Sudan University of Science &Technology

College of Graduate Studies

Component-based Gender Identification from Facial Images

تحديد الجنس القائم على مكونات صورة الوجه

This Thesis is submitted in fulfilment for

The degree of Doctor of Philosophy in

Computer Science

Prepared by

Salma Mohammed Osman

supervised by

Prof. Serestina Viriri

SEPTEMBER 2021

Component-based Gender Identification from Facial Images

BY

SALMA MOHAMMED OSMAN MOHAMMED MUSA

SUDAN UNIVERSITY FOR SCIENCE & TECHNOLOGY

Supervisor:

PROF. SERESTINA VIRIRI

SEPTEMBER 2021

تحديد الجنس القائم على مكونات صورة الوجه

Component-based Gender Identification from Facial Images

BY

SALMA MOHAMMED OSMAN MOHAMMED MUSA

A dissertation submitted in Partial 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

SEPTEMBER 2021

ABSTRACT

As one of the most flourishing applications of image analysis and understanding, face identification has gained significant attention, especially in the past several decades. In the field of image processing, face is one of the most important biometric traits and is becoming more popular for many application purposes now a days. This is a very important field of image processing because of its applications in many areas like security, monitoring, surveillance, commercial profiling, and human-computer interaction. Most previous researchers have been using whole face to classify gender. In this study many techniques for gender classification proposed and the experimental results have shown that the proposed techniques have high accuracy rate, fast computational time besides reduction in the processing time. The study applied in many situations (whole face and components of the face) of gender from facial images using different datasets (FERET, ESSEX and UOFG). Then two types of face detecting algorithms, namely viola & jones and Discriminative Response Map Fitting DRMF model, are applied. Many feature extraction techniques are also applied; global Discrete Cosine Transform (DCT), Block based Discrete Cosine Transform (BBDCT) and hybrid DCT Discrete wavelet Transform DWT) and Local Binary Pattern (LBP), Local Directional LDP), Local Ternary pattern (LTP) and proposed Dynamic Local Ternary Pattern (DLTP) then CNN), then KNN and SVM are used in the last step to classify the images. The experimental results show that the performance of recognition (classification) does not depend only on feature extraction approach but also on the other steps of the recognition process such as the pre-processing stages and the classification algorithm. Moreover, it has been proved that the whole face is not required for gender identification using facial images. This has confirmed that the proposed Dynamic local ternary pattern (DLTP) is an accurate and efficient feature extraction technique for gender identification. Finally, ResNet-50 is applied with CNN to extract the feature of facial images and obtained features in the last fully connected layer which are used as inputs to SVM classifier to produce the final classification result. The study shows that when we used FERET datasets and SVM classifier in the whole face, the accuracy is 95%; but when we used components of the face the accuracy is 98.55%.

However, the best accuracy when we used the proposed DLTP is about 98.90% as we see the accuracy gradually increased.

مستخلص البحث

يعتبر تحديد الوجة والتعرف عليه واحد من التطبيقات الأكثر ازدهارا لتحليل الصور وفهمها ، وقد اكتسب تحديده اهتماما كبيرا ، خاصة في العقود العديدة الماضية. في مجال معالجة صورة الوجه هو واحد من أهم السمات البيومترية وأصبحت أكثر شعبية لكثير من التطبيقات في الاونه الاخيره. خلال عدة سنوات، اكتسب تصنيف نوع الجنس من صور الوجه أهمية هائلة وأصبح مجالاً شائعاً للبحث. هذا الحقل مهم جدا من معالجة الصور بسبب تطبيقاته في العديد من المجالات مثل الأمن والرصد والمراقبة والتنميط التجاري والتفاعل بين الإنسان والكمبيوتر. في هذه الدراسة اقترحت تقنيات جديدة في التصنيف الجنساني نتجت عنها ارتفاع معدل الدقة ، والخطوات الحسابية السريعة وتقليل الوقت لعملية التصنيف. الدراسه كانت في حالات متعدده ابتدا بتحديد الجنس من الوجه ككل ثم تحديد الجنس من مناطق محدده من الوجه مثل (العيون، الانف، الفم ، الجبهه، الخدود ، الذقن) وتم اخذ صور الدراسه من عدد من مجموعات بيانات الصور مثل (FERET,ESEEX ,UOFG) ويتم ذلك بخوارزميات استخلاص او استخراج الوجه او مناطق محدده من الوجه مثل خوارزمية (viola & jones) ونموزج (DRMF)بعد ذلك تدخل هذه المناطق المجرده من الخلفيات وبقايا أجزاء الصوره الي خوارزميات استخلاص الخصائص والمميزات وهنا تم اخذ هذه الصفات بواسطة نوعين من خوارزميات استخراج الصفات عامه مثل (DCT, BBDCT and hybrid DCT DWT) ومحليه مثل LBP, LDP, LTP & DLTP)وأيضا في هذه الخطوه تم استخدام خوارزميه التعليم العميق(CNN) وفي الخطوه الاخيره و لمعرفه ما اذا كانت الصفات للصوره لمراه او لرجل تم استخدام خوارزميتي (KNN,SVM) للتصنيف. أظهرت النتائج التجريبيه ان تصنيف الجنس من صورة الوجه لايعتمد فقط علي علي نهج او خوارزميات استخراج الصفات ولكن علي بقية الخطوات الأخرى لعملية التصنيف مثل مرحلة المعالجه المسبقه للصور وعمليات استخرج الصفات بخوارزميات معينه . ووجد أيضا ان انه ليس من الضروري استخدام الوجه بالكامل لاكتمال عملية التصنيف وانه من الممكن التصنيف باستخدام مناطق معينه من الوجه. ومن الدراسه توصلنا الي ان الخورزميه التي تم تطورها (DLTP) من النمط المحلي الديناميكي هي تقنيه مميزه وفعاله ومناسبه تماما لعملية تحديد الجنس من صورة الوجه. وأخيرا استخدمت شبكة ال (CNN) التي تم تطبيقها علي الصور للتعرف علي الجنس من الوجه وذلك باستخدام خوارزمية التصنيف (SVM) معها بدلا من خوارزمية التصيف الاساسيه التابعه للشبكه واستخلاص النتيجه. وتظهر الدراسه انه عندما تم استخدام مجموعة البيانات (FERET) وخوارزمية التصنيف (SVM) في الوجه كاملا بلغت الدقه (95%)، وعندما استخدمنا مناطق معينه كانت الدقه (98.55%) ولكن افضل دقه ظهرت عندما استخدمنا الخوارزميه المحسنه المقترحه (DLTP) وكانت الدقه(98.90%) لاحظنا ان في كل مره تزداد الدقه تدريجيا.

ACKNOWLEDGEMENTS

In the Name of Allah, the Most Gracious, the Most Merciful.

First and foremost, I thank my God (Allah S.W.T) who supplies all my needs.

Many different people helped me with various parts of this thesis. Technical support, moral support, mental support. I want to thank my supervisor, Prof. Serestina Viriri, for his advice, support, and guidance. I would also like to thank my committee members, for the time and effort they spent to read and comment on my work. The most important “acknowledgement" goes to my parents and my husband; they both supported and pushed me when I needed motivation.

I would like to acknowledge the financial support rendered by the University of Gezira and the Faculty of Mathematical and Computer Science.

DECLARATION

I hereby declare that this dissertation 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.

Salma Mohammed Osman Mohammed Musa

Signature_____ Date_____

TABLE OF CONTENTS

LIST OF TABLES

Table No. Page No.

LIST OF FIGURES

Figure No. Page No.

LIST OF ABBREVIATIONS

2D(PCA)	Two dimensional Principal component analysis
3D	Three diminutions
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
FERET	Facial Recognition Technology
FLD	Fisher linear discriminate
HMM	Hidden Markov model
ICA	International component analysis
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
LVQ	Learning vector quantization
NIN	Network in network
NN	Neural network

PCA	Principal component analysis
RBF	Radial basis function
RBM	Restricted Boltzmann machine

CHAPTER ONE
GENERAL INTRODUCTION

1.1 INTRODUCTION

Human beings are identified in their daily life by their biometrics or security mechanisms. The security mechanisms that identify human beings are items such as keys, security discs or tags to obtain entrance into buildings, gates, and homes. Other security mechanisms are passwords and pins to obtain money from bank accounts, logging on to computers, cell phones and tablets, and social media or social networking such as Gmail, Facebook, and Twitter. These security mechanisms are unreliable as they can be lost or stolen. Hence, the use of security mechanisms that are based on biometrics is increasing as these features are more secure than others. The face is a significant human biometric feature. Faces give an access to the mechanisms which control human social and emotional lives (Ekman, et al., 1993). Face identification is a fascinating branch of object recognition which identifies or verifies human subjects in different scenes from a video source or a digital image. There are a lot of promising applications for a flourishing gender classification method like demographic data collection, human identification, computer vision approaches for monitoring people, smart human computer interface. Nevertheless, human being can easily perform gender classification tasks while machines require human brainpower to perform the tasks (Alam, et al., 2016). These days, the world is getting more dependent on machine. Therefore, gender identification has become a significant topic of research in computer vision and image processing aspect. Gesture detection, person recognition, face detection, motion capture and detection are all very reliable in authentication and security process which cause a high increase in their demand (Alam, et al., 2016). Combination of gender classification and face detection methods might appear as easy task. Nevertheless, the process is more complicated than it seems due to involvement of many aspects to be considered and gender classification has been presented in many research works as psychological literature but there is a small number of machine vision methods have been introduced (Isa, Nurul Zarina Md, 2010). This study focuses on component-based gender identification from facial images and the local facial features algorithm used to extract the gender from frontal. In this kind of feature extraction,

features from some facial points like face, eyes and nose are extracted. Then the vector of feature is passed to the classifier to specify the gender. Some researchers work in this area like (Qacimy, et al., 2014) and (Lawgali, Ahmed, 2005).

Supervised learning and unsupervised learning are the two types of learning, in supervised learning, the aim is to learn mapping from the input (instance) to an output (target or label) where a supervisor provides the correct values.

Deep learning is a branch of machine learning in which human brain fascinates artificial neural networks algorithms, acquire learning from many data. Just as human learn from past events, series of repeated task are performed by deep learning algorithm, it slightly adjusts it a little each time to increase the result, so that it trains computers to perform human natural activities: learn through examples. Deep learning is the main technology which enables applications like driverless cars, to understand a traffic indication and to realize a pedestrian. It is the key to voice control in consumer interest devices like tablets, TVs, phones and hands-free speakers (Serj, et al., 2018).

Convolutional neural networks (CNNs) have been broadly used in computer vision tasks. The results of CNNs have proven its robustness in object recognition localization in variant images. Recently, the applied deep learning techniques are emerged also in the analysis function of medical image. The bulk of interesting deep learning works that were presented to enhance the performance of the medical image analysis techniques have been applied either by modifying the architecture of the existing deep learning networks or proposing new ones (nchez-Delacruz, Eddy and Parra, Pilar Pozos, 2018).

In this study, we focus on how to improve gender identification using facial components (forehead, eyes, mouth, cheeks, nose, chin), and how to extract feature from each of these facial components. Some of these features are extracted by shape and others using textual features, then we combine the feature vector for these facial components to extract the total feature of the face. Also, two areas in machine learning and deep leaning were used and applied the three steps of any gender recognition (face detecting, feature extraction and classification) in machine and deep learning. First, the study applied the steps in the whole face and then applied in component of the face then compared the two results.

1.2 MOTIVATION

A human can easily determine gender but this is a great task for a computer to perform, however the rate of technological development in the field of artificial intelligence has led to huge machine dependence on computer vision techniques such as face and gesture detection to name a few. Most previous researchers have been using facial features individually and the face as a whole to classify gender, thus improving gender classification accuracy by including feature fusion. Fusing different feature extractors is the main desired outcome in this research work. Very few researchers have indulged into feature fusion to improve gender classification rates such as (Liu, Chun-yan and Fu, Bo and Huang, He-Jiao, 2014) …who classified the facial components and also used the hair. Hence this research is motivated by the need to find out if feature fusion does improve gender classification.

Fusion of features has not been researched as much, as there are some features from the facial components which have a very high classification rate hence fusing them with other features aims to increase the classification rate. We seek to determine first if fusing components will lead to an increase in classification accuracy as even with human detection certain facial components have a greater impact in determining ones gender.

The use of various facial components such as the eyes, nose, mouth, forehead and cheeks can all be used to determine gender and hence this research will show the facial component with the greatest accuracy when it comes to gender classification and fuse it with that with the lowest accuracy.

1.3 APPLICATIONS

There are several applications of gender identification such as:

Surveillance Systems – It is possible to build a human-computer interaction surveillance system to identify one's attributes such as ethnicity, age, and gender. There are many reasons that surveillance systems can be used; namely, terror-related crimes, law-enforcement, and security (kinen, Erno, 2007).

Security - Facial security is slowly replacing password logins on certain applications and computer systems. At Manchester University, researchers are trying hard to produce consumer-focused facial recognition technology for security applications. Many famous companies show their interest in incorporating the security facial recognition system into their products (Ng, Epin, 2015).

Image Tagging - Facebook's automatic tagging feature uses facial and gender recognition for users to suggest people they might want to tag in their photos. This has been seen to save people time when posting images on social media. Image tagging is available on applications such as Apple's iPhoto, Google's Picasa, and Facebook (Sinha, Chandrakamal, 2013).

General Identity Verification – gender and facial identification using facial images have general uses, including banking, electoral registration, electronic commerce, new-born's identity, passports, national identity cards and employee identification cards (Viola, Paul and Jones, Michael, 2001), (Nazir, et al., 2010) and (Mohamed, Salma and Nour, Nahla and Viriri, Serestina, 2018)

1.4 PROBLEM STATEMENT AND ITS SIGNIFICANT

The human face can be separated into different components. Several methods were introduced in the literature for human gender identification through facial images. The methods are applied to obtain the colour of the facial region using hue and saturation values of an image. Most of these methods use the whole face gender identification. Obtaining the components of the facial region, for example the mouth, eyes, nose, forehead, cheeks and chin, can improve the extraction of the gender from that image. A component image is shown in figure 1 for the purpose of subsequent discussion.

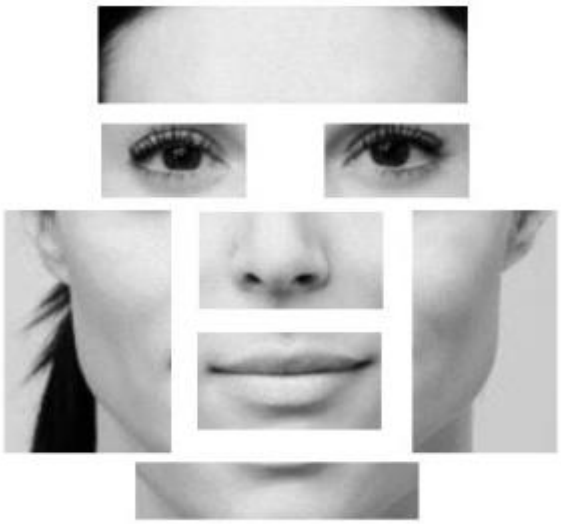


Figure 1: Extract of facial component (Viola, Paul and Jones, Michael, 2001)

Facial images are prone to variation in pose, illumination and expression (Viola, Paul and Jones, Michael, 2001) .Another issue observed with facial images is that they may also have sunglasses, glasses, hats and scarfs. These accessories make it difficult to ascertain the gender of that image.

The problems related with this research are:

- There is no standard algorithm for gender identification 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.

1.5 RESEARCH OBJECTIVES

The basic objectives of this research are to model a sufficiently robust and accurate model for human gender identification using facial component. This research precisely targets the fact of creating new techniques which could improve efficiently the

performance of facial data extraction and gender classification techniques. The specific objectives are:

- To identify accurately the human gender from facial component.
- To improve gender identification using different feature extraction techniques from facial component.
- To model framework for accurate facial component- based gender identification.

1.6 RESEARCH QUESTIONS

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

- Is it possible to identify gender facial images (efficiently) based on facial component accurately?
- Is it possible to improve the accuracy rate of gender identification using facial component?
- What facial components are most informative for determining the gender?

1.7 RESEARCH METHODOLOGY AND TOOLS

To carry out this research the following steps have guided the researcher:

1. Collecting the input images (the dataset which will be used)
2. Reprocessing of the dataset images using machine learning
- Face detections.
- Grey scale image
- Histogram equalization.
- Normalization.
- Convert all grey image to RBG when applied deep learning
3. Extract the feature using LBP local feature extraction technique in machine learning
4. Extract the feature in deep learning using reset_50 and reset_101
5. The last step is to input the vector of feature to the classifier, I will start with SVM in machine learning in the two cases (whole face and component)

6. In deep learning was applied SVM and soft max method in the two cases (whole face and component) to classify the gender.
7. Implementation of the two algorithms.
8. Evaluation of the overall solution.

1.8 RESEARCH ORGANIZATION

The organization of the following chapters in this research is as follows:

Chapter1: Presents introduction, problem statement, objective and motivation.

Chapter 2: presents the related literature review to the research topic, gives more details about the research problem, and critically investigates the existing solutions, which were proposed to address the research problem.

Chapter 3: presents the methods and techniques used in gender identification and the three main steps for gender identification (face recognition, feature extraction and classification)

Chapter 4: presents deep learning for visual understanding and the types of deep learning techniques

Chapter 5: explains Local Binary Patterns and CNN Algorithms**.**

Chapter 6: presents the research results, discussion about the steps used to reach results.

Chapter 7: presents the conclusion of the research besides the potential future works.

CHAPTER TWO
LITERATURE REVIEW

2.1 BACKGROUND

Face processing has been considered for long time as an essential module for many computer vision applications. The most interesting field of research in this area including face recognition are all depend on the classification of the gender and age of face objects. A person could be identified in order to get access to private facilities or to show aimed information in advertising according to individual demographic categories in public places through face analysis component (Mousa Pasandi, Mohammad Esmaeel, 2014).

The analysis of face images presents an efficient function in computer vision. It has been effectively applied in wide rang areas such as biometric, which is uses for robot interaction with human. Computer vision encounter the most challenging problem of gender classification. Pattern classification and feature extraction are the two main components of gender classification (Nazir, et al., 2010). However, solving this problem involves several techniques depending on facial images.

2.2 PREPROCESSING

The results of the systems or classifiers are affected by variation from illumination, poses in accuracies. The sensitivity can be reduced by performing some pre-processing. Some basic Pre-processing steps are involved (Ekman, et al., 1993), including

- Detection of facial portion (face recognition) and utilizing any algorithm detection to remove background.
- Image re-sizing by adjust the image size.
- Normalization of brightness using Histogram equalization function.
- Analysis (detect) the Component for Facial images.

2.3 FACE RECOGNITION

Face detection is a computer technology that utilized in different applications for human face identification in digital images. In addition, it also refers to the psychological process by which humans locate and attend to faces in a visual scene. There are many ways of detecting a face in a scene, which include both the simple and the difficult ones. Some common face detecting approaches are listed below. (Mousa Pasandi, Mohammad Esmaeel, 2014). There was a review of the existing techniques for face detection from a single image or colour image. The single image detection methods are categorized into four Categories (Liew, et al., 2016) shown in the following.

Knowledge-based methods these rule-based procedures encode human information which forms a distinguishable face. The relationships are usually restricted between facial features by the rules. These processes are mainly designed for face localization. The top-down methods and bottom-up methods are the two approaches in which knowledge based methods can be studied through them.

Feature invariant approaches these algorithms aim of finding architectural features which exist even when the pose, viewpoint, or lighting variation differ, and then use the features when finding faces. These methods are designed mainly for localizing faces.

Template matching methods many face standard patterns are stored which are used to describe the whole face or separated facial features. The relationships developed for detection between the stored patterns and an input image. These methods are functional for both face localization and detection. Template matching researches are classified into researches using deformable templates and researches using predefined templates as the two subdivisions.

Appearance-based methods the contrary to template matching, the models (or templates) are to undergo some training firstly among a group of training images, that expected to attain the facial appearance representative variation. These models are then used for detection. Moreover, these methods are specially designed to be used in face detection. Appearance-based methods have been used by many approaches such as, distribution-based methods, Eigen face, neural networks, sparse network of winnows,

support vector machines, Naive Bayes classifiers, inductive learning, hidden Markov model and information theoretical approach.

Although a lot of techniques are applied for recognizing faces, however, the Eigen face, Elastic Graph Matching, Local Feature Analysis, 3D Morphable Model and Active Appearance Model are the most common techniques that are used. Eigenfaces (Belhumeur, et al., 1997), (Collins, et al., 2010), (Lu, Xiaoguang, 2003) and (Moghaddam, Baback and Pentland, Alexander P, 1994). Effectively represent faces pictures by using Principal Component Analysis (PCA). Eigenfaces have several forms which are used for other face recognition as a base. The Eigenfaces do not use the methods that involve normal human recognition but finding correspondences between faces with the fewest controlled environments achieves reasonable results.

(Lin, et al., 2006) and (Milborrow, et al., 2010) used facial images as a technique for recognising gender, ethnicity and ages. To perform facial recognition this technique combined Gabor filter, Adaboost learning and SVM classification. Facial feature extraction is performed through the uses of Gabor filters and Adaboost learning and after features extraction, SVM classifiers are applied for facial recognition of soft feature biometrics. These are physical, behavioural technique or adhered human characteristics. This technique had achieved high accurate results and good performance which was due to the fact that the application of the pre-processing step, the Gabor Filter, improved the performance further.

On the other hand, (Zhang, et al., 2015) in their experiments used some techniques that tried to improve facial recognition by using soft biometrics of the facial region. It was known that soft biometrics can only be treated with the use of an algorithm but not independently; in many cases soft biometrics are considered as an extra component. The algorithm of SVM and Adaboost were used to encode the soft biometrics in order to do facial recognition. It was found that the performance of these experiments increased by 11%, with few soft biometrics' limits.

Another technique was discussed by (Ahonen, Timo and Hadid, Abdenour and Pietik, 2004), (Chen, et al., 2013) and (Klare, et al., 2012) to perform facial recognition, and it was named as Local Binary Patterns (LBP) which both shape and textural information

are taken into account to represent facia l image. This model is free from common error as it gained a 98% accuracy for facial recognition. This technique is achieved by dividing a grey scale image into smaller areas then applying an LBP 3 x 3 mask to the area. Then the comparison of the middle pixel value and the pixel values around it; if the pixel value is lower than the middle pixel, then the original value is changed to 0; if the pixel value is higher, then the original value is changed to 1. Then, reading from left to right, a vector is obtained with only 1's and 0's and the binary values obtained are extracted onto a histogram. Each smaller area is then plotted onto a histogram with its binary value and all these are then concatenated together. The image recognition is carried out by using Nearest Neighbour Classifier and dissimilarity measure of Chi Squared applied to the concatenated histogram value.

2.4 COMPONENT ANALYSIS FOR FACIAL RECOGNITION

The component-based facial recognition approach entails the extraction of parts of the face in order to perform facial recognition. This approach is insensitive to image variations, like face rotations. Automatic extraction and validation of the facial components are the main problem facing this approach. This approach is considered to be robust when it's performed without human interaction.

(Heiselet, et al., 2001) And (Huang, et al., 2002) proposed a method to detect and recognise facial components automatically. They started by moving an object window of a certain size over the input image. Thereafter, there is a selection of 14 points of reference in the object window according to their 3D correspondences from a morphable model. The algorithm then `draws small rectangles around the reference points which have been selected. The facial component detection is performed by looking for the maximum output of the smaller rectangular area, in which linear Support Vector Machine (SVM) is used to classify every component. There with position for the coordinates of the position of the maximum output of each component classifier. In order to extract the feature vectors, the Haar transform is applied on the frontal faces accurately.

Local Feature (LF) analysis is also a considerable technique (Han, et al., 2013), (Lu, Xiaoguang, 2003) which is used for facial biometric technologies, such as eye analysis.

These facial biometric technologies allow changes in facial expressions and aging. Local Feature analysis extracts a group of geometric metrics and distances by using facial images. The features used are eyes, eye browns, nose, jaw line, mouth and cheeks. The performance of this technique varies due to environmental differences and the resolution of images. Elastic Graph Matching Local Features (Jafri, Rabia and Arabnia, Hamid R, 2009), (Lu, Xiaoguang, 2003) are placed on certain locations on the face. The distances between these points are calculated and the higher weights go to the more important points. Elastic Graph Matching (Lu, Xiaoguang, 2003) is constant to affine transformations and localized changes in facial expressions, after which the same images at separate angles are then matched to the same point on the image. Each node is then put onto a graph which is called Elastic Bunch Graph Matching. This technique improves recognition and therefore, is considered to be more effective in finding the differences in posture and facial expression.

(Atharifard, Ali and Ghofrani, Sedigheh, 2011) and (Ng, et al., 2012), (Phillips, et al., 1998) proposed an efficient component face detection algorithm that depends on coloured features. This algorithm had great benefits as it reduced computational time. It was discovered while applying component detectors on facial images that it took very few minutes to detect faces. Extra pixels on the image were all removed including non-skin regions, were all removed so that component detectors would be applied includes some areas that are not skin. The eyes and mouth as the facial components were detected through the use of two models which were colour space that extracted the RBG colour from the image and skin colour segmentation, using a Gaussian model in order to extract skin colour. It was presented that these two models performed well even if the eyes and mouths were obstructed or not detected. This proves that these models are highly accurate than other techniques as they have the ability to localise missed or obstructed facial components.

The common disadvantage of this approach is that it needs human intervention to detect and extract the components; this is done by using geometric consideration and assumption the location on the facial components. This approach is only performed on frontal view images. It was observed that the approach consumes high computational time when using a large number of techniques to discover the location of facial components.

2.5 FEATURE EXTRACTION

Extraction of facial features is one of the most difficult steps in the face identification. However, extracting of several facial features is an important sub-task of gender classification. There are two categories of gender classification approaches based on the features extract: Global facial features and local facial features (Mohamed, Salma and Nour, Nahla and Viriri, Serestina, 2018). Feature extraction quality contribute to increase facial recognition performance. This is the main purpose of various techniques proposed.

Local feature extraction is our focus in this research; face representation concerning the local feature is the classical approach to modelling face processing. Some features like angle, distances, and facial part relationships are usually extracted. Despite invariability of geometrical representation of faces to rotation, scale and tilt, this approach is sensitive to changes in facial expressions and lighting. Specifying a sufficient minimum set of features for face representation is the most considerable challenge of face geometric features representation. It is observed that the textural characteristics of the face are ignored when using geometric features, which represents an essential prompt of gender classification (Jain, et al., 2005).

(Sun, et al., 2002) used genetic features to present gender classification from frontal face images. They applied the Principal Component Analysis (PCA) for representing every individual image as a feature vector in a low dimensional space and regarding genetic algorithms. There are some eigen vectors that are not needed for classification facial feature information were ignored in order to extract some subsets of features from the low dimensional representation. Subset selection of genetic algorithm feature was used for comparing the utilized classifiers like neural network, Bayes, SVM, and LDA.

(Jain, et al., 2005) used frontal facial images to address the gender classification problem, then they used the FERET facial database (250 males and 250 female) to develop gender classifiers with higher performance compared to the preceding gender classifiers. Independent Component Analysis (ICA) and different classifiers were used to carry out the experiment. The results of the experiment show a 96% accurate rate

achieved using the Support Vector Machine (SVM) in International Component Analysis (ICA) space.

(Lawgali, et al., 2015) presented a technique to attain the discriminating feature for handwritten digit, based on DCT and DWT. Experiments were carried out to prove this technique using ADBase database which contains 70,000 digits written by 700 different writers and categorized into two, training 60,000 digits and testing with 10,000 digits. All the normalized digit image are disintegrated into one level by Haar wavelet. In order to extract DCT coefficient, DCT is applied on the low-frequency sub-band (LL) of DWT image. These coefficients were fed to the ANN in the classification stage. It is observed from the result the accuracy is 97.25% for DCT and 98.32% for DWT respectively.

(Nazir, et al., 2010) Introduced an effective technique for gender classification, using the frontal facial images database of Stanford University Medical Students (SUMS). He used Viola and Jones face detection technique to segment the face part of the image. The illumination effect is normalized by performing Histogram equalization. Discrete Cosine Transform (DCT) is then K-nearest neighbour classifier (KNN) is used for classification. The experimental results indicate that the proposed approach achieves as high as 99.3% gender classification accuracy. The performance of the proposed method is tested with several selected training for set size testing and the results achieved are higher compared to other methods.

(Sinha, Chandrakamal, 2013) They used Precise Patch Histogram (PPH) which was extracted from the Active Appearance Model (AAM) standard points for gender evaluation. Statistics that the non-parametric (AAM) was used to describe the training image attributes and a patch library was constructed in the training phase. The results obviously prove that PPH improves accuracy of the global facial features.

(Rai, Preeti and Khanna, Pritee, 2014) Used Gabor features based (2D) 2PCA recognition to propose gender classification system robust to occlusion and introduces a gender classification system that works for non-occluded face images to face images occluded up to 60%. Local information of the face, generating by categorizing the face image into M _ N sub-images. Features are subsequently calculated for every sub-image

by using (2D) PCA on each invariant illumination, real Gabor space was generated using Gabor filter and Support Vector Machine (SVM) was used for classification. The system experiment delivers an accuracy of over 90%. It also overcomes the higher occlusion conditions by providing a least of 86.8% accuracy. Nevertheless, while keeping small sized feature vector, the system most importantly further in increasing the accuracies. (Du, et al., 2014) used another technique which is the multi-level fusion scheme for ethnicity identification. Local Binary Patterns (LBP) and HSV binning were the two techniques that were fused in order to extract the lower level features. HSV binning is applied after the cheek area is cropped from the colour facial image that is inputted. This involves the extraction of a histogram of the image in which the hue, saturation and value are obtained and concatenated into multiple bins. The image is converted to grayscale which is obtained by applying LBP, which is a 3 x 3 pixel window that is applied to each pixel in the facial image, once HSV binning is applied. The concatenation of colour and texture feature is the division of the LBP image into 4x2 blocks in which the histogram feature vector is extracted. The feature vector needs to have a normalisation technique applied, due to the uneven mixture of colour and texture feature; this was the division by range method. Once the feature vector is normalised, classification by the K-nearest neighbour and Support Vector Machine is applied. The ethnicity classes of European, Oriental and African were classified. The multi-level fusion scheme achieved above average results for ethnicity identification.

