the model.

Table 4.5: Training accuracy with varying number of hidden neurons for the five classifiers using LBP features

Hidden Neurons	Disgust	Happiness	Repression	Surprise	Others
50	83.53	90.34	89.66	92.39	75.57
100	93.18	97.39	97.84	98.07	90.23
150	99.77	100	100	100	99.2
200	100	100	100.00	100	100
250	100	100	100	100	100
300	**100**	**100**	**100**	**100**	**100**
350	100	100	100	100	100
400	100	100	100.00	100	100

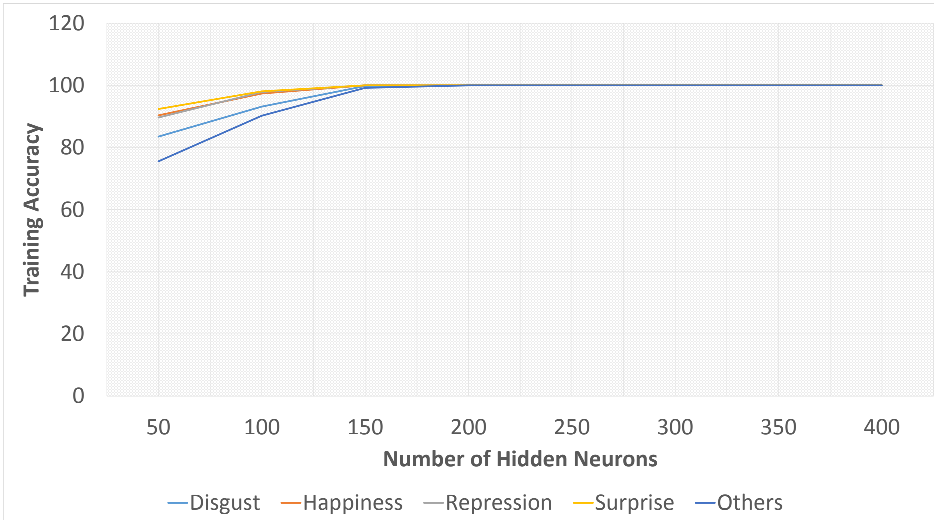

Figure 4.5: Optimisation of hidden neurons using LBP features

ELM Training on Image Sequences using LBP-TOP Features

Optimisation of parameters was performed by recording training accuracy at varying LBP-TOP radii values in x, y and t planes. These values were varied between 1 and 4 for x and y planes while variation for t plane was between 2 and 4 as presented in Table 4.6. Training results showed that the highest average training accuracy (97.57%) using ELM was achieved at $R_x = 1$, $R_y = 1$ and $R_t = 3$.

To select optimal number of hidden neurons, average training accuracy was recorded between 50 and 500 at an interval of 50 for each classifier. The highest average training accuracy (97.54%) was achieved at 200 hidden neurons as presented in Table 4.7.

Table 4.6: Optimisation of LBP-TOP radii values with ELM model. Accuracy was recorded in percentage (%)

R_x, R_y, R_t	**Disg.**	**Happ.**	**Rep.**	**Surp.**	**Others**	**Average**
1, 1, 2	92.07	91.74	99.67	99.02	99.24	96.35
1, 1, 3	**94.13**	**94.35**	**99.67**	**99.68**	**100**	**97.57**
1, 1, 4	98.37	98.37	92.17	92.66	100	96.24
2, 2, 2	94.78	100	96.85	91.74	100	96.67
2, 2, 3	94.13	99.57	97.72	91.41	100	96.57
2, 2, 4	93.8	100	97.18	91.52	100	96.5
3, 3, 2	93.59	99.68	93.58	90.65	93.04	94.11
3, 3, 3	94.35	99.57	93.48	90.44	93.04	94.17
3, 3, 4	94.13	99.57	93.91	90.65	93.15	94.28
4, 4, 2	93.69	99.57	93.59	91.09	92.72	94.13
4, 4, 3	92.07	92.39	96.96	90.33	99.68	94.28
4, 4, 4	92.5	92.72	97.17	90.65	100	94.61

Table 4.7: Training accuracy at varying number of hidden neurons for the five classifiers using LBP-TOP features

Hidden Neurons	**Disg.**	**Happ.**	**Repr.**	**Surp.**	**Others**	**Average**
50	83.48	88.48	85.11	94.89	92.82	88.96
100	90.65	91.74	92.39	98.48	98.59	94.37
150	94.02	94.35	99.67	99.68	99.68	97.48
200	**94.13**	**94.24**	**100**	**99.67**	**99.68**	**97.54**
250	93.81	93.91	100	99.67	99.68	97.52
300	93.70	93.91	100	99.68	99.68	97.39
350	94.13	94.13	100	99.67	99.68	97.39
400	94.13	94.02	100	99.68	99.68	97.5
450	93.48	94.02	100	99.67	99.57	97.35
500	93.8	93.91	100	99.57	99.57	97.37

4.6 Final Model Architecture

In the next chapter (Chapter 5) testing performance based on optimal training results achieved from the SVM and ELM models will be presented. Summary of details on resulting ELM model using both LBP and LBP-TOP features are presented in Table 4.8. These details include type of model (regression/classification), number of input and output weights, number of input and output neurons and labels. Optimal LBP-TOP radii values in x, y and t planes were 1, 1 and 3 respectively based on highest average training accuracy of 94.00% using SVM model as presented in Table 4.4. Optimal LBP-TOP training accuracy was also achieved at $R_x = 1$, $R_y = 1$ and $R_t = 3$.

For ELM-based model training using LBP features from apex and onset frames, optimal average training accuracy of 100% was achieved with 300 hidden neurons. (See Table 4.5). This selection was made based on training accuracy recorded for apex frames at varying number of hidden neurons between 50 and 400 at an interval of 50. For ELM-based model using image sequences and LBP-TOP features, 200 hidden neurons were optimal with an average training accuracy of 97.54%. (Table 4.7).

