WestVirginiaUniversity
THE RESEARCH REPOSITORY @ WVU

Graduate Theses, Dissertations, and Problem Reports

2016

Gender Estimation from Fingerprints Using DWT and Entropy

Sneha Nagabhyru

Follow this and additional works at: https://researchrepository.wvu.edu/etd

Gender Estimation from Fingerprints Using DWT and Entropy

Sneha Nagabhyru

Thesis submitted to the college of engineering and mineral resources at West Virginia University in partial fulfillment of the requirements for the degree of

Master of Science

in

Electrical Engineering

Bojan Cukic, Ph. D., Chair

Matthew Valenti, Ph. D.

Jeremy Dawson, Ph.D.

Lane Department of Computer Science and Electrical Engineering

Morgantown, West Virginia

2016

Abstract

Gender Estimation from Fingerprints using DWT and Entropy

Sneha Nagabhyru

Gender estimation from fingerprints have wide range of applications, especially in the field of forensics where identifying the gender of a criminal can reduce the list of suspects significantly. Although there have been quite a few research papers in the field of gender estimation from fingerprints most of those experiments used a lot of features but were only able to achieve poor classification results. That being the motivation behind the study we successfully proposed two different approaches for gender estimation from fingerprints and achieved high classification accuracy.

In this study we have developed two different approaches for gender estimation from fingerprints. The dataset used consists of 498 fingerprints of which 260 are male and 238 are female fingerprints. The first approach is based on wavelet analysis and uses features obtained from a six level discrete wavelet transform (DWT). Classification is performed using a decision stump classifier implemented in weka and was able to achieve a classification accuracy of 95.38% using the DWT approach. The second approach uses wavelet packet analysis and extracted the Shannon entropy and log-energy entropy from the coefficients of wavelet packet transform and provided a classification accuracy of 96.59% on the same dataset using decision stump classifier implemented in weka.

This work is   dedicated to my Family for their Continuous support and Encouragement

Special thanks to my friends who have believed in me and helped me in completing this work

Table of Contents

List of Figures

List of Tables

1.Introduction

1.1 Motivation

The use of fingerprints for identification and classification purpose is referred to as dactyloscopy. With the remarkable amount of information that is transferred through the internet every single day, it is necessary to make sure that only authorized people have access to important information. Biometrics authentication systems are becoming widely popular these days because compared to traditional authentication systems like passwords biometric systems they provide higher security. Although soft biometric traits like gender provide limited information they make the identification and verification process easier. Especially in the field of forensic anthropology finding out the gender from a fingerprint found at the crime scene can reduce the number of suspects quite significantly and can be a valuable help to solving the case. Hence, the development of an automated system for gender estimation from fingerprint using a simple algorithm and good accuracy is very useful to the field of biometrics.

1.2 Goal

The goal of this study is to develop an automated algorithm for gender estimation from fingerprints that uses few features to provide a good classification accuracy.

1.3 Contribution

The algorithm described in this thesis provides an efficient algorithm for gender estimation from fingerprints. Two different methods are proposed for gender estimation based on discrete wavelet transform and entropy. Both these algorithms give promising results for gender estimation by using very few features for classification.

1.4 Organization

The remaining sections of this thesis are organized as follows. Chapter 2 gives a brief introduction to biometrics while discussing why biometric authentication systems are important and what characteristics make a biometric trait, it also discusses how the various patterns in fingerprints are classified and how they can be used for verification purposes. It also describes wavelet analysis and wavelet packet entropy in detail Chapter 2 ends with information on all the previous work that has been conducted on gender estimation from fingerprints. Chapter 3 gives a detailed information on the algorithm using discrete wavelet transform used for gender estimation from fingerprint. Chapter 4 explains an entropy based algorithm for gender estimation from fingerprints. Chapter 5 describes the conclusions made from the work done in this thesis and also describes any scope for future work.

2. Related Work:

2.1 Biometrics: An Overview

In our day to day lives we encounter various instances where the identity of a person needs to be confirmed that is, confirmation of whether the person is who he/she claims to be. For instance an identity card to enter a building, driving license as a proof of age or a passport to enter into a country. These are all a means to establish the identity of an individual which helps to verify if the person is who they claim to be. This confirmation of identity is of significant importance especially in this present society where important information when misplaced in the hands of wrong people can cause a lot of harm. Hence the task of authentication of the identity of the individual before sharing any information with them is very important. Passwords and Identity cards were the traditional authentication systems used to establish one's identity but these systems had many drawbacks. For instance, complex passwords can be forgotten by users or simple passwords can be very easily hacked or the password can be revealed to an unauthorized user or a person may lose his/her identity card which can again grant access to an unauthorized user which leads to a breach in security. Hence with the increasing concerns for security and also advancements in the field of communication and computing, the need for robust user authentication systems is ever growing. This is why development in the field of biometrics is very important.

Biometrics is identifying a specific individual based on the distinctive physical or behavioral traits of that person [1]. Any physical or behavioral trait must possess the following seven qualities to some extent to be considered a biometric characteristic:[2].

1) Universality: Every individual should have this characteristic [2].

2) Distinctiveness: The characteristic must be sufficiently different between two individuals [2].

3) Permanence: For a given individual the characteristic should not vary significantly over time [2].

4) Collectability: It should be possible to measure the characteristic quantitatively for an individual [2].

5) Performance: It is defined as how the system must perform based on the characteristic to achieve the required level of accuracy and speed [2].

6) Acceptability: It indicates the degree to which people are ready to accept this characteristic in their daily lives [2].

7) Circumvention: Indicates how easily the biometrics system will be mistaken by using fraudulent methods [2].

Some of the physical traits are:

- Fingerprint
- Iris
- Lips
- Retina
- Palmprint
- DNA
- Face
- Hand geometry

Some of the behavioral traits include:

- Speech patterns.
- Keystrokes
- Signature
- Gait

As mentioned above biometrics uses various traits such as face, fingerprint, iris, voice, keystrokes etc to identify a person as shown in Fig2.1. So, the question is why limit ourselves with just one trait- fingerprint? Among all these different traits fingerprint biometrics is popular and widely used because:

- Fingerprints are unique for every person
- Fingerprints can be easily collected.
- Number of sources(every person has 10 fingers)
- Low cost of fingerprint biometric system and
- Fingerprint as a biometric trait is universally accepted
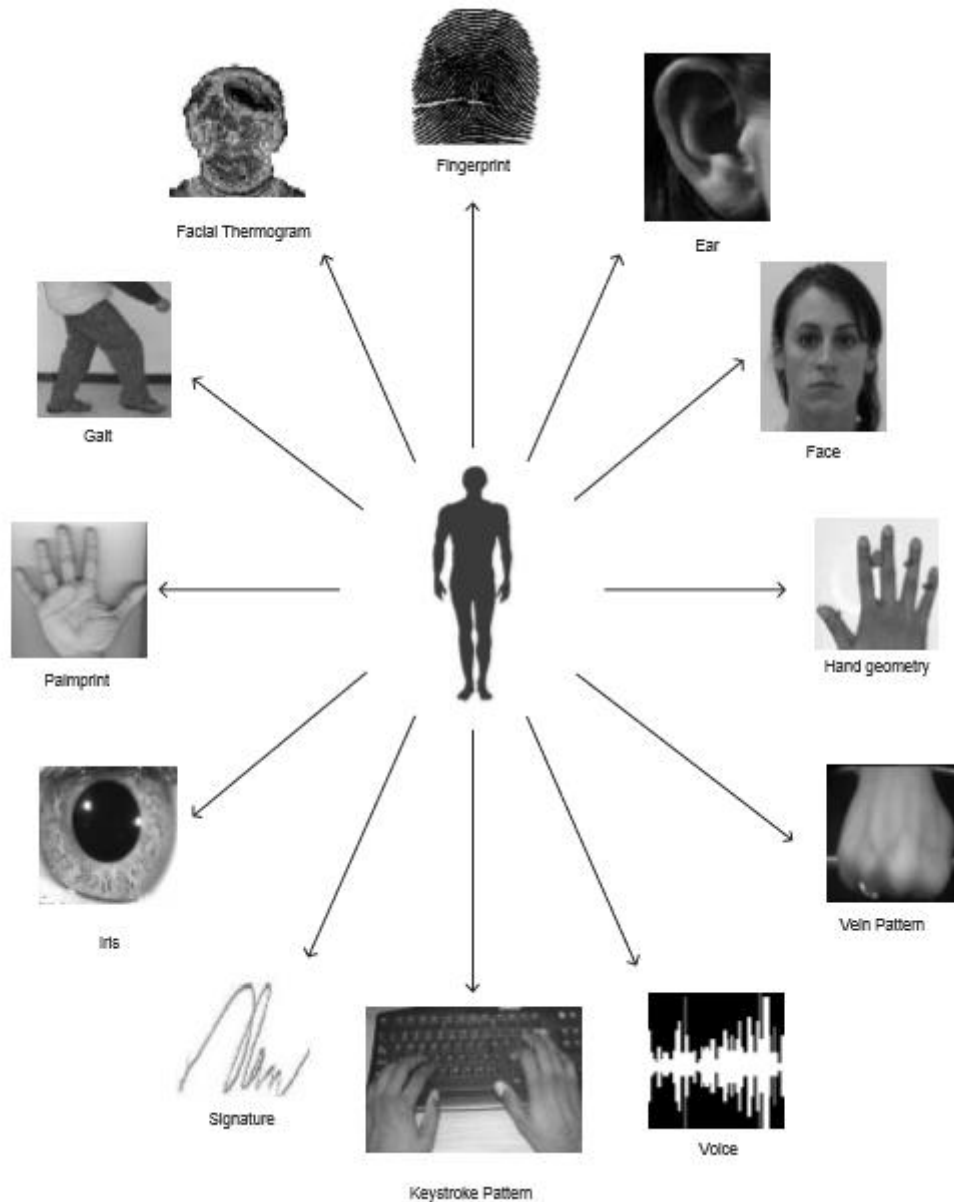
Fig 2.1: An illustration of the various biometric traits to identify an individual (borrowed from [1])

A lot of information can be extracted from fingerprints. Previous research has shown that useful information regarding age and gender of an individual can be extracted from fingerprints [3 4]. However, in this thesis only gender estimation from fingerprint is discussed. Gender estimation from fingerprints has many applications in biometrics, especially in the field of forensic anthropology [5]. For instance, identifying the gender of

a fingerprint collected from a crime scene helps to reduce the number of suspects quite significantly. That is, if the fingerprint found at a crime scene is identified as a male we can automatically eliminate all females from the suspect list. Various methods have been proposed for gender estimation which have been discussed in detail in the next chapter.

2.1.1 Performance measures in Biometrics

The performance of various biometric systems is most commonly measured in terms of False Accept Rate (FAR) and False Reject Rate(FRR).

False Accept Rate (FAR):

A FAR is the probability that the system incorrectly matches the features of the input biometric sample with those already present in the system database. FAR is also referred to as a false negative error and is an example of type II error in statistics.

False Reject rate (FRR):

A FRR is the probability that a biometric system incorrectly does not match the features of the biometric sample with those already present in the system database. FRR is also referred to as false positive error and is an example of type I error in statistics.

2.1.2Security Issues with Biometrics.

With the advancement in the field of biometrics there has been a huge improvement in the field of information and identity security. Undoubtedly biometrics provides us with higher security when compared to traditional methods of identity authentication like passwords etc. However, biometrics authentication system has its own limitations as well as privacy concerns.

Unlike passwords that can be changed when hacked or an ID card that can be bought gain when lost by submitting the necessary documents, biometric traits such as a fingerprint pattern of a person are unique and do not change in one's lifetime hence, they

cannot be replaced. And with the increasing use of fingerprint authentication systems fingerprints are stored in lots of databases. If by any chance the fingerprint of an individual is compromised there is no way to change that because biometrics are closely associated with the biological characteristics of an individual. Some of the concerns with a biometric authentication system are:

- At times, it is possible that more information is divulged than what needs to be known which is 'risky'. For instance if a study is collecting DNA samples for the purpose of studying the differences between male and female DNA. However access to ones DNA can give a lot of additional information about the person such as any genetic disorders that person might have which he/she might have wanted to keep a secret [5].

- Also in cases when a person intends to keep his identity confidential, biometrics might deny the person their privacy [5]. For example if a person is using an alias for legal reasons, comparison of fingerprints might give away information about the actual identity of that person.

- It is often possible to obtain biometric information from an individual, like the face without the knowledge of the person. As a consequence if a person wants to remain anonymous in a particular situation his privacy is denied by that biometric information.

.

Some methods proposed to handle the privacy concerns associated with biometrics are: To store the fingerprint as an encrypted signal rather than the fingerprint itself or to delete useful information from the fingerprint such that the new fingerprint does not give any information about the individual who owns it [6][7]. Hence an efficient biometric system must have a good trade-off between privacy and security.

2.2 Fingerprint Anatomy:

Before we dive into the topic of gender estimation from fingerprints it is relevant to know about the various patterns in a fingerprint and also how the usage of fingerprint as an

authentication system in biometrics came to be. In our day to day lives we touch a number of things: be it a cup, a door, a book, or a computer and every time we do so we are leaving our finger prints on them. The fingerprint of each and every one is unique and is never repeated. Even identical twins who have same DNA do not have the same fingerprints. The pattern of a finger print will not change during the lifetime of an individual. However, sometimes the fingerprint may slightly vary due to aging of the person or even sometimes disappear due to accidents like burns which are some special cases. These aside, because of this unique characteristic of fingerprints, fingerprint analysis has wide range of applications in the field of biometrics.

This knowledge that no two individuals have the same fingerprints was first used by Sir Francis Galton in 1900 to solve criminal cases in Scotland Yard in England. He implemented this knowledge of the uniqueness of fingerprints by comparing the fingerprints at the crime scene with that of the suspect. He based his work mostly on the observations by Sir Edmund Henry and hence, their approach is called Galton-Henry system. Later in 1904 a paper titled 'Comparative Fingerprints' was published by Juan Vucetich, working for a police department in Argentina [8]. To this day his technique is widely popular and is used in most of the Spanish speaking countries.

All of these systems proposed in the early 1900's are quite similar to one another. It was Edmund Henry who first recognized that all fingerprints could be described as having patterns [8]. These patterns were described as loops, arches and whorls. Eight basic patterns were later developed from these shapes and contours which are still used widely used by the Federal Bureau of Investigation (FBI) during fingerprint analysis.

