Authenticating Mobile Users Using Keystroke Dynamics

A Dissertation Report Submitted

By

Kawkab-un-nissa

(11506406)

To

Department of Computer Science & Engineering

In partial fulfillment of the Requirement for the

Award of the Degree of

Master of Technology in Information Technology

Under the guidance of

Mr. Baljit Singh Saini

(May 2017)

TOPIC APPROVAL PERFORMA

School of Computer Science and Engineering

Program : P173::M.Tech. (Information Technology) [Full Time]

COURSE CODE : INT546 **REGULAR/BACKLOG :** Regular **GROUP NUMBER :** CSERGD0266

Supervisor Name : Baljit Singh Saini **UID :** 15359 **Designation :** Assistant Professor

Qualification : _____ **Research Experience :** _____

SR.NO.	NAME OF STUDENT	REGISTRATION NO	BATCH	SECTION	CONTACT NUMBER
1	Kawkab Un Nissa	11506406	2015	K1520	9697012088

SPECIALIZATION AREA : System Programming **Supervisor Signature:** _____

PROPOSED TOPIC : Authenticating mobile users using keystroke Dynamics

Qualitative Assessment of Proposed Topic by PAC		
Sr.No.	**Parameter**	**Rating (out of 10)**
1	Project Novelty: Potential of the project to create new knowledge	7.40
2	Project Feasibility: Project can be timely carried out in-house with low-cost and available resources in the University by the students.	7.60
3	Project Academic Inputs: Project topic is relevant and makes extensive use of academic inputs in UG program and serves as a culminating effort for core study area of the degree program.	7.20
4	Project Supervision: Project supervisor's is technically competent to guide students, resolve any issues, and impart necessary skills.	8.00
5	Social Applicability: Project work intends to solve a practical problem.	7.20
6	Future Scope: Project has potential to become basis of future research work, publication or patent.	7.40

PAC Committee Members		
PAC Member 1 Name: Prateek Agrawal	UID: 13714	Recommended (Y/N): Yes
PAC Member 2 Name: Pushpendra Kumar Pateriya	UID: 14623	Recommended (Y/N): Yes
PAC Member 3 Name: Deepak Prashar	UID: 13897	Recommended (Y/N): Yes
PAC Member 4 Name: Kewal Krishan	UID: 11179	Recommended (Y/N): Yes
PAC Member 5 Name: Anupinder Singh	UID: 19385	Recommended (Y/N): NA
DAA Nominee Name: Kanwar Preet Singh	UID: 15367	Recommended (Y/N): Yes

Final Topic Approved by PAC: **Authenticating mobile users using Keystroke Dynamics(What is Dynamics. Clarity required in topic)**

Overall Remarks: Approved (with major changes)

PAC CHAIRPERSON Name: 11011::Dr. Rajeev Sobti **Approval Date:** 26 Oct 2016

ABSTRACT

Computer is very useful and powerful asset that has tremendously improved the way we live. We completely depend on computing systems in order to process and store confidential and sensitive data, leading to increased data exposure and consequently, to attacks. So with the increase in the dependence upon computers and computer networks authentication has become quite important. Authentication is a process of identification of users and is one of the five pillars of Information Assurance (IA). Best way to authenticate users is by the use of biometric technologies. Biometrics is the science that deals with the study and practice of authenticating the individuals based on their behavioral traits or physiological attributes. One of the biometric techniques used for user authentication is Keystroke dynamics. Keystroke dynamic come under the behavioral characteristics of a person, and is used to recognize users by their typing mannerism and rhythm. In this report we discuss how a new dataset of 40 subjects was created. What features were extracted from the data and how the feature extraction was carried out. We also present the concept of authenticating mobile phone users under different postures: sitting and walking using single profile of a particular user instead of using each profile for each position. This report also reflects the comparison between error rates achieved under different postures in order to gain insight which position is best suited and reliable for the authentication of mobile phone users. Apart from this, we also compared the results and analyzed the performance of two different classifiers while using some distinguished keystroke timing features under different situations. It was found that Random forest algorithm provide the best results.

Keywords: Authentication, Biometrics, Keystroke dynamics, Physiological, Behavioral, Information Assurance, FAR, FRR.