(Kumar, et al., 2009) developed the technique of using a large group of low level features, where the low level features were used to train the SMV classifier with RBF kernels in order to obtain facial features. Some low level features from original datasets that are related more to facial features were applied as the input for the SVM classifier. This technique which uses low level features recorded reasonable facial recognition rates of 31.68%.

2.6 CLASSIFICATION

Classification aims is to classify data (patterns) into many classes or categories. The target variable is finding the constancy in the input. Artificial Neural Network (ANN), Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Decision Trees,

Bayesian Networks (BNs) and the Hidden Markov Model (HMM) are the common supervised machined-learning techniques (Liew, et al., 2016)

(Tariq, et al., 2009) used silhouetted face profiles with computer vision techniques, in attempt to conduct gender and ethnicity. The facial images used were between the ages of 18 and 30 years for both male and female excluding. The facial images of males who had beards and moustaches. The shape context was obtained for each facial image that was used. The shape context is a distinct group of sample identifies from both the external and internal outlines of the object. After obtaining the shape context, the shape distance is calculated, which uses a weighted sum between the points chosen and the measure of bent energy. After the calculation of all points chosen on the test profile, K-nearest Neighbour is applied to the points in order to carry out the classifications.

A combining system of the cascaded face detector system that was presented earlier by Viola and Jones with discrete Adaboost-based for ethnicity and gender classification was later developed by (Yang, Ming-Hsuan and Moghaddam, Baback, 2000). This system enables several pre-processing and calculations which must be done to be shared by both the gender classifier and the face detector. (Ravi, S and Wilson, S, 2010) Presents an unprecedented face detection and gender classification strategy in colour images with deviating background. The RGB image is converted into the YCbCr colour space by the proposed algorithm for skin region detection in the face image then converted into grayscale image for facial feature detection. The study provides Support Vector Machine (SVM) used for classification. The evaluated threshold 0.07 identifies the facial gender containing in the given input image is a Male or a Female. Linear Support Vector Machine achieved the best classification rate.

(Tolba, Ahmad S, 2001) Applied the use of different neural network classifiers: a learning vector quantization (LVQ) network and a radial basis function (RBF) network. The results of gender identification show a high accuracy of 100% and 98.04% in the case of a LVQ network and an RBF network respectively. With exclusion of hair information, the LVQ classifier achieved the highest result of 95.1% correct identification. It was proved based on studies that the LVQ model is not as fast as the RBF model in learning the task. The proposed system is much faster (106 MS /face for

RBF network and 326 MS/face for LVQ) and achieves the best recognition rates (98.04% and 100% for RBF and LVQ) compared to other recent systems.

An appearance-based method for facial image gender identification was proposed by (Moghaddam, Baback and Yang, Ming-Hsuan, 2002) they used nonlinear SVM classifier and compared the results with traditional classifiers and modern techniques such as Radial Basis Function (RBF) networks and large ensemble-RBF classifiers, during the performance the comparison done with low-resolution and higher resolution images when classifying.

(kinen, Erno and Raisamo, Roope, 2008) Studied the advanced methods of gender classification comparably with the aim of finding out reliable and exact methods using a WWW and FERET image database. The study possessed comparable and all-inclusive classification results of gender classification methods combined with automatic real-time face detection with manual face normalization. Also resulted to increase classification precision and instructions to perform classification experiments in addition to the combination of gender classifier outputs which increases classification accuracies arithmetically. Then instruction to perform classification experiments, knowledge on the strengths and weaknesses of the gender classification methods. However, the study showed that a high classification rate was uncertain with the inclusion of hair was in the face images.

(Khalil-Hani, Mohamed and Sung, Liew Shan, 2014) Suggested an approach where they used a convolutional neural network (CNN) for actual time gender classification regarding facial images. SUMS and AT&T were used as databases. The proposed CNN architecture reduced the number of processing layers in the CNN to just four layers which led to a reduction of the design complication. A second-order back propagation learning algorithm with annealed global learning rates was used to train the network, it revealed accuracies of 98.75% and 99.38% for SUMS and AT&T respectively. The neural network has an the ability to processing and classifying a 32 * 32 pixel face image in less than 0.27 MS, which is similar to a very high output of over 3700 images per second. Training converges within less than 20 epochs. These results similar to the highest performance in classification, it was confirmed that the proposed CNN is an efficient real-time to overcome difficulties of gender recognition.

(Akanchha, Gour and others, 2016) Performed gender and age discovery through fingerprint. Discrete Cosine Transforms (DCT) and Discrete Wavelet Transforms (DWT) Coefficients were utilized for extraction fingerprint image features and, and classified by using KNN. The dataset consists of 100 fingerprint images of various age including both male and female fingerprints. The result reveals an accurate rate of 90% for age and gender.

(Qacimy, et al., 2014) worked on the evaluation of the feature extraction efficiency of four variants of the discrete cosine transform (DCT upper left corner (ULC) coefficients, DCT zigzag coefficients, block based DCT ULC coefficients and block based DCT zigzag coefficients) to capture features in order to attain greater classification accuracy for recognition of handwritten digit . The reference database MNIST and a modified MNIST database were used to assess the method. After that they used a pre-processing step that removes the non-information bearing areas. An SVM classifier was then used to assess features sets generally in relation to reduction and accuracy rate. The accuracies obtained from the results are (98.66%, 98.71%, 98.73%, and 98.76%) for DCT upper left corner (ULC) coefficients, DCT zigzag coefficients, block based DCT ULC coefficients and block based DCT zigzag coefficients respectively. The study shows that, the highest accuracy belongs to block based DCT zigzag coefficients.

(Ozbudak, et al., 2010) used three levels Discrete Wavelet Transform (DWT) to decompose face images and used Principal Component Analysis (PCA) to reduce the dimensions and Fisher Linear Discriminate (FLD) for gender recognition to achieve an accuracy of 93%.

(Nagi, et al., 2008) Performed an extraction of useful features out of Cambridge ORL face database by using hybrid feature extraction algorithms Discrete Wavelet Transform (DWT) and Discrete Cosine Transform (DCT). They applied level 2 Harr Wavelet decomposition in their work, then they used DCT on LL2 image sub-band. The Support Vector Machine (SVM) was used for facial recognition. The results revealed that the classifications contain 98.90% accuracy of DWT-DCT-SVM for face recognition.

(Sun, et al., 2006) Carried out an extraction of LBP histograms from local facial regions, using Adaboost method to classify face images taken in constrained environment and achieved an accuracy of 95.75%.

(Cui, et al., 2013) Studied a discriminative LBP-Histogram (LBPH) bins for classifying gender. The selected LBPH bins delivered a compact facial representation and attained accuracy of 94.81% on LFW database.

Six facial components (hair, forehead, nose, mouth, eyes and clothing) were used by (Li, et al., 2012) to determine gender from occluded faces. The experiments were performed on FERET and BCMI (self-collected) face databases. The overall accuracy was above 95% for non-occluded face while the occluded face reached 90% using a fivefold cross validation method.

(Rai, Preeti and Khanna, Pritee, 2014) Proposed the use of facial components in position of full face in a multi view gender classification system. They achieved an accuracy of (over 95% and almost 89%) for non-occluded faces and artificially occluded faces respectively using CAS-PEAL database.

2.7 INTRODUCTION TO DEEP LEARNING (DL) IN NEURAL NETWORKS (NNS)

Single-task learning is welcomed and broadly applied in machine learning for image processing (Sun, et al., 2013), (Heisele, Bernd and Blanz, Volker, 2006). Nevertheless, the effectiveness of this learning is only focusing on main purpose information, and less concerned with other connected information. That would not be easy when classifying complex objects with different sizes, shapes, orientations and outlines (Heiselet, et al., 2001) in reality, like face detection and object recognition. The 3D information (colour and depth) is a suggested method to ease complex object when classifying by adding distance to make the object of interest stereo. In addition, depth data guides to detect faces or object recognition, particularly in cases where images are rotated, overlapped, exposed to illumination condition or even deformed by noise. Hopefully, combining the depth information and 2D texture images is the hopeful method to improve recognition accuracies (Huang, et al., 2002).Surface based methods mostly use the 3D information in face recognition. (Wu, et al., 2013) Use calculation of the distance between curves in

the same level, which are classified by HMM (Hidden Markov Models) in representing every point in face with its comparable facial level curve. Colombo, A., Cusano C. and Schettini R. used curvature analysis as a developed method in face detection. (Chang, et al., 2005) and (Chen, et al., 2017) compared a PCA in face detection on their dataset. Accuracies of 51.74% and 58.25% were obtained respectively. Similarly, methods used in 3D object recognition (Berretti, et al., 2013), (Lei, et al., 2014) are mostly regard to crafted features, which demands high analysis of the object of interest. Additionally, the extracted features are disputable, as a result of their limitation to different knowledge background. Deep learning algorithms can help in reducing such limitations. Deep learning methods use raw data only to learn hierarchical features to improve performance in face recognition (Buchala, et al., 2005), facial key point (Gutta, Srinivas and Wechsler, Harry, 1996) and object detection (Davis, Jesse and Goadrich, Mark, 2006) However, up to the date there is no much systematic study received in deep learning methods in face recognition that base on both depth and 2D images. Moreover, the suggested deep learning methods in image recognition are mostly based on single-task, that not really effective for complex objects. Multi-task learning is a multi-structure model, which involve both single and secondary task. Single-task concentrates on training using the information of the main application, while secondary-task learns features from relative information, this multi-task is somehow connected to the main purpose of this study. If we aim of detecting face through the use of multi-task model, relative information for example may be milestone on the face in the process. If we combine the features learned from relative information and main information, we could be able to achieve the main application by enhancing the rate accuracy. A multi-task model has been applied to the neural networks for classification (Caruana, Rich, 2012). However, the features extracted are not hierarchical as the network is shallow.

Deep learning is a subset of machine learning which strives to learn high-level abstractions in data by using hierarchical structures. It is a newly created approach and has been broadly put into use in traditional artificial intelligence areas like semantic parsing (Bordes, et al., 2012), transfer learning (Ren, Jimmy SJ and Xu, Li, 2015) natural language processing (Mikolov, et al., 2013) computer vision (Ciregan, et al., 2012), (Krizhevsky, et al., 2012) and many more. Deep learning today is booming for three important reasons: the significant lowered cost of computing hardware, the

dramatic increase in abilities of chip processing (e.g., GPU units) and the considerable progress in the machine learning algorithms (Bengio, Yoshua, 2009). Recently many deep learning approaches have been reviewed and analysed extensively (Can, et al., 2019), (Bengio, Yoshua, 2013) and (Deng, Li, 2014) is one of the reviewers who highlighted the main inspirations and technical contributions of deep learning in a historical timeline pattern, while (Schmidhuber, Jurgen, 2015) scrutinized the challenges facing researchers in deep learning and proposed some visionary research directions. Deep networks can make appropriate extraction while connectively performing distinction and these networks have proven their great achievement in computer vision tasks (Bengio, et al., 2013). Different researchers have embraced deep learning methods in the last ImageNet Large Scale Visual Recognition Challenge (ILSVRC) competitions (Russakovsky, et al., 2015) and achieved highest accuracy scores. The questions fundamental credit assignment problem are as follow; Are the modifiable components of a learning system responsible for failure or success? What changes improve their performance (Minsky, M, 1963) .In order to solve this universal problem, there are general credit assignment methods which are highly efficient in various theoretical senses. However, the recent survey, will focus on the narrower, but now commercially significant to subarea of Deep Learning (DL) in Artificial Neural Networks (NNs). There are many simple connected processors in a standard neural network (NN) named as neurons, every neuron produces a sequence of real number activations. Input neurons derive activation amidst sensors observing the environment; while others derived their activation through weighted connections from preceding functional neurons. The environment may get influenced by some neurons by actuating actions. Learning or credit assignment is responsible for finding weights which enable the NN exhibit desired behaviour like driving a car. Such behaviour may demand long normal chains of computational steps, based on the challenge and how the neuron are related in which each step changes (often in a non-linear way) the network collective activation. Deep Learning is to assign credit perfectly many such steps. Shallow NN-like models with few such stages have been known for long years ago. From about 1960s and 1970s Models have achieved many successive nonlinear layers of neurons. A sufficient classified descent system for teacher-based Supervised Learning (SL) in discrete, distinguishable networks of arbitrary depth called back propagation (BP) was designed in the 1960s and 1970s, and was later used in NNs in 1981. Despite that BP

based on training of deep NNs with many layers, however, in the late 1980s it was happened to be practically difficult and became a definitive topic of research by the early 1990s. DL through the help of Unsupervised Learning (UL) became practically capable to some extent, e.g. Within the 1990s and 2000s there were several obvious progresses for purely supervised DL. Finally, from 2000s till date, deep NNs have fascinated researcher's concentration, particularly by outperforming alternative machine learning methods like kernel machines in various significant applications. As matter of fact, since 2009, supervised deep NNs have participated in several official international pattern recognition competitions and won, they got the first superhuman visual pattern recognition outcomes in restricted areas (Burges, et al., 1999), (Vapnik, Valdimir N, 1995). Deep NNs are now relevant for the more general field of Reinforcement Learning (RL) without any overseeing teacher. Both feed forward (acyclic) NNs (FNNs) and recurring (cyclic) NNs (RNNs) were succeeded in several competition. In some ways, RNNs are considered as the deepest among all NNs their general computers stronger compared to FNNs, and have got the ability to develop and process memories of arbitrary orders of input patterns in principle (Siegelmann, Hava T and Sontag, Eduardo D, 1991). Contrarily to classical methods for automatic consecutive program combination (Balzer, Robert, 1985) (Deville, Yves and Lau, Kung-Kiu, 1994) (Soloway, Elliot, 1986). RNNs are able to learn combined sequential programs and compare information processing naturally and efficiently, utilizing the weighty parallelism seen as extremely important to sustain the rapid declination of computation cost that have been observed for the last 75 years.

2.8 COMPONENT ANALYSIS FOR FACIAL RECOGNITION

In order to perform facial recognition, the component-based facial recognition approach entails the extraction of parts of the face. This approach is insensitive to image variations, like facial rotation. The automatic extraction and validation of the facial component are the expected problems with this approach; this approach needs to be robust to do these without any human interaction. (Buchala, et al., 2005) (Toderici, et al., 2010) proposed a method to detect and recognise facial components automatically. Firstly an object window of a certain size is slid over the input image. There after the

selection of 14 points of reference in the object window according to their 3D correspondences from a morphable model. Then small rectangles are drawn around the selected reference points by the algorithm. The facial detection components are performed by finding the highest output of the smaller rectangular area using linear Support Vector Machine (SVM) for classifying each component. Then the recorded position coordinates of the highest output of each component classifier is taken with the position. Lastly, the application of Haar transforms on the frontal faces for the feature vector attaining.

This approach needs human intervention in detecting and extracting the components which are considered as the drawback of this approach, the component detection and extraction are done by using geometric consideration and assumption on the location on the facial components. This approach is performed on frontal view images only. It was seen that it consumes high computational time when using a large number of techniques to ascertain the location of facial components.

2.9 DRAWBACKS OF CURRENT METHODS AND CONTRIBUTIONS

(Gutta, Srinivas and Wechsler, Harry, 1996) Presented that a person's face will exhibit various changes, such as the face smiling, the face frowning and the eyes being closed. These affect the way the computer vision system recognizes gender, ethnicity and age in the image. The accuracy of facial recognition is affected by accessories worn by the subject, such as hat, scarf, eye glasses or sun glasses. Illumination, lighting and image quality (like blurring, noise and resolution) change the subject and affect recognition. The accuracy of recognition also get affected in images where the orientation of the subject's head changes. Despite all the encouraging performances deep learning has successfully carried out, the research literature has identified different crucial challenges likewise the essential tendencies, that are defined here as theoretical understanding, vision equivalent to human, time complicacy , Training with restricted data, stronger models . Many studies (Le, et al., 2011), (Colombo, et al., 2006) have proved that when classifying a person's gender, the whole face is used.

2.10 CONCLUSION

A comprehensive description of ethnicity and facial recognition is included in this chapter. We reviewed information on component-based extraction techniques which was discussed. We presented the pathways of our research with the identification of the drawbacks in current methods and contributions. In the next chapter, the algorithms and feature extraction techniques for ethnicity identification are defined and explained.

CHAPTER THREE
METHOD AND TECHNIQUE

3.1 INTRODUCTION

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

Image representation can take many forms in computer science. It mostly refers to show how the transmittable information like colour, is digitally coded and how the image is stored, i.e., the way it is structured an image file. There are several open or individualized benchmarks to build, manipulate, store and exchange digital images. They describe arrangement of image files, algorithms of image encoding and format of further information; all this is named metadata. Separately, the visual content of the image can also participate in its representation. New approaches standards of representation have been provided by this recent concept and were all gathered into the discipline named *content-based image indexing* (Pianykh, Oleg S, 2009).

Face recognition tasks are divided into two: face identification and face verification. The identification methods depend on known faces in finding the identities of unknown faces, while the verification methods verify or disapprove two sameness faces. Known faces are set in a dataset named as gallery set while probe set is for unknown faces. There are several important stages for the general processing pipeline of the face recognition system, these stages are described as follows:

3.2 FACE DETECTION

Face detection is the first stage of facial recognition, which is concerned with finding the facial location in video frame or image and passing it to the next step. Viola and Jones is facial detection technique that searches the facial part is used to detect the region of interest (Du, et al., 2014). Image pixel values are converted to grey scale and the histogram equalization with adaptive parameters is used to improve the contrast. Here, is the detailed explanation: if we let ***f*** be a certain image represented as ***mr*mc***

matrix of integer pixel intensities ranging from 0 to $L - 1$, in which L is the number of expected intensity values, often 256. Let p signify the normalized histogram of f with a bin for all expected intensities (Du, et al., 2014).

3.2.1 Detecting facial components

To carry out feature extraction of either the individual components or the face as a whole, the face should first be detected. The Viola Jones algorithm (Viola, Paul and Jones, Michael, 2001) has three main modules which make it efficient which are the integral image which efficiently computes the sum of values in a rectangle subset, the adaboost learning which combines many weak hypotheses and the attentional cascade. Of great importance when looping through the datasets is the use of the globe function as shown below which an extract of looping over the facial images is, converting the image to grayscale among others taking in a vector of file-names and a path to the dataset. Once the face has been detected, we start by detecting the eyes, nose, mouth using Viola and Jones (Viola, Paul and Jones, Michael, 2001).The forehead, cheeks are detected using face geometrics. The pseudo code for detecting facial features as shown in fallowing algorithm.

```
Algorithm : Detecting Facial Components
function Detect Facial features;
Input : faces, cascades (eyes, nose ,mouth), labels
Output: detected(eyes, nose, mouth, forehead, cheeks)
for each image i in dataset do
    set paths to dataset and cascades;
    if file path or face cascade path are empty then
        print error message;
    convert image to gray scale;
    detectface ;
    draw bounding box on face(ROI);
    save ROI;
    for each face detected do
        if eye cascade is not empty then
            if eyes are being detected in ROI then
                mark center of eyes and draw bounding rectangle;
                crop and save eyes image;
                use eye width and height to approximate cheek location ;
                crop and save cheeks;
        if nose cascade is not empty then
            draw rectangle around nose;
            crop and save nose;
        if mouth cascade is not empty then
            if mouth is in ROI then
                draw bounding box around mouth;
                crop and save the mouth;
        draw forehead bounds using corner center of eyes and half of eye width
          distance;
        crop and save forehead;
    end
end
```

3.2.2 Face normalization

Face normalization is the second stage for facial recognition after face detection. Face normalization module performs preparation for the next stages. It consists of two components: the first one is the geometric normalization element that concerns with the rotation whose roles are rotating and scaling the face to the same position among all images; the second is the photometric normalization element which adjusts illuminating conditions.

3.3 FEATURE 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 (Timo, Ahonen, 2004). 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.

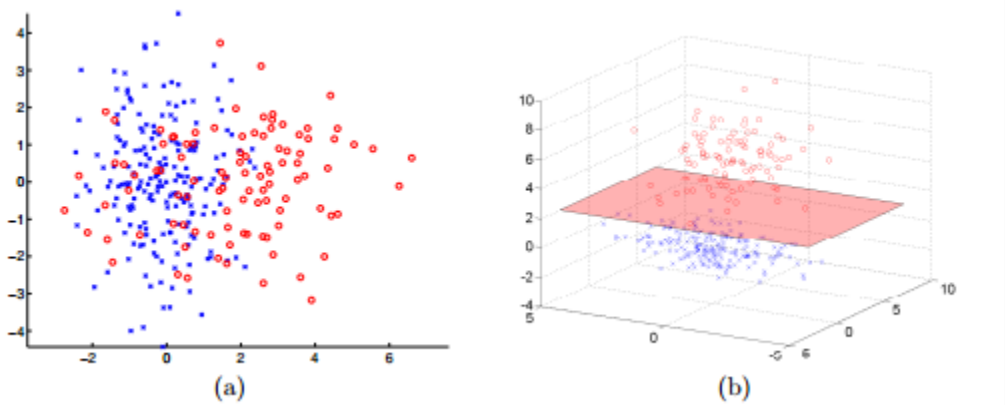


Figure 2: Samples arrangement in the input space and into feature space (Timo, Ahonen, 2004)

Separating samples in input space is quite difficult, while they are linearly separable in feature space. We can approximately divide the features into two categories: low-level features (e.g., LBP (Timo, Ahonen, 2004), (Shen, Linlin and Bai, Li, 2006) and high-level features which are calculation of low-level features. The high-level features are more informative and more strongly formed of variations.

3.3.1 Dimensionality reduction

The extracted features with face recognition method are of high dimensionality which is seen as the main problem of this method. For that reason, techniques that reduce dimensionality are preferable. Thus, the dimensionality reduction is a critical step in many advanced methods (Chen, et al., 2013), (Hussain, et al., 2012), (Cui, et al., 2013), (Tan, Xiaoyang and Triggs, Bill, 2007) and the linear subset of a space projection is considered as one of the widely used techniques.

In this stage, dataset images are converted into matrices of size 32*32. Then, the DCT zigzag is applied (Qacimy, et al., 2014) to the whole image matrix, next the DCT coefficients are sorted based on zigzag scan order and then arranged into vector with size 1*1024. Figure3 illustrates DCT Coefficients in a Zigzag Fashion.

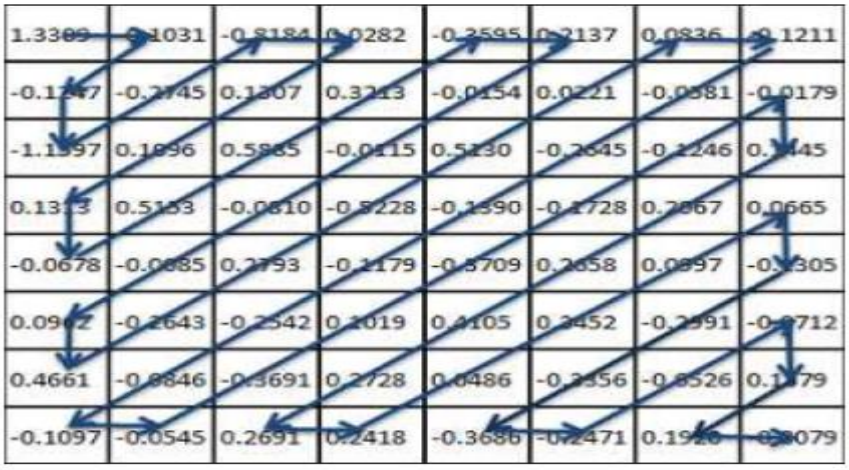


Figure 3: Selecting DCT coefficient in zigzag fashion (Qacimy, et al., 2014)

3.3.2 Block based DCT Zigzag

In this method, the image matrix is broken down into 16 8x8 blocks and after that DCT is applied to every block starting from left to right then from top to bottom. For both methods 16, 32, 128 and 64 DCT coefficients will be selected (Mohamed, Salma and Nour, Nahla and Viriri, Serestina, 2018). Figure4 illustrates block based DCT zigzag

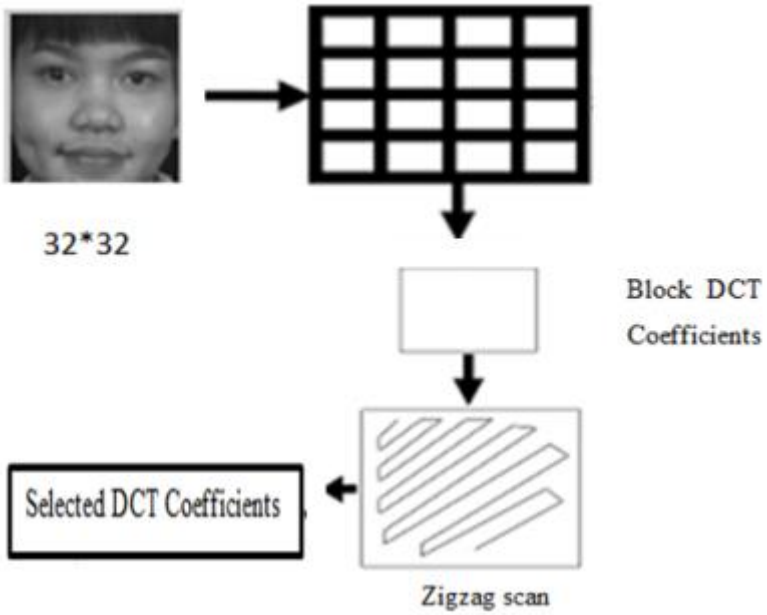


Figure 4: Zigzag DCT based block (Mohamed, Salma and Nour, Nahla and Viriri, Serestina, 2018)

3.3.3 Discrete Wavelet Transform (DWT)

DWT is a technique that is used for image feature extraction. At each stage of decomposition, a low-pass filter (LPF) and a high-pass filter (HPF) are applied to each image column to break it down into one low-frequency sub-band (LL) and three high frequency sub-bands (LH, HL, HH) (Akanchha, Gour and others, 2016). The one level decomposition of DWT is shown in figure 5.

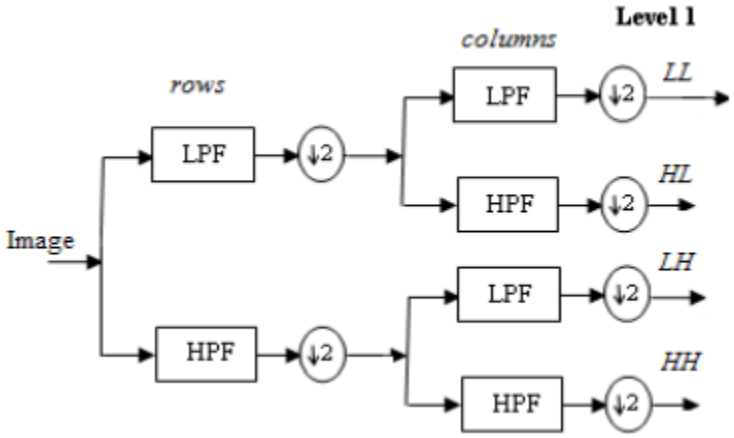


Figure 5: DCT decomposition in one level (Akanchha, Gour and others, 2016)

3.3.4 Hybrid DWT and DCT zigzag

This method follows the same idea of the above-mentioned method. First, there is extraction of the coefficients of the lowest frequency range in sub-bands, using the two level Haar DWT Zigzag (Mohamed, Salma and Nour, Nahla and Viriri, Serestina, 2018). Consequently, there is a selection of the LL sub-band element with the maximum image energy for feature extraction. Second, the DCT is used for the LL2 sub-band. After that, extraction of the higher value DCT coefficients matching to the low frequency is performed in a zigzag feature extraction process and stored in a vector. Figure 6 illustrates the process.

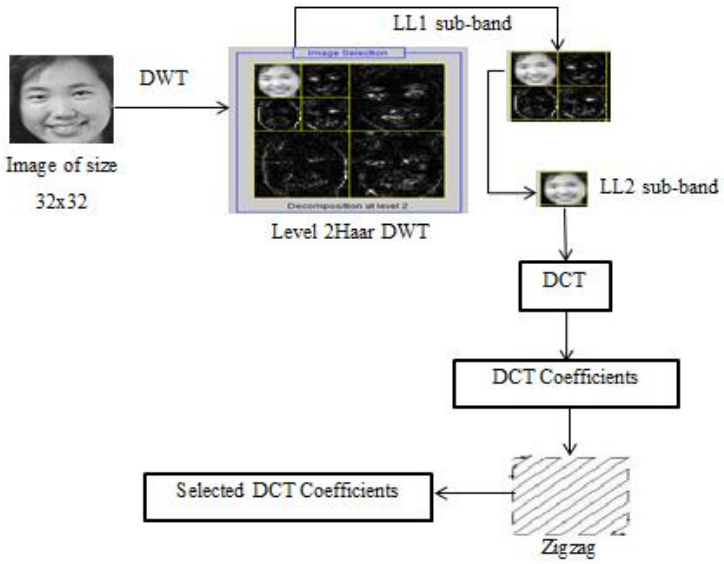


Figure 6: Hybrids DCT and DWT zigzag feature extraction process (NAZIR, ET AL., 2010)

3.3.5 Local Binary Pattern

The Local Binary Pattern (LBP) operator was introduced as a regional descriptor-based approach to describe texture by (Ojala, et al., 1996). (Ahonen, et al., 2004)) later proposed this pattern into facial recognition area.

