Optimum Feature Selection for Signature Verification using Different Classifiers

A Dissertation

Submitted in partial fulfillment of the requirements for the award of degree of

Master of Engineering

In

Electronic Instrumentation and Control

Submitted By:

Aman Kataria

(801151002)

Under the esteemed guidance of

Mr. Mandeep Singh
Assistant Professor
Department of Electrical and Instrumentation Engineering

ELECTRICAL AND INSTRUMENTATION ENGINEERINGDEPARTMENT

THAPAR UNIVERSITY

(Established under the section 3 of UGC act, 1956)

PATIALA, 147004, Punjab, India

JUNE 2013

DECLARATION

I hereby certify that the work which is being presented in the dissertation entitled as **"Optimum Feature Selection for Signature Verification using Different Classifiers"** in partial fulfillment of award of degree of **Master in Engineering in Electronic Instrumentation and Control** submitted in Electrical Instrumentation Department, Thapar University, Patiala, is an authentic record of my own work carried out under the supervision of **Mr. Mandeep Singh**, Assistant Professor, Department of Electrical Instrumentation Engineering, Thapar University, Patiala.

Aman Kataria

Date 15.07.13

Aman Kataria

(801151002)

I certify that the above statement made by the student is correct to the best of my knowledge and belief.

Date 15.07.13

(Mr. Mandeep Singh)

Assistant Professor, EIED

Thapar University,

Patiala

Countersigned by

S. Ghosh

Dr. Smarajit Ghosh

Head of Department, EIED

Thapar University

Patiala, Punjab

Dr. S.K. Mohapatra

Dean of Academic Affairs

Thapar University

Patiala, Punjab

ACKNOWLEDGMENT

I would like to express my sincere gratitude to my supervisor, **Mr. Mandeep Singh,** Assistant Professor (EIED), for all his guidance and invaluable advises throughout the progress. He has stimulated my interest in Biometrics and Data mining by inspiring me for doing research on this topic.

I would also like to thank **Dr. Smarajit Ghosh**, Professor & Head, Electrical and Instrumentation Engg. Department and **Dr. S.K. Mohapatra**, Dean, Academic Affairs for giving an opportunity to work in this regard.

Also I would like to thank my parents and my friends who were always there for me. It would not have been possible without their continuous support and encouragement.

(Aman Kataria)

ABSTRACT

Signature verification is an important parameter in the biometrics area as it is one of the behavioral biometric traits, which are widely used as a means of personal verification. Therefore, they require efficient and meticulous methods of authenticating the users. Features used in this work are employed for performance analysis using different combinations among them. For optimum feature selection Fisher's Discrimination ratio (FDR) is calculated for each individual feature and according to the calculated FDR, best feature is identified. Three classifiers: k-Nearest Neighbor (k-NN) classifier, Random Forest classifier and Naïve bayes classifier are used to verify whether the signatures are real or forgery. The results of all the classifiers are compared.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 OVERVIEW

Numerous features of human physiology are used to ratify a person's identity. The science of ascertaining and verifying the identity with respect to different characteristics traits of human being is called biometrics. The characteristic traits can be extensively classified into two categories i.e. physiological and behavioral. Measurement of physical features for personal identification is an age old practice which dates back to the Egyptians era [1]. But it was not until 19th century that the study of biometrics was extensively used for personal identification and security related issues. A person can be identified on the basis of different physiological and behavioral traits like fingerprints, faces, iris, hand geometry, gait, ear pattern, voice recognition, keystroke pattern and thermal signature. Biometric identifiers are the idiosyncratic, measurable characteristics used to tag and describe individuals. Biometric identifiers are often categorized as physiological versus behavioral characteristics. Physiological characteristics are related to the shape and contours of the body. Examples include, but are not limited only to fingerprint, face recognition, DNA, Palm print, hand geometry, iris recognition, retina and odor/scent. Behavioral characteristics are related to the behavior of a person, including but not limited to typing rhythm, gait, and voice. Signature has been a detecting and particularizing feature for person identification. Even today an proliferating number of transactions, remarkably related to financial and business are being authorized via signatures. Approaches to signature verification fall into two categories according to the acquisition of the data: On-line and Off-line. On-line data records the motion of the stylus (which is also part of the sensor) while the signature is produced, and includes location, and possibly velocity, acceleration and pen pressure, as functions of time. Online systems use this information captured during acquisition. These dynamic characteristics are specific to each individual and sufficiently stable as well as repetitive. Off-line data is a 2-D image of the signature. Off-line processing is intricate due to the scarcity of stable dynamic characteristics. Difficulty also lies in the fact that it is hard to segment signature strokes due to highly snazzy and unconventional writing styles. The nature and the assortment of the writing pen may also affect the nature of the signature obtained. The non-repetitive nature of variation of the signatures, because of age, illness, geographic location and perhaps to some extent the emotional state of the person, accentuates the problem. A robust system has to be designed which should not only be able to consider these factors but also classify various types of forgeries [2, 3].

1.2 OBJECTIVE OF THE DISSERTATION

The objective of this thesis is to classify the signatures into real and forgery type. Also Study of various signature classification techniques and implementing k-NN classifier. In this work comparison k-NN classifier with other classifiers like Naïve bayes and Random forest have been done along with their accuracy. Also literature survey of k-Nearest neighbor in different areas has been mentioned in this work.

1.3 ORGANIZATION OF THE DISSERTATION

The structure of the dissertation is as follows:

Chapter 1 This chapter describes the introduction part of the work (Signature classification) with the comparison among the different classifiers.

Chapter 2 This chapter depicts the literature survey of Signature verification using different approaches.

Chapter 3 This chapter describes the introduction to the biometrics along with the techniques and methods of authentication.

Chapter 4 In this chapter, various biometrics parameters are discussed along with their advantages and real time applications.

Chapter 5 This chapter presents overview of the parameters of signatures, types of signature along with their verification methods also types of forged signatures has been mentioned.

Chapter 6 This chapter presents the method of extraction of features of signature upon which performance analysis has been done.

Chapter 7 This chapter forms an objective of thesis. The work is carried out using K-Nearest neighbor algorithm and other classifiers along with the different combinations of the features are used for classification of signatures.

Chapter 8 This section deals with analysis of all the conclusions and discussions which has been done in the presented work so far with a brief overview on the future possibility of the remaining work in the same domain.

Chapter 2

Literature Review

This chapter deals with various research works that were carried in past and contributed to this field of Signature verification. Different approaches were deployed to carry out this work. Some of the related research works carried out in this regard is described below.

2.1 NEURAL NETWORKS APPROACH

Jose F. Velez, (2003) presented a approach which is widely used in signature verification systems, power, ease of use, capabilities in earning and generalizing are the main reasons to use this approach. When using this approach we have to structure the Neural Network (NN) by removing features from signers' samples and learning the relationship between the signature and its class. There are many ways to structure the NN training, but a very simple approach is to firstly extract the feature set representing the signature (details like length, height, duration, etc.), with several samples from different signers. The second step for the NN is to learn the relationship between a signature and its class (either "genuine" or "forgery"). Once this relationship has been learned, the network can be presented with test signatures that can be categorized as belonging to a particular signer. NNs therefore are highly suitable to modeling global aspects of handwritten signatures. In this method, the Back Propagation Neural Network is used as a tool to estimate the performance of the proposed method. The system has been tested with 70 test signatures from different persons revealing the recognition rate of 91.34% [4].

2.2 HIDDEN MARKOV MODELS APPROACH

E. J. R. Justino, (2003) Used an approach known as Hidden Markov Model (HMM) which is one of the most extensively used models for sequence analysis in signature verification. Handwritten signature is a sequence of vectors of values related to each point of signature in its route. A system was developed that automatically authenticates offline handwritten signatures using the Discrete Radon Transform (DRT) and a hidden Markov model (HMM). Given the robustness of the algorithm and the fact that only global features are considered, the system achieves an Equal Error Rate (EER) of 18% when only high-quality forgeries (skilled forgeries) are considered and an EER of 4.5% in the case of only casual forgeries. Average Accuracy obtained was 84.13% [5].

2.3 STATISTICAL APPROACH

Debnath Bhattacharyya (2007), presented a Statistical Approach for offline and written signature verification. Generally, it follows the concept of Correlation Coefficients which refers to the departure of the two variables from independence. In signature verification system, average signature (template) is calculated from previously collected signatures, stored in knowledge base, when new input signature is read, correlation concept is followed to find the distance between the test signature and average signature, then to decide if it is accepted or rejected. The algorithm proposed has the flexibility of choosing the number of signatures, i.e., no-of-Sign for testing purpose to generate a signature as a Avg-Sign containing the specialized mean features set from the test signature set. After collecting signatures for testing, the algorithm converts them into a set of 2D arrays of binary data values-0 and 1. From these binary arrays using statistical methods of calculating the expected mean an average data set. The Recognition scheme is based on an extensive Statistical Analysis of Correlation Coefficient between bivariate data set. In implementation of proposed algorithm to constant factors carry major impact on the validity of the method and the strength of the verification lies in the efficiency of selection of these constant parameters, namely AvgSign, Threshold Value and decision value. Average Accuracy obtained was 89.93% [6].

2.4 STRUCTURAL APPROACH

Ramachandra (2009) presents the new method in which structural features are extracted from the signature's contour using the (MDF) and its extended version: the Enhanced MDF (EMDF) and further two neural network-based techniques and Support Vector Machines (SVMs) are investigated and compared for process of the signature verification. The main idea in structural pattern recognition is the representation of patterns by means of symbolic data structure such as trees, graphs and strings. When a forged signature comes, its symbolic representation is compared with prototypes stored in database. In other words, Structural approach is based on the relational organization of the low-level features into higher-level structures, and then theses structures are matched with models stored in database. Structural features use Modified direction and transition Distance Feature (MDF) which extracts the transition locations and are based on the relational organization of low-level features into higher- level structures. The classifiers were trained using genuine specimens and other randomly selected signatures taken from the publicly available database of 3840 genuine signatures from 160 volunteers and 4800 targeted forged signatures. A Distinguishing Error Rate (DER) of 17.78% was obtained with a SVM whilst keeping the false acceptance rate for random forgeries (FARR) below 0.16%. Geometric features of the signature in fixed-point arithmetic for offline verification were calculated. The proposed features are then checked with different classifiers, such as the Hidden Markov Models, the Support Vector Machines, and the Euclidean distance verifier. The results show that HMM works slightly better than SVM and the distance Euclidean verifier, but, bearing in mind that the SVM and Euclidean distance-based verifiers can be programmed in a fixed-point microprocessor, the results encourage us to follow the SVM research line in order to built a smart card capable of detecting a simple forgery. Average Accuracy obtained was 76.6 % [7].

2.5 WAVELET- BASED APPROACH

In general, the multi-resolution wavelet transform can decompose a signal into low pass and high pass information. The high pass information usually represents features that contain sharper variations in time domain. Hou and Feng proposed (2005) the method uses a wavelet-based transformation to extract the inflections of the signature curves by using different scale wavelet transforms in the curvature signature signals transformation. After analysis, the proper scale is selected. The zero-crossings points are mined and are taken as the inflections of the signature. Then this signature curves are divided into several parts, i.e. the strokes, according to the above inflections. The distance between the two corresponding strokes can be measured with Dynamic Time Warping algorithm. In the end, the training algorithm of the signature verification system also the verification method of the signatures is also introduced. The experimental result shows that this method is superior to those methods that match the whole signature curves. Samaneh and Mohsen (2009), presented a method for offline Persian signature identification and verification based on image registration and fusion. Discrete Wavelet Transform (DWT) is applied on the preprocessed signature to get high frequency sub-images, then an image reduction and fusion methods are used to create a feature matrix from sub-images. In verification phase, for test signature; the feature matrix is compared with all feature matrixes stored in knowledge base using Euclidian distance. And then upon the specified threshold, the test signature would be accepted or rejected Larkins and Mayo (2008) , presented an offline signature verification method based on Adaptive Feature Thresholding (AFT). They converted the signature image to binary feature vector; by using the above conversion, the comparison was more accurate. That vector was based on gradient direction for each pixel from across a signature. Gradient direction gave a global features level; Spatial Pyramids were used to express a signature at deep levels, Equimass sampling grid with Spatial Pyramids were combined to improve the structural features. In classification phase, DWT and graph matching methods were used. Average Accuracy obtained by wavelet based approach was 64.4 % [8].

Summary

According to the above literature survey Neural approach gave the best results and Wavelet based approach gave the unsatisfactory results. Research in the improvement of Hidden Markov Model is in progress. On the other hand statistical approach employ methods to perform some of the statistical concepts like the relation, deviation, etc between two or more data items to find out a specific relation between them. Results obtained by Structural approach are satisfactory.

Chapter 3

Biometrics

3.1 INTRODUCTION

This chapter deals with the discussion of Biometrics Humans have used body characteristics such as face, voice, and gait for thousands of years to recognize each other. Alphonse Bertillon, chief of the criminal identification division of the police department in Paris, developed and then practiced the idea of using a number of body measurements to identify criminals in the mid-19th century. Just as his idea was gaining popularity, it was obscured by a far more significant and practical discovery of the distinctiveness of the human fingerprints in the late 19th century. Soon after this discovery, many major law enforcement departments embraced the idea of first “booking” the fingerprints of criminals and storing it in a database. Later, the leftover fingerprints at the scene of crime could be “lifted” and matched with fingerprints in the database to determine the identity of the criminals [9].

Although biometrics emerged from its extensive use in law enforcement to identify criminals it is being increasingly used today to establish person recognition in a large number of civilian applications. Biometric is used to identify the identity of an input sample when compared to a template, used in cases to identify specific people by certain characteristics. Possession based: using one specific "token" such as a security tag or a card and knowledge-based: the use of a code or password. Standard validation systems often use multiple inputs of samples for sufficient validation, such as particular characteristics of the sample. This intends to enhance security as multiple different samples are required such as security tags and codes and sample dimensions. So, the advantage claimed by biometric authentication is that they can establish an unbreakable one-to-one correspondence between an individual and a piece of data.

The authentication system has been utilized in various areas in the industry as well as in military and in the e-commerce etc. Unlike contemporary, passwords and ID cards have been used to authentication to systems. However, security can be easily break in these systems when a password is leaked to an unauthorized user or a card is stolen by an impostor; further, simple passwords are easy to guess (by an impostor) and difficult passwords may be hard to recall (by a genuine user). The emergence of biometrics has addressed the problems that plague traditional verification methods. Biometrics refers to the automatic identification (or verification) of an individual (or a claimed identity) by using certain physiological or behavioral traits associated with the person. By using biometrics it is possible to establish an identity based on 'who you are', rather than by 'what you possess' (e.g., an ID card) or 'what you remember' (e.g., a password). Biometric systems make use of fingerprints, hand geometry, iris, retina, face, hand vein, facial thermo grams, signature, voiceprint, palm print, etc. to establish a person's identity [9]. While biometric systems have their limitations, they have an edge over traditional security methods in that it insignificantly difficult to lose, steal or forge biometric traits [10].

3.2 TECHNIQUES OF BIOMETRIC AUTHENTICATION

A person can be identifies on the basis of "what he discerns? (Password etc)", "what he possess? (ID card, key etc)" and "what he is? "(Biometrics)". On the basis of above facts the authentication methods are classified in the following three ways as shown in the Figure 1.

a) Knowledge based Authentication method.

b) Token based Authentication method.

c) Biometrics Authentication method.

The knowledge based and token based authentication systems are called conventional authentication system [11].

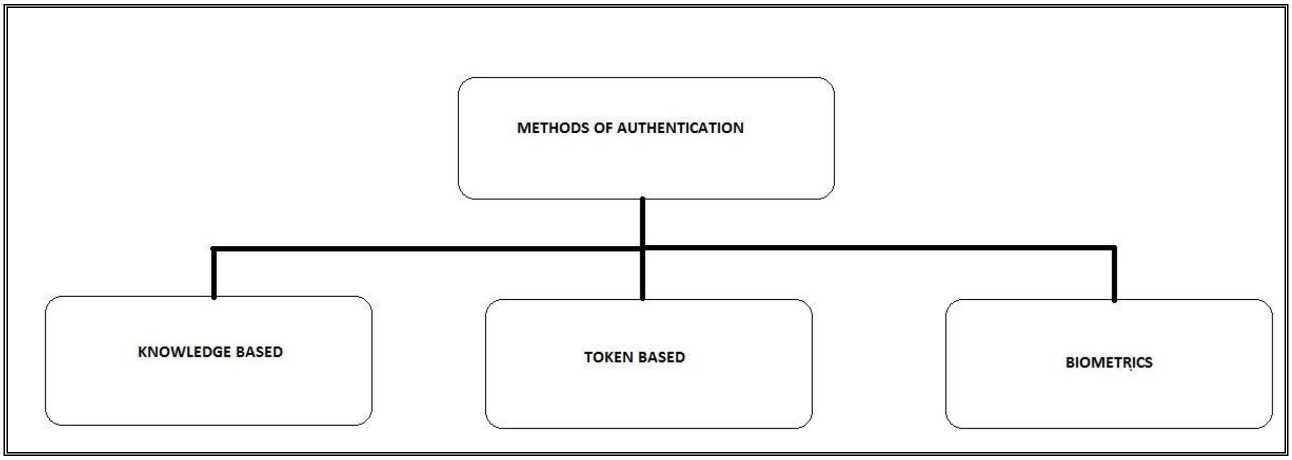


Figure 1 Methods of authentication

3.2.1 KNOWLEDGE BASED AUTHENTICATION METHOD

This type of authentication rely upon the knowledge of a person i.e. what he possess? It comes under the traditional authentication system. It is also known as password-based method. In the password-based scheme the user submits a password, which is normally passed through a one-way hash function. This, along with the identity is saved in a database during enrollment. During verification, the user submits the password, whose hash value is calculated, the hash value retrieved from the database. A “Yes" decision is the output if and only if the two hashes are the same. Otherwise a “no” decision is the output. The diagrammatical representation of this method is shown below by the Figure 2 and Figure 3. Figure 3 shows the enrollment process of knowledge based authentication method and Figure 3 shows the verification process of knowledge based system [12].

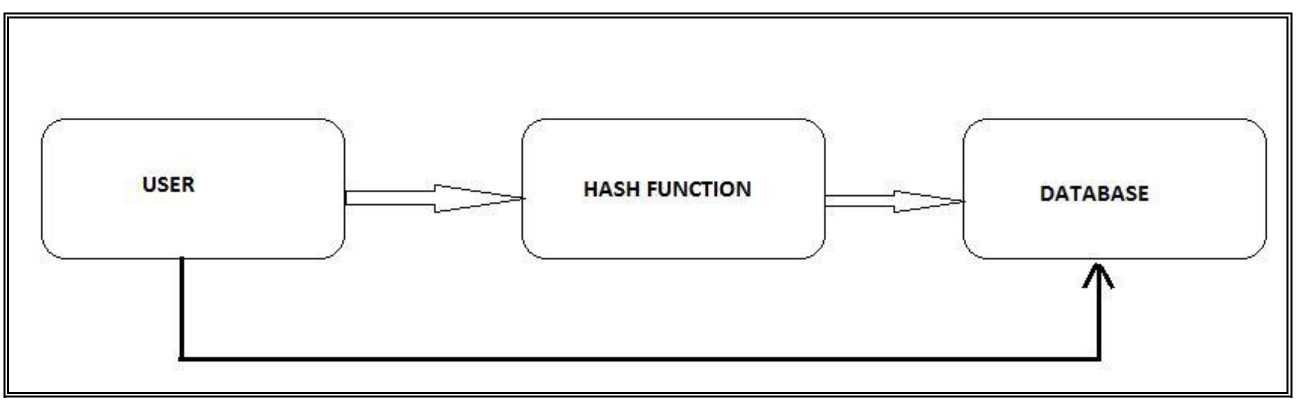


Figure 2 Enrollment process of knowledge based authentication method

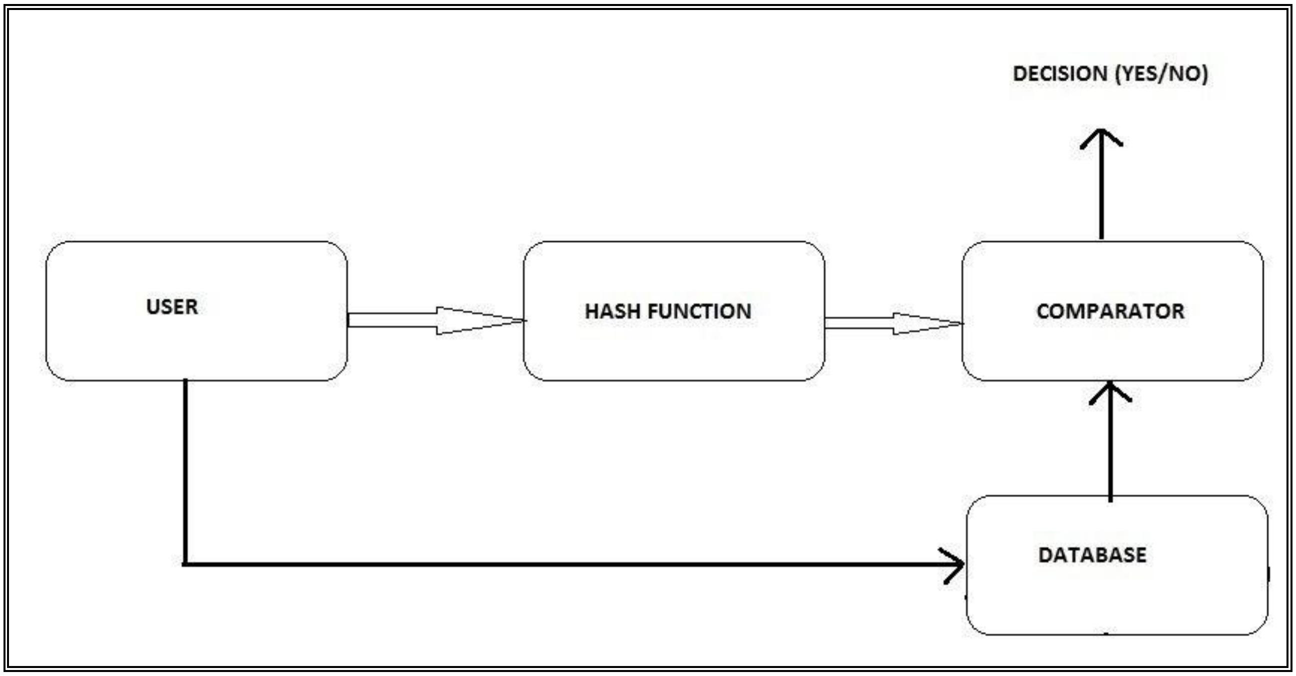


Figure 3 Verification process of knowledge based authentication method

3.2.2 TOKEN BASED AUTHENTICATION METHOD

This type of authentication rely upon the knowledge of a person i.e. what he possess? It also comes under the traditional authentication system. It is also known as token-based method. In the Token-based scheme, the user's identity and credentials (e.g., her birth certificate, diploma, signature) are verified by the authenticating establishment (e.g., a university registrar's office), which stockpile this information in a database. This token can also embodies entities to bind the token only with the authenticating establishment, such as an embossed seal, specific logo, or bar code to eliminate illegal reproduction of the token [13].