2.2.1 Ridge Characteristics

A fingerprint is a composition of ridges and valleys on the top epidermal layer of the skin. Where ridges are curved lines consisting of the upper layer segments of the skin and similarly valleys consist of the lower layer segment. Furthermore, a ridge ending is where the ridge curve ends and the point where a ridge divides into two is called a ridge bifurcation. The points at which the ridge structure changes are known as minutiae points.

These minutiae points are very useful in matching a fingerprint to a specific individual. Figure 2.2 shows the various ridge characteristics used to identify the minutiae [9].
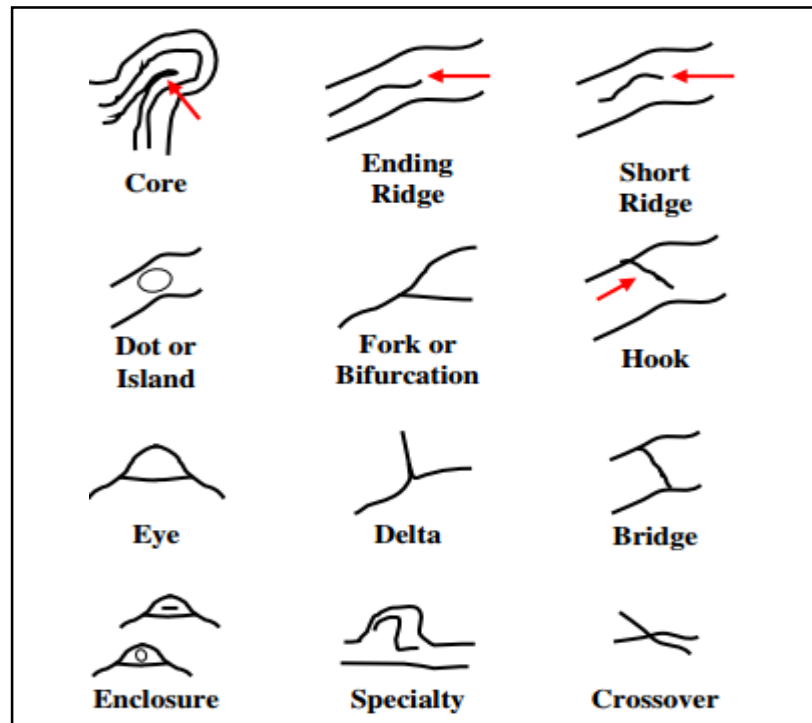
Fig 2.2: Various Ridge characteristics in a fingerprint(borrowed from [9])

2.2.2 Fingerprint Patterns

The patterns in fingerprints are basically divided into 3 basic types: Loop, Arch, Whorl as shown in fig 2.3 Which are further classified as follows:

1) Loop
   - Ulnar loop
   - Radial loop

2) <u>Whorls</u>

- Plain whorl
- Double loop whorl
- Central pocket whorl
- Accidental whorl

3) <u>Arches</u>

- Plain arch
- Tented arch

| (a) | (b) | (c) |
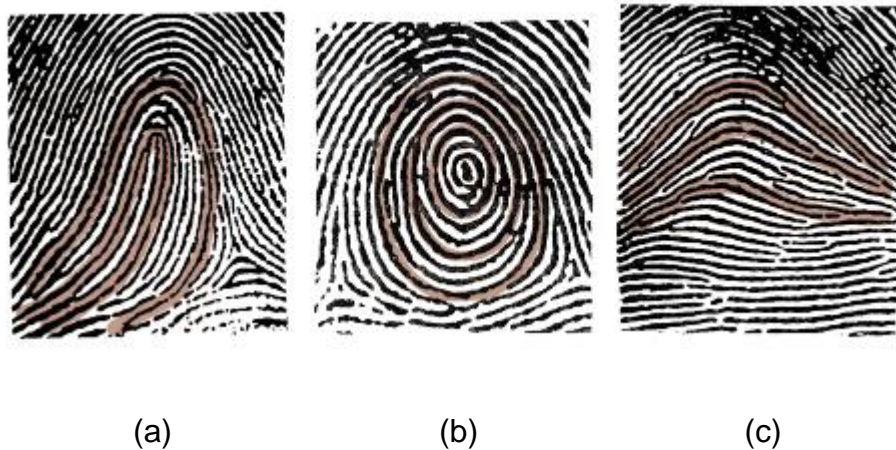
Fig 2.3: (a) Loop,(b) whorl and (c) arch patterns (borrowed from [9])

<u>Loops:</u>

The loop pattern is defined as one in which a ridge that enters from one side of the print recurves and exits from the same side. A loop is usually the result of one core and one delta. Loops are further divided into two types: ulnar loop and radial loop. If the loop points toward the little finger it is called ulnar loop and if the loop points toward the thumb finger it is called radial loop as shown in fig 2.4 Approximately 60% of the pattern types in fingerprints are accounted for by loops.
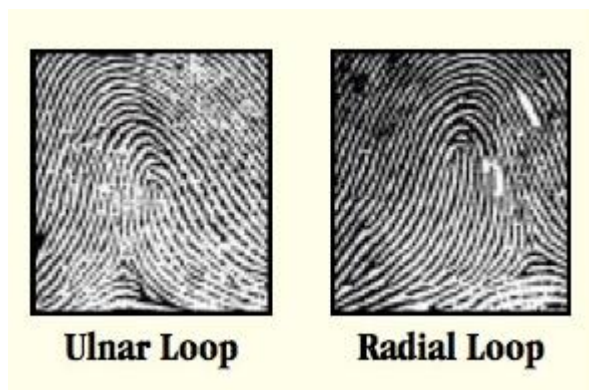
Fig 2.4: A ulnar loop and a radial loop pattern (borrowed from [9])

Whorls:

Whorls form circular or spiral patterns. There are four different types of whorls as shown below in fig 2.5: plain whorl, central pocket loop, double loop and accidental whorl. A whorl consists of one or more cores and at least two deltas. A plain whorl consists of a simple spiral or circular pattern consisting of concentric circles. A double loop as the name suggests comprises of two different loops together that often look like the shape 'S'. The central pocket loop is identified by a circular pattern usually at the center of the print. And an accidental loop is an irregularly shaped circular pattern. About 35% of the patterns in fingerprints are accounted for by whorls.
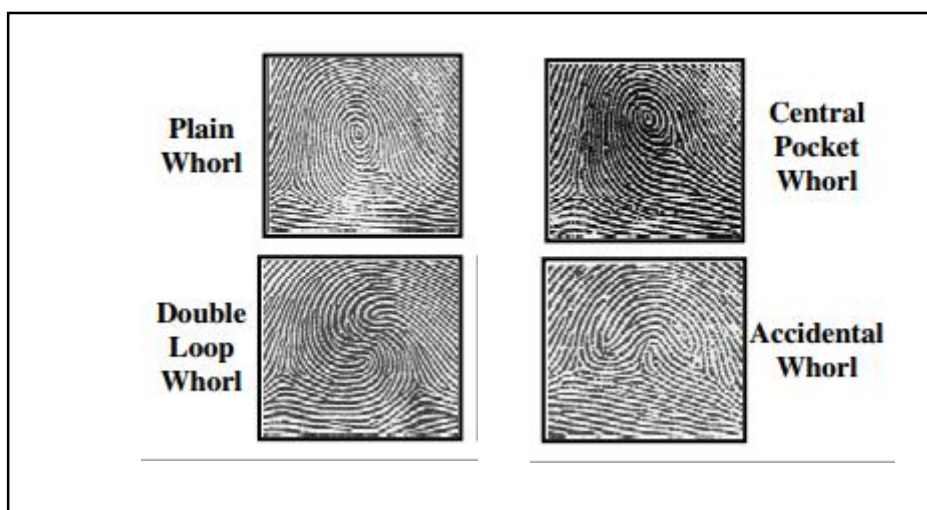
Fig 2.5: Different types of whorls in fingerprints (borrowed from [9])

Arches:

Arches are the least common type of fingerprint patterns which are identified by a ridge entering from one side of the print and exiting from the opposite side. There are two different types of arch patterns: Plain arch and Tented arch as shown below in fig 2.6 .The difference between tented arch and a plain arch is that in a tented arch there is a sharp spike in the ridge at the center before exiting the print which gives it a tent like appearance in comparison to the smoothly exiting ridge of a plain arch. Only 5% of the fingerprint patterns are arches.

(a)   (b)
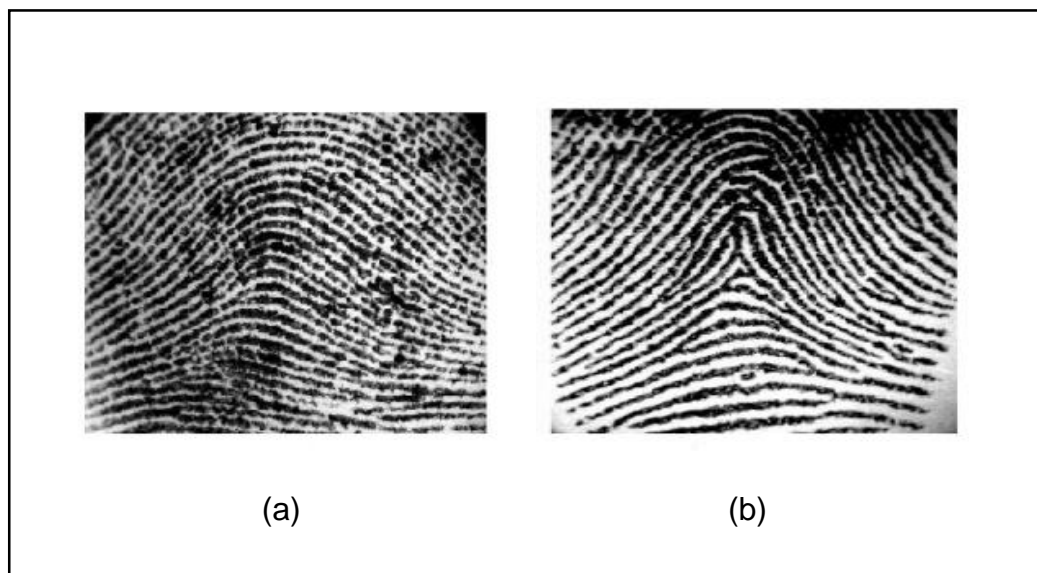
Fig 2.6: (a) Plain arch and (b) Tented arch (borrowed from [9])

These are the different basic pattern types in fingerprint images. Usually in fingerprint analysis the general patterns in the fingerprint such as loop, whorl and arch are used only to make initial comparisons and later on the exact fingerprint match is identified by comparing the specific minutiae in the fingerprints. Hence minutiae play a very important role in fingerprint biometrics. This is also confirmed in the later chapters of the thesis where minutiae count is extracted from the fingerprints accounts for a very important feature used for gender estimation from fingerprints.

2.3 Fingerprint Acquisition Techniques

Depending on how fingerprints are captured fingerprint acquisition is broadly classified into two different categories

- Off-line or inked images
- Live-scan or inkles images

Offline fingerprint images:

Offline fingerprint images are usually collected by spreading black ink onto any individual's finger, the image is then obtained by pressing and then rolling the inked finger onto a paper card. This card is then scanned to produce a digital image of the fingerprint. Although this type of offline fingerprinting is not a very convenient process and also cannot be used for biometric systems which need real time processing, it is still very popular in forensics where latent prints are collected from the crime scenes.

Live scan fingerprint images:

Live-scan fingerprints are acquired when a digital image of the fingerprint is sensed over an electronic fingerprint device (fingerprint sensor). Because these fingerprint images are collected directly in digital format, live-scan images can be used in biometric systems that use real time processing. There are various types of fingerprint sensors like optical, thermal, capacitance, ultrasonic, piezoelectric etc that are used to generate a live-scan fingerprint images. Below is an overview of the most commonly used fingerprint sensors.

Ultrasonic sensors:

An ultrasonic sensor consists of two main parts: a sender and a receiver. The sender usually is a piezoelectric transducer that generates high frequency sound waves which penetrate the outer epidermal layer of the skin of the fingertip and the reflected signal is then measured by the piezoelectric materials present in the receiver. Because the epidermal layer of the skin has the same characteristic as the fingerprint pattern, the computation of the reflected wave can be used to format a fingerprint image. One of the

most important advantages of an ultrasonic sensor is that it does not need a clean sensing surfaces, therefore any noise related disturbance is removed from the fingerprint image.

Capacitance:

 Capacitance sensors use the ability of the body to store electric charge (capacitance) to form fingerprint images. In capacitance type of sensors each array pixel of the sensor acts as one plate of the parallel plate capacitor and the conductive dermal layer acts as the other plate and the non-conductive epidermal layer acts as the dielectric. When a finger is placed on the sensor the ridges and valleys give different capacitance values because of the difference in distances. These different capacitance patterns for different ridges and valleys are then used to generate a fingerprint image as output.

Optical sensors:

Optical sensors are one of the most popular methods used to capture live-scan fingerprints, in fact even in this thesis we have used this type of sensor to collect fingerprint images. In optical type live scan sensors, visible light is used to capture the digital print of a fingerprint image. The topmost layer of the sensor is a sensing or touch surface where the finger will be placed. The finger placed on the touch surface is illuminated using a light emitting phosphor layer present right below the touch surface and the reflected light is then used to capture a digital fingerprint image using a charge coupled device. One of the main disadvantages of an optical sensor is that the quality of a fingerprint image is dependent on the quality of skin on the finger, for example a greasy or dirty finger produces a bad quality fingerprint image and similarly a clean finger produces a high quality fingerprint image.

2.4 Collecting and Analyzing Fingerprints

As we have mentioned earlier one of the main applications of gender estimation of fingerprints is in field of forensics. So, it would be appropriate for us to learn more about how the fingerprints are collected at crime scenes and also how they are analyzed. For the fingerprints to be analyzed it is first necessary to locate and collect the fingerprints from a crime scene.

Location of fingerprints:

At any crime scene, the collection of fingerprints is a relatively easy task if the fingerprints are visible to the naked eye. However, this is not always the case, in situations when the fingerprints are not visible to the naked eye a very careful and extensive search is required to locate the fingerprints. There are three different types of fingerprints and the amount of time and effort Investigators must put into to locate the fingerprint usually depends on the type of the fingerprint [43].