Table 4.8: Details on ELM model built for each classifiers

Model details	**Apex Frames**	**Frame Sequence**
Number of Samples	220	230
Feature Vector Size	1×256	1×177
ELM Type	Classification	Classification
Activation Function	Sigmoid	Sigmoid
Number of Input Neurons	256	177
Number of Output Neurons	2	2
Label	1 (Positive Samples),	1 (Positive Samples),
	0 (Negative Samples)	0 (Negative Samples)

4.7 Summary

In this chapter, details on data samples, their labellings and preparation for the experiments were presented. The process involved in extraction of both LBP features from apex frames and temporal (LBP-TOP) features from image sequences were presented. Final feature vectors used as input for the classifier were derived from the feature extraction process. Details on both SVM and ELM models were presented and selection of optimal parameters based on training accuracy for SVM and ELM model was performed.

Chapter 5

Experimental Results, Analysis and Discussion

5.1 Introduction

In this chapter, results obtained from all experiments are presented, analysed and discussed. The results include test performance for SVM and ELM models using LBP features from apex and onset micro-expression frames. It also includes test performance for SVM and ELM models using LBP-TOP features from micro-expression image sequences. Based on the results, comparative analysis was carried out to show which of the feature extraction and classification models performed better. Results from the study are also discussed in relation to results obtained from reviewed literature.

5.2 Micro-expression recognition using SVM

Classification process involved loading grey-scale image samples into each of the five classifiers. The grey-scale images were read and features (LBP/LBP-TOP) extracted from them. Feature extraction process resulted in a feature vector size of 1×256 and 1×177 for LBP and LBP-TOP features respectively. For each of the five classifiers, testing was performed via a one versus all classification approach by assigning label 1 to positive samples and label 0 to negative samples using the trained SVM model.

5.2.1 SVM-based Classification using LBP Features

For this experiment, a total of 220 samples (57 disgust, 25 happiness, 27 repression, 25 surprise and 88 others) were divided into training and test sets using five-fold cross validation. Number of training samples used for each classifier was within the range of 174 and 176 while number of test samples was within the range of 44 and 46 samples. After LBP features were extracted and training performed, classification of the micro-expression samples was performed on the test subsets for each of the five classifiers. Test results were recorded on both onset and apex frames. The mean accuracy, precision, recall and F-score performances were recorded for each classifier as presented in Tables 5.1, 5.2, 5.3 and 5.4 respectively. Overall average testing accuracy, precision, recall and F-score results achieved were 94%, 82.61%, 89.31% and 85.67% respectively using SVM model.

Table 5.1: Testing accuracy (%) of SVM using LBP features from apex and onset frames

Class	Apex Frames	Onset Frames	Average
Disgust	90.88	90.46	90.67
Happiness	96.36	96.36	96.36
Repression	95.42	95	95.21
Surprise	96.37	95.92	96.15
Others	90.47	92.76	91.62
Average	**93.9**	**94.1**	**94.0**

Table 5.2: Precision Results (%) of SVM using LBP features from apex and onset frames

Class	Apex Frames	Onset Frames	Average
Disgust	81.82	81.82	81.82
Happiness	80	80	80
Repression	80	80	80
Surprise	83.33	83.33	83.33
Others	87.5	82.68	87.87
Average	**82.53**	**82.68**	**82.61**

Table 5.3: Recall Results (%) of SVM using LBP features from apex and onset frames

Class	Apex Frames	Onset Frames	Average
Disgust	81.82	90	85.91
Happiness	80	80	80
Repression	100	80	90
Surprise	100	100	100
Others	87.5	93.75	90.63
Average	**89.86**	**88.75**	**89.31**

Table 5.4: F-score Results (%) of SVM using LBP features from apex and onset frames

Class	Apex Frames	Onset Frames	Average
Disgust	81.82	85.72	83.77
Happiness	80	80	80
Repression	88.89	80	84.45
Surprise	90.91	90.91	90.91
Others	87.5	90.91	89.21
Average	**85.82**	**85.51**	**85.67**

5.2.2 SVM-based Classification using LBP-TOP Features

For this experiment, a total of 230 image sequence samples (59 disgust, 30 happiness, 24 repression, 25 surprise and 92 others) were divided into training and test sets using a five-fold cross validation. Since one vs all classification was used, we had a total of five classifiers. Number of training samples used for each classifier was within the range of 184 and 186 while number of test samples was within the range of 44 and 46 samples. After LBP-TOP features were extracted and training performed, classification was performed on the test subsets using feature vectors from LBP-TOP and trained SVM model for each of the five classifiers. Test results were recorded using test subsets (samples) for each of the five classifiers. Performance of the classifiers was evaluated using accuracy, precision, recall and F-score in Table 5.5. Average performances for the five classifiers were also recorded. An average precision of 92.85% was achieved for the class classified which shows the exactness of the classifiers while average recall of 93.94% was achieved which shows the completeness of the classifiers.

Table 5.5: Test results for LBP-TOP with SVM model including accuracy, precision, recall and F-score for five classifiers (measured in %)with radius x, y and t = 1, 1 and 3 respectively

Measure	**Disg.**	**Happ.**	**Repr.**	**Surp.**	**Others**	**Average**
Accuracy	91.3	95.74	100	100	95.65	96.54
Precision	78.57	85.71	100	100	100	92.85
Recall	96.88	83.33	100	100	89.47	93.94
F-Score	86.77	42.86	100	100	94.44	84.81

5.2.3 Comparative Results from LBP and LBP-TOP Features using SVM Model

Average testing accuracy from apex and onset frames using LBP features were compared with average testing accuracy from image sequences using LBP-TOP. The comparison is illustrated using a column chart as shown in Figure 5.1. In terms of their testing accuracy, results showed that LBP-TOP features with SVM-based model had higher average test performance when compared with the use of LBP features using SVM-based model. Average testing accuracy for LBP on apex frames was 94% while average testing accuracy for LBP-TOP was 96.54%.