In a supervised authentication application (where a human companion is available), the user's card and her credentials (e.g., signature) are verified for constancy by a (human) supervisor. If these data match, the supervisor accepts the user. In an unsupervised application, just the existence of the card is monitored. So we can say that supervised authentication is more efficient and accurate than unsupervised authentication system. The diagrammatically representation of this scheme is given below by the help of Figure 4 and Figure 5 respectively [14].

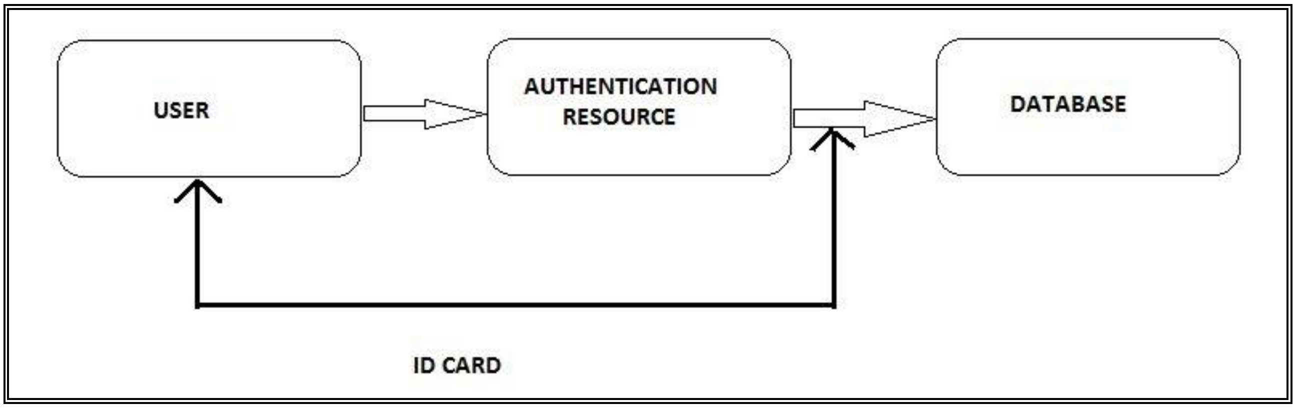


Figure 4 Enrollment process of Token based authentication method.

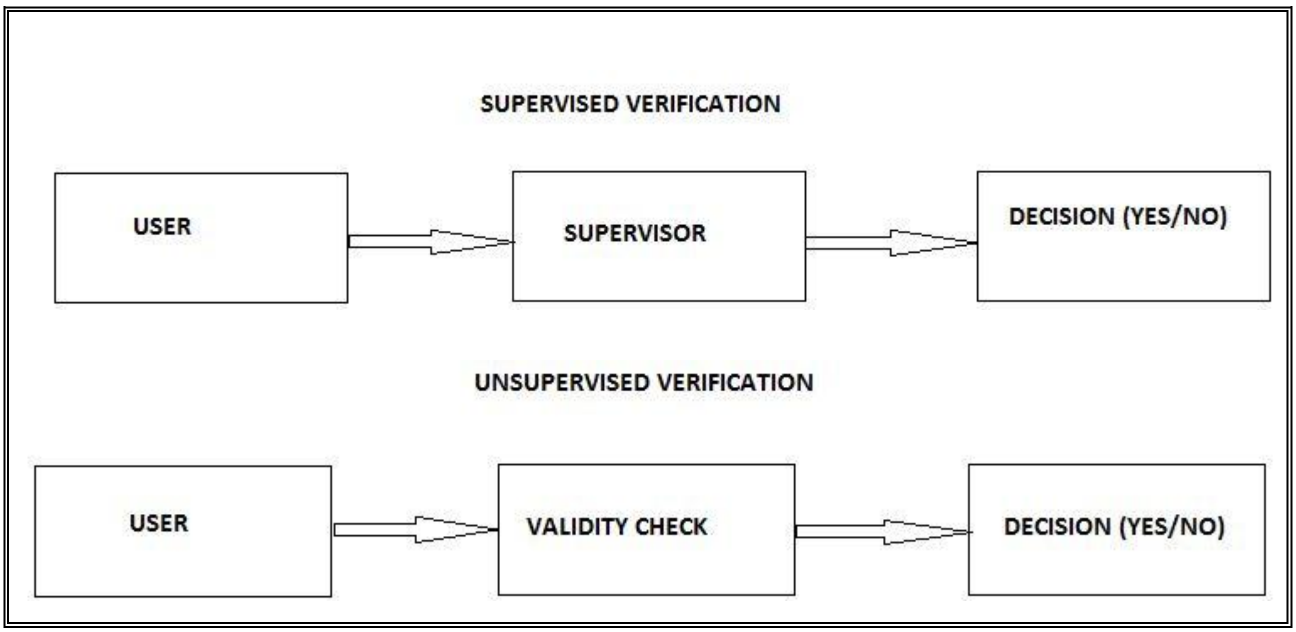


Figure 5 Verification process of Token based Authentication method

3.2.3 BIOMETRICS AUTHENTICATION METHOD

Biometrics is basically a pattern reorganization system that is used to identify or verify users based on his or her unique physical characteristics. Traditional techniques for identification and

verification are based on ‘What you know’ such as passwords, PINS which can be forgotten and ‘What you have’ such as tokens, ID which can be stolen or lost. In the enrollment process for an individual, using a biometric to be identified or verified, a copy of the individual’s signature has to be stored in a database at the enrollment stage. Figure 6 presents the block-diagram of the enrollment stage. Biometric identification is based on physiological or behavioral characteristics that provide information about ‘Who you are’. And the output is the person identity. Figure 7 presents the block diagram of identification. Biometric verification is based on physiological or behavioral characteristics that provide the decision accepted or rejected. And the output is “yes” if accepted and “No” if rejected. Figure 8 presents the block diagram of verification [15].

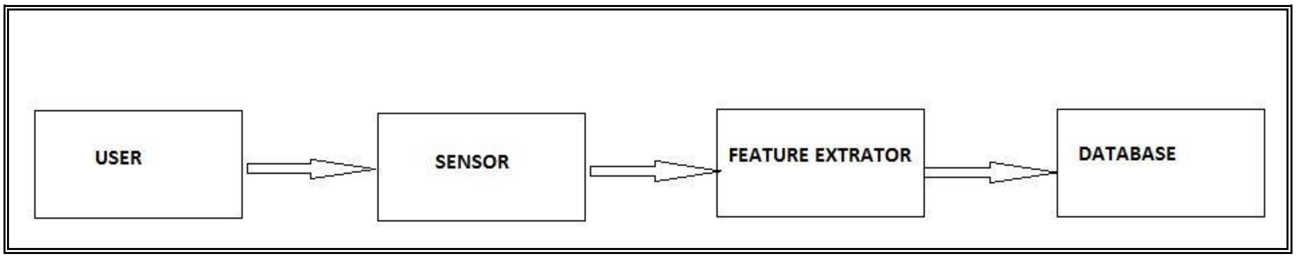


Figure 6 Enrollment process of Biometrics authentication method

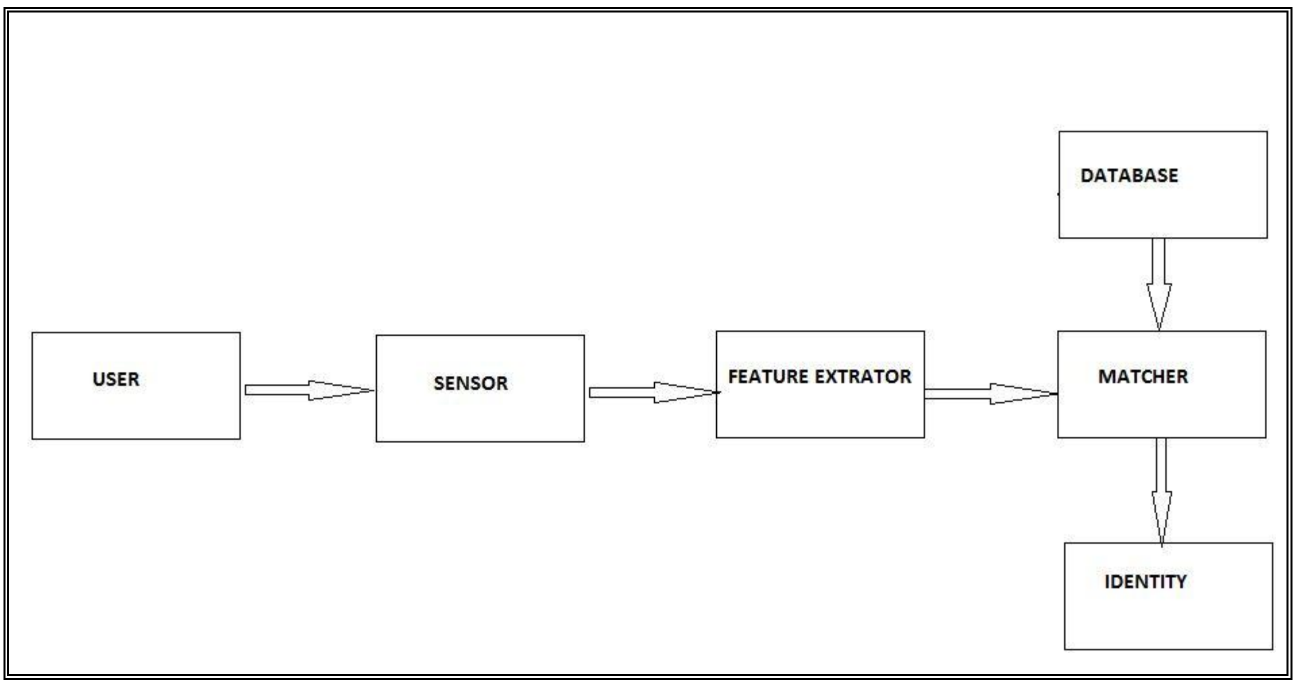


Figure 7 Identification process of Biometrics authentication method

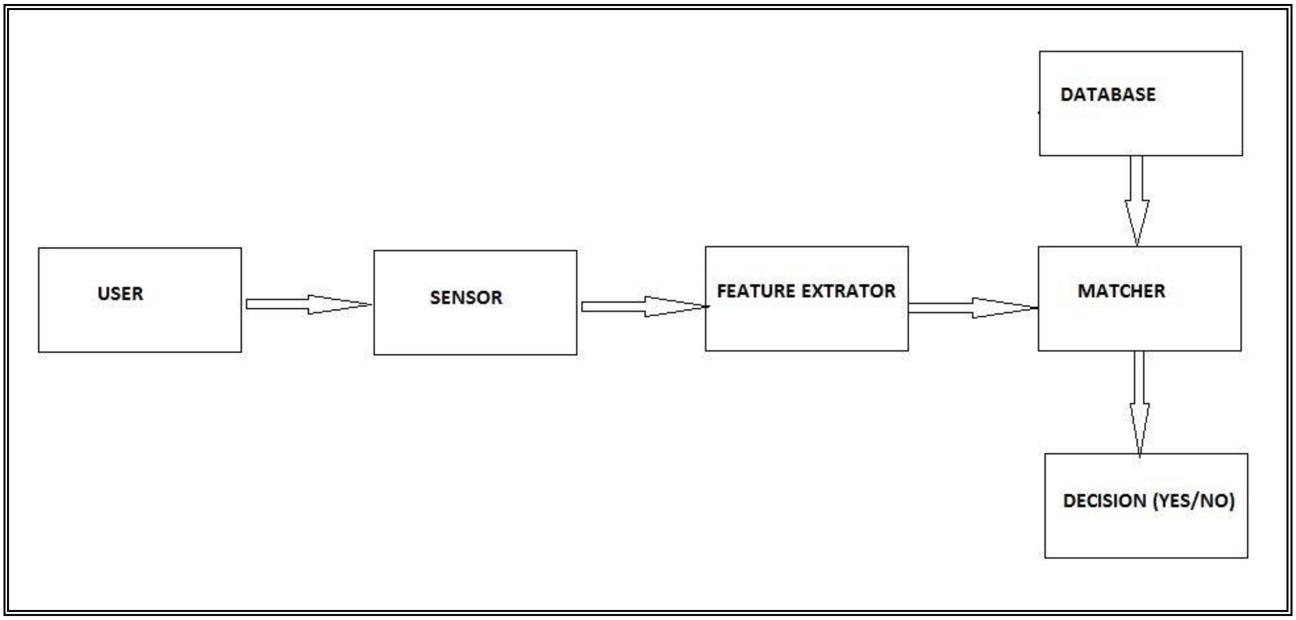


Figure 8 Verification process of Biometrics authentication method

3.3 BASIC STRUCTURE OF A BIOMETRIC SYSTEM

Every biometric system consists of four basic modules. These module are given below.

3.3.1 ENROLLMENT MODULE

The enrollment module registers individuals into the biometric system database. During this phase, a biometric reader scans the individual's biometric characteristic to produce its digital representation [9].

3.3.2 FEATURE EXTRACTION UNIT

This module processes the input sample to generate a compact representation called the template, which is then stored in a central database or a smartcard issued to the individual.

3.3.3 MATCHING UNIT

This module compares the current input with the template. If the system performs identity verification, it compares the new characteristics to the user's master template and produces a score or match value (one to one matching). A system performing identification matches the new characteristics against the master templates of many users resulting in multiple match values (one to many matching).

3.3.4 DECISION MAKER

This module accepts or rejects the user based on a security threshold and matching score.

3.4 BIOMETRIC SYSTEM PERFORMANCE PARAMETERS

The performance evaluation of a biometric system depends on two types of errors – matching errors and acquisition errors. The matching errors consist of the following [16]:

3.4.1 FALSE ACCEPTANCE RATE (FAR)

False acceptance rate (FAR) is a measure of biometric accuracy. It represents the probability that a given biometric system will accept an incorrect input as a positive match. FAR is calculated by testing known biometric templates against a large database or simply we can say that mistaking biometric measurements from two different persons to be from the same person.

3.4.2 FALSE REJECTION RATE (FRR)

The false reject rate (FRR) is a measure of the probability that a biometric system will incorrectly reject an input as a negative match or simply we can say that mistaking biometric measurements from the same person to be from two different persons.

FAR and FRR rates have an inverse relationship with one another. In other words, the more selective the biometric system is (better FAR rate), the more likely that the system will also begin to occasionally reject the correct fingerprint [17].

The acquisition errors consist of the following:

3.4.3 FAILURE TO CAPTURE RATE (FTC)

Proportion of attempts for which a biometric system is not able to capture a sample of desired quality.

3.4.4 FAILURE TO ENROLL RATE (FTE)

Proportion of the user population for which the biometric system is not able to generate reference templates of sufficient quality. This includes those who, for physical or behavioral reasons, are unable to present the required biometric feature. All of the above are used to calculate the accuracy and performance of biometric system. Biometric systems like any authentication system are not completely foolproof. It has its own drawbacks. While a biometric is a unique identifier, it is not a secret and biometrics, once lost is lost forever (Lack of secrecy and non-replace ability) .

Chapter 4

Various classifications and features of Biometrics

4.1 CLASSIFICATION OF BIOMETRICS.

As shown in Figure 9 Biometrics has two categories-:

a) Physiological Biometrics

b) Behavioral Biometrics

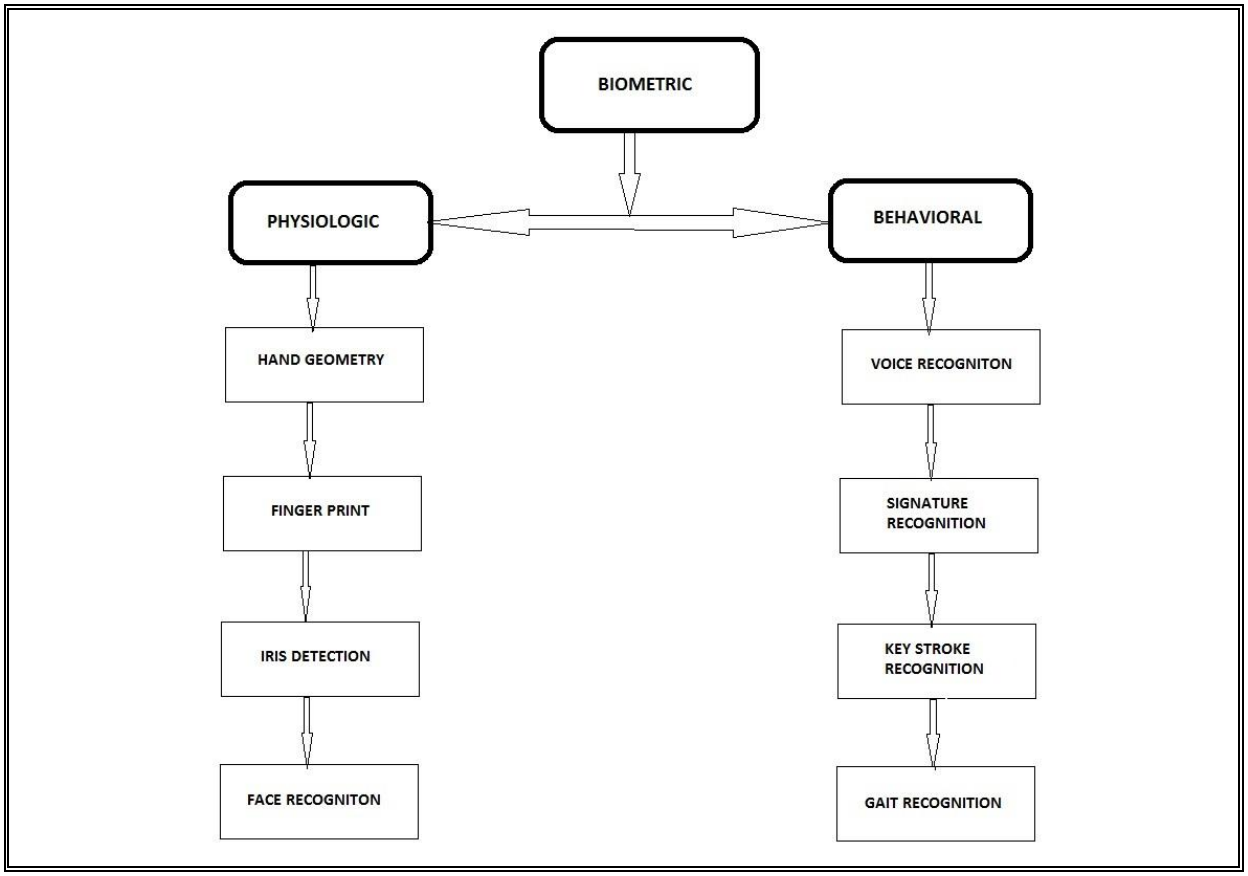


Figure 9 Classification of Biometrics.

4.2 PHYSIOLOGICAL BIOMETRICS

Physiological biometrics are based on a person's physical characteristics and depends upon the physical appearance of the human body or Shape which are assumed to be relatively unchanging such as fingerprints, iris patterns, retina patterns, facial features, palm prints, or hand geometry, nose, chin, eyes, face and lips, odor or DNA deoxyribonucleic acid etc. are the examples of physiological biometrics as shown in Figure 10 [18].

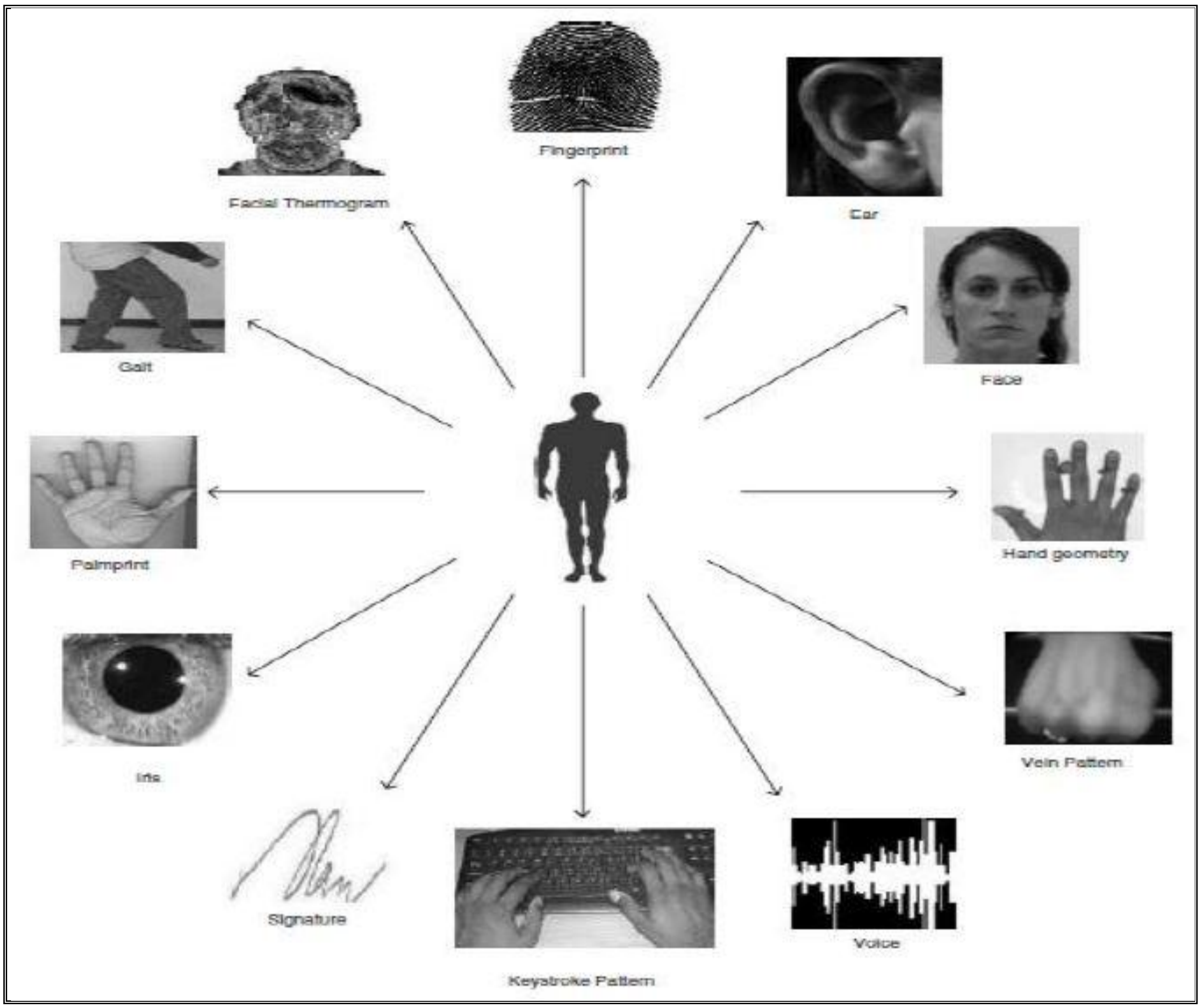


Figure 10 Biometric traits

4.2.1 Finger print recognition

Humans have utilized fingerprints for personal identification for many centuries and the matching accuracy using fingerprints has been shown to be very eminent and distinguished [19]. A fingerprint is an impression of the friction ridges of all or any part of the finger. A friction ridge is a raised portion of the on the palmer (palm) or digits (fingers and toes) or plantar (sole) skin, consisting of one or more connected ridge units of friction ridge skin. These ridges are sometimes known as "dermal ridges" or "dermal ". The traditional method exploited the ink to get the finger print onto a piece of paper. This piece of paper is then scanned using a traditional scanner. Fingerprints of identical twins are completely different and so are the prints on each finger of the same person. Now in modern approach, live finger print readers are being used. These are based on optical, thermal, silicon or ultrasonic principles [20]. It is the oldest and fundamental of all the biometric techniques. Optical finger print reader is the most common at present. They are based on reflection changes at the spots where finger papilar lines touch the reader surface .All the optical fingerprint readers comprise of the source of light, the light sensor and a special reflection surface that changes the reflection according to the pressure. Some of the readers are fitted out with the processing and memory chips as well . The finger print obtained from an Optical Fingerprint Reader is shown in Figure 11.

