Sudan University of Science &Technology

College of Graduate Studies

A Development of Directional Gradient Local Ternary Pattern for Facial Expression Recognition

تطوير نموذج النمط الثلاثي النحداري المتدرج لتمييز تعابير الوجه

This thesis is submitted in fulfilment for the degree of Doctor of Philosophy in

Computer Science

Prepared by

Nahla Nour

Supervised by

Prof. Serestina Viriri

MAR 2022

A Development of Directional Gradient Local Ternary Pattern for Facial Expression Recognition

BY

Nahla Abdellatif Elmobark Nour

SUDAN UNIVERSITY FOR SCIENCE & TECHNOLOGY

Supervisor:

PROF. SERESTINA VIRIRI

MAR2022

تطوير نموذج النمط الثلاثي الأنحداري المتدرج لتمييز تعابير الوجه

A Development of Directional Gradient Local Ternary Pattern for Facial Expression Recognition

BY

Nahla Abdellatif Elmobark Nour

A dissertation submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

in

Computer Science

College of Computer Science and Information Technology

Sudan University of Science & Technology

ABSTRACT

Humans can easily determine ones' gender, identity and ethnicity with highest accuracy as compared to face expression recognition.. This makes development of automatic face expression recognition techniques that surpass human performance an attractive yet challenging task. Facial expressions recognition requires extraction of robust and reliable expression discriminative features. Local binary patterns (LBP) sensitivity to noise makes it insufficiently reliable in capturing expression discriminative features. Although a local ternary pattern (LTP) is insensitive to noise, it uses a single static threshold for all images regardless of varied image conditions. Local directional patterns (LDP) uses k directional responses to encode image gradient and disregards not only central pixel in the local neighborhood but also 8 − kdirectional responses. Every pixel in an image carry subtle information. Discarding 8 − k directional responses lead to lose of discriminative texture features. Each pair are compared and the bit corresponding to the maximum value in the pair is set to 1 while the resultant binary code is converted to decimal and assigned to the central pixel. Local ternary directional patterns (LTDP) first get the difference between neighboring pixels and central pixel in 3×3 image region. These differential values are convolved with Kirsch edge detectors to obtain directional responses. These responses are normalized and used as probability of an edge occurring towards a respective direction. An adaptive threshold is applied to deriveLTDP code. The LTDP code is split into its positive and negative LTDP codes. Histograms of negative and positive LTDP encoded images are concatenated to obtain texture feature.. This study proposes fusion of different facial features to enhance their discriminative power. Experimental results show that face component comparing with whole face achieve lower MAE compared to single feature performance. Finally the study used Alex-Net, Vgg-16 & Resent which applied on images to extract the feature of facial images and obtained features in the last fully connected layer which are used as input to SVM classifier for producing the final classification result. The study shows that when this study used CK & JEFFE datasets and SVM classifier in whole face the accuracy is 99.3% but when it used facial component the accuracy is 99.7% this results show that the accuracy gradually increas

مستخلص البحث

تعابير الوجه هي عملية تلقائيه لا مفر منها ولا يمكن السيطرة عليها وتؤثر باستمرار الشكل والمظهر العام لوجه الإنسان. يمكن للبشر بسهولة تحديد- تحديد الجنس والهوية والعرق بأعلى دقة مقارنة بـتعابير الوجه . هذا يجعل تطوير تقنيات تقدير الحاله التلقائي يتجاوز الأداء البشري مهمة جذابة ولكنها صعبة. تقدير تلقائي لتعبيرات الوجه يتطلب استخراج ميزات تمييزية قوية وموثوقة للحالة الشخص . ثنائي محلي حساسية الأنماط (LBP) للضوضاء تجعلها غير موثوقة بدرجة كافية في التقاط التعابير والسمات التمييزية. على الرغم من أن الأنماط الثلاثية المحلية (LTP) غير حساسة للضوضاء ، يستخدم حدًا ثابتًا واحدًا لجميع الصور بغض النظر عن حالة الصورة المتنوعة

تستخدم أنماط الاتجاه المحلية (LDP) استجابات اتجاهية لتشفير الصورة ويتجاهل ليس فقط البكسل المركزي في الحي المحلي ولكن أيضًا 8 الاستجابات الاتجاهية. كل بكسل في الصورة يحمل معلومات دقيقة k - .

تحصل أنماط الاتجاه الثلاثية المحلية (LTDP) أولاً على الاختلاف بين وحدات البكسل المجاورة والبكسل المركزي في منطقة الصورة 3 × 3. هؤلاء يتم تحويل القيم التفاضلية باستخدام كاشفات الحافة Kirsch للحصول على الاتجاه استجابات. يتم تطبيع هذه الاستجابات واستخدامها كاحتمال لحدوث حافة تتجه نحو اتجاه معين. يتم تطبيق عتبة تكيفية للاشتقاق كود LTDP. يتم تقسيم رمز LTDP إلى أكواد LTDP الإيجابية والسلبية. النتائج التجريبية التي تم اتباعها من خلال الانحدار لتعبيرات الوجه الدقيقه باستخدام الدعم الانحدار المتجه (SVR). كان متوسط الخطأ المطلق (MAE) والنتيجة التراكمية (CS) تستخدم لتقييم أداء واصفات الوجه المقترحة. وتظهر الدراسه انه عندما تم استخدام مجموعة البيانات CK و JEFFE و استخدمنا الخوارزميه المحسنه المقترحه (DGLTP)وخوارزمية التصنيف (SVM) في الوجه كاملا بلغت الدقة (99.3%)، وعندما استخدمنا مناطق معينه كانت الدقه (99.7%) لاحظنا ان في كل مره تزداد الدقه تدريجيا. اما بإستخدام CNN باستخدام طريقه (88.2%) كانت النتيحه Vgg-16 وادخال SVM كمصنف في الطبقة الاخيرة.

ACKNOWLEDGEMENTS

I thank the God of all creations for the gift of life and good health, guidance and protection He has unconditionally granted me.

I am obliged to Masinde Muliro University of Science and Technology for offering me a chance to pursue this research under proficient guidance, support and advice of Prof. J.-R. Tapamo and Prof. A. O. Adewumi. I am indebted to Prof. J. R. Tapamo for literally teaching me Image Processing, skills that helped me settle in this hitherto strange field. This research would never have been done smoothly without superb technical support of the crazy tech man Soren Greenwoods, thank you Soren.

I am grateful to my dear mother Niemat Altaher, a courage's Mum who taught me that whenever you step into the unknown, you either find a solid base or you learn to fly. And I must admit in most cases, I learnt how to fly. Thank you mum for your support, love and guidance without which I would never have found a definition of myself. I am indebted to my husband Muazer Mustafa who allowed me to travel miles to pursue this study. Your love and emotional support is very much acknowledged! Among the best, you two, also my Sister Nada for helping and supporting me I wish her all the best.

Finally but not least, I thank my friend and colleges for their support and guides for their support, critic and comments that helped shape not only this thesis but also my mind.

DECLARATION

I hereby declare that this thesis is the result of my own investigation, except where otherwise stated. I also declare that it has not been previously or concurrently submitted as a whole for any other degrees at Sudan University of Science and Technology or other institutions.

Student Name: Nahla Nour

Signature________________ Date 23\02\2021

LIST OF PUBLICATIONS

I, Nahla Nour, declare that the following are publications from this thesis

1. N. Nour, S. Viriri, “Facial Expression Recognition Using Directional Gradient Local Ternary Patterns”, Multi-disciplinary Trends in Artificial Intelligence, Springer LNCS, vol. 11909,, pp. 87-96, 2019.
2. SM. Osman, N. Nour, S. Viriri, “Component-Based Gender Identification Using Local Binary Patterns”, Computational Collective Intelligence, Springer LNAI, vol. 11683, pp. 307-315, 2019.
3. Mohamed, S., Nour, N. and Viriri, S., 2018, March. Gender identification from facial images using global features. In 2018 Conference on Information Communications Technology and Society (ICTAS) (pp. 1-6). IEEE.
4. A conference paper entitled, " Facial Expression Recognition : A Survey of the State-of-the-art " has been accepted and in :on Information Communications Technology and Society (ICTAS),2019
5. A conference poster entitled, " Facial Expression Recognition : A Survey of the State-of-the-art " has been accepted and in :SCCSIT 17 The 5th Sudan Conference on Computer Science and Information Technology

Also I, declare that the following are accepted from this thesis

1. A conference paper entitled, " Component-Based Using Directional Gradient Local Ternary Pattern " has been accepted and in :The Advances in Science, Technology and Engineering Systems Journal (ASTESJ), 2020 conference paper entitled, “Face Expression Recognition using Convolution Neural Network
2. face expression recognition using convolution neural network (CNN) models" has been accepted and in : International Journal of Grid Computing & Applications (IJGCA), 2020

TABLE OF CONTENTS

LIST OF TABLES

List of Figures

LIST OF ABBREVIATIONS

2D(PCA)	Two-dimensional Principal component analysis
3D	Three diminution
AAM	Active appearance model
ANN	Artificial neural network
BN	Bayesian network
BP	Back propagation
CNN	Convolution natural network
DBN	Deep Boltzmann network
DCT	Discrete cosine transform
DL	Deep learning
DLTP	Directional local ternary pattern
DWT	Discrete wavelet transform
FLD	Fisher linear discriminate
HMM	Hidden Markov model
KNN	K-Nearest Neighbour

LBP	Local binary pattern
LBPH	local binary pattern histogram
LDP	Local directional pattern
LFW	Labelled Faces in the Wild
LTP	Local ternary pattern
NIN	Network in network
NN	Neural network

CHAPTER 1
INTRODUCTION

OVERVIEW

Extraction of human emotions from facial expression has significantly won researchers attention in computer vision area. Recently, several attempts have been made by researchers to enhance facial detection and tracking. So that better human communication interaction (HCI) applications will be possibly invented if we can develop computers' detection and facial expression analysis like those of humans. The applications could be used in games to improve our way of interaction with virtual characters, it could even lead to an increase in our general interpretation of human emotions. In recent years, humans' facial expression recognition has become a very common topic of research in the area of computer vision.

Faces have great impact in communication; therefore, the area of facial expression recognition (FER) has been progressed, and obviously contributes to (HCI). More research need to be conducted in FER on order to upgrade common (HCI). We frequently use facial expressions to communicate with others in non-verbal communication, which are subtle parts of the overall communication process. Understanding facial expressions and their meanings is an important part of communication. Whenever there is verbal and not-verbal cues between two or more people to get to a mutual understanding definitely there is communication. Some factors like (different languages) complicate this process, which makes understanding one another very difficult. A lot of communication in business aspects occurs through electronic messages such as text messages and emails, and it is sometimes difficult hard to understand the real intended context or meanings of electronic messages perfectly. However, speech in facial expressions always depends on understanding facial expressions, such as happiness, sadness, and anger, likewise other non-verbal communication. Today doctors, anthropologists, neuroscientists and psychologists have shown much interest to our ability to continue recognizing familiar faces and subtle facial gestures. This is simply because both facial expression and human brain work equivalently.

Although there are many face expression associated with human training, but facial appearance has the most common characteristic that is reliable for people recognize one’s expression. Prior perceptible information of one’s expression, identity, gender, mood, age and ethnicity are derived from human face. In[1] asserts that attributes obtained from human facial appearance like expression, mood and feel have scientific impact in interpersonal behaviour and considered as an important contextual cue in social network [2], [3] as well as social perception during craniofacial development.[4] .If facial appearance discovers psychological characteristic, it will definitely show an impression. These are signs of great expected application of automatic face expression recognition system.

An image face expression needs a high level of information. Extraction of facial expression feature is extremely important in science, as a result that the performance of face recognition depends on how quality the features are[5] . A number of feature extraction techniques for facial expression like active appearance model (AAM) [6], Texture Based[7] , Active Shape Model , Skin colour Based[8], Feature Based[9], Template Matching[10], and Knowledge Based, have been introduced in the literature.

Pattern recognition is the scientific area of machine learning which aims to classify data (patterns) into many categories [11]. Face expression is not a classical classification problem. This problem can be approached as a multi-class classification problem [12], [7], [13], [14]or a regression problem [15],[16], [11] or combination of consideration and regression [17], [19]. Illumination [20]pose [21]and noise[22] are factors affecting database research of face expression recognition. In this thesis, we combined deep learning, classification and regression as the approached face expression recognition system.

MOTIVATION

Despite the richness and attraction of application aspect of automatic face expression recognition, Difficulties and challenges facing automatic expression recognition are revealed by the existing information from observation area. Expression gives reflection of feeling and thought of the face which affected by some factors such as Illumination[20], noise [21] and pose[22]. Here are the 5 motivating questions of this study 1)

What kind of features a machine can use to collect almost zero error in automatic face expression recognition? 2) What techniques are needed to create automatic face expression recognition system that is strongly formed to obtain face expression? 3) What are the previous performances and achievement of face expression recognition? 4) What can be done to improve automatic face expression recognition and overcome challenges and difficulties? 5) Is expression affected by facial components like forehead, cheeks, eyes, nose and mouth? And how facial component-based expression features extraction. There have been concentration on the above questions in image processing in the research community of computer vision. Thus, this study aims to answer these questions with an intention of increasing performance and accuracies of face expression recognition.

PROBLEM STATEMENT

There are not many proposed face expression techniques in literature[12], [7]. This may be ascribed to face expression recognition not being classical classification problem, light[20], variability[21], the pose of the face and noise[22].

A 100% recognition accuracy has never been achieved by any Face expression technique[11]. This make face expression available research problem. The challenges that arise in designing a portable computer vision system include illumination variation, computation efficient and random noise. To end this, this research can extract either local (geometric-based) features or global (appearance-based) features. In geometric based method this research extract feature information using shape, distance and position of facial components and appearance-based feature extraction method using appearance information such as pixel intensity of face image [2]. The most appearance-based features extractions methods that are used include local binary pattern (LBP) [3], [4], [5] [6], Local directional pattern (LDP) [5] [6] [7], local ternary pattern (LTP) [7] and Gabor wavelet transform (GWT) [5], [8]. LBP was proposed by [9], which divided the facial image into local regions, it used grey-level intensity value to encode the texture of the image that leads to efficient computational. This is achieved using CK datasets 90.1 and 83 for testing 6-classes and 7-classes respectively. Nevertheless, it poorly performed in illumination variation and random noise in addition to high error rate when the background.

The problems related with this research are:

- There is no standard algorithm for face expression recognition from facial component images.
- The coordinate points in the face may vary from one algorithm to another algorithm (some algorithms take the feature from the face globally and others take it locally as point).
- Major characteristics of an image such as colour, texture, and shape are not considered while producing the output.

RESEARCH OBJECTIVES

The main aim of this research is to model a sufficiently robust and accurate model for human expressions recognition using facial component. This research precisely targets the fact of creating new techniques which could improve efficiently the performance of facial data extraction and expression recognition techniques. The specific objectives are:

1. To analyse previous studies in face expression recognition: factors affecting expressions, algorithms

used for expression recognition, results achieved, face expression challenges, image representation techniques for expression modelling, expression recognition evaluation protocols.

2. To find out the trend of facial landmarks displacement across face expression.
3. To develop a model for face expression recognition based on fused facial and facial component features and ensemble of classifier.
4. To define LDP variant that consider significant orientation responses in encoding image gradient rather than top $_k$ responses.
5. To extend LDP to a ternary directional pattern based on pixel differential values and an adaptive threshold for encoding image gradient.
6. To investigate effect of local pooling, within feature selectors, on performance of CNN in face expression recognition.
7. To evaluate the proposed model.

RESEARCH QUESTIONS

This research will be conducted to answer the following questions in order to realize the expected objectives:

- Is it possible to recognize expression from images (efficiently) based on facial component accurately?
- Is it possible to improve the accuracy rate of facial expression recognition using facial component?
- What facial components are most informative for determining the accurate expression?

CONTRIBUTION OF THE THESIS

The contributions of the proposed work are as follows:

1. Dual-feature fusion technique is proposed in this work for effective and efficient classification of facial expressions in the unconstrained environment.
2. The proposed framework is based on local and global features, which make the proposed framework robust to change in occlusions, illumination, and noise.
3. Feature selection process is used to obtain the discriminative features, where the redundant features are discarded. The reduction in feature vector length also reduces the time complexity.
4. Using facial parts as input to reduce the size of input and FER processing time when compared to other FER architectures that used the whole face for facial expression recognition.
5. Improve existing deep learning methods and come out with more accurate FER model.

SCOP OF THE STUDY

This research focuses on feature extraction of both face and facial component for expression recognition. Face expression recognition systems contain many modules such as user interface, facial image acquisition, image pre-processing, and expression feature extraction and expression recognition. This study focuses on face facial component detection and expression feature extraction process with light regard based on other modules. Face expression recognition are approaches as hierarchical multi-class classifier. The extraction of both shape and texture features from face and (JEFFE, Ck) dataset. This dataset contains 213 images of 7 facial expressions (6 basic facial expressions + 1 neutral) and CK dataset contain 593 image sequences (327 sequences with discrete emotion labels). Support vector machine (SVM) and Artificial Neural Network (ANN) are both used as machine learning techniques for recognizing expression from the extracted facial expression features.

RESEARCH ORGANIZATION

The rest of this thesis is organized as follows: **Chapter 2** gives general background and application of face expression and also reviews related work in face expression recognition and feature extraction techniques. **Chapter 3** gives detailed discussion of methods and techniques for face detection, feature extraction and classification. **Chapter 4** discussed local pattern for face expression in wide scope. **Chapter 5** presents the proposed framework for face expression. It presents face expression used approaches, algorithm used for facial component localization, directional improve gradient local ternary pattern (D-ILTP). **Chapter 6** presents a comprehensive analysis and discussion of result including comparison with related work. **Chapter 7** conclusion and future work.

CHAPTER 2
LITERATURE REVIEW

INTRODUCTION

In the recent years, human's facial expression recognition has become one of the most operating research area in computer vision. Facial expression gives a significant contribution to understand human communication, this expression has been categorized into seven basic emotions (e.g., neutral, happiness, fear, anger, sadness, surprise and disgust,). Face expression recognition has achieved an amazing improvement in the recent years [9], [16], [17], [20]. This chapter gives presentation of facial expression process analysis, developed techniques for face expression, with algorithm analysis of face expression recognition and face recognition. It also gives comparative analysis of preceding research in facial expression recognition.

FACE EXPRESSION BACKGROUND

Human frequently use facial expressions a lot with each other in nonverbal communication, and these are the most delicate parts to analyze in general process of communication. Understanding facial expressions and their meanings need to be understood in order to make communication meaningful. Whenever people exchange verbal and non-verbal clues to reach a common understanding definitely there is communication. But some factors sometimes complicate the process like (language differences and others) which make understanding difficult among people. In businesses, a lot of communication in business aspects occurs through electronic messages such as text messages and emails and understanding the real meanings or context of electronic messages perfectly is sometimes difficult. Whereas speech in facial expressions and other non-verbal communication, depends on understanding basic facial expressions like anger, sadness and happiness. Today doctors, anthropologists, neuroscientists and psychologists have shown much interest to our skill to continue recognizing common faces and subtle facial gestures. This is simply because human brain and facial expression work in an equivalent manner. Facial expression shows human emotional situation, which tend to be easy for human being to identify one's expression with these features, but it seems to a difficult task in computer vision.

2.2.1 Application of face expression area

A social dimension is brought to human computer interaction by the agents of computer animation and robots strengthen human thinking in new dimension about how computers could be utilized in human's daily activities [1]. There are several important real word application that are used to characterize facial appearance variability over expression. Computer based face recognition is used to determine one's expression. Face expression recognition has been applied in several areas. The most common application areas include;

2.2.1.1 Human Interaction Communication application

Therefore, there must be a development of robust real-time perceptive advanced technologies in order to fulfil the idea of face to face interactive machines [23].

2.2.1.2 Patient Monitoring

Different states of health affect one's facial expression, thus, a health care framework could benefit from recognition system of facial expression [24]. A Smart City is excelling in multiple key areas; economy, environment, energy and living, but healthcare aspect is also a very important service. A smart healthcare system has to be a significant component of a developed city, so that patients’ feelings can be continually monitored and proper actions will be taken as needed. Moreover, to give stakeholders automatic feedback from patients without needing to ask them for feedback [25].

2.2.1.3 Clever Marketing

Humans can widely categorize people patronizing a store into four categories [26] Firstly, the Browsers, who just like walking round the store and wasting their own time and the time of the shop assistant. Secondly, there are some come to get information concerning some products or just come to price the product but with aim of purchasing in the future. Thirdly, some customers are for real, they want to purchase but they are not properly handled by the shop assistant and therefore the deal is not done. The last category is about of the Buyers who come to store particularly to purchase some products and they instantly do. It is difficult for some sales personnel to distinguish the potential customers among all the categories and therefore design selling strategies tailored to those categories for sales personnel to allocate time to such customers.

2.2.1.4 Emotional Analysis

Both verbal and nonverbal interactions are combined to form human communication. Human can exchange interactions through body gestures, facial expressions and other nonverbal cues. This is especially true in the communications of emotions[26]. Presently, a range of remote patient monitoring systems (RPMS) are being developed to look after patients at home instead of being in the expensive hospital environment. These systems enable remote monitoring by health professionals without much medical intervention. Nevertheless, these systems are still less effective than human communication based on one on one. The face and facial features can give faster patient cognitive and emotional states of patients than electrical signals and facial expression is considered as one of the most effective features of RPMS [25].