Using these average values, a t-test was conducted to show if the use of apex and onset LBP features is significant than the use of LBP-TOP features from image sequences. T-test results showed a probability value (p-value) of 0.2438 at a confidence interval of 95%. This means that there was no significant difference in their average mean. Based on this t-test, we cannot ascertain which feature extraction approach is more efficient. We can suggest that LBP-TOP features with SVM-based model results in a higher performance when compared with LBP features with SVM model.

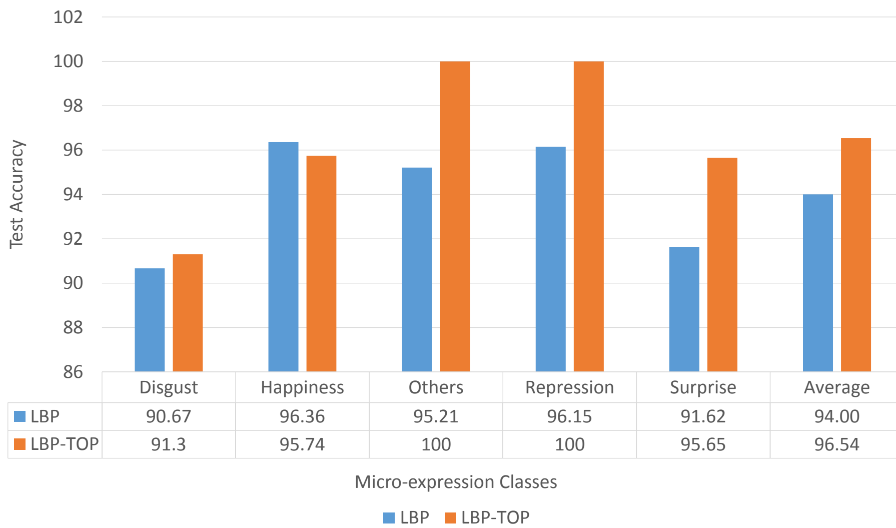

Figure 5.1: Comparative results for LBP on apex frames and LBP-TOP on entire image sequence using SVM

5.3 Micro-expression Recognition using ELM

Classification process involved loading grey-scale image samples into the classifier. Grey-scale images were read and features (LBP/LBP-TOP) were extracted from them. Feature extraction process resulted in a feature vector size of 1×256 and 1×177 for LBP and LBP-TOP respectively. For each of the five classifiers, testing the ELM model was conducted through a one versus all classification by assigning 1 to positive samples and label 0 to negative samples.

5.3.1 ELM-based Classification on Apex Frames using LBP Features

A total of 220 onset and apex samples (57 disgust, 25 happiness, 27 repression, 25 surprise and 88 others) were divided into training and test sets using a five-fold cross validation. Number of training samples used for each classifier varied between the range of 174 and 176 while number of test samples was within the range of 44 and 46 samples. After LBP features were extracted and training performed, classification of the micro-expression samples was performed on the test subsets for each of the five classifiers using the trained ELM model. Test performances were recorded on both onset and apex frames. The mean accuracy, precision, recall and F-score from the test results were recorded for each classifier. Results are presented in Tables 5.6, 5.7, 5.8 and 5.9. Overall average testing accuracy, precision, recall and F-score performances achieved were 94.08%, 87.86%, 90.99% and 89.01% respectively using SVM model.

Table 5.6: Testing accuracy for LBP on apex frames and onset frames using ELM model with 300 hidden neurons

Class	**Apex Frames (%)**	**Onset Frames (%)**	**Average**
Disgust	91.75	91.74	91.75
Happiness	94.54	95.91	95.23
Repression	95.03	94.55	94.79
Surprise	97.74	97.30	97.52
Others	91.38	90.85	91.12
Average	**94.09**	**94.07**	**94.08**

Table 5.7: Precision Results for LBP on apex frames and onset frames using ELM model with 300 hidden neurons

Class	**Apex Frames (%)**	**Onset Frames (%)**	**Average**
Disgust	81.82	81.82	81.82
Happiness	80	100	90
Repression	80	80	80
Surprise	100	100	100
Others	87.5	87.5	87.5
Average	**85.86**	**89.86**	**87.86**

Table 5.8: Recall Results for LBP on apex frames and onset frames using ELM model with 300 hidden neurons

Class	**Apex Frames (%)**	**Onset Frames (%)**	**Average**
Disgust	100	100	100
Happiness	100	80	90
Repression	80	80	80
Surprise	100	100	100
Others	87.5	82.35	84.93
Average	**93.5**	**88.47**	**90.99**

Table 5.9: F-score Results for LBP on apex frames and onset frames using ELM model with 300 hidden neurons

Class	**Apex Frames (%)**	**Onset Frames (%)**	**Average**
Disgust	90	90	90
Happiness	88.89	88.89	88.89
Repression	80	80	80
Surprise	100	100	100
Others	87.5	84.85	86.18
Average	**89.28**	**88.75**	**89.01**

5.3.2 ELM-based Classification on Image Sequences using LBP-TOP Features

A total of 230 image sequence samples (59 disgust, 30 happiness, 24 repression, 25 surprise and 92 others) were divided into training and test sets using a five-fold cross validation. Number of training samples used for each classifier was within the range of 184 and 186 while number of test samples was within the range of 44 and 46 samples. After LBP-TOP features were extracted and training performed, classification was performed on the test subsets using feature vectors from LBP-TOP and trained ELM model for each of the five classifiers.

Test results were recorded using test subsets for each of the five classifiers. Performance of the classifiers was evaluated using accuracy, precision, recall and F-score in Table 5.10. Average performance of the five classifiers was also recorded. Performances of the classifiers were evaluated using accuracy, precision, recall and

F-score as presented in Table 5.10. Confusion matrices for the five classifiers are presented in Figure 5.2. An average precision of 93.51% was achieved for the class classified which shows the exactness of the classifiers while average recall of 94.66% was achieved which shows the completeness of the classifiers.