Figure 11 Bitmap of Optical Fingerprint

The size of optical finger is around 10.1x9.8x14.8. It is tough to minimize them much more as the reader has to comprise the source on light reflection surface and light sensor. Optical Silicon

Fingerprint Sensor is based on the capacitance of finger [21]. DC-capacitive finger print sensor consists of rectangular arrays of capacitors on a silicon chip. One plate of the capacitors is finger, other plate contains a tiny area of metallization on the chips surface son placing finger against the surfaces of a chip, the ridges of finger print are close to the nearby pixels and have high capacitance to them [22]. The valleys are more distant from the pixels nearest them and therefore have lower capacitance. Ultrasound finger print is newest and least common. They use ultrasound to monitor the figure surfaces, the user places the finger on a piece of glass and the ultrasonic sensor moves and reads whole finger print. This process takes 1 or 2 seconds [23].
Finger print matching techniques can be placed into two categories. One of them is Minutiae based and the other one is Correlation based. Minutiae based techniques find the minutiae points first and then map their relation placement on the finger. Correlation based techniques require the precise location of a registration point and are affected by image translation and rotation [24, 25].

4.2.2 Face Recognition

Face recognition is a nonintrusive method, and facial images are probably the most common biometric characteristic used by humans to make a peculiar and unique recognition. It is the most natural means of biometric identification [26]. A facial recognition technique is an application of computer for automatically identifying or verifying a person from a digital image or a video frame from a video source. The face is comprised of many distinct micro and macro elements. The macro elements include the mouth, nose, eyes, cheek bones, chin, lips, forehead, and ears. The micro features include the distances between the macro features, or a reference feature and the size of the feature itself [27]. The most prevailing approaches to face recognition are based on either the position or shape of facial attributes such as the eyes, eyebrows, nose, lips and chin, and their spatial relationships, or the global analysis of the face image that represents a face as a weighted combination of a number of canonical faces. Face recognition plays a crucial role in many applications such as security systems, credit card verification and criminal identification. Also, unseen to the human eye is the fact that our bodies and faces radiate heat, which can be measured by using infrared cameras [28]. All these features can be used by face biometric systems to help identify and authenticate someone. Although an ability to infer intelligence or character from facial appearance is suspect, the human ability to recognize faces is remarkable. The first step of human face identification is to extract the relevant features from facial images. Research in the field primarily intends to generate sufficiently reasonable familiarities of human faces so that another human can correctly identify the face [29]. The question naturally arises as to how well facial features can be quantized. If such a quantization is possible, then a computer should be capable of recognizing a face, given a set of features. Facial metric technology relies on the manufacture of the specific facial features (the system usually look for the positioning of eyes, nose and mouth and distances between these features), shown in Figure 12 and Figure 13.

Figure 12 Recognition procedure

The face region is rescaled to a fixed pre-defined size (e.g. 160-110 points). This normalized face image is known the canonical image. Then the facial metrics are computed and stored in a face template. The typical size of such a template is between 4 and 6 KB, but there exist systems with the size of the template as small as 96 bytes. The figure for the normalized face is given in as Figure 13.

Figure 13 Normalized Face

The Eigen Face method shown in Figure 14 on the next page is based on categorizing faces according to the degree of it with a fixed set of 100 to 150 Eigen faces [30]. The Eigen faces that are created will seem to be as light and dark areas that are arranged in a specific pattern. This pattern shows how different features of a face are singled out. It has to be evaluated and scored. There will be a pattern to evaluate symmetry, if there is any style of facial hair, where the hairline is, or evaluate the size of the nose or mouth [31]. Other Eigen faces have patterns that are less simple to identify, and the image of the Eigen face may look very little like a face. This technique is in fact similar to the police method of creating a portrait, but the image processing is automated and based on a real picture [32]. Every face is assigned a degree of fit to each of 150 Eigen faces, only the 40 template Eigen faces with the highest degree of fit are necessary to reconstruct the face with the accuracy of 99 percent. The whole thing is done using Face Recognition software [33, 34].

Figure 14 Normalized Face

4.2.3 Hand Geometry recognition

It is based on the fact that nearly every person's hand is shaped diversely and that the shape of a person's hand does not modify after certain age. These techniques include the approximation of length, width, thickness and surface area of the hand. Basically Hand geometry consists of two sides; one is Palm side and Dorsum side or top side or back side of the hand [35] as shown in Figure 15. Hand Geometry recognition is the method of personal authentication and available for over twenty years. To achieve personal authentication, a system may measure either physical characteristics of the fingers or the hands. Hand geometry has gained acceptance in a wide range of applications. It can frequently be found in physical access control in commercial and residential applications, in time and attendance systems and in general personal authentication applications. Hand geometry recognition systems may provide three kinds of services, Verification, Classification and Identification. For verification, the user provides his/her identity along with the hand geometry and the system verifies his/her identity. For classification, the user does not provide any identity information but is known to be legitimate. For identification, the user does not provide any identity information other than the hand geometry and may be an intruder. Without hand geometry we can’t identify a person. The system tries to identify the individual or deny access [36]. Various method are used to measure the hands-Mechanical or optical principle [37].

Figure 15 Hand geometry scanner

There are two sub-classifications of optical scanners. Devices from first category create a black and white bitmap image of the hand’s shape. This is easily done using a source of light and a black and white camera. The bitmap image is processed by the computer software. Only 2D characteristics of hand can be utilized in this case. Hand geometry systems from other category are more complicated. They use special guide marking to portion the hand better and have two (both vertical and horizontal) sensors for the hand shape measurements. So, sensors from this category handle data of all 3D features. Figure 15 and 16 shows the hand geometry system. Some of hand geometry scanners produce only the video signal with the hand shape. Image digitalization and processing is then done in the computer to process those signals in order to obtain required video or image of the hand [38].

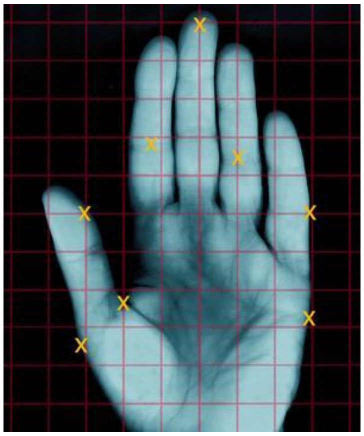


Figure 16 Acquired image of hand

4.2.4 Iris Detection

The human iris is an annular part between pupil (black portion) and cornea shown in Figure 17. Iris is an inner organ part of human body. The structure of human Iris contains five layers of fiber like tissues [39]. These tissues are very complex and reveal in various forms. The surface of iris also contains a complex structure such as crystals, thin threads, spot con caves, radials, furrows, stripes etc. Iris is a place where our nerve systems are situated and it gives information about human body [40].This recognition method uses the iris of the eye which is colored area that surrounds the pupil. Iris patterns are unique and are obtained through video based image acquisition system. Each iris structure is featuring a complex pattern. This can be a combination of specific Characteristics known as corona, crypts, filaments, freckles, pits, furrows, striations and rings [41]. An IRIS image is shown in Figure 17.

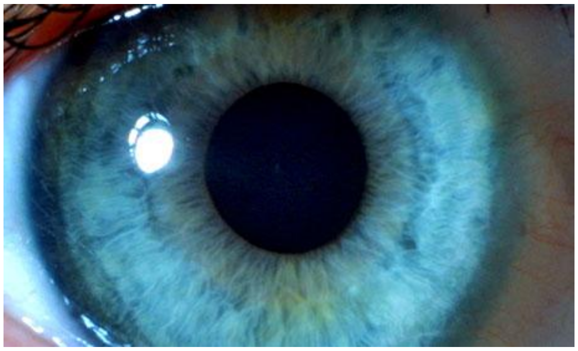


Figure 17 Image of an IRIS

The iris pattern is taken by a special gray scale camera in the distance of 10- 40 cm of camera. Once the gray scale image of the eye is obtained then the software tries to locate the iris within the image. If an iris is found then the software creates a net of curves covering the iris. Based on the darkness of the points along the lines the software creates the iris code. Here, two influences have to take into account [42]. First, the overall darkness of image is influenced by the lighting condition so the darkness threshold used to decide whether a given point is dark or bright cannot be static, it must be dynamically computed according to the overall picture darkness. Secondly, the size of the iris changes as the size of the pupil changes. Before computing the iris code, a proper transformation must be done. In decision process, the matching software takes two iris codes and compute the hamming distance based on the number of different bits. The hamming distances score (within the range 0 means the same iris codes), which is then compared with the security threshold to make the final decision. Computing the hamming distance of two iris codes is very fast (it is the fact only counting the number of bits in the exclusive OR of two iris codes). We can also implement the concept of template matching in this technique. In template matching, some statistical calculation is done between a stored iris template and a produced. Depending on the result decision is taken. Iris-based recognition is the most widely used biometrics among various biometric techniques (face, fingerprint, palm vein, signature, palm print, iris, etc.) because of its unique, stable, and noninvasive characteristics. It is the process of recognizing an individual by analyzing the random pattern of the iris and comparing it with that of reference in the database (DB) [43].

4.3 Behavioral Biometrics

Behavioral biometrics is related to behavior, nature and expression of human or person like heat of the body and they provide a number of advantages over traditional biometric technologies. They can be collected non-obtrusively or even without the knowledge of the user. Collection of behavioral data often does not require any special hardware and is so very cost effective. While most behavioral biometrics are not unique enough to provide reliable human identification, they have been shown to provide sufficiently high accuracy identity verification Many biometrics techniques like keystroke, signature, voice recognition, ear pattern, gait recognition and palm print recognition come under behavioral biometrics in the biometric brochure as shown in the

Figure 2. Even eye blinking, driving style and lip movement also fall under this category these characteristics are referred to as traits, indicators, modalities [44].

4.3.1 Keystroke Recognition

Keystroke dynamics utilizes the rhythm and manner in which an individual types characters on a keyboard which can cope with trained typists as well as the amateur two-finger typist.. The original keystroke data contain the pressing key time and releasing key time shown in Figure 18 and 19, from which two kinds of features are extracted, flight time and dwelling time. The flight time is defined as the time difference between one key release and the following key press. The dwelling time is the time difference between the press and release of one key. These systems should be cheap to install as all that is needed is a software package [45] .The technology, which measures the time for which keys are held down, as well as the length between strokes, takes advantage of the fact that most computer users evolve a method of typing which is both consistent and idiosyncratic – especially for words used frequently such as a user name and password [46]. When registering, the user types his or her details nine times so that the software can generate a profile. Future login attempts are measured against the profile which, the current claim is that it can recognize the same user's keystrokes with 99 per cent accuracy, using what is known as a behavioral biometric [47].

Figure 18 Keystroke recognition

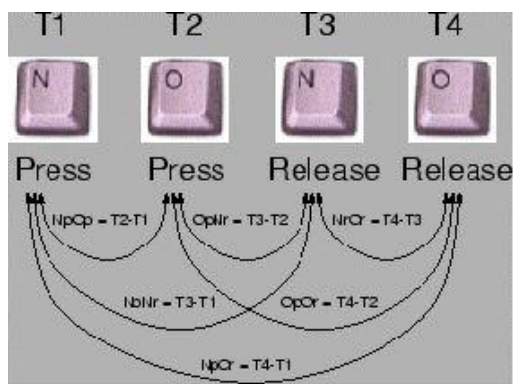


Figure 19 Key recognition mechanism

4.3.2 Voice recognition

Voice or speech recognition is an ability of a machine or program to receive and interpret dictation, or to understand and carry out spoken commands. Strictly speaking, voice is also a physiological trait because every person has a different pitch, but voice recognition is mainly based on the study of the way a person speaks, commonly classified as behavioral as shown in Figure 20.

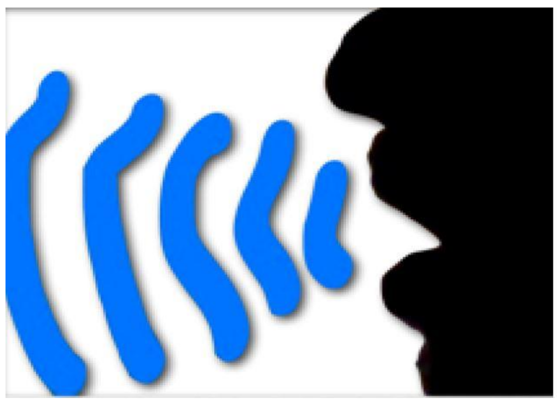


Figure 20 Speech/Voice Recognition

Speech recognition, which can be classified into identification and verification, is the process of automatically recognizing who is speaking on the basis of individual information included in speech waves [48]. This technique makes it possible to use the speaker's voice to verify their identity and control access to services such as voice dialing, banking by telephone, telephone shopping, database access services, information services, voice mail, security control for confidential information areas, and remote access to computers. On the other hand, speaker verification is the process of accepting or rejecting the identity claim of a speaker. Most applications in which a voice is used as the key to confirm the identity of a speaker are classified as speaker verification [49]. Speaker recognition methods can also be divided into text-dependent and text-independent methods. The former requires the speaker to say key words or sentences having the same text for both training and recognition trials, whereas the latter do not rely on a specific text being spoken [50]. Voice is also physiological trait because every person has different pitch, but voice recognition is mainly based on the study of the way a person speaks, commonly classified as behavioral Speaker verification focuses on the vocal characteristics that produce speech and not on the sound or the pronunciation of speech itself. The vocal characteristics depend on the dimensions of the vocal tract, mouth, nasal cavities and the other speech processing mechanism of the human body [50]. It doesn't require any special and expensive hardware. Speaker recognition uses the acoustic features of speech that have been found to differ between individuals. These acoustic patterns reflect both anatomy (e.g. size and shape of the throat and mouth) and learned behavioral patterns.(e.g. voice pitch, speaking style) [51].

Speaker recognition system employs three styles of spoken input and they are listed below.

(a) Text dependent
(b) Text prompted
(c) Text independent

Text dependent involves selection and enrollment of one or more voice passwords. Text prompted is used whenever there is concern of imposters. Various technologies used to process and store voice prints include hidden Markov models, pattern matching algorithms, neural networks, metric representation and decision tree. Some technology also uses "anti maker" techniques, such as cohort models, and world models [52].Voice changes due to aging also need to be addressed by recognition Systems. Capture of the biometric is seen as non-invasive. The technology needs additional hardware by using existing microphones and voice transmission technology allowing recognition over long distances via ordinary telephones [53, 54].

4.3.3 Signature recognition

Biometric signature recognition systems as shown in Figure 21 will measure and analyze the physical activity of signing, such as the stroke order, the pressure applied and the speed as it is based on the dynamics of making the signature, rather than a direct comparison of the signature itself afterwards. Some systems may also compare visual images of signatures, but the core of a signature biometric system is behavioral, i.e. how it is signed rather than visual, i.e. the image of the signature. Handwritten signature is one of the most widely used personal methods for identity verification. As a symbol of consent and authorization, especially in the prevalence of credit cards and bank cheques, handwritten signature has long been the target of fraudulence. The most obvious and important advantage of this is that a fraudster cannot glean any information on how to write the signature by simply looking at one that has been previously written. There are various kinds of devices used to capture the signature dynamics. These are either traditional tablets or special purpose devices. Tablets capture 2D coordinates and the pressure [55], as shown in Figure 22. Therefore, with the growing demand for processing of individual identification faster and more accurately, the design of an automatic signature verification system faces a real challenge [56]. Special pens are able to capture movements in all three dimensions. Tablets have two significant disadvantages First, the resulting digitalized signature looks different from the usual user signature. Secondly, while signing the user does not see what he or she has already written. He/she has to look at the computer monitor to see the signature. This is a considerable drawback for many (inexperienced) users. Some special pens work like normal pens, they have ink cartridge inside and can be used to write with them on paper [57, 58].

Figure 21 Signature recognition

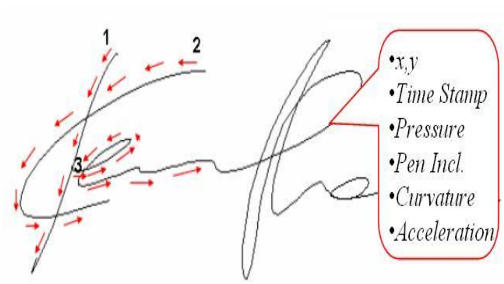


Figure 22 Parameters of signature

4.4 Applications of Biometrics

Biometric authentication is highly reliable, because physical human characteristics are much more difficult to forge then security codes ,passwords ,hardware keys sensors ,fast processing equipment and substantial memory capacity , so the system are costly. Biometric based authentication applications include workstation and network access, single sign-on, application logon, data protection, remote access to resources, transaction security, and Web security. The promises of e-commerce and e-government can be achieved through the utilization of strong personal authentication procedures. Secure electronic banking, investing and other financial transactions, retail sales, law enforcement, and health and social services are already benefiting from these technologies. Biometric technologies are expected to play a key role in personal authentication for large-scale enterprise network authentication environments, Point-of-Sale and for the protection of all types of digital content such as in Digital Rights Management and Health Care applications. Utilized alone or integrated with other technologies such as smart cards, encryption keys and digital signatures, biometrics is anticipated to pervade nearly all aspects of the economy and our daily lives. For example, biometrics is used in various schools such as in lunch programs in Pennsylvania, and a school library in Minnesota. Examples of other current applications include verification of annual pass holders in an amusement park, speaker verification for television home shopping, Internet banking, and users' authentication in a variety of social services [59].

4.5 Advantages of Biometrics

1. Increased securities provide a convenient and low cost additional tier of security.
2. Reduces fraud by employing hard-to-forge technologies and materials. For e.g. minimize the opportunity for ID fraud, buddy punching.
3. Reduces password administration costs, Eliminate problems caused by lost IDs or forgotten passwords by using physiological attributes. For e.g. prevent un-authorized use of lost and found, stolen or "borrowed" ID cards.
4. Integrates wide range of biometrics solution and technologies, customer applications and database into a robust and scalable control solution for facility and network access.
5. It is possible to establish an identity based on who you are, rather than, by what you possess, such as an ID card or what your remember, such as password.
6. It plays a very significant role in area of time and attendance and also helps in saving the cost and time.

Chapter 5

Types of Signatures

This thesis puts to the classifying methods for hand-written signature verification. Signature verification is a requisite task in society, as signatures have a well naturalized and consented place as a formal means of personal verification. Due to this, they are used in government, legal and commercial transactions and they are the most consented method of identity verification. An predictable side-effect of signatures is that they can be tapped for the purpose of simulating a document's authenticity. Because of this, the identification of forgeries is one of the oldest forms of forensic science, with references being made to it in Roman Law as far back as the 3rd Century AD, and forgery being a statutory offence in Britain in the 13th Century. Even today, signature forgeries are still a usual problem, with the total global cost to businesses, financial institutions and individuals being untold. Signature verification comes as a form of behavioral biometric rather than an anatomical Biometric such as a fingerprint or an iris pattern. Although biometrics in the area of signature matching is not as widely explored as other forms of biometrics, it should not be surprising that signature authentication is as important, if not more, as fingerprint and other forms of biometrics. Signature classification and verification reduces the risk of a forged signature being accepted as a unfeigned. The problem that arises though is that validation of the signature often needs to be carried out instantaneously, so that completion of the intended process can be permitted. This commonly occurs when dealing with monetary transactions, especially via credit cards and bank. The types of automatic signature classification and verification considered in this thesis are suitable for assuaging this problem, as a human examiner need only be consulted when the authenticity of a signature is disputed. . In any event, our signature and the way it is signed can be considered a unique enough identity verification methodology for a variety of personal transactions. With technological advancements, rapid increments in computing power have taken place. This has led to the ability of computer machines performing complex and computationally intensive algorithms at a faster rate [59].

5.1 Signatures as a Biometric

Signatures are a behavioral biometric, which like writing, walking and talking, is a attribute or trait that is learnt over time. Behavioral biometrics differs from physiological Biometrics as they measure a person's physical features, such as the face, fingerprint or retina, and tend to remain relatively stable over time. The behavioral characteristics of a person gravitate to change over both the short and long terms due to health, physical state and ageing, which makes them more unmanageable to distinguish than the physical characteristics. This is especially true for signatures, which can also vary depending on the signatory's mental state, fatigue, and their writing position. As a result of this high unevenness in the signing process, the verification of a signature is not a niggling task, for both human experts, and especially computers, which is the focus of this thesis [60].