DECLARATION

I hereby declare that the research work reported in the dissertation entitled **"AUTHENTICATING MOBILE USERS USING KEYSTROKE DYNAMICS"** in partial fulfillment of the requirement for the award of Degree for Master of Technology in Information Technology and Engineering at Lovely Professional University, Phagwara, Punjab is an authentic work carried out under supervision of my research supervisor Mr. Baljit Singh Saini. I have not submitted this work elsewhere for any degree or diploma.

I understand that the work presented herewith is in direct compliance with Lovely Professional University's Policy on plagiarism, intellectual property rights, and highest standards of moral and ethical conduct. Therefore, to the best of my knowledge, the content of this dissertation represents authentic and honest research effort conducted, in its entirety, by me. I am fully responsible for the contents of my dissertation work.

.

Date:

Name: Kawkab-un-nissa

Registration No: 11506406

CERTIFICATE

This is to certify that the work reported in the M.Tech Dissertation entitled "**AUTHENTICATING MOBILE USERS USING KEYSTROKE DYNAMICS"**, submitted by **Kawkab-un-nissa** at **Lovely Professional University, Phagwara, India** is a bonafide record of her original work carried out under my supervision. This work has not been submitted elsewhere for any other degree.

Signature of Supervisor

Date: **Name:**

Counter Signed by:

1) **Concerned HOD:**
   HoD's Signature: _____
   HoD Name: _____
   Date: _____
2) **Neutral Examiners:**

   External Examiner

   Signature: _____
   Name: _____
   Affiliation: _____
   Date: _____

   Internal Examiner

   Signature: _____
   Name: _____
   Date: _____

ACKNOWLEDGEMENT

I owe a debt of deepest gratitude to my thesis supervisor, **Mr. Baljit Singh Saini**, Department of Computer Science And Engineering, for his guidance, support, motivation and encouragement throughout the period this work was carried out. His readiness for consultation at all times, educative comments, concern and assistance even with practical things have been invaluable.

I would like to thank the Almighty and my parents who were always there as a source of support throughout the period of this work. I would also like to thank the people who helped us to complete this work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

The increasing use of computers that have marked ubiquitous influence in modern society have made our lives facile to a significant extent, while making us reliant on computers and computer networks. This significantly improved network services but it also introduced new risks and threats to computer system thus jeopardizing the security of these systems. New challenges abound with the passing time. Like with the wide dissipation of digital identities, security issues became more conspicuous. In order to overcome these issues more refined methods were deployed for user authentication. Authentication is a process to verify whether claimed physical identities of people and computers digital identity are valid [4]. Different authentication techniques provide varied levels of security and none of these technologies provide complete security to a system. Users can be authenticated by one of the following authentication policies [4]:

- Knowledge based: In knowledge-based authentication users provide something known to them like password or Personal Identification Number (PIN). The problem faced by this method is that hacker can crack any password and would appear to be authorized user.
- Token: The token or object based authentication depends on something that is under the ownership of user, for example, smart card or a PIN provided for identification. Such authentication methods are not so convenient because smart cards or tokens are susceptible to theft and users may also forget their PIN or make some typographical errors
- Biometric authentication depends on something that a user is (e.g. fingerprints, face recognition, iris scan etc.). Here users provide their physical attribute. Biometrics is considered to be a strong alternative as it can't be borrowed, stolen, or forgotten. Thus this method is becoming globally acceptable. Biometric characteristics can be either behavioral or physiological. Physiological characteristics are the physical parameters of a certain body part. Examples are, iris scanning, fingerprints, face recognition, retina scanning etc. Behavioral characteristics are associated with the mannerisms of a person. Such as signature recognition, keystroke dynamics etc.

The problem associated with the use of knowledge based or Token based authentication mechanisms is that we don't have any idea whether the user is legitimate user or not but we can just verify whether the user possesses correct information or not. Biometrics has arisen as a promising authentication system. One of the most useful user authentication methods of biometrics is *keystroke dynamics* .Our emphasis is on this technique. Keystroke dynamics being a behavioral characteristic, endeavours to achieve user authentication by monitoring and analyzing their typing pattern through the keyboard. It takes into consideration keystroke duration, keystroke latency, force of keystrokes and other such attributes while users are typing. This method greatly enhances the computer security. Keystroke dynamics can be performed either using static text or dynamic text. A static text analysis is limited to the fixed expressions while as dynamic text can analyze any free text. The former requires less effort to implement while as latter gives the better performance.