The LBP generation approach in (Ahonen, et al., 2004) is described as follows. It starts with the application of LBP operator to build LBP map. For each pixel of the image, the operator thresholds is ***3×3*** pixels: for each neighbourhood pixel, we task a binary number1 to the pixel if its grey scale value is higher than the centre pixel value; else, we task a 0 to the pixel. After that, all the binary numbers should be piled together into one vector and considered as centre pixel label. Figure7 shows the diagram of the encoding scheme. The centre pixel is labelled as 01011010 in binary or 90 in decimal. One of the extensions of this basic operator allowing neighbourhoods of arbitrary size and numbers as presented in figure8. We used the notation *LBPP;R* to signify an LBP operator in which *P* points are sampled on a circle of radius *R*. The bilinear interpolation shall be applied to attain its value in the absence of sampling point from the centre of pixel.

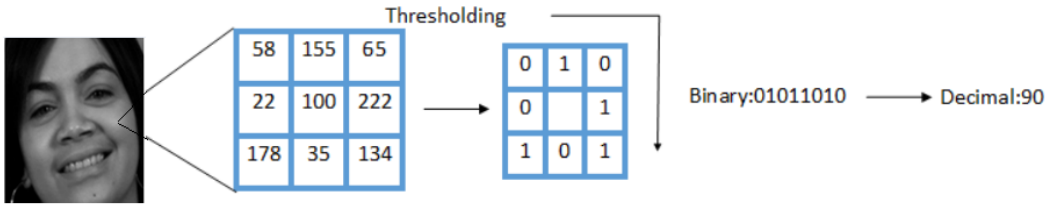


Figure 7: The centre pixel is labelled as 01011010 in binary or 90 in decimal (Ahonen, et al., 2004)

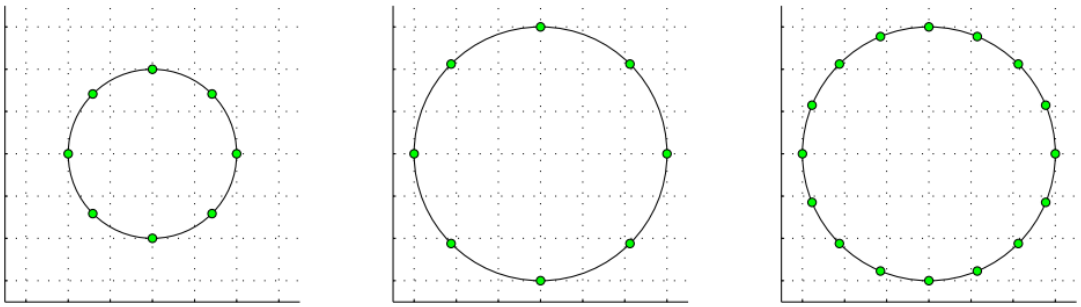


Figure 8: LBP operators which can be denoted as LBP8;2, LBP8;3, LBP16;3 (Ahonen, et al., 2004)

Another notable extension of this basic operator is the uniform pattern that is signified as LBP u2. An LBP label is called uniform when at most two bitwise transitions from 0 to 1 or vice versa. Then considering it as a circular. Some examples are shown in figure 3.7. For the operator of 8 sampling points, there are 58 uniform patterns. According to (Ahonen, et al., 2004), there are around 90% pattern in LBP8; and 1 is uniformed. In this case, LBP labels can be more encoded into 59 numbers: non-uniform pattern get one and the rest for uniform pattern.

The second stage of (Ahonen, et al., 2004) is LBP histogram feature generation. The LBP image is divided into different non-overlapped cells, and the histograms are computed in each cell which is known as histogram function H:

$$Hi = X\ B\ ((LBPP;R\ u2\ (x; y)) = i)\ ji\ 2\ [0; n]$$

Here, various encoded LBP labels are signified, where (x, y) are the circle centre coordinates,

$$B(v)=\begin{cases}1, & when\ v\ is\ true\\ 0, & when\ v\ is\ false\end{cases}$$

Several LBP label numbers at a precise area are counted by this histogram function and then the result is piled into one string. The overall LBP histogram descriptor of this image is the focus of the histograms of all cells. The illustration in figure 9 below shows all the approach. The LBP pseudo code is presented in algorithm below:

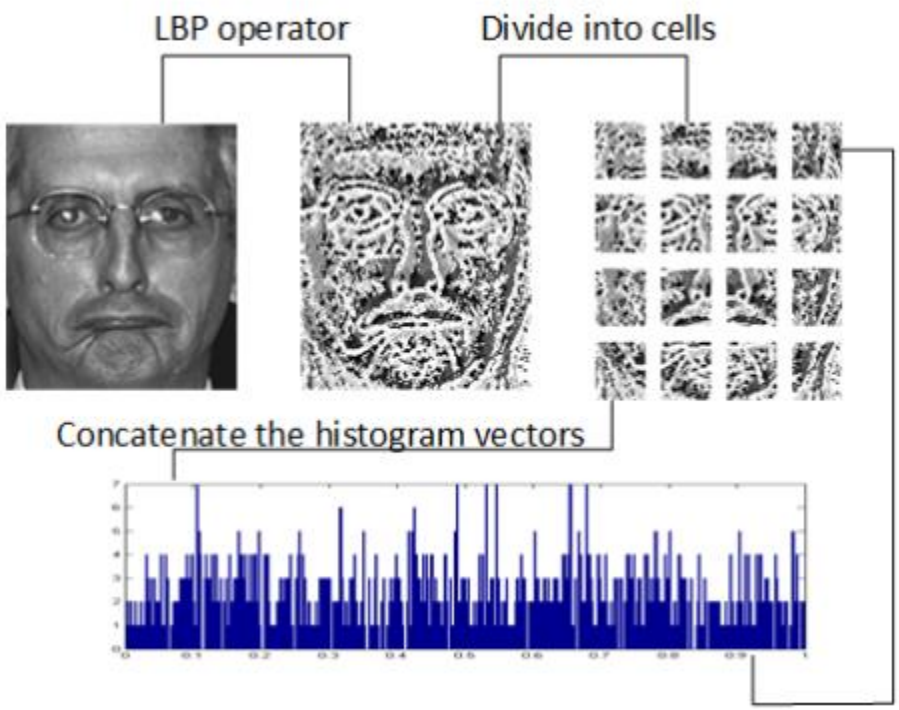


Figure 9: LBP descriptor extraction pipeline (Ahonen, et al., 2004)

Algorithm Original LBP-TOP implementation.

Data: Video data V, where $[H, W, T] = size(V)$. R_X, R_Y and R_T are radiu neighborhood along X, Y and T direction respectively, and P_{XY}, P_{XT} and P_{YT} are numbers of neighborhood points on XY, XT and YT plane respectively.

Result: LBP occurrence histograms on XY, XT and YT plane, namely $HIST_{XY}$, $HIST_{XT}$, and $HIST_{YT}$ respectively.

```
for i = R_T to T − R_T − 1 do
    for j = −R_T to R_T do
        Frame_{H×W×(j+R_T+1)} = V_{H×W×t}, where t = i + j + 1;
    end
    for yc = 1 to H do
        for xc = 1 to W do
            CenterVar = Frame(yc, xc, R_T);
            // Compute LBP on XY plane for center pixel with
               neighborhood (P_XY, R_X, R_Y).
            LBP_CenterVar = LBP_PIXEL(CenterVar, P_XY, R_X, R_Y);
            Update HIST_XY with the value of LBP_CenterVar;
            // Compute LBP on XT plane for center pixel with
               neighborhood (P_XT, R_X, R_T).
            LBP_CenterVar = LBP_PIXEL(CenterVar, P_XY, R_X, R_T);
            Update HIST_XT with the value of LBP_CenterVar;
            // Compute LBP on YT plane for center pixel with
               neighborhood (P_YT, R_Y, R_T).
            LBP_CenterVar = LBP_PIXEL(CenterVar, P_XY, R_Y, R_T);
            Update HIST_YT with the value of LBP_CenterVar;
        end
    end
end
```

3.3.6 Local Directional Pattern (LDP)

Local Directional Pattern (LDP) is an eight-bit binary code which is ascribed to each pixel of an input image. The relative edge response value is compared when calculating LDP of a pixel in different directions. In this sense, we use Kirsch masks (Jabid, et al., 2010) in calculating eight directional edge response value of a specified pixel in eight orientations differently (M0~M7) centred on its own location. Figure10 presents the masks.

$$\begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{bmatrix} \quad \begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 3 \end{bmatrix} \quad \begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix} \quad \begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}$$

$East(M_0)$ $NorthEast(M_1)$ $North(M_2)$ $NorthWest(M_3)$

$$\begin{bmatrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{bmatrix} \quad \begin{bmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{bmatrix} \quad \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ 5 & 5 & 5 \end{bmatrix} \quad \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & 5 \\ -3 & 5 & 5 \end{bmatrix}$$

$West(M_4)$ $ThouthWest(M_5)$ $Thouth(M_6)$ $Thouth\,East(M_7)$

Figure 10: Kirsch edge masks in eight direction (Jabid, et al., 2010)

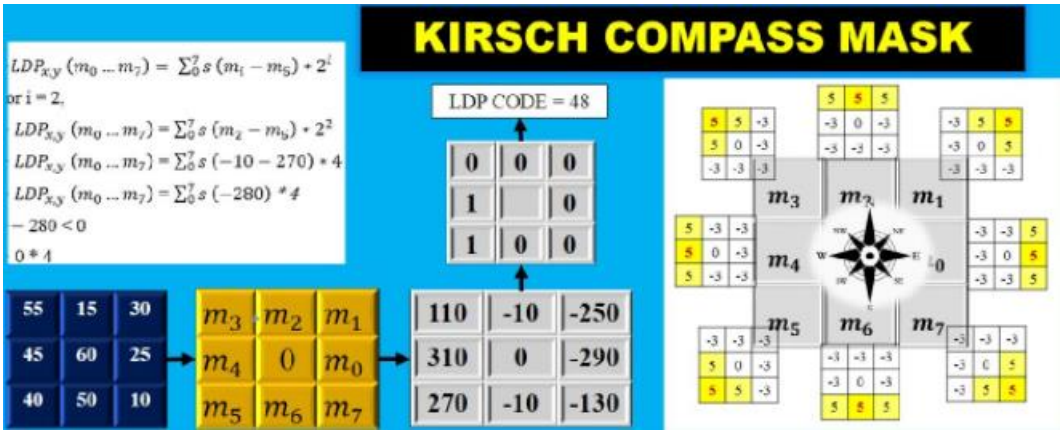


Figure 11: Kirsch compass mask (Jabid, et al., 2010)

To apply eight masks, we attain eight edge response values *m0, m1, …, m7*, each representing the edge significance in its particular direction. The response values of some surpass others, their importance is not equal in all directions. The response values are high when edge or corner is present in particular directions. Our interest is in knowing the ***k*** which most leading and noticeable directions is for the generation of the LDP. Therefore, we look for the highest ***k*** values |***mj***| and set them to 1, and set the other ***8-k*** bit of 8- bit LDP pattern to **0**.

$$C\left[f(x,y)\right] := (c_i = 1) \qquad if\ 0 \le i \le 7\ and\ m_i \ge \psi$$
$$where\ \psi = k^{th}(M); \qquad M = \{m_0, m_1, ..., m_7\} \qquad (2)$$

It was proven that LDP patter has more stability when there is noise. For instance, figure13 displays the real image and the similar image but with the addition of Gaussian white noise. When the noise was added, the 5th bit of LBP is changed from 1 to 0, hence, LBP pattern changed to a non-uniform code from being uniform. LDP pattern produces the same pattern value despite the presence of that noise and varied illumination changes (Jabid, et al., 2010).

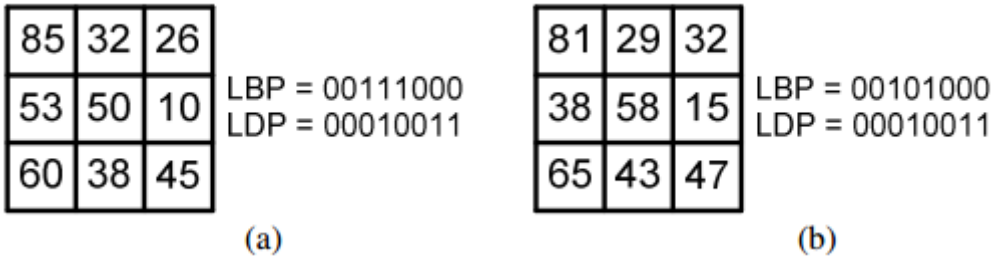


Figure 12 : Image stability using LBP and LDP (JABID, ET AL., 2010)

Histogram of LDP

Once we encoded an image with the LDP operator we get an encoded image *I*. We use k=3 which generates 56 noticeably different values in our encoded image. So histogram *H* of this LDP labelled image *I(x,y)* is a 56 bin histogram and can be defined in equation 3 as follows:

$$H_i = \sum_{x,y} P\left(I_L(x,y) = C_i\right), \quad C_i = i^{th}\ LDP\ pattern\ (0 \le i < 56)$$
$$where\ P(A) = \begin{cases} 1, & if\ A\ is\ tru \\ 0, & if\ A\ is\ false \end{cases} \qquad (3)$$

3.3.6.1 Face Representation using LDP

An LDP histogram represents each face. It contains fine detailed information of an image, such as corner, spot, edges and other local texture features. But the histogram that is computed on the whole face image encodes only the affairs of the micro-patterns without knowing anything about their positions. Facial images are divided into small regions $R_0, R_1, \ldots, R_n$ in order to integrate some degree of locational information and extract the LDP histograms HR_i from each region R_i. These n LDP histograms are concatenated to get a spatially combined LDP histogram which performs in the given face image as a global face feature. Figure14 presents the process (Jabid, et al., 2010).

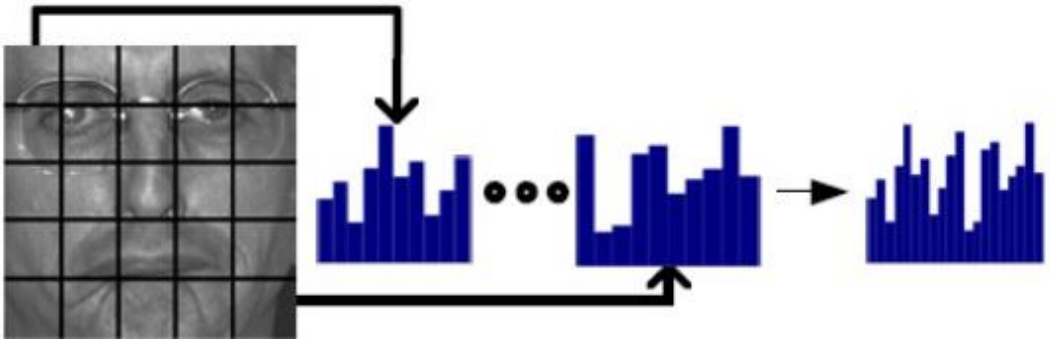


Figure 13: Facial image representation using enhanced histogram (2010 ,.ET AL ,ABID])

3.3.6.2 Face Recognition using LDP

While processing facial recognition, face features are extracted from the face. The aim is the comparison of the LDP encoded feature vector from single person with all other candidate's feature vector with a Chi-Square dissimilarity measure. It is understandable that the candidate that is selected with lowest measured value signify the presence of a match. It is proved from face physiology knowledge that some portions of the face have got more discrimination capacity, like eye, eye-brow, mouth, nose etc. Consequently, a weighted chi-square measure that gives various weight in different face block region is

used. Weighted Chi square dissimilarity measure between two spatially encoded LDP histograms ***SLH1*** and ***SLH2*** is defined in equation 4:

$$\chi_w^2\left(SLH^1, SLH^2\right) = \sum_{i,j} W_i \frac{\left(SLH_{i,j}^1 - SLH_{i,j}^2\right)}{\left(SLH_{i,j}^1 + SLH_{i,j}^2\right)} \qquad (4)$$

where the index ***i*** indicates the region number,

j refers to bin number of that region and

w_i represent the weight of region ***i***.

3.3.7 Local Ternary Pattern (LTP)

LTP is introduced in the literature by researchers (Tan, Xiaoyang and Triggs, Bill, 2007), (Tan, Xiaoyang and Triggs, Bill, 2010) to solve the noise related challenges of LBP. LTP is a ternary or 3-valued code. In LTP, we used a lag limit value '1' to compare the neighbourhood pixel values with a central pixel. According to this comparison, the neighbourhood values will assign one of the three values +1 or 0 or -1, as given in equation (5) (Tan, Xiaoyang and Triggs, Bill, 2010).

$$LPT\left(T_i\right) = \begin{cases} 1 & p_i \geq \left(i_c + l\right) \\ 0 & \left|P_i - i_c\right| < l \\ -1 & p_i \leq \left(i_c - l\right) \end{cases} \qquad (5)$$

where ***p*** represents the grey level intensity values of the neighbouring pixels and the central pixel respectively, ***l*** represents the lag limit value and represents the one of the ternary value assigned to the neighbouring pixel i. Figure15 shows the illustration of LTP encoding with l=5.

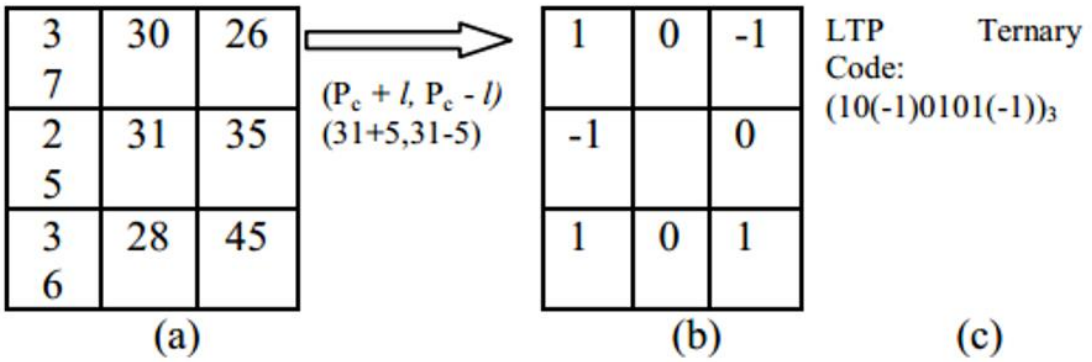


Figure 14: The transition of LTP with i=5 (Tan, Xiaoyang and Triggs, Bill, 2007)

The LTP and ternary representing symbol produce a whole of *0* to *3n-1* valued codes and also additional number of ULBPs were produced by the LTP and it leads to lots of difficulty in the systems involving histograms. To make the complexity simple and easy some of the researchers (Mohamed, Abdallah A and Yampolskiy, Roman V, 2012) divided LTP into two different channels of LBP, named Positive or High LBP (LBP-H) and Negative or Low LBP (LBP-L) as illustrated in figure16 for ternary code [10(-1)0101(-1)]. The process is illustrated in figure16 is more extended in figure17. LTP problems related to complexity and dimensionality are solved by this representation of LBP-H and LBP-L. This phenomenon of using lag limitation value in LTP empowers the defiance to noise.

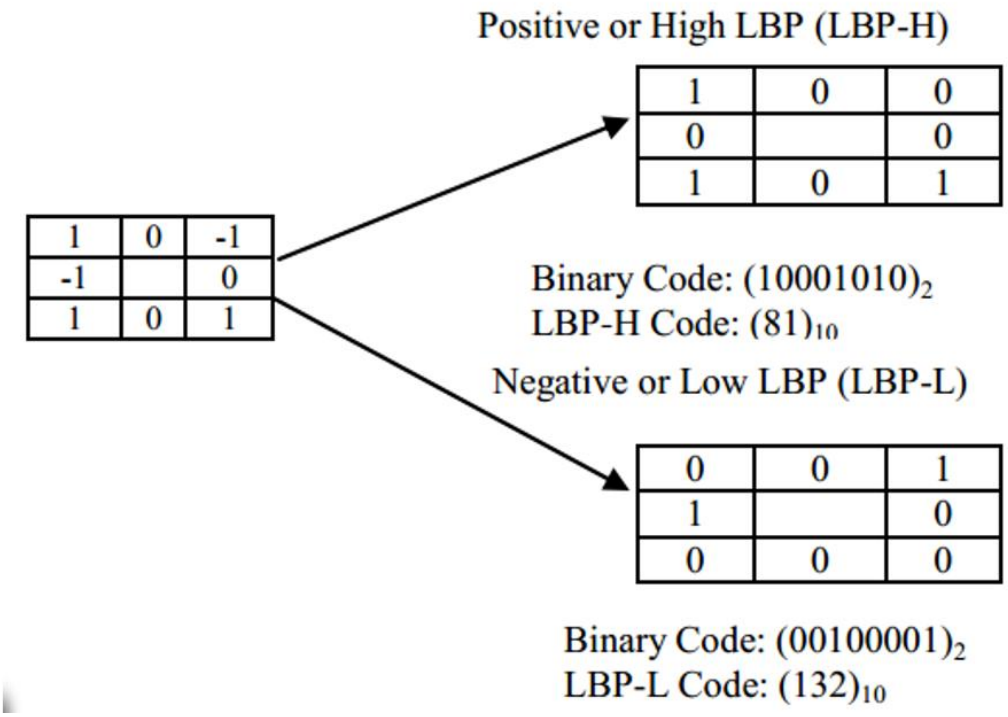


Figure 15: Splitting of LTP in to LBP_H and LBP_L (Tan, Xiaoyang and Triggs, Bill, 2007)

From the observation of figures14 and figure15 of LTP and figures16 and figure17 of LBP, it is proven that a small noise does not affect LTP, whereas the same amount of noise may affect LBP code drastically.

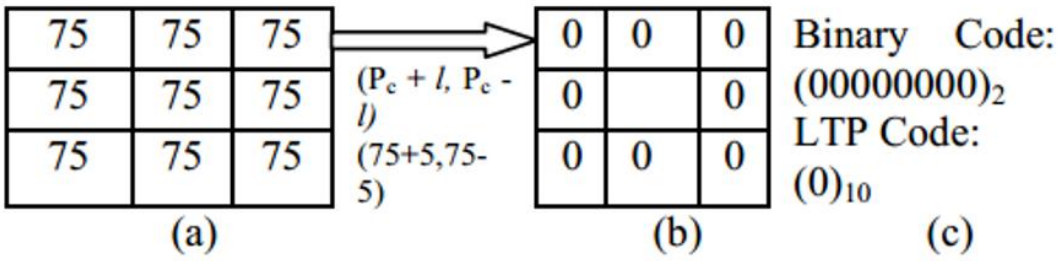


Figure 16: Encoding of basic LTP without noise (Tan, Xiaoyang and Triggs, Bill, 2007)

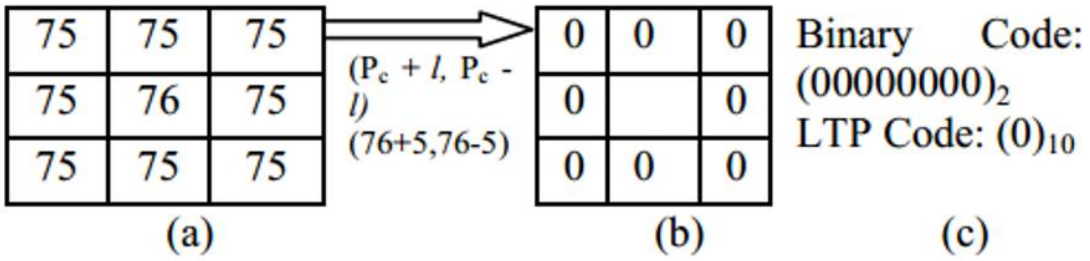


Figure 17: Encoding of basic LTP with noise (Tan, Xiaoyang and Triggs, Bill, 2007)

In an LTP, a small noise is not likely to convert a ULBP into a non ULBP, but conversion is more likely in LBP, that is understandable from figures18 and figure19. And if by chance the small noise spreads on more positions of the image, conversion of many ULBP‘s into NULBP is certain or under heterogeneous label. This affects and drastically deducts the total facial recognition rate and other classification problems.

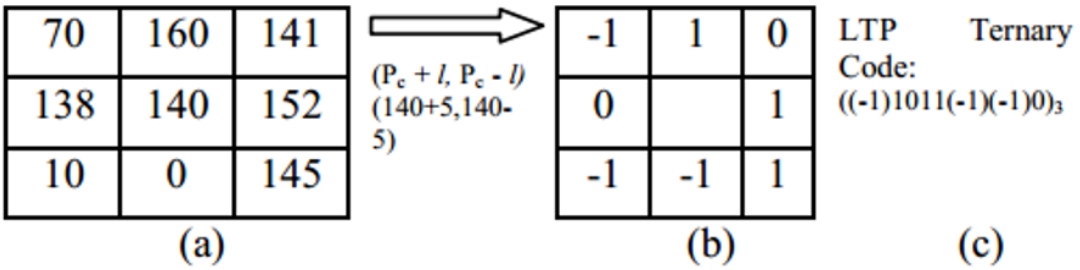


Figure 18: Transitions on LTP without noise (Tan, Xiaoyang and Triggs, Bill, 2007)

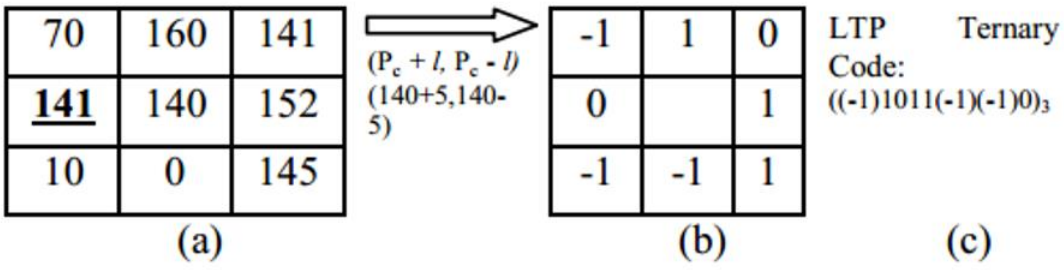


Figure 19: Transitions on LTP with noise (Tan, Xiaoyang and Triggs, Bill, 2007)

3.3.8 Res-Net Networks

Res-Net is the short form of Residual Networks which is a classic neural network used as most substantial method for various computer vision’s function. Resnet won the

Image Net competition in 2015. We could successfully train extremely deep neural networks with 150+layers through Resnet which is considered as the fundamental breakthrough. Training deep neural networks was very hard for the preceding models before ResNet because of the problem of vanishing gradients (Dhankhar, Poonam, 2019). ResNet first proposed the skip connection concept. Skip connection is illustrated in the below diagram. And we pile convolution layers on the right like the previous system, but now there is an addition of the original input to the output of the convolution block. This is known as skip connection. Skip connections function well here because of these reasons (Dhankhar, Poonam, 2019):

- They ease the difficulty of vanishing gradient by providing this shortcut as an alternative path in other to the gradient to flow.
- They enable the model to learn an identity task which guarantees a good performance for the higher layer, to perform as lower layer or even better but not less.

Figure 20: Resnet residual block (Dhankhar, Poonam, 2019)

3.3.8.1 Resnet-50

Resnet-50 is a present advanced convolutional neural network architecture. When there is an additional identity mapping capability with Resnet, it is similar in architecture to networks such as VGG-16 (Simonyan, Karen and Zisserman, Andrew, 2014). Resnet residual block diagram with skip connection is illustrated in figure 21.

3.3.8.2 Resnet-101

A pre-trained model is a kind of model that undergo training on a high standard dataset to address a problem with the same nature to the one we aim to solve. ResNet-101 is a convolutional neural network which has been trained on over one million images from

the Image Net database. The network consists of 347 layers and can possibly categorize images into 1000 object categories. An image input size of the network is 224*224 (He, et al., 2016)

3.4 PATTERN RECOGNITION (CLASSIFICATION)

Classification and Regression are the two methods of supervised learning. Figure22 illustrates the types of machine learning. Pattern recognition is the scientific discipline of machine learning which only aim to classify data (patterns) into many classes or categories. We applied this method in this thesis. No supervising teacher for unsupervised learning but only an input data. The target is finding the constancy in the input. Clustering method is a known method for unsupervised learning whose aim is to find clusters or groupings of input. The common supervised machine-learning techniques are Artificial Neural Network (ANN), Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Decision Trees, Bayesian Networks (BNs) and the Hidden Markov Model (HMM) (Khayam, Syed Ali, 2003). The figure below illustrates machine learning techniques are used extensively in the process of gender classification and face recognition (Mitchell, Tom M, n.d.). Facial classification is the last level of the face recognition channels through the features that have been extracted. The classifier assesses the level of facial similarity and decides based on the similarity.

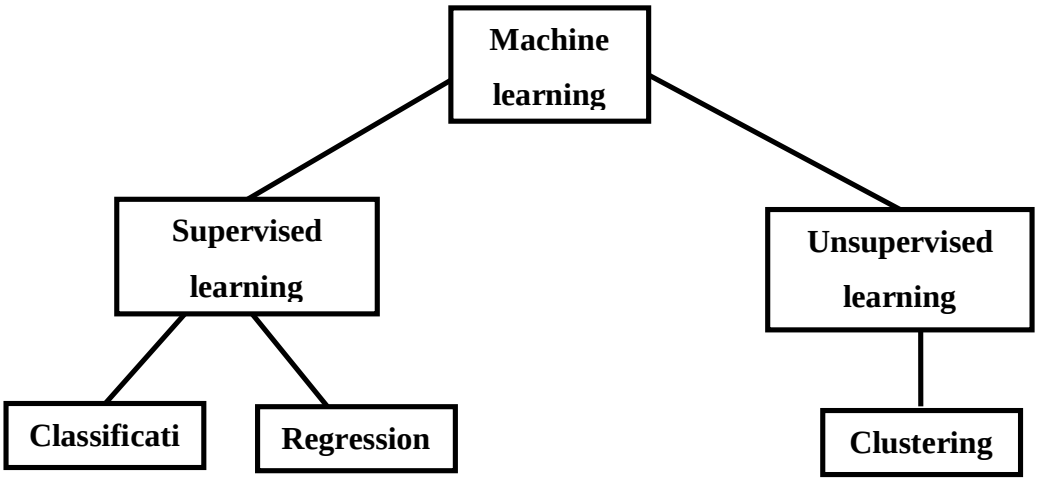


Figure 21: machine learning types (Mitchell, Tom M, n.d.)