2.2.1.5 Security and Surveillance

Crime surveillance and monitoring may need facial expression recognition to be aware of crimes before happening [27], also can be significant in controlling ATM money transfer fraud by mongering expression [28].

MACHINE LEARNING

Learning is divided into supervised learning and unsupervised learning. Supervised learning is a kind of learn whose aim is learning mapping from the input (instance) to an output (target or label) in which supervisor provides the correct values.

2.3.1 Supervised Learning

This section presents an overview on the three steps for face expression recognition for this research.

2.3.1.1 Face detection step

Face detection step has been at the top of discussion in the computer vision domains [29]. It is also a required first stage in facial recognition systems, whose purpose is to localize and extract the facial region out of the background [5]. Structural components, pose and facial expressions are the most common challenges of face detection. Several methods have been developed for detecting faces in an arbitrary scene [30], [31], [32], [33], [34], [35], [36] Nevertheless, most of those methods can only perform detection on frontal and near frontal views of faces. Heisele et al. [30] proposed a trainable system based on component to detect frontal and near frontal views of faces Rowley et al. [36] introduced a method based on neural network for frontal views of faces detection. This step includes image processing:

2.3.1.1.1 Acquisition of the image

A camera is used by many people for taking an image. But presently, it has become very easier to use a mobile camera in taking images, nevertheless; a dataset can also be used that is basically an image database (which is usually used in many cases for research purposes).

2.3.1.1.2 Image pre-processing

This step concentrates on deposing, which means an act removing noise from an image. We can perform this through histogram equalization, filtering/convolution, or morphological operation.

2.3.1.1.3 Segmentation of the region of the interest (ROI)

The foreground can be extracted for analysis and we leave the background.

2.3.1.2 Feature Extraction Step

Feature extraction is the next step after face detection, in which some significant features are extracted to describe the faces emotion. There are two primary characteristics features extraction which are appearance and geometry.

Due to specific emotions, such as wrinkles and bulges, an image filter applied in appearancebased methods to the whole face or s pecific facial regions to extract changes in texture. Well-known appearance-based methods include the use of local binary patterns (LBP)[36], [37], [38], [18], [38], [18], [39]. local directional patterns (LDP) [39], [40], [41].local ternary patterns (LTP) [41] Enhanced local texture feature sets for face recognition under difficult lighting conditions and Gabor wavelet transform (GWT) [39], [40], [41], [42].While LBP uses grey-level intensity [39] values to encode an image texture is very efficient in computation, a poor performance was shown in LBP when there is non-monotonic illumination variability and random noise, because LBP can be easily changed by a small change in grey-level values[43]. LDP uses a structure scheme for coding which is different from LBP scheme, instead of using the grey-level intensity values it uses lateral edge response values. It was proven that LDP performance surpasses LBP, due to its two-level discrimination coding scheme it tends to produce patterns with of consistency in uniform and near-uniform regions, and is heavily dependent on many noticeable edge direction parameters[44]. LTP which added an additional level of discrimination and introduced ternary codes which is different from binary codes in LBP, to solve the restriction. More recently, the proposed gradient local ternary pattern (GLTP) method [45] performs combination of the advantages of the preceding methods. GLTP uses a three-level discrimination ternary coding scheme to encode the texture of a gradient magnitude.

Geometric methods centralise on feature extraction, where the distances between certain points on the face are measured such as the distance between corners of the eye and mouth. We can also extract the shape of different facial components of the emotional changes. Feature extraction based on Geometric is observed to more difficult in implementation than appearance-based methods due to the variable size [37]. Using four control points, Bezier curves were used to accurately represent the shape of each facial component precisely in several expressions. In this method we use the distance and angle of the end points to describe each emotion was. Methods for 3D FER [46]and the latest 4D FER [65] have also been introduced with the emerging 3D imagery technology . Unlike traditional 2D methods, 3D FER perform emotional classification by extracting face geometric from 3D facial scans. 4D FER exists when temporal information is used in adjacent 3D facial to transmit variations. Though 3D/4D FER is still a fairly new term, study have proven its effect in overcoming the restrictions of illumination and pose variance in most 2D FER techniques [46].

Due to the large number of data that could be generated from a face, a significant part of the information is redundant. Programme effectiveness can be highly reduced by large feature sets particularly at the process of training a classifier. To reduce feature set size, Dimensionality reduction (DR) techniques are usually employed to eliminate redundant data, extremely improving the performance without accuracy reduction. Some well-known techniques for the reduction of dimensionality reduction are; principal component analysis (PCA) [47], [41], [46] independent component analysis (ICA) [37] linear discriminant analysis (LDA) and AdaBoost [48], [37]. In [63], multiple feature sets were combined with PCA were used for FER. A high accuracy was achieved for classification from the result of using multiple feature sets. With combination of the method[47] and PCA for dimensional reduction, there is a considerable level of improvement in performance compared to other methods based on appearance. In the same sense, in [49], in order to increase the classification accuracy, different transformation area feature sets were fused by using canonical correlation analysis (CCA). The two-dimensional PCA (2DPCA) [50], aiming to decrease the size of the feature vector generally, then the size of feature sets preceding classification had to be reduced by using a more effective variant of PCA. Recently in literature, additionally to the common methods based on appearance and geometric, there was proposal of other feature selection methods. A deep belief network (DBN) [51] was used to select, learn and classify features in [50], [52]. Nevertheless, there was no increase shown in recognition accuracy on the traditional methods from the result. Traditional methods are less cost compared to the computational cost of deep learning. More recently, several attempts were made on dynamic FER. Dissimilar to traditional FER methods, the main aim of dynamic FER is emotional determination from a sequence of images as opposed to a single static image. In[51], extracting dynamic expression spatial and temporal information required the use of atlas and sparse representation. Higher recognition accuracies were achieved when temporal information was used with spatial information compare to the results of static image FER. Computational with dynamic FER is also higher in expense based on the length of the sequence.

2.3.1.3 Classification Step

Creating a classifier depends on the extracted features in the previous step is the final step of FER. The features extracted are forwarded into a machine learning algorithm which attempts classifying them into separate emotion classes according to the similarities of feature data. After the classifier has been trained to input features to a specific emotion class. There was wide approval and acceptance of the first proposal of Ekman [53] that the all-round recognizable emotions are seven, as (surprise, joy, fear, disgust, sadness, anger and neutral emotion). Some well-known supervised classification methods are support vector machines (SVM) [38], [40],[45], [53],[37], K-nearest neighbours (K-NN) [37],[53], and neural networks (NN) [54], [55], [18]. When using SVM with a radial basis function (RBF) kernel for FER task [63] [18], [37] the performance surpasses other classifiers based on including alternative kernel SVMs [18], [37].

2.3.1.4 Limitation

When designing a portable computer vision system, some challenges like illumination variation, random noise and computation effective come up. To end this, we can extract either local features based on geometric or global features based on appearance can be extracted to address these challenges. Feature information is extracted in geometric-based method by using distance, shape and position of facial components, while appearance information such as pixel intensity of face image is used to extract information in appearance-based method [2]. The most appearance-based features extractions methods that are used include local binary pattern (LBP) [36], [38], [18],[39]. Local directional pattern (LDP) [38], [40], local ternary pattern(LTP) [41] and Gabor wavelet transform (GWT) [18], [56] . LBP was introduced by [42], where the facial image was divided into local regions and grey-level intensity value was used for texture encoding of the image which causes effective computation. When CK datasets were used the accuracies achieved were 90.1% for testing 6-classes and 83% for testing 7-classes. However, LBP performance is very poor in illumination variation, random noise and a large number of error when changes occur in the background [57]. In order to address this problem, LDP feature is prevailed in which the edge response values are computed in eight directions at the location of all pixels and a relative strength magnitude code is generated unlike LBP which use grey-level intensity. It is shown from [29] that recognition rate of 93.7 was for 6-classes and 88.4 for 7-classes. Despite that LDP performance was shown to be more effective than LBP but it produces inconsistence pattern in uniform and near uniform region. LTP was introduced to tackle these drawbacks of LDP by including a discrimination coding process. In addition to the uses of ternary code in LTP in replacement of binary code in LBP. A considerable improvement was observed in the recognition rate when 7-classes (88.9) are used. Gradient local ternary pattern (GLTP) was introduced in the recent works to combine the good qualities of the former techniques [44]. GLTP uses a three-level discrimination ternary coding schema of gradient magnitude value for encoding image texture [58] and successfully gained 99.3 with 6-classes and 97.6 with 7-classes as the recognition rate. Nevertheless, despite the high results achieved by these researches, there are still some disadvantages that are observed from the uses of static global threshold has effect on background and illumination variation, which lower the rate of accuracy. Moreover, the accuracy of Improve Gradient Ternary Pattern IGLTP is limited when just few components of the face are used to determine the expression and four cardinal directions (North, East, West, and South) are applied.

2.3.2 Un-supervised machine learning (Deep learning)

Deep learning is a new research field of machine learning system which uses Deep Neural Network (DNN) in classifying human face images into emotional categories. Convolutional neural networks (CNN) were broadly applied to address the challenges in classifying facial expression.

Studies have proven deep learning algorithms to be brilliant enough in computer vision tasks for identifying and classifying objects. Convolutional Neural Networks [57], [58] were evolved with the aim of easing the selecting feature processing and providing higher results than the preceding machine learning methods. CNN architectures have developed to such a state that they even in different images classification tasks, they surpass human performance. The Alex Net Convolutional Neural Network [59] consists of 5 convolution layers and 60 million parameters. Deep Learning networks also have an ability of classifying the facial expressions. Firstly, it is required to perform the facial region cropping from images to restrict external data. Provision of Alex Net involved Grayscale facial images as training data, and there was observation on the results.

2.3.2.1 Deep learning (State Of The Art)

Deep learning is a kind of machine learning in which a model learns for carrying out classification function direct from text, images, or sound. Deep learning is applied using a neural network architecture. The word deep indicates the number of layers in the network, more layers increase the depth of the network. Deep learning algorithm performs a repeated task, each time tweaking it a little to increase the result, so that it trains computer system to perform what comes to humans naturally: learning from examples [62]. Deep learning is the main technology for several applications like driverless cars, it enables them to distinguish pedestrians and to recognize a stop sign [63]. Recently, for a good reason deep learning has been getting lots of attention and giving results that were impossible before [64].

Deep learning models could get an accurate state-of-the-art that occasionally surpasses human ability. They have been trained by using a large number of identified data as well as neural network frameworks with number of layers [64]. Deep learning models have been trained by using large of identified data and neural network frameworks which automatically learn features from the data then manual extraction of features is not needed [65].

Recently, deep learning turned to be a favourite topic for researchers and it has won state-of-the-art success for several applications [60]. Deep learning aim to capture high-level abstractions of multiple nonlinear transformations and representations across hierarchical architectures. In this section, we briefly present some already used deep learning techniques for FER.

2.3.2.1.1 Convolution Neural Network (CNN)

CNN has been widely used in different applications of computer vision including FER. In early 21st century, in the FER literature many studies[61],[62] proved the CNN to be competent enough to face changes in location and scale, and to highly perform than the Multilayer perception (MLP) in terms of past hidden variations in face pose.[63] The CNN used to resolve the equality problem of the subject, rotation, encoding, and invariance of scale in the facial expression identification.

A CNN is consisting of three dissimilar types of layers: convolutional layers, pooling layers, and fully connected layers. The ***convolutional layer*** possesses a group of learnable filters to join together through the entire input image and gather various types of feature maps for activation. There are three advantages for convolution task: local connectivity, which learns close connections between neighbouring pixels; weight sharing on the same feature map, which highly decreases the number of parameters to be learned; and shift-invariance to object location. The two most widely used nonlinear down-sampling methods for invariance to translation are ***average pooling and max pooling***. Basically the ***fully connected layer*** is involved at the end of the network for the assurance of completion of all neurons in the layer connected to activations in the past layer and for transformation of the 2D feature maps into 1D feature maps for further feature representation.

Both Alex-Net and VGG-Net [61] were winning architectures for ImageNet competition. Though both networks use similar conv-relu-pool-fc structure, they are different in many aspects. The most obvious difference is their depth. Alex-Net has 8 layers while VGG-Net has 16 layers. Configurations and features of some these models that have been applied to FER are listed in Table 2. 1:

Table 2. 1 Comparison of CNN models and their achievements

	Alex-Net [59]	VGG-Net [64]	Google-Net [65]	Res- [

Year	2012	2014	2014	2015
# of layers	5+3	13/6+3	21+1	151+
Kernel size	11,5,3	3	7,1,3,5	7,1,3,

Moreover, beside of these networks, there are also several noted derived systems. In [66], [67],region-based CNN (R-CNN) [68] was employed to the learn FER functions. In [69], Faster R-CNN[70] was used for facial expression classification through the generation of proposals for high quality regions. Moreover, Ji et al .[71] Introduced 3D CNN to capture adjacent information frames of multiple-encoded motion for action recognition through 3D convolutions. Tran et al. [68].

2.3.2.2 Limitation

The main difficulties in FERs like obtaining optimum pre-processing, feature extraction and classification, most especially under variable conditions of input data, the head pose, environmental disorder and illumination, various causes of variation in face. Problems remain even when deep learning is applied to FER despite its feature learning ability. Firstly, many training data are required by deep neural networks to be free from over fitting. Nevertheless, the subsisting facial expression databases are insufficient for training the well-known variations in pose, illumination and occlusions are common in unconstrained scenarios of facial expression. These factors are connected with facial expressions longitudinally and hence empower the deep network necessity to overcome variation in the large intra-class and to train active expression precise symbolizations. Deep learning methods have achieved successful application for extracting and classification features, in precise Convolutional Neural Networks (CNN) architectures which are biologically inspired multi-stage one that learned automatically hierarchies of non-variant features [72].

CHAPTER 3

METHODS AND TECHNIQUES

INTORODUCTION

Different methods and techniques are used to obtain the expression recognition of the facial image and to identify the correct expression from the image. In this chapter, the methods and techniques for facial expression recognition are generally discussed.

IMAGE REPRESENTATION FOR FACE MODELLING

This section presents various image representation approaches for recognizing facial expression. From previous studies they can model expression recognition through the use anthropometric data, Active Appearance Model (AAM) parameters, expression Pattern Subspace, manifold learning, appearance features or a hybrid of two or more modelling technique. An overview of these modelling techniques is presented in the subsequent sections.

3.2.1 Anthropometric model

Anthropometric model for facial expression concentrates on measuring distance between facial points. Face anthropometry is a study in which size and proportion of human face are measured [73]. Farkas [73] descriped face anthropometric according to measurement taken from 57 landmark points on human faces.

Some useful points for face description are shown in Fig 3.1. This research can identify landmark points by observing their respective anthropometric name. For instance, eye inner corner is ^{en} for endocanthier while front of the ear is for tragion. Farks defined five measurement between landmarks; shortest distance, axial distance, tangential distance, angle of inclination and angle between locations[74]. Farkas defined total number of 132 facial measurement [73], meanwhile he paired some similar measurement on both sides of the face. This research can take the measurement by hand or 3D scanners [75], [73].

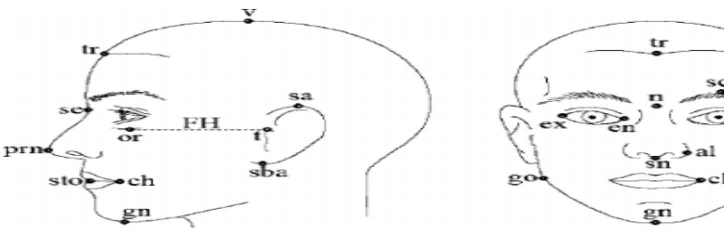


Fig 3. 1 Anthropometric points on the face [76].

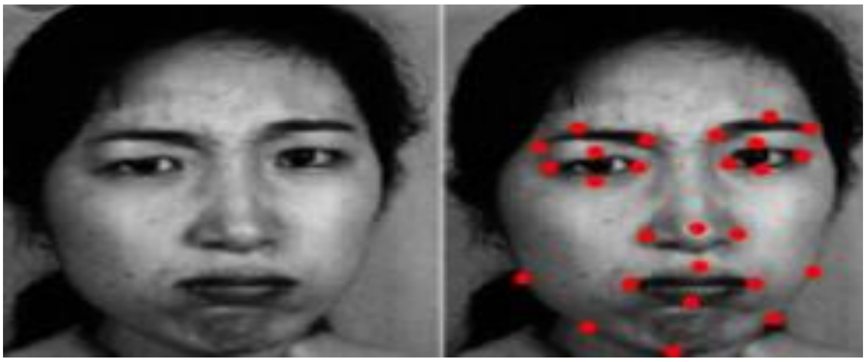
These measurements could be taken for face expression recognition. Ration of distance between facial landmark like eyes, nose, cheeks, mouth, and forehead are measured across expression.

$$R` = R\,(1+k\,(1-\cos\Theta)) \qquad 3.1$$

Where R is the initial radius of the circle, θ is the initial angle formed with the vertical axis, k is a parameter that increases with time. The mathematical formulation in equation above is not severally used in expression recognition for not encoding head profile [77] whereas it is difficult to estimate head profiles from 2D facial images [78].In addition, in some expressions without significant changes in facial shape, anthropometric models cannot be used for expression modelling. This approach is suitable in use only for frontal face images due to the sensitivity of distance between landmarks to head poses. This technique has not been experimented on a large public available database, with few number of reported studies in the literature working on few private datasets. This approach is also limit by considering only distance between facial landmarks without considering facial appearance. Measurements and landmark points defined by Farkas in [79], which often guide modelling technique are from people in one ethnic group (European) and may not be able to represent all other races.

3.2.2 Active Shape Model

Active Shape Model (ASM) [22] is a statistical model which describes characters of an object shape. ASM builds a model by learning patterns of variation from a training set of correctly annotated images. On a contrary to Active Contour Models (ACM), ASMs have got the ability of capturing natural variability of images of the same class [80]ASM are precise to images of the object classes they represent. A generation of landmark points signify face image shape. Points at corners of face and facial landmark Boundaries are the better landmark points. This research can determine these points by using proper 2D landmark algorithm like one introduced in [81]To reduce the variance in distance between points with the same value the groups of points are aligned automatically. These number of landmark points must be competent enough to give general shape of the face images. Each face is then represented by a forecast number of landmark points based on complicacy of the facial shape and the coveted level of descriptive information. Examination of spatial statistics points with label provides a Point Distribution Mode (PDM). PDM provides main locations of points and a set of parameters that control major variability modes detected from the training set. Providing such a model and test image, choosing values for each parameter in which we detect the appropriate model to the image requires image interpretation. ASM enables initial rough guess of best shape,



orientation, scale and position which is developed by comparison of hypothesized model instance to image data and using model and image differences to contort to shape. ASM is almost as same as AAM but contrast in the sense that instances in ASM can only deform in according to detected variations from the training set. ASM is not mostly used when recognizing expression therefore, it is essential to carry out more investigation adopting this modelling strategy.

3.2.3 Active Appearance Model

Active Appearance Models (AAM) [22] are statistical codding models for facial image. Using Principal Component Analysis (PCA), AAM learns shape model and intensity model from a set of training images. AAMs have been widely used to model facial shape for face recognition, face verification, gender estimation, age estimation and others. AAM regards both facial shape and texture contrast to anthropometric models which only regards shape parameters.

Fig 3. 2 Facial shape and appearance annotation [81]

AAMs most suitable for age estimation modelling at all stages right from infancy till old age due to the consideration of both shape and texture. AMM label each test image with a definite age label from continuous age range, this enables the approaches to present exact age estimations. Annotated set of training images are identified with points, definite facial main features are necessitated in Fig 3. 3

If 2D landmark algorithm is used appropriately like the one introduced in then we can determine the points [81]. Before building a model statistical shape the points are represented as a vector and aligned. Then each training image is deformed in order to match the annotated points with points of mean shape and obtain a shape-free image patch. The shape-free raster is pushed into a texture vector which is normalized through a linear transformation. After normalization a texture model is developed by using Principal Component Analysis (PCA). Finally, shape and texture connections are learned to provide a combined appearance model as detailed in [82].

Active Shape Models (ASM) are faster compared to AAM [22]. We can get details of AAM application in [22].AAM face encoding regards both shape and texture contrast to anthropometric techniques that consider shape only. This resulted in making AAM approaches appropriate for recognizing facial expression since both texture and shape features are necessitated in getting exact expression recognition. Nevertheless, showing that expression patterns can be modelled as a quadratic function and highlight effect of outliers in expression recognition needs evidence.

3.2.4 Appearance Model

Appearance models more specific in modelling facial appearance by using shape, wrinkle and texture features for recognizing facial expression, face verification, face recognition, gender estimation and age estimation among other tasks. Image is represented by vectoring both shape and texture[83]. Appearance models are more like AAM [84]that uses shape and texture of the face to build a statistical model. Both local and global texture, shape and wrinkle features are extracted and modelled for expression recognition. Texture and shape has been used for expression recognition [85], [86]. Ace expression recognition using appearance features can be improved by performing gender estimation prior since males and females exhibit varied ageing patterns.