5.2 Verification Methods

Currently, organizations that rely on signatures do not actively check the authenticity of each signature. Instead, in place measures are relied upon that attempt to restrict the occurrence of forgeries, with any suspect signatures requiring visual confirmation in all instances. The introduction of automatic procedures have the potential of further reducing the possibility of forgeries slipping In signature verification, there are two overall methods that are employed for determining whether a signature is a genuine or a forgery. The first and most common method is manual verification, in which a human determines the validity of a signature. The second method involves the use of a computer for automatic verification, and is becoming more practicable as research advances [61, 62].

5.3 Human Forensic Document Scrutiny

Document probe is a forensic science that has a number of areas of expertise, which include the identification of handwriting and signatures, and the composition of inks, papers and materials from which documents are produced. A specialist in this field is known as a forensic document examiner, an occupation that requires an panoptic amount of training in order to become proficient. In general, the majority of forensic document examiners are employed by professional laboratories. When research scholars unite these laboratories, they will first work for up to two years alongside highly experienced document examiners before handling their own case. For appropriate trained examiner, the examination of a document is not carried out by simply looking at the handwriting, but also requires an array of equipment and techniques that are developed to extract as much information as possible from the document without damaging or altering it. The examiner, when experimenting on a signature, will also use a chart of elemental characteristics such as ticks, coordinates, smoothness of curves, pressure change of pen, spacing and slant to help colligation of the signature [63].

5.4 Automatic Verification

Automatic signature verification is a method that until now has not gained universal acceptance, but could be used as a surrogate for or a tool used by a forensic document examiner. This is because the viability of using automatic methods is more suited for activities such as boarding a train, entering a secure geographical location and performing financial transactions. For these activities, verification of the signature class is determined algorithmically by going through a sequence of steps. These steps transform the signature from a piece of writing on a document into a set of features that can be used to classify the signature. These features differ significantly from the ones used in the manual method. This is because a computer cannot visually analyze or comprehend the entirety of a signature the same way that a human can. A computerized algorithm treats with a signature at the pixel level, with features being developed from the pixels and the way these pixels pertain to each other. The methods used in automatic verification produce a range of results that can deviate based on a number of factors related to the choice of algorithms. Until now the unsurpassed classification precision achieved is approximately 90% – 92% on a general database of given number of signatures for different persons [64].

5.5 Learning Approach

In signature verification, there are generally two learning strategies that are used to determine a signature's class. These learning strategies may be depend on person or may not. The person-dependent approach involves building a classifier using genuine signatures from a single person. Essentially it measures the distance in feature space between all genuine signatures taken from this person. These distances then constitute the distribution of the genuine signatures in feature space. This distribution can then be used to calculate a probability or similarity by finding the distance between an unknown signature and each genuine. If this similarity is within the bounds that is set by the known distribution, then the unknown signature is considered to be a genuine, otherwise it is classified as a forgery. Person-independent is a more general approach to signature verification, and accounts for all varieties of verification that involves multiple classes. It differs from person-dependent in the fact that the classifier is built from two or more classes, where these classes can be either genuine or forgeries from as many people as desired. This approach at the basic level is a nearest neighbor approach that calculates a likelihood ratio of an unknown signature belonging to a particular class. If the likelihood is above a certain threshold, then the signature is deemed to be a genuine, otherwise it is considered to be a forgery.

The research that was carried out in this thesis primarily focused on the person dependent approach, though the effect that machine learning has on the person independent approach was investigated [65,66].

5.6 Signature Capture

Signature capture is the first step required in being able to determine a signature's authenticity, that is, whether it is genuine or a forgery. Capturing a signature is the process of obtaining information about the signature, which can include both spatial and dynamic information. This information is turned into a form that can be interpreted and processed by a computer. The spatial information is the visual aspects of a signature, such as its shape and size, while the dynamic information relates to the writing process of the signature and defines aspects such as the speed and pressure used. Once this has been accomplished, the defining features can then be extracted from the signature. This section looks at the two data acquisition methods used for capturing written signatures and the data that is extracted from them; these are the dynamic (on-line) and static (off-line) methods. The remaining chapters will be focusing on static signatures and the techniques that are related to off-line signature verification [67].

5.7 Types of signatures

A signature verification system and the techniques using to solve this problem can be divided into two classes: online and off-line. In an online system, a signature data can be obtained from an electronic tablet and in this case, dynamic information about writing activity such as speed of writing, pressure applied, and number of strokes is available. In off-line systems, signatures written on paper as has been done traditionally are converted to electronic form with the help of a camera or a scanner and obviously, the dynamic information is not available. In general, the dynamic information represents the main writing style of a person. Since the volume of information available is less, the signature verification using off-line techniques is relatively more difficult. Our work is concerned with the techniques of off-line signature verification. The static information derived in an off-line signature verification system may be global, structural, geometric or statistical [68].

5.7.1 Dynamic Signatures

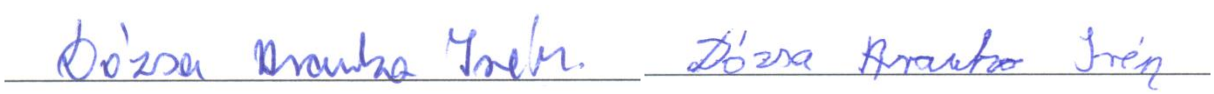
Dynamic signatures capture both the spatial and the dynamic information of a signature. The dynamic features that are captured include one or more of the following attributes as the signature is written: acceleration and velocity; the position of the pen; the pressure that is applied; and the pen inclination. Dynamic information is not limited to these factors as it can include any information that is relevant to the writing process. The capture of a signature's dynamic information requires the use of specialized hardware such as a digitizing tablet or an electronic pen that captures the written attributes and converts them into process able data. The advantage of using dynamic signatures for verification is that they produce a greater number of defining features than their static counterparts, making forged signatures easier to detect. The added advantage of digitally capturing this data is that there is typically less pre-processing required than static signatures, making feature extraction easier. Using this method to capture the writing process also has its associated disadvantages. The main disadvantage is that these tools are not very common and as such, cannot be utilized for every situation where signature verification is required. As well as this, their use is unnatural for the user, which could negatively affect the writing of the signature. More recent approaches exploit the use of a video camera which is focused on the writing of the signature and can be carried out using ordinary pen and paper. This allows the handwriting to be recovered from its spatio-temporal representation, which is given by the sequence of images that are produced [69].

5.7.2 Static Signatures

A static signature is the visual representation of a signature that has been written out in its entirety. Because of this, a static signature is only captured once the writing process has been completed, allowing for visual aspects such as the size, slope and curvature to be extracted. It is these visual aspects and their derivatives that allow similar signatures to be grouped together and differentiated from others that do not belong in the same group. A signature is generally written on paper using a standard pen, meaning that the signature is not initially in a digital format. Because of this, the signature has to be captured and transformed into a format that can be processed by a computer. The conversion of a signature to a digital format is carried out either by taking a photograph or scanning the signature. Both of these methods produce a digital image from which defining features of the signature can then be extracted. In society, a signed hardcopy of a document is the general requirement for binding that document to the signatory. For this purpose, a digitally captured signature is not always sufficient. As well as this, static signatures are much more prevalent than digitizing tablets and electronic pens, as they only require a standard pen and paper. The use of static signatures does pose some disadvantages for automatic verification though. One disadvantage is the fact that once a signature has been captured, it will usually require pre-processing in order to remove unwanted artifacts, and normalize the size and binarisation of the signature image. Another disadvantage is that static signatures do not produce as much data as dynamic signatures, causing the verification of the signature's validity to be based solely on the extracted features [70].

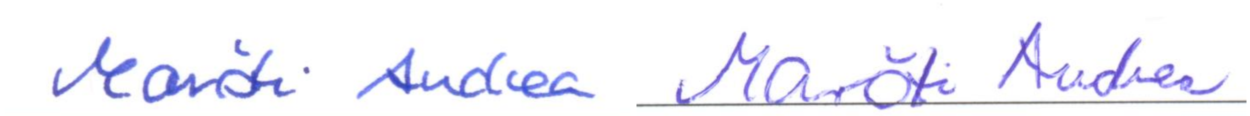
5.8 Forgery Types

In signature verification, forged signatures can be broken up into three different categories. These categories are based on how similar a forgery is in relation to the genuine signature and are known as random, simple and skilled. A random forgery is one in which the forger does not know the signer's name or signature shape [71]. A simple forgery is produced knowing the name of the original signer but not what their signature looks like. A skilled forgery is a close imitation of the genuine signature and is produced by a forger who has seen and practiced writing the genuine signature [72]. It is these skilled forgeries that this thesis will focus on for signature verification. There are three different types of forgeries to take into account. First one is **random forgery** which is written by the person who doesn't know the shape of original signature. The second, called **simple forgery**, which is represented by a signature sample, written by the person who knows the shape of original signature without much practice. The last type is **skilled forgery**, represented by a suitable imitation of the genuine signature model [73]. Each type of forgery requires different types of verification approach. Hybrid systems have also been developed. Figure 23 shows the different types of forgeries and how much they vary from original signature [74].



a) Original Signature b) Forged signature

a) Original Signature b) Forged signature



a) Original Signature b) Forged signature

Figure 23 shows the different types of forgeries

Chapter 6

Material and Methodology

6.1 DATABASE

The data base used in the work is obtained from the Signature verification competition held in 2004. This data base set contains signature data collected from 40 users. For each user, there are 20 genuine signatures and 20 skilled forgeries. Each genuine and forgery signature is stored in a separate text file. The file names are in the format "USERx_y.txt", where x (1-40) indicates the user and y (1-40) indicates one signature instance of the corresponding user, with the first 20 (1-20) representing genuine signatures and the rest (21-40) representing skilled forgeries provided by the other users. In each text file, the signature is simply represented as a sequence of points. The first line stores a single integer which is the total number of points in the signature [75].

6.2 METHODOLOGY

The algorithm developed has been applied on the dataset as shown in Figure 24 is as follows:

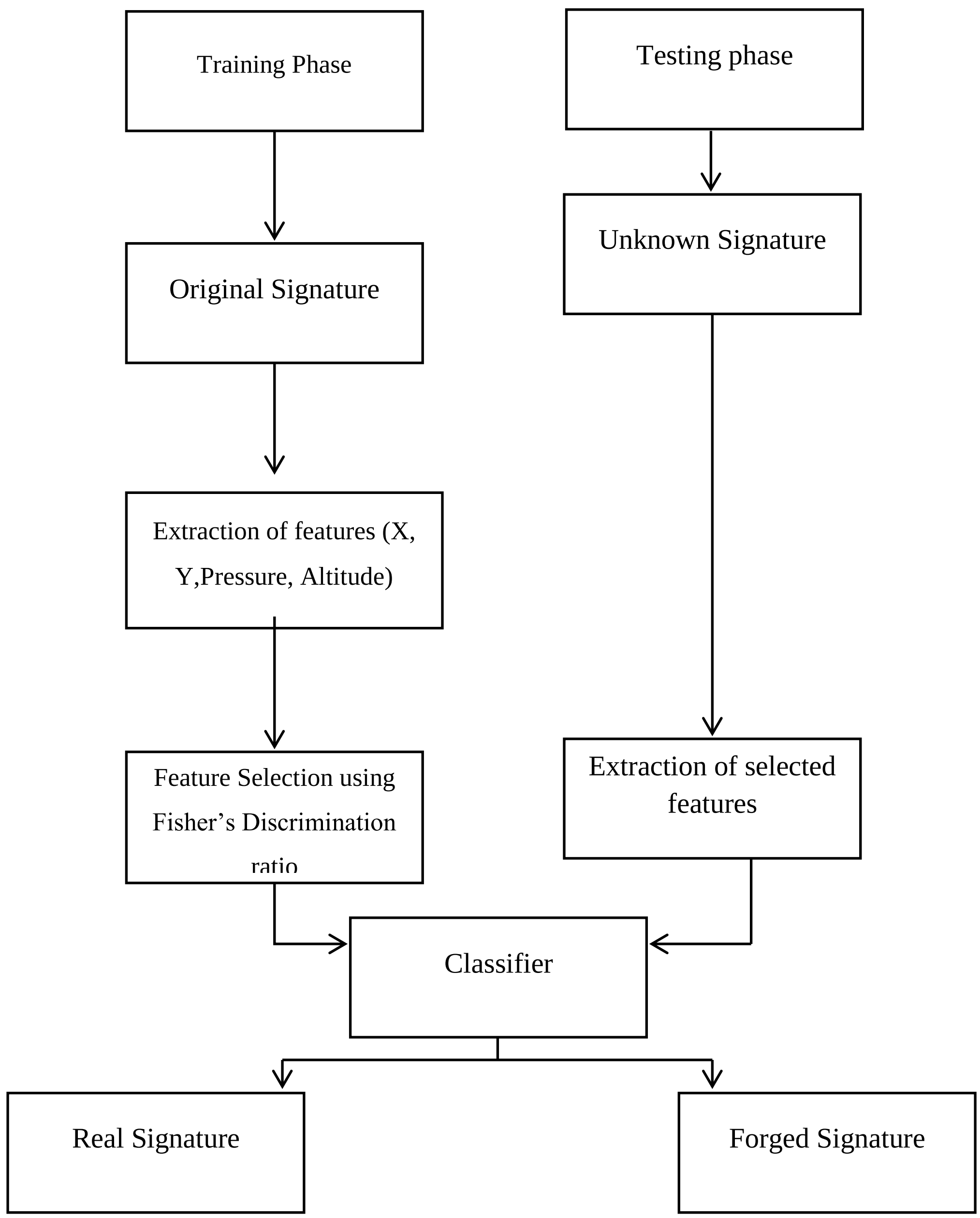


Figure 24 Flowchart showing the methodology to be followed for the classification of the Signatures.

6.2.1 Extracting Features from a Signature

In the first step of the feature extraction signature is done with use of writing tablet as shown in Figure 25

Figure 25 Signature obtained from writing pad for feature extraction

After obtaining the signature the features like X coordinate, Y coordinate , Azimuthal Angle of the pen, Altitude of the pen, Pressure of tip of pen on the writing pad, Pen Status (up/down) and Time stamp(time required for writing the signature) can be obtained by using Dynamic time warp algorithm.

The objective of the pre-processing phase is to get rid of noise from the input signature. Pre-processing contains six steps: In the first step Noise is removed by smoothing of data based on local regression using weighted linear least squares and a 2nd degree polynomial model is used to eliminate the potential sensor noise. Each smoothed data value is determined by neighboring data points defined within the predefined span based on a regression weight function. Figure 26 and 27 shows the signature before smoothing and after smoothing respectively. The regression weight for each data point in the span is given by the tri-cube function as shown in the Equation (1):

$$\omega_i = \left(1 - \left| \frac{x - x_i}{d(x)} \right|^3\right)^3 \qquad (1)$$

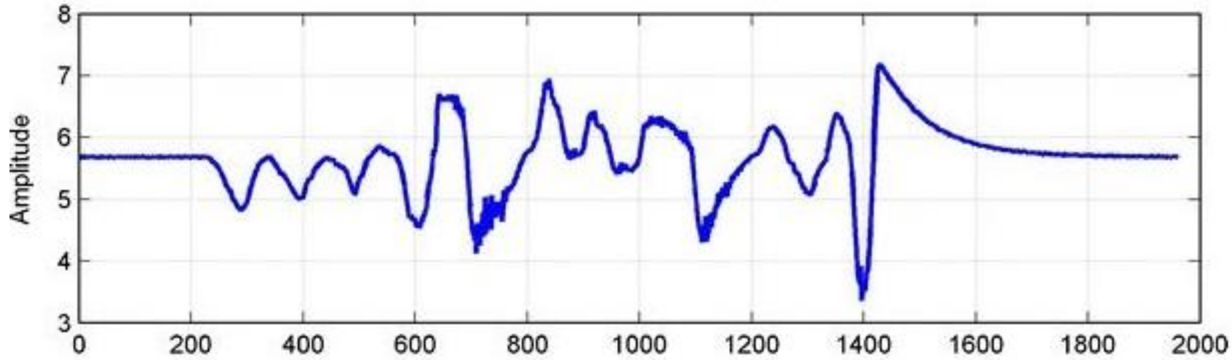


Figure 26 Signature before smoothing

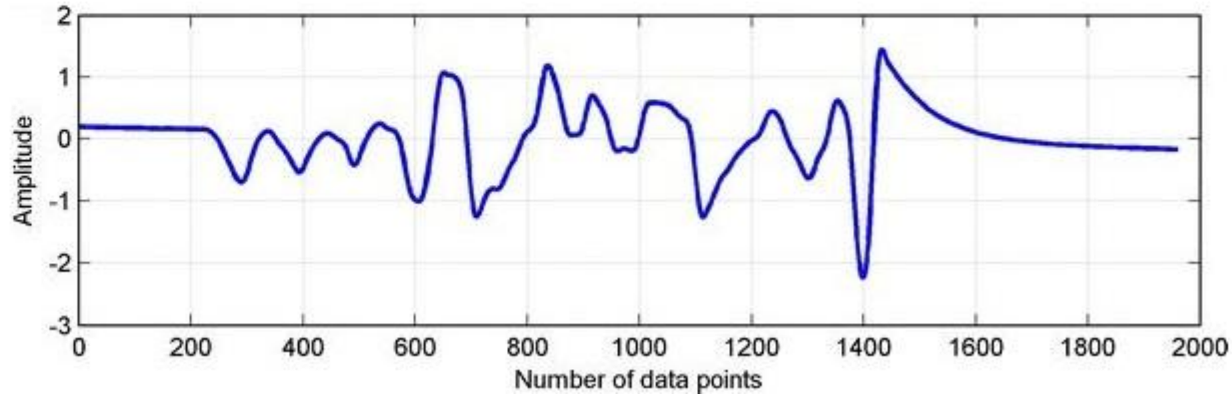


Figure 27 Signature after smoothing

In the second step Normalization of amplitude is done In order to partially compensate large personal variations of time series in the amplitude domain, normalization of the time series is done in such a way that data is normalized to [−1, 1], [0, 1] or z-score . The data normalization of amplitude values in [−1, 1], [0, 1] and z-score are denoted by $norm_1$, $norm_2$ and $norm_3$, respectively, is achieved due to the Equations. (2), (3) and (4) and Figure 28, 29 and 30 show the result of signature data after Normalization of the signature using $Norm_1$, $Norm_2$ and $Norm_3$ respectively.

$$\mathrm{Norm}_1:\ \mathrm{R(t)} = \frac{\mathrm{R(t)}}{\mathrm{Max}\ (|\mathrm{R(t)}|)} \qquad (2)$$

$$\mathrm{Norm}_2\text{: } R(t) = \frac{R(t) - R(t)_{min}}{R_{max} - R_{min}} \tag{3}$$

$$\mathrm{Norm}_3\text{: } R(t) = \frac{R(t) - \mathrm{mean}\ (R(t))}{\mathrm{RTD}\ (R(t))} \tag{4}$$

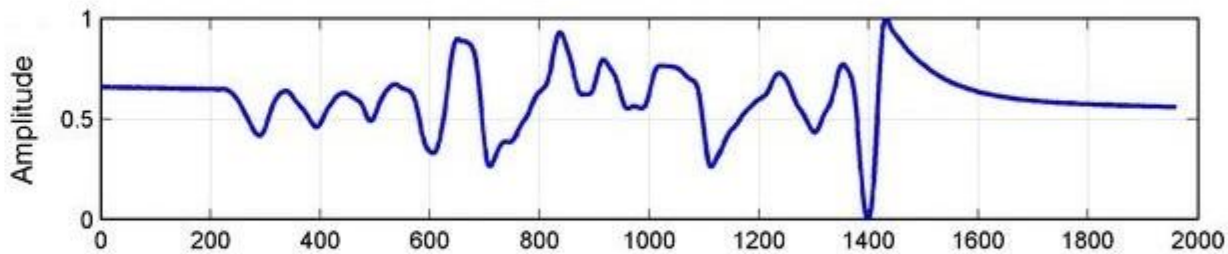


Figure 28 Signature data after Norm_1

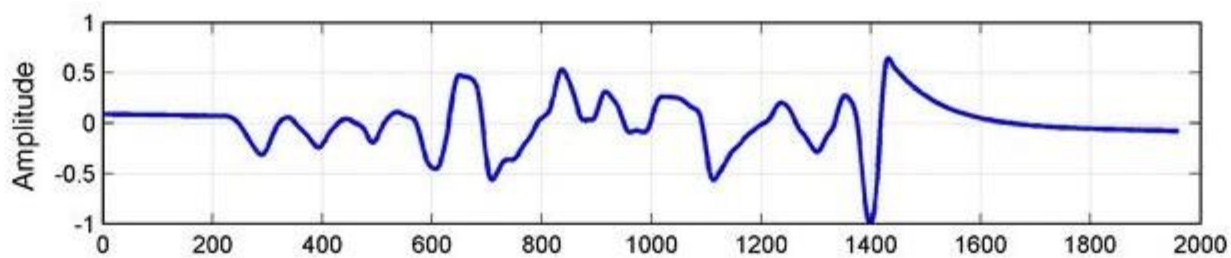


Figure 29 Signature data after Norm_2

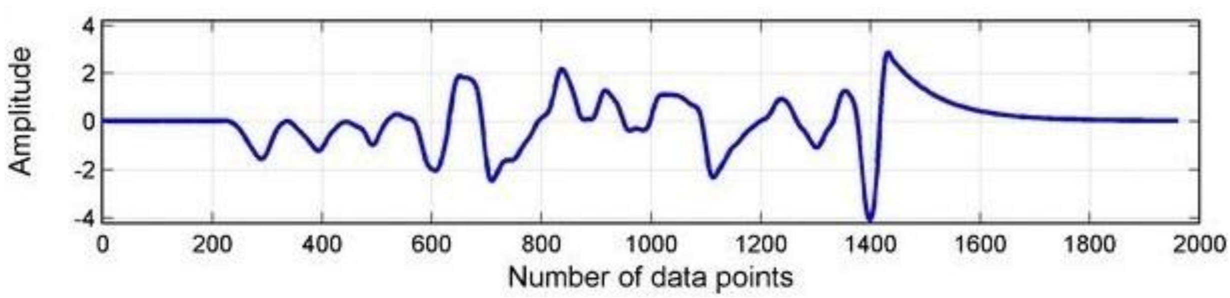


Figure 30 Signature data after Norm_3

In the next step in order to compensate the complexity of DTW, the low pass filtered time series are down-sampled to reduce the sampling rate or, the size of the database. Filtering is performed by an eight-order low-pass Chebyshev Type I filter. It filters time series effectively by doubling the filter order in both the forward and reverse directions in order to remove all phase distortions. Dimension normalization which is about the size of signature, so in the fourth step Signature size can be normalized according to one of the dimensions (width or height), which does not completely remove size characteristic of a writer. It is also known that people does not equally scale their signatures with respect to width and height, Re-Sampling which is eliminating the redundant points of signatures. Once two signatures compared with respect to their shape they must be re-sample to extract more consistent shape features, but since DTW algorithm the method which is going to use to implement this System, so that, there is a need to consider that DTW-based systems may resample the signature into an equal-distant point sequence before string matching. Smoothing that used the method which is going to use to implement this project, so that, there is a need to consider that DTW-based systems may resample the signature into an equal-distant point sequence before string matching. Figure 31 and 32 shows the two genuine signatures of unequal lengths and normalized to equal lengths respectively [76].