The three different types of fingerprints according to forensic science are:

- Patent prints – These type of fingerprints are visible to the naked eye and therefore it is very easy to detect them at a crime scene. These fingerprints usually occur when the surface of the finger is coated with any substance that can leave a visible impression such as oil, ink or paint.
- Plastic prints – Plastic prints occur when fingerprints are pressed in wax, tar, soap or any similar substance that leaves a three dimensional impression of the fingerprint. Although these fingerprints are visible to the naked eye just like patent prints they are not very common.
- Latent prints – Latent prints are the most commonly found fingerprints at crime scenes. Unlike patent and plastic prints which are visible to the naked eye these prints are invisible and so, location of these type of fingerprints is not as easy. When we touch any kind surface the natural oils and residue on the fingers leave a deposit on the surface that is a copy of the ridges and valleys present on the finger which are referred to as latent prints.

During the investigation of a crime scene the first step usually involves looking for the visible patent and plastic prints. Then for the location of latent prints different methods such as using powder techniques for non-porous surfaces or chemical techniques (like silver nitrate) for porous surfaces are employed. The next step involves lifting the fingerprint from the crime scene, this collected print is then compared with the different fingerprints already present in the fingerprint database and the suspect is identified if the fingerprint at the crime scene matches with any one of the fingerprints at the database.

2.5 Wavelet Theory

2.5.1 Families of Wavelet

Wavelets are small wave like oscilllations that decay quickly and also have an average value of zero . Unlike sinusoidal waves that extend fover infinite space wavelets have a definite beginning and end. Sinusoidal waves are usually good at describing stationary signals because they are smooth and predictable. However for signals that have pulses or anamolies instead of sinusoidal waves, wavelets are better at describing them because they are irregular, of limited duration and often not symmetrical [42]. Fig 2.7 below shows a comparision between a sine wave and a wavelet. Notice how the frequency of the wavelet varies over its length making it suitable to analyze non stationary signals.

(a)   (b)
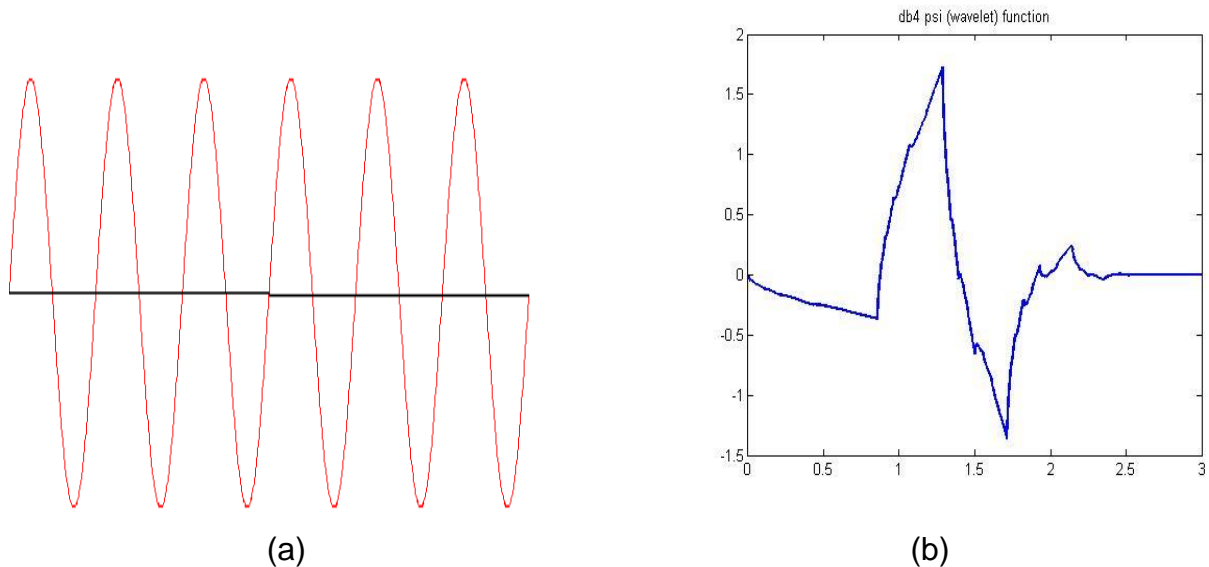
Fig 2.7 (a) portion of sinusoidal sine wave and (b) a finite length wavelet (borrowed from [54][53])

There are a lot of different types of wavelets which can be used for both continuous and discrete analysis. Some examples of wavelets used for discrete analysis include orthogonal wavelets and B-spline biorthogonal wavelets and some examples of wavelets used for continuous analysis include Morlet, Meyer and paul wavelets. The choice of a wavelet for any given problem depends on the nature of the application and also the characteristics of the given dignal or image. Some of the most commonly used wavelets and wavelet families will be discussed below:

17

<u>Haar wavelet:</u>

Haar is a very simple wavelet function proposed by Alfred Haar in 1909. A Haar wavelet resembles a step function and is discontinuous. Therefor it is useful for analysis of signals with sudden transitions. A diagramatic representation of a Haar wavelet can be seen in Fig 2.8.
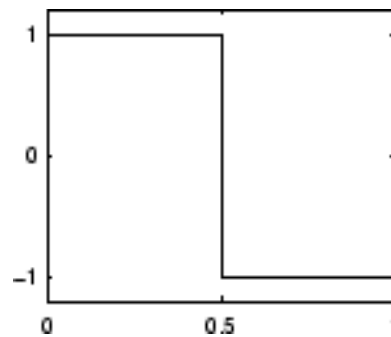
Fig 2.8 Haar Wavelet (borrowed from [52])

<u>Meyer wavelet:</u>

Meyer wavelet shown in Fig 2.9 was proposed by Yves Meyer. It is an orthogonal wavelet which is infinitely differentiable. Meyer wavelet is a symmetrical wavelet and is defined in the frequency domain.
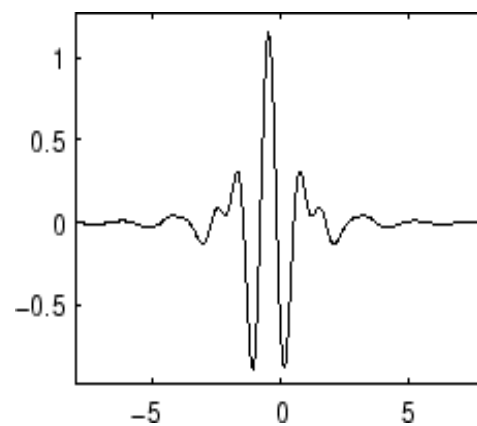
Fig 2.9 Meyer Wavelet (borrowed from [52])

Mexican hat wavelet:

Also referred to as the ricker wavelet shown in Fig 2.10 , Mexican hat wavelet is derived from a a function that is proportional to the second derivative function of the Gaussian probability density function and also it does not have a scaling function.
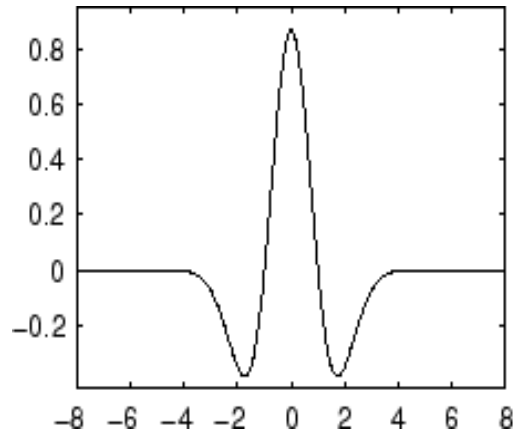
Fig 2.10 Mexican hat wavelet (borrowed from [52])

Daubechies wavelet:

Daubechies wavelets proposed by Ingrid Daubechies are the most popularly used wavelets. Daubechies wavelets are a family of orthogonal wavelets defining DWT and are characterized by a maximal number of vanishing moments for some given support.Different wavelets in the Daubechies family are shown in Fig 2.11.
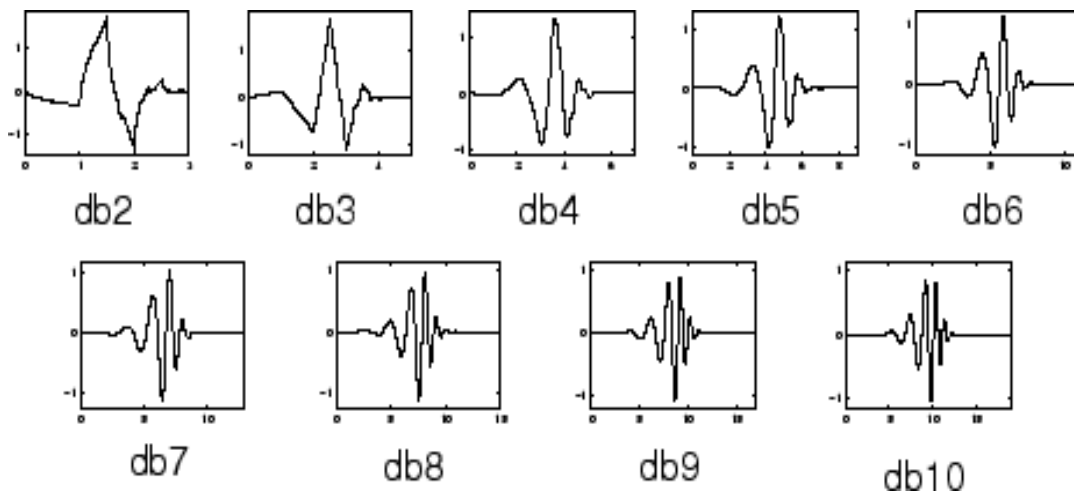
Fig 2.11 Daubechies family of wavelets (borrowed from [52])

2.5.2 Wavelet Analysis

When working with different type of signals we often notice that the information contained within the signal may not always be available to us in a desirable form suitable to be used to solve the problem we are dealing with. Traditionally Fourier analysis was used to deal with such problems. The main branch of mathematics which led to wavelet theory first began with the theories of frequency analysis by Joseph Fourier, which is often now referred to as Fourier synthesis [23]. He asserted that any 2π- periodic function $f(x)$ can be represented as a sum of sine and cosine functions.

$$f(x) = a_0 + \sum_{k=1}^{\infty}(a_k \cos kx + b_k \sin kx) \tag{1}$$

Where,

$$a_0 = \frac{1}{2\pi}\int_0^{2\pi} f(x)\, dx \;,\; a_k = \frac{1}{\pi}\int_0^{2\pi} f(x) \cos(kx)\, dx \;,\; b_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\, \sin(kx)dx \;.$$

However analyzing the signals with Fourier analysis has some drawbacks, because Fourier analysis uses sines and cosines as bases to appropriate any given function and since both sine and cosine functions are non-local as they extend up to infinity Fourier analysis is well suited for approximating stationary signals but is not suitable for approximation of transient signals (signals with sharp spikes). Fourier transform, one of the most common methods of frequency analysis analyzes a function in time domain for its frequency content [23].

Wavelet analysis is similar to Fourier analysis in the aspect that it uses basis functions to approximate any given functions. However, unlike in a Fourier analysis which uses sine and cosine functions for approximation wavelet analysis uses wavelets that are generated by the dilations and shifting of a fixed function called the mother wavelet [24]. A wavelet is a small wave like oscillation that starts with zero amplitude and then the amplitude increases and quickly decays back to zero again [25]. Unlike the big sine and cosine waves that extend to infinity in Fourier transform, wavelets are small oscillations. The mathematical conditions equivalent to wavelet are as follows:

$$\int_{-\infty}^{\infty}|\psi(t)|^2dt < \infty; \tag{2}$$

$$\int_{-\infty}^{\infty}|\psi(t)|dt = 0 ; \tag{3}$$

$$\int_{-\infty}^{\infty}\frac{|\hat{\psi}(w)|^2}{|w|}dw < \infty ; \tag{4}$$

Where $\hat{\psi}(w)$ is the Fourier transform of $\psi(t)$ and equation 4 is called the admissibility condition.

2.5.3 Wavelet transform:

The idea of wavelet transform was first proposed by Jean Morlet in 1982. Morlet considered wavelets as a family of functions constructed from the translations and dilations of a single function called the "mother wavelet" $\psi(t)$ [23]. They are defined by

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right), \ a,b \in \mathbb{R}, a \neq 0 \tag{5}$$

In equation 5, $'a'$ is a scaling parameter or scale which measures the degree of compression and $'b'$ is a translational parameter which determines the time location of the wavelet [26]. The wavelet is a compressed version of the mother wavelet if $|a| < 1$ and mainly corresponds to higher frequencies. If $|a| > 1$ the wavelet correspond to lower frequencies and $\psi_{a,b}(t)$ has larger time-width than $\psi(t)$. Thus wavelets have time-widths adapted to their frequencies.

One of the main advantages of wavelet transform is that it provides localization in both time and frequency domain while the Fourier transform is localized only in the frequency domain. Some of the other advantages of wavelet analysis are : In wavelet analysis it is possible to obtain a good approximation of a given function by using just a few wavelet coefficients, Also wavelet analysis is computationally fast and finally one of the most important advantage of wavelet analysis that prompted us to opt for wavelet transform is that wavelets have the ability to separate very fine details in a signal, small wavelets can isolate very small details in the signals and similarly large wavelets can identify coarse details in the given signal [26].

2.5.4 Wavelet Packet Entropy:

The wavelet Packet transform decomposes both the approximation and details of the signal and therefore the signal can be represented in a lot of different ways and the number of binary subtrees maybe very large. So based on the design objective it is important to select the 'best tree' (optimal decomposition) from all the different bases obtained from wavelet packet decomposition. Entropy is a common method used to find the optimal decomposition with applications in signal and image processing. Entropy is a measure of the randomness in a signal, entropy based wavelet packet decomposition is proposed by Coifman and Wickerhauser [22]. There are different entropy types such as Shannon, log energy, sure, threshold etc [22]. The entropy 'E' must be an additive cost function such that $E(0) = 0$ [28]. and

$$E(s) = \sum_i E(s_i)$$

The wavelet entropy (WE) is for the observed signal in $/^p$ norm with $p \geq 1$ can be expressed as [28],

$$E(s_i) = |s_i|^p$$

and

$$E(s) = \sum_i |s_i|^p$$

Where the coefficients of the signal $'s'$ in an orthonormal basis are represented by s_i. WE provides us with useful information about the underlying dynamical process related to the signal because it is a measure of the uncertainty in the signal [29]. If the value of entropy is greater than 1 this means that the component can provide more information relevant to the signal and needs to be decomposed further to obtain simple frequency component of the signal [27].