After features are extracted and associated with expression label, they are used to estimate expression with either the use of a regression model or classification. Efficiency of LBP [87]in characterizing texture has made it popular in extracting appearance features for expression recognition. LBP has been used in [88] and scored accuracy of 80% in expression recognition when used with nearest neighbour classifier, and accuracy of 80%-90% when with AdaBoost classifier [89]. Gao an Ai [90] reported higher results when using Gabor filter appearance feature extraction technique for expression recognition compared to LBP technique. BIF [91] is also used in models based on appearance as used in [42]. This proves BIFs greater performance in expression recognition. Spatially Flexible Patch (SFP) introduced in [51] and [43] is another feature descriptor that can be used for characterizing appearance for expression recognition. Other techniques that can be used to build appearance models for expression recognition are Linear

Discriminant Analysis (LDA) and Principal Component Analysis (PCA). Detailed description of these techniques is presented in feature extraction techniques section.

3.2.5 Hybrid model

What is the best modelling approach for face expression recognition? The question is really difficult to answer since each of the modelling approaches discussed above have their essential strengths and limitations. Getting answer to this question requires one to try using several modelling approaches on the images, then their performance should be compared. Comparison of several modelling approaches, including their strengths and limitations would identify the best among them. Some modelling approaches complement each other and can be combined to form a hybrid modelling approach. Hybrid face expression modelling perform different modelling technique combination to exploit strengths of each technique. With combination of different modelling techniques, accuracies of face expression recognition are expected to be boost and robust. In this research can combine these model in parallel or a hierarchical manner and results from different models combined for final face expression recognition.

FACE EXPRESSION EXTRACTION TECHNIQUES

Feature selection includes various algorithms such as:

3.3.1 Gabor filter

Denis Gabor proposed Gabor filter in 1946 [92], it has been widely applied for wrinkle, edge and texture feature extraction because of its ability to determine orientation and magnitude of wrinkles [93]. Gabor filter has been considered as the best texture descriptor in object recognition, segmentation, tracking of motion, and image registration [94]. Gabor features was also used in age estimation [95] and it was proved to be more efficient texture descriptor comparing to LBP. Since wrinkles appear as edge-like components with high frequency, Gabor edge analysis technique has been commonly used for wrinkle features extraction. Sobel filter [122], [123] Hough transform [96] and active contours [61]are some of most typically used texture edge descriptors. Though, edges in a face image also contain noise like hair, moustache beards and shadows. To reduce the effect of this noise, [93]introduces the use of dominant orientation of wrinkles to be put into consideration when extracting wrinkle features. 2D spatial domain Gabor is defined as Eq 3. 1

$$g(x,y) = \left(\frac{1}{2\pi\sigma_x\sigma_y}\right) exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma 2_x} + \frac{y^2}{\sigma 2_y}\right) + 2\pi jWx\right]$$

Eq 3. 2

Where σ_x and σ_y are the standard deviations of the distribution along x and y axes respectively and W is the sinusoidal radial frequency. Fourier transform of the function in Eq 3. 3 is expressed as

$$\text{G (u,v) = exp}\left[-\frac{1}{2}\left(\frac{(u-W)^2}{\sigma_u^2} + \frac{v^2}{\sigma_u^2}\right)\right]$$

Eq 3. 4

Where $\sigma_u = \frac{1}{2}\pi\sigma_r$ and $\sigma_v = \frac{1}{2}\pi\sigma_y$. Gabor filter bank is obtained by rotating and dilating g(x,y). We could express the general equation for creation of Gabor filter bank as

Gb (r , y) = $a^{-m} g\ (\bar{r}, \bar{y})$ Eq 3. 5

Where

$\bar{r}$ = r cos 0+y Eq 3. 6

&

$\bar{y}$ = r sin 0 + y cos 0 Eq 3. 7

$$where\ 0_k = \pi \frac{(k-1)}{n}, k = 1,2,3 \ldots n$$

where n is the number of orientations used and a^{-m} is filter scale for $m = 0,1,2 \ldots s$ for S scales. Designing Gabor wavelets prevent redundancy in the frequency domain as in Eq 3. 8

$$\sigma_u = \frac{\left(\left(\frac{U_h}{U_t}\right)^{\frac{1}{(1-a)}} - 1\right) U_h}{\left(\left(\frac{U_h}{U_t}\right)^{\frac{1}{(a-1)}} + 1\right)}$$

Eq 3. 9

where

$$\sigma_v = \tan\left(\frac{\pi}{2k}\right) \left[U_h - 2In\left(\frac{\sigma_u^2}{U_h}\right)\right] \left[2\ In\ 2 - \frac{(2In\ 2)^2\ \sigma_u^2}{U_h^2}\right]^{0.5}$$

Eq 3. 10

3.3.2 Liner discrimination

Linear Discriminant Analysis (LDA) [97], [98] is a technique for extracting features that searches for the best discriminant features among classes. Providing a set of features that are independent, LDA creates a linear combination of these features in such sense of achieving largest mean differences between classes. LDA expresses two measures; with-in class scatter matrix, given by Eq 3. 11 :

$$S_w = \sum_{j=1}^{c} \sum_{i=1}^{N_j} \left(r_i^j - \mu\right)\left(r_i^j - \mu_j\right)^T$$

Eq 3. 12

Where r_i^j is i^{th} sample of class j, μ_j is the mean of class j, c is number of classes and N_j is the number of samples class j; and between-class scatter matrix, given byDenis Gabor proposed Gabor filter in 1946 [92], it has been widely applied for wrinkle, edge and texture feature extraction because of its ability to determine orientation and magnitude of wrinkles [93]. Gabor filter has been considered as the best texture descriptor in object recognition, segmentation, tracking of motion, and image registration [94]. Gabor features was also used in age estimation [95] and it was proved to be more efficient texture descriptor comparing to LBP. Since wrinkles appear as edge-like components with high frequency, Gabor edge analysis technique has been commonly used for wrinkle features extraction. Sobel filter [122], [123] Hough transform [96] and active contours [61]are some of most typically used texture edge descriptors. Though, edges in a face image also contain noise like hair, moustache beards and shadows. To reduce the effect of this noise, [93]introduces the use of dominant orientation of wrinkles to be put into consideration when extracting wrinkle features. 2D spatial domain Gabor is defined as Eq 3. 1

$$S_b = \sum_{j=1}^{c} \left(\mu_j - \mu\right)\left(\mu_j - \mu\right)^T$$

Eq 3. 13

Where μ is the mean of all classes. A major aim of LDA is maximizing between class scatter matrix while minimizing within-class scatter matrix.

This is done by maximizing the ration $\frac{det|S_b|}{det|S_w|}$. Given that S_w is non-singular, studies have shown that this ratio is maximized when column vectors of projection matrix are the eigenvectors of $S_w^{-1}S_b$. S_w maximum rank is N-c with N samples and c classes. Hence, this demands N=t+c samples to affirm that S_w does not change singular, where t is the dimensionality of imput data. The number of samples N is almost always smaller than t, making the scatter matrix S_w singular. To address this issue [99]and [100] proposed projecting input data to PCA subspace, by reducing dimensionality to N − c, or less than it before application of LDA. LDA and PCA are most commonly used method for extracting appearance features in pattern recognition [101]. Therefore, we adopted LDA to extract global face appearance features for age-group estimation.

3.3.3 Local Binary Pattern

Texture features have been broadly used as techniques in age estimation [67]. Local Binary patterns (LBP) is a texture description technique which detects microstructure patterns like lines, edges, spots and flat areas on the skin [46]. We use LBP for texture description of face detection, gender classification, face recognition, age estimation, tracking of face and facial components. Gunay and Nabiyev [107] used LBP for texture feature characterization for age estimation. The result was 80% accuracy on FERET [118] dataset the use of nearest neighbour classifier and accuracy of 80 – 90% on FERET and PIE datasets using AdaBoost classifier [119]. A binary number is formed by interlinking all eight bits. And we convert the resulted binary number to a decimal and assigned to centre pixel as its LBP code.

Ojala et al. [45] observed that when 8 neighbours are used with radius 1, 90% of all patterns formed uniform patterns. The original LBP operator captures dominant features with large scale structures to a certain extent. The operator was latter further in capturing texture features with neighbourhood of various radii [45]. Neighbourhood was defined by a set of sampling pixels distributed evenly along circle circumference centred at the pixel to be labelled. Bilinear interpolation of points which do not fall in the pixels is done to enable any radii and any number of sampling pixels.

MACHINE LEARNINIG

It is divided into: supervised and unsupervised learning. The aim of supervised learning is to learn mapping from the input (instance) to an output (target or label) in which a supervisor provided its correct values [102]

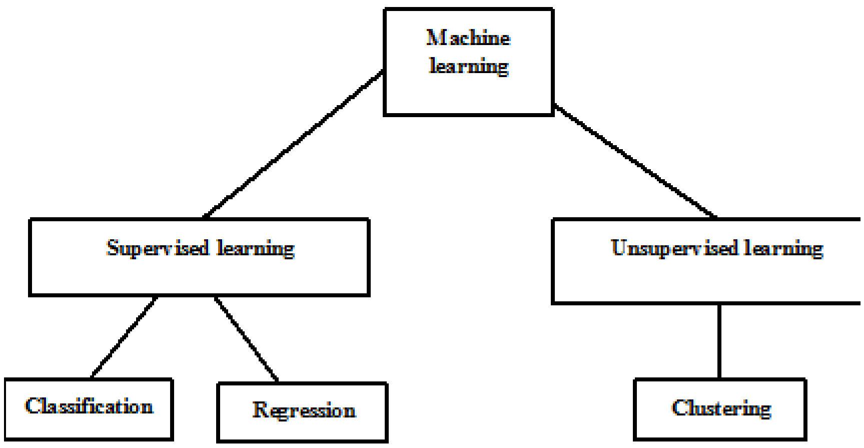


Fig 3. 4 Illustrates types of machine learning.

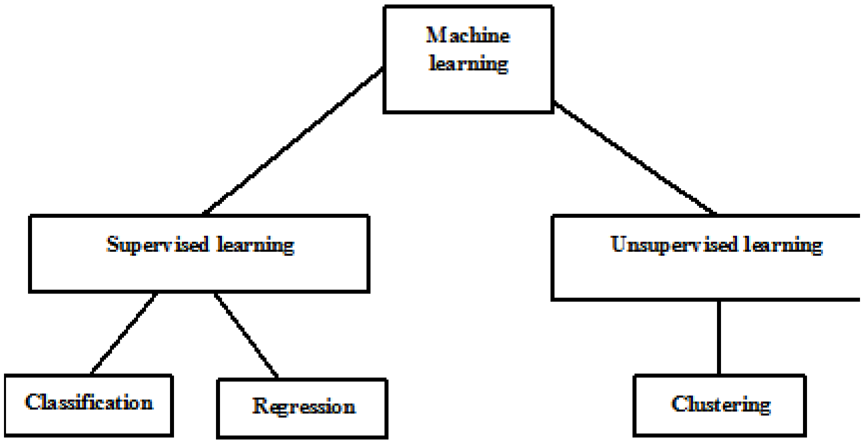


Fig 3. 4 Machine learning types [103]

3.4.1 Supervised machine learning

It’s the last step of a facial expression recognition system that recognizes the class of expression according well selected features. There are common classifier methods such as the Artificial Neural Network (ANN), Support Vector Machine (SVM), and Decision Tree (DT), Bayesian Networks (BNs), K-Nearest Neighbour (KNN), and Hidden Markov Model (HMM). Target and label of sample data are unknown in unsupervised learning. Unsupervised learning methods are designed to summarize the key features of the data and to form the natural clusters of input patterns when given a cost function. The most well-known methods of unsupervised learning are k-means clustering, hierarchical clustering, and self- organizing map. It is hard to evaluate unsupervised learning, because it has no control, consequently, it does not have labelled data for testing [30].

3.4.1.1 Support Vector Machine (SVM)

It is another machine learning technique which can be classified as Kernel-based methods. The main of SVM is minimize an upper bound of the generalization error by maximizing the margin between the separating hyper planes as presented in Fig 3. 5

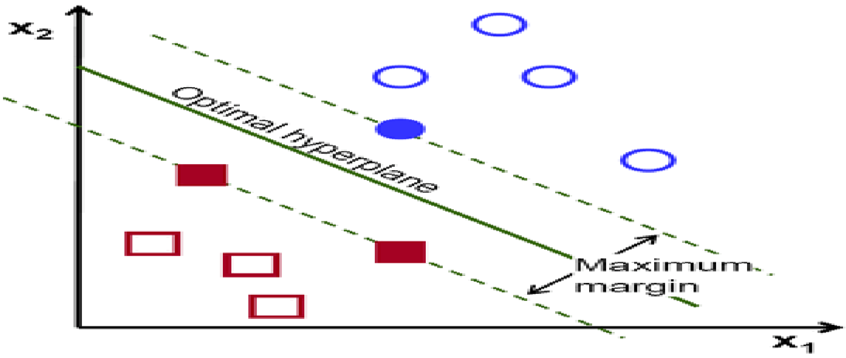


Fig 3. 5 6mmaximizes the Margin between the Hyper Plane and the Data [104]

It depends on the statistical learning theory and it was created to address binary classification challenges. A supervised classification function includes separation of data into testing and training set. Each training set representative contains the class label(Y) and the features(X). The aim of the SVM is to produce a model, according to the training data, which indicates the target values of the test data given only the test data features, the training set of instance-label pairs (x_i, y_i), i = 1… L where xi ε $R^{\wedge n}$ and y ε (1,-1) or (0, 1). The training vectors xi is mapped into a higher or infinite dimensional space. The SVM works on linear separating hyper plane with the maxim margin in this higher dimensional space by the use of the kernel function. Kernel functions are the determined components before the training stage. Which are central concepts for many learning tasks. SVM and the kernel architecture are used currently in a various areas, including multimedia information retrieval, Bioinformatics and pattern recognition. Choosing the most suitable kernel is closely related to the problem and data set because it based on our aimed model. Next section shows the common types of kernel function:

Linear

$$K(xi, xj) = (xi.xj)^T \qquad 3.2$$

Radial basis function (RBF)

$$K(xi, xj) = exp(-ɣ\|xi - xj\|^2) \;\; ɣ > 0. \qquad 3.3$$

Sigmoid

$$\mathbf{K}(\mathrm{xi}, \mathrm{xj}) = \tanh(ɣ.\mathrm{xi}.\mathrm{xj} + \mathrm{r}). \qquad 3.4$$

Polynomial

$$(xi, xj) = (ɣ.\mathrm{xi}.\mathrm{xj} + \mathrm{r})^d, ɣ > 0 \qquad 3.5$$

3.4.1.2 Artificial Neural networks

Artificial Neural Networks are very effective computational models. Which have been applied in different areas like computing, medicine, engineering, economics, and many others. ANN is a computational model that active in function to the human brain. It is composed by a set of artificial neurons (known as processing units) that are interconnected with other neuron, these neurons based on the neural network weights. The word network in neural network indicates the interconnection between neurons existing in different layers of a system. McCulloch and Pitts proposed the first artificial neuron in 1943 in form of formal model. They proved that this neuron model could perform any computational task using a finite number of artificial neurons and synaptic weights adjustable [105]. A neural network structure is formed by an “input” layer, one or more “hidden” layers, and the “output” layer. The neurons in the input layer receive and transfer the data to neurons in the first hidden layer through the weighted connections. A mathematical processing of data is performed and transferred the result to the neurons in the next layer. The network’s output is provided by the neurons in the last layer .The $\mathrm{j}^{-\mathrm{th}}$ neuron in hidden layer processes of the incoming data (xi) by: (i) calculating the weighted sum and adding a “bias” term (θ_j) based on Eq3. 6 :

$$netj = \sum_{i=1}^{m} xi * wij + \theta i \quad j = (1,2, \dots., n) \tag{3. 7}$$

ANN can be defined based on the following three features[106]:

1. The Architecture indicating the number of layers and the number of nodes in each layer.
2. The learning mechanism (supervised learning or unsupervised learning) applied for updating the weights of the connections.
3. The activation functions used in various layers.

A radial basis function network is an artificial neural network that uses radial basis functions as activation functions. The input radial basis functions and neuron parameters are linearly combined as the output of the network see Fig 3. 7.

3.4.1.3 Neural Network Algorithms

A radial basis function network is an artificial neural network that uses radial basis functions as activation functions. The input radial basis functions and neuron parameters are linearly combined as the output of the network.

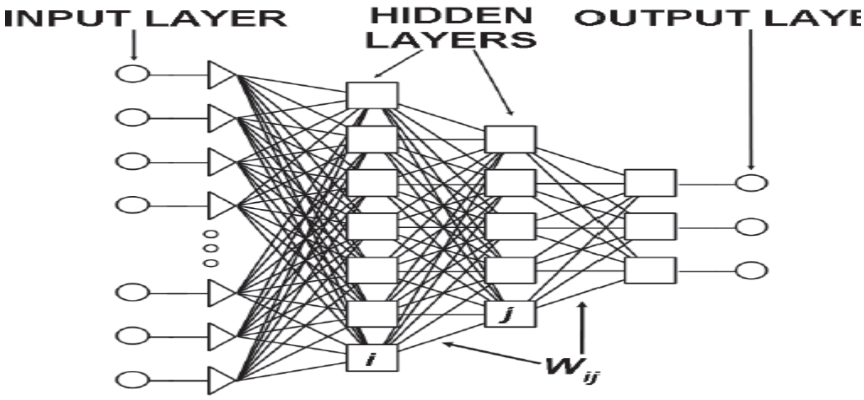


Fig 3. 7 General structure of a neural network with two hidden layers [106].

3.4.1.3.1 A self-organizing map

A self-organizing map (SOM) is also kind of Artificial Neural Network that is trained to produce a low-dimensional (typically two-dimensional) with the use of unsupervised learning, representation of the input space of the training samples and named a map. Self-organizing maps are different from other artificial neural networks in a way that the input space topological properties are maintained by using a neighbourhood function. There is an idea of combination of the SOM with the elastic net algorithm to overcome Euclidean issues like the travelling salesman problem [107].

3.4.1.3.2 Back Propagation Neural

Back propagation ANNs represented by nonlinear networks. The back propagation (BP) algorithm used for training multilayer networks by means of error propagation through several calculations. It uses a gradient descent technique to minimize the sum of squared approximation errors.

3.4.1.3.3 Hop-field Neural

Hop-field networks are used in solving optimization problems. Although there is no layer structure to the architecture, and the weights are constants and symmetric. Hop-field networks are a fully interlinked system of N neurons. Information about the memories or stable states of the network are stored by the weights of the network. Each neuron has a state xi which is bounded between 0 and 1. Neurons are updated according to a different equation, and due to the state stability of the network, an energy function is reduced after some time [108].

3.4.2 Unsupervised learning (Face expression using deep learning method)

Deep learning is a new research field of machine learning system which uses Deep Neural Network (DNN) in classifying human face images into emotional categories. Convolutional neural networks (CNN) were broadly applied to address the challenges in classifying facial expression.

Studies have proven deep learning algorithms to be brilliant enough in computer vision tasks for identifying and classifying objects. Convolutional Neural Networks [57], [58] were evolved with the aim of easing the selecting feature processing and providing higher results than the preceding machine learning methods. CNN architectures have developed to such a state that they even in different images classification tasks, they surpass human performance. The Alex Net Convolutional Neural Network [59] consists of 5 convolution layers and 60 million parameters. Deep Learning networks also have an ability of classifying the facial expressions. Firstly, it is required to perform the facial region cropping from images to restrict external data. Provision of Alex Net involved Grayscale facial images as training data, and there was observation on the results.

Human facial expressions have been recognized in several cases by different methods mainly with the use of the Facial Action Coding System (FACS) [109] as a guide. We need to firstly detect action units from images before we predict expressions based on FACS, using action detection methods like facial feature point tracking, dense flow tracking or using grey value changes. This extracted data extracted is then forward propagated through the recognition classifier for facial expression. These approaches are expensive in terms of computation and numbers of different blocks are required before having actual prediction. This research introduce a new method which implements a deep learning algorithm facial expression recognition and the image product of cropped faces and their components are used.

Many tasks have been done to recognize facial expressions using the Facial Action Coding System. This has been obviously shown in works like [110], [111], [112]. FACS classifies emotions based on different Action Units (AU) on the human face. Various action units make up a single expression. These approaches have also showcased high results. Other approaches like Artificial Neural Networks, Hidden Markov Model (HMM) and Bayesian Networks have also been used for facial expression recognition [113], [114]. This research rather used a

Convolutional Neural Network instead for classifying these images and the result evaluation. The key features described by the FACS is not applicable in CNN. The general approach used for facial expression recognition need to be explored in order to understand the deep learning approach. First of all this research perform image registration. Registration uses face detection algorithms to find out the faces in the image in order to be localized [115] after that, the localized faces are matched to a template image. After that, feature extraction algorithms such as the Histogram of Oriented Gradients (HOG) [116], Local Binary Patterns (LBP) [7], and Gabor filters [117] are applied to these registered images. The algorithm produces a feature vector which is then fed to the classifier for classification. Deep Convolutional Neural Networks on the other hand do not require us to define the feature extraction algorithms to be used. The network itself during training learns the weights, biases and the kernels to be convoluted with the image for feature extraction [106]. Such approaches have been used in the past in [118], [119], [120] . These approaches are highly variable when used in the CNN architecture. This research present the commonly used methods in FER:

3.4.2.1 Alex-Net

The architecture consists of eight layers: five convolutional layers and three fully-connected layers. But this isn't what makes AlexNet special; these are some of the features used that are new approaches to convolutional neural networks. In Alex-Net, the input is an image of size 227x227x3. After Conv-1, the size of changes to 55x55x96 which is transformed to 27x27x96 after MaxPool-1. After Conv-2, the size changes to 27x27x256 and following MaxPool-2 it changes to 13x13x256. Conv-3 transforms it to a size of 13x13x384, while Conv-4 preserves the size and Conv-5 changes the size back go 27x27x256. Finally, MaxPool-3 reduces the size to 6x6x256. This image feeds into FC-1 which transforms it into a vector of size 4096×1. The size remains unchanged through FC-2, and finally, this research get the output of size 1000×1 after FC-3.