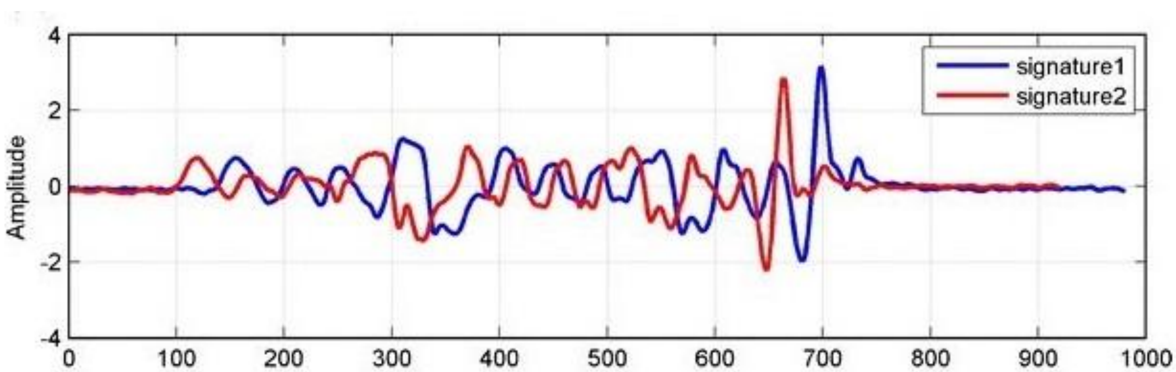


Figure 31 Two genuine signatures of unequal length

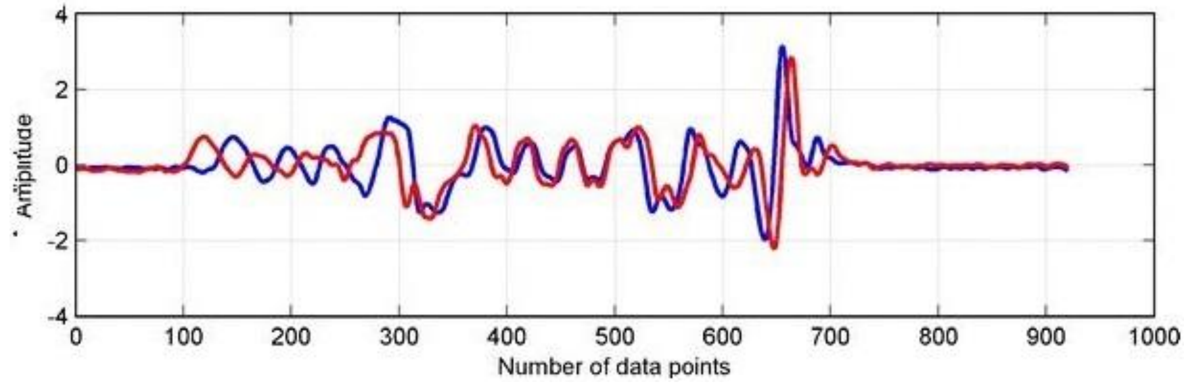


Figure 32 Two genuine signatures normalized to equal length

Next step is the Feature Extraction which is one of the most significant processes in signature verification. In online signature verification, representation of data is done by sequence of points. Thus, the features are extracting from a sequence of points. Feature extraction has been played an important role through the verification process. However, choosing the related features is still quite an unplanned activity that researchers still trying to find the best features set [77].Features used in this research include, X-coordinate , X-coordinate , Pressure, Azimuth angle, Pen status, Altitude , and time stamp. Each person has an original shape of signature, but it is big deal to find two same signatures of a signer. The differentiations of these two signatures are in either shape or dynamic features[78]. From comparing these features, a string will be assembling. These new strings which come from the extracted feature will compare together to reach the matching points. There are many different algorithm proposed for feature matching. One of the most effective algorithms is DTW which is a matching technique that is used to align two sequences of different length [79, 80]. The general algorithm for DTW is

$$D(m, n) = d(rm, tn) + \{min (D(m, n-1), D (m-1, n-1), D (m-1, n)\}$$

6.3 Dynamic time warping

(DTW) is an algorithm for measuring similarity between two sequences which may vary in time or speed. In general, DTW is a method that allows a computer to find an optimal match between two given sequences. The sequences are "warped" non-linearly in the time dimension to determine a measure of their similarity independent of certain non-linear variations in the time dimension.

6.3.1 How DTW works?

Dynamic time warping (DTW) is a time series alignment algorithm developed originally for speech recognition. It aims at aligning two sequences of feature vectors by warping the time axis iteratively until an optimal match (according to a suitable metrics) between the two sequences is found. Consider two sequences of feature vectors [81, 82]:

$A = a_1, a_2, a_3, a_4 \ldots\ldots a_i \ldots\ldots a_n$

$B = b_1, b_2, b_3, b_4, \ldots\ldots b_j, \ldots\ldots b_m$

The two sequences can be arranged on the sides of a grid, with one on the top and the other up the left hand side. Both sequences start on the bottom left of the grid as shown in Figure 33.

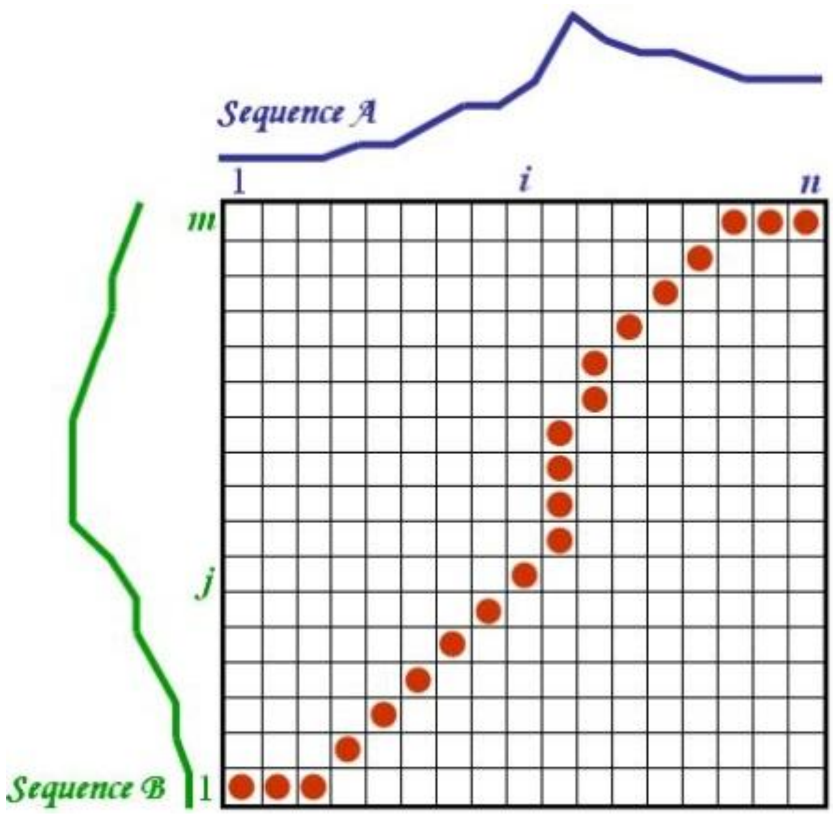


Figure 33 Algorithm of DTW

Inside each cell a distance measure can be placed, comparing the corresponding elements of the two sequences. To find the best match or alignment between these two sequences one need to find a path through the grid which minimizes the total distance between them. The procedure for computing this overall distance involves finding all possible routes through the grid and for each one computes the overall distance. The *overall distance* is the minimum of the sum of the distances between the individual elements on the path divided by the sum of the weighting function. The weighting function is used to normalize for the path length. It is apparent that for any considerably long sequences the number of possible paths through the grid will be very large [83].

6.4 FEATURES EXTRACTED FOR SIGNATURE

The following table shows the different features extracted from the signature.

Sr. No.	Features	Description
1)	X-coordinate	Value of pen cursor on the X-axis.
2)	Y-coordinate	Value of pen cursor on the Y-axis.
3)	Pen status	Value of pen whether it is touching the writing pad or not. If it is touching then the status is down otherwise up.
4)	Time stamp	Average time taken by pen in X and Y Axis.
5)	Azimuth	The horizontal angle or direction of pen.
6)	Altitude	Angle upward toward the positive z-axis.
7)	Pressure	The average pen tip pressure over the entire signature.
8)	H1	Hybrid feature obtained by squaring the Altitude value

9)	H2	Hybrid feature obtained by cubing the Altitude value
10)	H3	Hybrid feature obtained by squaring the Pressure value
11)	H4	Hybrid feature obtained by cubing the Pressure value

6.5 CLASSIFICATION

Classification is a task of training a classification model that maps each attribute set x to one of the predefined class labels y. The Figure 34 shows a general approach of a classification problem. First, a training set consisting of records whose class labels are known are provided. The training set is then used to train the classification model, which is eventually used to label the testing set [84].

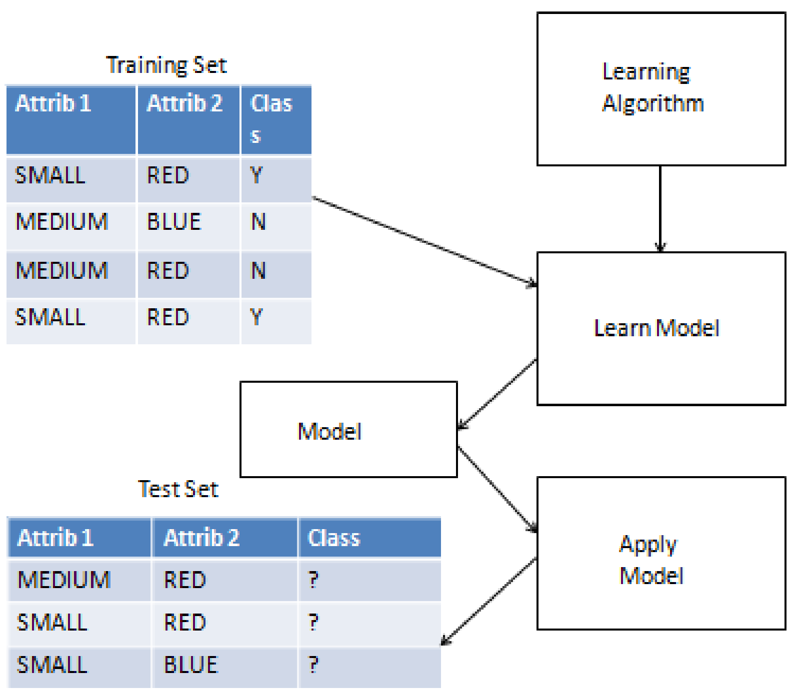


Figure 34 General classification approach.

There is no single classification approach which can be used for every problem. A classification model is to be chosen is problem specific and accuracy of the result differs from approach to approach. Three classification models are used in this work. They are

1) K-Nearest Neighbor (K-NN)

2) Random Forest

3) Bayesian

The following sections explain each of them briefly.

6.5.1 K-NEAREST NEIGHBOR (K-NN)

This is one of the oldest, simplest and most popular classification algorithms used in data mining and pattern recognition field. K-NN algorithm tries to find k-nearest points to the testing point and labels(maps) the testing point to a class which appeared maximum number of times in the circle enclosing k-training points. This algorithm is also called as Lazy learning method and it requires very less or no training. The main steps of the algorithms are

1) Define k-value

2) Train the set

3) Distance measurement

4) Labeling.

The Figures 35 and 36 shows a general approach to label (map, classify) a given data set (point) from a set of predefined labels.

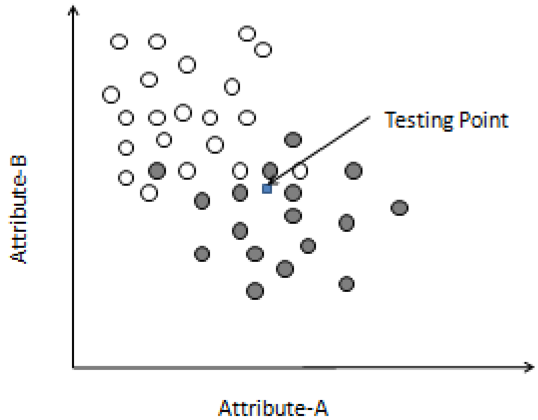


Figure 35 Scatter plot of two classes.

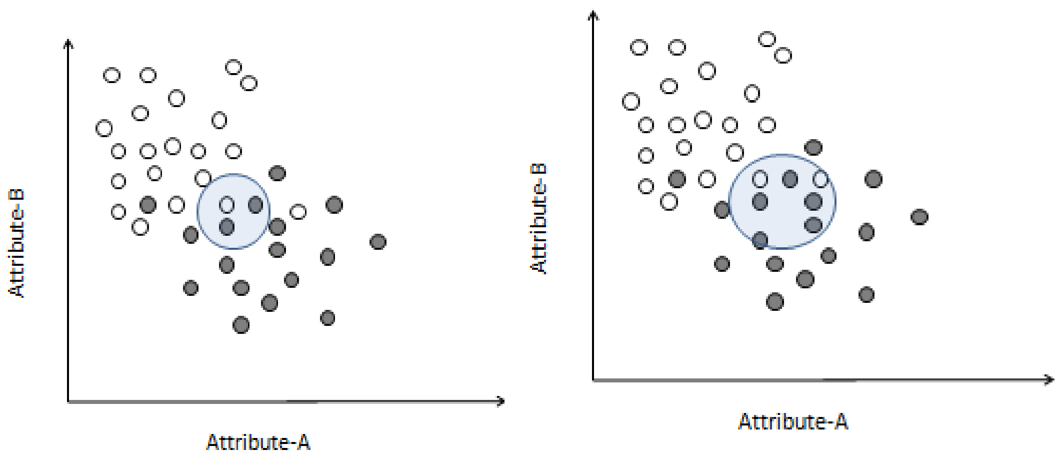


Figure 36 Decision making for different k-values

Figure 36 shows a scatter-plot of two-classes, in which unfilled-'dots ' belong to Class YES and filled-'dots' belong to class NO. The square 'dot' is a point which is to be labeled. Actually the testing dot here belongs to the class YES.

Let's see how k-NN algorithm labels the testing set in the following section.

Case1. k=3: Here the value of k is 3, i.e. we are considering 3 nearest points to classify the testing set. We find that out of 3-nearest points, two dots belong to the class NO and one-point belongs to the class YES. Majority is the class NO. Hence the model labels the testing set as class YES.

Case2. k=7: Here the value of k is 7, i.e. we are considering 7 nearest points to classify the testing set. We find that out of 7-nearest points, five dots belong to the class NO and two-point belongs to the class YES. Majority is the class NO. Hence again the model labels the testing set as class NO.

As we have seen, more the number of k-values more the training set it uses for training the model. In the first case the probability of the testing set belonging to class NO was 0.66. But in the second case the probability of labeling the test set to class NO was increased to 0.71, which is more accurate. Hence we can understand a very important fact that, more the value of k more the accurate is the model.

6.5.2 NAÏVE BAYES

The Naïve Bayes classifier technique is based on the Bayesian theorem and is particularly suited when the dimensionality of the inputs is high. Despite of its simplicity, Naïve Bayes can often outperform more sophisticated classification methods. To demonstrate the concept of Naïve Bayes classification, consider the example displayed in the figure below. The objects can be classified as either GREEN or RED. Our task is to classify new cases as they arrive, i.e., decide to which class label they belong, based on the currently existing objects [85].

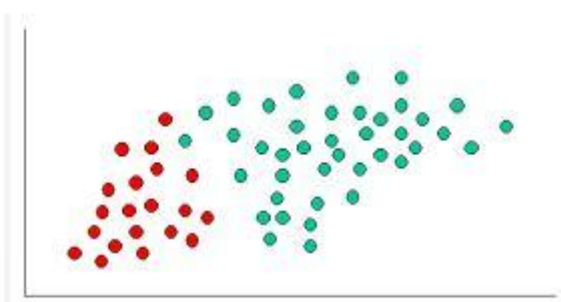


Figure 37 Naïve Bayes example illustration

Since there are twice as many GREEN objects as RED, it is reasonable to believe that a new case (which is unobserved yet) is twice as likely to have membership GREEN rather than RED. In the Bayesian analysis, this belief is known as the prior probability. Prior probabilities are based on previous experience, in this case the percentage of GREEN and RED objects, and often used to predict outcomes before they actually happen.

$$\text{Probability for the GREEN} = \frac{Number\ \ of\ Green\ \ objects}{Total\ \ number\ \ of\ objects}$$

$$\text{Probability for the RED} = \frac{Number\ \ of\ Red\ objects}{Total\ \ number\ \ of\ objects}$$

Since there are total of 60 objects, 40 of which are GREEN and 20 RED, our prior probabilities for class memberships are:

$$\text{Probability for the GREEN} = \frac{40}{60}$$

$$\text{Probability for the RED} = \frac{20}{60}$$

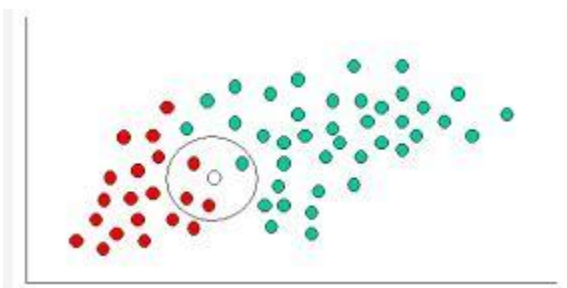


Figure 38 Probability calculated

6.5.3 RANDOM FOREST

Random forest is a type of recursive partitioning decision tree methods which involve an ensemble of classification (or regression) trees that are calculated on random subsets of the data, using a subset of randomly restricted and selected predictors for each split in each classification tree. In this way, random forests are able to better examine the contribution and behavior that each predictor has, even when one predictor's effect would usually be overshadowed by more significant competitors in simpler models (e.g., simple or mixed effect regression models). Furthermore, the results of an ensemble of classification/regression trees have been shown to produce better predictions than the results of one classification tree on its own [86]

6.6 FEATURE SELECTION

To select the optimum features for signature verification Fisher's Discriminate Ratio (FDR) is one of such widely used techniques in pattern recognition and statistical analysis of data. Some of the other techniques are – Principal Component Analysis, Factor Analysis, and Independent Component Analysis [87].

6.6.1 FISHER'S DISCRIMINATION RATIO (FDR)

Consider two cases shown in figure 4.3. In the first case, the standard deviation within the Classes is high and classes slightly overlapping in the feature space. In the second case, the standard deviation is less and class separation is also high. Higher class separation means larger distance between the mean of the classes. Consider the mathematical expression of FDR,

$$\mathbf{FDR} = \frac{|\mu_1 - \mu_2|}{\sqrt{\sigma_1^2 - \sigma_2^2}} \qquad \mathbf{(5)}$$

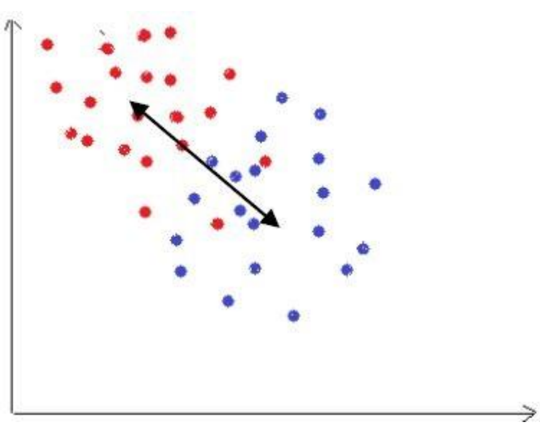


Figure 39 Feature space case 1

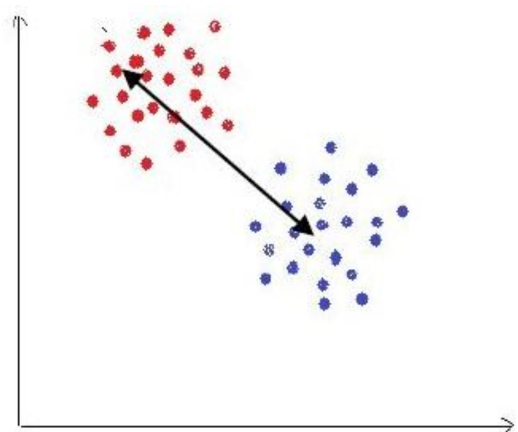


Figure 40 Feature space case 2

As per observations from the figures below it can be observed that, higher difference between the means, and lower the standard deviation within the higher is the discrimination between the two given classes. Hence higher is the FDR, better is the feature for classification and eventually the respective feature gives better classification accuracy. In case 1, the values are highly scattered around the mean as show in the Figure 39. In case 2, the values are not scattered as much in case 1 and hence gives higher FDR value compared to case 1. Hence feature sets shown in Figure 40 give higher classification accuracy and better discrimination between two classes.

Chapter 7

Results

This chapter deals with the results of the work at various levels. The results are presented as:

1. On the basis of different combinations of the features
2. On the basis of different classifiers
3. Confusion matrix
4. ROC Curve

7.1 BASIC APPROACH

Initially the data set is created which has some original signatures and some forged signatures, Classification is done on the basis of combinations of different features and the accuracy is calculated for the different classifiers like k-nn (for different values of n), naïve bayes, and random forest.

In this thesis 3 data sets have been used which comprises of mixture of genuine and forged signatures. Sample of input data has been showed below which has 7 features and 4 additional features which are modified version of the existing features. Hybrid features used are the originated from the altitude and pressure feature. Hybrid features used are H1, H2, H3 and H4 in which H1 is obtained by squaring the Altitude feature values, H2 is obtained by cubing the Altitude feature, H3 is obtained by squaring the pressure feature value and H4 is obtained by cubing the pressure feature value. ROC curve and confusion matrix has been showed of the best combination of features and the best classifier among all the classifiers.