1.1. BIOMETRICS

Biometric authentication is a method that uses either the behavioral characteristics of a user such as handwriting, voice, signature, keystroke dynamics etc. or unique physiological attributes like palm, face, iris etc. to identify or authenticate a particular user. Biometrics is an excellent way to verify identity of a user because unlike passwords or keys, biometric attributes cannot be overheard, stolen or lost thus it is a foolproof way for determining someone's identity. Physiological biometrics are the attributes that are intrinsic or by default present in users or naturally grown traits and behavioral biometrics are mannerisms or characteristics that one has acquired or learned. Biometrics, the behavioral characteristics and physical traits, make each of us quite distinctive and is natural and justified choice to verify identity of a user. Using physiological characteristics i.e., static, such as fingerprints, is a genuine way for identifying a user because no two people on the earth can have completely same physiological characteristics. Behavioral traits that are the non-static biometric attributes reflect person's mannerisms. Behavioral attributes include the amplitude and pitch in our voice, our way of signing, and even the way we type. Biometric systems recognize a living person by encompassing both physiological and behavioral characteristics. Since biometric technologies provide ample security as compared to traditional methods during authentication, they are increasingly gaining popularity. Also attention is tremendously shifting towards the use of such biometric techniques

for identification that encompasses or includes both physiological and behavioral traits of a person in order to have more reliable identification methods. In the future, the biometrics will not only eradicate the usage of passwords, PINs and ID cards. Rather, they will provide quite higher security levels, accountability and more trustworthy identification than passwords, PINs and ID cards, especially in case where security is of prime importance.

Biometric systems are divided into two phases: First is the Enrollment phase and second is the authentication or verification phase. During the first phase i,e. the enrollment phase biometric data is collected from the user, processed and then stored in a database as a reference file or template as shown in Fig.1. This template is then used for subsequent user authentication operations by the system. Next is the authentication or verification phase. In this phase the biometric data that is acquired from the user is processed and the authentication decision is based on whether the outcome of second phase matches to the already stored reference templates or not[4].

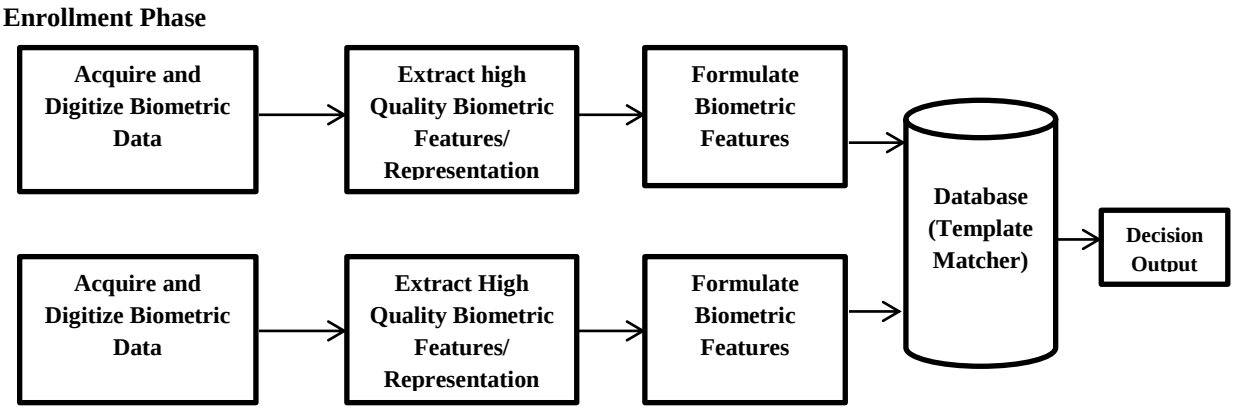


Fig. 1. Biometric System [4].

During Performance evaluation the basic types of errors that are encountered are as follows:

(i) False Acceptance Rate (FAR): It gives us the proportion of imposters or invalid users erroneously accepted as legitimate or genuine users.