3.4.1 K-Nearest Neighbour (KNN)

KNN classifier is a most suitable method for people classification according to their images according to its lesser time consumption in execution and higher accuracy compared to the rest commonly used methods including Kernel method and Hidden Markov Model.

Despite being proved through studies that some methods like Adaboost algorithms and SVM are having higher accuracy than KNN classifier, but KNN classifier is better in the sense of quick performance and domination than SVM. The nearest neighbour classification is the simplest classification scheme in the image space. Under this scheme, we assign the label of the nearest point in the learning set to recognize an image in the test set, in which the distances are measured in image space. To determine the closeness between the data points in KNN, the Euclidean distance metric is often used. A distance is assigned between all pixels in a dataset. Distance is described as the Euclidean distance between two pixels. The Euclidean distance is shown in equation 6 below:

$$d(x_{i1}, x_{i2}) = \sqrt{\sum_{i=1}^{n} (x_{i1} - x_{i2})^2} \quad (2.7) \qquad (6)$$

This Euclidean distance is used by default in a KNN classifier. But we can determine the measurement of the distance between two features depending on one of the distance cosine and correlation (Kaur, Manvjeet and others, 2012).

The K-nearest neighbour algorithm (KNN) is a method for object classification according to the nearest training examples among the feature space. K-NN is a type of instance-based learning, or lazy learning where the function is only approximated locally, and all the calculation is withheld until the classification time. The KNN algorithm is definitely among the simplest of all machine learning algorithms: an object classification is performed by a majority vote of its neighbours, with the object being assigned to the class most common amongst its KNN (where k is a positive integer, and

typically small). The object is simply assigned to the class of its nearest neighbour if k is equal to 1 (Kaur, Manvjeet and others, 2012). The KNN algorithm steps are as follows:

- Determine parameter K, which is the number of neighbours.
- Calculated and query the distances between the instance and its neighbours (K parameter)
- Consociate the distances then decide nearest neighbours according to the K^{th} minimum distance.
- Collect the category of the nearest neighbours.
- Use simple majority of the category of the nearest neighbours as the prediction value of the query instance.

3.4.2 Fuzzy- KNN

Keller (Keller, et al., 1985) proposed a new classification approach to address the objectionable features of simple K-NN by incorporating the Fuzzy set theory with K-NN classifier named as Fuzzy KNN (Mohamed, Salma and Nour, Nahla and Viriri, Serestina, 2018). K-NN and Fuzzy K-NN both having the same searching methodology. However, every data point in K-NN belongs to a single class, and that is the majority class (Ding, et al., 2007), whereas the data points of Fuzzy K-NN belong to multi class and these classes enjoy multiple membership functions. In Fuzzy K-NN method, the fuzzy class membership $u_i(p)$ is tasked to the test instance based on the following equation 7:

$$u_i(x) = \frac{1/\|x - Z_i\|^{2/(m-1)}}{\sum_{j=1}^{c} \left(1/\|x - Z_j\|^{2/(m-1)}\right)} \tag{7}$$

where ***m*** is a fuzzy strength parameter that is used in controlling the efficient magnitude of distance of the neighbours from the test instance, k identifies the number of nearest neighbours, and ***C*** is the number of classes.

3.4.3 Support Vector Machines

Support Vector Machines (SVM) is a supervised learning algorithm that is used for binary classification. Given a set of training samples with labels $\boldsymbol{Fx_i,y_j}$ in which x_i signifies the sample vector and y_i is class label. SVM seeks a hyper plane $\mathbf{w} \cdot \mathbf{x} - b = 0$ in hyperspace H which splits the samples based on their labels (a) in figure23 Here w symbolizes the weight vector and b is the bias. If bias is a varied infinite number of hyper planes can be obtained with the same weight vector. Among all the possible representations of the hyper plane, the maximum and minimum value of b present a kind of hyper planes which function in minimizing the hyper plane distances and samples from one class. Here, support vector is the sample(s) which is closest to the hyper plane support vector (Barrena, Jordina Torrents and Valls, Dom{\`e}nec Puig, 2014). As a matter of fact, w and b are scaled to proper values to represent the two hyper planes as follow:

$$W.\ X - b_1 = 1$$

$$W.\ X - b_2 = -1$$

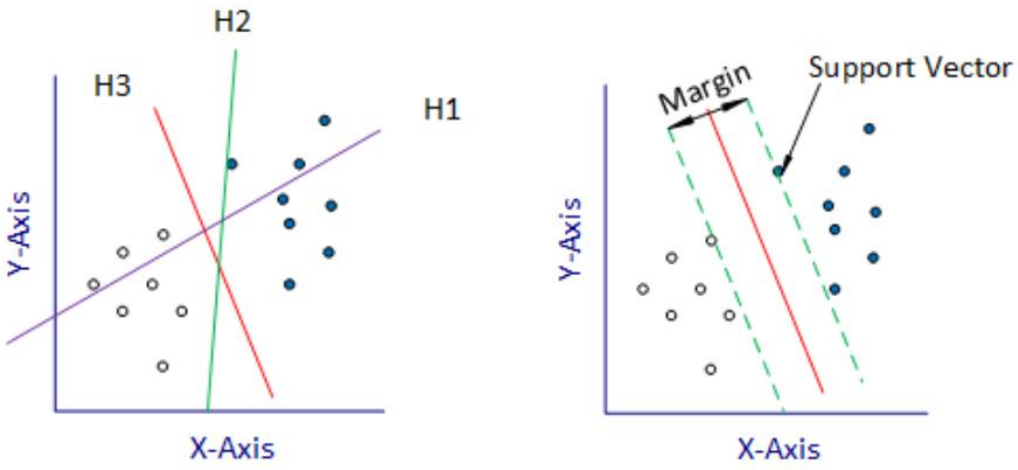


Figure 22: H1, H2 and H3 represent three hyper planes to split classes. H1 fails on the separation, while H2 successfully does the separation. H3 optimizes the separation better than them all (Barrena, Jordina Torrents and Valls, Dom{\`e}nec Puig, 2014)

The distance between these two hyperactive planes is $\frac{2}{||w||}$ which is called margin. Intuitively, the optimum separation should increase the margin when we are having no preceding knowledge of the distribution (see part (b) in figure 3.21)

We can write the optimization problem as follows in equation (8).

$$\min L(w) = \|w\| \text{ subject to } y_i(w \cdot x_i - b) \geq 1 \tag{8}$$

This is a problem of Lagrangian optimization and can be solved using Lagrange

Multipliers α_i

$$f(x) = \text{sgn}\left(\sum_{i=1}^{m} \alpha_i y_i K(x_i, x) + b\right)$$

$$\text{sgn}(v) = \begin{cases} 1, & if\ v \geq 0 \\ -1 & if\ v < 0 \end{cases}$$

where α_i and b are found by using SVC learning algorithm and

for the linear separation in the input space, $K(xi; x) = xi \cdot x$.

Kernel trick technique is employed for non-linear separation. Samples are projected into feature space which is linearly separable. Here are some of popular kernels presented as follows (Barrena, Jordina Torrents and Valls, Dom{\`e}nec Puig, 2014).

Linear

$$K(xi, xj) = (xi.xj)^T$$

Radial basis function (RBF)

$$K(xi, xj) = exp(-\gamma||xi - xj||^2) \ \gamma > 0.$$

Sigmoid

$$K(xi, xj) = tanh\ tanh\ (\gamma.xi.xj + r).$$

Polynomial

$$K(xi, xj) = (\gamma.xi.xj + r)^d, \gamma > 0.$$

3.5 CROSS VALIDATION

Cross-Validation is a statistical process which evaluates and compares learning algorithms by categorizing data into two parts; the first part is used for model training and the other is used for model validation.

The data is randomly separated in k-fold cross-validation into ***k*** equally or almost equally sized folds or segments. Then we perform ***k*** iterations of training repetitively and a different fold of the data is held up for validation for each repetition while the remainder ***k-1*** folds are used for learning as illustrated in figure 24. We can perform evaluation or comparison of learning algorithms in each repetition by using cross-validation. One or more learning algorithms use *k-1* folds of data to train one or more models, afterwards the trained models perform the task of data predictions in the

validation fold. We trace the performance of each learning algorithm on each fold by using accuracy or other predetermined performance standard (Refaeilzadeh, et al., 2009).

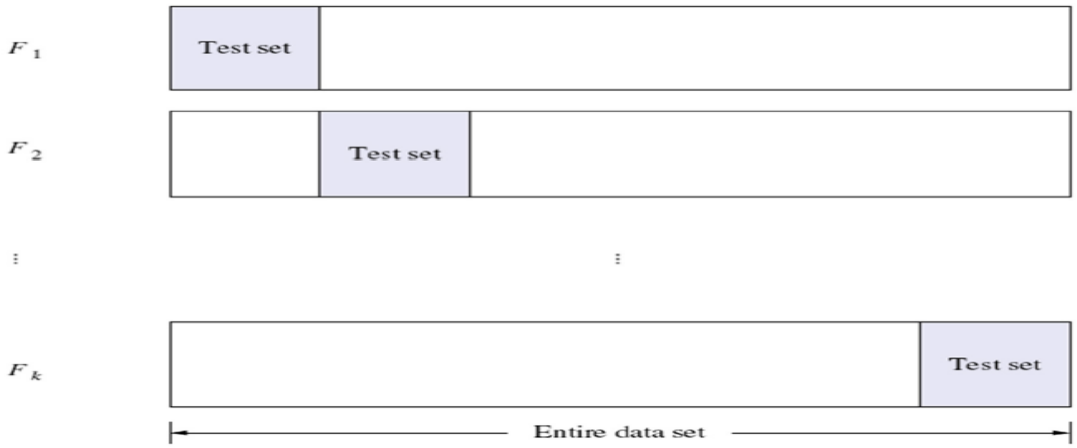


Figure 23: N fold cross validation

3.6 CONCLUSION

This chapter discussed the different methods and techniques used in this research work, including the formulation of the gender identification framework. Feature extraction techniques are classified in this framework to obtain the resulting gender. The results and discussions of this research work are discussed in chapter 5 and 6.

CHAPTER FOUR
DEEP LEARNING FOR VISUAL UNDERSTANDING

4.1 INTRODUCTION

Deep learning is a subdivision of machine learning whose aim is to utilize hierarchical architectures in learning high-level abstractions in data. It is a recent approach but has been broadly used in traditional artificial intelligence domains, such as semantic parsing (Bordes, et al., 2012), computer vision (Dan C and Meier, Ueli and Schmidhuber, 2012), natural language processing (Mikolov, et al., 2013), transfer learning (Krizhevsky, et al. ،2012) and others. There are three significant reasons behind the fast improvement of deep learning recently:

- the dramatically increased chip processing competencies (e.g., GPU units),
- the considerable progress in the machine learning algorithms and
- the significantly lesser cost of hardware computation (Deng, Li, 2014).

Recently several approaches of deep learning have undergone several reviews and discussions extensively. (Bengio, Yoshua, 2013), (Deng, Li, 2014), (Bengio, et al., 2013) studied the difficulties of deep learning in terms of research and presented a few visionary directions. Deep networks have been considered as prosperous networks for computer vision tasks due to their ability of extracting suitable features during connective performance of discrimination (LeCun, Yann, 2012). Deep learning methods were widely adopted in recent Image Net Large Scale Visual Recognition Challenge (ILSVRC) competitions (Russakovsky, et al., 2015) by many researchers and achieved high accuracy score (Burkert, et al., 2015). This research is expected to be beneficial to computer vision and multimedia researchers who aim to known about the state-of-the-art in deep learning in computer vision and to general neural computing as a whole. It provides a general description of different deep learning algorithms and their applications, particularly those that can be employed in the computer vision field.

In recent years, many researchers have extensively examined deep learning in computer vision domain and because of that, many connected approaches have come up.

Generally, these methods can be divided into four categories based on the basic method they are derived from: Convolutional Neural Networks (CNNs), Restricted Boltzmann Machines (RBMs) (Deng, Li, 2014), Auto encoder and Sparse Coding.

The categorization of deep learning methods and some of their important representation works is shown in figure25. We shall briefly review each deep learning methods with their most recent achievement one after the other.

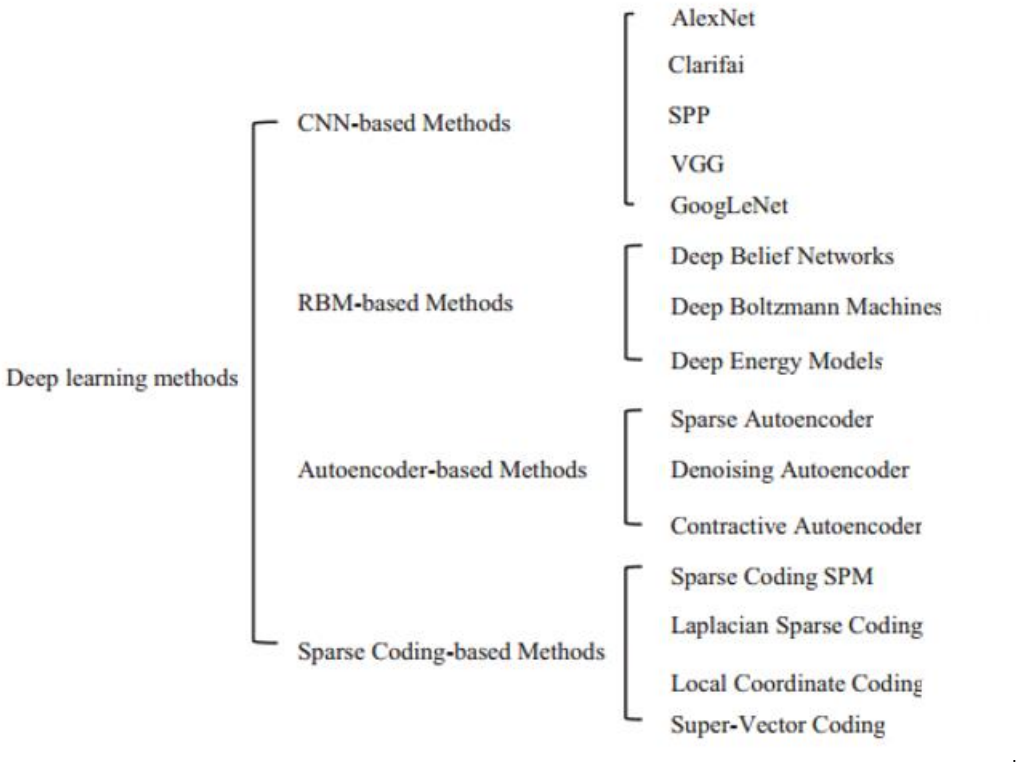


Figure 24: Categorization of deep learning methods and their representative works ,(Bengio, Yoshua, 2013)

4.2 CONVOLUTIONAL NEURAL NETWORKS (CNNS)

The CNN is considered as one of the most remarkable deep learning approaches in which multiple layers are robustly trained (LeCun, et al., 1989)). It has proved its high effectiveness as it is the most commonly used in different computer vision applications. The pipeline of the general CNN structure is illustrated in figure 4.2 generally, a CNN consists of three major neural layers, which are convolutional layers, pooling layers, and fully connected layers. Each kind of layer plays separate roles different from the

others. A general CNN architecture for classifying images (Krizhevsky, et al., 2012) is presented layer-by layer in figure 26. A forward stage and backward stage are the two stages for the network training process. First, the forward stage plays the main stage of representing the input image with the present parameters (weights and bias) in every layer. After that, the predictive output is used for the loss cost computation with the ground truth labels. Second, the backward stage computes the gradients of each parameter with chain rules, according to the loss cost. Then we update all the parameters depending on the gradients, and they are finally prepared for the next forward computation. The network learning stopped after confirming the repetition sufficiency of the forward and backward stages. Then functions of each of the three layers as well as the recent developments of each, and after that we briefly discuss the commonly used training strategies of the networks. Finally, we present various common CNN models, derived models, and conclude with the current tendency for using these models in real applications.

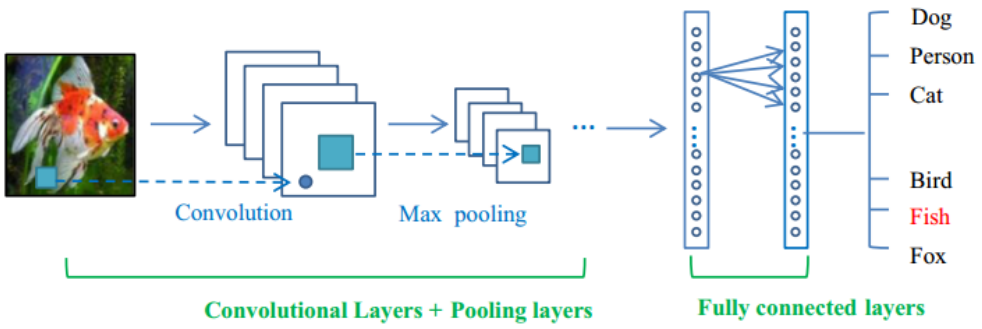


Figure 25: The encoding of CNN architecture (Krizhevsky, et al., 2012)

4.2.1 Types of layers

Generally, a CNN is a hierarchical neural network whose convolutional layers alternate with pooling layers, then followed by some fully connected layers. This section presents the function of these three layers and a brief discussion on the current developments that came out in research on those layers.

4.2.1.1 Convolutional layers.

In the convolutional layers, a CNN uses different kernels convolving the whole image and the average feature maps, and generates various feature maps, as figure 27 shows that. There are three major merits of the convolution operation weights sharing mechanism to reduce the quantity of parameters in the same feature map, local connectivity learns correlations among neighbouring pixels and the convolution is always invariable to the object location (Zeiler, Matthew D, 2013).

As a result of these advantages, the convolution operation has been used in the position of the fully connected layers instead, to speed the learning process by some noted research studies (Szegedy, et al., 2015). The Network in Network (NIN) method is an involving approach of controlling the convolutional layers. The traditional convolutional layer is replaced as main idea of this approach with a view of multilayer perception containing multiple fully connected layers with nonlinear activation functions, thereby replacing the linear filters with nonlinear neural networks (Lin, Min and Chen, Qiang, 2013). This method has achieved remarkable results in image classification.

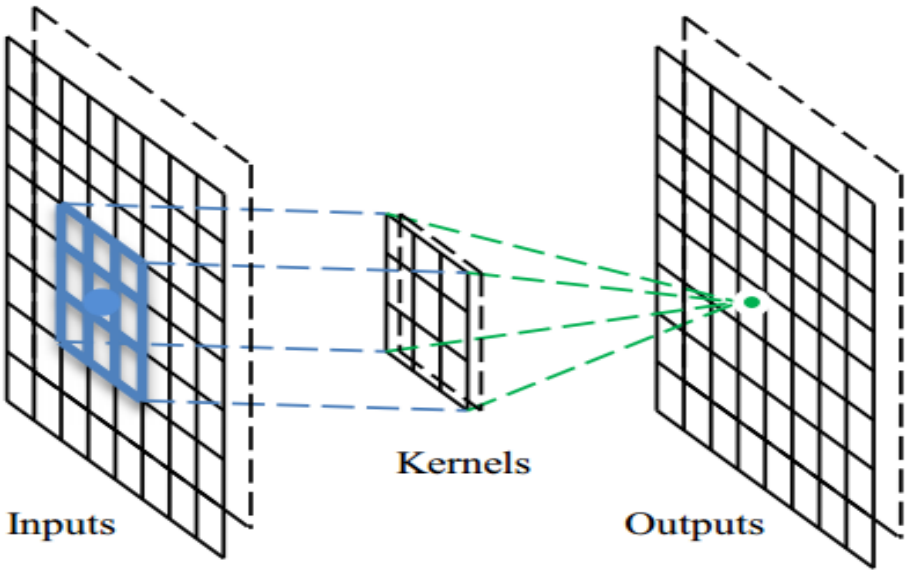


Figure 26: The operation of convolution layer (Szegedy, et al., 2015)

4.2.1.2 Pooling layers.

In general, the next layer after a convolutional layer is a pooling layer which is used to reduce the dimensions of feature maps and network parameters. Pooling layers are the same as convolutional layers as they are both invariant in translation because neighbouring pixels are considered in their computations. It is observed that max pooling as well as average pooling are the most noted strategies used among the layers. With a max pooling operator that has 2×2 size and stride 2, for 8×8 feature maps, the output reduces to 4×4 dimension. (Boureau, et al., 2010) presented a detailed theoretical analysis of the performances of max pooling and average pooling. (Scherer, et al., 2010), furthermore, compared between the two pooling operations and realized that max-pooling can lead to faster convergence and enhance generalization and pick superior constant features. In recent years, various fast GPU implementations of CNN variants have been introduced, and many of them use max-pooling strategy (Cire{\c{s}}an, et al., 2011). The pooling layers are studied in a wide range among the three layers. There are three common related approaches to the pooling layers, in which each has different purpose.

Four.2.1.2.1 Stochastic pooling

A recognized disadvantages of max pooling is the over fitting of the training set as a result of its high sensitivity, which makes it difficult to be well generalized in testing samples (Zeiler, Matthew D, 2013) With the of aim solving this drawback, (Zeiler, Matthew D and Fergus, Rob, 2013) proposed a stochastic pooling approach to address this challenge of max pooling. The proposal replaces the ordinary deterministic pooling operations with a stochastic process, by selecting the activation randomly in each pooling region based on a multinomial distribution. It has got the similar value and use of standard max pooling whereas it has several copies of the input image, each of them has small local deformations. This stochastic pooling with this nature is helpfully in solving the over fitting issue.

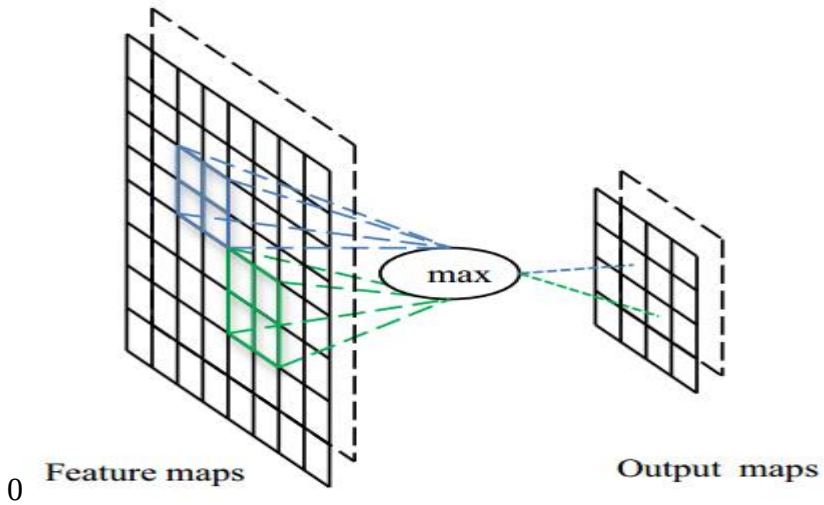
Four.2.1.2.2 Spatial pyramid pooling (SPP)

The methods based on CNN needs a fixed-size input image. This limitation may cause uncompleted accuracy of image recognition of an arbitrary size. (He, et al., 2015)

utilized the general CNN architecture but put a spatial pyramid pooling layer in position of the pooling layer to avoid the limitation in CNN-based method. The spatial pyramid pooling can extract fixed-length representations from arbitrary images (or regions), thus collecting a flexible solution for controlling different sizes, scales, aspect ratios, and this can be applied for improving the CNN performance.

Four.2.1.2.3 Def-pooling

(Ouyang, et al., 2014) discovered handling deformation as a basic problem in computer vision, for the object recognition task particularly. Despite the usefulness of max pooling and average pooling in dealing with deformation, they failed to learn the deformation constraint and geometric model of object parts. So, to handle the deformation more effectively, they proposed a new deformation constrained pooling layer, named as def-pooling layer, to enhance the deep model by learning the deformation of visual patterns. The traditional max-pooling layer can be replaced by this layer at any information abstraction level. There could be a combination of different pooling strategies to increase the performance of a CNN, due to their purpose and procedure differences. Figure28 illustrates how max pooling operates.



0

Figure 27: The operation of max pooling layer (Ouyang, et al., 2014)

4.2.1.3 Fully Connected Layers.

As figure 28 shows the final pooling layer in the network, there are conversions of the 2D feature maps into an 1D feature vector by several fully connected layers, for more features representing symbols, as figure 29 shows, there is a similarity between fully connected layers in traditional neural network in terms of performance and consist of almost 90% of the CNN parameters. The vector can either be feed forwarded into certain number categories for image classification (Krizhevsky, et al., 2012) or it should be considered as a feature vector in follow-up processing (Girshick, et al., 2014). It is not common to change the fully connected layer's structure, but an example came in the transferred learning approach (Oquab, et al., 2014), which maintained the parameters learned by ImageNet (Krizhevsky, et al., 2012). The last fully connected layer was substituted by two new fully connected layers in order to be adapted to the new visual recognition tasks. These layers consist of several parameters, which led to a high computational effort for training them, and this is considered as the disadvantage of these layers. Therefore, removing these layers or reducing the connection with a specified method is a hopeful and generally applied direction. For example, Google Net (Boureau, et al., 2010) built a wide and deep network while keeping the computational budget constant, by switching to sparsely connected architectures from fully connected.

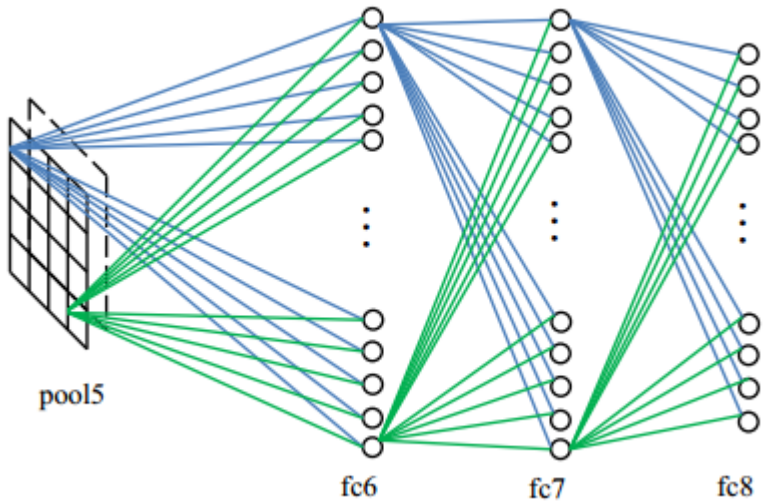


Figure 28: The operation of the fully-connected layer (Boureau, et al., 2010)

Four.2.1.3.1 Training strategy

One of the advantages of deep learning is its ability of building deep architectures to learn more abstract information compared to shallow learning. However, there might be another problem like over fitting resulting from the large number of parameters introduced. In recent years, a number of regularization methods have arisen in defence of over fitting, in which the above-mentioned stochastic pooling is involved, and many other regularization techniques that may influence the training performance that will be presented in this section.

Four.2.1.3.2 Dropout and Drop Connect.

(Hinton, et al., 2012) proposed dropout but it was deeply explained by (Baldi, 2013). During the training of each case, half of the feature detectors are omitted randomly by the algorithm for prevention complex co-adaptations on the training data and improving the generalization ability. This method got more improvement by some researchers (Ba, Jimmy and Frey, Brendan, 2013), and (Warde-Farley, et al., 2013). Research by (Warde-Farley, et al., 2013) precisely analysed the efficiency of dropouts and claimed that it is an extremely efficient ensemble learning method. Drop connect (Wan, et al., 2013) is a very noted generalization derived from Dropout, which rather drops weights instead of the activations in a random manner. It is shown from experiments that it can achieve competitively or even higher results on a variety of standard benchmarks, despite being a bit slower. No-Drop, Dropout and Drop Connect networks are all compared in figure30 (Wan, et al., 2013).

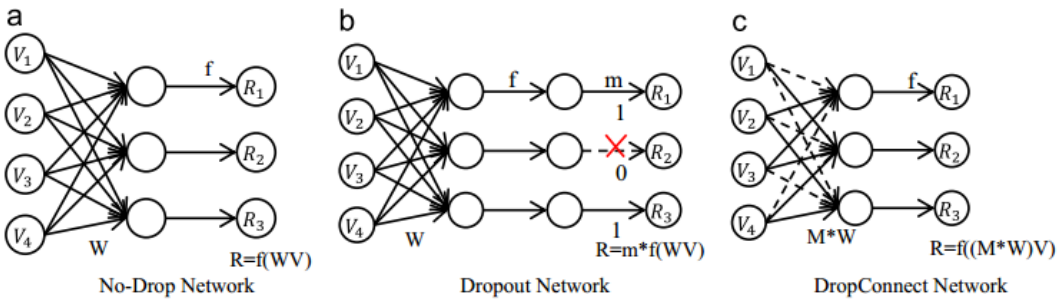


Figure 29: Comparison of No drop dropout connect net (a) No-Drop Network, (b) Dropout Network and (c) Drop Connect network (Wan, et al., 2013)

Four.2.1.3.3 Data augmentation.

Data augmentation is often utilized when a CNN is applied to visual object recognition, to collect additional data without an introduction of extra labelling costs. The common Alex Net (Krizhevsky, et al., 2012) used two noticeably different forms of data augmentation: the first form of data augmentation generates image translations and horizontal reflections, and the second form consists of changing the intensities of the RGB channels in training images. Alex-Net was considered as a base model by (Howard, Andrew G, 2013) and improved the colour constancy by lengthening image crops with extra pixels and additional colour manipulations then added. This method has been broadly applied by researchers on the field in recent years (Szegedy, et al., 2015) and (He, et al., 2015). (Dosovitskiy, et al., 2013) introduced an unsupervised feature learning approach depending on data augmentation. It firstly starts by sampling a set of image patches randomly and declares each of them as a substitute class, then expanded these classes by applying transformations related to scale, translation, colour and contrast. In the end, the substitute classes were discriminated by a trained CNN. Remarkable results showed on a various classification task from the feature learnt by network. Apart from the traditional methods like trimming, scaling, and rotating (Wu, et al., 2015) further adopted colour casting, vignetting and lens distortion techniques, where more training examples were produced with wide coverage.