7.2 SAMPLE DATA

Table 1 Sample of input data

X-Coordinate	Y-Coordinate	Time stamp	Pen status	Azimuth	Altitude	Altitude^2	Altitude^3	Pressure^2	Pressure^3	Pressure	Signature
8047	6694	31303424	1	1500	520	270400	140608000	518400	373248000	720	Real
8126	6532	31303434	1	1500	520	270400	140608000	592900	456533000	770	Real
8163	6270	31303444	1	1510	510	260100	132651000	635209	506261573	797	Real
8163	5840	31303454	1	1520	510	260100	132651000	638401	510082399	799	Real
8127	5299	31303464	1	1520	510	260100	132651000	644809	517781627	803	Real
8095	4685	31303474	1	1530	500	250000	125000000	659344	535387328	812	Real
8095	4191	31303484	1	1540	500	250000	125000000	664225	541343375	815	Real
8116	3820	31303494	1	1540	500	250000	125000000	674041	553387661	821	Real
8157	3531	31303504	1	1540	500	250000	125000000	682276	563559976	826	Real
8180	3390	31303514	1	1550	500	250000	125000000	682276	563559976	826	Real
8192	3333	31303524	1	1560	510	260100	132651000	674041	553387661	821	Real
8186	3333	31303534	1	1560	510	260100	132651000	582169	444194947	763	Real
8116	3375	31303544	1	1550	520	270400	140608000	404496	257259456	636	Real
8004	3481	31303555	1	1560	530	280900	148877000	180625	76765625	425	Real
6022	4506	31303655	0	1520	530	280900	148877000	318096	179406144	564	Real
6121	4508	31303665	1	1520	530	280900	148877000	444889	296740963	667	Real
6224	4658	31303675	1	1510	510	260100	132651000	559504	418508992	748	Real
6581	4840	31303685	1	1500	510	260100	132651000	594441	458314011	771	Real
7064	4954	31303695	1	1510	500	250000	125000000	630436	500566184	794	Real
7630	5014	31303705	1	1510	500	250000	125000000	628849	498677257	793	Real
8179	5039	31303715	1	1510	490	240100	117649000	619369	487443403	787	Real
8634	5105	31303725	1	1510	490	240100	117649000	633616	504358336	796	Real
9025	5167	31303735	1	1520	490	240100	117649000	577600	438976000	760	Real
9176	5227	31303745	1	1530	490	240100	117649000	409600	262144000	640	Real

X-Coordinate	Y-Coordinate	Time stamp	Pen status	Azimuth	Altitude	Altitude^2	Altitude^3	Pressure^2	Pressure^3	Pressure	Signature
9214	5265	31303755	1	1530	490	240100	117649000	99225	31255875	315	Real
7885	6088	31303835	0	1550	500	250000	125000000	260100	132651000	510	Real
7543	5669	31303845	1	1560	500	250000	125000000	349281	206425071	591	Real
7436	5588	31303855	1	1560	500	250000	125000000	380689	234885113	617	Real
7360	5566	31303855	1	1570	490	240100	117649000	380689	234885113	617	Real
7372	5403	31303865	1	1560	480	230400	110592000	274576	143877824	524	Real
2184	5575	27499976	0	1410	390	152100	59319000	215296	99897344	464	Real
2313	5691	27499986	1	1410	390	152100	59319000	222784	105154048	472	Real
2339	5932	27499996	1	1410	390	152100	59319000	225625	107171875	475	Real
2417	6026	27500006	1	1410	390	152100	59319000	225625	107171875	475	Real
2528	6069	27500016	1	1410	390	152100	59319000	228484	109215352	478	Real
2660	6079	27500026	1	1450	360	129600	46656000	241081	118370771	491	Real
2787	6106	27500036	1	1500	330	108900	35937000	272484	142236648	522	Real
2919	6158	27500046	1	1560	310	96100	29791000	319225	180362125	565	Real
3024	6228	27500056	1	1590	280	78400	21952000	346921	204336469	589	Real
3109	6211	27500066	1	1590	280	78400	21952000	384400	238328000	620	Real
3210	6133	27500076	1	1590	290	84100	24389000	407044	259694072	638	Real
3283	6014	27500086	1	1590	300	90000	27000000	432964	284890312	658	Real
3336	5786	27500096	1	1590	310	96100	29791000	480249	332812557	693	Real
3354	5489	27500106	1	1570	330	108900	35937000	494209	347428927	703	Real
3344	5145	27500116	1	1560	340	115600	39304000	508369	362467097	713	Real
3315	4682	27500126	1	1560	350	122500	42875000	532900	389017000	730	Real
3285	4249	27500136	1	1540	370	136900	50653000	534361	390617891	731	Real
3251	3837	27500146	1	1530	390	152100	59319000	521284	376367048	722	Real
3213	3549	27500156	1	1530	400	160000	64000000	524176	379503424	724	Real
3195	3362	27500166	1	1510	410	168100	68921000	529984	385828352	728	Real
3184	3243	27500176	1	1500	430	184900	79507000	412164	264609288	642	Real
3184	3182	27500186	1	1490	440	193600	85184000	173889	72511713	417	Real

X-Coordinate	Y-Coordinate	Time stamp	Pen status	Azimuth	Altitude	Altitude^2	Altitude^3	Pressure^2	Pressure^3	Pressure	Signature
2292	3922	27500306	0	1410	490	240100	117649000	380689	234885113	617	Real
2536	3989	27500316	1	1410	490	240100	117649000	383161	237176659	619	Real
2602	4173	27500326	1	1410	490	240100	117649000	379456	233744896	616	Real
2788	4426	27500336	1	1410	490	240100	117649000	384400	238328000	620	Real
3054	4537	27500346	1	1410	490	240100	117649000	373321	228099131	611	Real
3286	4552	27500356	1	1410	490	240100	117649000	390625	244140625	625	Real
3473	4502	27500366	1	1420	480	230400	110592000	418609	270840023	647	Real
3630	4399	27500376	1	1400	480	230400	110592000	429025	281011375	655	Real
3713	4218	27500386	1	1400	480	230400	110592000	442225	294079625	665	Real
3713	3985	27500396	1	1420	470	220900	103823000	436921	288804781	661	Real
3665	3634	27500406	1	1420	470	220900	103823000	432964	284890312	658	Real
3577	3279	27500406	1	1420	470	220900	103823000	430336	282300416	656	Real
3494	2917	27500416	1	1430	480	230400	110592000	429025	281011375	655	Real
3415	2671	27500426	1	1430	480	230400	110592000	448900	300763000	670	Real
3386	2503	27500436	1	1430	480	230400	110592000	488601	341532099	699	Real
3386	2436	27500446	1	1430	480	230400	110592000	552049	410172407	743	Real
3408	2411	27500456	1	1430	480	230400	110592000	620944	489303872	788	Real
3482	2411	27500466	1	1430	480	230400	110592000	654481	529475129	809	Real
3645	2475	27500476	1	1440	480	230400	110592000	654481	529475129	809	Real
3866	2611	27500486	1	1430	490	240100	117649000	643204	515849608	802	Real
4115	2797	27500496	1	1430	490	240100	117649000	579121	440711081	761	Real
4426	3036	27500506	1	1440	490	240100	117649000	380689	234885113	617	Real
4777	3249	27500516	1	1440	490	240100	117649000	431649	283593393	657	Real
6022	5181	27500657	0	1530	470	220900	103823000	268324	138991832	518	Real
6184	5240	27500667	1	1560	450	202500	91125000	284089	151419437	533	Real
6195	5340	27500677	1	1580	440	193600	85184000	313600	175616000	560	Real
6187	5250	27500687	1	1610	420	176400	74088000	384400	238328000	620	Real
6229	5001	27500697	1	1630	420	176400	74088000	447561	299418309	669	Real

X-Coordinate	Y-Coordinate	Time stamp	Pen status	Azimuth	Altitude	Altitude^2	Altitude^3	Pressure^2	Pressure^3	Pressure	Signature
6275	4645	27500707	1	1630	420	176400	74088000	492804	345948408	702	Real
6275	4282	27500717	1	1630	420	176400	74088000	499849	353393243	707	Real
6234	3963	27500727	1	1640	410	168100	68921000	514089	368601813	717	Real
6214	3673	27500737	1	1650	410	168100	68921000	525625	381078125	725	Real
6195	3509	27500747	1	1640	420	176400	74088000	524176	379503424	724	Real
6175	3384	27500757	1	1640	420	176400	74088000	446224	298077632	668	Real
6153	3285	27500767	1	1640	430	184900	79507000	223729	105823817	473	Real
5773	4812	27500857	0	1550	510	260100	132651000	313600	175616000	560	Real
6166	5097	27500867	1	1550	500	250000	125000000	396900	250047000	630	Real
6343	5399	27500877	1	1550	500	250000	125000000	452929	304821217	673	Real
6523	5814	27500887	1	1560	500	250000	125000000	476100	328509000	690	Real
6705	6031	27500897	1	1560	500	250000	125000000	499849	353393243	707	Real
6908	6119	27500907	1	1580	490	240100	117649000	534361	390617891	731	Real
7074	6119	27500917	1	1590	480	230400	110592000	555025	413493625	745	Real
7214	6030	27500927	1	1600	470	220900	103823000	620944	489303872	788	Real
7359	5826	27500937	1	1610	470	220900	103823000	683929	565609283	827	Real
7464	5547	27500947	1	1610	470	220900	103823000	765625	669921875	875	Real
7542	5112	27500957	1	1610	460	211600	97336000	819025	741217625	905	Real
7577	4610	27500967	1	1610	460	211600	97336000	831744	758550528	912	Real
7562	4063	27500977	1	1620	470	220900	103823000	840889	771095213	917	Real
7513	3604	27500987	1	1620	470	220900	103823000	887364	835896888	942	Real
7407	3253	27500997	1	1620	470	220900	103823000	870489	812166237	933	Real
7315	2974	27501007	1	1620	480	230400	110592000	889249	838561807	943	Real
7231	2766	27501017	1	1610	480	230400	110592000	879844	825293672	938	Real
7163	2643	27501027	1	1620	500	250000	125000000	879844	825293672	938	Real
7118	2578	27501037	1	1620	500	250000	125000000	786769	697864103	887	Real
7089	2573	27501047	1	1610	500	250000	125000000	556516	415160936	746	Real
7041	2624	27501057	1	1600	510	260100	132651000	176400	74088000	420	Real

X-Coordinate	Y-Coordinate	Time stamp	Pen status	Azimuth	Altitude	Altitude^2	Altitude^3	Pressure^2	Pressure^3	Pressure	Signature
5109	3610	27501157	0	1570	540	291600	157464000	354025	210644875	595	Real
5143	3624	27501167	1	1560	530	280900	148877000	434281	286191179	659	Real
5203	3735	27501177	1	1550	520	270400	140608000	484416	337153536	696	Real
5451	3866	27501187	1	1540	520	270400	140608000	528529	384240583	727	Real
5823	3972	27501197	1	1550	510	260100	132651000	562500	421875000	750	Real
6302	4089	27501207	1	1550	500	250000	125000000	624100	493039000	790	Real
6815	4233	27501217	1	1540	500	250000	125000000	628849	498677257	793	Real
7351	4392	27501227	1	1550	490	240100	117649000	633616	504358336	796	Real
7713	4524	27501237	1	1550	490	240100	117649000	616225	483736625	785	Real
7879	4595	27501247	1	1550	490	240100	117649000	492804	345948408	702	Real
7913	4626	27501257	1	1560	480	230400	110592000	259081	131872229	509	Real
7342	5089	27501307	0	1590	490	240100	117649000	137641	51064811	371	Real
6989	5116	27501317	1	1610	490	240100	117649000	218089	101847563	467	Real
6872	5098	27501327	1	1610	490	240100	117649000	270400	140608000	520	Real
6769	5030	27501337	1	1620	500	250000	125000000	280900	148877000	530	Real
6711	4928	27501347	1	1620	490	240100	117649000	288369	154854153	537	Real
6684	4744	27501357	1	1620	490	240100	117649000	219024	102503232	468	Real
3183	4804	60610112	0	1550	390	152100	59319000	82369	23639903	287	Forged
3100	4837	60610122	1	1520	410	168100	68921000	90601	27270901	301	Forged
3047	4874	60610132	1	1500	420	176400	74088000	87616	25934336	296	Forged
2996	4912	60610142	1	1470	430	184900	79507000	92416	28094464	304	Forged
3003	4862	60610152	1	1440	440	193600	85184000	94249	28934443	307	Forged
3032	4737	60610162	1	1430	450	202500	91125000	100489	31855013	317	Forged
3075	4562	60610172	1	1410	450	202500	91125000	98596	30959144	314	Forged
3129	4305	60610182	1	1400	460	211600	97336000	103684	33386248	322	Forged
3200	4014	60610192	1	1400	460	211600	97336000	106929	34965783	327	Forged
3282	3701	60610202	1	1390	460	211600	97336000	116964	40001688	342	Forged
3363	3517	60610212	1	1380	470	220900	103823000	119716	41421736	346	Forged

X-Coordinate	Y-Coordinate	Time stamp	Pen status	Azimuth	Altitude	Altitude^2	Altitude^3	Pressure^2	Pressure^3	Pressure	Signature
3430	3396	60610222	1	1380	470	220900	103823000	134689	49430863	367	Forged
3488	3317	60610232	1	1380	480	230400	110592000	159201	63521199	399	Forged
3526	3288	60610242	1	1370	480	230400	110592000	193600	85184000	440	Forged
3560	3288	60610252	1	1360	490	240100	117649000	216225	100544625	465	Forged
3591	3304	60610262	1	1360	490	240100	117649000	221841	104487111	471	Forged
3627	3366	60610272	1	1350	500	250000	125000000	219024	102503232	468	Forged
3664	3475	60610282	1	1340	500	250000	125000000	214369	99252847	463	Forged
3690	3615	60610292	1	1340	500	250000	125000000	207936	94818816	456	Forged
3707	3791	60610302	1	1330	510	260100	132651000	212521	97972181	461	Forged
3721	3924	60610312	1	1340	510	260100	132651000	219024	102503232	468	Forged
3737	4029	60610322	1	1340	510	260100	132651000	216225	100544625	465	Forged
3751	4029	60610332	1	1340	510	260100	132651000	219024	102503232	468	Forged
3763	3963	60610342	1	1340	510	260100	132651000	215296	99897344	464	Forged
3799	3854	60610352	1	1340	510	260100	132651000	218089	101847563	467	Forged
3840	3698	60610362	1	1340	510	260100	132651000	215296	99897344	464	Forged
3897	3581	60610372	1	1340	510	260100	132651000	225625	107171875	475	Forged
3963	3490	60610382	1	1340	510	260100	132651000	245025	121287375	495	Forged
4031	3447	60610392	1	1340	510	260100	132651000	247009	122763473	497	Forged
4098	3447	60610402	1	1340	510	260100	132651000	271441	141420761	521	Forged
4165	3486	60610412	1	1330	510	260100	132651000	274576	143877824	524	Forged
4231	3570	60610422	1	1330	510	260100	132651000	277729	146363183	527	Forged
4288	3708	60610432	1	1330	510	260100	132651000	272484	142236648	522	Forged
4331	3883	60610442	1	1330	510	260100	132651000	276676	145531576	526	Forged
4364	4086	60610452	1	1330	510	260100	132651000	276676	145531576	526	Forged
4394	4283	60610462	1	1340	500	250000	125000000	276676	145531576	526	Forged
4426	4475	60610472	1	1340	500	250000	125000000	272484	142236648	522	Forged
4450	4578	60610482	1	1330	510	260100	132651000	262144	134217728	512	Forged
4469	4624	60610492	1	1330	510	260100	132651000	159201	63521199	399	Forged

X-Coordinate	Y-Coordinate	Time stamp	Pen status	Azimuth	Altitude	Altitude^2	Altitude^3	Pressure^2	Pressure^3	Pressure	Signature
4494	4631	60610502	1	1330	510	260100	132651000	8281	753571	91	Forged
4922	4435	60610633	0	1350	490	240100	117649000	106929	34965783	327	Forged
4986	4333	60610643	1	1360	490	240100	117649000	106929	34965783	327	Forged
4959	4388	60610653	1	1370	480	230400	110592000	112225	37595375	335	Forged
4967	4314	60610663	1	1370	480	230400	110592000	116964	40001688	342	Forged
4999	4214	60610673	1	1380	490	240100	117649000	119716	41421736	346	Forged
5033	4098	60610683	1	1380	490	240100	117649000	137641	51064811	371	Forged
5066	4003	60610693	1	1380	490	240100	117649000	149769	57960603	387	Forged
5093	3926	60610703	1	1380	490	240100	117649000	157609	62570773	397	Forged
5117	3862	60610713	1	1380	500	250000	125000000	159201	63521199	399	Forged
5135	3822	60610723	1	1380	500	250000	125000000	157609	62570773	397	Forged
5150	3800	60610733	1	1380	500	250000	125000000	158404	63044792	398	Forged
5162	3790	60610743	1	1400	510	260100	132651000	137641	51064811	371	Forged
5162	3790	60610753	1	1400	510	260100	132651000	15376	1906624	124	Forged
4870	5032	60610893	0	1380	490	240100	117649000	21316	3112136	146	Forged
4870	5032	60610903	1	1390	480	230400	110592000	44521	9393931	211	Forged
4870	5032	60610913	1	1390	480	230400	110592000	51984	11852352	228	Forged
4870	5092	60610923	1	1390	480	230400	110592000	54289	12649337	233	Forged
4877	5084	60610933	1	1390	480	230400	110592000	56169	13312053	237	Forged
4898	5053	60610943	1	1400	480	230400	110592000	57600	13824000	240	Forged
4929	5007	60610953	1	1400	480	230400	110592000	57121	13651919	239	Forged
4957	4970	60610963	1	1400	480	230400	110592000	58564	14172488	242	Forged
4981	4939	60610973	1	1420	480	230400	110592000	58564	14172488	242	Forged
5002	4913	60610983	1	1420	480	230400	110592000	56169	13312053	237	Forged
5023	4890	60610993	1	1420	480	230400	110592000	44521	9393931	211	Forged
5593	4699	60611104	0	1440	490	240100	117649000	43264	8998912	208	Forged
5702	4793	60611114	1	1430	500	250000	125000000	62001	15438249	249	Forged
5719	4885	60611124	1	1430	500	250000	125000000	90601	27270901	301	Forged

X-Coordinate	Y-Coordinate	Time stamp	Pen status	Azimuth	Altitude	Altitude^2	Altitude^3	Pressure^2	Pressure^3	Pressure	Signature
5726	4920	60611134	1	1430	500	250000	125000000	103684	33386248	322	Forged
5726	4929	60611144	1	1430	500	250000	125000000	114921	38958219	339	Forged
5720	4922	60611154	1	1430	500	250000	125000000	126736	45118016	356	Forged
5685	4902	60611164	1	1430	500	250000	125000000	132496	48228544	364	Forged
5631	4833	60611174	1	1430	500	250000	125000000	138384	51478848	372	Forged
5573	4731	60611184	1	1430	500	250000	125000000	156816	62099136	396	Forged
5545	4629	60611194	1	1430	500	250000	125000000	159201	63521199	399	Forged
5537	4524	60611204	1	1430	500	250000	125000000	176400	74088000	420	Forged
5563	4414	60611214	1	1420	500	250000	125000000	184900	79507000	430	Forged
5627	4322	60611224	1	1420	500	250000	125000000	193600	85184000	440	Forged
5717	4241	60611234	1	1420	500	250000	125000000	205209	92959677	453	Forged
5840	4178	60611244	1	1430	510	260100	132651000	216225	100544625	465	Forged
5948	4119	60611254	1	1430	510	260100	132651000	219024	102503232	468	Forged
6047	4064	60611264	1	1430	510	260100	132651000	221841	104487111	471	Forged
6126	4003	60611274	1	1430	510	260100	132651000	223729	105823817	473	Forged
6176	3942	60611284	1	1430	510	260100	132651000	215296	99897344	464	Forged
6206	3865	60611294	1	1440	510	260100	132651000	218089	101847563	467	Forged
6201	3779	60611304	1	1430	520	270400	140608000	230400	110592000	480	Forged
6165	3695	60611314	1	1430	520	270400	140608000	241081	118370771	491	Forged
6108	3624	60611324	1	1430	530	280900	148877000	248004	123505992	498	Forged
6045	3562	60611334	1	1430	530	280900	148877000	254016	128024064	504	Forged
6002	3503	60611344	1	1420	540	291600	157464000	259081	131872229	509	Forged
5971	3464	60611354	1	1440	540	291600	157464000	255025	128787625	505	Forged
5971	3439	60611364	1	1440	540	291600	157464000	249001	124251499	499	Forged
5995	3431	60611374	1	1440	540	291600	157464000	259081	131872229	509	Forged
6032	3464	60611384	1	1430	550	302500	166375000	260100	132651000	510	Forged
6117	3524	60611394	1	1430	550	302500	166375000	257049	130323843	507	Forged
6236	3651	60611404	1	1430	550	302500	166375000	252004	126506008	502	Forged

X-Coordinate	Y-Coordinate	Time stamp	Pen status	Azimuth	Altitude	Altitude^2	Altitude^3	Pressure^2	Pressure^3	Pressure	Signature
6377	3829	60611414	1	1430	550	302500	166375000	246016	122023936	496	Forged
6499	4042	60611424	1	1430	550	302500	166375000	242064	119095488	492	Forged
6612	4282	60611434	1	1430	550	302500	166375000	241081	118370771	491	Forged
6720	4549	60611444	1	1430	550	302500	166375000	239121	116930169	489	Forged
6818	4834	60611454	1	1430	550	302500	166375000	239121	116930169	489	Forged
6903	5104	60611464	1	1430	550	302500	166375000	242064	119095488	492	Forged
6980	5345	60611474	1	1430	550	302500	166375000	251001	125751501	501	Forged
7039	5568	60611484	1	1440	540	291600	157464000	265225	136590875	515	Forged
7070	5723	60611494	1	1440	540	291600	157464000	284089	151419437	533	Forged
7081	5816	60611504	1	1460	540	291600	157464000	310249	172808693	557	Forged
7081	5869	60611514	1	1460	540	291600	157464000	321489	182284263	567	Forged
7064	5869	60611524	1	1470	530	280900	148877000	322624	183250432	568	Forged
7040	5805	60611534	1	1470	530	280900	148877000	327184	187149248	572	Forged
6988	5703	60611544	1	1480	530	280900	148877000	339889	198155287	583	Forged
6912	5502	60611554	1	1490	520	270400	140608000	350464	207474688	592	Forged
6819	5202	60611564	1	1490	520	270400	140608000	345744	203297472	588	Forged
6735	4836	60611574	1	1490	520	270400	140608000	349281	206425071	591	Forged
6682	4503	60611584	1	1500	510	260100	132651000	354025	210644875	595	Forged
6650	4279	60611594	1	1500	510	260100	132651000	363609	219256227	603	Forged
6656	4128	60611604	1	1510	510	260100	132651000	379456	233744896	616	Forged
6667	4061	60611614	1	1510	510	260100	132651000	380689	234885113	617	Forged
6683	4030	60611624	1	1510	510	260100	132651000	379456	233744896	616	Forged
6698	4023	60611634	1	1510	510	260100	132651000	401956	254840104	634	Forged
6715	4039	60611644	1	1490	520	270400	140608000	407044	259694072	638	Forged
6766	4128	60611654	1	1490	520	270400	140608000	398161	251239591	631	Forged
6837	4267	60611664	1	1490	520	270400	140608000	396900	250047000	630	Forged
6912	4454	60611674	1	1480	530	280900	148877000	400689	253636137	633	Forged
6967	4571	60611684	1	1470	530	280900	148877000	407044	259694072	638	Forged

7.3 PERFORMANCE ANALYSIS

When performance analysis was done with 5 data sets following results were obtained.