(ii) False Rejection Rate (FRR): It measures the percentage of authorized users incorrectly denied and categorized as imposters.

(iii) Equal Error Rate (ERR): It is the error value obtained when both FRR and FAR values are equivalent to each other. It is also known as Cross Error Rate (CER).Performance is inversely proportional to the ERR i,e. lower the ERR , better the performance.

Depending upon the sensitivity of algorithms the FAR and FRR vary according to the graph shown in Fig. 2. Increase in one rate causes decrease in another.

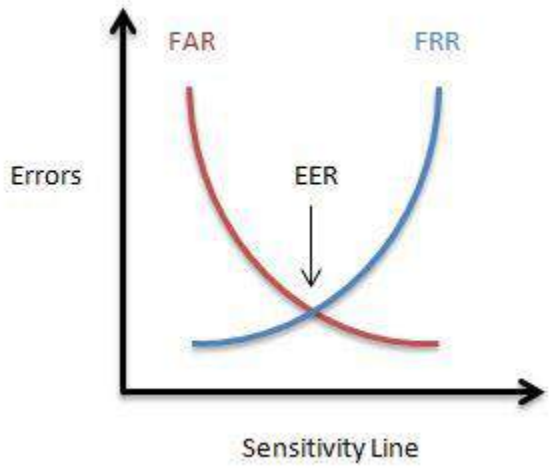


Fig. 2. FAR, FRR and ERR [3].

1.2. KEYSTROKE DYNAMICS

Keystroke dynamics is one of the biometric techniques established on behavioral measurement that focuses on verifying the identity of users based on the characteristics such as key hold time or duration of a keystroke, inter-keystroke times i,e. latency of keystrokes, force of keystrokes, typing error etc or the way a particular user types The typing way and characteristics of user are analyzed by the keystroke dynamics from the keyboard. Thus models, representing the typing rhythm and mannerisms of the user are defined. Then, these models are used for user recognition, in such a way that if rhythm of typing of a typist or user is different from the reference template already stored, such users are classified as intruders every typist has unique typing patterns. The keystroke dynamics is obviously advantageous in computer environment as it increasingly enhances computer security.

Keystroke dynamics can be classified into two types – structured or static text and free or dynamic text [3]. Static analysis analyzes an individual's keystroke behavior at certain points in the system on predetermined phrases. For example, when a user logs in to a system, his/her typing pattern is analyzed when he/she types the password and user-id. The use of a particular phrase that is common for all the users of the system can also be involved. In systems where

there is no scope for further text entry, static text entry can be deployed. For example, when a user logs in order to check his bank accounts online, usually there is no further scope of text entry while as Dynamic analysis involves periodic or constant monitoring of keystroke behavior. Firstly, it is checked when a user logs in the system and continues thereafter. For example, if a person is surfing the net, certain websites maybe visited frequently by the user. A list of the frequently occurring websites and the user's typing behavior while entering the string can be stored. Here, a training phase would be required where the user types a particular string several times in order to build a model for that string. When user types, the string along with its timing information is recorded which can then be used for authentication. However, due to its intrusive nature, dynamic monitoring may lead to privacy issues. Marsters [45] proposed a solution in his thesis, that is, only the quadgraphs were collected by him and then stored the data in a matrix instead of an ordered log. The recovery of the keystroke log is discouraged by this method thereby improving the privacy of the data

Keystroke dynamics involve two distinctive processes: feature extraction and classification of extracted features [3]. In the first process features are extracted or acquired from the user for authentication purposes. These features are extracted in order to represent user behavior in keystroke dynamics and in the second process the extracted features are classified using various algorithms such as neural networks, machine learning algorithms etc. to determine whether the extracted features match the reference template of user or not. Based on this user is either granted or denied access to the particular system.

1.2.1. FEATURES

Keystroke dynamics encompasses various approaches to authenticate the user. But before discussing these approaches let us take on the features that are extracted from raw typing data. Computer has the ability to record the moment or the point at which key was pressed, which key was pressed and other such features while typing. It can also record the time duration for which key was pressed and when was it released. While typing, all this information can be stored and used for feature extraction of user as shown in Fig. 3.