Table 5.10: Test results for LBP-TOP with ELM model showing accuracy, precision, recall and F-score for five classifiers (measured in %)with radius x,y and t = 1, 1 and 3 respectively, 8 neighbouring points and 200 hidden neurons

Measure	**Disg.**	**Happ.**	**Repr.**	**Surp.**	**Others**	**Average**
Accuracy	95.65	95.65	100	97.83	98.71	97.57
Precision	91.67	83.33	100	97.56	95	93.51
Recall	91.67	83.33	100	100	100	94.66
F-Score	91.67	83.33	100	98.76	97.44	94.24

	Predicted (1)	Predicted (0)
Actual (1)	11	1
Actual (0)	1	33

A

	Predicted (1)	Predicted (0)
Actual (1)	5	1
Actual (0)	1	39

B

	Predicted (1)	Predicted (0)
Actual (1)	5	0
Actual (0)	0	41

C

	Predicted (1)	Predicted (0)
Actual (1)	5	1
Actual (0)	0	41

D

	Predicted (1)	Predicted (0)
Actual (1)	19	1
Actual (0)	0	27

E

Figure 5.2: Confusion matrices for five classifiers used; labelled as A,B,C,D and E representing disgust, happiness, repression, surprise and others respectively

5.3.3 Comparing Test Results from LBP and LBP-TOP Features using ELM

Average testing accuracy from apex and onset frames using LBP features was compared with average testing accuracy from image sequences using LBP-TOP. The comparison was based on test results from ELM model. This is illustrated using a column chart (Figure 5.3). Results showed that an average testing accuracy for LBP on apex frames using ELM was 94.08% while average testing accuracy for LBP-TOP on temporal data using ELM was 97.57%. Using these average values, a t-test was conducted to show if the use of LBP-TOP features on image sequences is significantly better than using apex and onset LBP features. T-test results gave a p-value of 0.0466 at a confidence interval of 95%. This means that there was a significant difference in their average mean. Based on the t-test results, we

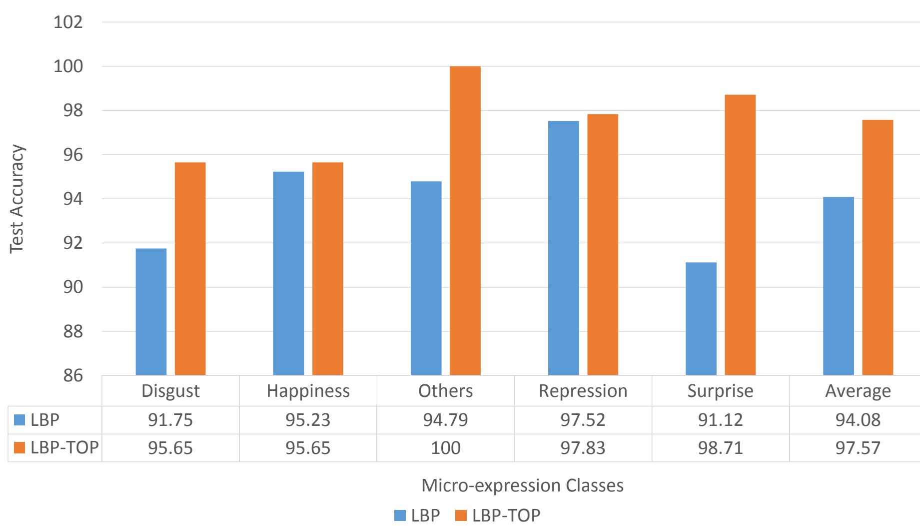

Figure 5.3: Comparative results for LBP on apex frames and LBP-TOP on entire image sequence using ELM

can deduce that LBP-TOP features on micro-expression image sequences is more relevant when compared with using LBP on apex and onset frames only using ELM model. As opposed to an assertion made by Liong et al. [36] and [37] that apex frames alone is enough to effectively recognize micro-expressions, results from this study showed that static feature extraction using apex frames only might not be effective for micro-expression recognition when ELM model is used. However, with SVM models we were not able to fully ascertain this fact. Although, based on the testing accuracy LBP-TOP outperformed LBP. This partly answered one of our research questions on the use of temporal feature extraction for effective micro-expression recognition.

5.3.4 Comparative Test Results for SVM and ELM using LBP-TOP features

A comparison was made between SVM and ELM models (uisng LBP-TOP features) based on their average testing accuracy and F-score. LBP-TOP features with SVM model yielded an average testing accuracy and F-score of 96.54% and 84.81% respectively while LBP-TOP features with ELM model yielded an average accuracy and F-score of 97.57% and 94.24% respectively. Results show that ELM outperformed SVM in terms of average testing accuracy and F-score measures for the five classifiers as seen in Figure 5.4 and Figure 5.5.

Based on this comparative analysis, we cannot draw a conclusion that SVM model is actually better than ELM. For this reason, a significance test was performed to check if SVM's average test accuracy and F-score is significantly different from ELM's test accuracy and F-score. T-test results for accuracy showed a p-value of

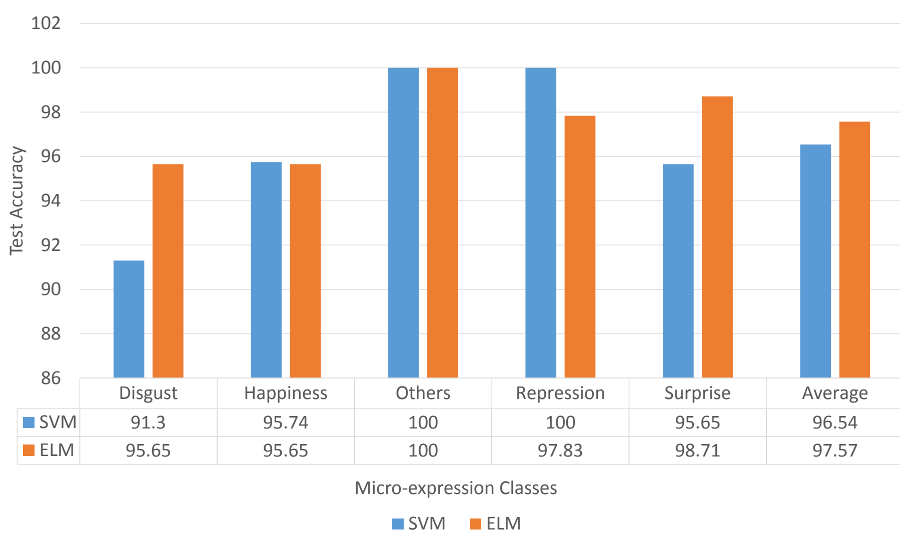

Figure 5.4: Comparative results for SVM and ELM using LBP-TOP features

0.5903 at 95% confidence interval. T-test results for F-score also gave a p-value of 0.4243 at 95% confidence interval. This means that there was no significant difference between these two average means. Hence, we cannot conclude that ELM