Table 2 Data set with "User 1" has his/her 10 real signatures and 20 Forgeries of his/her own

	Features of Signatures	**Accuracy obtained using various classifiers used**						
		K-nn (k=1)	K-nn (k=3)	K-nn (k=5)	K-nn (k=7)	K-nn (k=9)	Naïve Bayes	Random Forest
1.	X,Y Coordinates	74.9%	**77.5%**	77.5%	78.2%	78.4%	76.8%	77.0%
2.	X-Y ,Azimuthal	89.9%	**90.8%**	90.3%	90.4%	90.6%	82.2%	90.8%
3.	X-Y,Altitude	83.5%	**85.2%**	85.5%	84.3%	84.6%	76.4%	85.3%
4.	X-Y,Azimuthal,Pressure	95.6%	**94.7%**	93.9%	92.8%	92.5%	81.0%	93.7%
5.	X-Y,Azimuthal,Altitude	96.6%	**95.3%**	94.2%	93.6%	92.4%	82.1%	95.0%
6.	X-Y,Pressure,Altitude	91.6%	**91.4%**	90.2%	89.8%	89.9%	77.4%	90.5%
7.	X-Y,Pre,Alt,H1,H2,H3,H4	92.3%	**92.0%**	89.7%	89.4%	89.0%	76.5	90.7%
8.	X-Y,Pen Status	75.1%	**77.9%**	77.6 %	78.6%	78.7%	76.9%	78.0%
9.	X-Y, Azimuthal,Altitude,Pressure	97.9%	**97.0%**	95.7%	94.7%	94.7%	82.3%	95.9%
10.	X-Y,Azimuthal,Altitude,Pressure,Time stamp	99.8%	**99.7%**	99.6%	99.3%	99.1%	99.0%	99.89%

Table 3 Data set in with “User 1” has his/her 10 real signatures and 20 Forgeries of other user say “User 6”

Sr. No.	**Features of Signatures**	**Accuracy obtained using various classifiers used**						
		K-nn (k=1)	K-nn (k=3)	K-nn (k=5)	K-nn (k=7)	K-nn (k=9)	Naïve Bayes	Random Forest
1.	X,Y Coordinates	77.1%	**82.7%**	82.4%	82.9%	82.9%	78.6%	80.4%
2.	X-Y ,Azimuthal	86.7%	**87.4%**	86.2%	85.8%	85.5%	80.2%	88.0%
3.	X-Y,Altitude	90.6%	**91.2%**	90.2%	89.8%	89.5%	80.7%	91.4%
4.	X-Y,Azimuthal,Pressure	90.3%	**89.3%**	88.7%	88.0%	87.6%	80.1%	91.8%
5.	X-Y,Azimuthal,Altitude	92.6%	**91.6%**	90.9%	90.6%	90.1%	82.4%	93.9%
6.	X-Y,Pressure,Altitude	92.2%	**92.0%**	92.1%	91.3%	91.0%	80.9%	93.0%
7.	X-Y,Pre,Alt,H1,H2,H3,H4	92.8%	**92.5%**	91.7%	91.2%	90.7%	81.5%	92.9%
8.	X-Y,Pen Status	77.7%	**81.2%**	81.5 %	81.9%	82.2%	78.2%	80.5%
9.	X-Y, Azimuthal,Altitude,Pressure	94.2%	**92.9%**	92.9%	92.0%	91.8%	82.3%	95.6%
10.	X-Y, Azimuthal,Altitude,Pressure,Time stamp	99.6%	**99.69%**	99.5%	99.4%	99.3%	99.4%	99.8%

Table 4 Data set in which a 3 different User's 10 real signatures and 3 different user's 20 forged signature are used.

Sr. No.	**Features of Signatures**	**Accuracy obtained using various classifiers used**						
		K-nn (k=1)	K-nn (k=3)	K-nn (k=5)	K-nn (k=7)	K-nn (k=9)	Naïve Bayes	Random Forest
1.	X,Y Coordinates	76.5%	**81.1%**	82.2%	81.9%	82.5%	80.9%	79.4%
2.	X-Y ,Azimuthal	93.1%	**93.4%**	93.0%	92.7%	92.4%	80.2%	93.0%
3.	X-Y,Altitude	91.3%	**91.6%**	91.0%	90.9%	90.4%	80.9%	90.8%
4.	X-Y,Azimuthal,Pressure	96.9%	**96.2%**	95.8%	95.4%	95.2%	84.1%	96.0%
5.	X-Y,Azimuthal,Altitude	98.8%	**98.4%**	98.0%	97.6%	97.6%	85.8%	98.2%
6.	X-Y,Pressure,Altitude	95.5%	**94.8%**	94.2%	93.9%	93.5%	80.9%	94.2%
7.	X-Y,Pre,Alt,H1,H2,H3,H4	95.2%	**94.9%**	94.4%	94.2%	93.6%	69.6%	94.3%
8.	X-Y,Pen Status	76.7%	**80.1%**	81.3%	82.1%	82.3%	80.9%	79.6%
9.	X-Y, Azimuthal,Altitude,Pressure	99.0%	**98.5%**	98.1%	98.0%	98.0%	87.4%	98.4%
10.	X-Y, Azimuthal,Altitude,Pressure, Time stamp	99.9%	**99.8%**	99.8%	99.7%	99.6%	98.9%	99.8%

7.3.1 Analysis of different Data sets

Performance analysis of 18 data sets was carried out for which k-NN classifier (at k=3) gave the best results.

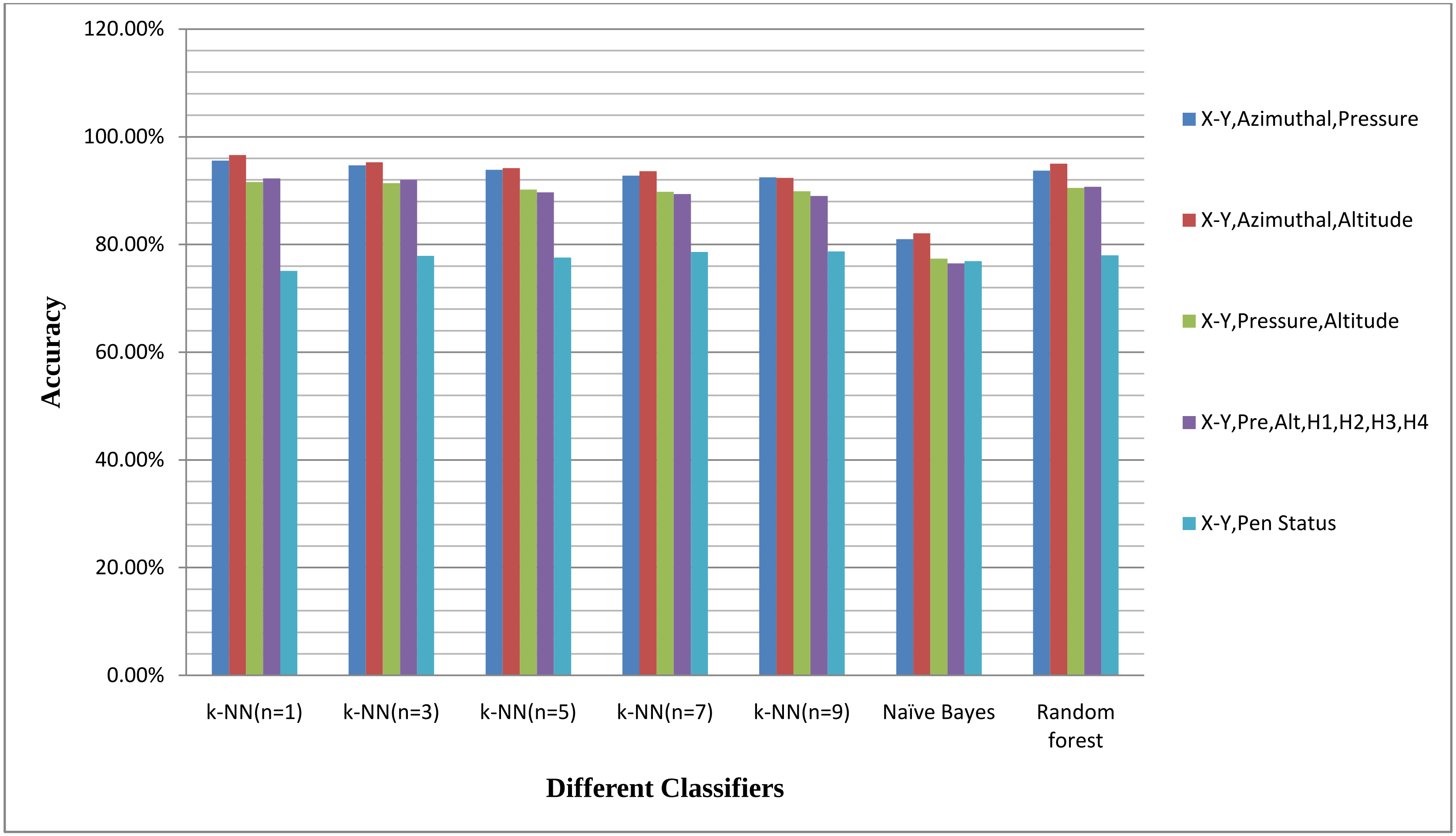


Figure 41 Performance analysis graph of 18 Data sets (5 combinations of features)

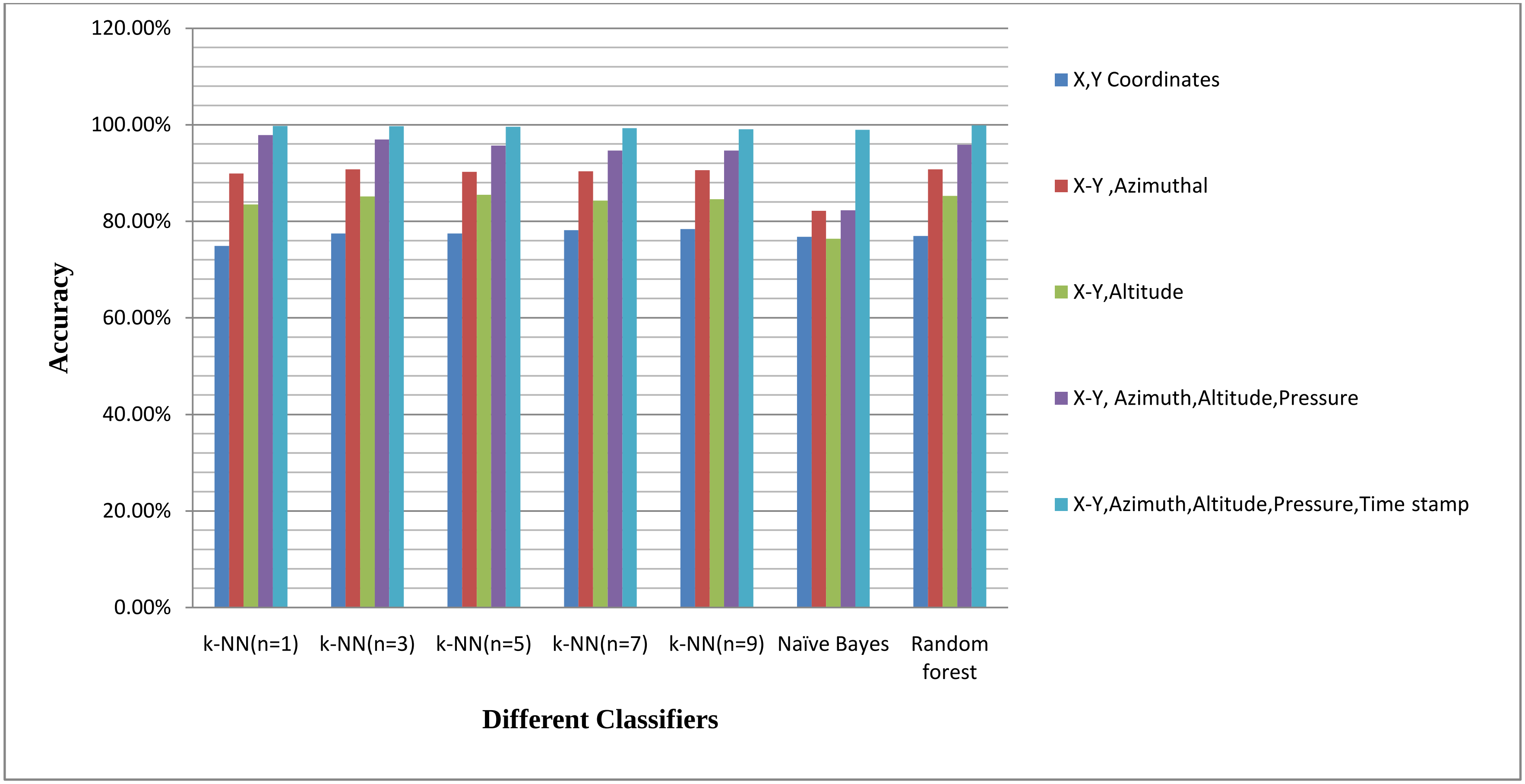


Figure 42 Performance analysis graph of 18 Data sets (5 combinations of features)

7.4 CONFUSION MATRIX OF DATA SETS

A confusion matrix is a specific table layout that allows visualization of the performance of an algorithm. Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class. The name stems from the fact that it makes it easy to see if the system is confusing two classes (i.e. commonly mislabeling one as another).

Table 5 Confusion Matrix

Actual class	**Predicted Class**	
	Real	Forged
Real	True Positive	False Negative
Forged	False Positive	True Negative

Table 6 Confusion matrix of Data set 1

Actual class	**Predicted Class**	
	Real	Forged
Real	1056	0
Forged	10	2967

Table 7 Confusion matrix of Data set 2

Actual class	**Predicted Class**	
	Real	Forged
Real	965	0
Forged	16	3832

Table 8 Confusion matrix of Data set 2

Actual class	**Predicted Class**	
	Real	Forged
Real	2947	0
Forged	47	12535

7.5 CHARACTERSTICS OF PERFORMANCE ANALYSIS

Characteristics like True positive rate, false positive rate and Precision been calculated.

Table 9 Characteristics table of Data set 1

Class	**TP Rate**	**FP Rate**	**Precision**	**Recall**	**F-Measure**
Real	0.995	0.000	1.000	0.994	0.994
Forged	1.0	0.012	0.996	0.998	0.998
Weighted Average	0.997	0.009	0.997	0.997	0.997

Table 10 Characteristics table of Data set 2

Class	**TP Rate**	**FP Rate**	**Precision**	**Recall**	**F-Measure**
Real	0.992	0.000	1.000	0.998	0.998
Forged	1.0	0.003	0.998	0.998	0.998
Weighted Average	0.996	0.039	0.999	0.998	0.998

Table 11 Characteristics table of Data set 3

Class	TP Rate	FP Rate	Precision	Recall	F-Measure
Real	0.992	0.000	1.000	0.993	0.997
Forged	1.0	0.007	0.998	1.0	0.999
Weighted Average	0.996	0.006	0.999	0.999	0.999

7.6 RECEIVER OPERATING CHARACTERISTICS

A receiver operating characteristic (ROC), or simply ROC curve, is a graphical plot which illustrates the performance of a classifier system as its discrimination threshold is varied. It is created by plotting the fraction of true positives out of the positives (TPR = true positive rate) vs. the fraction of false positives out of the negatives (FPR = false positive rate), at various threshold settings.

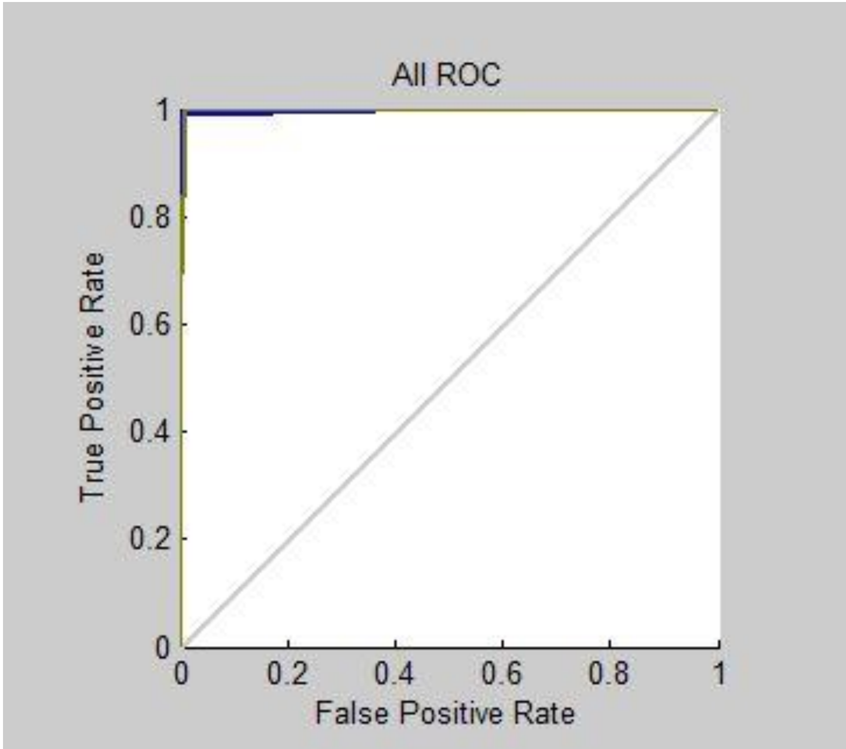


Figure 43 Receiver operating characteristics of optimum features of Data set 1

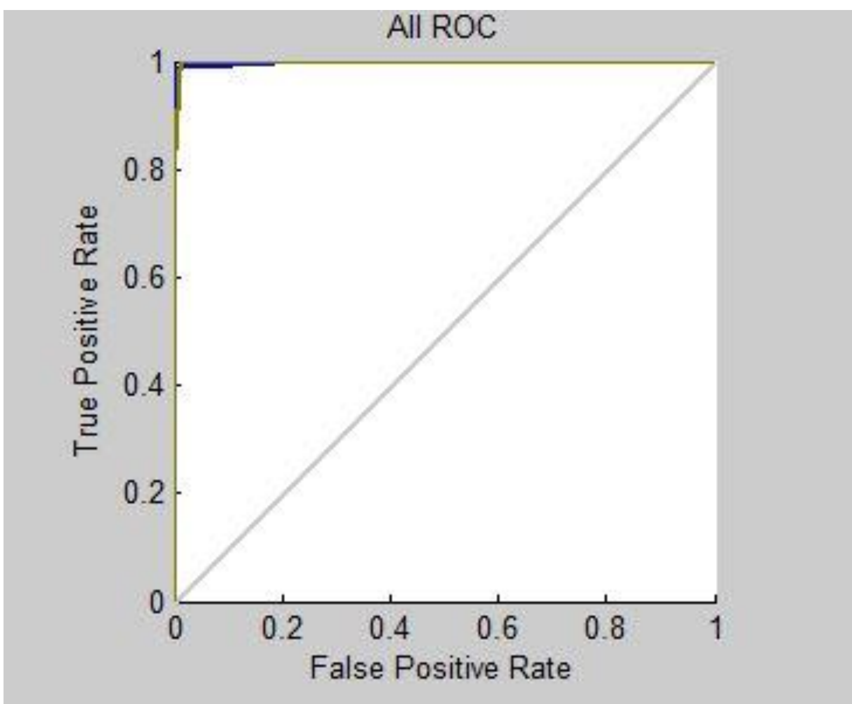


Figure 44 Receiver operating characteristics of optimum features of data set 2

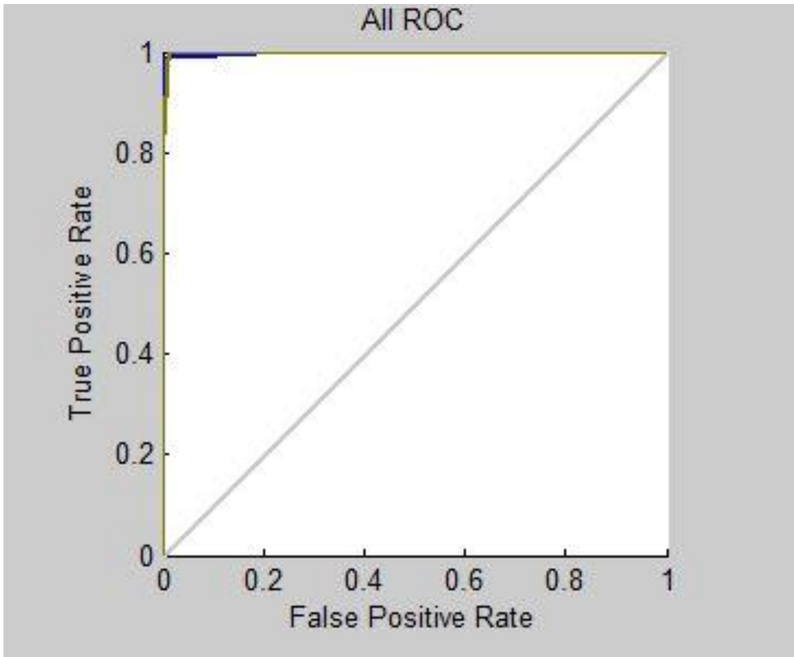


Figure 45 Receiver operating characteristics of optimum features of data set 3

7.7 FEATURE SELECTION

For the optimum feature selection of the fisher's discrimination method is used in which mean and standard deviation is calculated and by using the Fisher's Discrimination ratio formula (Equation 5) and by observing the values of FDR calculated priority of discriminatory power of each feature is obtained. Thus the feature which can discriminate the real and forged signature can be obtained by using this method.