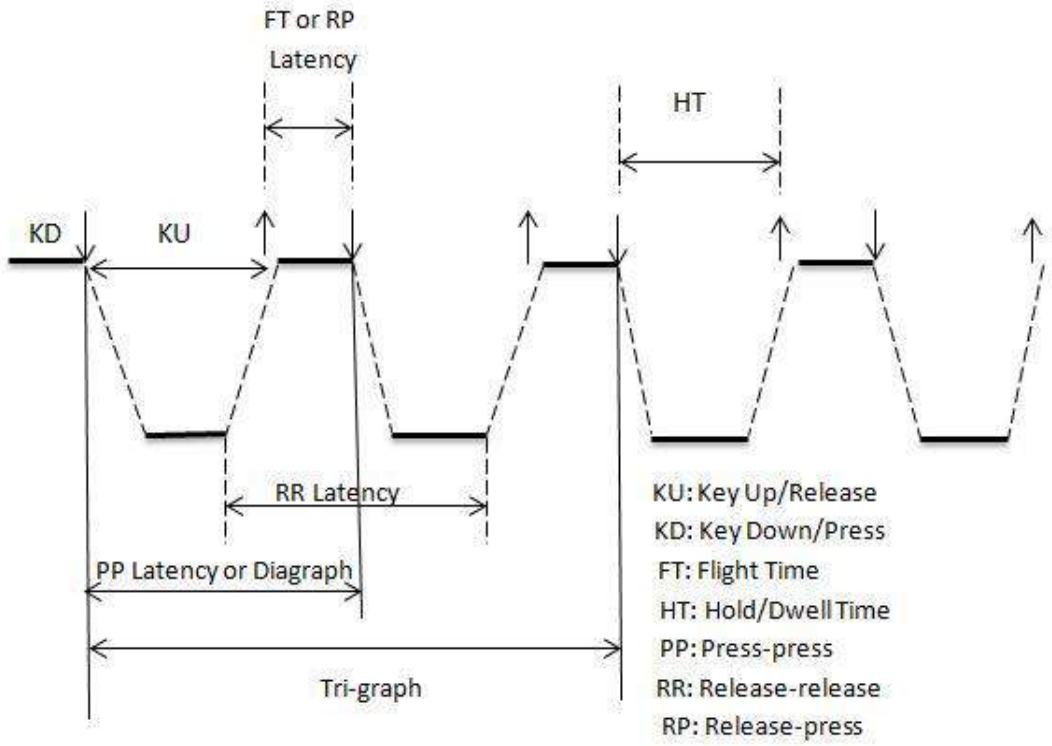


Fig. 3 Keystroke Timing Information [2].

Latency is the most frequently used feature by researchers. Latencies are of three types: release to press (RP), press to press (PP) and release to release (RR) latencies [2]. Diagraph is also considered as press to press latency by various researchers. Some researchers call release to press time as flight time. Since the system can log the time of each key press, it is easy to extract such features from the raw information. The time between the presses and releases of alternate keys is called as Trigraph. N-graph features have also been used by some researchers to determine authentic users. It is observed that trigraph features give better performance than digraphs or n-graphs. Dwell time or key hold time is the time for which each keystroke was pressed. The combination of hold and flight times is called Keystroke latency. Hold times are more important as compared to inter-key times. Moreover, the total duration taken to type a string can also be measured. Using above features, sub features can also be derived such as typing speed, force applied to a key etc. Minimum/maximum typing speed, standard deviation and mean of these features which constitute secondary features can also be derived.

1.2.2. ERROR MATRIX

In the authentication/verification phase, data from the user is obtained and processed to extract the biometric features. Then the comparison between the acquired features and the template stored in the database is performed. An algorithm is then used to check whether the former matches the latter or not and accordingly it is determined whether the user is authenticated or not. Identification is one-to-many process because a submitted sample is checked against all the biometric reference files in database [11].

The two error rates used to determine the performance of biometric authentication systems are FAR and FRR. FAR is the measurement of imposters incorrectly allowed as legitimate users

$$FAR = \frac{\text{Number of incorrect acceptances}}{\text{Total number of imposter attempts made}} * 100 \qquad \text{(i)}$$

FRR gives us the percentage of genuine users inaccurately categorized as imposters.

$$FRR = \frac{\text{Number of incorrect rejections}}{\text{Total number of genuine attempts made}} * 100 \qquad \text{(ii)}$$

EER is the error value when FAR and FRR rates are same. Lower the EER, better the performance. FAR is preferred in the systems where the security is not paramount while as FRR is preferred in the systems where security is major concern [2].