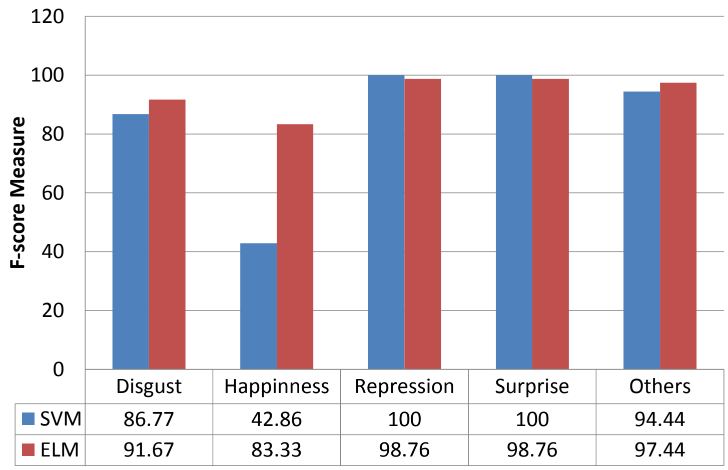

Figure 5.5: Comparative results for SVM and ELM F-score performance using LBP-TOP features

is actually better than SVM. In terms of learning speed ELM was found to learn faster than SVM. This is further discussed in the next section.

5.3.5 Comparative Results on SVM and ELM Learning Speed

In the previous experiments, a conclusion could not be drawn to show that ELM performs better than SVM in terms of recognition rates. The major difference that was identified between these models is their learning speed. To ascertain this, training times for each class of micro-expression was recorded and the average training time for the five classes of micro-expressions calculated. Results were recorded using both SVM and ELM. Average training time using SVM was 0.3405 seconds while ELM had an average training time of 0.0499 seconds. Hence, we can deduce that ELM learns faster than SVM. The difference between the training time of SVM and ELM for the five classes of micro-expressions using LBP-TOP features is presented in Table 5.11.

Table 5.11: Training time (seconds) of SVM and ELM using LBP-TOP features

Model	Disg.	Happ.	Repr.	Surp.	Others	Average
SVM	0.2808	0.3806	0.2434	0.2995	0.3182	0.3405
ELM	0.0468	0.0499	0.0593	0.0530	0.0406	0.0499

5.4 Our Study Vs Past Studies

Results obtained from this study were compared with results from past related studies. Table 5.12 and Table 5.13 shows the comparative results using apex/static frames and image sequences respectively. A similar study was carried out by Yan et al. [3] while evaluating the performance of their micro-expression algorithm using LBP-TOP for feature extraction and SVM for classification. In their work, images were divided into blocks before applying LBP-TOP feature extraction on each block which resulted in an accuracy of 63.41%.

In this study images were used in holistic manner which means that images were not divided into blocks before extraction of features to reduce computational complexity as expressed by Guo et al. [45].

Furthermore, results from LBP-TOP with SVM model gave an average testing accuracy of 96.54% while results from LBP-TOP features with ELM model gave an average testing accuracy of 97.57%. These two results showed higher recognition performances when compared with the study in [3]. From the study of Guo et al. [56], CBP-TOP, a variant of LBP which compares the centre pixel with pairs of neighbouring pixels was used for feature extraction. Classification was performed using ELM and an average test performance of 82.07% was achieved on the five classes of micro-expression used. Our study outperforms Guo et al's [56] study in terms of recognition accuracy.

From our results, we can also deduce that recognition of micro-expressions using LBP-TOP features (holistic) for SVM and ELM on CASME II database outperformed other methods. Even though the difference in performance of SVM and ELM was not statistically significant, ELM provided a faster learning model when compared with SVM. The learning speed of ELM was believed to contribute to the overall performance of the micro-expression recognition process.

Table 5.12: Comparative results with past studies that used static features

Author(s)	**Feature Ext.**	**Classification**	**Database**	**Performance**
Wu et al. [26]	Gabor Filters	GentleSVM	Cohn Kanade	85.42% Accuracy
Lyons et al. [36]	LBP & Bi-WOOF	SVM	SMIC & CASME II	61% F-score for SMIC and 62% for CASME II
Lyons et al. [37]	CLM & LBP	None	CASME II	Over 20% improvement when compared with baseline method used in [32]
Our Study	**LBP**	**ELM**	**CASME II**	**94.08% Accuracy, 85.67% F-score**
Our Study	**LBP**	**SVM**	**CASME II**	**94.14% Accuracy, 89.01% F-score**

Table 5.13: Comparative results with past studies that used temporal features

Author(s)	**Feature Ext.**	**Classification**	**Database**	**Performance (%)**
Wang et al. [52]	DTSA	ELM	CASME	46.90
Yan et al. [3]	LBP-TOP	SVM	CASME II	63.41
Guo et al. [45]	LBP-TOP	NN	SMIC	65.83
Guo et al. [56]	CBP-TOP	ELM	CASME	82.07
Our Study	**LBP-TOP**	**SVM**	**CASME II**	**96.54 Accuracy, 92.85 Precision, 93.94 Recall & 84.81 F-score**
Our Study	**LBP-TOP**	**ELM**	**CASME II**	**97.57 Accuracy, 93.51 Precision, 94.66 Recall & 94.24 F-score**

5.5 Summary

This chapter presents the test results achieved using the two trained models (SVM and ELM). Test results in the form of accuracy, precision, recall and F-score were calculated and recorded for all experiments. Comparative analysis was carried out to show the difference in performance of LBP/LBP-TOP and SVM/ELM. Results were also discussed in relation to past related literature.