3.4.2.2 VGG-Net

The VGG network architecture was proposed by Simonyan and Zisserman[121], Very Deep Convolutional Networks for Large Scale Image Recognition. This network is known by its simplicity, using only 3×3 convolutional layers piled on top of each other in increasing depth[64]. It has (224 * 224) as the fixed size RGB image was given as input to this network which means that the matrix was of shape (224,224,3).

3.4.2.3 Res-Net

It is a short form of Residual Networks which is a classic neural network, it has been used as an essential model for many computer vision tasks. And it won in the competition of ImageNet challenge in 2015. The fundamental improvement with ResNet was enabling us to train extremely deep neural networks with 150+layers successfully. This research use ResNet-50 and ResNet-101 for The ResNet-50 model contain 5 stages each with a convolution and Identity block. There are 3 convolutional layers for each convolution block and each identity block also has 3 convolutional layers. Over 23 million trainable parameters are observed for ResNet-50.

FACE EXPRESSION EVALUATION PROTOCOL

Evaluation protocol evaluates system test, system performance measurement and criteria for testing data selection. Independency of training data and representative is good description of a good validation strategy should be independent of training data and representative [122]. Expression recognition technique needs to be confirmed by using the past unseen data to overcome over fitting expression recognition technique and improve its generalization ability. Cross validation is a commonly used strategy for expression recognition evaluation. In cross validation, data is separate into two subsets; one segment is used to train or learn expression recognition model and the other segment is used for the model evaluation. In traditional cross-validation, training and validation datasets must cross-over in successive rounds in sense that each data point has equal chance of being validated or evaluated against the other. The basic form of validation is holdout.

Holdout strategy is the simplest and computational effective strategy [123] used to validate expression recognition techniques. The dataset is randomly separate into two sets; training subset and validation subset. Commonly training subset consist of $\frac{2}{3}$ of the original data and remaining $\frac{1}{3}$ samples constitute validation subset. Expression recognition model is latter fitted by using the training subset and validated on the test subset. In this strategy training and validation are only once for the model. Despite preference of this method is preferred and its very short time consuming in computation, but its evaluation based on the data in respective subsets which results into high variance therefore, the way of dividing the datasets affects the evaluation results [124]. Another validation strategy commonly used is repeated random sub-sampling (RSS) [125], [126]. In RSS validation technique, the holdout strategy is repeated for several times and results averaged. The dataset is randomly split into two subsets (train and validation) with certain number of samples for each phase of validation. For each data split, expression recognition model is retrained on train subset and validated using test subset. The advantage of this strategy over k-fold validation is that the size of training and validation is not in anyways depending on the number of validation iterations. Nevertheless, the drawback of this strategy is that some samples might be selected repetitively while others might not be selected at all which lead to overlapping of validation subsets [127]. But with significantly large number of iterations done, RSS is could possibly achieve higher results as k-fold validation [128].

Cross validation [127] is a standard statistical technique used for model generalization ability and widely applied in classification and regression challenges [129] . It involves dividing dataset into two subsets, one subset is used for training an estimator and the other subset is used for testing an estimator [130]. Cross validation is used to evaluate generalization method of model to initially hidden data [127], [131]. Cross-validation strategies can be categorized into two: i) exhaustive (compute all possible ways of data splitting) and ii) non-exhaustive (does not compute all possible ways of data splitting). Exhaustive cross-validation algorithms include Leave-One-Out (LOO) and Leave-P-Out (LPO) while non-exhaustive include k-fold and repeated random subsampling (RSS) [124], [132]. Cross-validation [133] contain averaging multiple holdout validation results from different subsets of data.

K-fold cross-validation is the basic form of cross-validation.

The rest forms of cross validation are considered as just but special cases of k-fold cross-validation or involve repeated rounds of k-fold validation. In k-fold cross-validation [133] , original data is randomly separated into k equal subsets. Then, perform k repetitions of training and validation in sense that in each repetition, a different fold of data is reserved for validation while the last k-1 are used for a model learning. The assessed error is the mean of all validation errors. Confidence range of the estimate can be approximated by standard deviation of those errors. The major benefit of k-fold cross-validation is that all samples will be eventually used for both learning and validating a model. Common value of k used in various techniques is 10 as a compromise betthis researchen accuracy and efficiency. A stratified cross-validation is commonly used to boost the estimation accuracy [127].

Leave-One-Out (LOO) [133], [130] is a special type of cross-validation that given a dataset with C classes, $C - 1$ validation experiments are performed. For each experiment, data from $C - 1$ classes is used for training and data from one class that was left out is used for validation. Hence, given a dataset of S subjects from expression $0 \rightarrow A_n$, LOO cross validation will perform $S - 1$ validation experiments. In each experiment i, facial images of subject S_i are used for validation while images of the rest $S - 1$ subjects are used for learning a model. In this approach, images of each subject will be used for both training and validation. This way, the technique is validated in the same way as its application scenario where the subject whose expression is to be estimated is previously unseen in the system. Although LOO has never been proved of being bias, but because of its high variation, the given estimation may sometimes unreliable [134]. Leave-p-Out (LPO) [135] with $p \in \{1, 2, 3 \ldots$

, n − 1} successively leaves out every possible subset of p data samples to be used for validation. In expression recognition, given a set of images of N subjects, LPO can be used by leaving out images of p where $p \leq (N - 1)$ subjects to be used for validation and use images of $N - p$ subjects for training. Elisseef and Pontil [136] proved that LPO cross-validation is more trusted of being unbiased than LOO. LPO will have n k iterations where n is the number of images. These iteration of LPO are far higher than $n - 1$ iterations in LOO, leading to high computation time. There is similarity between $p = 1$ of LPO and the one of LOO. LOO and LPO compare to other method are exhaustive cross-validation strategies. More information on LPO can be found in [137]. Detailed information on cross-validation can be found in [135] and [138].

Efron and Tibshirani proposed Bootstrap strategy [139] . Bootstrap is preferably used when a dataset is small [140]. In this strategy, a bootstrap set is formed by uniformly sampling, with replacement, n instances from the original data to make a training set. The remaining unselected samples are used for testing set. Changing of the value n of selected samples is from fold to fold is likely. Since data is sampled with replacement, the probability of any data unselected sample given by $(1 - \frac{1}{n})^{n} \approx e^{-1} \approx 0.368$. Chances of a data sample being selected into a train set is $(1 - 0.368) = 0.632$. Therefore, expected number of distinct samples appearing in the train set is $0.632 \times n$. Since error estimate obtained by using test data will be too pessimistic (since only 62.3% of instances are used for training), error is calculated as error $= 0.632 \times e\ 0 + 0.368 \times e_{bs}$ where e 0 is rate of error obtained from bootstrap sets not having the instance being predicted (test set error) and e_{bs} is the error obtained on bootstrap sets themselves, both averaged over all data samples and bootstrap samples. Estimate accuracy is directly proportional to number of times the process is repeated. Further information on bootstrap validation technique can be found in [141]. Bootstrapping increases the variance that could happen in every fold which makes bootstrap more realistic strategy of the real application situation [141]. This validation strategy is not commonly used in face expression recognition.

Mostly, a dataset is separated into 3 subsets: validation subset, training subset and testing subset [192]. In this approach, the validation subset is used to tune the system to determine termination point of the[128] training phase when the training subset starts over fitting. The testing subset is used for validating the trained model with the use of data samples not initially in validation and training subsets. Kiline and Uysal [189] introduced a technique on splitting the dataset with samples from specific subjects rotationally left out of training and validation sets. Budka and Gabrys [122] introduced a density-preserving sampling (DPS) technique which excludes the necessity of the procedures of repeating error estimation by dividing dataset into subsets that are guaranteed to be representative of the population the dataset is drawn from. These new introduced approaches of model validation could be tested in face expression recognition problem and compare the results with other known methods. Cross-validation and bootstrap strategies are mostly used when there is limited data such that holdout strategy is not enough for data representativeness in both training and test sets. With abundant data with stable distribution over time, single stratified random separated could give demanded representativeness [122].

When intending to compare metric performance of two or more learning algorithms Salzberg [142] introduced the use of k-fold cross-validation followed by appropriate hypothesis testing in replace of their average accuracy comparison. This strategy can be more of useful in comparing two techniques of face expression recognition. In each iteration of validation, Absolute Error (AE) for each estimated age is defined as in Eq 3. 8

$$AE = |a_i - \bar{a}_i| \qquad 3.\ 9$$

Where

a_i is the ground truth age and $\bar{a}_i$ is the estimated expression. After all validation iterations, Mean Absolute Error (MAE) is defined as the average of all absolute errors as in Eq 3. 10 :

$$MAE = \frac{1}{n}\sum_{i=1}^{N} |a_i - \bar{a}_i| \qquad 3.\ 11$$

Where

N is total number of test images, a_i is the truth expression of image i, and $\bar{a}_I$ the estimated expression of image i. Despite the commonly use of this performance evaluation, it only provides the technique general performance for all expressions instead of giving estimation performance for specific expression. This approach could be slightly modified such that it gives MAE of each age and General MAE of the technique.

Given a set of testing images $a_1^{n\ 1}$, $a_2^{n\ 12}$,………., $a_k^{n\ k}$ belonging to k ages to be estimated with n_i representing number of test images known to belong to age a_i, MAE for every expression can be defined as in Eq 3. 12:

$$MAE_k = \frac{1}{n}\sum_{i=1}^{n} |a_k - \bar{a}_k| \qquad 3.\ 13$$

Where

$\bar{a}$ k is the image age estimated i of expression a_k and n is the number of tested images that belong to expression a_k. This will give age specific performance of age estimation technique. Overall MAE can be found by summing all the MAE for all ages tested and divide by summing up number of test images with each expression as

$$MAE_{TOTAL} = \sum_{i=1}^{k} \frac{\text{MAEk} \times ni}{N} \qquad 3.\ 14$$

Where

$N = n_1 + n_2 + \ldots\ldots + n_k$

This research evaluates the performance of expression recognition techniques based on MAE. The expression estimation performs better when MAE is small. MAE only shows average performance of the expression estimation technique. When some images are missing in the training data, MAE is the suitable measure of expression estimation [143].

Additionally to those techniques for evaluation this research can meajer the performance by specificity, sensitivity, and F1 score. Sensitivity measures the proportion of actual positives samples that correctly identify, where the percentage of expression that is correctly classified as correct one. Hence, it is computed based on the following definition in Eq3. 15:

$$\text{Sensitivity} = \frac{TP}{TP+FN} \qquad 3.\ 16$$

Where TP (true positive) is the number of images which have been successfully detected, and FN (false negative) is the number of images which have been detected through the method. Differently, specificity measures the percentage of indicated negatives samples, in which the percentage normal images correctly classified as normal. In this manner, specificity computation as in Eq 3. 17 :

$$\text{Specificity} = \frac{TP}{TP+FN}$$

3. 18

Where TN (true negative) is the number of normal successfully classified patients, and FP (false positive) is the number of normal falsely classified patients as Pneumonia. And, F1-score measures the average F1 score through various class labels which is computed as in Eq 3.19

:

$$\text{F1} = 2 \times \frac{PPV \times TPR}{PPV + TPR}$$

3. 20

Where

$$\text{PPV} = \frac{TP}{TP+FP}$$

3. 21

And

$$\text{TPR} = \frac{TP}{TP+FN}$$

3. 22

The accuracy was the fraction of the predicted labels which were correctly computed as in Eq 3. 23 :

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad 3.\ 24$$

SUMMERY

This research gives comprehensive review of various techniques and approaches used for expression recognition. There has been enormous effort from both academia and industry dedicated towards modelling expression estimation, designing of algorithms, face expression dataset collection and protocols for evaluating system performance. Image representation and estimation techniques are the main issues to be considered in facial expression recognition. AAM provide a parametric modelling for face representation. A face represents a set of shape and texture parameters learnt from a face image. AAM can represent several faces expression since model parameters encode both facial shape and texture. AAM is often used in line based on regression estimation approaches. Anthropometric face representation encode change in facial shape.

In facial representation Anthropometric approaches are considered very important to capture facial shape changes. Expression manifold learning entwines expression feature extraction and dimensionality reduction. Expression manifold can be used in approaches based on both classification and regression. Appearance models often extract facial features that can be used in expression recognition approach based on classification or regression. These features represent facial appearance. These features could be texture, shape or wrinkle. Feature extraction techniques like LDP, LBP, PCA, Gabor, LDA and BIF are the commonly used for appearance face modelling.

Expression recognition can be either classification-based, regression-based or hybrid of both classification and regression. Size, face image representation and dataset expression distribution guide the choice between regression and classification. Regression-based and classification-based are both suitable to be used for enormous dataset with sequential expression-labels, while classification-based is more suitable for datasets with only expression labels. Both classification and regression can be combined in a hierarchical manner.

Cumulative Score (CS) or Mean Absolute Error can be used to evaluate techniques for expression recognition. MAE is appropriate when some ages are missing in the training set while CS is used when training dataset has samples at almost every age. Overall performance of the system is represented by CS. In practice, both MAE and CS are used because different techniques, datasets may be biased for evaluation. LOPO and Cross-Validation are the most frequently used as evaluation protocols[146].

CHAPTER 4
LOCAL PQTTREN AND DEEP LEARNIG

4.1 Local Pattern for features extraction

A feature is the numbering system used for numbers of an image representation, which can be computed directly from either intensity image or from other image features. The extracted features are well-conditioned to variations and easy to classify compared to intensity images . Feature extraction as a mathematical context is a projection from input space into features space. Figure2 illustrates samples that are arranged from the input space.

4.1.1 Local Binary Pattern LBP

It's is one of the most known algorithm in local facial features. This algorithm has been extensively applied in facial expression recognition [14], [15], [16]. Ojala et al. firstly introduced the LBP operator in generic form for feature extraction [16]. The LBP processing for feature extraction for the certain image begins by the LBP operator selecting a local neighbourhood around each pixel, thresholding the pixels of the neighbourhood at the centre value and the code will be constructed for each pixel. The resulting binary value image patch is latter used as local image descriptor [17]

Fig 4. 1 illustrates the basic LBP operator.

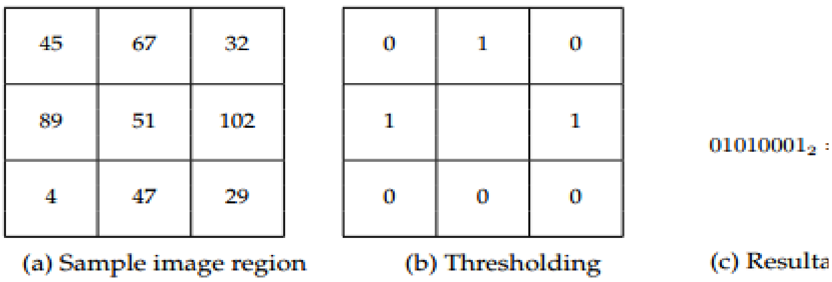


Fig 4. 1The basic LBP operator [16].

Where

g_c sygnifies the grey value of the centre pixel (x_c ,y_c) and g_p are grey values of P equally spaced pixels on the circumference of a circle with radius R.

4.1.2 Local Directional pattern LDP

There was an observation that local binary pattern (LBP) [17] was unstable to image noise and illumination variation. Jabid et al. [18] introduced local directional pattern (LDP) due to its robustness to image noise and non-monotonic variation in illumination.

LDP compute 8-bit binary code for each pixel in the image by comparing edge response for each pixel in different orientation rather than comparing raw pixel intensive as LBP. Kirsch edge detector [19], Prewitt edge detector [20], Sobel edge detector [21] and Scherr edge detector [22] are commonly used as edge detector [138]. Kirsch edge detector among others has been known to detect different directional edge responses with higher accuracy than others because of all eight neighbours it considers [23].

4.1.2.1 Kirsch mask operator

Kirsch is first -order derivation edge detector that get image gradient by convolving 3x3 image region with a set of masks. Kirsch defined a non-linear edge detector technique as [24]:

$$P(x, y) = max \quad \left\{1, \max_{k=0} \|5S_k - 3T_k\|\right\}$$

4. 1

Where

$$S_{k= P_k + P_{k+1} + P_{k+2}}$$

4. 2

&

$$T_{k} = P_{k+3} + P_{k+4} + P_{k+5} + P_{k+6} + P_{k+7}$$
4. 3

Where a in k_a is evaluated as a = a mod 8 and P_k [k = 0,1,2...,7] are eight neighbouring pixels of P (x, y) shown in Fig 1a. Image gradient towards a particular direction is found by convolving 3×3 image region with the respective mask M_k shown in Fig 4. 2

-3	-3	5
-3	0	5
-3	-3	5

East (M_0)

-3	5	5
-3	0	5
-3	-3	-3

East (M_1)

5	5	5
-3	0	-3
-3	-3	-3

East (M_2)

5	5	-3
5	0	-3
-3	-3	-3

East (M_3)

5	-3	-3
5	0	-3
5	-3	-3

East (M_4)

-3	-3	-3
5	0	-3
5	5	-3

East (M_5)

-3	-3	-3
-3	0	-3
5	5	5

East (M_6)

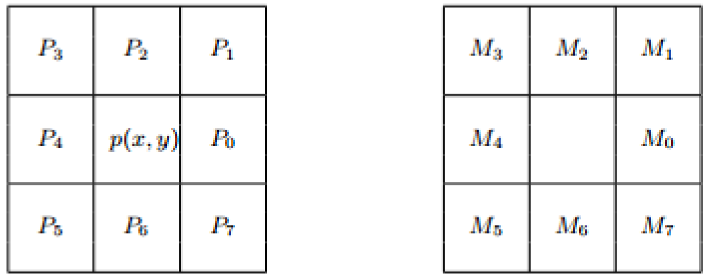
-3	-3	-3
-3	0	5
-3	5	5

East (M_7)

Fig 4. 2The Kirsch gradient in a direction is found by convolving 3x3image region with the respective masks Mk[147].

4.1.3 Local Ternary pattern LTP

Generally, the LBPs are considered to be appropriate for texture classifications and not affected by light illumination effect, because they are invariant, specifically to monotonic grey-level transformations [147]. Nevertheless, the LBPs are known to be sensitive to noise, particularly in near-uniform image regions.



(a) Eight neighbours of pixel p(x, y)

(b) Eight directional Kirsch mask positions

Fig 4. 3: (a) Eight neighbours of pixel p (x, y) and (b) corresponding Kirsch mask[147]

Additionally, they are also sensitive to weak illumination gradients. The factors of LBPs drawback is due to them threshold at exactly the central pixel value, which is signified here as i_c. In this section, the LBP is extended to 3-value codes namely -1, 0, +1. The LBP mathematical procedures are illustrated as shown in equation.

$$S'\ (u, i_c, t) = \begin{cases} 1, & u\ \geq i_c + t \\ 0, & |u - i_c| < t \\ -1, & u \geq\ i_c\ - t \end{cases}$$

4. 4

4.1.4 Gradient Local Ternary pattern GLTP

Gradient local ternary pattern (GLTP) is facial texture feature based on local appearance and was introduced by Ahmed and Hossain [19]. GLTP calculates the gradient magnitudes of local neighbourhoods within the image and quantizing the values into three different discrimination levels in encoding the local texture of face expression. The resulting local patterns are used as facial feature descriptors. The main aim of GLTP is solving the problems of common features based on appearance in LBP

[9,35] and LDP [14] through combination of the benefits of the Sobel-LBP [36] and LTP [16] operators. GLTP uses more efficient gradient magnitude values as opposed to grey levels with a three-level encoding scheme to discriminate between smooth and highly textured facial regions. This ascertain the generation of consistent texture patterns even with the presence of illumination variation and random noise. Firstly, this research apply the Sobel-Feldman operator to obtain horizontal and vertical approximations of the derivatives of the source image f(x,y) [36, 37]. A Sobel operator convolves the source image with horizontal and vertical masks to obtain the horizontal (Gx) and vertical (Gy) approximations of the derivatives accordingly (see Fig.1a–c). We combine Gx and Gy to find the gradient magnitude (Gx,y) for every pixel using the formula:

$$g(x) = f(x) \odot h(x) = \int_{-\infty}^{\infty} f(s)h(x - s)ds, g(x,y) \qquad 4.5$$

4.1.4.1 Sobel operator

The Sobel-Feldman [37] operator used with GLTP is often inaccurate in computation of the gradient magnitude of an image. This is because this operator only computes an approximation of the derivatives of the image.

4.1.5 Improve Gradient Local Ternary pattern IGLTP

A common limitation of most appearance-based methods of FER, including GLTP, is that the number of features extracted from images tends to be very large. Unfortunately, most of these features are likely to form excess information because all image regions are not guaranteed to contain the same amount of discriminative data. Classification accuracy and efficiency could be reduced due to such large feature set. GLTP also utilizes the inaccurate Sobel operator for the gradient magnitude image computation when more accurate gradient operators could have been used. This section presents enhancement of the GLTP method. These include the use of a more accurate Scharr gradient operator, dimensionality reduction for the size of the feature vector size reduction.