1.2.3. APPROACHES

After extracting the features and creating the reference templates, users are classified based on the similarities and dissimilarities among the templates. Researchers classified typists by using simple pattern to complex pattern recognition algorithms. In some cases, combination of these methods have been used. There are four major categories of classification algorithms, described below.

1.2.3.1 Statistical Algorithms

The simplest statistical method comprises of calculating the standard deviation and mean of the features in the template. Which can then be used for comparison using distance measures such as Euclidean distance, absolute distance and weighted absolute distance etc., hypothesis testing, t-

tests [2]. While using absolute distances only a False Rejection Rate (FRR) of 16.36% and False Acceptance Rate (FAR) of 0.25% have been achieved [13]. Vector analysis classifies users with 95% accuracy [12]. Although some researchers have presented impressive results but it should be taken into consideration that the number of samples in these experiments are not much. Since keystroke depends on subject's behavior, the features come to be non-linear in nature. So, using linear and statistical approaches may not produce vigorous results. Another pitfall of using statistical algorithms is the lack of a training stage which can be quite useful to discover the patterns in the keystroke data. . Table 1 shows the work done by researchers towards development of authentication and identification systems using statistical methods.

Table 1: Statistical methods with corresponding results.

S.NO.	Study	Method	FAR	FRR	EER
1.	Gaines et al.	Mean and standard deviation	4%	0%	-
2.	Umphress and Williams	Mean, standard deviation and diagraphs	17%	30%	-
3.	Markov	Hidden Markov Model	-	-	3.6%
4.	Francesco et al.	Degree of disorder(Timing information)	4%	-	0.01%
5.	Markov	N-graphs	0.005%	5%	-

1.2.3.2 Neural Networks

Neural networks are adaptive statistical and non-linear data modeling tools that have been inspired by the biological interconnection of neurons. The two ways in which weights can be assigned or learned are supervised learning and unsupervised learning [2]. Back propagation is one among the popular methods used in supervised learning and the popular method used in unsupervised learning is the Hopfield neural network. Various other algorithms such as Sum of Products (SOP), weightless neural networks, perceptron and Adaline have been used for classifying users depending on their keystroke dynamics. A way to classify inter-character times

by using an artificial neural network was presented by Obaidat and Macchiarolo [14]. Three different neural network architectures were tested during the investigation phase: back-propagation, sum-of-products and hybrid sum-of-products. From experiments, it was found that hybrid sum-of-products performed better than other architectures and achieved an identification rate of 97.8%. Yong et al. [15] suggested using weightless neural networks for classification of users. Data was scaled before discretizing it into linear and non-linear intervals. It was also observed that the non-linear intervals gave better results as compared to linear intervals. Table 1 lists some of the major works that has been undertaken by researchers towards developing authentication and identification systems using neural networks. Many researchers have successfully used neural networks with good results. An advantage of Neural networks is that they can handle many parameters. However, they can be slow not only during the application phase but also in training. It is difficult to decide in neural networks as to which features are important for classification. This could be a challenge in case of continuous keystroke authentication where results are typically desired in real time

Table 2: Few Neural Network approaches.

S.No.	Study	Method	FAR	FRR	ERR
1.	Angela and Sharon	Perceptron Algorithm	-	-	2%
2.	Bleha and Obaidat	Linear perceptron	9%	8%	-
3.	Cho et. Al.	Auto associative neural networks	1%	0%	-

1.2.3.3 Pattern Recognition Techniques

It is the scientific field that aims at the classification of patterns into various classes. Various pattern recognition techniques are used for keystroke dynamics classification and feature selection. Pattern recognition uses patterns and objects and then classifies them into certain categories based on different algorithms [44]. It uses simple machine learning algorithms such as the clustering and nearest neighbor algorithms to much complex algorithms such as FLD i,e. Fishers linear discriminant, data mining, Bayes classifier, SVM (support vector machine) and

graph theory .A three step approach was used by Yu and Cho [17] to improve the performance of keystroke identification. The support vector machine novelty detector achieved an average error rate of 0.81%. Giot et al. [18] proposed a method to verify the identity of users by using a support vector machine and achieved an accuracy rate of 95%. SVM is a supervised learning algorithm which reflects vigorous results for both identification and authentication. This is quite useful algorithm on basis of which future algorithms should be benchmarked. One of the advantages of using probabilistic learning algorithm is that along with the decision made they provide the confidence value. Probabilistic learning algorithms ignore outputs with low confidence value thus reducing error propagation challenge. In addition to it, unsupervised learning techniques automatically identify patterns in the data.