Four.2.1.3.4 Pre-training and fine-tuning.

Pre-training is a process where the networks with pre-trained parameters are initialized instead of parameters being set randomly. It is quite common in CNN-based models, because of the merits that it has an ability of accelerating the learning process and improve the ability of generalization. (Erhan, et al., 2010) have performed several simulations on the present algorithms to confirm why pre-trained networks perform better than a network that is trained traditionally. Due to the excellent performance, Alex-Net (Krizhevsky, et al., 2012) presented to the public, a number of approaches choose Alex-Net trained on Image Net2012 as their baseline deep model (He, et al., 2016), (Girshick, et al., 2014), (Oquab, et al., 2014) and use fine-tuning of the

parameters based on their precise function. However, there are many approaches (He, Kaiming and Sun, Jian, 2015), (Ouyang, et al., 2014) and (Long, et al., 2015) that present higher performance by training on other models, e.g. Google Net (Szegedy, et al., 2015), and the method of (Simonyan, Karen and Zisserman, Andrew, 2014). Fine-tuning is an extremely important stage for model upgrading to be fitted to tasks and datasets. Generally, class labels are required by fine-tuning when training new dataset, which are utilized for the loss function computation. In such situations, all layers of the new model should be initialized according to the pre-trained model, such as Alex-Net (Krizhevsky, et al., 2012), excluding the last output layer which depends on the number of class labels of the new dataset and for such reason there will be a random initialization. Nevertheless, occasionally obtaining class labels for any new dataset is very hard. Therefore, (Yoo, et al., 2015) proposed a similarity learning objective function to be applied as the loss functions without class labels, so the back reproduction can perform normally and allow the model to be amended one after the other. Several research describe how to transfer the pre-trained model transitions. Another new way is proposed to calculate the quantity of the degree in which a certain layer is general or specific, namely how well the layer features transfer from one task to another. There was a conclusion that a network initialization along with transferred features from almost any layer number could provide an increase to the generalization performance after being fine-tuning into a new dataset. There are other common methods apart from the regularization methods we described above, such as weight decay, weight tying and others. Weight decay performs by adding an additional term to the cost function for parameter penalizing, preventing them from modelling the training data in an exact way and for that reason it helps in generalizing to new examples (Krizhevsky, et al., 2012). Weight tying deduct the number of parameters in Convolutional Neural Networks (Gu, et al., 2018), which enables models to learn good representations of the input data. Note that we can possibly combine these regularization techniques for training to raise the performance due to being non-mutually exclusive.

4.2.2 CNN architecture

Some well-known CNN models have emerged after the new developments of CNN schemes in the computer vision field. This section firstly describes the typically used CNN models, and their characteristics and applications are then summarized. All the

configurations and the basic contributions of some common CNN models. Alex-Net (Krizhevsky, et al., 2012) is a very important CNN structure, which possesses five convolutional layers and three fully connected layers. After the input of one fixed-size (224×224) image, the network would repeatedly convolve and combine the activations, then the results would be forwarded into the fully connected layers. The network was trained on ImageNet and incorporated different regularization techniques, such as dropout, data augmentation, etc. Alex-Net was the winner of the ILSVRC2012 competition (Xie, Saining and Tu, Zhuowen, 2015), and surge an interest in deep convolutional neural network architectures. However, there are two main disadvantages of this model,

1. it is not flexible; it demands a resolution of the image to be fixed.
2. the reason behind the high performance is not clearly understandable.

(Zeiler, Matthew D, 2013)) proposed a novel visualization technique to offer penetration into the inner workings of the medium feature layers. They could realize through the uses of these visualizations the architectures that made Alex-Net (Krizhevsky, et al., 2012) greater on the ImageNet classification standard, and the resulting model, Clarifai, won highest performance in the ILSVRC2013 competition. (He, Kaiming and Sun, Jian, 2015) proposed a new pooling strategy for the demand of a fixed resolution, i.e. spatial pyramid pooling, to remove the restriction of the image size. The accuracy of a various published CNN architectures could be increased by the resulting SPP-net although they are different in terms of designs. Aside from the commonly used configuration of the CNN architecture (five convolutional layers with three fully connected layers), there are also approaches trying to examine deeper networks. (Simonyan, Karen and Zisserman, Andrew, 2014) in contrast to Alex-Net, added more convolutional layers to boost the depth of the network and take advantage of very small convolutional filters in all layers. Similarly, (Szegedy, et al., 2015) introduced a Google Net model, that achieved leading performance in the ILSVRC2014 competition. (Russakovsky, et al., 2015) also has quite a deep structure (22 layers). Despite that, various models have achieved the top-tier classification performances, there is no limitation to image classification of CNN-related models and applications. New architectures have been derived from these models to overcome other difficult tasks, such as object detection, semantic segmentation, etc. RCNN (Regions with CNN

features), (Girshick, et al., 2014) and FCN (fully convolutional network) (Long, et al., 2015) are the two well-known derived frameworks, which are mainly designed for object detection and semantic segmentation respectively. The main idea of RCNN is to collect multiple object proposals and perform feature extraction from each proposal using a CNN, and then use a category specific linear SVM to classify each candidate window. The "recognition using regions" paradigm received encouraging performance in detecting object and has gradually become the general course development for recent promising object detection algorithms (Gidaris, Spyros and Komodakis, Nikos, 2015). Nevertheless, the performance of RCNN depend on the specification of the object location, and that may restrict its strength, in addition to the generation and processing of many proposals that could also decrease its effectiveness. Current developments (Girshick, et al., 2014), (Zhang, et al., 2015) are mostly concentrating on these two areas. RCNN uses the CNN models as feature extractor and does not include any change to the networks, unlike FCN which independently introduces a technique to modify the CNN models as fully convolutional nets. The modifying technique is effective in removing the limitation of image resolution and could provide efficient correspondingly sized output. Despite the fact that FCN could also be used in other application like image classification and edge detection etc., and presented mostly for semantic segmentation, (Yoo, et al., 2015), (Xie, Saining and Tu, Zhuowen, 2015). Apart from designing different models, there are also a few number characteristics shown by usage of these models.

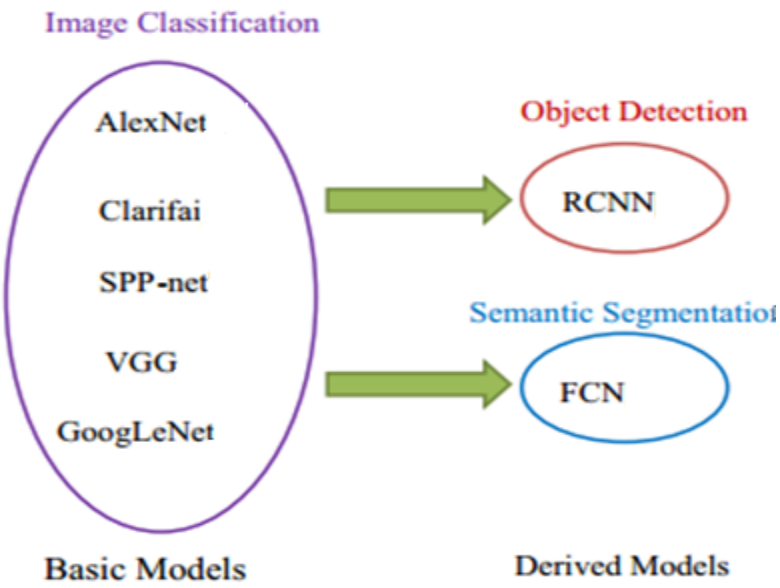


Figure 30: CNN basic model (Yoo, et al., 2015)

4.2.2.1 Large networks.

The obvious target of all models is to enhance the performance of CNNs by boosting their sizes, which involves increasing the depth (the number of levels) and the width (the number of units at each level) (Szegedy, et al., 2015). Both Google Net (Szegedy, et al., 2015) and VGG (Simonyan, Karen and Zisserman, Andrew, 2014) described above, adopted quite large networks, 22 layers and 19 layers, accordingly, proving that image recognition accuracy benefit from size increasing. Collectively training multiple networks could enable higher performance compared to one. Some researchers (Ouyang, et al., 2014) and (Wang, et al., 2014) also combined different deep architectures in cascade fashion to design large networks, and the later networks utilized the output of the former networks, as illustrated in figure32. The cascade architecture can be used in handing different tasks, and there may be changing of tasks in the function of the preceding networks (i.e., the output). For example, (Wang, et al., 2014) used two connected networks to extract objects, and the first network is used for object localization. Hence, the output is the corresponding object coordinator. (Sun, et al., 2013) introduced three-level carefully designed convolutional networks for facial key point's detection. The first level provides highly robust first evaluations, as the remaining two levels fine-tune the earlier prediction. (Ouyang, et al., 2014) also took

up a multi-stage training scheme that was introduced by (Zeng, et al., 2013), i.e., in dealing with misclassified sample classifiers at the earlier stages which are collective task with the classifiers at the present stage.

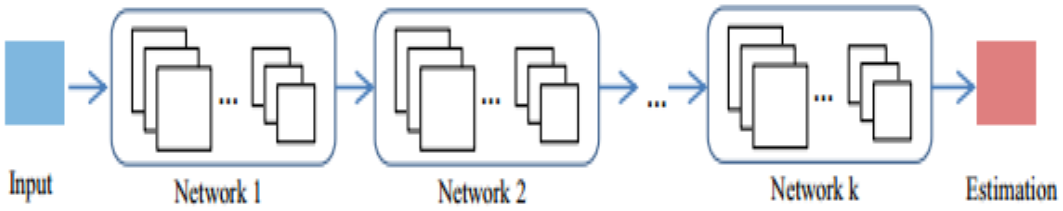


Figure 31: Complaining deep structure in cascade model (Zeng, et al., 2013)

4.2.2.2 Multiple networks.

In recent applications, combining the results of multiple networks is another tendency entirely, where each network can independently perform the work, instead of creating a separate structure and collectively training the networks inside for task implementation, as figure33 clarifies that. (Miclut, Bogdan, 2014) introduced some penetrations concerning the way of generating the final results when series of scores have been received. (Ciregan, et al., 2012) introduced a method named Multi-Column DNN (MCDNN) prior to the Alex-Net (Krizhevsky, et al., 2012), which combines various DNN columns and make their predictions to be average. This model has achieved human competitive results on tasks, such as handwritten digits recognition and traffic signs recognition. (Ouyang, et al., 2014) also recently carried out an experiment to evaluate the performance of model combination strategies. This experiment learnt 10 models with different settings and made them averaging scheme combination. It was proved from the results that models collected through this strategy are highly diverse and complementary to each other in increasing the detection results.

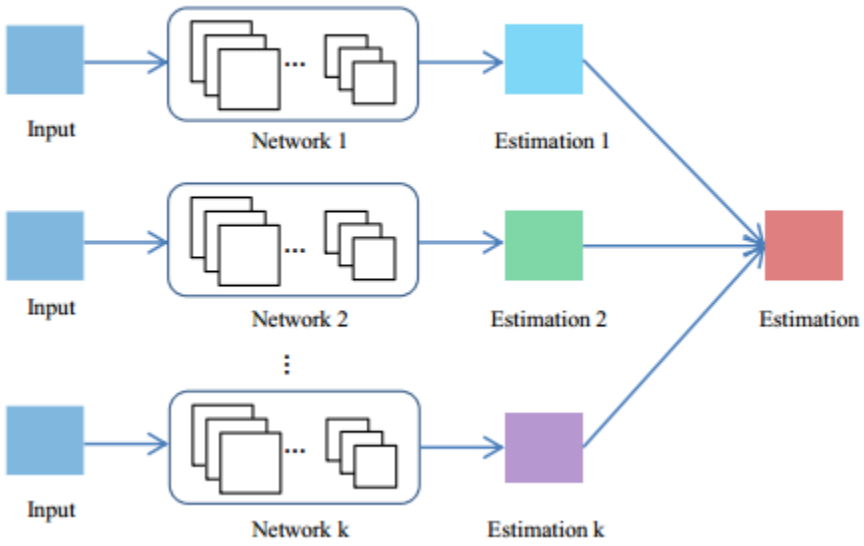


Figure 32: Combining the results of multiple networks (Ouyang, et al., 2014)

4.2.2.3 Diverse networks.

Other than modifying the CNN architecture, some researchers go far with the aim of introducing information from other sources, e.g., shallow structure combination, integrating contextual information, as illustrated in figure 34. Shallow methods can give additional insight into the problem. In the literature, we can get examples of combining shallow methods and deep learning frameworks (Weston, et al., 2012), i.e., consider a deep learning method for feature extraction and input these features to the shallow learning method, e.g., an SVM. RCNN method (Girshick, et al., 2014) has been considered as one of the most representing symbol and successful algorithms, which feeds the highly distinctive CNN features into an SVM to finalize the detection task for objects. Besides that, we can combine deep CNNs and Fisher Vectors (FV) (Simonyan, Karen and Zisserman, Andrew, 2014) to significantly boost the image classification accuracy and they both complement each other. An object detection task sometimes gets contextual information available, and it can possibly incorporate global context information with the information from the bounding box. NUS which won the Image Net Large Scale Visual Recognition Challenge 2014 (ILSVRC2014), interlinked all the raw detection scores and combined them with the outputs from a traditional classification framework by using context refinement (Song, et al., 2011). Similar to

that was (Ouyang, et al., 2014) who also took the 1000-class image classification scores as the contextual features for object detection.

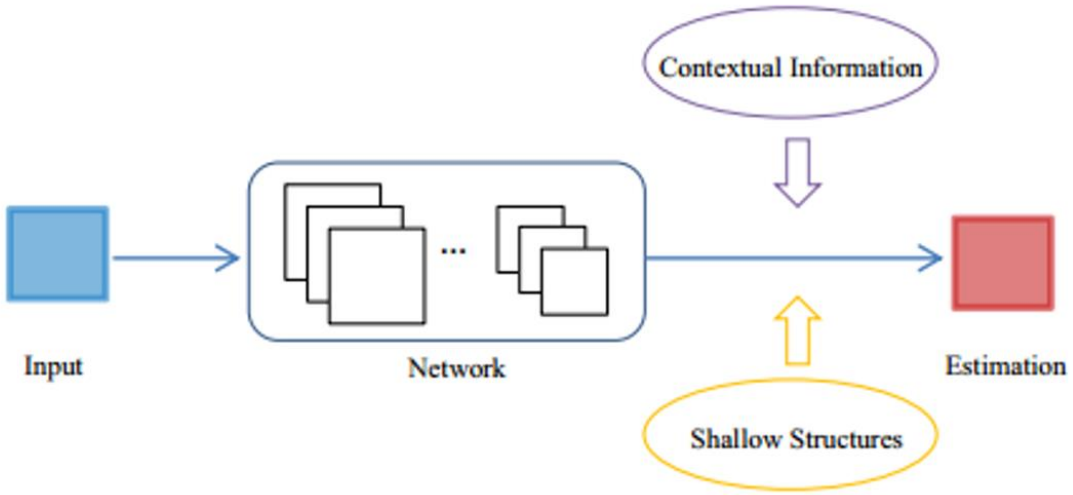


Figure 33: Combination a deep network with information from other sources (Song, et al., 2011)

4.3 RESTRICTED BOLTZMANN MACHINES (RBMS)

A Restricted Boltzmann Machine (RBM) was proposed by (Hinton, Geoffrey E and Sejnowski, Terrence J and others, 1986) which is a generative stochastic neural network. It is a bit different from the Boltzmann Machine, with the limitation of forming the visible units and hidden units as a bipartite graph. This restriction enables more efficient training algorithms, most especially the gradient-based contrastive divergence algorithm (Carreira-Perpinan, M and Hinton, G, 2005). Hinton (Hinton, Geoffrey E, 2012) presented much explanation and mentioned a way of training RBMs practically. Further work in (Cho, et al., 2011) discusses the main challenges to be faced while training RBMs, the reasons for the challenges and presented a new algorithm, which has an enhanced gradient and an adaptive learning rate, to address those challenges. A noted progress of RBM can be found in (Nair, Vinod and Hinton, Geoffrey E, 2010). The model approximates the binary units with noisy rectified linear units to maintain information about relative intensities as information goes round through multiple layers of feature detectors. The refinement is broadly applied in other different approaches based on CNN and functions well as it does in these models (Krizhevsky, et al., 2012),

and (Zeiler, Matthew D and Fergus, Rob, 2013). When utilizing RBMs as learning modules, the following deep models can be composed: Deep Belief Networks (DBNs), Deep Boltzmann Machines (DBMs) and Deep Energy Models (DEMs). The three models were compared in figure 35. At the first two of the layers, DBNs have undirected connections which form an RBM but with directed connections to the lower layers. DBMs have undirected connections within all layers of the network. DEMs have stochastic hidden units at the top hidden layer and have deterministic hidden units for the lower layers (Ngiam, et al., 2011).

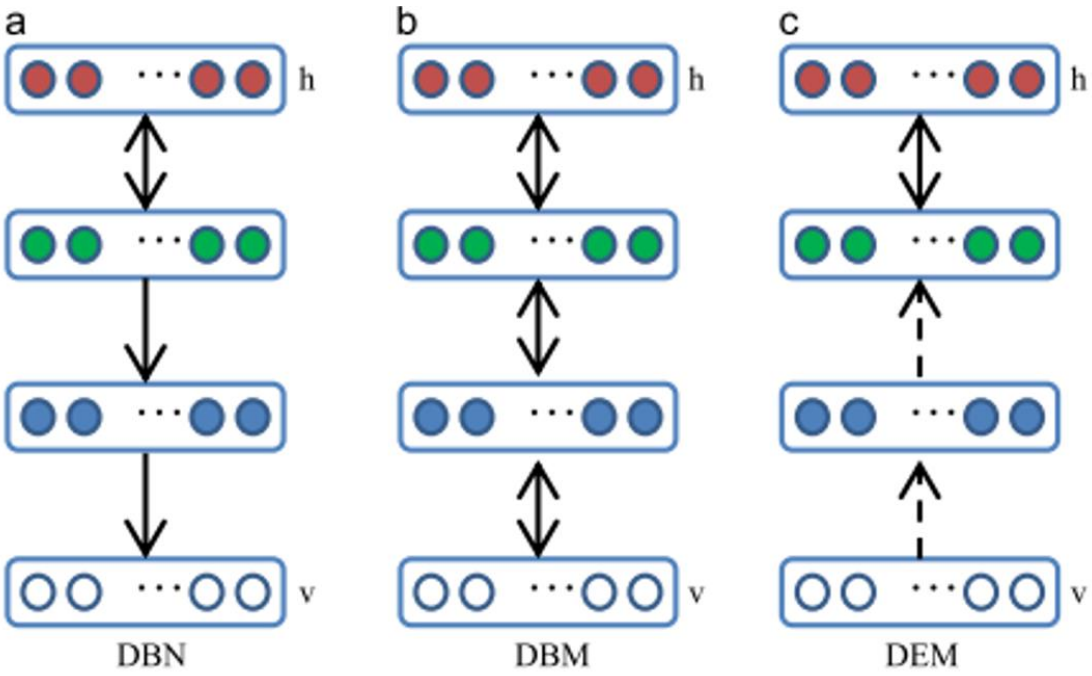


Figure 34: Deep Belief Networks (DBNs) (Ngiam, et al., 2011)

The Deep Belief Network (DBN) was introduced by (Hinton, Geoffrey E, 2012) and considered as a significant advance in deep learning. It gives a collective probability distribution over observable data and labels because it is a probabilistic generative model. A DBN first utilized an effective layer-by-layer greedy learning strategy to initialize the deep network, and then fine-tunes all the weights jointly with the expected outputs. There are two main advantages of greedy learning procedure (Arel, et al., 2010):

(1) it gathers a suitable initialization of the network, overcoming the challenge in parameter selection which prevent the local optima to a certain level.
(2) it is an unsupervised procedure where class labels are not required; therefore, labelled data is unnecessary for training.

Nevertheless, creating a DBN model involves training several RBMs which makes a computationally expensive task, and maximum-likelihood training is uncertain approximately to further optimize the model (Bengio, et al., 2013). DBNs attracted researchers' concentration on deep learning and because of that, several variants were created (Lee, et al., 2008). (Nair, Vinod and Hinton, Geoffrey E, 2010) developed a modified DBN where the top-layer model utilizes a third-order Boltzmann machine for object recognition. The model in (Lee, et al., 2008) learned a two-layer model of natural images using sparse RBMs, in which the first layer learns local, oriented, edge filters, and the second layer captures a variety of contour features as well as corners and junctions. (Lee, et al., 2008) applied two strategies to improve the robustness against occlusion and random noise. The first strategy takes advantage of sparse connections in the first layer of the DBN to regularize the model, while the second is to develop a probabilistic de-noising algorithm. A drawback of DBNs is that when they are applied to computer vision task, they do not consider the 2D structure of an input image. So, the Convolutional Deep Belief Network (CDBN) was introduced (Lee, et al., 2009). to address this problem. CDBN introduced convolutional RBMs to utilize the spatial information of neighbouring pixels, generating a translation invariant generative model that scales well with high dimensional images. The algorithm was further extended in (Huang, et al., 2012) and succeeded having excellent performance in face verification.

4.3.1 Deep Boltzmann Machines (DBMs)

The Deep Boltzmann Machine (DBM), introduced by (Hinton, et al., 2012), is also a deep learning algorithm where the units are rearranged in layers. Comparing DBM to DBNs, whose top two layers form an undirected graphical model and lower layers form a directed generative model, and the DBM has undirected connections over its architecture. The DBM is also a subdivision of the Boltzmann family like the RBM, but DBM is different because it has multiple layers of hidden units, with units in odd-numbered layers being conditionally independent from even-numbered layers, unlike

RBM. Given the visible units, calculating the posterior distribution over the hidden units is no longer controllable, because of the relations between the hidden units. When we train the network, all layers of a specific unsupervised model would be trained together by a DBM, and instead of increasing the likelihood directly, the DBM maximizes the lower bound on the likelihood using a stochastic maximum likelihood (SML). (Younes, Laurent, 1999) use Markov chain Monte Carlo (MCMC) method to update one or few between each parameter update. To avoid this problem, DBM involve a greedy layer-wise training strategy into the layers. Figure 35 illustrate the comparison of the three models (Ngiam, et al., 2011) (a) DBN, (b) DBM and (c) DEM. To look at their Method, Characteristics, Advantages, Drawbacks and References, the top two layers of DBN (Hinton, Geoffrey E and Salakhutdinov, Ruslan R, 2006) are undirected in connection and the lower layers are directed in connection. The characteristics of DBN are:

1. A DBN properly initializes the network, which prevents the process to a reasonable extent.

2. Using an unsupervised training eliminates the need of labelled data for training.

Creating a DBN model is expensive in term of computation because of the involvement of the initialization process. (Lee, et al., 2008), (Lee, et al., 2009) undirected connections between all layers of the network deals more robustly with ambiguous inputs by incorporating top–down feedback. The joint optimization is time-consuming (Hinton, Geoffrey E, 2012) , (Cho, et al., 2011), (Montavon, et al., 2012) (Goodfellow, et al., 2009) (Ngiam, et al., 2011) the lower layers have deterministic hidden units and stochastic hidden units at the top, hidden layer produces better generative models that is allowed to adapt to the training of higher layers. The learnt initial weight may not have good convergence (Carreira-Perpinan, Miguel and Wang, Weiran, 2014) (Elfwing, et al., 2015) when the DBM network is been pre-trained, it is a lot similar to the DBN (Bengio, et al., 2013). The collective learning has achieved significant progress, both in terms of likelihood and the classification performance of the deep feature learner. Nevertheless, a main drawback of DBMs is that the time complicacy of approximate inference is too high compared to DBNs, which leads to the impractical of the collective optimization of DBM parameters for large datasets. In order to develop the effectiveness of DBMs, an approximate inference algorithm was proposed by some researchers

(Salakhutdinov, Ruslan and Larochelle, Hugo},, 2010), which uses a special "recognition" model for initializing the values of the latent variables in all layers, therefore, the inference is effectively accelerated. There are also various approaches aiming to enhance the efficiency of DBMs. The improvements can either occur at the pre-training step (Hinton, Geoffrey E and Salakhutdinov, Russ R, 2012), (Cho, et al., 2013) or when training (Montavon, et al., 2012), (Goodfellow, et al., 2009). For example, (Montavon, et al., 2012) proposed the cantering trick to improve the invariability of a DBM and increase the level of generating and discriminating. The multi-prediction training scheme (Goodfellow, et al., 2013) was used to collectively train the DBM which surpasses the preceding methods in image classification presented in (Goodfellow, et al., 2013).

4.3.2 Deep Energy Models (DEMs)

The Deep Energy Model (DEM) was proposed by (Ngiam, et al., 2011), which is the latest approach for training deep architectures. DEM is dissimilar to DBNs and DBMs which both share the property of having multiple stochastic hidden layers, while the DEM has a single layer of stochastic hidden units for efficient training and inference. Deep feed forward neural networks are utilized to model the energy landscape and it has the ability of training all layers simultaneously. (Ngiam, et al., 2011) use hybrid Monte Carlo (HMC) to train the model. There are also other options including score matching, contrastive divergence, and others. Such work can be found in (Elfwing, et al., 2015). Despite that CNNs is more suitable than RBMs in terms of vision applications, there are still many good examples adopting RBMs to vision tasks. (Eslami, et al., 2014) proposed the Shape Boltzmann Machine, to be in charge of the task of modelling binary shape images, which learns high quality probability distributions over object shapes, for both realism of samples from the distribution and generalization to new examples of the same shape class. The CRF and the RBM were combined by (Kae, et al., 2013) to model both local and global structure in face segmentation, which has been consistent in reducing the error in face labelling. New deep structure has been introduced for phone recognition (Dahl, et al., 2010) where a Mean-Covariance RBM feature extraction module with a standard DBN were

combined. This approach tackles the problems of representational ineffectiveness of GMMs and a significant restriction of preceding work by applying DBNs to phone recognition.

4.4 AUTOENCODER

The autoencoder is one of the artificial neural networks specially made to be used for learning efficient encodings (Liou, et al., 2014). An autoencoder is trained to reconstruct its own inputs X instead of training the network to predict some target value Y given input X. Consequently, the output vectors and the input vector have similar dimension. Figure36 shows the general process of an autoencoder. The autoencoder is optimized during the process by minimizing the reconstruction error, and the similar code is considered as the learned feature. In general, a single layer is not able to get the discriminative and representative features of raw data. Recently researchers have utilized the deep autoencoder, which progresses the code learnt from the previous autoencoder to the next level, to get their task accomplished. The deep autoencoder was first proposed by (Hinton, Geoffrey E and Salakhutdinov, Ruslan R, 2006), and is still widely studied in many papers (Zhou, et al., 2014). A deep autoencoder is often trained with a variant of back-propagation, e.g., the conjugate gradient method. Though deep autoencoder is often reasonably effective, but the presence of errors in the first few layers could cause ineffectiveness of this model. This may cause the network to learn to reconstruct the average of the training data. Pre-training the network with initial weights that approximate the final solution is the proper approach to solve this problem. (Hinton, Geoffrey E and Salakhutdinov, Ruslan R, 2006) proposed to make the representation as “constant” as possible with respect to the changes in input. In the next sections, we describe three important variants: sparse autoencoder, denoising autoencoder and contractive autoencoder.

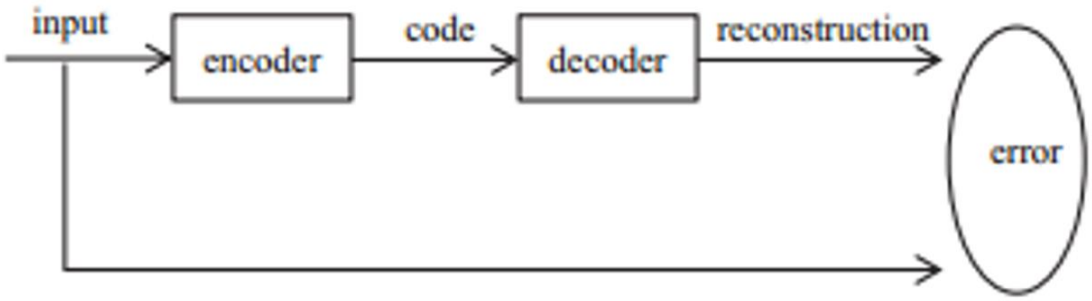


Figure 35: The pipeline of autoencoder (Liou, et al., 2014)

4.4.1 Spars Autoencoder

A sparse autoencoder intends sparse feature extraction from raw data. We can achieve the sparsity of the representation either by penalizing the hidden unit biases (Ranzato, et al., 2007), (Lee, et al., 2008) (Goodfellow, et al., 2009) or by directly penalizing the output of hidden unit activations (Le, et al., 2011), (Zou, et al., 2011). There are several potential advantages for sparse representations (Ranzato, et al., 2007).

1) The likelihood that different categories will be easily separable is increased when using high-dimensional representation, similar to SVMs theory.

2) Simple interpretation of the complex input data from sparse representations in terms of a number of "parts";

3) Biological vision uses sparse representations in early visual areas (Simoncelli, Eero P, 2005).

The common difference between sparse autoencoder and other models is a nine layer locally connected sparse autoencoder with pooling and local contrast normalization (Le, Quoc V, 2013) this difference allows the system to train a face detector without having to label images as containing a face or not.

4.4.2 Denoising autoencoder

Vincent proposed a model called denoising autoencoder (DAE) (Vincent, et al., 2008) (Vincent, et al., 2010) to increase the robustness of the model, which has the ability of recovering the correct input from a corrupted version. Hence, it forces the model to capture the structure of the input distribution. The process of a DAE is shown in figure37.