4.1.5.1 Schnarr operator.

Schnarr [38] proposed a new operator that uses an optimized filter for convolution based on minimizing the weighted mean-squared angular error in the Fourier domain. The Scharr operator and the Sobel operator are similar in term of speed but Scharr ensures much greater accuracy when calculating the derivatives of an image.

4.1.2 Challenge in features extraction using local pattern

When designing a portable computer vision system, some challenges like illumination variation, random noise and computation effective come up. To end this, this research can extract either local features based on geometric or global features based on appearance can be extracted to address these challenges. Feature information is extracted in geometric-based method by using distance, shape and position of facial components, while appearance information such as pixel intensity of face image is used to extract information in appearance-based method [2]. The most appearance-based features extractions methods that are used include local binary pattern (LBP) [36], [38], [18],[39]. Local directional pattern (LDP) [38], [40], local ternary pattern(LTP) [41] and Gabor wavelet transform (GWT) [18], [56] . LBP was introduced by [42], where the facial image was divided into local regions and grey-level intensity value was used for texture encoding of the image which causes effective computation. When CK datasets were used the accuracies achieved were 90.1% for testing 6-classes and 83% for testing 7-classes. However, LBP performance is very poor in illumination variation, random noise and a large number of error when changes occur in the background [57]. In order to address this problem, LDP feature is prevailed in which the edge response values are computed in eight directions at the location of all pixels and a relative strength magnitude code is generated unlike LBP which use grey-level intensity.

CHAPTER 5
PROPOSED FRAMEWORK

This chapter presents an overview on face expression recognition models adopted for this research. Techniques used for expression feature extraction are also described and dataset used presented.

Table 5. 1Facial expression database

Database	Facial expression	Number of Subjects	Number of images/video	Gray /Color	Reso Fran
Extended Cohn-Kanade Dataset (CK+)[5]	neutral, sadness, surprise, happiness, fear, anger, contempt and disgust	123	593 image sequences (327 sequences having discrete emotion labels)	Mostly gray	640*

Japanese Female Facial Expressions (JAFFE)	neutral, sadness, surprise, happiness, fear, anger, and disgust	213	213 static images	Gray	256*
FERG-3D-DB (Facial Expression Research Group 3D Database) for stylized characters	angry, disgust, fear, joy, neutral, sad, surprise	4	39574 annotated examples	Color	
MMI Database		43	1280 videos and over 250 images	Color	720*

FACE EXPRESSION DATASETS

A facial expression database is a collection of images or video clips with facial expressions of a range of emotions. Most of the databases are usually based on the basic emotions theory which assumes the existence of six discrete basic emotions (anger, fear, disgust, surprise, joy, sadness). In **Error! Reference ource not found.**there are some details of the facial expression databases.

RESARCH METHDOLOGY

This study presents Hierarchical face Expression Recognition Model (FER). It applies both machine & deep learning methods. **Error! Reference source not found.**shows the structure of roposed hierarchical Face Expression.

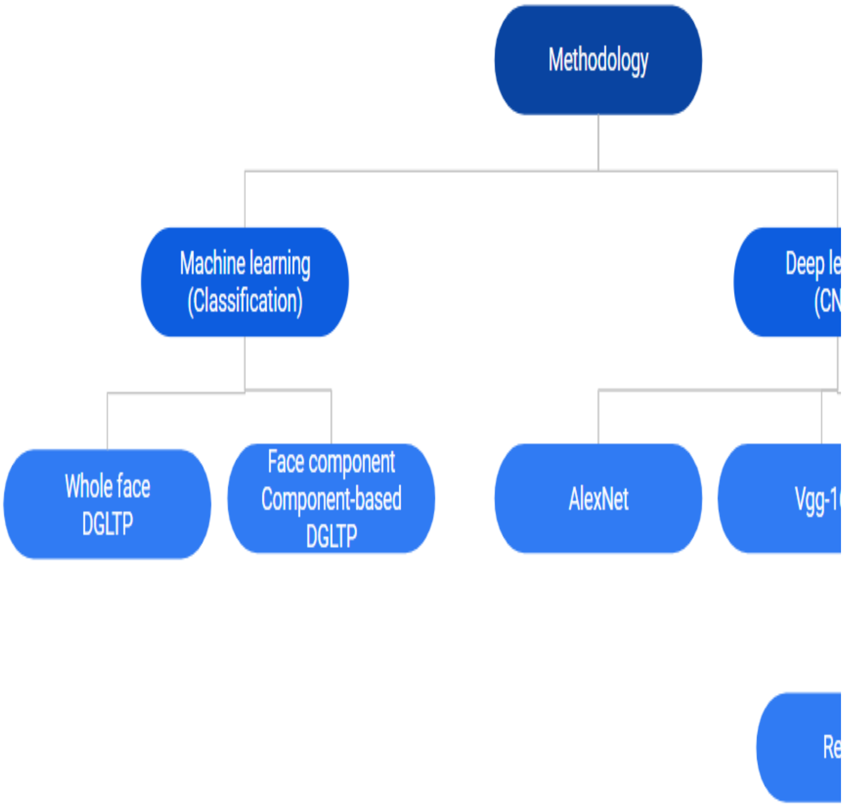


Fig 5. 1 Hierarchical Face Expression Recognition Methodology

5.2.1 Face Expression Recognition Using Directional Gradient Local Ternary Pattern (DGLTP)

This section present Directional Gradient Local Ternary Patterns model (DGLTP) and Component-Based DGLTP. The modules that make up DGLTP are image pre-processing, feature extraction; feature enhancement and features classification se

Face detection.

This study used Viola and Jones detector for Face detection. In some situations the face was detected using Haar-cascade [157]. Fig 5. 2 shows input image and detected face. Expression features are extracted from each of the grey-scale face image used for expression recognition.

5.2.1.1 Feature Extraction Using DGLTP Model

Firstly, 8-direction of the derivatives of the source image f(x,y) are obtained by applying the kirsch (kernels) for 8 orientations. The orientations of these edges are defined as East, North East, North, North West, West, south West, South, and South North approximations of the derivatives respectively.

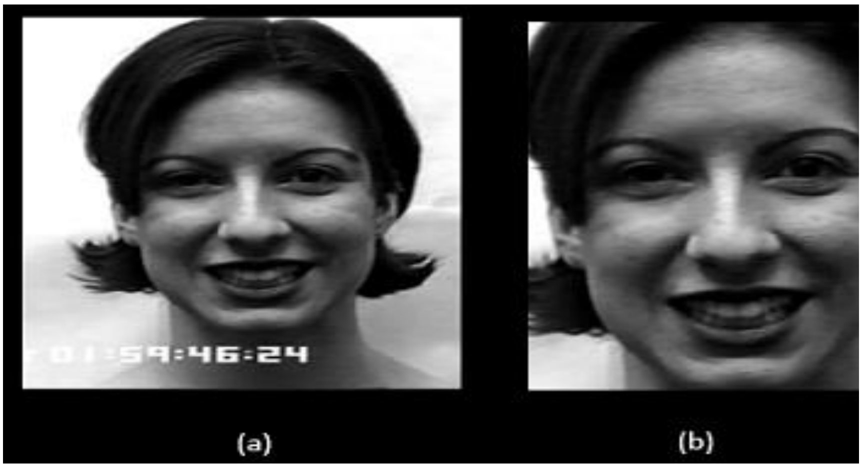


Fig 5. 2 (a) Grey-scale input image (b) Detected and cropped face

Then this research calculate the 4-main directional which is (max, min, Std, Avg) the algorithm below illustrate our methodology. After that determines the differences in pixel intensities between central pixel and its neighbouring pixels which shown in Fig 5. 3. Responses are then normalized before being used to generate DLTDP code. Min max normalization is done as Eq 5.1

$$x_i^{norm} = \frac{xi - min}{\max - min}$$

5. 1

where

x_i is the absolute value of respective responses for i = 0, 1, 2, . . . 7, min and max are minimum and maximum responses respectively and $x^{norm}{}_i$ is the normalized value of x_i . The normalized responses are in the range of 0.0 and 1.0 which signify the probability of an edge from the central reference pixel stretching towards respective direction.

Threshold τ is set to ±0.1667 deviation from 0.50 value. The value 0.5 is selected as offset reference value for τ because it shows equal chance of there being an edge or not.

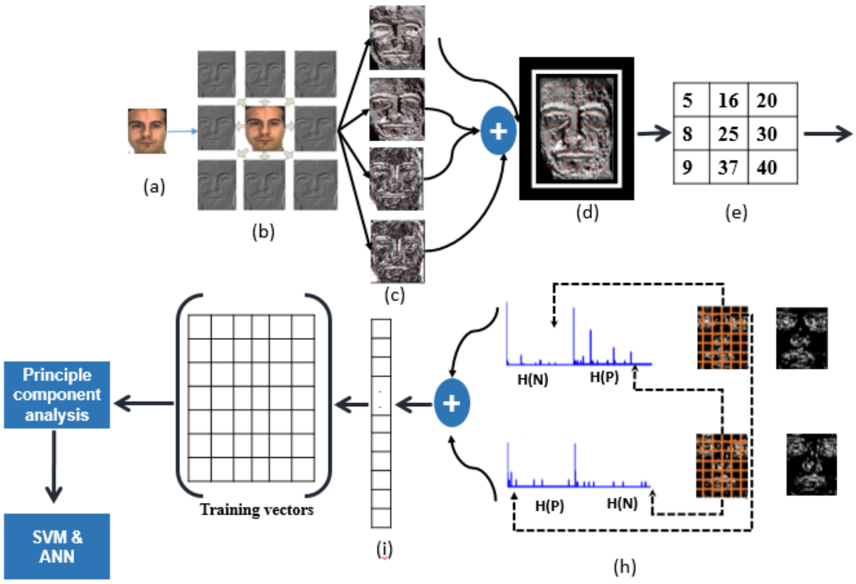


Fig 5. 3 Feature Extraction Using DGLTP Model

The value of τ is chosen to ensure the probability space is divided into 3 equal segments, one for each ternary bit. If the normalized response value is greater or equal to 0.5+ τ , its corresponding bit is set to +1, if the normalized response value is less or equal to 0.5 − τ , its corresponding bit is set to −1, and the corresponding bit is set to 0 if the normalized response is between 0.5 − τ and 0.5 + τ as in Eq 5. 2

$$f(x_i) = \begin{cases} 1 & if \quad r_i^{norm} \geq \; 0.50 \; + \; T \\ 0 \quad if \;\; 0.50 \; - \; T < & if \;\; r_i^{norm} < 0.50 + T \\ -1 & if \qquad r_i^{norm} \leq 0.50 - T \end{cases}$$

5. 3

The presence of an edge towards a particular direction is signified by not only significant differential directional response towards that direction but also significant differential directional response of one of its neighbouring direction. A differential directional response is significant if its value d is greater than ¯ m = 0.5 × m + τ where m is the maximum differential directional response of the local region. Differential directional responses closer to ¯ m are coded as being invariant relative to central pixel hence there corresponding bit set to 0. The differential directional response further away below ¯ m = 0.5 × m − τ are coded as having a negative image gradient hence there corresponding bit set to −1 and those further away above ¯ m are coded as having positive image gradient hence there corresponding bit set to 1. Each DLTDP is split into its corresponding negative and positive segments as shown in Fig 5. 4

0	0	1
1		1
1	1	0

00010011=19

0	1	0
0		0
1	1	0

(a) Positive LDP code

(b) Negative LDP code

Fig 5. 4 Resultant LDP codes from the DGLTP code

These codes are converted to decimal and assigned to corresponding central pixel of positive and negative DLTDP encoded images respectively. A histogram is generated for both negative and positive DLTDP encoded images as in Eq **Error! eference source not found.**

$$H_i = \sum_{x,y} I(f(x,y) = i), i = 0,2,\ldots,2^p - 1 \qquad 5.4$$

Where p is the number of patterns that can be encoded by the LDP operator (positive and negative) and see Eq 5.4

$$I(a) = \begin{cases} 1, & if\ a\ is\ TRUE \\ 0 & otherwise \end{cases} \qquad 5.5$$

The positive and negative histograms are concatenated and used as a feature vector for expression recognition. The dimensionality of the resultant feature vector is reduced using PCA. These two histograms are concatenated to form final DLTDP feature vector. Experimental results of face expression recognition using DLTDP are presented in Later sections. The proposed DLTDP technique is published in [158].

5.2.2 Face Expression Recognition Using Component-Based DGLTP

Fast ROI detection technique that uses Haar-Cascade classifiers to detect eyes, nose and mouth and geometric approximations to locate cheek bone area and forehead is proposed. Haar-cascade

[277] and LBP-cascade [279] are used for face detection. Once the face is detected, the technique start by detecting eyes using Viola and Jones [216] Haar-Cascade eye detection classifier. High speed and high detection rate of Viola and Jones classifier guided its choice for this study. The rectangular region around, and including the eyes is detected and cropped. Haar-Cascade classifier [280] is used to detect nose and mouth separately. The bounding rectangles around the nose and the mouth are found and ROIs related to them are detected. Due to high chances of false detections for nose and mouth, threshold is used to ensure that the mouth is always below the nose. The center point between the two detected eyes ρ is calculated. The Euclidean distance d between the centre of gravity of the two bounding rectangles around the eyes is calculated. Nose-bridge is located as a circular region centred at point p with radius 0.25 x d. Forehead is defined as rectangular region above eyes with top-left corner at point (x_1, y_1) bottom-right corner at point (x_2, y_2) Coordinate values x1 and x2 correspond to x coordinates of left eye and right eye respectively. Coordinate values and y1 and y2 are experimentally derived from geometric information of eyes. The top-left corner of the cheek bone area is found as point (x,y+h) where x and y are the x and y coordinate of the top-left corner of respective eye, h is the height of the eye. Bottom-right corner of cheek bone area is found as point (x +w , y + 1.5 x h) where w is the width of the respective bounding rectangle around the eye. Shifting factor of y-coordinate is determined experimentally and is found to be closer to 1 for top-left corner and varies between 1.5 to 2.0 for bottom right corner. As shown in Algorithm 1, the contribution of this work is the localization of left and right cheek bone areas, forehead and nose-bridge using geometric and spatial information of fiducial landmarks. Figure 3.4 the facial landmarks localized and how they are

annotated in this study. For each landmark, a rectangular bounding box shown in Fig 5. 5is returned.

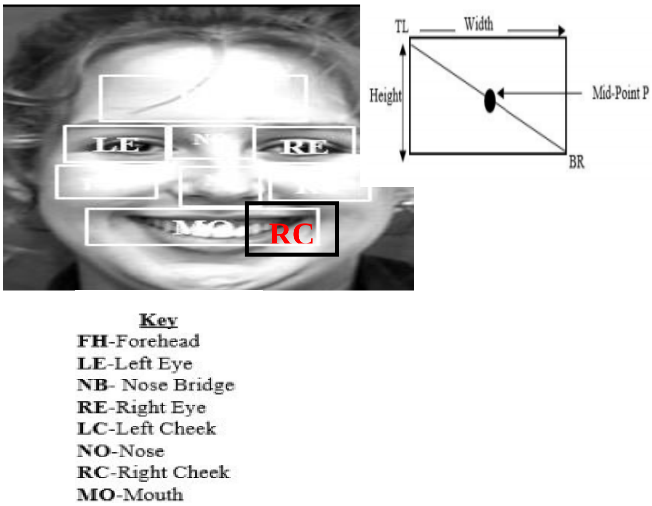


Fig 5. 5Annotation of landmarks to be located

ɪp-left corner (TL) and bottom-right corner (BR) of fiducial ɪdmark rectangles are tracked because they are used to ɔmetrically approximate TL and BR of secondary landmarks (LC, : and FH). Top-left corner of RE is considered as RE.TL and its ɔrdinates are RE.T L_x and RE.T L_y respectively.

ft and right eyes are detected using Haar-cascade classifier [216] d geometric and spatial information of LE is used to detect LC d that of RE to detect RC. Distance d between LE and RE is lculated. The mid-point p between LE and RE is gotten and used locate Nose Bridge. Nose Bridge is located as rectangular region ıt inscribe a circle of radius 0.25 x d centred at p. The centre of : denoted as LE.C and centre of RE denoted as RE.C are derived. rehead is located as a rectangular region above eye-brows and htly bounded between eye centres. The y-coordinates of the :ehead top left corner F H.T L and bottom right F H.BR are ıpirically derived from the y-coordinates of left eye centre and ,ht eye centre respectively. The y coordinate of top left corner of ·ehead is found to vary between $LE.C_y$ −(1.40 × h) and $LE.C_y$ 1.50 × h) where h is height of LE.F $H.BR_y$ is found to vary tween $RE.C_y$ − (0.50 × h) and $RE.C_y$ − (0.60 × h) where h is ight of RE.

ter an image has been pre-processed, facial part of the image is tomatically detected and cropped. There are a number of face tection algorithms that can be used such as Viola and Jones [35] ıar-cascade classifiers[157]. After detection of face, facial ıtures can be extracted using holistic or local feature extraction :hniques. Facial landmarks like eyes, nose, mouth, cheeks, nose- .dge, forehead, eye-brows could be located in the detected face. ›listic and local feature extraction techniques are applied on the ıole face or the detected landmarks for expression features traction.

5.2.2.1 Feature extraction for **Component-Based DGLTB**

opping was performed on the eyes, nose, nose-bridge, cheek and ›uth regions. However, in [146] cropping was only performed on : eye, nose and mouth in addition to [18] cropping was only rformed on the eye and mouth regions with the nose being cluded. Algorithm below illustrate how we use face component to me up with great performance.

5.2.2.1.1 DGLTP Algorithm

Source image i.e. pre-processing cropped face Vector of DGLTP

1. Crop 8-component using landmark localization

approach.

2. Compute the 8-direction derivative approximation of image using Kirich operator.
3. Take the MAX, MIN, AVG and STD response of each of two responses for each orientation (horizontal, vertical, left diagonal, right-diagonal).
4. Apply thredshold ±t o f around centre gradient t value G_c in 3×3 neighbourhood to determine $S_{DG\ LT\ P}$ code for each component.
5. For each component compute positive P-DG LT P and negative N-DG LT P code image representation from $S_{DG\ LT\ P}$ values.
6. Split code into m × n regions.
7. Compute positive $H_{P\text{-}DG\ LT\ P}$ and negative $H_{N\text{-}DG\ LT\ P}$ DGLTP histogram for each region.
8. Concatenate positive HPG LT P and negative HNG LT P GLTP histogram for each region to form feature vector.

5.2.3 Face Expression Using Deep Leaning

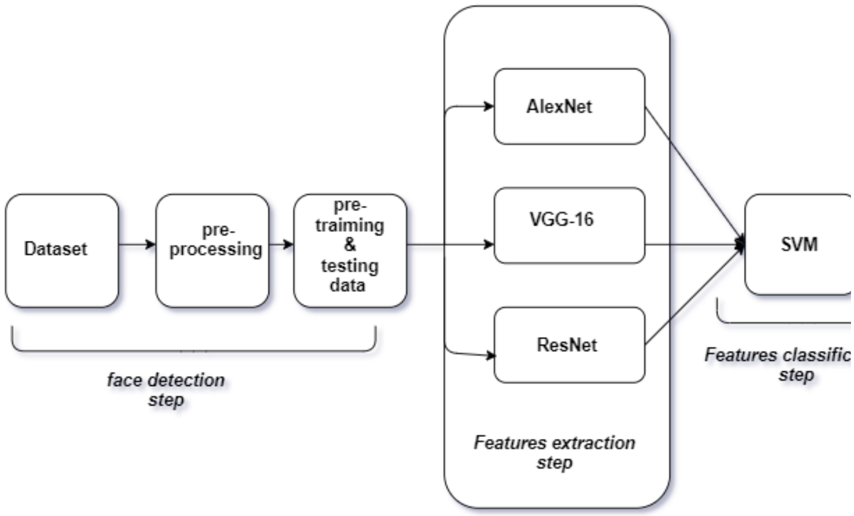
approach combines standard methods such as Viola Jones method for face detection/extraction [159], [160], convolutional neural network for feature extraction and Support vector machine for classification.

THE proposed Face expression recognition method used is based on a four model

Fig 5. 6 The methodology for CNN methods architecture that is able to recognize facial expressions Fig 5. 6shows the process diagram of the proposed facial expression recognition system which is divided into three stages- Image Pre-Processing which involves Face detection which is divided into three stages-

Image Pre-Processing which involves Face detection using viola\Jones algorithm, feature extraction and feature classification steps.

The Face detected cropped and extracted and then used as the input into the inception layer of the CNN models. The Training phase involves the feature extraction and classification using SVM classifier. It is expected that using these models will reduce training time and increase the probability of high level feature extraction thereby increasing system accuracy.



5.2.3.1 Image pre-processing

Pre-processing is the first step of the proposed system and it uses Viola-Jones face detection framework for face detection. The face was first detected, cropped and face components were extracted and normalized 224 x 224 x 3 pixel size for Res-Net50 and Res-Net101 CNN Models, also 128 x128 pixels for Alex-Net CNN model. When image scale is reduced helps to reduce the information that has to be learned by the network is also reduced and ensure faster training and memory cost is less [37]. The detected face component were applied as the input into the inception layer of the CNN to make them compatible with the input size of CNN Pre-Trained Models and convert any grayscale images to RGB images using an augmented Image Data store to resize image and convert image to RG.