The Shannon entropy [30] introduced by Shannon is defined as

$$E(s) = \sum_i s_i^2 \log s_i^2.$$

The log energy entropy is defined by

$$E(s) = \sum_i \log s_i{}^2$$

The threshold entropy is defined as

$E(s) = \#\{i \ such \ that \ |s_i| > p\}$, the number of times the coefficient is larger than a threshold.

The SURE entropy is defined as

$$E(s) = n - \#\{i \ such \ that \ |s_i| \leq p\} + \ \sum_i min(s_i{}^2, p^2) \ .$$

2.6 Feature Selection

Feature reduction is an important aspect of any classification problem. Because of the advancements in the field of computers and information technology in the present day we

have large amounts of data available to us. From this wide network of information available to us it is important to separate noisy, redundant and unreliable data, so that we are only left with data that is relevant to the problem at hand and helps us to achieve high accuracy for a classification problem. There are two different methods to reduce the attribute (feature) space for a feature set: dimensionality reduction and feature selection [32]. Although both methods try to reduce the number of attributes in the feature set, dimensionality reduction method does so by producing linear combination of the original attributes of the data set whereas feature selection creates a new feature subset by removing some of the attributes from the original feature set.

The process of identifying the most relevant attributes (features) and removing any attributes (features) redundant or irrelevant for the classification purpose is referred to as feature selection [31]. By reducing the number of features in the feature set feature selection not only helps deal with the problem if 'curse of dimensionality', but also designs a simpler and better understandable classification model [32 ,33].

There are three different categories in feature selection algorithms:

- Embedded methods
- Wrapper methods and
- Filter methods

Wrapper methods:

Wrapper methods search through the feature set and create various combinations of features (feature subsets) and evaluates how each combination affects the classification accuracy. The feature space is usually either searched forwards, that is by adding attributes one by one to an initially empty attribute set or backwards, by removing features one by one from the complete set of attributes to determine how each feature affects the classification accuracy [32]. The drawback of this type of wrapper methods is that they are dependent on the specific type of classifier to evaluate the quality of a set of features and also because of the exhaustive search wrapper methods are usually computationally complex and time taking. An example of wrapper method is recursive feature eliminating algorithm.

Embedded methods:

While the classification model is being created embedded methods learn which of the attributes best contribute to the accuracy of the classifier, most common example of this type of embedded methods are regularization methods. Regularization methods also known as penalization methods, introduce additional constraints into the optimization of a predictive algorithm and bias the model towards lower complexity [34].

Filter methods:

Filter based feature selection methods are independent of the classifier. Filter based selection algorithm uses the statistical correlation between a set of features and a target feature to assign a score to each feature [31]. The features are then ranked based on the score and are either included or excluded from the feature set. Some examples for filter based methods are chi squared method, information gain etc. Because the filter based

methods are independent of the classifier they are faster, scalable and computationally simpler than wrapper based methods.

Information gain is a filter type of feature selection method and is widely used for attribute selection. The worth of an attribute is evaluated by measuring the information gain with respect to class in information gain attribute evaluation [35].

InfoGain(Class,Attribute) = H(Class) - H(Class | Attribute).

In information theory and probability theory information gain is a measure of difference between any two probability distributions. Let Y be a feature with entropy

$$H(Y) = -\sum_{y \in Y} p(y) \log_2(p(y))$$

Where (y) : marginal probability density for Y.

According to the values of second feature X, if the observed values of Y in the training dataset S are partitioned and the entropy of Y with respect to the partitions induced by X is less than the entropy of Y prior to partitioning, then there is a relationship between features Y and X [36]. The entropy of Y after observing X is:

$$H(Y/X) = -\sum_{x \in X} p(x) \sum_{y \in Y} p(y/x) \log_2(p(y/x))$$

Where (y/x) : conditional probability of y given x .

In a training set S where entropy is used as a measure of impurity in the set, information gain(IG) can be used a measure reflecting additional information about Y provided by X that represents the amount by which the entropy of Y decreases [36].

$$IG = H(Y) - H(Y/X) = H(X) - H(X/Y)$$

From the above equation we can see that IG is a symmetrical measure and hence the information gained about Y after observing X is the same as the information gained about X after observing Y.

2.7 AdaBoost

Machine learning deals with the concept of how information that is already learnt from a set of training data can be applied to new or unseen data. A lot of research is done in machine learning to improve the accuracy of classification of this new data. One of such successful attempts led to the development of ensemble learning [46]. Unlike in ordinary machine learning approach which uses just one learning algorithm obtained from the training set on the new data, ensemble leaning uses multiple learning algorithms combined to improve the accuracy of classification of the new data. There are three different methods in ensemble learning:

- Boosting: Boosting is an ensemble learning method based on the idea that a highly accurate prediction rule can be made by combining many weak and inaccurate rules [45].

- Bagging: Bagging also referred to as bootstrap aggregating is an ensemble method that takes different samples from the training data and generates a learning algorithm for each of these training data samples. The results of these multiple learning algorithms are then combined.

- Stacking: Stacking also called blending is an ensemble method where multiple different algorithms are prepared on the training data and a meta classifier is prepared that learns how to take the predictions of each classifier and make accurate predictions on unseen data [47].

We will only discuss about the boosting method of ensemble learning in detail in this thesis. As already mentioned above in machine learning the approach of boosting is based on the idea that a highly accurate learning algorithm can be made by combining many weak and inaccurate learners [45]. The AdaBoost [48 49] algorithm is one of the most influential, practical and one of the most widely used boosting algorithm to this day. This adaptive Boosting algorithm, named AdaBoost was proposed by Freund and Schapire which won them the Godel prize, one of the most prestigious awards in the field of theoretical computer science in 2003. Unlike other boosting algorithms AdaBoost did

not need information about the error bound of the base learners (weak learning algorithms) which is usually unknown in practice, and because of this advantage AdaBoost has many practical applications in problems related to object detection, text analysis, data mining and etc [50].

Given: $(x_1, y_1) \ldots \ldots \ldots (x_m, y_m)$, where $x_i \in \chi$ and $y_i \in \{-1, +1\}$

Initialize: $D_1(i) = \frac{1}{m}$ for $i = 1, \ldots, m$

For $t = 1, \ldots . T$:

- Train weak Learner Using distribution D_t.
- Get weak hypothesis $h_t: \chi \rightarrow \{-1, +1\}$.
- Aim: select h_t with low weighted error

$$\epsilon_t = P_{ri \sim D_i}[h_t(x_i) \neq y_i.$$

- Choose $\alpha_t = \frac{1}{2} \ln(\frac{1-\epsilon_t}{\epsilon_t})$.
- Update:

$$D_{i+1}(i) = \frac{D_t(i)}{z_t} \times \begin{cases} e^{-\alpha t} \ if \ h_t(x_i) = y_i \\ e^{\alpha t} \ if \ h_t(x_i) \neq y_i \end{cases}$$

$$= \frac{D_t(i) \exp(-\alpha_t y_t h_t(x_i))}{z_t}$$

Where Z_t is a normalization factor (chosen so that D_{t+1} will be a distribution).

Output the final hypothesis:

$$H(x) = sign \ (\textstyle\sum_{t=1}^{T} \alpha_t h_t(x)).$$

Fig 2.12 The pseudocode for AdaBoost (borrowed from [49])

From the pseudocode for AdaBoost in Fig 2.12 it can be seen that the input of the algorithm are a set of m training samples $(x_1, y_1) \ldots \ldots (x_m, y_m)$ where the x_i's belong to some domain χ and the labels $y_i \in \{-1, +1 \}$. For each round $t = 1, 2 \ldots \ldots, T$, AdaBoost calls on a Weak learning algorithm or a weak learner. To maintain a distribution or a set

of weights over a training set is an important step in this algorithm and the weight of this distribution on training sample i in the round t is denoted by $D_t(i)$. All the weights are initially equal but after each round the weights of the samples that are classified incorrectly are increased and the weak learner is made to focus on the hard examples in the training set[49].The purpose of the weak learner is to find a weak hypothesis $h_t : \chi \to \{-1, +1\}$, with low weighted error ϵ_t for the distribution D_t [51]. The final hypothesis H is calculated as a weighted majority vote of the T weak hypotheses where the weight assigned to each h_t is α_t. Which is equivalent to saying that the sign of a weighted combination of the weak hypotheses is computed by the final hypothesis H.

$$F(x) = \sum_{t=1}^{T} \alpha_t h_t(x) \ .$$

2.8 Cross Validation

Cross validation is a model evaluation method whose goal is to predict the outcome of an independent dataset, based on the results of statistical analysis on a given dataset. 10 fold cross validation which is used to train the various classifiers used in this study is described in detail below

10- fold cross validation:

- In a 10-fold cross validation, the dataset is divided into 10 subsets each of size N/10 where 'N' is the number of data points.
- Out of these 10 subsets, the classifier is trained on 9 subsets and tested on the one remaining subset.
- The process is repeated 10 times and the mean accuracy is calculated.
- The advantage of using cross validation is that every data point gets to be on the test set only once and gets to be on the training set 9 times, therefore it does not matter how the dataset is divided.
- The only disadvantage of this method is that the algorithm has to run from scratch 10 times and therefore 10 times as much computation is required to make an evaluation.

2.9 Previous Research

Some of the previous research for gender estimation from fingerprints has been discussed briefly in this chapter and has been tabulated in table 2.1

Ahmed Badawi et al. proposed a method for fingerprint based gender classification. The dataset used consists of 10-different fingerprints collected from 2200 individuals consisting of 1100 males and 1100 females. The features that have been extracted to analyze the fingerprints are ridge to valley thickness ratio (RTVTR), white lines count, ridge count, pattern type, pattern type concordance and ridge count asymmetry. The RTVTR values were calculated automatically and the values of white lines count, ridge count and pattern type were extracted manually. Their values showed that females have higher RTVTR and white lines count compared to males whereas males tend to have higher ridge count than females. Classification using Fuzzy C-Means (FCM), Linear Discriminant Analysis (LDA) and Neural Network (NN) achieved a classification accuracy of 80.39%, 86.5% and 88.5% respectively [4].
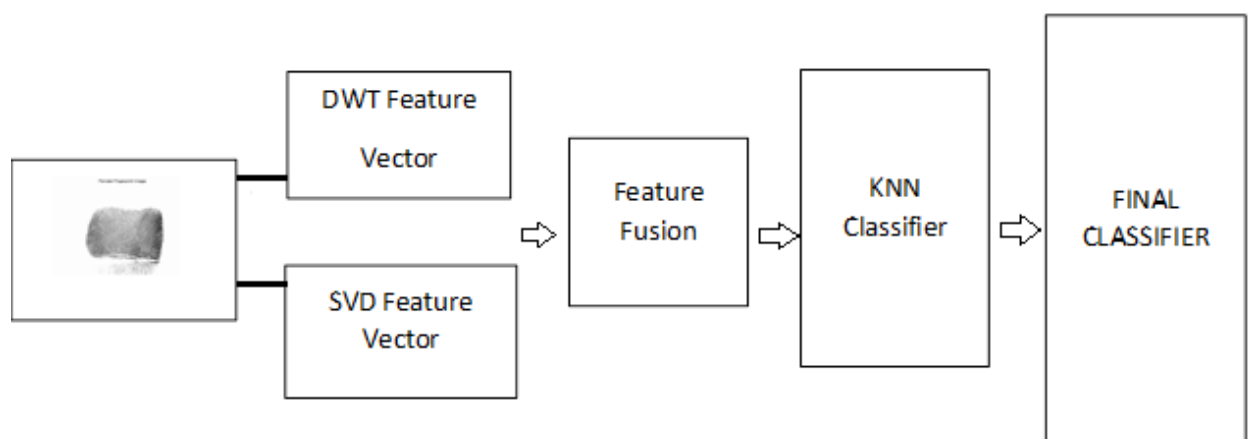
A similar method for gender classification from fingerprints was proposed by Manish Verma et al. The dataset consisted of 400 fingerprints of which 200 were male and 200 were female. The ridge to valley thickness ratio, ridge width and ridge density were the features extracted and classification was performed with a SVM classifier .The proposed approach was able to correctly classify male and females with an accuracy of 91% [10]. Jen feng wang et al proposed a method for gender based classification with fingertip features .Fingerprints were obtained from 115 adults of which 57 were male and 58 were female. Finger size was used as a feature for classification along with ridge count and ridge density. All the features were extracted from the left little finger. The highest accuracy from the proposed approach was 86% obtained using both finger size and ridge count as the features for classification [11].

A study was conducted by Sudesh Gundadin to establish a relationship between sex and fingerprint ridge density. For this purpose 500 individual fingerprints consisting of 250 male and 250 female fingerprints with age ranging from 18-60 years were collected. From all the ten fingers the fingerprints ridges were counted in the upper portion of the radial border of each print and mean value was calculated. The study shows that fingerprint

ridge of ≤13 ridges for 25 mm^2 is more likely to be male and finger print ridge of ≥ 14 ridges for 25 mm^2 is more likely to be female. This study was successfully able to support the hypothesis that women tend to have a statistically greater ridge density than men [12].

Gnanasivam P et al. (2011) proposed a method for gender estimation from fingerprints using Frequency domain analysis by using Fast Fourier Transform (FFT), Discrete Cosine Transform (DCT) and power Spectral Density (PSD) to analyze the fingerprints. The dataset used consisted of 400 individuals of varying age and gender. Gender is identified by comparing the frequency domain calculations with a predetermined threshold. The accuracy of the results obtained is 91.67% and 84.69% for male and female respectively. The proposed approach had an overall accuracy of 88.28% [13].

Gnanasivam P et al. (2012) also proposed a method for gender classification using discrete wavelet transform (DWT) and singular value decomposition (SVD) as shown in fig 2.13 [14] For a n level DWT there are $3n + 1$ sub-bands. The energy of all the sub-band coefficients of DWT combined with the spatial features of non-zero singular values obtained from SVD which is obtained from fingerprint images, is used for classification. Classification is performed using a K nearest neighbor (KNN) classifier. For a dataset consisting of 3570 fingerprints in which 1980 were male and 1590 were female, finger wise gender classification achieved an accuracy of 94.32% for female left hand little finger and 95.46% for male left hand index finger. The proposed method was able to achieve an overall accuracy of 88.28% for gender classification [14].