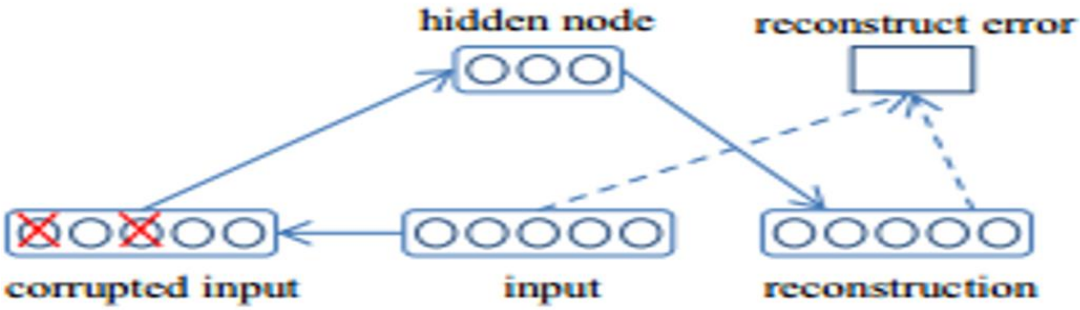


Figure 36: De noising autoencoder (Vincent, et al., 2010)

4.4.3 Contractive autoencoder

Contractive autoencoder (CAE) was introduced by (Rifai, et al., 2011). It came into existence after the DAE and shared the same motivation of learning robust representations (Bengio, Yoshua, 2013). A DAE injects noise in the training set, to make the whole mapping robust, while a CAE obtained strength by adding an analytic contractive penalty to the reconstruction error function. Although (Bengio, et al., 2013) stated the notable differences between DAE and CAE, (Alain, Guillaume and Bengio, Yoshua, 2014) suggested some close relations between DAE and a form of CAE; a DAE with small corruption noise can be valued as a type of CAE where the contractive penalty is on the whole reconstruction function rather than just on the encoder. In the Unsupervised and Transfer Learning Challenge both DAE and CAE were successfully used (Glorot, et al., 2012).

4.5 SPARSE CODING

The sparse coding is purposely used for learning an over-complete set of basic functions to describe the input data (Glorot, et al., 2012). There are many benefits of sparse coding (Yu, et al., 2009), (Yang, et al., 2009):

(1) It is more efficient in reconstructing the descriptor with the use of multiple bases and capture the correlations between the same descriptors which share bases.

(2) The sparsity enables the representation to attain most important properties of images.

(3) It is in line with the biological visual system, which argues that sparse features of signals are useful for learning.

(4) Image patches belong to sparse signals as proved by image statistics study.

(5) Patterns with sparse features are more linearly separable.

CHAPTER FIVE
LOCAL BINARY PATTERNS AND CNN ALGORITHMS

5.1 INTRODUCTION

In this chapter, we proposed two simple algorithms to provide a novel structure for gender identification. Generally, we firstly described the gender identification when applying on the whole face using global feature extraction techniques in details and using KNN to classify the gender and get the first result. Next, we used the local feature extraction technique LBP, LDP, LTP and a parameter is added to LTP to become DLTP to identify the gender and produce the other result. Then deep learning techniques Resnet-50 and Resnet-101 and the SVM were used to achieve the other result.

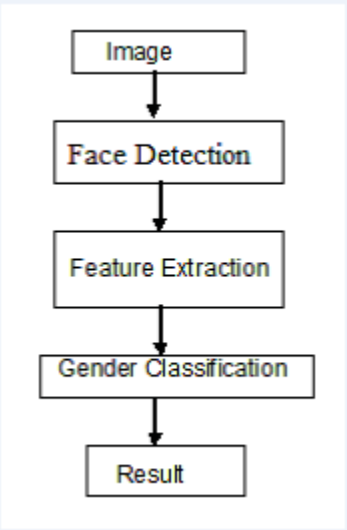


Figure 37: The general system for gender identification

5.2 GENDER IDENTIFICATION FROM FACIAL IMAGES USING GLOBAL FEATURES

The performance of global feature extraction techniques (DCT, Block DCT, DWT and Hybrid DCT-DWT) is evaluated in terms of classification accuracy of KNN, Fuzzy KNN and SVM techniques using the two facial datasets, namely FERET and ESSEX. It was proven by the experimental results that Hybrid DCT-DWT got a high accuracy

when we used the KNN as a classification technique on the ESSEX dataset. However, when FERET dataset was used, the DCT attained higher accuracy. As well as this, the Hybrid DCT-DWT got the higher accuracy when using F-KNN and SVM.

The ESSEX data is held in four directories (faces94 , faces95 , faces96 , grimace). The dataset that exists in faces94 subdirectory was used. This database contains 153 images with different objects and green backgrounds, and each image has a .jpg format. There are 20 images for every object, which is 3060 images for the whole database. For some subjects, images were taken at separate times with lighting variation, facial expression (eyes open, eyes closed, smiling face, and frowning face) and facial details were varied (with glasses and without glasses). Each image has the size of 180×200 pixels. The figures below illustrate the sample.

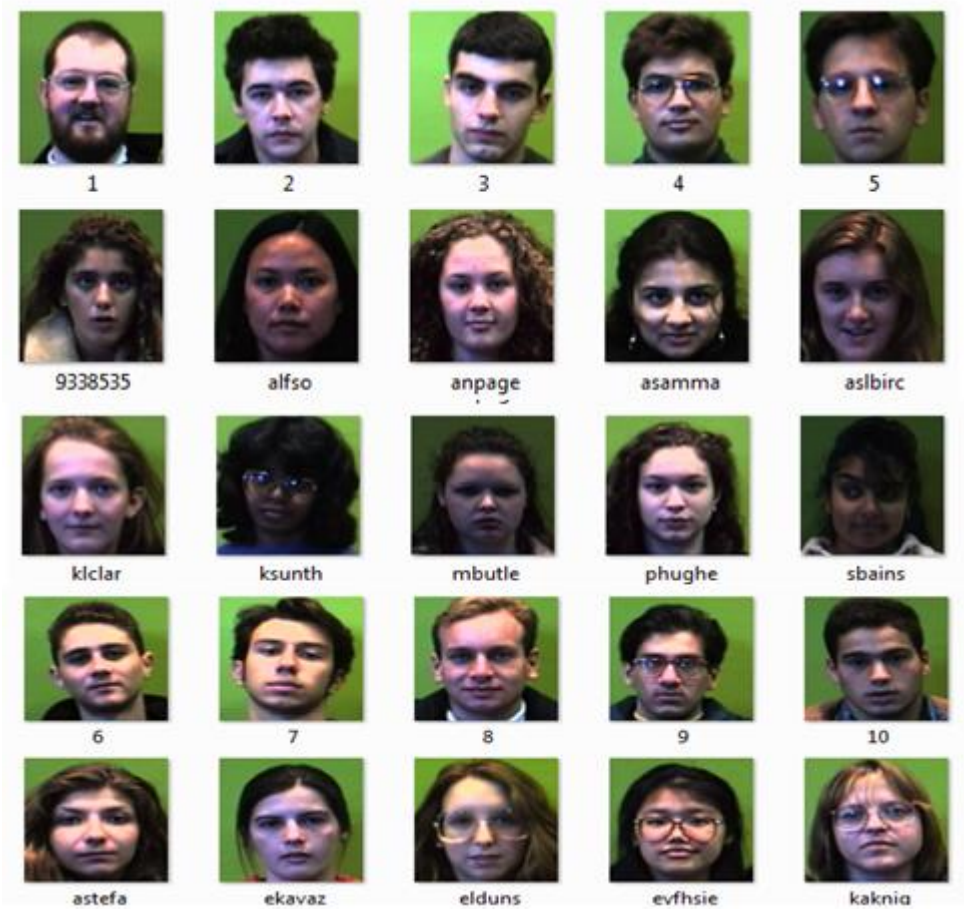


Figure 38: Sample of ESSEX dataset

FERET: (Face Recognition Technology) the FERET program intended to show a large database of facial images that was collected separately from the algorithm developers. The images were collected in a semi-controlled environment. To manage a degree of persistency all over the database, a similar physical setting was used in each photography session. Because there must be equipment similarity in every session, some gathered images on different dates were varied. The collection of the FERET database happened in 15 sessions between August 1993 and July 1996. The database consists of 1564 sets of images with a total of 14,126 images that includes 1199 individuals and 365 duplicate sets of images. A duplicate set is a secondary set of images of a person already in the database and was usually taken on a different day. Many years passed for some individuals between their first and last sittings, and some subjects were snapped multiple times. This time lapse was essential because it allowed researchers to study for the first time the changes in a subject's appearance that happened within the years.

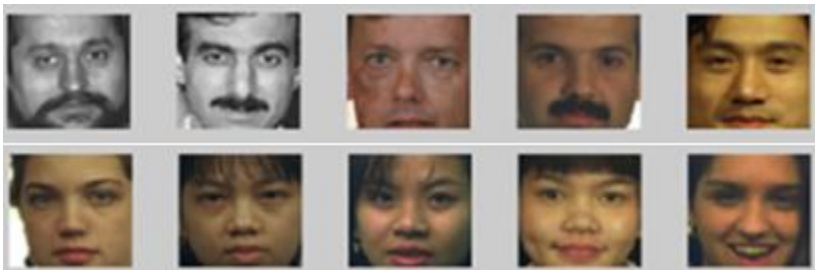


Figure 39: Sample of FERET dataset

5.3 COMPONENT-BASED GENDER IDENTIFICATION USING LOCAL BINARY PATTERNS

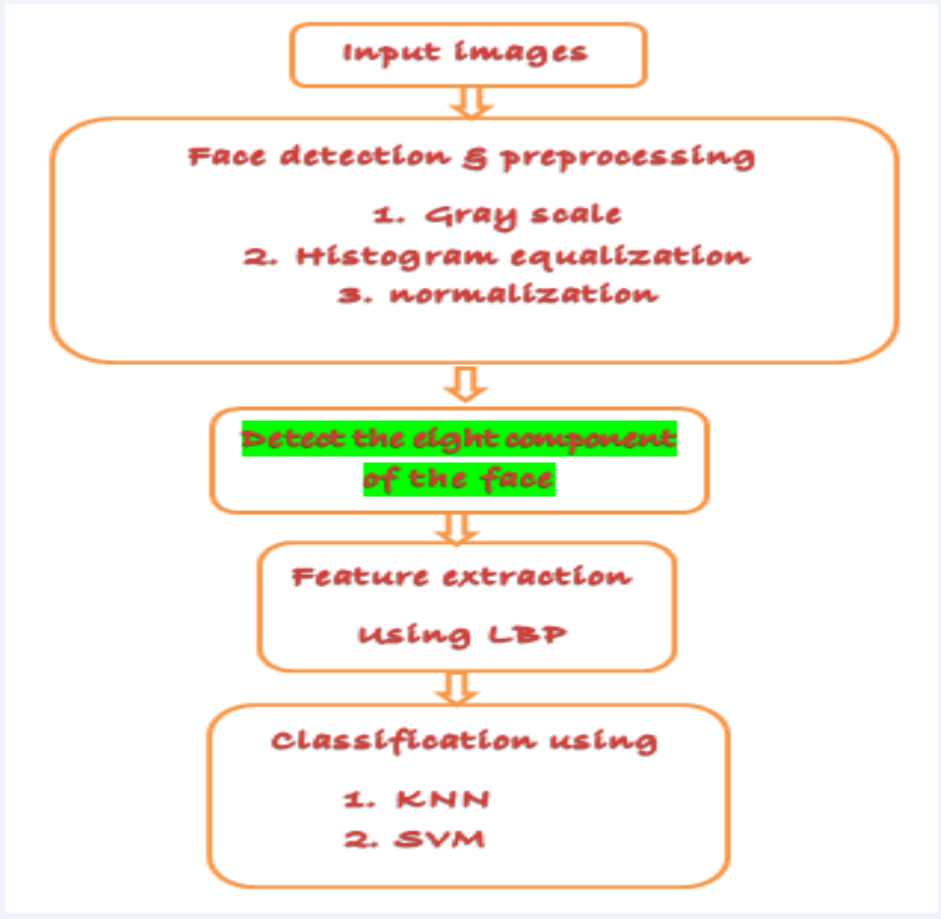


Figure 40: The proposed method when using LBP and SVM

In this part, ESSEX, FERET and UOFG frontal face databases were the facial datasets used. The ESSEX database consists of facial images with 153 different objects (every object has 20 images) with a green background. The FERET database consists of 485 images for training and 416 images for testing. The image has the size of 256 × 384 pixels with grey scale and colour images. The UOFG dataset contains 10,000 frontal images within the age group of 18 to 23 years old. Detection of the whole face and some regions of the facial comparison were performed, using Viola and Jones face detection technique and DRMF model, LBP feature extraction technique and two classifiers KNN and SVM.

Figure 41: Sample of UOFG dataset

5.4 DYNAMIC LOCAL TERNARY PATTERNS FOR GENDER IDENTIFICATION USING FACIAL COMPONENTS

In this study, the data was collected from LFW frontal face databases, which contain 500 images, 250 for males and 250 for females. There are also colour and grey scale images. There was performance of facial detection on the eight regions of the face (eyes, nose, mouth, cheeks, chin, and forehead) using the Viola and Jones face detection technique. The DRMF model, LBP, LDP and DLTP feature extraction techniques were applied, and the classifiers SVM was used.

Figure 42: Sample of LFW database

The normalized responses are in the range of 0.0 and 1.0, which signifies the probability of edge from the central reference pixel stretching toward respective direction.

$$p(x_i) = \begin{cases} 1 & if x_i^{norm} \geq 0.05 + \epsilon \\ 0 & if 0.50 - \epsilon < x_i^{norm} < 0.50 + \epsilon \\ -1 & if x_i^{norm} \leq 0.05 - \epsilon \end{cases}$$

0.0	0.72	0.89
0.21	0	0.44
1	0.93	0.03

(a) Normalized pixel

-1	1	1
-1	0	0
-1	1	-1

(b) Assigning DLTP code at £= 0.5±0.1667

5.5 GENDER CLASSIFICATION MODEL BASED ON DEEP CONVOLUTIONAL NEURAL NETWORK

This section presents the experiment of the CNN solution for gender classification. The introduced CNN was applied in Windows with the configuration of Intel Core i5- CPU @ 2.7 GHz with 8GB. MATLAB (2019a) tool was applied in evaluating the method and performing the feature selection and classification task. The learned image features were extracted from a pre-trained CNN for feature extraction. We used the preceding layer of classification layer, named ‘fc1000’, for feature extraction by using the activation method. Then we used those features to train and test SVM classifier. On the other side, the layers to the new classification task were done by replacing the fully connected layer by a new fully connected layer, so that it had the same size as the number of classes in the new data based on dataset classes. This replaced the classification layer with a new classification layer with one without class labels. The Trained Network automatically sets the output classes of the layer at training time. The results demonstrated that pre trained CNN models +SVM were more accurate than transfer learning model.

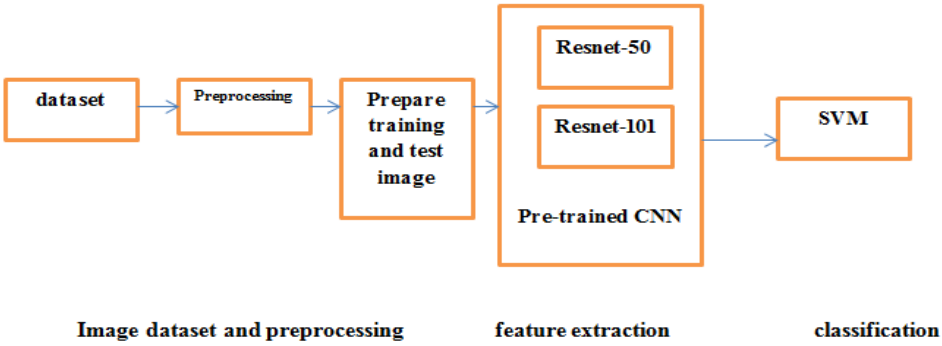


Figure 43: The proposed when using deep learning

CHAPTER SIX
RESULTS AND DISCUSSIONS

6.1 INTRODUCTION

In this chapter, the performance of the facial images was analysed with the different classification techniques. The results achieved in this research are also shown and discussed. We described the performance of the classification model through a table called the confusion matrix and the table is shown below figure 46 depicted confusion matrix.

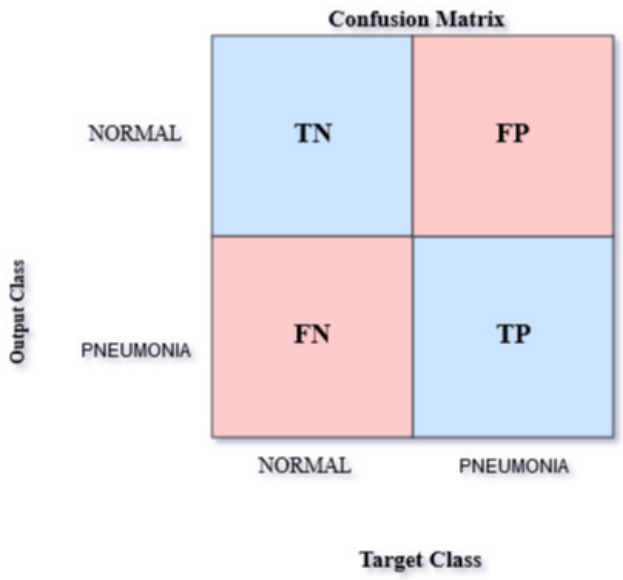


Figure 44: The confusion matrix

(1) True Positives (TP): The number of cases where the model predicted yes and the person has Pneumonia,
(2) True Negatives (TN): the number of cases where the model predicted no, and the person does not have Pneumonia.
(3) False positives (FP): Model predicted yes, but actually they do not have Pneumonia.

(4) False Negatives (FN): Model predicted no, but actually they do have Pneumonia.

The proportion of actual positives samples that is correctly classified is given by sensitivity measures and is calculated as in equation 15

$$\text{Sensitivity} = \frac{TP}{TP+FN} \qquad (15)$$

In addition, for our study, we will be focusing on specificity as it measures the proportion of identified negatives samples, in which the percentage of normal images are correctly classified as normal, and it is calculated as in equation 16

$$\text{Specificity} = \frac{TP}{TP+FN} \qquad (16)$$

Equation 17 computes F1-score, which measures the average F1 score through different class labels:

$$\text{F1} = 2 \times \frac{PPV \times TPR}{PPV+TPR} \qquad (17)$$

The accuracy was the fraction of the predicted labels that are correctly computed as shown in equation 18

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \qquad (18)$$

The programming environment is explained in section 6.2 and the dataset defined in chapter 5 Thereafter, the Feature Extraction Technique is discussed in chapter 3 followed by the classification techniques for identification in chapter 3 also.

6.2 PROGRAMMING ENVIRONMENT

The system was implemented and run with Intel Core (TM) i7-4770S @ 3.10 Ghz and 8.00GB RAM. The system was implemented in MATLAB 2017 using open CV. All these plugins are in the public domain and are used for MATLAB image processing. We have also developed separate image processing operations, some of which are in the region of interest detection, bounding box drawing, and many more. The plugins were used in sequence such that the output of one image will serve as the input to another image or plugin. A statistical MATLAB package was used for the SVM and KNN classification.

6.3 OUR REACHED RESULT IN DETAILS

In this study, the result achieved in many levels beginning from survey of gender identification to the whole face generally after that split the face to the facial component using machine learning techniques and finally using deep learning to identify the gender from face images.

6.3.1 The state-of-the-art

There are several works in gender identification system as illustrated in table1.

Table 1: The survey of gender identification

Author, year	Techniques	Classification	Dataset	Accuracy
Tolba 2001 [8]	PCA	LVQ ,RBF	Facial images	100% ,98.04%
Amit at.al 2005 [10]	ICA	SVM	FERET	96%
Timo at.al 2008 [35]	LPQ , LBP	NN, Chi square distance	CMU PIE , FRGC 1.0.4	99.2% , 92.7%
J. Wu at. al 2008 [37]	PGSFS	RFNM	ND,FERET	91:67%
Luis 2010 [31]	LIB	SVM	FERET ,UND	86.78%, 86.34%
M. Nazir at.al 2010 [12]	DCT	KNN	SUMS	99.3%
Ravi ,Wilson 2010 [34]	RGB , YCbCr	SVM	NUB	threshold 0.07
J. Zheng at.al 2011 [36]	Gray, Gabor, (LBP) , (MLBP), (LGBP)	SVMAC	CAS-PEAL	97.2%
Chia Shih 2012 [17]	(AAM), (PPH)	POPFT	LFW ,FERET	84.2%,86.5%
Yasmina at al 2013 [32]	Gray ,PCA and LBP	1-NN, PCA + LDA ,SVM	FERET, PAL	97.2% 94.06%, 88.57%
P. Rai, P. Khanna 2014 [18]	(2D)PCA	SVM	FERET	98.4%
Kalam , Guttikonda 2014 [33]	LBP ,LDP ,PPBTF ,GWT	ratios threshold	T& T	95.6%
Shan Sung at.al 2016 [9]	CNN	NN	SUMS,T& T	98.75% , 99.38%

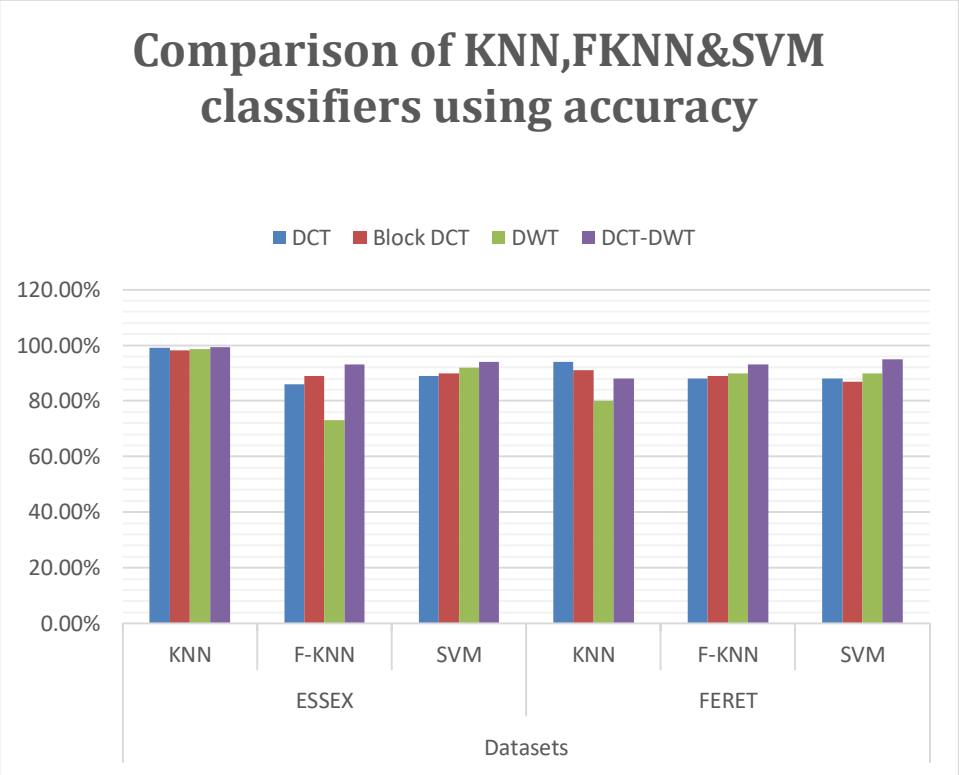
Firstly, pre-processing, face detection, feature extraction and classification were applied to review the general system with the gender identification method. Performing gender classification on pixels is more expensive; therefore, gender classification prefers to extract face features instead of working on pixels directly. Also, it provided a survey on several preceding algorithms of gender identification, which were proposed earlier by researchers for further development in the classification field. It also presents a general explanation of some of the proceeding research in gender identification. Gender identification in entirely unrestricted settings makes it a very fascinating task in all its steps.

6.3.2 Gender Identification from Facial Images using Global Features

This research evaluated the efficiency of four methods of (DCT Zigzag, Block Based DCT Zigzag, DWT Zigzag and Hybrid DWT-DCT Zigzag) using feature extraction algorithms to realize useful features for the purpose of attaining maximum classification accuracies. Two datasets ESSEX and FERED frontal face images were applied for evaluation. The results have proven that the Block Based DCT Zigzag coefficients achieved highest accuracy compared to other feature extraction methods. The performance of recognition (classification) is not only based on feature extraction approaches, but also on the other stages of the recognition processes like the pre-processing stage and the classification algorithm. Table2 below illustrates the result achieved.

Table 2: Comparison of three classifiers using accuracy

	Datasets					
Method	**ESSEX**			**FERET**		
	KNN	**F-KNN**	**SVM**	**KNN**	**F-KNN**	**SVM**
DCT	**99.2%**	**86%**	**89%**	**94%**	**88%**	**88%**
Block DCT	**98.2%**	**89%**	**90%**	**91%**	**89%**	**87%**
DWT	**98.6%**	**73%**	**92%**	**80%**	**90%**	**90%**
DCT-DWT	**99.3%**	**93%**	**94%**	**88%**	**93%**	**95%**



6.3.3 Component-Based Gender Identification Using Local Binary Patterns

A gender classification method was introduced according to the facial components. The results achieved show that using the whole face to identify gender from facial images is not necessary. The Discriminative Response Map Fitting model for facial component detection which are eyes, forehead, cheeks, mouth, nose, and chin was applied. The experimental results performed on the frontal face dataset of LFW showed 98.90% as an accuracy rate, and the forehead, eyes, and mouth were considered as having the highest distinctive rate. The proposed DLTP surpassed other related state-of-the-art techniques for extracting features regarding gender identification accuracy. For further work, it is envisioned that this model could extensively further to facial recognition and other facial characterization generally such as age detection to determine the most distinguishable and time-invariant facial components. Table 2 and figure 46, show the outcome rate of the gender classification. The experimental results prove the uses of UOFG dataset with SVM classifier on 4 facial components (forehead, eyes, nose, and mouth) having the highest recognition. Generally, the highest recognition was achieved without applying the method on the whole face but just some component. This adduced that some facial components are more distinguishable for gender identification of

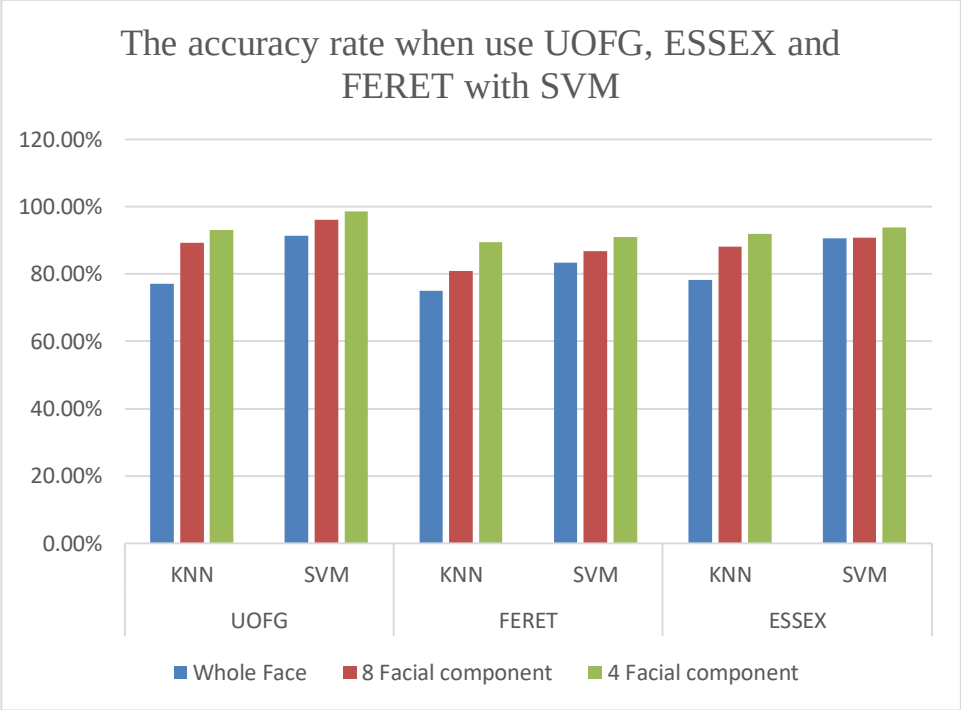
human. There is not much problem with time complexity, because of multi-processing ability that is presently available.

The image bellow illustrates the sample of image from FERET dataset (the original image) and the same image when some pre-processing was included (face detection, grey scale, histogram equalization and facial components detection).

Figure 45: (1) Original image (2) resizing (3) gray scale (4) histogram (5) component face

Table 3: The accuracy rate when use UOFG, ESSEX and FERET with SVM

DATASET	FERET		UOFG		ESSEX	
Classifier	**KNN**	**SVM**	**KNN**	**SVM**	**KNN**	**SVM**
Whole Face	**77.09%**	**91.35%**	**75.08%**	**83.44%**	**78.22%**	**90.53%**
8 Facial component	**89.25%**	**96.16%**	**80.90%**	**86.85%**	**88.21%**	**90.85%**
4 Facial component	**93.12%**	**98.55%**	**89.55%**	**91.00%**	**92.03%**	**93.93%**



6.3.4 Dynamic Local Ternary Patterns for Gender Identification using Facial Components

Here Dynamic Local Ternary Pattern (DLTP) was developed as a technique to improve gender identification accuracy. The results show there are some regions in the face is not necessary. Also, discriminative Response Map Fitting model was used for detecting facial component which are eyes, forehead, cheeks, mouth, nose and chin. The experimental results performed on the frontal face dataset of LFW showed 98.90% as an accuracy rate, and the forehead, eyes, and mouth are considered as having the highest distinctive rate. The proposed DLTP surpassed other related state-of-the-art techniques for extracting features regarding gender identification accuracy. For further work, it is envisioned that this model can extensively further facial recognition and other facial characterization generally such as age detection to determine the most distinguishable and time-invariant facial components.