Chapter 6

Conclusion and Future Work

6.1 Introduction

This chapter seeks to present a summary of all that this study entails. It also summarises how the research questions asked have been answered. Limitations encountered are also discussed and future areas of research presented based on these limitations.

6.2 Conclusion

This study has been able to show that it is possible to recognise micro-expressions automatically and achieve promising results through the use of temporal feature extraction technique (LBP-TOP) and a machine learning algorithm with an efficient and very fast learning speed (ELM). Two data formats were used for the experiments. The first format includes onset and apex frames extracted from CASME II micro-expression samples. The second data format include temporal image sequences from CASME II micro-expression samples.

The first set of experiments were conducted by extracting LBP features from onset and apex frame samples using SVM and ELM for classification. The second set of experiments were conducted by extracting LBP-TOP features from micro-expression image sequence samples using SVM and ELM for classification. Results showed that LBP-TOP features with SVM had a higher average testing accuracy of 96.54% when compared with LBP features and ELM which resulted in an average testing accuracy of 94.00%. However, the difference in their average accuracy was not statistically significant.

Results also showed that using LBP-TOP features from image sequence samples with ELM had a higher average test accuracy (97.57%) while using LBP features from apex and onset frames gave average test accuracy of 94.08% using ELM. An independent t-test was conducted on these two average results from LBP with ELM and LBP-TOP with ELM and a p-value of 0.0466 was achieved with $p < 0.05$ confidence interval. This t-test result suggests that there is a significant difference between their average means and hence we say LBP-TOP features with ELM is better than LBP features with ELM for micro-expression recognition.

Since the two models resulted in high test performance using LBP-TOP features, an independent t-test was performed on test results for LBP-TOP features with SVM and LBP-TOP features with ELM. The purpose was to check which machine learning algorithm performed better for micro-expression classification. T-test results showed that there was no significant difference between these two results. Hence, we cannot conclude that one algorithm performed better than the other. One advantage of ELM over SVM as stated earlier is their fast learning speed [69], [74]. Results from average training time for ELM outperformed results from SVM training time using LBP-TOP features.

Based on this results, we were able to draw the following conclusions and answer our research questions as stated below:

1. Recognition of micro-expression using temporal features is more effective than recognition of micro-expression using static features. However, this depends on the classification model used and the number of samples.

2. Supervised machine learning algorithms (SVM and ELM) have the ability to successfully allocate micro-expressions into their appropriate classes.

3. Exteme learning machine algorithm learns faster than SVM.

6.3 Future Work

Future work should include employing a better means of reducing LBP-TOP feature vector size and optimising the features (feature selection). In this study, we reduced the feature vector size from 1×256 to 1×177 by applying uniform patterns. We were not sure whether this reduction to 177 features were efficient in improving classification. Hence, we propose the use of optimisation algorithms such as genetic algorithm or ant colony in the future. This will help to optimise the features used for micro-expression recognition and hence improve performance as identified in [75]. The use of different kernels for SVMs and different activation functions for ELMs is also an area of interest that can be further studied.

Bibliography

[1] I. P. Adegun and H. B. Vadapalli, "Automatic recognition of micro-expressions using local binary patterns on three orthogonal planes and extreme learning machine," in *Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech), 2016*. IEEE, 2016, pp. 1–5.

[2] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with Gabor Wavelets," in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*. IEEE, 1998, pp. 200–205.

[3] W.-J. Yan, X. Li, S.-J. Wang, G. Zhao, Y.-J. Liu, Y.-H. Chen, and X. Fu, "CASME II: An improved spontaneous micro-expression database and the baseline evaluation," *PloS one*, vol. 9, no. 1, p. e86041, 2014.

[4] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 6, pp. 915–928, 2007.

[5] W.-J. Yan, S.-J. Wang, Y.-J. Liu, Q. Wu, and X. Fu, "For micro-expression recognition: Database and suggestions," *Neurocomputing*, vol. 136, pp. 82–87, 2014.

[6] X. Li, T. Pfister, X. Huang, G. Zhao, and M. Pietikainen, "A spontaneous micro-expression database: Inducement, collection and baseline," in *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*. IEEE, 2013, pp. 1–6.

[7] W.-J. Yan, Q. Wu, Y.-J. Liu, S.-J. Wang, and X. Fu, "CASME database: A dataset of spontaneous micro-expressions collected from neutralized faces," in *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*. IEEE, 2013, pp. 1–7.

[8] "Introduction to Support Vector Machines," http://docs.opencv.org/2.4/doc/tutorials/ml/introduction_to_svm/introduction_to_svm.html, accessed: 2017-07-24.

[9] I. Cohen, N. Sebe, A. Garg, L. S. Chen, and T. S. Huang, "Facial expression recognition from video sequences: Temporal and static modeling," *Computer Vision and image understanding*, vol. 91, no. 1, pp. 160–187, 2003.

[10] Y.-l. Tian, T. Kanade, and J. F. Cohn, "Recognizing action units for facial expression analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 2, pp. 97–115, 2001.

[11] E. Vural, M. Bartlett, G. Littlewort, M. Cetin, A. Ercil, and J. Movellan, "Discrimination of moderate and acute drowsiness based on spontaneous facial expressions," in *Pattern Recognition (ICPR), 2010 20th International Conference on*. IEEE, 2010, pp. 3874–3877.

[12] J. Whitehill, M. Bartlett, and J. Movellan, "Automatic facial expression recognition for intelligent tutoring systems," in *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on*. IEEE, 2008, pp. 1–6.

[13] L. Su and M. D. Levine, "High-stakes deception detection based on facial expressions," in *Pattern Recognition (ICPR), 2014 22nd International Conference on*. IEEE, 2014, pp. 2519–2524.

[14] E. A. Haggard and K. S. Isaacs, "Micromomentary facial expressions as indicators of ego-mechanisms in psychotherapy," in *Methods of research in psychotherapy*. Springer, 1966, pp. 154–165.