Fig 2.13 DWT and SVD based gender Classification system (borrowed from [14])

Marasco et al. proposed a method for gender and age estimation from fingerprints by exploiting quality and texture features. In the proposed approach features are extracted from 40 different frequency bands for image quality representation and image texture properties were captured using Local Binary Pattern (LBP) and Local Phase Quantization (LPQ) operators. For 500 different users the proposed approach was able to achieve a prediction accuracy of 89.1% for age and 88.7% for gender estimation [15]. Wadhwa et al. proposed a method for gender estimation using ridge to valley area, entropy of fingerprint image and root mean square (rms) of the DCT coefficients as features used for classification. The novelty of this solution lies in the fact that the determination of age and gender is independent from the pressure i.e. finger prints thickness or ridge/valley thickness [16].

Rijo Jackson et al. used frequency domain analysis to estimate gender from fingerprints. The fingerprints collected from 400 different persons of varying age and gender were analyzed using 2D Discrete Wavelet Transforms (DWT) and Principal Component Analysis (PCA). A minimum distance method has been used for classification to achieve an overall success rate in gender classification of around 70% [17].

H.Agarwal et al. proposed a method for gender classification using multi-class SVM. The features extracted here were ridge thickness, RTVTR and ridge density. This approach has used a multi class SVM as classifier which was able to overcome the problems in SVM binary classifier. The results of the proposed approach showed that by using a multi class SVM classifier, the classification accuracy increases with the increase of sample size of the dataset. For a dataset consisting of 600 individuals (300 male and 300 female) the proposed approach was able to achieve a classification accuracy of 91%[18].

Samta Gupta et al. proposed a method for gender classification using DWT and artificial neural network. The proposed approach is a combination of two different methods for gender classification. First, wavelet transformation is applied to the fingerprint to extract the energy values of each sub-band by performing wavelet decomposition up to 5 levels. Next, back propagation artificial neural network algorithm was used for gender identification. In this approach, a network is created with six hidden layers consisting of varying number of neurons in each layer. With a database consisting of 550 fingerprints

of which 275 were male and 275 were female fingerprints, this approach has achieved a classification accuracy of 91.45% [19].

Table 2.1: Previous Research on gender estimation from fingerprints

S.No	Author/ Title	Features	Classification method	Result
1	A. Badawi, M. Mahfouz, R. Tadross, and R. Jantz "Fingerprint - based gender classification"[4] 2006.	ridge count, ridge thickness to valley thickness ratio (RTVTR), white lines count, and ridge count asymmetry, and pattern type concordance	Fuzzy - C Means (FCM), Linear Discriminant Analysis (LDA), and Neural Network (NN).	80.39%, 86.5%, and 88.5% using FCM, LDA, and NN, respectively.
2	Manish Verma and Suneeta Agarwal.'' Fingerprint Based Male - Female Classification.''[10] 2008	ridge width, ridge thickness to valley thickness ratio (RTVTR), and ridge density.	SVM	91%
3	Jen feng wang, et al. "Gender Determination using Fingertip Features". [11] 2008.	Ridge count, ridge density, and finger size.	Structural analysis	86% accuracy is obtained by using ridge count and finger size features together.
4	P. Gnanasivam and S. Muttan, "Gender Identification Using	Frequency domain analysis	Fast Fourier Transform (FFT), Discrete	88.28%

	Fingerprint through Frequency Domain Analysis, 2011		Cosine Transform (DCT) and power Spectral Density (PSD)	
5	Gnanasivam, P. and Muttan, S., "Fingerprint Gender Classification using Wavelet Transform and Singular Value Decomposition.,"(2012)	Discrete wavelet transform(DWT) and Singular Value Decomposition(SVD)	KNN classifier	88.28%
6	Marasco, E., Lugini, L., and Cukic, B., "Exploiting Quality and Texture Features to Estimate Age and Gender from Fingerprints," (2014).	Quality and texture features using Local Binary Pattern and Local Phase Quantization operators	Principal Component Analysis(PCA)	88.7% accuracy for gender estimation and 89.1% for age estimation
7	Rijo Jackson Tom, T.Arulkumaran , "Fingerprint Based Gender Classification Using 2D Discrete Wavelet Transforms and Principal Component Analysis"[17] 2013.	Frequency domain analysis.	2D Discrete Wavelet Transforms (DWT) and Principal Component Analysis (PCA).	70%
8	Heena Agrawal et al. "Fingerprint Based Gender Classification	Features extracted were; ridge width, ridge thickness to	Muliclass SVM	91%

		valley thickness ratio (RTVTR), and ridge density.		
9	Samta Gupta et al. "Fingerprints based gender classification using Discrete Wavelet Transform and Artificial Neural Network", 2014[19]	Frequency domain analysis.	DWT & Backpropagation ANN	91.45%

3. Wavelet-Transform Approach

3.1 Dataset

Our dataset consists of index and thumb fingerprints extracted from 498 users out of which 238 are female and 260 are male .All the fingerprints are collected from a live – scan optical sensor and have a resolution of 500 dpi.

3.2 Preprocessing:

As mentioned in the previous chapter 'fingerprint' is the most popular and widely used biometric characteristic and hence it is very important to have a good quality fingerprint image with clear distinction between the various ridges to have better performance of the classifier. However in real life applications it is not always possible to get good quality images because, various factors like scars, temperature and humidity, improper contact with the sensor that corrupt the image and incorporate unwanted noise into the fingerprint image.

The main aim of image preprocessing is to improve the image data and remove any unwanted information so as to enhance the image features that are important for further processing. So, for any biometric system that uses fingerprints image preprocessing is a very important step that can help improve the classification accuracy significantly.

The preprocessing algorithm that is implemented in this thesis has two main steps

- Normalization
- Segmentation

Normalization: The input gray scale image is first normalized to have zero mean and unit standard deviation. The purpose of normalization is to have uniform illumination throughout the image. Hence, any noise we might have had in the images due to non-

uniform lightning can be reduced because of normalization. The normalization of any image $f(x,y)$ can be calculated as follows:

$$g(x,y) = \frac{f(x,y) - M_f(x,y)}{\sigma_f(x,y)}$$

Where,

$f(x,y)$: the original image;

$M_f(x,y)$: the estimation of mean of $f(x,y)$ and

$\sigma_f(x,y)$: an estimation of variance of $f(x,y)$.

Segmentation:

Segmentation of a fingerprint image helps to remove any unwanted back ground information other than the region of interest (ROI). In this approach a threshold relative to the unit standard deviation of the normalized fingerprint image is first specified. For the segmentation of the fingerprint image the image is first divided into blocks of size 16×16 and the standard deviation of each of the blacks is calculated. If the calculated standard deviation is greater than that of the threshold value only then is the block deemed as a part of the fingerprint image. This segmentation function used identifies the ridge regions of a fingerprint image and returns a mask identifying this region. Finally the mask image is renormalized so that the ridge regions have zero mean and unit standard deviation.

The preprocessing of the fingerprint image is implemented and the output image is shown in fig 3.1 implemented.

Female Fingerprint Image

(a)   (b)
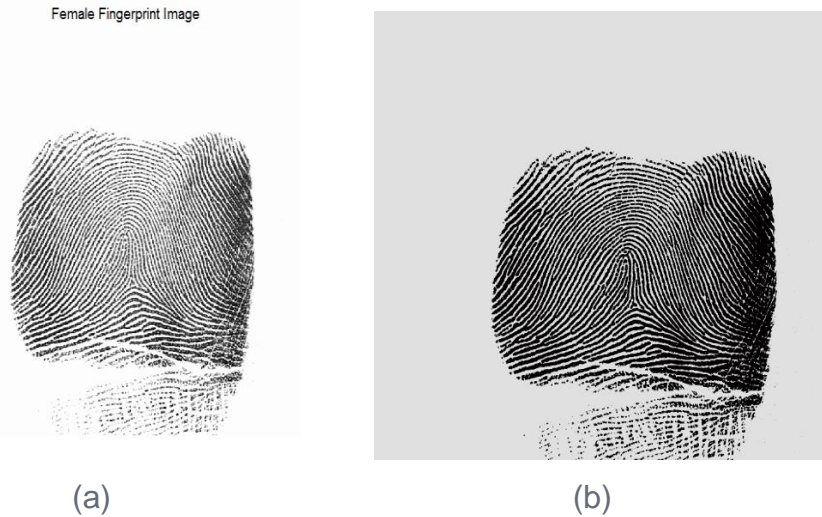
Fig 3.1 Fingerprint image (a) Before Preprocessing (b) After Preprocessing

3.3. 2-D Discrete Wavelet Transform:

Wavelets are widely used in image processing for the purpose of image de-noising, image compression, feature extraction and also for facial recognition. Discrete Wavelet Transform (DWT) is a wavelet transformation for which the wavelets are discretely sampled. The DWT of a signal x is calculated by passing it through a series of filters. First, the samples are passed through a low pass filter with impulse response g resulting in the convolution of the two as shown in the equation 6 below.

$$y[n] = (x * g)[n] = \sum_{k=-\infty}^{\infty} x[k]g[n-k] \,. \qquad\qquad (6)$$

The signal is also simultaneously decomposed using a high pass filter h . Now, only half of the each filter output is used to characterize the signal which results in the time resolution to be halved in this decomposition. Hence the frequency resolution has been doubled as each output has half the frequency band of the input [20]. This decomposition is repeated to further increase the frequency resolution and the approximation coefficients are decomposed with high and low pass filters and then down-sampled .

37

The sub-bands obtained as the output of a two dimensional DWT of an image are localized in frequency and orientation [14]. Hence by decomposing the images into different frequency ranges, it is possible to isolate the frequency components that are introduced by any deformations within the image or by external factors such as noise in the sub-bands. This allows even the minute changes in the image to be isolated particularly in the high frequency sub-bands. Therefore, DWT is a very useful tool to design a classification algorithm. During the wavelet transformation, frequency components in the image will be filtered out as approximations and details. The approximations are the high-scale, low-frequency components of the signal. The details are the low-scale, high-frequency components [21].

A level one decomposition of an image produces four sub-bands which are usually referred to as: low-Low (LL), low-high (LH), high-low (HL) and high-high (HH) where different properties of the image are represented in each of the sub-bands. Further decomposition is repeated only on the low-low sub-band of the image as most of the energy of an image is in the low frequencies (dyadic decomposition). Hence for a *n*- level DWT of an image there are *(3*n)+1* sub-bands [14]. In our approach we have performed a six level wavelet decomposition resulting in 19 sub-bands as shown in Fig 3.2.
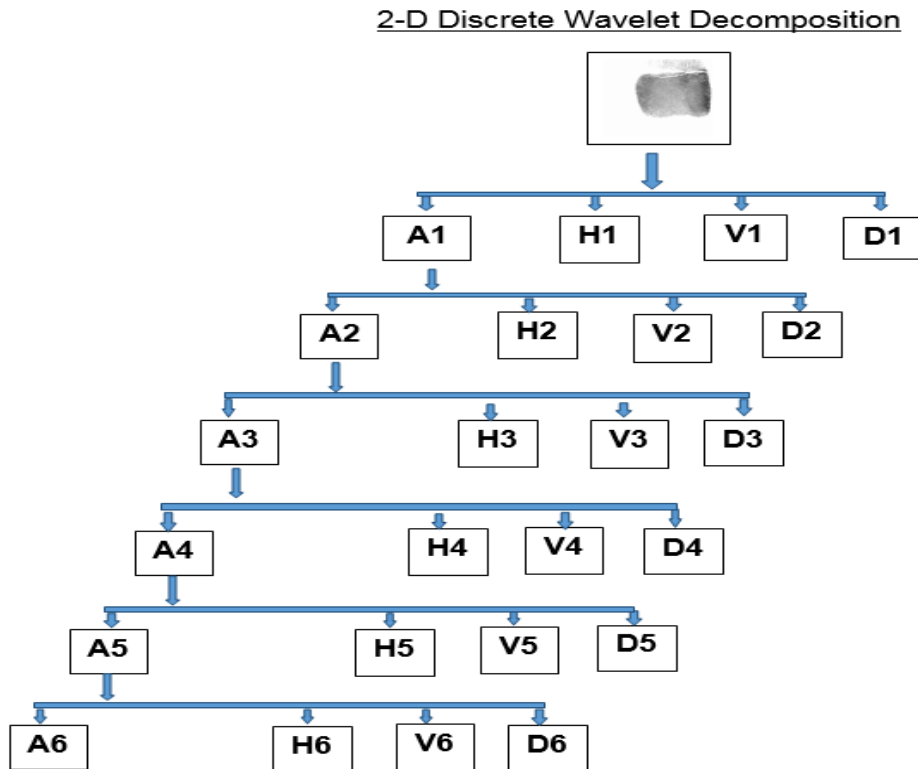
Fig 3.2 Six level 2-D DWT

3.4 Features Extracted

A total of 23 features have been used for classification out of which 10 features are discarded according to their ranking obtained after performing Info gain attribute evaluation.

The list of the 13 features that are used for classification are listed as follows:

- Standard deviation values of the noise in the image.
- The minutiae count values of the fingerprint.
- The log energy entropy values
- The NFIQ values as a representation of the image quality.

- Features 5 to7 are the standard deviation of detail coefficients of the first level of wavelet decomposition.
- Features 8 and 9 are the standard deviation of horizontal and diagonal detail coefficients of the second level of wavelet decomposition.
- Features 10 and 11 are the standard deviation of horizontal and diagonal detail coefficients of the third level of wavelet decomposition.
- Features 12 and 13 are the standard deviation of horizontal and diagonal detail coefficients of the fourth level of wavelet decomposition.

And the features that have been discarded because of their low info gain values are as follows:

- The values of Shannon Entropy.
- Features 2 to 4 are the standard deviation of horizontal, vertical and diagonal detail coefficients of the fifth level of wavelet decomposition.
- Features 5 to 7 are the standard deviation of horizontal, vertical and diagonal detail coefficients of the sixth level of wavelet decomposition.
- The standard deviation of vertical detail coefficients of the second level of wavelet decomposition.
- The standard deviation of diagonal detail coefficients of the third level of wavelet decomposition.
- The standard deviation of vertical detail coefficients of the fourth level of wavelet decomposition.