Table 12 Mean and Standard Deviation of Data set 1

Data set 1(4033 samples)	Real samples		Forged samples	
Feature	μ	σ	μ	σ
x	2202.512	5787.235	5386.402	1949.369
y	4299.323	1204.305	4939.674	1224.631
Time Stamp	27523452	17486.83	54035032	10581413
Azimuth	1515.413	97.4331	1310.816	87.02485
Altitude	514.2683	67.29432	519.334	51.17059
Pressure	661.3096	171.1617	390.8612	179.5693
pen status	0.944653	0.228764	0.954034	0.20951

Table 13 FDR of Features of Data set 1

Data set 1	
Feature	FDR
x	0.52137
y	0.37282
Time Stamp	**2.50548**
Azimuth	1.566125
Altitude	0.05992
Pressure	1.090186
Pen status	0.03024

Table 14 Mean and Standard Deviation of Data set 2

Data set 2(4813 samples)	Real samples		Forged samples	
Feature	μ	σ	μ	σ
x	5787.235	2202.512	5604.422	1909.702
y	4299.323	1204.305	5110.686	1209.988
Time Stamp	27523452	17486.83	56392789	16366179
Azimuth	1515.413	97.4331	1375.254	118.3697
Altitude	514.2683	67.29432	532.8413	72.84913
Pressure	661.3096	171.1617	487.6903	201.2057
Pen status	0.944653	0.228764	0.960229	0.195453

Table 15 FDR of Features of Data set 2

Data set 2	
Feature	FDR
x	0.062712
y	0.47527
Time Stamp	**1.76396**
Azimuth	0.914202
Altitude	0.18728
Pressure	0.657253
Pen status	0.05177

Table 16 Mean and Standard Deviation of Data set 3

Data set 3(15529 samples)	Real samples		Forged samples	
	μ	σ	μ	σ
x	6029.298	2096.527	5898.899	2058.677
y	4872.277	1065.586	4363.752	1227.192
Time Stamp	32797298	1065071	47299812	18045329
Azimuth	1162.484	74.79508	1359.111	96.62336
Altitude	582.0699	62.81681	563.8853	86.65938
Pressure	579.4913	158.7901	516.9807	193.5456
Pen status	0.970139	0.170232	0.972514	0.163521

Table 17 FDR of Features of Data set 3

Data set 3	
Feature	FDR
x	0.044379
y	0.312888
Time Stamp	**1.60919**
Azimuth	0.80228
Altitude	0.169899
Pressure	0.249695
Pen status	0.01006

Chapter 8

Conclusion and Future Scope

8.1 CONCLUSION

In this work various combinations of features of signatures were studied like using different classifiers like k-Nearest neighbor classifier (at different threshold values), Naïve bayes, and random forest on 18 data sets ,each data set having 40-50 signatures.

When features X-coordinate and Y-coordinate were used to verify the signatures, the accuracy obtained was around 75% using k-NN (at threshold value 3), Naïve bayes and Random forest on 18 data sets. k-NN at threshold value 3 gave the best results and classifier Naïve bayes gave the worst results.

- Among 11 features used in this work optimum features were selected according to their accuracy obtained.
- The four new features were formulated using pressure and altitude feature, they were obtained by applying mathematical operations on pressure and altitude like squaring and cubing the values of them. There was increment in resultn by approximately 2%.
- After going through different combinations highest accuracy was obtained up to 99% by combining six features: X-coordinate, Y-coordinate, Azimuth angle, Altitude, Pressure, and Time stamp. k-NN at threshold value 3 gave the best accuracy.
- For the feature selection **Fisher's Discrimination ratio** is calculated for each feature using different data sets and according to the FDR obtained for each feature it is observed that Time Stamp was rated with **Highest** FDR and Pen status was rated with **Lowest** FDR.

8.2 FUTURE SCOPE

In future it is expected to interface this work with hardware. The main motive is to obtain best accuracy by deploying minimum number of features in the experiment. Also some other approaches can be employed like Neural networks, Artificial intelligence and Advanced computational mathematics (Regression and correlation). It can be used as biometric for human identification along with iris detection and finger print recognition.

References

[1] S. Prabhakar, S. Pankanti, and A. K. Jain, “Biometric recognition: Security and privacy concerns,” IEEE Security Privacy Mag., vol. 1, no. 2, pp. 33–42, 2003.

[2] R. Sabourin, G. Genest, and F. Preteux, ”Off-Line Signature Verification by Local Granulometric Size Distributions”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 9, 1997.

[3] G. Rigoll, A. Kosmala, ”A Systematic Comparison Between On-Line and Off-Line Methods for Signature Verification with Hidden Markov Models”, In Proceedings of International Conference on Pattern Recognition, vol. 2, pp. 1755-1757, 1998.

[4] J Edson, R. Justino, F. Bortolozzi and R. Sabourin, "An off-line signature verification using HMM for Random, Simple and Skilled Forgeries", Sixth International Conference on Document Analysis and Recognition, pp.1031-1034, Sept.2001. 211-222, Dec 2000.

[5] J.F. Velez, A.Sanchez, and A.B. Moreno, 2003. Robust Off-Line Signature Verification using Compression Networks and Position Cuttings.

[6] Debnath Bhattacharyya, Samir Kumar Bandyopadhyay, Poulami Das, Debashis Ganguly, Swarnendu Mukherjee, “Statistical approach for offline handwritten signature verification”, Journal of Computer Science March 01, 2008.

[7] Ramachandra, Ravi, Raja, Venugopal, Patnaik,"Signature Verification using Graph Matching and Cross-Validation Principle", Int. J. of Recent Trends in Engineering (IJRTE), Vol. 1 (1), pp. 57-61, 2009.

[8] Xiufen Ye, Weiping Hou, Weixing Feng,“Off-line Handwritten Signature Verification", with Inflections Feature”, Proc. of the IEEE International Conference on Mechatronics & Automation, Canada , July 2005

[9] J.L.Wayman, “Fundamentals of biometric authentication technologies”, International Journal of Image and Graphics, vol. 1, no. 1, pp. 93–113, 2001.

[10] L.O. Gorman, “Seven issues with human authentication technologies”, in Proc. of Workshop on Automatic Identification Advanced Technologies (AutoID), (Tarrytown, New York), pp. 185–186, Mar 2002.

[11] R. W. Frischholz and U. Dieckmann, “Bioid: A multimodal biometric identification system,” IEEE Comput., vol. 33, pp. 64–68, 2000.

[12] Schneier and Bruce, “Inside risks: the uses and abuses of biometrics,” Communications of the ACM, Volume 42 Issue 8, August 1999

[13] P. Ambalakat, ”Security of Biometric Authentication Systems”, 21st Computer Science Seminar SA1-T1-1.

[14] M. A. Dabbah, Woo, W. L., and Dlay S. S., "Secure Authentication for Face Recognition," In Proc. of IEEE Symposium on Computational Intelligence in Image and Signal Processing, USA, pp. 121 – 126, Apr. 2007

[15] Im S., Park H., Kim Y., Han S., Kim S., Kang C., and Chung C., “A Biometric Identification System by Extracting Hand Vein Patterns”, Journal of the Korean Physical Society, Korean Publication, Volume 38, Issue 3, pp. 268-272, Mar. 2001

[16] A. Kumar, D. C. Wong , H. C. Shen, and A. K Jain., “Personal Verification using Palmprint and Hand Geometry Biometric”, In Proc. of 4th International Conference on Audio- and Video-based Biometric Person Authentication, Guildford, UK, , pp. 668 – 678, Jun. 2003

[17] S. Ben-Yacoub, Y. Abdeljaoued, and E. Mayoraz, “Fusion of Face and Speech Data for Person Identity Verification,” IDIAP, Martigny, Switzerland, Res. Paper IDIAP-RR 99–03, 1999.

[18] J. Daugman, “Recognizing persons by their Iris patterns,” in Biometrics: Personal Identification in a Networked Society, Eds. Norwell, MA: Kluwer, pp. 103–121, 1999.

[19] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain, “FVC2002: Fingerprint verification competition,” in Proc. Int. Conf. Pattern Recognition (ICPR), Quebec City, QC, Canada, pp. 744–747, Aug. 2002

[20] A. Ross, S. Dass, and A. K. Jain, “A deformable model for fingerprint matching”, Journal of Pattern Recognition, Elsevier, Volume 38, No. 1, pp. 95–103, Jan. 2005

[21] T. Matsumoto, H. Hoshino, K. Yamada, and S. Hasino, “Impact of artificial gummy fingers on fingerprint systems”, In Proc. of SPIE, Volume 4677, pp. 275–289, Feb. 2002

[22] R. Haralick, L. Watson, “A facet model for image data,” Comput. Vision Graphics Image Prcoess, 15, 113-129, 1981.

[23] A. K. Jain, A. Ross, and S. Pankanti, “Biometric: A Tool for Information Security”, IEEE Trans. Information Forensics and Security, Volume 1, No. 2, pp. 125–144, Jun. 2006

[24] R. Cappelli, D. Maio, D. Maltoni, J. L. Wayman, and A. K. Jain, “Performance evaluation of fingerprint verification systems”, IEEE Trans. Pattern Anal. Mach. Intell., Volume 28, issue 1, pp. 3–18, Jan. 2006

[25] A. K. Jain, L. Hong, and R. Bolle, “On-line fingerprint verification”, IEEE Transactions on Pattern Recognition and Machine Intelligence, Volume 19, No. 4,, pp. 302–314, Aug. 1996.

[26] M. A. Dabbah, W. L. Woo, and S. S. Dlay, "Secure Authentication for Face Recognition," In Proc. of IEEE Symposium on Computational Intelligence in Image and Signal Processing, USA, pp. 121 – 126, Apr. 2007

[27] M. Golfarelli, D. Maio, and D. Maltoni, “On the error-reject tradeoff in biometric verification systems,” IEEE Trans. Pattern Anal. Machine Intell., vol. 19, pp. 786–796, July 1997.

[28] A. J. Goldstein, L.D. Harmon, And A. B. Lesk, “Identification Of Human Faces,” in Proc. IEEE, pp. 748-760, May 1971.

[29] N. D. Haig, “How Faces Differ-A New Comparative Technique,” Perception 14, pp. 601-615, July 1985.

[30] G. Rhodes, “Looking At Faces: First-Order And Second Order Feature As Determinants Of Facial Appearance,” Perception 17, pp. 43-63, 1988.

[31] L. Hong and A. K. Jain, “Integrating faces and fingerprints for personal identification”, IEEE Trans. Pattern Anal. Mach. Intell., Volume 20, No. 12, pp. 1295–1307, Dec. 1998.

[32] A. Ross and R. Govindarajan, “Feature level fusion using hand and face biometrics”, In Proc. of SPIE Conf. Biometric Technology for Human Identification II, pp. 196–204, Mar. 2005.

[33] S. Z. Li and A. K. Jain, Eds., Handbook of Face Recognition. New York: Springer Verlag, 2004.

[34] F. Cardinaux, C. Sanderson, and S. Bengio, "User Authentication via Adapted Statistical Models of Face Images", IEEE Transaction on Signal Processing, Volume 54, Issue 1, pp. 361 – 373, Jan. 2006.

[35] B. Mathivanan, Dr. V. Palanisamy, Dr. S. Selvarajan, "An Efficient Hand Image Segmentation Algorithm for Hand Geometry based Biometrics Recognition System," International Journal of Computer Application, Vol. 35, No. 10, Dec 2011.

[36] Yaroslav Bulatov, Sachin Jambawalikar, Piyush Kumar, Saurabh Sethia, "Hand recognition using geometric classifiers."

[37] R. Sanchez-Reillo, C. Sanchez-Avila, and A. Gonzales-Marcos, "Biometric identification through hand geometry measurements,", IEEE Trans. Pattern Anal. Mach. Intell., Volume 22, Issue. 10, pp. 1168–1171, Oct. 2000

[38] E. Kukula, S. Elliott, "Implementation of Hand Geometry at Purdue University's Recreational Center: An Analysis of User Perspectives and System Performance", In Proc. of 35th Annual International Carnahan Conference on Security Technology, UK, pp. 83 – 88, Oct. 2001.

[39] Zuraini Othman, Anton Satria Prabuwono, "Preliminary Study on Iris Recognition System: Tissues of Body Organ in Iridology," IEEE EMBS Conference on Biomedical Engineering & Sciences (IECBES), 30 Nov. to 2 Dec. 2010.

[40] Sanjay R. Ganorkar, Ashok A. Ghatol, "Iris Recognition: An Emerging Biometric Technology", In Proc. of the 6th WSEAS International Conference on Signal Processing, Robotics and Automation, Greece, pp.91 – 96, Feb. 2007.

[41] J. L. Wayman, A.K. Jain, D. Maltoni, and D. Maio, Eds., "Biometric Systems: Technology, Design and Performance Evaluation", New York: Springer Verlag, 2005.

[42] J. Daugman, "The importance of being random: statistical principles of iris recognition", Journal of Pattern Recognition, Elsevier, Volume 36, No. 2, pp. 279–291, Feb. 2003.

[43] Amol D. Rahulkar and Raghunath S. Holambe, "Half-Iris Feature Extraction and Recognition Using a New Class of Bi-orthogonal Triplet Half-Band Filter Bank and Flexible k-out-of-n: A Post-classifier," IEEE Trans. On Information Forensics and Security, Vol. 7, No.1, Feb 2012.

[44] Anil K. Jain, Patrick Flynn, And Arun A. Ross, Handbook of biometrics. USA: Springer, 2008.

[45] S. Hocquet, J. Ramel, H. Cardot, “Fusion of Methods for Keystroke Dynamic Authentication”, In Proc. of 4th IEEE Workshop on Automatic Identification Advanced Technologies, USA, pp. 224 – 229, Oct. 2005.

[46] F. Monrose and A. Rubin, “Authentication via keystroke dynamics”, In Proc. of 4th ACM Conference on Computer and Communications Security, Switzerland, pp. 48–56, Apr. 1997.

[47] Yilin Li, Baochang Zhang, Yao Cao, Sanqiang Zhao, YongshengGao, Jianzhuang Liu, “Study on the BeiHang Keystroke Dynamics Database,” IEEE, 2011.

[48] J. L. Wayman, “Fundamentals of Biometric Authentication Technologies”, International Journal of Image and Graphics, World Scientific Publication, Volume 1, No. 1, pp. 93-113, Jan. 2001.

[49] A.K. Jain, A. Ross, and S. Prabhakar, “An introduction to biometric recognition,” IEEE Trans. Circuits Syst. Video Technology, Special Issue Image- and Video-Based Biomet., Volume 14, Issue 1, pp. 4–20, Jan. 2004.

[50] S. Furui, "Recent Advances in Speaker Recognition", In Proc. of First International Conference on Audio and Video based Biometric Person Authentication, UK, pp. 859-872, Mar. 1997.

[51] Petricia Melin, Jerica Urias, Daniel Solano, Miguel Soto, Miguel Lopez, and Oscar Castillo, “Voice Recognition with Neural Networks, Type-2 Fuzzy Logic and Genetic Algorithms,” Advance online publication, Aug. 2006.

[52] B. Kar, B. Kartik, P.K Dutta,. “Speech and Face Biometric for Person Authentication”, In Proc. of IEEE International Conference on Industrial Technology, India, pp. 391 – 396, Dec.2006.

[53] A. Eriksson and P. Wretling, “How flexible is the human voice? A case study of mimicry,” In Proc. Of European Conference on Speech Technology, Rhodes, Greece, pp. 1043–1046, Sep. 1997

[54] S. Furui, "Recent Advances in Speaker Recognition", In Proc. of First International Conference on Audio and Video based Biometric Person Authentication, UK, pp. 859-872, Mar. 1997.

[55] Banshidhar Majhi, Y Santhosh Reddy, D Prasanna Babu, “Novel Features for Off-line SignatureVerification” International Journal of Computers, Communications & Control, Vol. I, No. 1, pp. 17-24, 2006

[56] R. Plamondon and S.N. Srihari, "Online and Offline Handwriting Recognition: A Comprehensive Survey",IEEE Tran. on Pattern Analysis and Machine Intelligence, vol.22 no.1, pp.63-84, Jan.2000.

[57] J.J. Brault and R. Plamondon, "Segmanting Handwritten Signatures at Their Perceptually Important Points", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.15, no. 9, pp.953-957, Sept.1993.

[58] B. Fang, C.H. Leung, Y.Y. Tang, K.W. Tse, P.C.K. Kwok and Y.K. Wong, "Off-line signature verification by the tracking of feature and stroke positions", Patter Recognition 36, pp. 91–101, 2003.

[59] S.Madabusi, V.Srinivas, S.Bhaskaran,M. Balasubramanian, “On-line and off-line signature verification using relative slope algorithm”, IMS , pp. 11-15, 2005.

[60] V.Nguyen, M.Blumenstein, V. Muthukkumarasamy G. Leedham, “Off-line Signature Verification Using Enhanced Modified Direction Features in Conjunction with Neural Classifiers and Support Vector Machines”, ICDAR, pp. 734 – 738, 2007.

[61] F.A.Fernandez, J. Fierrez, M. M.Diaz, J.O.Garcia, “Fusion of Static Image and Dynamic Information for Signature Verification”, ICIP , pp. 2725-2728, 2009.

[62] S. Chen and S. Srihari, “Use of Exterior Contour and Shape Features in Off-line Signature Verification”, ICDAR, pp. 1280-1284, 2005

[63] Bence Kovari, Gergely Kiss, and Istvan Albert , “Stroke Matching for Off-line Signature Verification based on Bounding Rectangles”, 23rd International Symposium on Computer and Information Sciences, pp. 1-5, 2008.

[64] Katsuhiko Ueda, “Investigation of Off-line Japanese Signature Verification Using a Pattern Matching’, Proc. of the 7th’ ICDAR-2003.

[65] D. Impedovo, G. Pirlo, Automatic signature verification: the state of the art, IEEE Transactions on Systems, Man, and Cybernetics—Part C 38 (5) 609–635, 2008.

[66] J. Ruiz-Shulcloper and W.G. Kropatsch (Eds.) "Offline Signature Verification Using Local Interest Points and Descriptors" CIARP 2008, LNCS 5197, pp. 22–29, 2008. © Springer-Verlag Berlin Heidelberg 2008..

[67] Nemcek, W.F., & Lin, W.C. Experimental investigation of Automatic Signature Verification. IEEE Transactions on Systems, Man and Cybernetics, 4, 121-126, 1974

[68] R. Bajaj, & S. Chaudhury, Signature verification system using multiple neural classifiers. Pattern Recognition, 30(1), 1-7, 1997.

[69] el-Yacoubi, A. Justino, E.J.R., Sabourin, R. & Bortolozzi, F. Off-line signature verification using HMMS and cross-validation. In Proceedings of Ninth IEEE Workshop on Neural Networks for Signal Processing (pp. 859-868), 2000.

[70] V.S. Nalwa. Automatic on-line signature verification. Proceedings of the IEEE, 85(2):215-239, 1997.

[71] H. Sakoe and S. Chiba, Dynamic programming algorithm optimization for spoken word recognition, IEEE Trans. on Acoustics, Speech and Signal Processing 26 (1), 43-48, 1978.

[72] Jayadevan R, Shaila Subbaraman and P M Patil, Verification of Hand Printed Signature Images Using Discrete Dyadic Wavelet Transform, Proceedings of IEEE Second International Conference on Industrial and Information Systems, Srilanka, 341-345, 2007.

[73] R. Plamondon and G. Lorette. Automatic signature verification and writer identification - the state of the art. Pattern Recognition, 22(2):107-131, 1989.

[74] F. Leclerc and R. Plamondon. Automatic signature verification: the state of the art - 1989-1993. International Journal of Pattern Recognition and Artificial Intelligence, 8(3):643-660, 1994.

[75] A. Jain, R. Bolle, and S. Pankanti. Biometrics: Personal Identification in Networked Society. Kluwer Academic Publishers, Boston, MA, USA, 1999.

[76] A.K. Jain, F.D. Griess, and S.D. Connell. On-line signature verification. Pattern Recognition, 35(12):2963-2972, 2002.

[77] H. Baltzakis and N. Papamarkos, A new signature verification technique based on a two-stage neural network classifier, Engineering Application of Artificial Intelligence, N±14, PP.95-103, 2001

[78] V. Nguyen and M. Blumenstein and G. Leedham, “Global Features for the Off-Line Signature Verification Problem”. 10th International Conference on Document Analysis and Recognition, 2009

[79] A.C. Ramachandra, J.S. Rao, K.B. Raja, K.R. Venugopla, L.M. Patnaik. “Robust Offline Signature Verification Based On Global Features », 2009 IEEE International Advance Computing Conference (IACC 2009), Patiala, India, 6-7 March 2009

[80] K. Huang and H. Yan, “Off-line signature verification Based On Geometric Feature Extraction and Neural Network classification”, Pattern Recognition, Elsevier Science, Vol. 30, No. 1, pp. 9-17,1997.

[81] A. Gilperez and F. Alonso-Fernandez and S. Pecharroman and J. Fierrez and Javier Ortega-Garcia, Off-line Signature Verification Using Contour Features

[82] F. Vargas, M. Ferrer, C. Travieso and J. Alonso, “Off-line Handwritten Signature GPDS-960 Corpus”, Ninth ICDAR 2007, Vol. 2, pp. 764-768, Sept. 2007.

[83] V. S. Babu and P. Viswanath, “Rough-fuzzy weighted k-nearest leader classifier for large data sets,” *Pattern Recognition*, vol. 42, pp. 1719–1731, 2009.

[84] D. Wettschereck, T.G. Dietterich. “An Experimental Comparison of the Nearest Neighbor and Nearesthyperrectangle Algorithms,” Machine Learning, 9: 5-28, 1995.

[85] V. Svetnik, A. Liaw, C.Tong, Variable selection in random forest with application to quantitative structure-activity relationship. In Proceedings of the 7th Course on Ensemble Methods for Learning Machines; Intrator, N., Masulli, F., Eds.; Springer-Verlag: Berlin, 2004.

[86] B. V. Dasarathy, Nearest Neighbor (NN) Norms, NN Pattern Classification Techniques. IEEE Computer Society Press, 1990.

[87] R. A. Fisher, “The Precision Of Discriminant Functions. Annals of Human Genetics”,10: 422–429. doi: 10.1111/j.1469-1809.1940.tb02264.x, 1940.