5.2.3.2 Prepare Training and Test Image Sets

Separate the sets into two, the separation should be randomized to avoid being bias. Pick 85% of the selected images from each set for the training data and the remaining 15%, for the test data. It takes one day for training and 4 hour for testing.

5.2.3.3 CNN Pre-Trained Models

This research applied four models CNN pre-trained convolution neural networks as the proposed facial expression recognition methods, which are Alex-Net, VGG, ResNet-101 and ResNet-5 and compared the four methods with state of the art methods. A pre-trained model is a kind of model which has been trained on a large standard dataset to address a problem similar to the one we aim to address. The process of proposed facial expression recognition system divided into three stages; Image Pre-Processing where we used Viola Jones algorithm for Face and facial parts detection, facial Feature extraction and feature classification using CNN and SVM.

5.2.3.3.1 Alex-Net Model

It was winning architectures for ImageNet competition. It uses large filter and stride size at the first layer. It is uses 8-layer network, with an image input size of 48 by 48.It takes one day for training and 4 hours for testing.

5.2.3.3.2 VGG-19 Model

Our VGG-19 has an input of 48x48 RGB image. The image is passed through a bunch of convolution layers, where 3x3 filters are used. It contains 19 weight layers which involves 16 convolutional layers with 3x3 filter size and 3 fully connected layers and followed by stack of convolutional layers. Each first two of 3 fully connected layers has 4096 channels, the third performs 7-way ILSVRC classification and therefore contains seven channels (one for each class). SVM layer is the final layer. It takes one 22 hours for training and 5 hours for testing.

5.2.3.3.4 Res-101 Model

It is a convolutional neural network that is trained on more than a million images from the ImageNet database. The network contains 347layers and has an ability to classify images into 1000 object categories. The network has an image input size of 224 by 224. It takes two day for training and 7 hour for testing.

5.2.3.3.5 Res-50 Model

It is a convolutional neural network that is trained on over a million images from the ImageNet database. The network is consisting of 177 layers and can classify images into 1000 object categories. The network has an image input size of 224-by-224. It takes 18 hours for training and 5 hour for testing.

5.2.3.4 Train A Multiclass SVM Classifier Using pre-trained CNN

Features extract the learned image features from a pre-trained CNN for feature extraction, the precedent layer to classification layer called fc1000 was used for feature extraction through the activation method. These features were then used to train and test SVM classifier.

5.2.3.5 Evaluation protocol

The performance of a method, in facial expression images analysis, is typically measured by specificity, sensitivity, and F1 score Sensitivity measures the proportion of actual positives samples that correctly identify, in which the percentage of expression image that is correctly classified as the same expression.

5.3 Summary

This chapter presented hierarchical face expression recognition models adopted together with feature extraction techniques and experimental dataset. Expression recognition process consists on whole face expression recognition using DGLTP, Component-Based DGLTP and face expression recognition using deep learning models. Feature extraction techniques used for face expression are describe the chapter concludes by presenting experimental dataset used for expression recognition.

6 CHAPTER SIX (RESULT AND DISSCUSSION)

6.1 Introduction

This chapter present and discuss results of the proposed methods for face expression recognition using enhanced facial and facial component features. In this recognition approach, the first step is localization of facial landmarks like eyes, mouth, nose, cheeks, nose-bridge and forehead. Thereafter, facial expression features are extracted from both holistic face and the localized landmarks. These features are fused and used for face expression recognition. Mouth, nose, nose-bridge, eyes and forehead are cropped from detected images and used to model expression recognition that show how fiducial landmarks drift from each other across age. Expression features from the face and facial parts are used for face expression recognition modelling. In addition to use four models in deepening for recognizing face expression.

6.2 Landmark localization

This section presents results for the experiments described in the previous sections. To ensure accurate results, the images were pre-processed before feature extraction. Two forms of pre-processing were implemented: cropping the face from the image and cropping multiple facial components from the image. Face, fiducial and secondary landmarks are automatically detected. Shows a sample output of a face image that was loaded, face region detected and regions of interest (landmarks) cropped. To remove the background and other edge-lying obscurities, the subject’s face was cropped from the original image based on the positions of the eyes. For the CK+ & JEFFE datasets, 68 landmark locations were provided for each image, each of which represents a point on the face as shown in Fig 6. 1.

Fig 6. 1 Sixty-eight facial landmarks [38]

There were 1892 faces detected out of 2004 images in the experimental dataset, which correspond to 94.41% accuracy. This detection rate could be attributed to adverse image conditions of some images in the dataset. Landmark detection was performed in these 1094 detected faces and results are shown in Table 6. 2

Table 6. 2Regions of interest detection accuracy

ROI	Detection	Total	Accuracy
Face	1892	2004	94.41%
FH	1690	1892	89.32%
RE	1732	1892	91.54%
LE	1704	1892	90.06%
RC	1732	1892	91.54%
LC	1704	1892	90.06%
NO	1724	1892	91.12%

NB	1698	1892	89.75%
MO	1714	1892	90.60%

there was a 94.41% accuracy on the face detection since 1094 aces were detected from the 2004 face images. The results show that 3436 eyes were detected from the possible 3784 of which 1732 are right eyes and 1704 are left eyes. The lower rate of eye detection is attributed to poor image condition, pose and face orientation. The same results as for the eye were observed for the cheeks since spatial information and dimensions of the eye were used to detect the cheeks. Therefore, for every true eye there was a true cheek and consequentially for every false eye there was a false cheek. Using geometric information of eyes, 1698 nose-bridge regions were detected.

The results obtained are promising as 1742 (91.12%) of the 1892 possible noses were accurately detected and 1714 (90.60%) of the possible 1892 mouths were detected. Performance of mouth classifier is adversely affected by beards, image orientation and pose among other image conditions. However, it was observed that mouth classifier is robust to slight expression around the mouth region. There were 1690 (89.32%) correctly detected foreheads out of the possible 1892 foreheads.

6.3 Face expression recognition using DGLTP model

This section presents results for experiment described in previous sections.

6.3.1 Results on the JEFFE dataset

In this section, results are reported on CK & JEFFE datasets using the method mention in section 4 with a SVM classifier. The results are summarized in Tables Table 6. 3shows a significant improvement in recognition accuracy over traditional IGLTP for the JAFFE dataset. The largest improvement in recognition accuracy was seen during 7-class cross-validation testing, where Directional IGLTP outperformed IGLTP by 0.4% (see Fig 6. 2 & Fig 6. 3).

4;

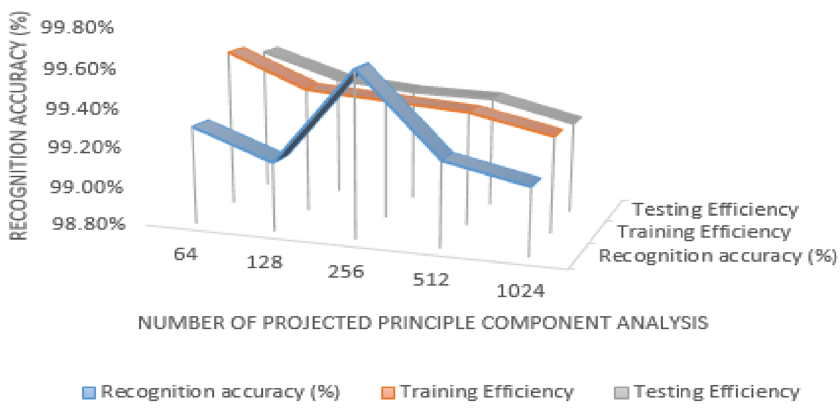


Fig 6. 2 The effect of extracting different faces expression on recognition accuracy.

Fig 6. 3 The effect of cropping faces using Viola and Jones.

The confusion matrices for cross-validation on the JAFFE dataset provided in Table 6. 4 & Table 6. 5.This research observe that, lastly, the compression had been made between proposed methods and other methods mention in the literature review. For measuring the accuracy this research used cross-validation testing procedure.

Table 6. 4 Confusion matrix (%) of cross-validation testing method for 6-classes emotional in JEFFE dataset using DGLTP

	Joy	Sur	Ang	Fear	Dis	Sad	N
Joy	98.4	0	0	0.5	0	0	
Sur	0	97.2	0	1.5	0.1	0	
Ang	0	0.3	96.4	1.2	2.1	0	
Fear	0	2.3	0.1	92.9	0	1.4	
Dis	0	0	0.5	0.2	98.8	0.5	
Sad	0	0	0.1	4.3	4	90.5	
Neu	0	0.3	0.4	0	0	0.2	9

Table 6. 5Confusion matrix (%) of cross-validation testing method for 6-classe emotional in JEFFE dataset using DGLTP

	Joy	Sur	Ang	Fear	Dis	Sad
Joy	98.4	0	0	0.5	0	0
Sur	0	97.2	0	1.5	0.1	0
Ang	0	0.3	96.4	1.2	2.1	0
Fear	0	2.3	0.1	92.9	0	1.4
Dis	0	0	0.5	0.2	98.8	0.5
Sad	0	0	0.1	4.3	4	90.5
Neu	0	0.3	0.4	0	0	0.2

In Table 2 and Table 3 the study illustrates the recognition rate of GLTP, IGLTP and our proposed method, for six and seven class expression in JEFFE dataset. Our observation is the proposed method outperform better than traditional GLTB and IGLTP. The results show that when this research combine the proposed enhancements in IGLTP such as the use of the accurate kirsch mask operator, PCA for dimentinolaty reduction and four max value direction, further increase the recognition rate of IGLTP. This research use cross-validation test from above on the JAFFE dataset. This research see that our proposed method, Directional IGLTP, shows a significant improvement in recognition accuracy over traditional IGLTP for the JAFFE dataset. The 7-clas cross-validation testing has been achieved the largest improvement in recognition accuracy, where Improved DGLTP outperformed IGLTP by 0.4%. These result are shown in Table 6. 6

.

Table 6. 6Confusion matrix (%) of cross-validation testing method for 7-classes emotional in JEFFE dataset using DGLTP

Methods	**6-classes**	**7-classes**
LBP [39]	90.1%	88.3%
LDP [18]	93.7%	88.4%
LTP [146]	93.6%	88.9%
GLTP [146]	97.2%	91.7%
IGLTP [146]	99.3%	97.6%
Our DGLTP	**98.8%**	**99.7%**

6.3.2 Compression to result in literature

In Table 6. 6 it Shows comparison between common appearance-based methods of FER and other state-of-the art methods in current literature on the CK dataset. To be more accurate and fair in our comparison, the results had been taken for LBP, LDP, LTP and GLTP from [36]

6.4 Face expression recognition using Component-DGLTP model

This section presented and discussed the results of proposed face expression recognition by using fused holistic and features of facial component.

6.4.1 Experimental setup

This research used our model for recognition of facial expression and this research studied the performance accuracy on public dataset (see Fig 6. 4), which widely used for such task. The experiment carried out a 10-fold cross validation where each dataset is categorized into 10 clusters randomly. 9 clusters were used as a classifier training dataset, while the only remaining cluster was used as testing dataset. This research used common supervised learning classifier namely SVM with default parameters and neural network. Further, this research used gradient decent to train our model.

Fig 6. 4 Sample prototypic expressions from (a) CK+ dataset and (b) JAFFE dataset

6.4.2 Results on the datasets

Table 6. 5 shows a comparison between using the whole face,-facial parts (eyes, mouse and nose) and also the 8- facial parts (left eye, right eye, nose-bridge, left cheek, right cheek, nose and mouth). The results show that 8-facial components get higher recognition accuracies for feature extraction. In addition to illustrate Max, Min, Avg and Std for our proposed method and this research found that high accuracy when this research use Max. Table 6. 7,Table 6. 8,Table 6. 9 &

Table 6. 10 present the convolution matrix for them.

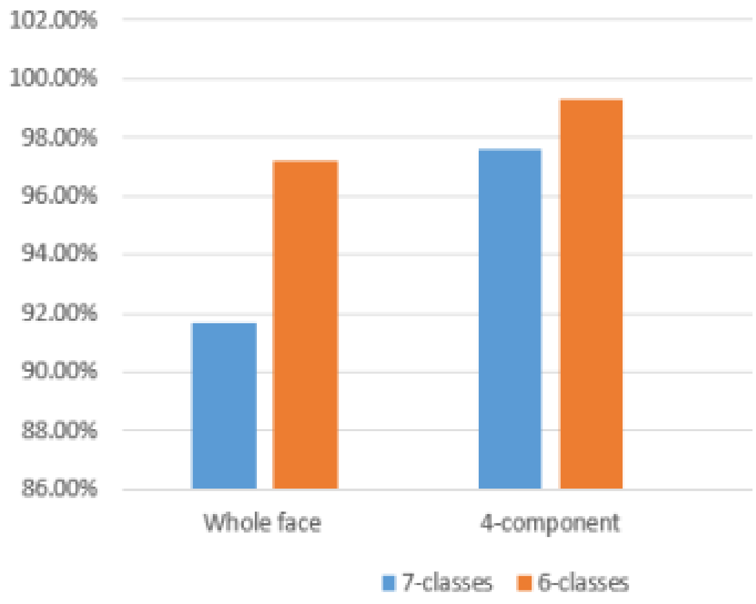


Table 6. 5The effect of extracting different sets of components on recognition accuracy in datasets

It is a quite challenging task to compare face expression recognition differences in various approaches of face expression. This research carry out performance comparison in expression recognition in this paper with similar expression in previous approaches. LBP results is the least one, where LTP and LDP results were 0.4% higher. GLTP results is 2.2% to 3.5% higher than all of these methods. As for our method, this research found it the best of them all and its results was 99.7 % and 98.9% for 6 classes and 7 classes respectively. All this improvement is due to the fact that this research get rid of all the redundancy in the image and took the most significant important parts that can help in recognizing the expression.

Table 6. 7 Confusion matrix (%) of CV method for 7-classes face expression in CK+ Dataset using Avg-DGLTP

	Joy	**Sur**	**Ang**	**Fear**	**Dis**
Joy	98.4	0	0	0.5	0
Sur	0	97.2	0.7	1.5	0.1
Ang	0.1	0.2	96.5	1.2	2.7
Fear	2	0.3	0.1	92.9	2
Dis	0	1.2	0.5	0.2	98.8
Sad	0	0	0.1	4.3	4
Neu	0	0.3	0.4	0	0

Table 6. 8Confusion matrix (%) of CV method for 7-classes

face expression in CK+ Dataset using Mx-DGLTP

	Joy	**Sur**	**Ang**	**Fear**	**Dis**	
Joy	99.4	0	0	0.5	0	0
Sur	0	94.2	0.7	1.5	0.1	0
Ang	0	0.3	95.4	1.2	2.7	0
Fear	0	2.3	0.1	94.9	0	1
Dis	0	1.2	0.5	0.2	91.8	0
Sad	0	0	0.1	4.3	4	8
Neu	0	0.3	0.4	0	0	0

Table 6. 9 Confusion matrix (%) of CV method for 7-classes face expression in CK+Dataset using Min-DGLTP

	Joy	**Sur**	**Ang**	**Fear**	**Dis**
Joy	97.4	0.5	0.1	0.5	0
Sur	0	96.2	0.7	1.5	0.1
Ang	0.1	0.2	96.4	1.2	2.7
Fear	2	0.3	0.1	92.5	2
Dis	0	1.2	0.5	0.2	97.8
Sad	0	0	0.1	4.3	4
Neu	0	0.3	0	0	0

Table 6. 10 Confusion matrix (%) of CV method for 7-classes face expression in CK+Dataset using Avg-DGLTP

	Joy	Sur	Ang	Fear	Dis
Joy	98.4	0.5	0.1	0.5	0
Sur	0	97.2	0	1.5	0.1
Ang	0.1	0.2	96.4	1.2	2.7
Fear	2	0.3	0.1	92.9	2
Dis	0	1.2	0.5	0.2	98.8
Sad	0	0	0.1	4.3	4
Neu	0	0.3	0	0	0

6.4.3 Comparison to results in literature

As show in Table 6. 12 this research observed that overall face expression recognition has improved from 90.1%, 93.7%, 93.6% 97.2%, and 99.3% reported in [40] respectively to 99.7%. This significantly proves that the performance of expression recognition is improved when both global and local features with features fusion and decision fusion are used.

Table 6. 11 Comparison with methods in literature

Methods	6-classes	7-classes
LBP [39]	90.1%	88.3%
LDP [18]	93.7%	88.4%

LTP [146]	93.6%	88.9%
GLTP [146]	97.2%	91.7%
IGLTP [146]	99.3%	97.6%
Our DGLTP	**98.8%**	**99.7%**

6.5 face expression recognition using deep learning

This section presents results for experiment described in section 5.3. The proposed system runs on Intel Core i5- CPU @ 2.7 GHz with 8GB. MATLAB 2019a tool was used to evaluate the method and perform the feature selection and classification task. The network is trained for classification using the training subset of 85% while the testing subset of 15% is used to test the probability that a facial image belongs to a particular facial expression class.

6.5.1 Training of the CNN is achieved in a supervised manner

Using the standard across-validation algorithm. The average recognition accuracy is used to evaluate the performance of the network. Image pre-processing took an average of 0.1s and image classification took an average of 0.2s. The recognition accuracy achieved when using Alex-net, VGG-Net, Res-Net101, Res-Net50 methods is 88.2%, 84.2%, 81.6%, 80.9% respectively. This research compared the average recognition accuracy of the proposed methods with other methods for facial expression recognition. The other methods on the CK+ database are shown in Table 6. 12 it was proven that the proposed methods achieved a higher recognition accuracy comparing to other existing methods in literature.

5.6 Summary

This research proposed an improved technique, Directional Gradient Local Ternary Pattern (DGLTP) for accurate Face Expression recognition. The results achieved show that the whole face is not necessarily required for face expression recognition from facial images. The landmark model is used to extract the component of the faces (forehead, eyes, nose, cheeks, mouth and chin). The experimental results conducted on the whole face expression dataset, showed an accuracy rate of about 98.90%. The proposed DGLTP outperformed other related state-of-the-art feature extraction techniques, for face expression accuracy.

Fig 6. 6 Confusion matrix (%) of cross-validation testing method for 7-classes emotional in CK dataset using Alex-Net

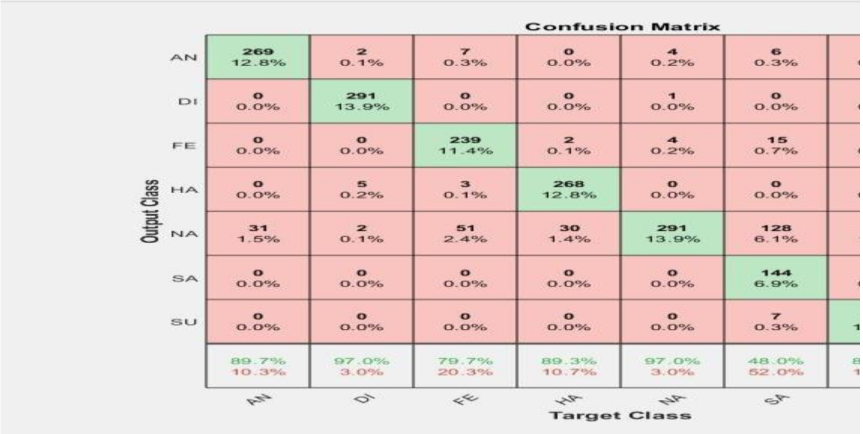


Fig 6. 7 Confusion matrix (%) of cross-validation testing method for 7-classes emotional in CK dataset using Vgg-16

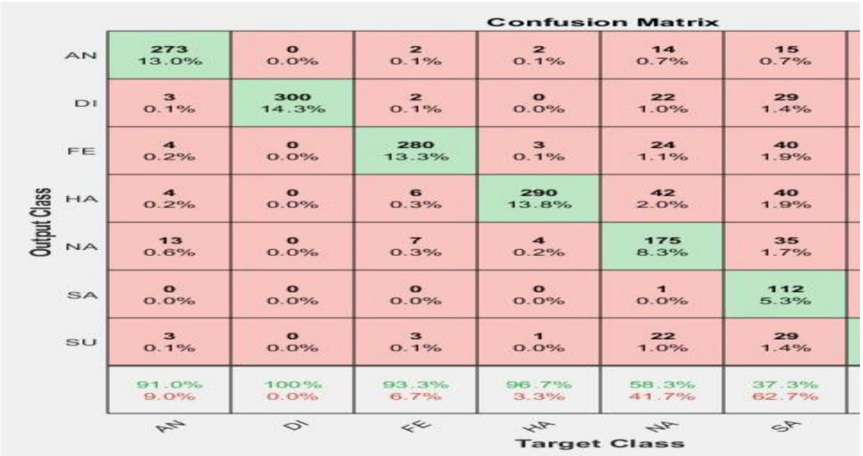


Fig 6. 8 Confusion matrix (%) of cross-validation testing method for 7-classes emotional in CK dataset using ResNet 50

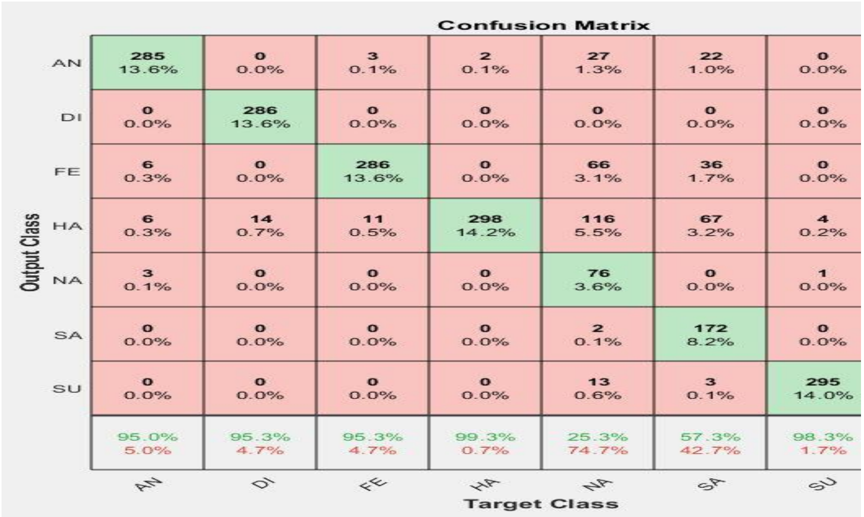


Fig 6. 9 Confusion matrix (%) of cross-validation testing method for 7-classes emotional in CK dataset using Res-Net 101

Table 6. 12 CNN methods compared with previous methods

Methods	Accuracy
AlexNet	61.7%
Vgg-16	76.7%
ResNet	63.1%
Our Alex Net	**88.2%**
Our Vgg-16	**84.2%**
Our ResNet	

	81.6%

7 CHAPTER SEVEN
(CONCOLUSION AND FUTURE WORK)

This chapter summarises the findings from the research reported in this thesis with respect to the objectives listed in Chapter 1. It then outlines its contributions to the expression recognition research domain. Finally, it offers directions for future work inspired by the results of the experiment performed so far.