The NFIQ and minutiae count values are obtained from the National Institute of Biometrics Software of NIST which has a minutiae detection algorithm which detects the minutiae in the fingerprint, their quality and generates an image quality map. The NIST software also generates the NFIQ values which are used to assess the image quality. The NFIQ values are generally integers from 1 to 5 where 1 means the best quality image and 5 means the worst quality image. And the standard deviation values of the approximation and detail coefficients are obtained by performing a six level wavelet decomposition using DWT in MATLAB.

Classification for all the features extracted is implemented with the help of various classifiers such as naïve bayes, random forest and decision stump. All the classifiers are trained using 10-fold cross validation in Weka. Weka which stands for Waikato Environment for Knowledge Analysis is a open-source package for machine learning written in java developed by the University of Waikato. Weka is a collection of machine learning algorithms for data mining tasks which can be directly applied to a dataset. In addition to being suitable for developing new machine learning schemes weka also contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization [44].

3.5 Machine learning

Machine learning is a field of artificial intelligence that allows computers to learn and make predictions on data to improve future performance [37]. In this study we have used some basic machine learning algorithms for classification like the Naïve Bayes, Random Forest and Decision stump and from the classification accuracy we have observed that the Decision Stump algorithm performs better than both Naïve Bayes and Random Forest for our fingerprint data.

3.5.1 Naïve Bayes

The Naïve Bayes is a simple machine learning algorithm that is based on Baye's rule of conditional probability shown in equation

$$P_b(A/B) = \frac{P_b(B/A)\, P_b(A)}{P_b(B)}$$

Where,

$P_b(A)$ and $P_b(B)$ are probabilities of A and B that are independent of one another;

$P_b(A/B)$ is the conditional probability : probability of observing event A if B is true;

$P_b(B/A)$ is the probability of observing event B given that A is true.

By implementing Naïve Bayes learner for our set of features we were able to achieve a classification accuracy of 86.5462%.

3.5.2 Random Forest

The next algorithm used is the random forest algorithm. The random forest algorithm used in our study is developed by Breiman and Cultler [38]. The random forest algorithm is based on the development of a 'forest' of decision trees as its name suggests. This algorithm constructs a collection if decision trees with controlled variations by combining Breiman's 'bagging' idea with Tin Kam Ho's 'random subspace method' [39, 40, 41]. Random forests is a relatively complex algorithm than Naïve Bayes and also performs better than Naïve Bayes with a classification accuracy of 89.56%.

3.5.3 Decision Stump

The final and main algorithm used in our study is the Decision Stump. Decision stump is a weak learning algorithm that needs boosting using AdaBoost to improve classification results. A decision stump is basically a one level decision tree which is why it is often referred to as one stump.

We have boosted our decision stump learner with Adaboost and performed classification with the set of features we have extracted. As it can be seen from table 3.1. the level of decomposition has been gradually increased from third level DWT and it has been observed that the classification accuracy has increased with the increase in decomposition until six level DWT but from seven level DWT the accuracy of classification started to decrease which is why we have decided that a six level DWT is optimal for classification.

Table 3.1: Classification results for 3-6 levels of DWT using decision stump

| Decomposition	3rd level	4th level	5th level	6th level
results	91.3655%	92.3695%	93.1807%	95.3815%

At six level DWT with the features extracted and the decision stump algorithm for classification, we were able to achieve an accuracy of classification of 95.3815%. As can be seen from table 3.2 when the output of various machine learning algorithms for our set of features has been computed we get the best performance from the decision stump classifier.

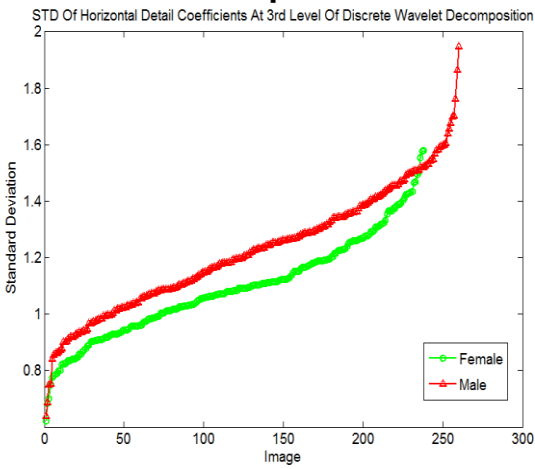Table 3.2: Comparison of the performance of various classification algorithms

S:No	Classification Algorithm	Iterations	Results
1	Decision stump	55	95.3815%
2	Decision Stump	10	94.9799%
3	Random Forest	55	89.5582%
4	Naïve Bayes	N/A	86.5462%
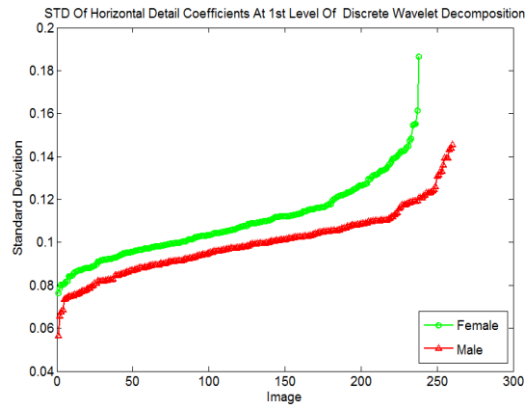
(a)
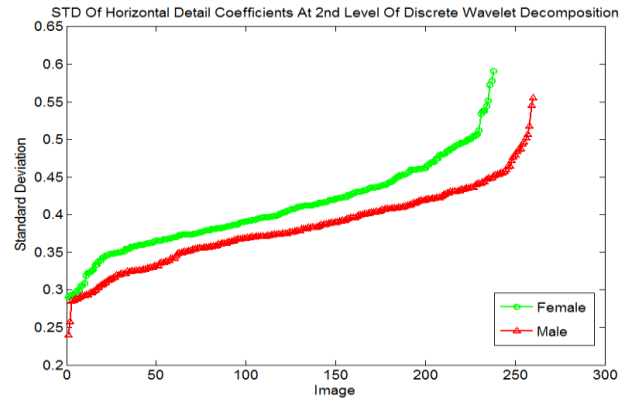
(b)

(c)
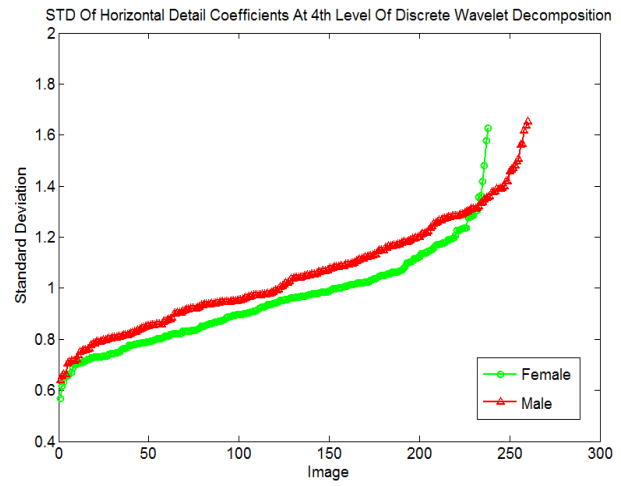
(d)
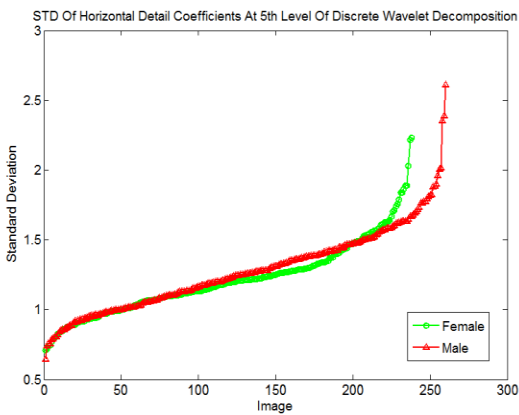
(e)
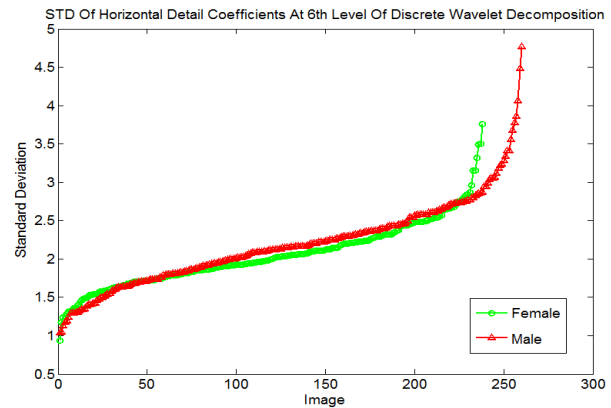
(f)

Fig 3.3: Standard deviation of horizontal detail coefficients for male and female.
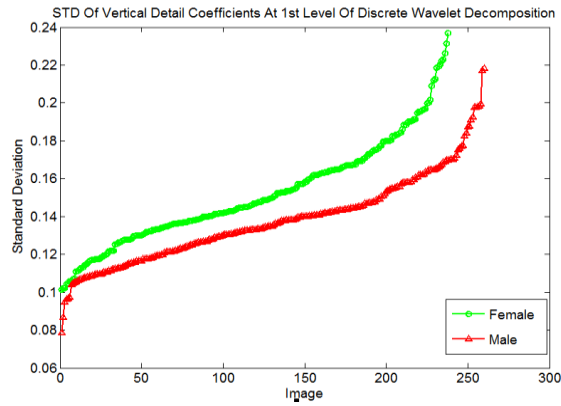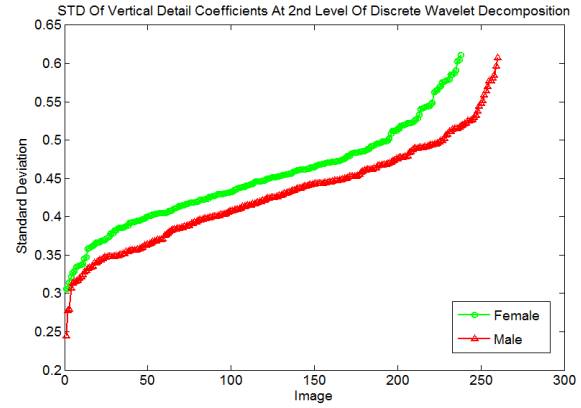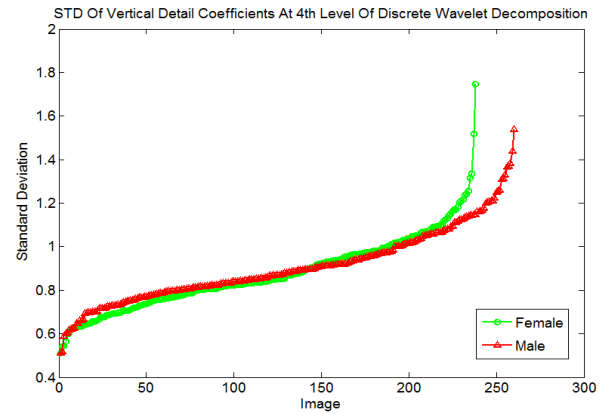
Fig 3.4 Standard deviation of vertical detail coefficients for male and female.

45

(a)   (b)
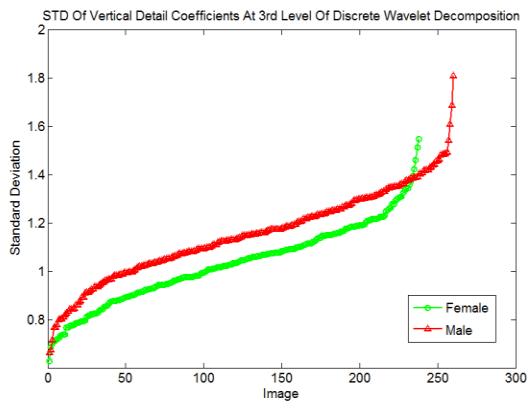
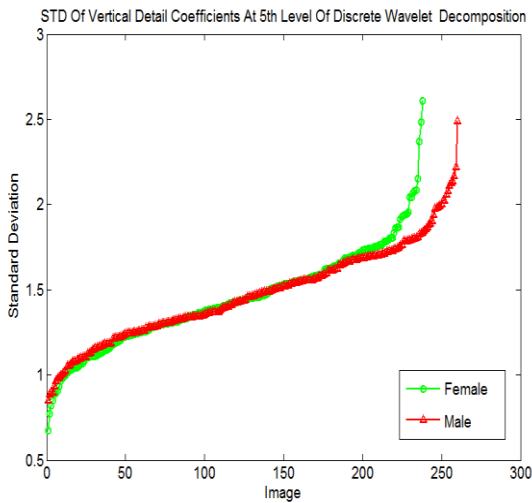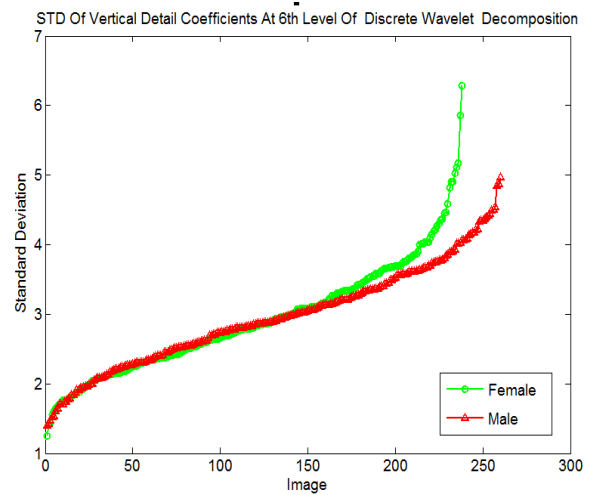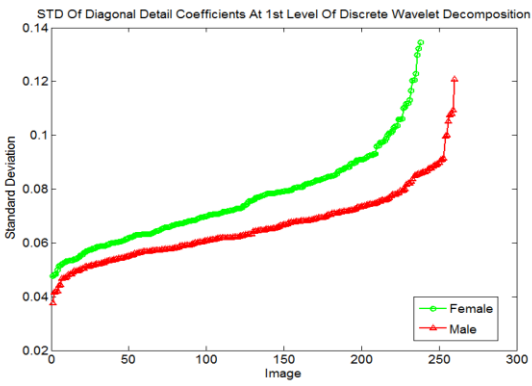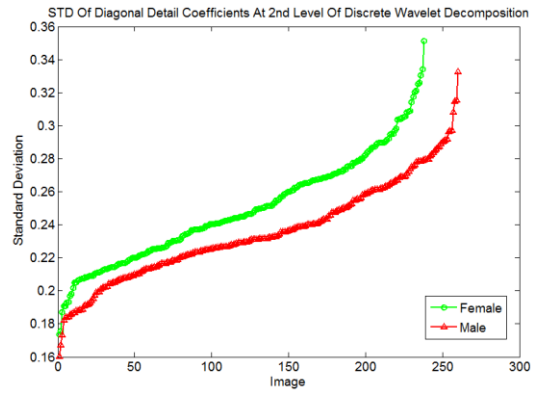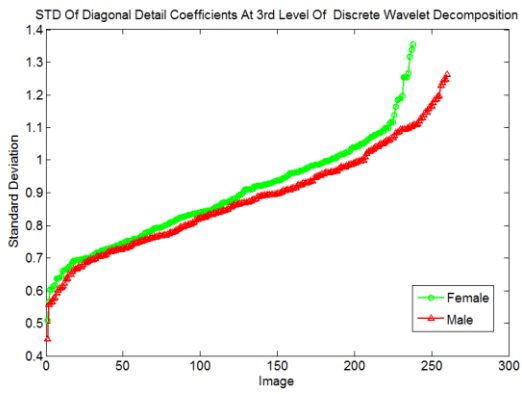(c)   (d)
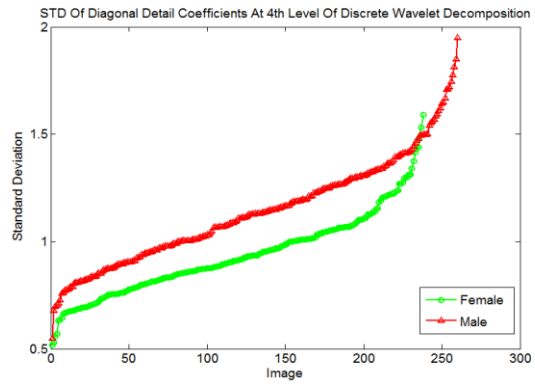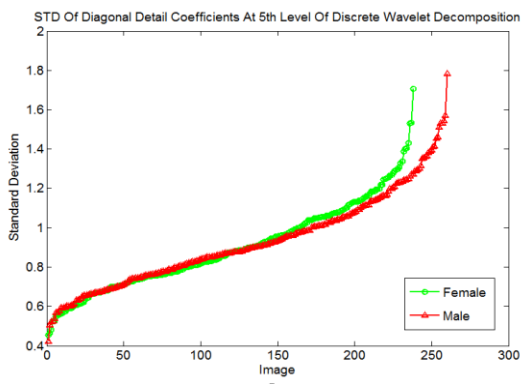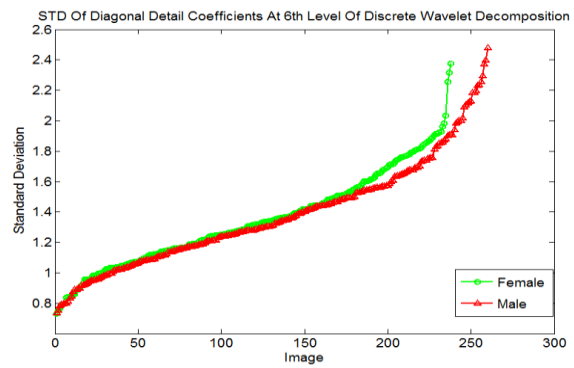
(e)   (f)