Table 4: The result component using four feature extraction techniques

Facial component	Feature Extraction Techniques			
	LBP	**LDP**	**LTP**	**DLTP**
Nose	77.09%	77.05%	78.04%	80.03%
Mouth	86.06%	85.35%	87.50%	87.90%
Eyes	86.12%	88.55%	89.40%	89.70%
Forehead	87.32%	90.35%	90.90%	90.10%
Chin	78.12%	81.55%	81.89%	84.02%
Cheeks	82.12%	82.50%	81.95%	84.40%

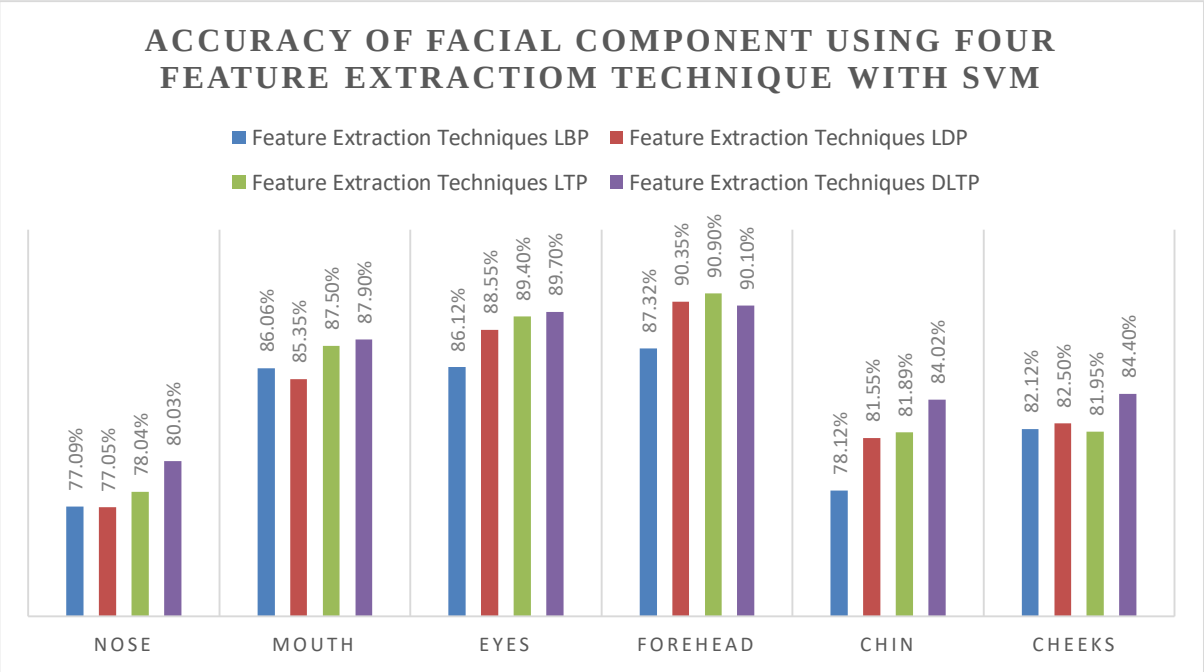


Table 4 compares our average gender identification rate with the results achieved by the connected studies on gender identification. From the results, we can see that this study has succeeded getting better results compared to other related works. In table 7 we can see advantages and disadvantages of LBP, LDP, LTP and proposed LTP

Table 5: Gender identification accuracy using DLTP and SVM

Facial Component	Accuracy
Forehead + Eyes	92.40%
Eyes + Mouth	92.67%
Eyes + Nose +Mouth	91.90%
Forehead +Eyes + Mouth	98.90%

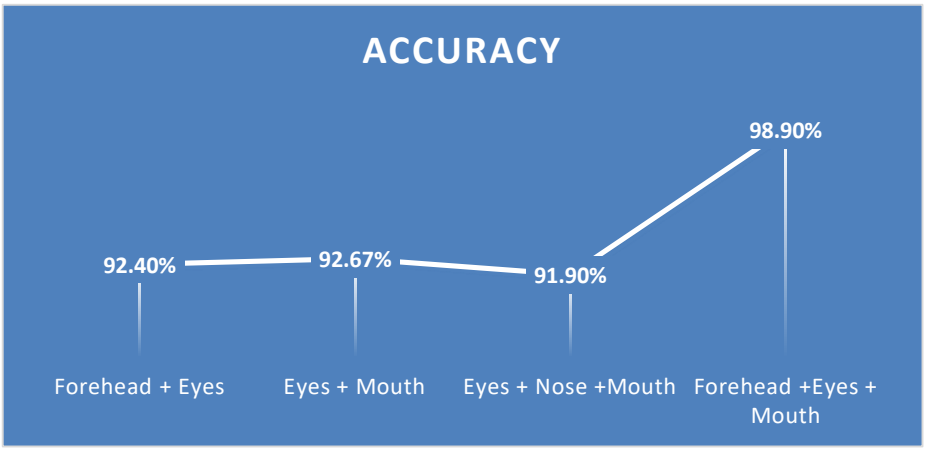


Table 6: Comparison feature extraction result

Method	Accuracy
LBP	90.11%
LDP	93.71%
LTP	93.61%
Proposed DLTP	**98.10%**

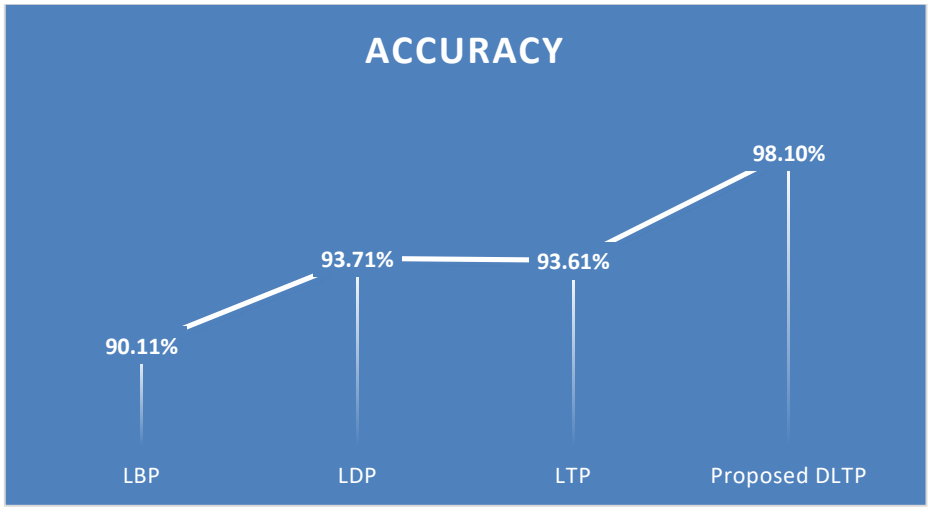


Table 7: The comparison of LBP, LDP, LTP and DLTP

Feature extraction method	Advantages	Disadvantages
LBP	Use gray-level intensity and computational efficient.	Extract the information using only two directions and random noise and high error when background change.
LDP	Directional edge response values and gery level intensity	Inconsistent in uniform and near uniform regions also heavily depend on the number of prominent edge parameters.
LTP	It is added an extra discrimination level	It has static threshold.
Proposed LTP	Increasing the direction of the pattern and it added dynamic threshold.	Need more tested for more images and it can be implemented with suitable hardware.

6.3.5 Gender classification model based on deep convolutional neural network

1. In this part, a methodology has been proposed to perform pre-training of convolution neural network (CNN) frameworks which were implemented for classification gender of humans from FERET facial images. The pre-trained CNN models were applied for extracting features (Resnet-50 and Resnet-101) then SVM were used to classify gender from face images. Using confusion matrix to evaluate the efficiency of pre-trained convolutional neural network for the classification based on accuracy, which have demonstrated the high recognition rates of the proposed pre-trained convolutional neural network. It gave a very accurate rate at about 98.60% for results in gender classification from whole region of face. The table 8 below display the differences of use in resnet-50 and resnet-101.
2. Table 8: the comparison of resnet-50 and resnet-101

Resnet-50	**Resnet-101**
It consists of about 177 layers Trained on 1.28 million training images in 1000 classes Used in object detection like face or bird detection	It consists of about 347 layers Trained on 1.00 million 1000 object categories Used in accurate system like brain tumour diagnoses

We showed that the pre trained CNN models + SVM obtained a more accurate rate when compared with CNN model + Soft-max classifier which obtained 90.3%. The result is illustrated in figure 48 below.

Figure 46: The CNN using the softmax classifier

Although CNN algorithm is just used in the second step (feature extraction step), it could obtain better results for gender recognition when transfer learning techniques are used. Transfer learning with a pre-trained model that is more relevant to the task, such as Resnet-50 and Reanet-101, can produce results in gender identification, which even might surpass human performance. We do not have to expend many days to train models from scratch to train techniques and study changes in network designs. Training step of this study done by using of a Resnet-50 and Resnet101. This study has identified the advantages presented by certain model designs, pre-trained weights, and training techniques. It has also showed that hierarchies of AI models offer promise and must be considered at the implementation of a classification system, as shown in figure 49 and table 8.

Table 9: The result using CNN

Method	Accuracy
Resnet50+SVM	98.6%
Resnet101+SVM	97.3%

The table below illustrate comparison of some study with the same database (FERET) and our study.

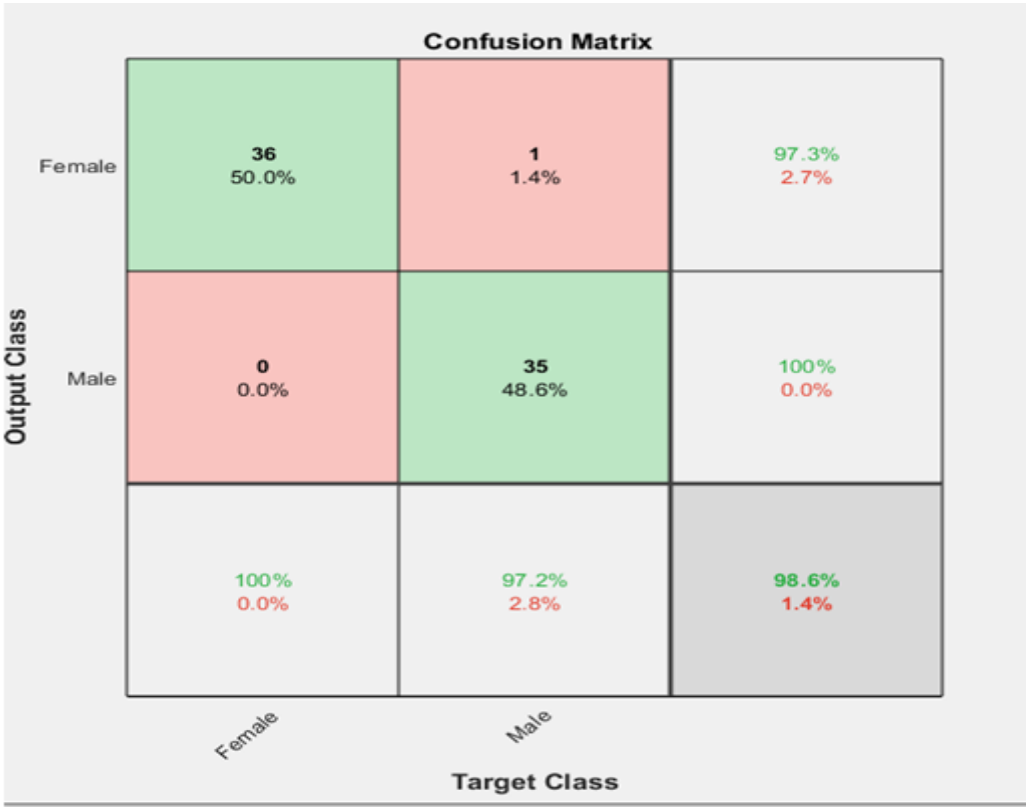


Figure 47: The result achieve using SVM

Table 10: Benchmarking results based on FERET face database

Study	Accuracy
Biradar, K.M et al. 2017 [22]	100%
Abikoye, O.C et al. 2019 [23]	91.84%
Shoyemi, I.O et al.2019 [24]	83.60%
Tiagrajah V. J et.al 2019 [29]	85%

Our study 2020	**98.60%**

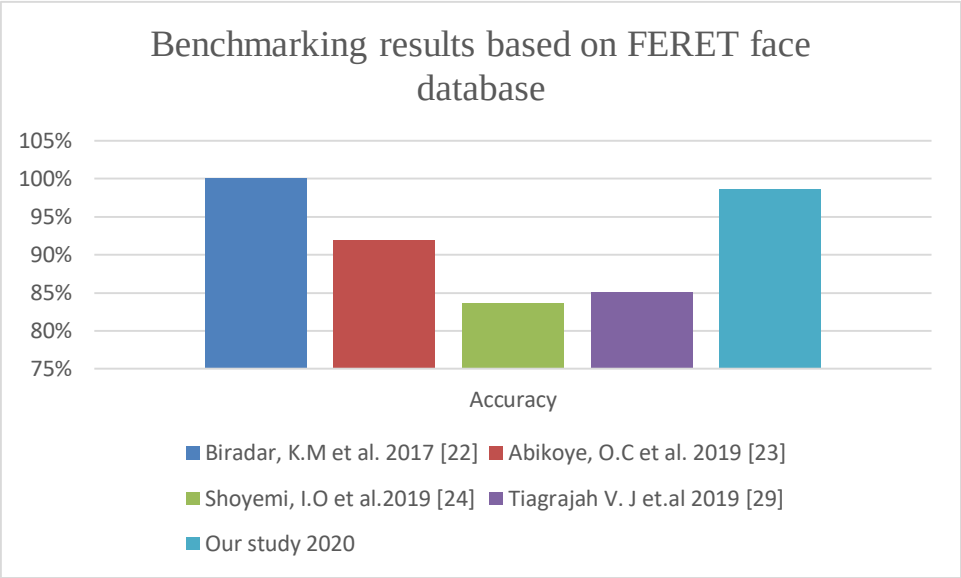


Table 11: the overall results

when using RERET datasets and SVM classifier with different feature extraction methods

Whole face	Component of the face	Proposed LTP	CNN
95%	**98.55%**	**98.90**	**98.60%**
(DCT&DWT)	**LBP**	**DLTP**	**Resnet-50**

6.4 CONCLUSION

In this chapter, the summary of all the steps and studies of gender identification have been presented in our study beginning from surveying of the state of the art, identifying the gender from whole face and identifying the gender from components of the face. The parameter to the components to improve the results was added. Lastly, deep learning to identify the gender is applied.

CHAPTER SEVEN
CONCLUSION AND FUTURE WORK

In this chapter we will show and summarize our experiment in gender identification until the conclusion of the study.

7.1 CONCLUSION

Firstly, pre-processing, face detection, feature extraction and classification steps were used to review the general system in gender identification method. Performing gender classification on pixels is more expensive; therefore, gender classification prefers to extract face features instead of working on pixels directly. Also, it provides a survey on several preceding techniques algorithms of gender identification, which were proposed earlier by researchers for further development in the classification field. It also presents a general explanation of some of the proceeding research in gender identification. Gender identification in entirely unrestricted settings makes it a very fascinating task in all its steps.

Secondly, feature extraction is an essential stage of gender classification problem like other numbers of data classification problems. This research evaluated the efficiency of four methods of (DCT Zigzag, Block Based DCT Zigzag, DWT Zigzag and Hybrid DWT-DCT Zigzag) feature extraction algorithms to realize useful features for the purpose of attaining maximum classification accuracies. All these methods were applied on ESSEX and FERED frontal face images datasets for evaluation. The results have proven that the Block Based DCT Zigzag coefficients achieved the highest accuracy compared to other feature extraction methods. The performance of recognition (classification) is not only based on feature extraction approaches, but also on the other stages of the recognition processes like the pre-processing stage and the classification algorithm.

Thirdly, a gender classification method based on the whole face and face component was presented. The data was passed to the face detector technique in the case of whole face. The Viola and Jones applied the crop to the region of the face and illuminated the background from the image. In the case of the face component, the image was passed with the DRMF model to detect sum of the region of interest rather than detecting the whole face. In the second step, the data detected in the first step was passed with the LBP texture feature extraction technique to extract the vector of the feature then it passed the vector to the classifier. We used the KNN and SVM classifier for this.

Next, via deep testing on the FERET frontal face dataset, it was confirmed that the dynamic local ternary pattern (DLTP) is an exact and effective technique to extract features most appropriate for the task of gender identification from face images. The improved DLTP method was applied, tested, and shown to further promote recognition accuracies and effectiveness. When we compared techniques from present literature, our proposed method was proven to outperform other progressive methods in terms of accuracy. In the coming time, improved DLTP can be combined with multiple feature sets to further enhance accuracy.

Lastly, here the methodology for the purpose of pre-training convolution neural network (CNN) architectures was proposed, which were used for classification of human gender classification from FERET facial images. We use Resnet-50 and Resnet-101 for feature extraction part and SVM in the classification part from facial images. The performance of pre-trained convolutional neural network applied to classify gender was evaluated based on accuracy from the Confusion Matrix. The results achieved have demonstrated the high-rate recognition of the proposed pre-trained network. It gave a very accurate rate of about 98.60% as a result of gender classification, the study showed that the combination of the pre trained CNN models with SVM achieved accurate rates.

Despite that, the CNN algorithm is just used in the second step (feature extraction step), but we can still acquire perfect results for gender recognition when transfer learning techniques were used. Transfer learning with a more relevant pre-trained model to the task, like resnet-50 and reanet-101, can produce better results in gender identification even possibly to exceed human performance. The training of the model is very fast so no need to spend more time to study variation in network designs and training

techniques. The study pointed out the benefits of the given model designs, pre-trained weights and training techniques. It has also shown that hierarchies of AI models offer promise and should be considered when applying a classification method.

7.2 FUTURE WORK

We worked on facial images mainly in gender identification. This topic is very current, so it needs additional study.

- The Sudanese dataset UOFG was collected and used in the study, and it achieved a good accuracy, so the recommendation is to finding a mechanism to enhance it to be standard and known in the field of image processing.
- Improve the DLTP; it can be combined with multiple datasets and another classifier to further improve accuracy rate.
- Work in deep learning using facial component to achieve accurate result and improve the general viewing in this area.
- Other CNN models can also be applied for more examination. Inception ResnetV2, InceptionV3, Alex-Net and Dense-Net are other proposed models.
- Here I just used the accuracy rate to evaluate our study, so it can be using another measurement method like execution time or robustness to increase the reliability of our study.

REFERENCES

Ahonen, Timo and Hadid, Abdenour and Pietik, 2004. Face recognition with local binary patterns. In: *European conference on computer vision.* s.l.:Springer, pp. 469--481.

Ahonen, Hadid, T. a., Pietik{\"a}inen, A. a. & Matti, 2004. Face recognition with local binary patterns. In: *European conference on computer vision.* s.l.:spriger, pp. 469--481.

Akanchha, Gour and others, 2016. Roy Increasing Acuuracy of Age and Gender Detection by Fingerprint Analysis Using DCT and DWT. *International Journal of Innovative Research in computer and communication Engineering.*

Akanchha, Gour and others, 2016. Roy Increasing Acuuracy of Age and Gender Detection by Fingerprint Analysis Using DCT and DWT. *International Journal of Innovative Research in computer and communication Engineering.*

Alain, Guillaume and Bengio, Yoshua, 2014. What regularized auto-encoders learn from the data-generating distribution. *The Journal of Machine Learning Research,* Volume 15, pp. 3563--3593},.

Alam, Rocky, M. M. a. & others, S. R. a., 2016. *Gender detection from frontal face images,* s.l.: BRAC university.

Arel, Rose, I. a., Karnowski, D. C. a. & P, T., 2010. Deep machine learning-a new frontier in artificial intelligence research [research frontier. *IEEE computational intelligence magazine,* Volume 5, pp. 13--18.

Atharifard, Ali and Ghofrani, Sedigheh, 2011. Robust component-based face detection using color feature. In: *Proceedings of the World Congress on Engineering.* s.l.:s.n., pp. 6--8.

Ba, Jimmy and Frey, Brendan, 2013. Adaptive dropout for training deep neural networks. In: *Advances in neural information processing systems.* s.l.:s.n., pp. 3084--3092.

Baldi, P. a. S. P. J., 2013. Understanding dropout. In: *Advances in neural information processing systems.* s.l.:s.n., pp. 2814--2822.

Balzer, Robert, 1985. A 15 year perspective on automatic programming. *IEEE Transactions on Software Engineering,* Volume 11, pp. 1257--1268.

Barrena, Jordina Torrents and Valls, Dom{\`e}nec Puig, 2014. Tumor Mass Detection through Gabor Filters and Supervised Pixel-Based Classification in Breast Cancer. *University Rovira, Virgil.*

Barrena, Jordina Torrents and Valls, Dom{\`e}nec Puig, 2014. Tumor Mass Detection through Gabor Filters and Supervised Pixel-Based Classification in Breast Cancer. *University Rovira, Virgil.*

Belhumeur, Hespanha, P. N. a., Kriegman, J. P. a. & J, D., 1997. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on pattern analysis and machine intelligence,* Volume 19, pp. 711-720.

Bengio, Yoshua, 2009. *Learning deep architectures for AI.* s.l.:Now Publishers Inc.

Bengio, Yoshua, 2013. Deep learning of representations: Looking forward. In: *International Conference on Statistical Language and Speech Processing.* s.l.:SPRIGER, pp. 1--37.

Bengio, Yoshua, 2013. Deep learning of representations: Looking forward. In: *International Conference on Statistical Language and Speech Processing.* s.l.:Springer, pp. 1--37.

Bengio, Courville, Y. a., Vincent, A. a. & Pascal, 2013. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence,* Volume 35, pp. 1798--1828.

Bengio, Courville, Y. a., Vincent, A. a. & Pascal, 2013. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence,* Volume 35, pp. 1798--1828.

Berretti, et al., 2013. Geometric histograms of 3D keypoints for face identification with missing parts. In: *Proceedings of the Sixth Eurographics Workshop on 3D Object Retrieval.* s.l.:s.n., pp. 57--64.

Bordes, et al., 2012. *Joint learning of words and meaning representations for open-text semantic parsing.* s.l., s.n., pp. 127--135.

Bordes, et al., 2012. Joint learning of words and meaning representations for open-text semantic parsing. In: *Artificial Intelligence and Statistics.* s.l.:s.n., pp. 127--135.

Boureau, et al., 2010. A theoretical analysis of feature pooling in visual recognition. In: *Proceedings of the 27th international conference on machine learning (ICML-10.* s.l.:s.n., pp. 111--118.

Buchala, et al., 2005. Principal component analysis of gender, ethnicity, age and identity of face images. *Procs of IEEE ICMI 2005.*

Burges, Scholkopf, C. J. a., Smola, B. a. & J, A., 1999. *Advances in kernel methods: support vector learning.* s.l.:MIT press Cambridge, MA, USA.

Burkert, et al., 2015. Dexpression: Deep convolutional neural network for expression recognition. *arXiv preprint arXiv:1509.05371.*

Can, Kocaman, R. a., Gokceoglu, S. a. & Candan, 2019. A convolutional neural network architecture for auto-detection of landslide photographs to assess citizen science and volunteered geographic information data quality. *ISPRS International Journal of Geo-Information,* Volume 8, p. 300.

Carreira-Perpinan, M and Hinton, G, 2005. On contrastive divergence learning. 10th Int. In: *Workshop on Artificial Intelligence and Statistics (AISTATS'2005).* s.l.:s.n.

Carreira-Perpinan, M and Hinton, G, 2005. On contrastive divergence learning. 10th Int. In: *Workshop on Artificial Intelligence and Statistics (AISTATS'2005.* s.l.:s.n.

Carreira-Perpinan, Miguel and Wang, Weiran, 2014. Distributed optimization of deeply nested systems. In: *Artificial Intelligence and Statistics.* s.l.:s.n., pp. 10--19.

Caruana, Rich, 2012. A dozen tricks with multitask learning. In: *Neural Networks: Tricks of the Trade.* s.l.:spriger, pp. 163--189.

Chang, Bowyer, K. I. a., Flynn, K. W. a. & J, P., 2005. An evaluation of multimodal 2D+ 3D face biometrics. *IEEE transactions on pattern analysis and machine intelligence,* Volume 27, pp. 619--624.

Chan, Kittler, C.-H. a. & Kieron, J. a. M., 2007. Multi-scale local binary pattern histograms for face recognition. In: *International conference on biometrics.* s.l.:Springer, pp. 809--818.

Chen, et al., 2013. Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification. In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* s.l.:s.n., pp. 3025--3032.

Chen, et al., 2013. Scalable face image retrieval using attribute-enhanced sparse codewords. *IEEE Transactions on Multimedia,* Volume 15, pp. 1163--1173.

Chen, et al., 2017. Vehicle type classification based on convolutional neural network. In: *2017 Chinese Automation Congress (CAC).* s.l.:IEEE, pp. 1898--1901.

Cho, Raiko, K. a., Ilin, T. a. & Alexander, 2011. Enhanced gradient and adaptive learning rate for training restricted Boltzmann machines. In: *ICML.* s.l.:s.n.

Cho, et al., 2013. A two-stage pretraining algorithm for deep boltzmann machines. In: *International Conference on Artificial Neural Networks.* s.l.:springer, pp. 106--113.

Cho, et al., 2013. A two-stage pretraining algorithm for deep boltzmann machines. In: *International Conference on Artificial Neural Networks.* s.l.:spriger, pp. 106--113.

Cire{\c{s}}an, et al., 2011. High-performance neural networks for visual object classification. *arXiv preprint arXiv:1102.0183.*

Cire{\c{s}}an, Meier, D. C. a., Schmidhuber, U. a. & J{\"u}rgen, 2012. Transfer learning for Latin and Chinese characters with deep neural networks. In: *The 2012 international joint conference on neural networks (IJCNN.* s.l.:IEEE, pp. 1--6.

Ciregan, Meier, D. a., Schmidhuber, U. a. & J{\"u}rgen, 2012. Multi-column deep neural networks for image classification. In: *2012 IEEE conference on computer vision and pattern recognition.* s.l.:IEEE, pp. 3642--3649.

Ciregan, Meier, D. a., Schmidhuber, U. a. & J{\"u}rgen, 2012. Multi-column deep neural networks for image classification. In: *2012 IEEE conference on computer vision and pattern recognition.* s.l.:s.n., p. 2012 IEEE conference on computer vision and pattern recognition.

Collins, et al., 2010. EigenBody: analysis of body shape for gender from noisy images. In: *International machine vision and image processing conference.* s.l.:s.n.

Colombo, Cusano, A. a., Schettini, C. a. & Raimondo, 2006. 3D face detection using curvature analysis. *Pattern recognition,* Volume 39, pp. 444--455.

Cui, et al., 2013. Fusing robust face region descriptors via multiple metric learning for face recognition in the wild. In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* s.l.:s.n., pp. 3554--3561.

Dan C and Meier, Ueli and Schmidhuber, 2012. *Transfer learning for Latin and Chinese characters with deep neural networks.* s.l., The 2012 international joint conference on neural networks (IJCNN), pp. 1--6.

Davis, Jesse and Goadrich, Mark, 2006. The relationship between Precision-Recall and ROC curves. In: *Proceedings of the 23rd international conference on Machine learning.* s.l.:s.n., pp. 233--240.

Deng, Li, 2014. A tutorial survey of architectures, algorithms, and applications for deep learning. *APSIPA Transactions on Signal and Information Processing,* 3(Cambridge University Press).

Deng, Li, 2014. A tutorial survey of architectures, algorithms, and applications for deep learning. *APSIPA Transactions on Signal and Information Processing,* 3(Cambridge University Press).

Deville, Yves and Lau, Kung-Kiu, 1994. Logic program synthesis. *The Journal of Logic Programming,* Volume 19, pp. 321--350.

Dhankhar, Poonam, 2019. ResNet-50 and VGG-16 for recognizing Facial Emotions. *International Journal of Innovations in Engineering and Technology (IJIET),* Volume 13, pp. 126--130.

Ding, Zhang, Y.-S. a., Chou, T.-L. a. & Kuo-Chen, 2007. Prediction of protein structure classes with pseudo amino acid composition and fuzzy support vector machine network. *Protein and peptide letters,* Volume 14, pp. 811--815.

Dosovitskiy, Springenberg, A. a., Brox, J. T. a. & Thomas, 2013. Unsupervised feature learning by augmenting single images. *arXiv preprint arXiv:1312.5242.*

Du, Salah, H. a., Ahmed, S. H. a. & O, H., 2014. A color and texture based multi-level fusion scheme for ethnicity identification. In: *Mobile Multimedia/Image Processing, Security, and Applications 2014.* s.l.:International Society for Optics and Photonics, p. 91200B.

Du, Salah, H. a., Ahmed, S. H. a. & O, H., 2014. A color and texture based multi-level fusion scheme for ethnicity identification. In: *Mobile Multimedia/Image Processing, Security, and Applications 2014.* s.l.:International Society for Optics and Photonics, p. 91200B.

Ekman, et al., 1993. Final report to NSF of the planning workshop on facial expression understanding. *Human Interaction Laboratory, University of California, San Francisco,* p. 378.

Ekman, et al., 1993. Final report to NSF of the planning workshop on facial expression understanding. *Human Interaction Laboratory, University of California, San Francisco,* p. 378.

Elfwing, Uchibe, S. a., Doya, E. a. & Kenji, 2015. Expected energy-based restricted Boltzmann machine for classification. *Neural networks,* Volume 64, pp. 29--38.

Erhan, et al., 2010. Why does unsupervised pre-training help deep learning?. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics.* s.l.:s.n., pp. 201--208.

Gidaris, Spyros and Komodakis, Nikos, 2015. Object detection via a multi-region and semantic segmentation-aware cnn model. In: *Proceedings of the IEEE international conference on computer vision.* s.l.:s.n., pp. 1134--1142.

Girshick, et al., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* s.l.:s.n., pp. 580--587.

Glorot, g. M. Y. a. et al., 2012. Unsupervised and transfer learning challenge: a deep learning approach. In: *Proceedings of ICML Workshop on Unsupervised and Transfer Learning.* s.l.:s.n., pp. 97--110.

Goodfellow, Courville, I. J. a., Bengio, A. a. & Yoshua, 2013. Joint training deep boltzmann machines for classification. *arXiv preprint arXiv:1301.3568.*

Goodfellow, I. a. M., Courville, M. a., Bengio, A. a. & Yoshua, 2013. Multi-prediction deep Boltzmann machines. In: *Advances in Neural Information Processing Systems.* s.l.:s.n., p. Advances in Neural Information Processing Systems.