7.1 Conclusion

This study proposes two improvements of GLTP operator and deep learning methods, Feature fusion strategy to improve discriminative power of extracted features is also proposed. A landmark localization technique and investigation of landmark displacement across facial expressions are also presented in this study.

A two stage fast facial landmark localization approach was developed. In the first step, fiducially landmarks like eyes, nose and mouth are detected using different Haar-cascade classifier. In the second phase, secondary landmarks like cheek-bone area, forehead and nose-bridge are located using geometric and spatial information of fiducially landmarks. Not incorporating machine learning makes the proposed technique fast without compromising accuracy. The proposed technique is appropriate for real time landmark localization for subsequent image processing tasks.

A Directional Gradient Local Ternary Pattern (DGLTP) calculates the difference in neighbouring pixel values in a 3 × 3 image region with the central pixel. These differences in pixel values are convolved with Kirsch masks to obtain eight directional responses. The normalized responses are used as probability of an edge occurring in a particular direction. The probability space is split into three subspaces and used to generate DGLTP codes. Experimental results on face expression recognition show that DGLTP performs better compared to LDP, LTP and LBP operators.

Lastly, the study proposed a methodology for the purpose of pre-training convolution neural network (CNN) architectures, which were used for classification face expression recognition from CK+ facial images. We applied the pre-trained CNN models to extract features (Alex-Net, Vgg-16, Res-Net 50, Res-Net 101) and support vector machine SVM was applied for expression recognition from facial images. The performance of pre-trained convolutional neural network applied to recognize the expression was evaluated based on accuracy from Confusion Matrix. The results achieved have demonstrated the high rate recognition of the proposed pre-trained network, it gave very accurate rate of about 88.6% as a result in face expression recognition, it showed that combination of the pre trained CNN models with SVM achieved high accurate rates.

7.2 Future work

As we see the oriented of the word, it depends on technologies, object detection and image identification. It is a big area for work here we work on facial images mainly in Face expression recognition. This study start from recognizing the expression from whole face using machine learning techniques also segmented the face to the component and get result then adding parameter to the model to improve the accuracy lastly using deep learning to enhance the rate of the accuracies. This topic is one of the hot areas topic so need addition studies because it is the renewing area. This field one of the important fields in computer science, since most application are now based on human feeling interested, so we are using it with facial component for future work based on deep learning methods to be more accurate and improve the general viewing in this area. Also this research recommended to used Sudanese dataset also we recommend finding mechanism to enhance it to be standard and known in field of image processing.

References:

[1] D. Berry, L. Zebrowitz-MeArthur, and T. Alley, “Social and applied aspects of perceiving faces,” 1988.

[2] H. T. Reis, W. A. Collins, E. Berscheid, and ; W Andrew Collins, “Psychological Bulletin The Relationship Context of Human Behavior and Development,” *psycnet.apa.org*, vol. 126, no. 6, pp. 844–872, 2000, doi: 10.1037/0033-2909.126.6.844.

[3] M. Knapp, J. Hall, and T. Horgan, *Nonverbal communication in human interaction*. 2013.

[4] M. Parrish, *Social work perspectives on human behaviour*. 2014.

[5] C. Zhang and Z. Zhang, “A Survey of Recent Advances in Face Detection,” 2010. Accessed: May 28, 2020. [Online]. Available: http://www.research.microsoft.com.

[6] Y. Tayal, P. Kumar Pandey, D. B. V Singh, and M. Gwalior, “International Association of Scientific Innovation and Research (IASIR) International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS) www.iasir.net Face Recognition using Eigenface,” *Citeseer*, 2013, Accessed: May 28, 2020. [Online]. Available: www.iasir.net.

[7] C. Shan, S. Gong, P. M.-I. and vision Computing, and undefined 2009, “Facial expression recognition based on

local binary patterns: A comprehensive study," *Elsevier*, Accessed: May 28, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0262885608001844.

[8] S. K. Singh, D. S. Chauhan, M. Vatsa, and R. Singh, "A Robust Skin Color Based Face Detection Algorithm," 2003. Accessed: May 28, 2020. [Online]. Available: http://iab-rubric.org/papers/tkjse.pdf.

[9] A. Punitha and M. Kalaiselvi Geetha, "Texture based Emotion Recognition from Facial Expressions using Support Vector Machine," 2013. Accessed: May 28, 2020. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.403.3906&rep=rep1&type=pdf.

[10] K. C. Yow and R. Cipolla, "Feature-Based Human Face Detection," 1996. Accessed: May 28, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0262885697000036.

[11] C. Ravat, S. S.-P. of the National, and undefined 2018, "Survey on different methods to improve accuracy of the facial expression recognition using artificial neural networks," *pdfs.semanticscholar.org*, Accessed: May 28, 2020. [Online]. Available: https://pdfs.semanticscholar.org/0a2a/b8add82f3546197d49c4213fbaf025faa2a6.pdf.

[12] I. Kotsia and I. Pitas, "Facial Expression Recognition in Image Sequences Using Geometric Deformation Features

and Support Vector Machines," *IEEE Trans. IMAGE Process.*, vol. 16, no. 1, 2007, doi: 10.1109/TIP.2006.884954.

[13] C.-W. Hsu and C.-J. Lin, "A Comparison of Methods for Multiclass Support Vector Machines," 2002. Accessed: May 27, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/991427/.

[14] C. H. Q. Ding and I. Dubchak, "Multi-class protein fold recognition using support vector machines and neural networks," 2001. Accessed: May 28, 2020. [Online]. Available: http://www.nersc.gov/.

[15] K. Naman, S. Zhenye, Z. B. Schölkopf, A. Smola, and K.-R. Müller, "NONLINEAR COMPONENT ANALYSIS AS A KERNEL EIGENVALUE PROBLEM."

[16] L. Chen, M. Zhou, W. Su, M. Wu, J. She, and K. Hirota, "Softmax regression based deep sparse autoencoder network for facial emotion recognition in human-robot interaction," *Inf. Sci. (Ny).*, vol. 428, pp. 49–61, 2018, doi: 10.1016/j.ins.2017.10.044.

[17] T. Kalsum, S. Anwar, M. Majid, S. M. Anwar, B. Khan, and S. M. Ali, "Emotion Recognition from Facial Expressions using Hybrid Feature Descriptors Stress Measurement View project Health Monitoring using wearable sensors View project Emotion recognition from facial expressions using hybrid feature descriptors," *IET*, vol. 12, no. 6, pp. 1004–1012, Jun. 2018, doi: 10.1049/iet-ipr.2017.0499.

[18] S. Tripathi and P. Suja, “Biographical notes: P. Suja obtained her BE in Computer Science and Engineering in 1998 from Bharathiyar University,” *Int. J. Adv. Intell. Paradig.*, vol. 7, no. 1, pp. 57–73, 2015, doi: 10.1504/IJAIP.2015.070349.

[19] S. Tripathi, S. Mariam Thomas, and V. K. Madan, “Emotion Recognition from Images Under Varying Illumination Conditions Fuzzy Logic Machine learning based controller View project Robotic Grasper based on an End-to-End Neural Architecture using Raspberry Pi View project Suja Palaniswamy Amrita Vishwa Vidy,” *Springer*, vol. 357, pp. 913–921, 2016, doi: 10.1007/978-3-319-18416-6_72.

[20] E. Sariyanidi, … H. G.-I. transactions on, and undefined 2014, “Automatic analysis of facial affect: A survey of registration, representation, and recognition,” *ieeexplore.ieee.org*, Accessed: May 28, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6940284/.

[21] W. Chao, J. Ding, J. L.-S. Processing, and undefined 2015, “Facial expression recognition based on improved local binary pattern and class-regularized locality preserving projection,” *Elsevier*, Accessed: May 28, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165168415001425.

[22] C. J. Taylor, G. J. Edwards, T. F. Cootes, and C. J. Taylor, “Face Recognition Using Active Appearance

Models. Image Analysis software architecture View project Nailfold View project Face Recognition Using Active Appearance Models," *Springer*, vol. 1407, pp. 581–595, 1998, doi: 10.1007/BFb0054766.

[23] M. Bartlett, G. Littlewort, … I. F.-2003 C. on, and undefined 2003, "Real time face detection and facial expression recognition: development and applications to human computer interaction.," *ieeexplore.ieee.org*, Accessed: May 27, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/4624313/.

[24] G. Muhammad, M. Alsulaiman, … S. A.-I., and undefined 2017, "A facial-expression monitoring system for improved healthcare in smart cities," *ieeexplore.ieee.org*, Accessed: May 27, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7941992/.

[25] H. Hosseini, Z. K.-T. 26th A. International, and undefined 2004, "Facial expression analysis for estimating patient's emotional states in RPMS," *ieeexplore.ieee.org*, Accessed: May 27, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/1403465/.

[26] G. S. Shergill, M. A. Sarrafzadeh, and A. Shekar, "COMPUTERIZED SALES ASSISTANTS: THE APPLICATION OF COMPUTER TECHNOLOGY TO MEASURE CONSUMER INTEREST-A CONCEPTUAL FRAMEWORK," 2008. Accessed: May 27, 2020. [Online]. Available: http://156.62.60.45/handle/10292/1715.

[27] 720 JC Chang - US Patent App. 10/429 and undefined 2004, “Facial surveillance system and method,” *Google Patents*, Accessed: May 27, 2020. [Online]. Available: https://patents.google.com/patent/US20040223629A1/en.

[28] R. Nagpal, P. Nagpal, and S. Malhotra, “BIOMETRIC TECHNIQUES AND FACIAL EXPRESSION RECOGNITION-A REVIEW,” 2012. Accessed: May 27, 2020. [Online]. Available: www.jgrcs.info.

[29] J. Thies, M. Zollhöfer, M. Nießner, L. Valgaerts, M. Stamminger, and C. Theobalt, “Real-time expression transfer for facial reenactment,” *ACM Trans. Graph.*, vol. 34, no. 6, 2015, doi: 10.1145/2816795.2818056.

[30] H. Gunes, M. P.-18th I. C. on Pattern, and undefined 2006, “A bimodal face and body gesture database for automatic analysis of human nonverbal affective behavior,” *ieeexplore.ieee.org*, Accessed: May 27, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/1699093/.

[31] H. Kobayashi, A. Tange, and F. Hara, “Real-time Recognition of 6 Basic Facial Expressions,” 1996. Accessed: May 27, 2020. [Online]. Available: https://www.jstage.jst.go.jp/article/jrsj1983/14/7/14_7_994/_article/-char/ja/.

[32] A. Pentland, B. Moghaddam, and T. Starner, “View-Based and Modular Eigenspaces for Face Recognition,” 1994. Accessed: May 27, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/323814/.

[33] H. Rowley, S. Baluja, T. K.-I. T. on pattern, and undefined 1998, “Neural network-based face detection,” *ieeexplore.ieee.org*, Accessed: May 27, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/655647/.

[34] T. Sim, S. Baker, … M. B.-T., and undefined 2003, “SHORT PAPERS-Face Database-The CMU Pose, Illumination, and Expression Database,” *[New York] IEEE Comput. Soc.*

[35] P. Viola, M. J.-I. journal of computer vision, and undefined 2001, “Robust real-time object detection,” *hpl.hp.com*, Accessed: May 27, 2020. [Online]. Available: https://www.hpl.hp.com/techreports/Compaq-DEC/CRL-2001-1.pdf.

[36] X. Wang, X. Liu, L. Lu, Z. S.-2014 I. 17th International, and undefined 2014, “A new facial expression recognition method based on geometric alignment and lbp features,” *ieeexplore.ieee.org*, Accessed: May 27, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7023829/.

[37] H. Bao, … T. M.-I. I. C. on C. and, and undefined 2014, “Feature extraction and facial expression recognition based on bezier curve,” *ieeexplore.ieee.org*, Accessed: May 28, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6984772/.

[38] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, “300 Faces in-the-Wild Challenge: The first facial

landmark localization Challenge,” *cv-foundation.org*, 2013, doi: 10.1109/ICCVW.2013.59.

[39] D. Huang, C. Shan, M. Ardabilian, Y. Wang, and L. Chen, “Local Binary Patterns and Its Application to Facial Image Analysis: A Survey,” *IEEE Trans. Syst. Man, Cybern. Part C Appl. Rev.*, pp. 1–17, 2011, doi: 10.1109/TSMCC.2011.2118750ï.

[40] T. Jabid, M. H. Kabir, and O. Chae, “Robust facial expression recognition based on local directional pattern,” *ETRI J.*, vol. 32, no. 5, pp. 784–794, Oct. 2010, doi: 10.4218/etrij.10.1510.0132.

[41] X. Tan and B. Triggs, “Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions,” in *Analysis and Modeling of Faces and Gestures*, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 168–182.

[42] H. Zhou, R. Wang, C. W.-I. Sciences, and undefined 2008, “A novel extended local-binary-pattern operator for texture analysis,” *Elsevier*, Accessed: May 27, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025508002715.

[43] “59. F Ahmed, MH Kabir, in Consumer Electronics (ICCE),... - Google Scholar.” https://scholar.google.com/scholar?hl=en&as_sdt=0%252C5&q=59.%2509F+Ahmed%252C+MH+Kabir%252C+in+Consumer+Electronics+%2528ICCE%2529%252C+2012+IEEE+International+Conference+on.+Directional+te

rnary+pattern+%2528dtp%2529+for+facial+expression+recognition+%2528IEEE%252C+Las+Vegas%25 (accessed May 27, 2020).

[44] F. Ahmed, E. H.-C. J. of Engineering, and undefined 2013, "Automated facial expression recognition using gradient-based ternary texture patterns," *hindawi.com*, Accessed: May 27, 2020. [Online]. Available: https://www.hindawi.com/journals/cje/2013/831747/abs/.

[45] S. Mahto, Y. Y.-I. J. A. R. Electr., E. I. Eng, and undefined 2014, "A survey on various facial expression recognition techniques."

[46] "1. SS Meher P Maben, in Advance Computing Conference... - Google Scholar." https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=1.+SS+Meher+P+Maben%2C+in+Advance+Computing+Conference+%28IACC%29%2C+2014+IEEE+International.+Face+recognition+and+facial+expression+identification+using+PCA+%28IEEE%2C+Gurgaon%2C+2014%29%2C+pp.+1093-1098&btnG= (accessed May 27, 2020).

[47] J. De Leeuw, "NONLINEAR PRINCIPAL COMPONENT ANALYSIS AND RELATED TECHNIQUES," 2011. Accessed: May 27, 2020. [Online]. Available: https://escholarship.org/uc/item/7bt7j6nk.

[48] E. El-Shazly, … M. A.-2015 11th I., and undefined 2015, "Efficient facial and facial expression recognition using canonical correlation analysis for transform domain

features fusion and classification," *ieeexplore.ieee.org*, Accessed: May 27, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7400630/.

[49] J. Yang, D. Zhang, … A. F.-I. transactions on, and undefined 2004, "Two-dimensional PCA: a new approach to appearance-based face representation and recognition," *ieeexplore.ieee.org*, Accessed: May 27, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/1261097/.

[50] Y. Lv, Z. Feng, C. X.-2014 I. C. on Smart, and undefined 2014, "Facial expression recognition via deep learning," *ieeexplore.ieee.org*, Accessed: May 27, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7043872/.

[51] Y. Guo, G. Zhao, M. P.-I. T. on Image, and undefined 2016, "Dynamic facial expression recognition with atlas construction and sparse representation," *ieeexplore.ieee.org*, Accessed: May 27, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7423736/.

[52] G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006, doi: 10.1162/neco.2006.18.7.1527.

[53] C. Hsu, C. L.-I. transactions on N. Networks, and undefined 2002, "A comparison of methods for multiclass support vector machines," *ieeexplore.ieee.org*, Accessed: May 28, 2020. [Online]. Available:

https://ieeexplore.ieee.org/abstract/document/991427/.

[54] S. Goyal, A. Upadhyay, … R. J.-S. C. and, and undefined 2018, “Real-life facial expression recognition systems: a review,” *researchgate.net*, Accessed: May 28, 2020. [Online]. Available: https://www.researchgate.net/profile/Vamseekrishna_Allam/publication/321976016_Defected_Ground_Structure_Switchable_Notch_Band_Antenna_for_UWB_Applications/links/5cee5398458515026a63b4eb/Defected-Ground-Structure-Switchable-Notch-Band-Antenna-for-UWB-Applications.pdf#page=313.

[55] “34. Goyal, Samta Jain, et al. "Real-Life Facial Expressio... - Google Scholar.” https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=34.%09Goyal%2C+Samta+Jain%2C+et+al.+%22Real-Life+Facial+Expression+Recognition+Systems%3A+A+Review.%22+Smart+Computing+and+Informatics.+Springer%2C+Singapore%2C+2018.+311-331.&btnG= (accessed May 27, 2020).

[56] S. Bashyal and G. K. Venayagamoorthy, “Recognition of facial expressions using Gabor wavelets and learning vector quantization,” *Eng. Appl. Artif. Intell.*, vol. 21, pp. 1056–1064, 2008, doi: 10.1016/j.engappai.2007.11.010.

[57] Y. Lecun, K. Kavukcuoglu, and C. Farabet, “Convolutional Networks and Applications in Vision.” Accessed: Jun. 01, 2020. [Online]. Available: http://www.cs.nyu.edu/.

[58] V. Mavani, … S. R.-P. of the I., and undefined 2017, “Expressions of Emotion Dataset [5]. We trained the network on CFEE dataset and tested the trained model on RaFD. 2. Related Work.” Accessed: Jun. 01, 2020. [Online]. Available: http://openaccess.thecvf.com/content_ICCV_2017_workshops/w40/html/Mavani_Facial_Expression_Recognition_ICCV_2017_paper.html.

[59] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks.” Accessed: Jun. 02, 2020. [Online]. Available: http://code.google.com/p/cuda-convnet/.

[60] L. Deng and D. Yu, “the essence of knowledge Deep Learning Methods and Applications Foundations and Trends ® in Signal Processing Deep Learning Methods and Applications,” *docs.huihoo.com*, doi: 10.1561/2000000039.

[61] S. Li and W. Deng, “Deep Facial Expression Recognition: A Survey.” Accessed: May 31, 2020. [Online]. Available: http://www.cse.oulu.fi/CMV/Downloads/Oulu-CASIA.

[62] “Head-pose invariant facial expression recognition... - Google Scholar.” https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Head-pose+invariant+facial+expression+recognition+using+convolutional+neural+networks%2C”+in+Proceedings+of+the+4th+IEEE+International+Conference+on+Multimoda

l+Interfaces.+IEEE+Computer+Society%2C+2002%2C+p.+529&btnG= (accessed Jun. 17, 2020).

[63] M. Matsugu, K. Mori, Y. Mitari, and Y. Kaneda, “Subject independent facial expression recognition with robust face detection using a convolutional neural network,” *Elsevier*, doi: 10.1016/S0893-6080(03)00115-1.

[64] K. Simonyan and A. Zisserman, “VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION,” 2015. Accessed: Jun. 04, 2020. [Online]. Available: http://www.robots.ox.ac.uk/.

[65] C. Szegedy *et al.*, “Going Deeper with Convolutions.” Accessed: Jun. 15, 2020. [Online]. Available: https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Szegedy_Going_Deeper_With_2015_CVPR_paper.html.

[66] B. Sun *et al.*, “Combining Multimodal Features within a Fusion Network for Emotion Recognition in the Wild,” 2015, pp. 497–502, doi: 10.1145/2818346.2830586.

[67] B. Sun Liandong Li Guoyan Zhou Jun He Bo Sun, L. Li, and G. Zhou, “Facial expression recognition in the wild based on multimodal texture features,” *spiedigitallibrary.org*, doi: 10.1117/1.JEI.25.6.061407.

[68] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation.” Accessed: Jun. 17, 2020. [Online]. Available: http://arxiv.