[15] P. Ekman and W. V. Friesen, "Non-verbal leakage and clues to deception," *Psychiatry*, vol. 32, no. 1, pp. 88–106, 1969.

[16] W.-J. Yan, Q. Wu, J. Liang, Y.-H. Chen, and X. Fu, "How fast are the leaked facial expressions: The duration of micro-expressions," *Journal of Nonverbal Behavior*, vol. 37, no. 4, pp. 217–230, 2013.

[17] M. Frank, M. Herbasz, K. Sinuk, A. Keller, and C. Nolan, "I see how you feel: Training laypeople and professionals to recognize fleeting emotions," in *The Annual Meeting of the International Communication Association. Sheraton New York, New York City*, 2009.

[18] J. K. Burgoon, D. P. Twitchell, M. L. Jensen, T. O. Meservy, M. Adkins, J. Kruse, A. V. Deokar, G. Tsechpenakis, S. Lu, and D. N. Metaxas, "Detecting concealment of intent in transportation screening: A proof of concept," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 1, pp. 103–112, 2009.

[19] S. Weinberger, "Airport security: Intent to deceive?" *Nature*, vol. 465, no. 7297, pp. 412–415, 2010.

[20] F. Xu, J. Zhang, and J. Wang, "Microexpression identification and categorization using a facial dynamics map," *IEEE Transactions on Affective Computing*, 2016.

[21] T. A. Russell, E. Chu, and M. L. Phillips, "A pilot study to investigate the effectiveness of emotion recognition remediation in schizophrenia using the micro-expression training tool," *British Journal of Clinical Psychology*, vol. 45, no. 4, pp. 579–583, 2006.

[22] P. Ekman, *Telling lies: Clues to deceit in the marketplace, politics, and marriage (revised edition).* WW Norton & Company, 2009.

[23] T. Pfister, X. Li, G. Zhao, and M. Pietikainen, "Recognising spontaneous facial micro-expressions," in *Computer Vision (ICCV), 2011 IEEE International Conference on.* IEEE, 2011, pp. 1449–1456.

[24] J. Whitehill, Z. Serpell, Y.-C. Lin, A. Foster, and J. R. Movellan, "The faces of engagement: Automatic recognition of student engagement from facial expressions," *IEEE Transactions on Affective Computing*, vol. 5, no. 1, pp. 86–98, 2014.

[25] L. D. Pool and P. Qualter, "Improving emotional intelligence and emotional self-efficacy through a teaching intervention for university students," *Learning and Individual Differences*, vol. 22, no. 3, pp. 306–312, 2012.

[26] Q. Wu, X. Shen, and X. Fu, "The machine knows what you are hiding: An automatic micro-expression recognition system," in *Affective Computing and Intelligent Interaction.* Springer, 2011, pp. 152–162.

[27] S. Polikovsky, Y. Kameda, and Y. Ohta, "Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor," in *Crime Detection and Prevention (ICDP 2009), 3rd International Conference on.* IET, 2009, pp. 1–6.

[28] X. Zhao and S. Zhang, "A review on facial expression recognition: Feature extraction and classification," *IETE Technical Review*, vol. 33, no. 5, pp. 505–517, 2016.

[29] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 971–987, 2002.

[30] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 886–893.

[31] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, 2009.

[32] W.-J. Yan, S.-J. Wang, Y.-H. Chen, G. Zhao, and X. Fu, "Quantifying micro-expressions with constraint local model and local binary pattern," in *Workshop at the European Conference on Computer Vision.* Springer, 2014, pp. 296–305.

[33] D. Cristinacce and T. F. Cootes, "Feature detection and tracking with constrained local models." in *BMVC*, vol. 1, no. 2, 2006, p. 3.

[34] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," in *European conference on computer vision.* Springer, 1998, pp. 484–498.

[35] T. F. Cootes and C. J. Taylor, "Active shape models—'smart snakes'," in *BMVC92.* Springer, 1992, pp. 266–275.

[36] S.-T. Liong, J. See, R. C.-W. Phan, and K. Wong, "Less is more: Micro-expression recognition from video using apex frame," *arXiv preprint arXiv:1606.01721*, 2016.

[37] S.-T. Liong, J. See, K. Wong, A. C. Le Ngo, Y.-H. Oh, and R. Phan, "Automatic apex frame spotting in micro-expression database," in *Pattern Recognition (ACPR), 2015 3rd IAPR Asian Conference on.* IEEE, 2015, pp. 665–669.

[38] X. Jia, S. Liu, and D. M. Powers, "Dynamic feature extraction for facial expression recognition based on optical flow," *Information Technology Journal*, vol. 12, no. 23, p. 7305, 2013.

[39] H. Fang, N. Mac Parthaláin, A. J. Aubrey, G. K. Tam, R. Borgo, P. L. Rosin, P. W. Grant, D. Marshall, and M. Chen, "Facial expression recognition in dynamic sequences: An integrated approach," *Pattern Recognition*, vol. 47, no. 3, pp. 1271–1281, 2014.

[40] J.-J. Lien, T. Kanade, J. F. Cohn, and C.-C. Li, "Subtly different facial expression recognition and expression intensity estimation," in *Computer Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference on.* IEEE, 1998, pp. 853–859.

[41] Y. Yacoob and L. S. Davis, "Recognizing human facial expressions from long image sequences using optical flow," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 18, no. 6, pp. 636–642, 1996.

[42] A. Sánchez, J. V. Ruiz, A. B. Moreno, A. S. Montemayor, J. Hernández, and J. J. Pantrigo, "Differential optical flow applied to automatic facial expression recognition," *Neurocomputing*, vol. 74, no. 8, pp. 1272–1282, 2011.

[43] T. Kanade, J. F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on.* IEEE, 2000, pp. 46–53.

[44] A. K. Davison, M. H. Yap, N. Costen, K. Tan, C. Lansley, and D. Leightley, "Micro-facial movements: An investigation on spatio-temporal descriptors," in *European Conference on Computer Vision.* Springer, 2014, pp. 111–123.