Goodfellow, et al., 2009. Measuring invariances in deep networks. In: *Advances in neural information processing systems.* s.l.:s.n., pp. 646--654.

Gu, J. a. W. Z. a. K. J. a. M. L. a. S. et al., 2018. Recent advances in convolutional neural networks. *Pattern Recognition,* Volume 77, pp. 354--377.

Gutta, Srinivas and Wechsler, Harry, 1996. Face recognition using hybrid classifier systems. In: *Proceedings of International Conference on Neural Networks (ICNN'96).* s.l.:IEEE, pp. 1017--1022.

Han, Otto, H. a., Jain, C. a. & K, A., 2013. Age estimation from face images: Human vs. machine performance. In: *2013 International Conference on Biometrics (ICB.* s.l.:IEEE, pp. 1--8.

He, Kaiming and Sun, Jian, 2015. Convolutional neural networks at constrained time cost. In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* s.l.:s.n., pp. 5353--5360.

Heisele, Bernd and Blanz, Volker, 2006. Morphable models for training a component-based face recognition system. In: *Face Processing, Advanced Modeling and Methods.* s.l.:s.n., pp. 439--462.

Heiselet, et al., 2001. Component-based face detection. In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001.* s.l.:s.n., pp. I--I.

Heiselet, et al., 2001. Component-based face detection. In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001.* s.l.:IEEE, pp. I--I.

He, et al., 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern analysis and machine intelligence,* Volume 37, pp. 1904--1916.

He, et al., 2016. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* s.l.:s.n., pp. 770--778.

Hg, R. a. J. et al., 2012. An rgb-d database using microsoft's kinect for windows for face detection. In: *2012 Eighth International Conference on Signal Image Technology and Internet Based Systems.* s.l.:IEEE, pp. 42--46.

Hinton, Geoffrey E and Salakhutdinov, Ruslan R, 2006. Reducing the dimensionality of data with neural networks. *science*, Issue American Association for the Advancement of Science, pp. 504--507.

Hinton, Geoffrey E and Salakhutdinov, Russ R, 2012. A better way to pretrain deep boltzmann machines. In: *Advances in Neural Information Processing Systems.* s.l.:s.n., pp. 2447--2455.

Hinton, Geoffrey E and Sejnowski, Terrence J and others, 1986. Learning and relearning in Boltzmann machines. *Parallel distributed processing: Explorations in the microstructure of cognition*, Volume 1, pp. 282-317.

Hinton, Geoffrey E, 2012. A practical guide to training restricted Boltzmann machines. In: *Neural networks: Tricks of the trade.* s.l.:Springer, pp. 599--619.

Hinton, et al., 2012. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580.*

Howard, Andrew G, 2013. Some improvements on deep convolutional neural network based image classification. *arXiv preprint arXiv:1312.5402.*

Huang, Blanz, J. a., Heisele, V. a. & Bernd, 2002. Face recognition using component-based SVM classification and morphable models. In: *International Workshop on Support Vector Machines.* s.l.:Springer, pp. 334--341.

Huang, Blanz, J. a., Heisele, V. a. & Bernd, 2002. Face recognition using component-based SVM classification and morphable models. In: *International Workshop on Support Vector Machines.* s.l.:Springer, pp. 334--341.

Huang, Lee, G. B. a., Learned-Miller, H. a. & Erik, 2012. Learning hierarchical representations for face verification with convolutional deep belief networks. In: *2012 IEEE Conference on Computer Vision and Pattern Recognition.* s.l.:s.n., pp. 2518--2525.

Hussain, Napol{\'e}on, S. U. a., Jurie, T. a. & Fr{\'e}deric, 2012. *Face recognition using local quantized patterns.* s.l.:s.n.

Isa, Nurul Zarina Md, 2010. *Gender Recognition Based on Facial Image Extraction,* s.l.: UMP.

Jabid, Kabir, T. a., Chae, M. H. a. & Oksam, 2010. Local directional pattern (LDP) for face recognition. In: *2010 digest of technical papers international conference on consumer electronics (ICCE).* s.l.:s.n., pp. 329--330.

Jafri, Rabia and Arabnia, Hamid R, 2009. A survey of face recognition techniques. *journal of information processing systems,* 5(Korea Information Processing Society), pp. 41--68.

Jain, Huang, A. a., Fang, J. a. & Shiaofen, 2005. Gender identification using frontal facial images. In: *2005 IEEE International Conference on Multimedia and Expo.* s.l.:IEEE, pp. 4--pp.

Kaur, Manvjeet and others, 2012. K-nearest neighbor classification approach for face and fingerprint at feature level fusion. *Int. J. Comput. Appl,* Volume 60, pp. 13--17.

Keller, Gray, J. M. a., Givens, M. R. a. & A, J., 1985. A fuzzy k-nearest neighbor algorithm. *IEEE transactions on systems, man, and cybernetics,* Volume 4, pp. 580--585.

Khalil-Hani, Mohamed and Sung, Liew Shan, 2014. A convolutional neural network approach for face verification. In: *2014 International Conference on High Performance Computing \& Simulation (HPCS).* s.l.:s.n., pp. 707--714.

Khayam, Syed Ali, 2003. The discrete cosine transform (DCT): theory and application. *Michigan State University,* Volume 114, pp. 1--31.

kinen, Erno and Raisamo, Roope, 2008. An experimental comparison of gender classification methods. *pattern recognition letters,* Volume 29, pp. 1544--1556.

kinen, Erno, 2007. *Face Analysis Techniques for Human-Computer Interaction.* s.l.:Tampere University Press.

Klare, et al., 2012. Face recognition performance: Role of demographic information. *IEEE Transactions on Information Forensics and Security,* Volume 7, pp. 1789--1801.

Krizhevsky, Sutskever, A. a., Hinton, I. a. & E, G., 2012. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems.* s.l.:s.n., pp. 1097--1105.

Krizhevsky, Sutskever, A. a., Hinton, I. a. & E, G., 2012. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems.* s.l.:s.n., pp. 1097--1105.

Kumar, et al., 2009. Attribute and simile classifiers for face verification. In: *2009 IEEE 12th international conference on computer vision.* s.l.:IEEE, pp. 365--372.

Lawgali, Ahmed, 2005. *Handwritten Digit Recognition based on DWT and DCT,* s.l.: University of Benghazi.

Lawgali, et al., 2015. *Handwritten Arabic character recognition: Which feature extraction method.* s.l.:University of Benghazi.

Le, Quoc V, 2013. Building high-level features using large scale unsupervised learning. In: *2013 IEEE international conference on acoustics, speech and signal processing.* s.l.:IEEE, pp. 8595--8598.

LeCun, Yann, 2012. Learning invariant feature hierarchies. In: *European conference on computer vision.* s.l.:Springer, pp. 496--505.

LeCun, et al., 1989. Gradient-based learning applied to document recognition. *Proceedings of the IEEE,* Volume 86, pp. 2278--2324.

Lee, Ekanadham, H. a., Ng, C. a. & Y, A., 2008. Sparse deep belief net model for visual area V2. In: *Advances in neural information processing systems.* s.l.:s.n., pp. 873--880.

Lee, Ekanadham, H. a., Ng, C. a. & Y, A., 2008. Sparse deep belief net model for visual area V2. In: *Advances in neural information processing systems.* s.l.:s.n., pp. 873--880.

Lee, et al., 2009. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In: *Proceedings of the 26th annual international conference on machine learning.* s.l.:s.n., pp. 609--616.

Lei, et al., 2014. An efficient 3D face recognition approach using local geometrical signatures. *Pattern Recognition,* Volume 47, pp. 509--524.

Le, et al., 2011. On optimization methods for deep learning. In: *ICML*. s.l.:s.n.

Le, Tang, V. a., Huang, H. a. & S, T., 2011. Expression recognition from 3D dynamic faces using robust spatio-temporal shape features. In: *Face and Gesture 2011*. s.l.:IEEE, pp. 414--421.

Liew, et al., 2016. Gender classification: a convolutional neural network approach. *Turkish Journal of Electrical Engineering \& Computer Sciences,* 24(The Scientific and Technological Research Council of Turkey), pp. 1248-1264.

Li, Lian, B. a., Lu, X.-C. a. & Bao-Liang, 2012. Gender classification by combining clothing, hair and facial component classifiers. *Neurocomputing,* Volume 76, pp. 18--27.

Lin, Min and Chen, Qiang, 2013. Shuicheng Yan: Network In Network. *arXiv preprint arXiv:1312.4400.*

Lin, Lu, H. a., Zhang, H. a. & Lihe, 2006. A new automatic recognition system of gender, age and ethnicity. In: *2006 6th world congress on intelligent control and automation.* s.l.:IEEE, pp. 9988--9991.

Liou, et al., 2014. Autoencoder for words. *Neurocomputing,* Volume 139, pp. 84--96.

Long, Shelhamer, J. a., Darrell, E. a. & Trevor, 2015. Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* s.l.:s.n., pp. 3431--3440.

Long, Shelhamer, J. a., Darrell, E. a. & Trevor, 2015. Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* s.l.:s.n., pp. 3431--3440.

Lu, Xiaoguang, 2003. Image analysis for face recognition. *Personal notes,* Issue Citeseer, p. 36.

Miclut, Bogdan, 2014. Committees of deep feedforward networks trained with few data. In: *German Conference on Pattern Recognition.* s.l.:springer, pp. 736--742.

Mikolov, et al., 2013. Distributed representations of words and phrases and their compositionality. In: *Advances in neural information processing systems.* s.l.:s.n., pp. 3111--3119.

Mikolov, et al., 2013. Distributed representations of words and phrases and their compositionality. In: *Advances in neural information processing systems.* s.l.:s.n., pp. 3111--3119.

Milborrow, Morkel, S. a., Nicolls, J. a. & Fred, 2010. The MUCT landmarked face database. *Pattern Recognition Association of South Africa,* Volume 201.

Minsky, M, 1963. Steps toward artificial intelligence. Comput. *Thought,* Volume 406, p. 450.

Mitchell, Tom M, n.d. *Machine Learning, volume 1 of 1.* 1997: McGraw-Hill Science/Engineering/-Math.

Moghaddam, Baback and Pentland, Alexander P, 1994. Face recognition using view-based and modular eigenspaces. In: *Automatic Systems for the Identification and Inspection of Humans.* s.l.:International Society for Optics and Photonics, pp. 12-21.

Moghaddam, Baback and Yang, Ming-Hsuan, 2002. Learning gender with support faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Volume 24, pp. 707--711.

Mohamed, Abdallah A and Yampolskiy, Roman V, 2012. Adaptive extended local ternary pattern (aeltp) for recognizing avatar faces. In: *2012 11th International Conference on Machine Learning and Applications.* s.l.:IEEE, pp. 57--62.

Mohamed, Salma and Nour, Nahla and Viriri, Serestina, 2018. Gender identification from facial images using global features. In: *2018 Conference on Information Communications Technology and Society (ICTAS).* s.l.:s.n., pp. 1--6.

Montavon, Orr, G. a., M{\"u}ller, G. a. & Klaus-Robert, 2012. *Neural networks: tricks of the trade.* s.l., springer.

Mousa Pasandi, Mohammad Esmaeel, 2014. *Face, Age and Gender Recognition using Local Descriptors.* s.l.:University of Ottawa.

Nagi, Ahmed, J. a., Nagi, S. K. a. & Farrukh, 2008. Pose Invariant face recognition using Hybrid DWT-DCT frequency features with Support Vector Machines. In: *Proceedings of the 4th International Conference Information Technology and Multimedia.* s.l.:s.n., pp. 99--104.

Nair, Vinod and Hinton, Geoffrey E, 2010. Rectified linear units improve restricted boltzmann machines. In: *ICML.* s.l.:s.n.

Nazir, et al., 2010. Feature selection for efficient gender classification. In: *Proceedings of the 11th WSEAS international conference.* s.l.:s.n., pp. 70--25.

nchez-Delacruz, Eddy and Parra, Pilar Pozos, 2018. Machine learning-based classification for diagnosis of neurodegenerative diseases. In: *LANMR.* s.l.:s.n., pp. 40--50.

Ng, Epin, 2015. *Gender classification from facial images.* s.l.:UTAR.

Ngiam, et al., 2011. Learning deep energy models. In: *ICML.* s.l.:s.n.

Ng, Tay, C. B. a., Goi, Y. H. a. & Min, B., 2012. Vision-based human gender recognition: A survey. *arXiv preprint arXiv:1204.1611.*

Ojala, Pietik{\"a}inen, T. a., Harwood, M. a. & David, 1996. A comparative study of texture measures with classification based on featured distributions. *Pattern recognition*, Volume 29, pp. 51--59.

Oquab, et al., 2014. Learning and transferring mid-level image representations using convolutional neural networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. s.l.:s.n., pp. 1717--1724.

Ouyang, W. a. L. P. a. Z. X. a. Q. S. a. T. Y. a. L. et al., 2014. Deepid-net: multi-stage and deformable deep convolutional neural networks for object detection. *arXiv preprint arXiv:1409.3505*.

Ozbudak, Tukel, O. a., Seker, M. a. & S, 2010. Fast gender classification. In: *2010 IEEE International Conference on Computational Intelligence and Computing Research*. s.l.:s.n., pp. 1--5.

Phillips, Wechsler, P. J. a., Huang, H. a. & J, J. a. R. P., 1998. The FERET database and evaluation procedure for face-recognition algorithms. *Image and vision computing*, 16(Elsevier), pp. 295--306.

Pianykh, Oleg S, 2009. *Digital imaging and communications in medicine (DICOM): a practical introduction and survival guide*. s.l.:Springer Science \& Business Media.

Qacimy, E., Kerroum, B. a., Hammouch, M. A. a. & Ahmed, 2014. Feature extraction based on DCT for handwritten digit recognition. *International Journal of Computer Science Issues (IJCSI)*, 11(Citeseer), p. 27.

Rai, Preeti and Khanna, Pritee, 2014. A gender classification system robust to occlusion using Gabor features based (2D) 2PCA. *Journal of Visual Communication and Image Representation*, Volume 25, pp. 1118--1129.

Rai, Preeti and Khanna, Pritee, 2014. A gender classification system robust to occlusion using Gabor features based (2D) 2PCA. *Journal of Visual Communication and Image Representation*, Volume 25, pp. 1118--1129.

Ranzato, et al., 2007. Efficient learning of sparse representations with an energy-based model. In: *Advances in neural information processing systems}.* s.l.:s.n., pp. 1137--1144.

Ranzato, et al., 2007. Efficient learning of sparse representations with an energy-based model. In: *Advances in neural information processing systems.* s.l.:s.n., pp. 1137--1144.

Ravi, S and Wilson, S, 2010. Face detection with facial features and gender classification based on support vector machine. *International Journal of Imaging Science and Engineering,* pp. 23--28.

Refaeilzadeh, Tang, P. a., Liu, L. a. & Huan, 2009. Cross-Validation. *Encyclopedia of database systems,* Volume 5, pp. 532--538.

Ren, Jimmy SJ and Xu, Li, 2015. On vectorization of deep convolutional neural networks for vision tasks. In: *Twenty-Ninth AAAI Conference on Artificial Intelligence.* s.l.:s.n.

Rifai, et al., 2011. Contractive auto-encoders: Explicit invariance during feature extraction. In: *Icml.* s.l.:s.n.

Russakovsky, O. a. D. J. a. S. H. a. K. J. a. S. S. a. M. et al., 2015. Imagenet large scale visual recognition challenge. *International journal of computer vision,* Volume 115, pp. 211--252.

Russakovsky, O. a. D. J. a. S. H. a. K. J. a. S. S. a. M. et al., 2015. Imagenet large scale visual recognition challenge. *International journal of computer vision,* Volume 115, pp. 211--252.

Russakovsky, O. a. D. J. a. S. H. a. K. J. a. S. S. a. M. et al., 2015. Imagenet large scale visual recognition challenge. *International journal of computer vision,* Volume 115, pp. 211--252.

Salakhutdinov, Ruslan and Larochelle, Hugo},, 2010. Efficient learning of deep Boltzmann machines. In: *Proceedings of the thirteenth international conference on artificial intelligence and statistics.* s.l.:s.n., pp. 693--700.

Scherer, M{\"u}ller, D. a., Behnke, A. a. & Sven, 2010. Evaluation of pooling operations in convolutional architectures for object recognition. In: *International conference on artificial neural networks.* s.l.:SPRINGER, pp. 92--101.

Schmidhuber, J{\"u}rgen, 1990. *Dynamische neuronale Netze und das fundamentale raumzeitliche Lernproblem.* s.l.:Technische Universit{\"a}t M{\"u}nchen.

Schmidhuber, J{\"u}rgen, 2015. Deep learning in neural networks: An overview. *Neural networks,* Volume 61, pp. 85--117.

Serj, et al., 2018. A deep convolutional neural network for lung cancer diagnostic. *arXiv preprint arXiv:1804.08170.*

Shen, Linlin and Bai, Li, 2006. A review on Gabor wavelets for face recognition. *Pattern analysis and applications,* Volume 9, pp. 273--292.

Siegelmann, Hava T and Sontag, Eduardo D, 1991. Turing computability with neural nets. *Applied Mathematics Letters,* Volume 4, pp. 77--80.

Simoncelli, Eero P, 2005. 4.7 statistical modeling of photographic images. *Handbook of Video and Image Processing,* 9(Academic Press).

Simonyan, Karen and Zisserman, Andrew, 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556.*

Sinha, Chandrakamal, 2013. Gender classification from facial images using PCA and SVM. *Master of Technology thesis, National Institute of Technology Rourkela.*

Soloway, Elliot, 1986. Learning to program= learning to construct mechanisms and explanations. *Communications of the ACM,* Volume 29, pp. 850--858.

Song, et al., 2011. Contextualizing object detection and classification. In: *CVPR 2011.* s.l.:IEEE, pp. 1585--1592.

Sun, Bebis, Z. a., Miller, G. a. & Ronald, 2002. On-road vehicle detection using Gabor filters and support vector machines. In: *2002 14th International Conference on Digital*

Signal Processing Proceedings. DSP 2002 (Cat. No. 02TH8628). s.l.:IEEE, pp. 1019--1022.

Sun, Wang, Y. a., Tang, X. a. & Xiaoou, 2013. Deep convolutional network cascade for facial point detectio. In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* s.l.:s.n., pp. 3476--3483.

Sun, Wang, Y. a., Tang, X. a. & Xiaoou, 2013. Deep convolutional network cascade for facial point detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* s.l.:s.n., pp. 3476--3483.

Sun, et al., 2006. Gender classification based on boosting local binary pattern. In: *International symposium on neural networks.* s.l.:springer, pp. 194--201.

Szegedy, C. a. L. W. a. J. Y. a. S. P. a. R. et al., 2015. Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* s.l.:s.n., pp. 1--9.

Tan, Xiaoyang and Triggs, Bill, 2007. Fusing Gabor and LBP feature sets for kernel-based face recognition. In: *International workshop on analysis and modeling of faces and gestures.* s.l.:springer, pp. 235--249.

Tan, Xiaoyang and Triggs, Bill, 2010. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE transactions on image processing,* Volume 19, pp. 1635--1650.

Tariq, Hu, U. a., Huang, Y. a. & S, T., 2009. Gender and ethnicity identification from silhouetted face profiles. In: *2009 16th IEEE International Conference on Image Processing (ICIP).* s.l.:IEEE, pp. 2441--2444.

Timo, Ahonen, 2004. Face recognition with local binary patterns. In: *Euro. Conf. on Computer Vision.* s.l.:s.n.

Tin, Sein, H. H. K. a. & others, M. M. a., 2011. Race identification for face images. *ACEEE Int. J. Inform. Tech,* Volume 1, pp. 35--37.

Toderici, et al., 2010. Ethnicity-and gender-based subject retrieval using 3-D face-recognition techniques. *International Journal of Computer Vision,* Volume 89, pp. 382--391.

Tolba, Ahmad S, 2001. Invariant gender identification. *Digital Signal Processing,* 11(Elsevier), pp. 222--240.

Tseng, Paul, 2010. Approximation accuracy, gradient methods, and error bound for structured convex optimization. *Mathematical Programming,* Volume 125, pp. 263--295.

Vapnik, Valdimir N, 1995. The Nature of Statistical L earning T heory. *New York: Springer2V erlag.*

Vincent, et al., 2008. Extracting and composing robust features with denoising autoencoders. In: *Proceedings of the 25th international conference on Machine learning.* s.l.:s.n., pp. 1096--1103.

Vincent, et al., 2010. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of machine learning research,* Volume 11.

Viola, Paul and Jones, Michael, 2001. Rapid object detection using a boosted cascade of simple features. In: *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001.* s.l.:s.n., pp. I--I.

Wang, et al., 2014. Deep joint task learning for generic object extraction. In: *Advances in neural information processing systems.* s.l.:s.n., pp. 523--531.

Wan, et al., 2013. Regularization of neural networks using dropconnect. In: *International conference on machine learning.* s.l.:s.n., pp. 1058--1066.

Warde-Farley, et al., 2013. An empirical analysis of dropout in piecewise linear networks. *arXiv preprint arXiv:1312.6197.*

Weston, et al., 2012. Deep learning via semi-supervised embedding. In: *Neural networks: Tricks of the trade.* s.l.:springer, pp. Neural networks: Tricks of the trade},.

W, T.-X. a. L., Lu, X.-C. a. & Bao-Liang, 2012. Multi-view gender classification using symmetry of facial images. *Neural computing and applications,* Volume 21, pp. 661--669.

Wu, Wang, Y. a., Ji, Z. a. & Qiang, 2013. Facial feature tracking under varying facial expressions and face poses based on restricted boltzmann machines. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* s.l.:s.n., pp. 3452--3459.

Wu, et al., 2015. Deep image: Scaling up image recognition. *arXiv preprint arXiv:1501.02876,* Volume 7.

Xie, Saining and Tu, Zhuowen, 2015. Holistically-nested edge detection. In: *Proceedings of the IEEE international conference on computer vision.* s.l.:s.n., pp. 1395--1403.

Yang, Ming-Hsuan and Moghaddam, Baback, 2000. Support vector machines for visual gender classification. In: *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000.* s.l.:IEEE, pp. 1115--1118.

Yang, et al., 2009. Linear spatial pyramid matching using sparse coding for image classification. In: *2009 IEEE Conference on computer vision and pattern recognition.* s.l.:IEEE, pp. 1794--1801.

Yoo, et al., 2015. Multi-scale pyramid pooling for deep convolutional representation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops.* s.l.:s.n., pp. 71--80.

Younes, Laurent, 1999. On the convergence of Markovian stochastic algorithms with rapidly decreasing ergodicity rates. *Stochastics: An International Journal of Probability and Stochastic Processes,* Volume 65, pp. 177--228.

Yu, Zhang, K. a., Gong, T. a. & Yihong, 2009. Nonlinear learning using local coordinate coding. In: *Advances in neural information processing systems.* s.l.:s.n., pp. 2223--2231.

Zeiler, Matthew D and Fergus, Rob, 2013. Stochastic pooling for regularization of deep convolutional neural networks. *arXiv preprint arXiv:1301.3557.*

Zeiler, Matthew D, 2013. *Hierarchical convolutional deep learning in computer vision.* s.l.:New York University.

Zeng, Ouyang, X. a., Wang, W. a. & Xiaogang, 2013. Multi-stage contextual deep learning for pedestrian detection. In: *Proceedings of the IEEE International Conference on Computer Vision.* s.l.:s.n., pp. 121--128.

Zhang, et al., 2015. Improving object detection with deep convolutional networks via bayesian optimization and structured prediction. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* s.l.:s.n., pp. 249--258.

Zhou, et al., 2014. Is joint training better for deep auto-encoders?. *arXiv preprint arXiv:1405.1380.*

Zou, Ng, W. Y. a., Yu, A. Y. a. & Kai, 2011. Unsupervised learning of visual invariance with temporal coherence. In: *NIPS 2011 workshop on deep learning and unsupervised feature learning.* s.l.:s.n., p. 3.

Appendix

The result obtained when using CNN with SVM with FERET dataset to identify the gender

```
outputFolder = fullfile('dataset11');

 rootFolder=fullfile(outputFolder,'train');

 categoreis={   'Female','Male'};




imds=imageDatastore(fullfile(rootFolder,categoreis),'labelsource','fo
ldernames');


tbl = countEachLabel(imds);




minSetCount = min(tbl{:,2});

maxNumImages = 240;




imds = splitEachLabel(imds, minSetCount, 'randomize');

countEachLabel(imds);

net = resnet50();




net.Layers(1);

net.Layers(end);
```

```
numel(net.Layers(end).ClassNames);

[trainingSet, testSet] = splitEachLabel(imds, 0.85, 'randomize');



imageSize = net.Layers(1).InputSize;

augmentedTrainingSet = augmentedImageDatastore(imageSize, trainingSet,
'ColorPreprocessing', 'gray2rgb');

% Get the network weights for the second convolutional layer

w1 = net.Layers(2).Weights;



% Scale and resize the weights for visualization

w1 = mat2gray(w1);

w1 = imresize(w1,5);



% Display a montage of network weights. There are 96 individual sets
of

% weights in the first layer.

% figure

% montage(w1)

% title('First convolutional layer weights')



featureLayer = 'fc1000';



trainingFeatures = activations(net, augmentedTrainingSet,
featureLayer, 'OutputAs', 'rows');

trainingFeatures = double(trainingFeatures);

testFeatures = activations(net, augmentedTestSet, featureLayer,
'OutputAs', 'rows');
```

```
testFeatures = double(testFeatures);

% Get training and test labels from the trainingSet and testSet

trainingLabels = trainingSet.Labels;%YTrain

trainingLabels = cellstr(trainingLabels);

testLabels = testSet.Labels; %YTest

testLabels = cellstr(testLabels);

rng(1);

t = templateSVM('Standardize',1);

Md1 = fitcecoc(double(trainingFeatures),
cellstr(trainingLabels),'Learners',t, 'FitPosterior',1,
'ClassNames',{'Female','Male'});

% CVMdl = crossval(Md1);

% loss = kfoldLoss(CVMdl);

predictedLabels = predict(Md1, testFeatures);

[predictedLabels,~,~,Posterior] = predict(Md1,testFeatures);

[X,Y,T,AUC,OPTROCPT,SUBY,SUBYNAMES] = perfcurve(testLabels,
Posterior(:,2), 'Male');

figure(1)

plot(X,Y);

xlabel('False positive rate')

ylabel('True positive rate')

title('ROC for Classification CNN')

AUC

grid on



testLabel = testSet.Labels;



confMat = confusionmat(testLabel, predictedLabels);

plotconfusion(testLabel, predictedLabels);
```

```
confMat = bsxfun(@rdivide,confMat,sum(confMat,2));

Acu=mean(diag(confMat));


///
```

The accuracy when using CNN with softmax

```
outputFolder = fullfile('DataSet');

 rootFolder=fullfile(outputFolder,'train');

 categoreis={   'Female','Male'};


imds=imageDatastore(fullfile(rootFolder,categoreis),'labelsource','fo
ldernames');




 tbl = countEachLabel(imds);




minSetCount = min(tbl{:,2});

maxNumImages = 240;

minSetCount = min(maxNumImages,minSetCount);


imds = splitEachLabel(imds, minSetCount, 'randomize');

countEachLabel(imds);

tbl = countEachLabel(imds);

net = resnet50();

numel(net.Layers(end).ClassNames);
```

```
[imdsTrain,imdsValidation] = splitEachLabel(imds,0.85,'randomized');



 if isa(net,'SeriesNetwork')
  lgraph = layerGraph(net.Layers);
else
  lgraph = layerGraph(net);
end
[learnableLayer,classLayer]= findLayersToReplace(lgraph);



 numClasses = numel(categories(imdsTrain.Labels));


if isa(learnableLayer,'nnet.cnn.layer.FullyConnectedLayer')
    newLearnableLayer = fullyConnectedLayer(numClasses, ...
        'Name','new_fc', ...
        'WeightLearnRateFactor',10, ...
        'BiasLearnRateFactor',10);


elseif isa(learnableLayer,'nnet.cnn.layer.Convolution2DLayer')
    newLearnableLayer = convolution2dLayer(1,numClasses, ...
        'Name','new_conv', ...
        'WeightLearnRateFactor',10, ...
        'BiasLearnRateFactor',10);
end
lgraph = replaceLayer(lgraph,learnableLayer.Name,newLearnableLayer);
newClassLayer = classificationLayer('Name','new_classoutput');
```

```
lgraph = replaceLayer(lgraph,classLayer.Name,newClassLayer);

figure('Units','normalized','Position',[0.3 0.3 0.4 0.4]);

plot(lgraph)

ylim([0,10])




inputSize = net.Layers(1).InputSize;

augimdsTrain    =    augmentedImageDatastore(inputSize(1:2),imdsTrain
,'ColorPreprocessing', 'gray2rgb');

augimdsValidation                                                    =
augmentedImageDatastore(inputSize(1:2),imdsValidation,'ColorPreproces
sing', 'gray2rgb');

miniBatchSize =128;

valFrequency = floor(numel(augimdsTrain.Files)/miniBatchSize);

options = trainingOptions('sgdm', ...

    'MiniBatchSize',miniBatchSize, ...

    'MaxEpochs',1, ...

    'InitialLearnRate',3e-4, ...

    'Shuffle','every-epoch', ...

    'ValidationData',augimdsValidation, ...

    'ValidationFrequency',valFrequency, ...

    'Verbose',false, ...

    'Plots','training-progress');




net = trainNetwork(augimdsTrain,lgraph,options);

[YPred,probs] = classify(net,augimdsValidation);
```

```
imdsValidationbels = imdsValidation.Labels;


confMat = confusionmat(imdsValidationbels, YPred);

plotconfusion(imdsValidationbels, YPred);



confMat = bsxfun(@rdivide,confMat,sum(confMat,2));


accuracy = mean(YPred == imdsValidationbels)
```