[69] J. Li *et al.*, “Facial expression recognition with faster R-CNN,” *Elsevier*, Accessed: Jun. 17, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050917303447.

[70] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.” Accessed: Jun. 17, 2020. [Online]. Available: https://github.com/.

[71] S. Ji, W. Xu, M. Yang, K. Y.-I. transactions on pattern, and undefined 2012, “3D convolutional neural networks for human action recognition,” *ieeexplore.ieee.org*, Accessed: Aug. 11, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6165309/.

[72] Y. LeCun, L. Bottou, … Y. B.-P. of the, and undefined 1998, “Gradient-based learning applied to document recognition,” *ieeexplore.ieee.org*, Accessed: Aug. 11, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/726791/.

[73] J. Kolar and E. Salter, “Craniofacial anthropometry: Practical measurement of the head and face for clinical, surgical, and research use,” 1997.

[74] T. Ojala, M. Pietikäinen, and T. Mäenpää, “Multiresolution Gray Scale and Rotation Invariant Texture Classification with Local Binary Patterns.” Accessed: May 31, 2020. [Online]. Available: http://www.ee.oulu.fi/research/imag/texture.

[75] K. Yan, W. Zheng, T. Zhang, Y. Zong, and Z. Cui, "Cross-database non-frontal facial expression recognition based on transductive deep transfer learning," Nov. 2018, Accessed: Jun. 17, 2020. [Online]. Available: http://arxiv.org/abs/1811.12774.

[76] L. Farkas, "Anthropometry of the Head and Face," 1994.

[77] F. Tsalakanidou, S. M.-P. Recognition, and undefined 2010, "Real-time 2D+ 3D facial action and expression recognition," *Elsevier*, Accessed: May 29, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320309004786.

[78] G. Sandbach, S. Zafeiriou, M. Pantic, L. Y.-I. and V. Computing, and undefined 2012, "Static and dynamic 3D facial expression recognition: A comprehensive survey," *Elsevier*, Accessed: May 31, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0262885612000935.

[79] L. G. Farkas, "Accuracy of anthropometric measurements: Past, present, and future," *Cleft Palate-Craniofacial J.*, vol. 33, no. 1, pp. 10–18, 1996, doi: 10.1597/1545-1569_1996_033_0010_aoampp_2.3.co_2.

[80] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.*, vol. 1, no. 4, pp. 321–331, Jan. 1988, doi: 10.1007/BF00133570.

[81] "Automatic construction of 2D shape models," *ieeexplore.ieee.org*, Accessed: May 29, 2020. [Online].

Available: https://ieeexplore.ieee.org/abstract/document/922703/.

[82] G. J. Edwards, A. Lanitis, C. J. Taylor, and T. F. Cootes, “Statistical Models of Face Images-Improving Specificity,” *Elsevier*, doi: 10.5244/C.10.14.

[83] D. Beymer, T. P.- Science, and undefined 1996, “Image representations for visual learning,” *science.sciencemag.org*, Accessed: May 31, 2020. [Online]. Available: https://science.sciencemag.org/content/272/5270/1905.short.

[84] T. F. Cootes, G. J. Edwards, and C. J. Taylor, “Active Appearance Models.” Accessed: May 29, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/927467/.

[85] I. Cohen, N. Sebe, A. Garg, L. S. Chen, and T. S. Huang, “Facial expression recognition from video sequences: temporal and static modeling,” *Comput. Vis. Image Underst.*, vol. 91, pp. 160–187, 2003, doi: 10.1016/S1077-3142(03)00081-X.

[86] S. Hosoi, E. Takikawa, M. K.-S. I. International, and undefined 2004, “Ethnicity estimation with facial images,” *ieeexplore.ieee.org*, Accessed: May 31, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/1301530/.

[87] H. Sadeghi, A. Raie, M. R. Mohammadi, A.-A. Raie, and M.-R. Mohammadi, “Facial expression recognition using

geometric normalization and appearance representation," *ieeexplore.ieee.org*, 2013, doi: 10.1109/IranianMVIP.2013.6779970.

[88] S. Zhang, X. Zhao, and B. Lei, "Facial Expression Recognition Based on Local Binary Patterns and Local Fisher Discriminant Analysis." Accessed: May 31, 2020. [Online]. Available: https://dl.acm.org/doi/abs/10.5555/2190016.2190019.

[89] R. Verma, M. D.-2013 26th I. C. Conference, and undefined 2013, "Fast facial expression recognition based on local binary patterns," *ieeexplore.ieee.org*, Accessed: May 31, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6567728/.

[90] T. Ahsan, T. Jabid, and U. P. Chong, "Facial expression recognition using local transitional pattern on gabor filtered facial images," *IETE Tech. Rev. (Institution Electron. Telecommun. Eng. India)*, vol. 30, no. 1, pp. 47–52, Jan. 2013, doi: 10.4103/0256-4602.107339.

[91] L. G.-2011 S. I. C. on I. and and undefined 2011, "Smile expression classification using the improved BIF feature," *ieeexplore.ieee.org*, Accessed: May 31, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6005972/.

[92] D. Gabor, "Theory of communication. Part 1: The analysis of information," *J. Inst. Electr. Eng. - Part III Radio Commun. Eng.*, vol. 93, no. 26, pp. 429–441, Nov. 1946, doi: 10.1049/ji-3-2.1946.0074.

[93] T. R. Almaev, M. Valstar, and M. F. Valstar, "Local Gabor Binary Patterns from Three Orthogonal Planes for Automatic Facial Expression Recognition Computerized Performance Assessment for Clinical Teams View project 'ADAM'-Anthropomorphic Design through Advanced Manufacturing View project Local Gabor Binary Patterns from Three Orthogonal Planes for Automatic Facial Expression Recognition," *ieeexplore.ieee.org*, 2013, doi: 10.1109/ACII.2013.65.

[94] F. J. Chang, A. T. Tran, T. Hassner, I. Masi, R. Nevatia, and G. Medioni, "Deep, Landmark-Free FAME: Face Alignment, Modeling, and Expression Estimation," *Int. J. Comput. Vis.*, vol. 127, no. 6–7, pp. 930–956, Jun. 2019, doi: 10.1007/s11263-019-01151-x.

[95] W. Gu, C. Xiang, Y. Venkatesh, D. Huang, H. L.-P. recognition, and undefined 2012, "Facial expression recognition using radial encoding of local Gabor features and classifier synthesis," *Elsevier*, Accessed: May 31, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320311002056.

[96] M. Siddiqi *et al.*, "Facial expression recognition using active contour-based face detection, facial movement-based feature extraction, and non-linear feature selection," *Springer*, Accessed: May 31, 2020. [Online]. Available: https://link.springer.com/article/10.1007/s00530-014-0400-2.

[97] R. A. FISHER, “THE STATISTICAL UTILIZATION OF MULTIPLE MEASUREMENTS,” *Ann. Eugen.*, vol. 8, no. 4, pp. 376–386, Aug. 1938, doi: 10.1111/j.1469-1809.1938.tb02189.x.

[98] E. Tzanakou, *Supervised and unsupervised pattern recognition: feature extraction and computational intelligence*. 2017.

[99] E. Aditya, J. Saputro, … N. R.-… on E. and, and undefined 2019, “Face Recognition Implementation System As A Media Access To Restricted Room With Histogram Equalization And Fisherface Methods,” *ieeexplore.ieee.org*, Accessed: May 31, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8909665/.

[100] D. L. Swets, J. (Juyang, and) Weng, “Using Discriminant Eigenfeatures for Image Retrieval,” 1996. Accessed: May 31, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/531802/.

[101] A. M. Martõ Ânez and A. C. Kak, “PCA versus LDA.” Accessed: May 31, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/908974/.

[102] P. Hu, D. Cai, S. Wang, A. Yao, and Y. Chen, “Learning Supervised Scoring Ensemble for Emotion Recognition in the Wild,” *dl.acm.org*, vol. 17, pp. 553–560, Nov. 2017, doi: 10.1145/3136755.3143009.

[103] R. Cui, M. Liu, and M. Liu, “Facial expression recognition based on ensemble of mulitple CNNs,” in

Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2016, vol. 9967 LNCS, pp. 511–518, doi: 10.1007/978-3-319-46654-5_56.

[104] A. Hamid, A. Mohammed, M. Khalfalla, and E. Elfaki, "Feature Extraction Based on Discrete Cosine Transform for Mammogram Image Classification," 2016. Accessed: Jun. 01, 2020. [Online]. Available: http://repo.uofg.edu.sd/handle/123456789/1257.

[105] "Tumor Mass Detection through Gabor Filters and Supervised... - Google Scholar." https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Tumor+Mass+Detection+through+Gabor+Filters+and+Supervised+Pixel-Based+Classification+in+Breast+Cancer&btnG= (accessed Jun. 01, 2020).

[106] F. Amato, A. López, E. María Peña-Méndez, P. Vaňhara, A. Hampl, and J. Havel, "Artificial neural networks in medical diagnosis," *J Appl Biomed*, vol. 11, pp. 47–58, 2013, doi: 10.2478/v10136-012-0031-x.

[107] K. Metaxiotis and J. Psarras, "Neural networks in production scheduling: intelligent solutions and future promises," *Appl. Artif. Intell.*, vol. 17, no. 4, pp. 361–373, Apr. 2003, doi: 10.1080/713827140.

[108] P. Sibi, S. A. Jones, and P. Siddarth, "ANALYSIS OF DIFFERENT ACTIVATION FUNCTIONS USING BACK PROPAGATION NEURAL NETWORKS," *J. Theor. Appl. Inf. Technol.*, vol. 31, no. 3, 2013, Accessed:

Jun. 01, 2020. [Online]. Available: www.jatit.org.

[109] P. Ekman, … D. M.-… A. C. S., and undefined 1997, "Facial expression in affective disorders," *books.google.com*, Accessed: Jun. 02, 2020. [Online]. Available: https://books.google.com/books?hl=en&lr=&id=fFGYs079-7YC&oi=fnd&pg=PA331&dq=4.%09P.+Ekman+and+Wallace+Friesen.+Facial+action+coding+system:+a+technique+for+the+measurement+of+facial+movement.+Palo+Alto:+Consulting+Psychologists+(1978).&ots=ipUbFgLAPf&sig=Id8As_HHAFhxa3Y6RXXwkjBnFxk.

[110] M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Fully Automatic Facial Action Recognition in Spontaneous Behavior." Accessed: Jun. 02, 2020. [Online]. Available: http://kolmogorov.sourceforge.net.

[111] M. Pantic, L. R.-I. T. on Systems, and undefined 2004, "Facial action recognition for facial expression analysis from static face images," *ieeexplore.ieee.org*, Accessed: Jun. 02, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/1298893/.

[112] Y. Tian, T. Kanade, and J. F. Colin, "Recognizing Action Units for Facial Expression Analysis," 2002, pp. 32–66.

[113] E. B. Michael Lewis, J. M. Haviland-jones, and L. Feldman Barrett, *HANDBOOK OF EMOTIONS Third Edition THE GUILFORD PRESS New York London*. 2008.

[114] H.-W. Ng, D. Nguyen, V. Vonikakis, and S. Winkler, "Deep Learning for Emotion Recognition on Small Datasets Using Transfer Learning," *dl.acm.org*, pp. 443–449, Nov. 2015, doi: 10.1145/2818346.2830593.

[115] P. Viola and M. J. Jones, "Robust Real-Time Face Detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, May 2004, doi: 10.1023/B:VISI.0000013087.49260.fb.

[116] N. Dalal, B. T. Histograms, and B. Triggs, "Histograms of Oriented Gradients for Human Detection," *ieeexplore.ieee.org*, pp. 886–893, 2005, doi: 10.1109/CVPR.2005.177ï.

[117] C. Liu and H. Wechsler, "Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition," 2002. Accessed: Jun. 02, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/999679/.

[118] S. V. Kozyrev, "Classification by ensembles of neural networks," *P-Adic Numbers, Ultrametric Anal. Appl.*, vol. 4, no. 1, pp. 27–33, Jan. 2012, doi: 10.1134/S2070046612010049.

[119] S. Ali Khayam, "The Discrete Cosine Transform (DCT): Theory and Application 1," 2003. Accessed: Jun. 01, 2020. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.184.6102&rep=rep1&type=pdf.

[120] G. Goswami, "Machine Learning approaches to

Keystroke dynamics-based authentication system with Enhanced Features," *pdfs.semanticscholar.org*, Accessed: Jun. 01, 2020. [Online]. Available: https://pdfs.semanticscholar.org/86c4/aa33259efda09e9493378feec15e24c79240.pdf.

[121] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning Spatiotemporal Features with 3D Convolutional Networks." Accessed: Jun. 17, 2020. [Online]. Available: https://www.cv-foundation.org/openaccess/content_iccv_2015/html/Tran_Learning_Spatiotemporal_Features_ICCV_2015_paper.html.

[122] M. Budka, B. G.-I. transactions on neural networks and, and undefined 2012, "Density-preserving sampling: robust and efficient alternative to cross-validation for error estimation," *ieeexplore.ieee.org*, Accessed: Jun. 02, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6360017/.

[123] U. Mlakar, I. Fister, J. Brest, B. P.-E. S. with Applications, and undefined 2017, "Multi-objective differential evolution for feature selection in facial expression recognition systems," *Elsevier*, Accessed: Jun. 02, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417417305134.

[124] B. U. Ca and Y. G. Fr, "No Unbiased Estimator of the Variance of K-Fold Cross-Validation Yoshua Bengio Yves Grandvalet," 2004. Accessed: Jun. 02, 2020. [Online]. Available:

http://www.jmlr.org/papers/v5/grandvalet04a.html?92f58540.

[125] R. R. Picard and R. D. Cook, “Cross-validation of regression models,” *J. Am. Stat. Assoc.*, vol. 79, no. 387, pp. 575–583, 1984, doi: 10.1080/01621459.1984.10478083.

[126] A. Jain, R. Duin, J. M. and machine intelligence, and undefined 2000, “Statistical pattern recognition: A review.” Accessed: Jun. 02, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/824819/.

[127] R. Kohavi, “A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection,” 1995. Accessed: Jun. 02, 2020. [Online]. Available: http://robotics.stanford.edu/~ronnyk.

[128] O. Kilinc, I. U.-2015 I. 14th I. C. on, and undefined 2015, “Source-aware partitioning for robust cross-validation,” *ieeexplore.ieee.org*, Accessed: Jun. 02, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7424464/.

[129] P. J. García-Laencina AE José-Luis Sancho-Gómez AE Aníbal R Figueiras-Vidal, “Pattern classification with missing data: a review,” *Springer*, vol. 19, no. 2, pp. 263–282, 2010, doi: 10.1007/s00521-009-0295-6.

[130] M. Stone, “Cross-Validatory Choice and Assessment of Statistical Predictions,” *J. R. Stat. Soc. Ser. B*, vol. 36, no. 2, pp. 111–133, Jan. 1974, doi: 10.1111/j.2517-6161.1974.tb00994.x.

[131] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. 2009.

[132] P. Refaeilzadeh, L. Tang, H. L.-E. of database, and undefined 2009, “Cross-Validation.,” *Springer New York*.

[133] S. Geisser, “The predictive sample reuse method with applications,” *J. Am. Stat. Assoc.*, vol. 70, no. 350, pp. 320–328, 1975, doi: 10.1080/01621459.1975.10479865.

[134] B. Efron, “Estimating the error rate of a prediction rule: Improvement on cross-validation,” *J. Am. Stat. Assoc.*, vol. 78, no. 382, pp. 316–331, 1983, doi: 10.1080/01621459.1983.10477973.

[135] J. Shao, “Linear model selection by cross-validation,” *J. Am. Stat. Assoc.*, vol. 88, no. 422, pp. 486–494, 1993, doi: 10.1080/01621459.1993.10476299.

[136] A. Elisseeff, “Leave-one-out error and stability of learning algorithms with applications Stability of Randomized Learning Algorithms Source Leave-one-out error and stability of learning algorithms with applications.” Accessed: Jun. 02, 2020. [Online]. Available: https://www.researchgate.net/publication/228597518.

[137] A. Celisse and S. Robin, “Nonparametric density estimation by exact leave-out cross-validation Nonparametric density estimation by exact leave-p-out cross-validation,” *Comput. Stat. Data Anal.*, vol. 52, pp. 2350–2368, 2008, doi: 10.1016/j.csda.2007.10.002.

[138] S. Arlot, A. C.-S. surveys, and undefined 2010, “A survey of cross-validation procedures for model selection,” *projecteuclid.org*, Accessed: Jun. 02, 2020. [Online]. Available: https://projecteuclid.org/euclid.ssu/1268143839.

[139] P. M. Dixon, J. Weiner, T. Mitchell-Olds, and R. Woodley, “Bootstrapping the Gini Coefficient of Inequality,” *Ecology*, vol. 68, no. 5, pp. 1548–1551, Oct. 1987, doi: 10.2307/1939238.

[140] S. Weiss and C. Kulikowski, “Computer systems that learn: classification and prediction methods from statistics, neural nets, machine learning, and expert systems,” 1991, Accessed: Jun. 02, 2020. [Online]. Available: https://dl.acm.org/citation.cfm?id=102700.

[141] R. W. Johnson, “An Introduction to the Bootstrap,” *Teach. Stat.*, vol. 23, no. 2, pp. 49–54, 2001, doi: 10.1111/1467-9639.00050.

[142] S. L. Salzberg, “On comparing classifiers: Pitfalls to avoid and a recommended approach,” *Data Min. Knowl. Discov.*, vol. 1, no. 3, pp. 317–328, 1997, doi: 10.1023/A:1009752403260.

[143] R. Bowen, C. A.- Gerontology, and undefined 2004, “Living and dying for sex,” *karger.com*, Accessed: Jun. 03, 2020. [Online]. Available: https://www.karger.com/Article/PDF/79125.

[144] M. Uddin, M. Hassan, A. Almogren, … A. A.-I., and undefined 2017, “Facial expression recognition utilizing

local direction-based robust features and deep belief network,” *ieeexplore.ieee.org*, Accessed: Jun. 03, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7867858/.

[145] G. Guo, R. Guo, X. L.-I. T. on A. Computing, and undefined 2013, “Facial expression recognition influenced by human aging,” *ieeexplore.ieee.org*, Accessed: Jun. 03, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6517179/.

[146] R. P. Holder and J. R. Tapamo, “Improved gradient local ternary patterns for facial expression recognition,” *Eurasip J. Image Video Process.*, vol. 2017, no. 1, 2017, doi: 10.1186/s13640-017-0190-5.

[147] N. Nour and S. Viriri, *Facial Expression Recognition Using Directional Gradient Local Ternary Patterns*, vol. 11909 LNAI. 2019.

[148] I. Namatēvs, “Deep Convolutional Neural Networks: Structure, Feature Extraction and Training,” *content.sciendo.com*, vol. 20, pp. 40–47, 2017, doi: 10.1515/itms-2017-0007.

[149] M. Lapata, P. Blunsom, A. K.-P. of the 15th C. of, and undefined 2017, *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*. .

[150] Y. Zheng, Q. Liu, E. Chen, Y. Ge, and J. L. Zhao, “Time series classification using multi-channels deep convolutional neural networks,” in *Lecture Notes in*

Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2014, vol. 8485 LNCS, pp. 298–310, doi: 10.1007/978-3-319-08010-9_33.

[151] N. Ann *et al.*, "Application of Deep Convolutional Neural Network for Automated Detection of Myocardial Infarction Using ECG Signals Trust for consensus in group decision under Social Network View project Development of computer methods for the study of ocular surface temperature View project U Rajendra Acharya Application of Deep Convolutional Neural Network for Automated Detection of Myocardial Infarction Using ECG Signals," *Elsevier*, 2017, doi: 10.1016/j.ins.2017.06.027.

[152] R. Liu, G. Meng, B. Yang, … C. S.-I. T. on, and undefined 2016, "Dislocated time series convolutional neural architecture: An intelligent fault diagnosis approach for electric machine," *ieeexplore.ieee.org*, Accessed: Jun. 04, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7797508/.

[153] N. Persaud *et al.*, "Awareness-related activity in prefrontal and parietal cortices in blindsight reflects more than superior visual performance," *Elsevier*, 2011, doi: 10.1016/j.neuroimage.2011.06.081.

[154] C. Affonso *et al.*, "Deep learning for biological image classification," *Expert Syst. Appl.*, vol. 85, pp. 114–122, Nov. 2017, doi: 10.1016/j.eswa.2017.05.039.

[155] Y. Jia *et al.*, "Caffe: Convolutional Architecture for Fast

Feature Embedding *," *dl.acm.org*, pp. 675–678, Nov. 2014, doi: 10.1145/2647868.2654889.

[156] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. 2016.

[157] V. Gudipati, O. Barman, … M. G.-… & A. (CT, and undefined 2016, "Efficient facial expression recognition using adaboost and haar cascade classifiers," *ieeexplore.ieee.org*, Accessed: Jun. 05, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7868250/.

[158] A. Kaehler and G. Bradski, *Learning OpenCV 3: computer vision in C++ with the OpenCV library*. 2016.

[159] M. J. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," *ieeexplore.ieee.org*, 2001, doi: 10.1109/CVPR.2001.990517.

[160] J. Lu, X. Fu, T. Z.-& Communications, C. & B. Data, and undefined 2017, "A smart system for face detection with spatial correlation improvement in IoT environment," *ieeexplore.ieee.org*, Accessed: Jun. 07, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8397550/.