Fig 3.5 : Standard deviation of Diagonal detail coefficients for male and female.

It can be observed from the values of standard deviation from Fig 3.3,3.4 and 3.5 that the separation between male and female finger print images becomes almost insignificant as we increase the level of decomposition. But for wavelet decomposition levels one, two and three there is a significant difference between the standard deviation values based on gender and this case holds true for all horizontal, vertical and diagonal detail coefficients.

4. Entropy Based Approach

4.1 Dataset

The dataset used consists of fingerprints from an internal database collected here at WVU. Fingerprints were taken from a total of 498 users out of which 238 are female and 260 are male. All the fingerprints are collected from a live –scan optical sensor and have a resolution of 500 dpi.

4.2 Preprocessing

Before classification, the image is subjected to preprocessing. Preprocessing of the image helps to remove any unwanted information from the image such as noise and also helps to improve the classification accuracy. Preprocessing involves normalizing the fingerprint image and segmentation of the ridge region. The function used here identifies the ridge regions of a finger print image and it returns a mask identifying this region .The intensity values of the images are normalized so that the ridge regions have zero mean and unit standard deviation.

(a)   (b)
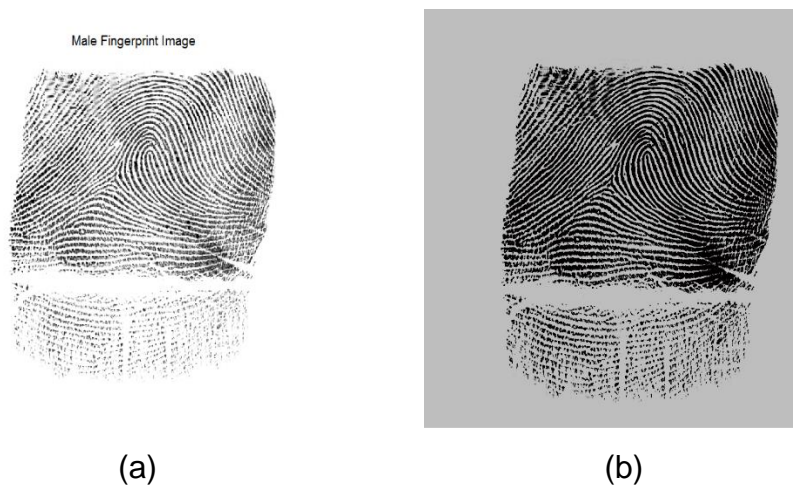
Fig 4.1 Male fingerprint image (a) Before Preprocessing (b) After Preprocessing

4.3 Entropy Approach

The wavelet packet decomposition is an extension of wavelet transform and therefore has the potential for more applications than that of simple wavelet transform. In the first level of a wavelet transform the signal is first decomposed into two sub-bands: an approximation and a detail. And for the second level of decomposition only the approximations are further decomposed into an approximation and a detail and the process is repeated for the next level and so on. Because the wavelet transform decomposes only the approximations of the signal, if we have any application at hand that contains significant information in higher frequency components we will not be able to extract any important features for classification by using just wavelet transform [27](As it has been explained in section 3.3 further decomposition in wavelet transform is applied only to the low-low sub-bands). In wavelet packet transform at the first level of decomposition we have two sub-bands: an approximation and a detail, similar to that of wavelet transform. But, at the second level of decomposition in a wavelet packet transform both the approximation as well as the detail sub-bands are further decomposed creating a full binary tree, this is the main difference between a wavelet transform (WT) and a wavelet packet transform (WPT). The WPT provides a better frequency resolution of the decomposed signal compared to WT. Hence, wavelet packet analysis not only gives better control over frequency resolution but also provides us with more features (i.e more information) about the signal than WT [28].

In this approach, WPD is performed on the input fingerprint image and the Shannon entropy and log energy entropy at each level of WPD is calculated as proposed by Coifman and Wickerhauser [22]. These values of entropy calculated are used as features for classification.

4.3.1 Shannon entropy

For every signal received, entropy is defined as the average amount of information in each signal which can also be stated as, entropy is an analysis of the amount of

uncertainty in the signal. Let $s = [s(0) \ s(1) \ s(2) \ldots\ldots s(N-1)]$ be a discrete random variable of finite length and with probability distribution function p(s) .The entropy of **s** is defined by

$$H(s) = -\sum_{i=0}^{N-1} p_i\ (s)\log_2(p_i(s)) \tag{7}$$

Where i indicates the ith discrete state. The Shannon entropy proposed by Shannon [10] is a useful standard for comparing and analyzing a probability distribution. For any distribution p(s), Shannon entropy (SE) is a measure of the information in the signal [22].

$$H_{SE}(s) = -E\{\log_2(p_i(s))\} \tag{8}$$

From this expression, we can define Shannon entropy as a measure of the average information content in **s**. In information theory Shannon entropy is defined as a measure of uncertainty in the system as follows [22].

$$H_{SE}(s) = -\sum_{i=0}^{N-1}(p_i(s))^2\ (\log_2(p_i(s)))^2 \tag{9}$$

4.3.2 Log Energy Entropy

Every image can be represented as a combination of different blocks, where each block is one source of the entire image. Log energy entropy is found to be more representative to the blocks of images.

Let $s = [s(0) \ s(1) \ s(2) \ldots\ldots s(N-1)]$ be a discrete random variable of finite length and with probability distribution function p(s). The log energy entropy of **s** is given by the following equation:

$$H_{LEE}(s) = -\sum_{i=0}^{N-1}(\log_2(p_i(s)))^2 \tag{10}$$

4.4 Features Extracted

The following set of features were extracted from each finger for classification purpose:

- 1 value corresponding to the minutiae count obtained from the NIST software.
- 24 values corresponding to the Log energy entropy of approximation and detail coefficients (horizontal, vertical and diagonal) at each level obtained from 6 level wavelet decomposition.
- 24 values of Shannon entropy obtained from approximation and detail coefficients (horizontal, vertical and diagonal) at each level obtained from 6 level wavelet decomposition.

The minutiae count values are obtained from the National Institute of Biometrics Software of NIST which has a minutiae detection algorithm which detects the minutiae in the fingerprint, their quality and generates an image quality map. And the values of Shannon entropy and log energy entropy are obtained by performing six level wavelet packet decomposition. We use MATLAB *wentropy* function to calculate the values of Shannon entropy and Log energy Entropy.

4.5 Classification

Before classification, we have performed attribute ranking by using info gain attribute evaluation on all the features and have only used the features which have info gain values greater than zero.
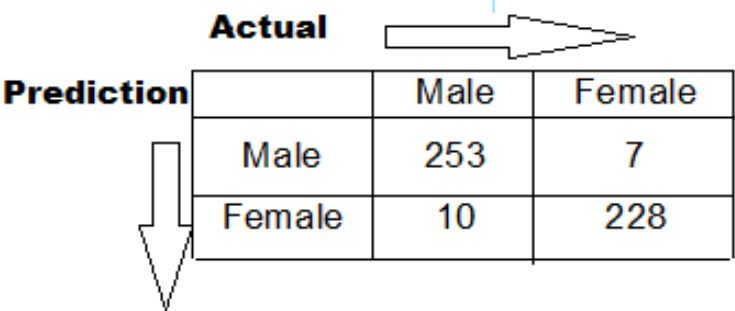
We have performed our experiments with different classification algorithms by training them using adaboost such as Random forest with 50 trees, Naïve Bayes and Decision stump, among which decision stump with 55 iterations with adaboost has produced the highest accuracy of 96.59%.

Adaboost has been used in order to decrease errors and improve accuracy. Adaboost is a machine learning algorithm which modifies the output of the algorithm with which it is being used, into the weighted sum of its initial output. We have used Ada-boost with various classification algorithms all of which have been trained with 10-fold cross validation. Weka 3.6 was used for classification. Weka which stands for Waikato Environment for Knowledge Analysis is a open-source package for machine learning written in java developed by the University of Waikato. Weka is a collection of machine learning algorithms for data mining tasks which can be directly applied to a dataset. In addition to being suitable for developing new machine learning schemes weka also contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization.[44]

4.6 Results

The confusion matrix which is the output of the the classification algorithm is shown in table 4.1 below . A confusion matrix also known as an error matrix is a visulaization of the performance the proposed algorithm. In the confusion matrix presented below each row represents the instances in an actual class and each column represents instances in a predicted class.

**Table4.1** Confusion matrix of Entropy Classification results

| Actual ⟹		Male	Female
Prediction ⇩	Male	253	7
	Female	10	228

As it can be seen clearly from table 4.1 Out of the 260 male users in the dataset the proposed classification algorithm was able to correctly classify 253 male users correctly and out of the 238 female users 228 of them were correctly classified. In terms of error classification 10 females were incorrectly classified as males and 7 males were iincorrectly classified as females. The output of the classier gives a classification of 263 males and 235 females in the given dataset instead of the actual 260 males and 238 females. A total of 17 users out of 498 were incorrectly classified by implementing the entropy approach for classification.