[45] Y. Guo, Y. Tian, X. Gao, and X. Zhang, "Micro-expression recognition based on local binary patterns from three orthogonal planes and nearest neighbor method," in *2014 Internation al Joint Conference on Neural Networks (IJCNN).* IEEE, 2014, pp. 3473–3479.

[46] S. Polikovsky, Y. Kameda, and Y. Ohta, "Detection and measurement of facial micro-expression characteristics for psychological analysis," *Kameda's Publication*, vol. 110, pp. 57–64, 2010.

[47] M. Shreve, S. Godavarthy, D. Goldgof, and S. Sarkar, "Macro-and micro-expression spotting in long videos using spatio-temporal strain," in *Automatic Face and Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on.* IEEE, 2011, pp. 51–56.

[48] S. Godavarthy, "Microexpression spotting in video using optical strain," *Masters Thesis, University of South Florida*, 2010.

[49] S.-J. Wang, W.-J. Yan, X. Li, G. Zhao, and X. Fu, "Micro-expression recognition using dynamic textures on tensor independent color space," in *Pattern Recognition (ICPR), 2014 22nd International Conference on.* IEEE, 2014, pp. 4678–4683.

[50] P. Ekman, W. V. Friesen, and J. C. Hager, "Facial action coding system (FACS)," *A technique for the measurement of facial action. Consulting, Palo Alto*, vol. 22, 1978.

[51] G. Warren, E. Schertler, and P. Bull, "Detecting deception from emotional and unemotional cues," *Journal of Nonverbal Behavior*, vol. 33, no. 1, pp. 59–69, 2009.

[52] S.-J. Wang, H.-L. Chen, W.-J. Yan, Y.-H. Chen, and X. Fu, "Face recognition and micro-expression recognition based on discriminant tensor subspace analysis plus extreme learning machine," *Neural processing letters*, vol. 39, no. 1, pp. 25–43, 2014.

[53] A. Cambridge, "The database of faces," *Cambridge University Computer Laboratory,[Online]. Available: http://www. cl. cam. ac. uk/research/dtg/attarchive/facedatabase. html.[Accessed 1 4 2012]*, 2002.

[54] A. Georghiades *et al.*, "Yale face database," *Center for computational Vision and Control at Yale University, http://cvc. yale. edu/projects/yalefaces/yalefa*, vol. 2, 1997.

[55] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE transactions on pattern analysis and machine intelligence*, vol. 23, no. 6, pp. 643–660, 2001.

[56] Y. Guo, C. Xue, Y. Wang, and M. Yu, "Micro-expression recognition based on CBP-TOP feature with elm," *Optik-International Journal for Light and Electron Optics*, vol. 126, no. 23, pp. 4446–4451, 2015.

[57] A. Vinciarelli, A. Dielmann, S. Favre, and H. Salamin, "Canal9: A database of political debates for analysis of social interactions," in *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops.* IEEE, 2009, pp. 1–4.

[58] M. Shreve, J. Brizzi, S. Fefilatyev, T. Luguev, D. Goldgof, and S. Sarkar, "Automatic expression spotting in videos," *Image and Vision Computing*, vol. 32, no. 8, pp. 476–486, 2014.

[59] S. M. Mavadati, M. H. Mahoor, K. Bartlett, P. Trinh, and J. F. Cohn, "DISFA: A spontaneous facial action intensity database," *IEEE Transactions on Affective Computing*, vol. 4, no. 2, pp. 151–160, 2013.

[60] X. Huang, S.-J. Wang, G. Zhao, and M. Piteikainen, "Facial micro-expression recognition using spatiotemporal local binary pattern with integral projection," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2015, pp. 1–9.

[61] V. Mayya, R. M. Pai, and M. M. Pai, "Combining temporal interpolation and DCNN for faster recognition of micro-expressions in video sequences," in *Advances in Computing, Communications and Informatics (ICACCI), 2016 International Conference on*. IEEE, 2016, pp. 699–703.

[62] Y. Wang, J. See, R. C.-W. Phan, and Y.-H. Oh, "Efficient spatio-temporal local binary patterns for spontaneous facial micro-expression recognition," *PloS one*, vol. 10, no. 5, p. e0124674, 2015.

[63] S.-J. Wang, W.-J. Yan, G. Zhao, X. Fu, and C.-G. Zhou, "Micro-expression recognition using robust principal component analysis and local spatiotemporal directional features," in *Workshop at the European Conference on Computer Vision*. Springer, 2014, pp. 325–338.

[64] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[65] T. K. Ho, "Random decision forests," in *Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on*, vol. 1. IEEE, 1995, pp. 278–282.

[66] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.

[67] L. Auria and R. A. Moro, "Support vector machines (svm) as a technique for solvency analysis," *German Institute for Economic Research*.

[68] C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data mining and knowledge discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[69] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: A new learning scheme of feedforward neural networks," in *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, vol. 2. IEEE, 2004, pp. 985–990.

[70] P. Ekman, W. V. Friesen, and J. C. Hager, "Facs investigator's guide," *A human face*, p. 96, 2002.

[71] P. Ekman and W. Friesen, "A technique for the measurement of facial movement," *Facial Action Coding System*, 1978.

[72] M. Pietikäinen, "Local binary patterns," *Scholarpedia*, vol. 5, no. 3, p. 9775, 2010.

[73] Y. Wang, F. Cao, and Y. Yuan, "A study on effectiveness of extreme learning machine," *Neurocomputing*, vol. 74, no. 16, pp. 2483–2490, 2011.

[74] G.-B. Huang, L. Chen, and C.-K. Siew, "Universal approximation using incremental constructive feedforward networks with random hidden nodes," *Neural Networks, IEEE Transactions on*, vol. 17, no. 4, pp. 879–892, 2006.

[75] M. Loderer and J. Pavlovičová, "Optimization of lbp parameters," in *ELMAR (ELMAR), 2014 56th International Symposium*. IEEE, 2014, pp. 1–4.