The results of the classifier can also be clearly seen from the plotted values of log energy entropy between the male and female fingerprints in Fig 4.2. It can be observed that there is a clear distinction between the log energy entropy values for both male and female with the values of log energy entopy being higher for male compared to female.
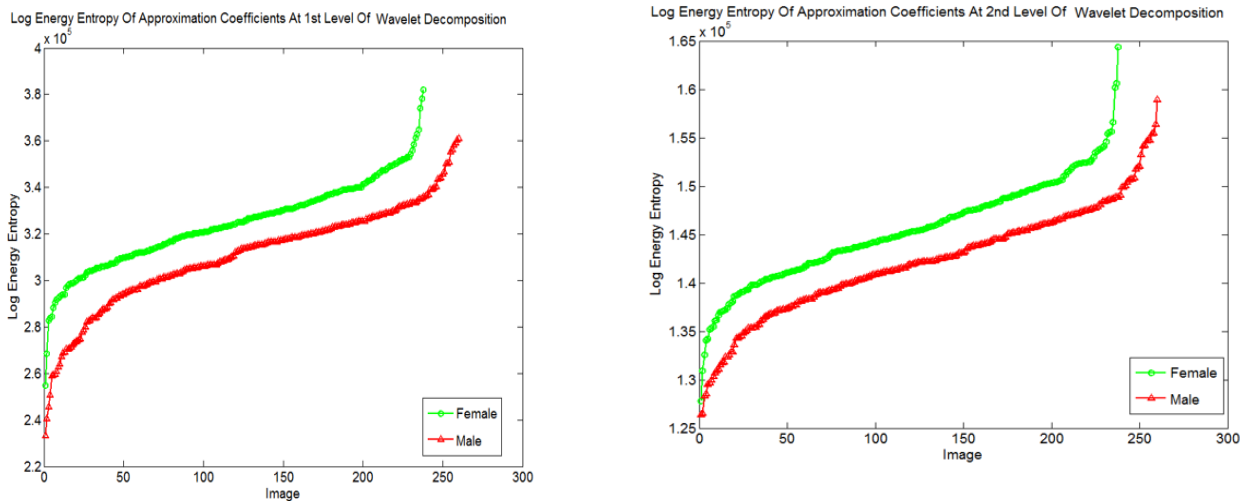

Fig 4.2: Male vs Female Plot of the Log Energy Entropy of Approximation coefficients at first and second level of wavelet decomposition

Like Log energy entropy it can be seen from Fig 4.3 that the fetures of Shannon entopy obtained by 6 six level WPD are also very distinctive for male and femal fingerprints and just like the log enegy entropy it can be observed that the values of Shannon entropy are are higher for female fingerprints than male fingerprints.
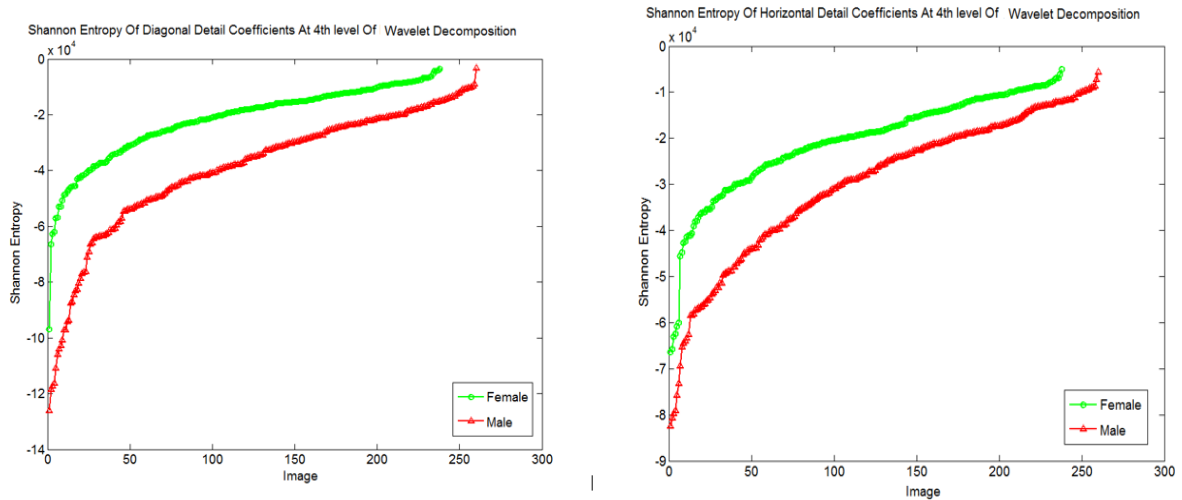
Fig 4.3: Male vs Female Plot of the Shannon Entropy of Diagonal and Horizontal detail coefficients at 4th level of wavelet decomposition

In addition to both Shannon entropy and log-energy entropy one other feature that has contributed significantly to the estimation of gender from fingerprints using Entropy approach is the fingerprint minutiae. Fig 4.4 shows a plot of the minutiae of our dataset.
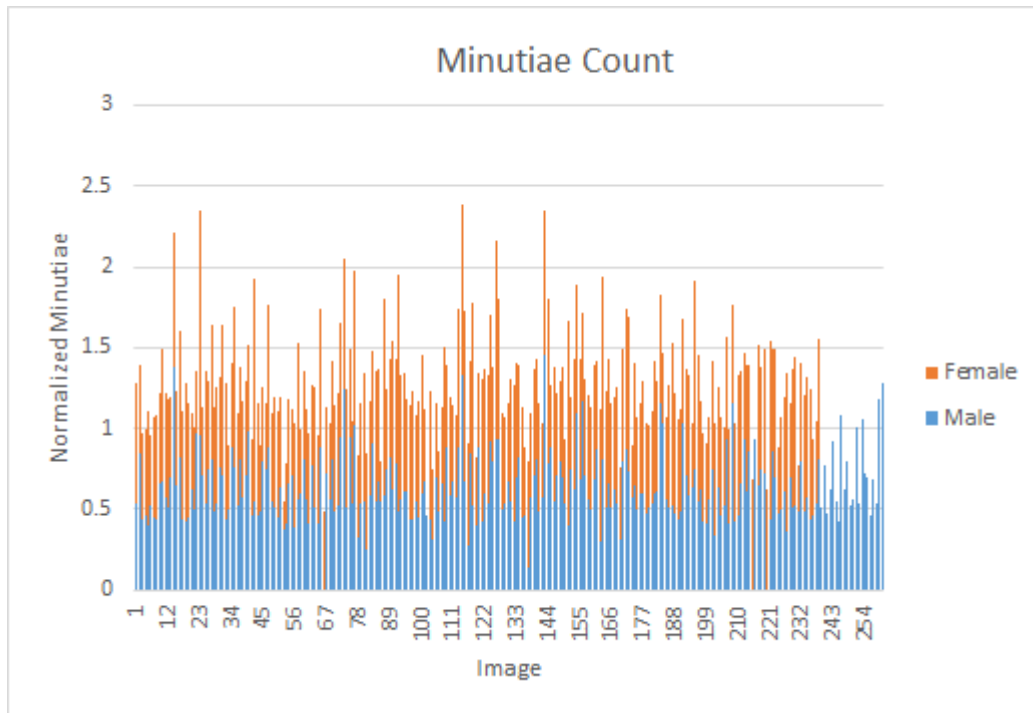
Fig 4.4: Plot of minutiae count of male and female fingerprints

5.Conclusions

5.1 Summary

Although Soft Biometric traits such as age, gender, ethnicity do not provide us with specific information about the gender of an individual they do provide useful information about the individual and have wide range of applications in the field of biometrics.

In this study we have proposed two different approaches to estimate the gender of an individual from fingerprint .The first approach uses features extracted from six level DWT and was able to achieve a classification accuracy of 95.38%.The second approach mainly uses features obtained from the values of Shannon and log energy entropy using six level WPD and has a classification accuracy of 96.59%.

5.2 Future Work

Although the two different algorithms proposed in this research were able to estimate gender from fingerprint data with an accuracy of more than 95%. The research may further be improved by fusion of the two approaches together.

Also it woulde be interesting to see if the same set of features can be used to estimate different biometric traits such as age and ethnicity from fingerprints.

References

[1] Anil K. Jain, Arun Ross, "Human Recognition Using Biometrics: An Overview," Ann. Telecommun., 2007, 62, (1), pp. 11–35.

[2] Anil K. Jain, Arun Ross, and Samil Prabhakar, "An introduction to biometric recognition," IEEE Transactions On Circuits And Systems For Video Technology, vol. 14, no. 1, pp. 4–21, 2004.

[3] P. Gnanasivam & Dr. S. Muttan "Estimation of Age through Fingerprints Using Wavelet Transform and Singular Value Decomposition" International Journal of Biometrics and Bioinformatics (IJBB), Volume (6): Issue (2): pp 58 - 67. 2012.

[4] A. Badawi, M. Mahfouz, R. Tadross, and R. Jantz "Fingerprint - based gender classification" The International Conference on Image Processing, Computer Vision, and Pattern Recognition, June 2006.

[5] Prabhakar, S., Pankanti, S., and Jain, A., "Biometric Recognition: Security and Privacy Concerns," IEEE Security & Privacy 1(2), 33–42 (2003)."

[6] E. Newton, L. Sweeney, and B. Malin, "Preserving Privacy by DeIdentifying Face Images," IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 2, pp. 232–243, 2005.

[7] Marasco, E., Lugini, L., and Cukic, B., "Exploiting Quality and Texture Features to Estimate Age and Gender from Fingerprints," SPIE Defense and Security 9075, 90750F–10 (2014).

[8] http://www.viewzone.com/fingerprintsx.html.

[9] http://sciencespot.net/Media/FrnsScience/fingerprintbasicscard.pdf

[10] Manish Verma and Suneeta Agarwal." Fingerprint Based Male - Female Classification." in Proceedings of the international workshop on computational intelligence in security for information systems (CISIS'08), Genoa, Italy, 2008, pp.251 – 257.

[11] Jen feng wang, et al, "Gender Determination using Fingertip Features". Internet Journal of Medical Update 2008 Jul – Dec; 3(2):22 – 8.

[12] S.Gungadin,"Sex determination from fingerprint ridge density", Int.J. Med. Update,2,2(2007), pp.1-4.

[13] P. Gnanasivam and S. Muttan, "Gender Identification Using Fingerprint through Frequency Domain Analysis," European Journal of Scientific Research, Vol. 59 No. 2, pp. 191-199, 2011.

[14] Gnanasivam, P. and Muttan, S., "Fingerprint Gender Classification using Wavelet Transform and Singular Value Decomposition.," International Journal of Biometrics and Bioinformatics (IJBB) 9(2) (2012).

[15] Marasco, E., Lugini, L., and Cukic, B., "Exploiting Quality and Texture Features to Estimate Age and Gender from Fingerprints," SPIE Defense and Security 9075, 90750F– 10 (2014).

[16] Ravi Wadhwa, Maninder Kaur, Dr. K.V.P. Singh "Age and Gender Determination from Finger Prints using RVA and dct Coefficients" IOSR Journal of Engineering (IOSRJEN) e-ISSN: 2250-3021, Issue 8 (August. 2013).

[17] Rijo Jackson Tom, T.Arulkumaran , "Fingerprint Based Gender Classification Using 2D Discrete Wavelet Transforms and Principal Component Analysis". International Journal of Engineering Trends and Technology, Volume 4 Issue 22 ISSN: 2231-5381, 2013.

[18] Heena Agarwal, Siddartha Choubey," Fingerprint based Gender Classification Using multi-class SVM", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3 Issue 8, August 2014

[19] Samta Gupta, A. Prabhakar Rao (2014). "Finger Print Based Gender classification using Discrete wavelet Transform & Artificial Neural Network", International Journal of Computer Science and mobile computing.

[20] Pallavi Chand, Shubhendu Kumar Sarangi," A Novel method for gender classification using DWT and SVD techniques", International Journal of Computer Technology & Applications, Vol 4 (3),445-449.

[21] https://en.wikipedia.org/wiki/Discrete_wavelet_transform

[22] R. Coifman and V. Wickerhauser, "Entropy-based algorithms for best basis selection," IEEE Trans. Informar. Theory, vol. 38, Mar. 1992

[23] Amara Graps. An introduction to wavelets. IEEE, 1995.

[24] D.T.L. Lee and A. Yamamoto, Wavelet Analysis: Theory and Applications, HP Journal, Dec 1994.

[25] https://en.wikipedia.org/wiki/Wavelet

[26] M. Sifuzzaman, M.R. Islam and M.Z. Ali "Application of Wavelet Transform and its Advantages Compared To Fourier Transform", Journal of Physical Sciences, Vol. 13, 2009, pp. 121-13.

[27] Shinde, A. D. (2004). *A wavelet packet based sifting process and its application for structural health monitoring*, Master Thesis, Faculty of Worcester Polytechnic Institute, pp. 22–32.

[28] Ekici, S., Yildirim, S., & Poyraz, M. (2008). Energy and entropy-based feature extraction for locating fault on transmission lines by using neural network and wavelet packet decomposition. Expert Systems with Applications, 34(4), 2937–2944.

[29] Osvaldo A. Rosso, Susana Blanco, Juliana Yordanova, Vasil Kolev, Alejandra Figliola, Martin Sch¨urmann, and Erol Bas¸ar. Wavelet entropy: a new tool for analysis of short duration brain electrical signals. Journal of Neuroscience Methods, 105(1):65 – 75, 2001.

[30] Shannon, C. E. Mathematical theory of communication. Bell Syst. Tech. J. 27:379–423, 623–656, 1948.

[31] P. Yildirim, "Filter Based Feature Selection Methods for Prediction of Risks in Hepatitis Disease," International Journal of Machine Learning and Computing, Vol. 5, No. 4, August 2015. DOI: 10.7763/IJMLC.2015.V5.517.

[32] Janecek, A.G.K., Gansterer, W.N., Demel, M.A., Ecker, G.F.: On the relationship between feature selection and classification accuracy. Journal of Machine Learning and Research 4, 90–105 (2008)

[33] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. Introduction to Data Mining. Addison Wesley, 1st edition, 2005

[34] http://machinelearningmastery.com/an-introduction-to-feature-selection/

[35] http://weka.sourceforge.net/doc.dev/weka/attributeSelection/InfoGainAttributeEval.html

[36] J. Novakovic, Using Information Gain Attribute Evaluation to Classify Sonar Targets, 17th Telecommunications forum TELFOR, 2009.

[37] https://en.wikipedia.org/wiki/Machine_learning.

[38] Leo Breiman and Adele Cutler, "Random forests," Technical report, Department of Statistics, University of California at
Berkeley, Berkeley, CA, 1999,

[39] Ho, Tin Kam. Random Decision Forest. Proc. of the 3rd Int'l Conf. on Document Analysis and Recognition, Montreal, Canada, August 14-18, 1995,278-282,1995

[40] Ho, Tin Kam. The Random Subspace Method for Constructing Decision Forests.IEEE Trans. on Pattern Analysis and Machine Intelligence 20 (8),
832-844, 1998

[41] Breiman, Leo. Random Forests. Machine Learning 45 (1), 5-32, 2001

[42] http://www.conceptualwavelets.com/docs/wavelets_ch1.pdf

[43] https://ncforensics.wordpress.com/2013/06/20/techniques-for-collecting-and-analyzing-fingerprints/

[44] http://www.cs.waikato.ac.nz/ml/weka/index.html

[45] http://rob.schapire.net/papers/explaining-adaboost.pdf

[46] Z.-H. Zhou. Ensemble learning. In S. Z. Li, editor, Encyclopedia of Biometrics. Springer, Berlin, 2008

[47]   http://machinelearningmastery.com/improve-machine-learning-results-with-boosting-bagging-and-blending-ensemble-methods-in-weka/.

[48] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 55(1):119–139, 1997.

[49] Y. Freund and R. E. Schapire. A short introduction to boosting. Journal of Japanese Society for Artificial Intelligence, 14(5):771–780, 1999.

[50] http://www.graphicon.ru/html/2005/proceedings/papers/vezhnevetz_vezhnevetz.pdf

[51] http://rob.schapire.net/papers/explaining-adaboost.pdf

[52] www.mathworks.com/help/wavelet/gs/introduction-to-the-wavelet-families.html

[53]latunyj.wordpress.com/2011/12/30/generating-daubechies-wavelet-function-psi-and-wavelet-scaling-phi-using-matlab/

[54] https://home.cc.umanitoba.ca/~krussll/138/sec4/acoust1.htm