Table 3: Pattern Recognition Techniques

S.No.	Study	Method	FAR	FRR
1.	Obaidat	Bayes decision rule	0.8%	2.1%
2.	Obaidat	Potential function	1.7%	1.9%
3.	Nick and Bojan	Decision tree, Naïve Bayes,voted perceptron and One R	14%	1%

1.2.3.4 Search Heuristics and combination of algorithms

Search heuristics like genetic algorithms are utilized in order to find an optimum solution. They form a part of evolutionary algorithms. An example where genetic algorithm is used is that of Ant colony optimization (ACO) . They also find application is areas such as bioinformatics. The application of computational technology to molecular biology is Bioinformatics. G. Azevedo et al [19] developed keystroke feature selection by using a hybrid system that is based on stochastic optimization algorithms like Genetic algorithm, on support vector machines and particle swarm optimization. The SVM verifier that utilizes a genetic algorithm as the evolutionary algorithm for feature selection gave a minimum error of 5.18% at a FAR of 0.43% and FRR of 4.75%. While using Particle Swarm Optimization (PSO) with a global and personal acceleration of 1.5, the minimum total error was 2.21% with a FAR of 0.41% and FRR of 2.07%

used a bioinformatics approach was used by Revett et al [20] to achieve a FAR of 0.1% and FRR of 0.1%. This algorithm is capable of handling 40000 samples and provides promising results. This algorithm gives promising results for a huge number of samples and it should be considered while benchmarking future algorithms. The advantage of using genetic algorithms is that they are capable of easily handling large databases. It also gives multiple solutions and is capable of handling multi-dimensional, non-continuous, non-parametrical and non-differential problems .Sometimes one or more classification techniques have been combined and used. A combination of neuro-fuzzy algorithms like Fuzzy-ARTMAP have been utilized to classify different subjects based on their keystroke dynamics. The effect of histogram equalization of time intervals on the keystroke performance based identification algorithms was tested by Montalvao et al [21]. Four algorithms were utilized for analysis –first one on static text, second on free and static text by using algorithm proposed by Monrose and Rubin [22], third on free text by using the algorithm that was proposed by Gunetti and Picardi [23] and last one on free text using a Markov chain algorithm in which the prior probability vectors are replaced by 2D and Woodard 1D histograms. After observing all the experiments it was seen that the histogram equalization of keystroke timing data led to an enhancement in equal error rate (EER). Table 4 lists all the major work that has been undertaken by researchers towards developing authentication and identification systems by using combination of algorithms and search heuristics.

Table 4: Hybrid Techniques

S.No.	Study	Method	Performance	ERR
1.	Obaidat elt al.	Pattern Recognition and Neural Network	Moderate	-
2.	Pin et al.	Gaussian probability density function and direction similarity measure	-	9.96%
3.	Azweeda et al.	Adaptive neural fuzzy inference system	Increases system ability to learn users keystroke pattern	-

CHAPTER 2
REVIEW OF LITERATURE

2.1. HMOG: New Behavioral Biometric Features for Continuous Authentication of Smartphone Users (Zdenka Sitova, Jaroslav Sedenka, Qing Yang, Ge Peng, Gang Zhou, Paolo Gasti Kiran, S. Balagani) : This paper discusses HMOG features for user authentication in touch screen based systems. HMOG stands for Hand Movement, Orientation, and Grasp. HMOG is a set of behavioral features that is used to authenticate the users of smartphone. HMOG features capture orientation dynamics and subtle micro-movements resulting from how a user holds, grasps and taps on the smartphone. Data was collected from 100 users under two conditions: walking and sitting. Authentication Equal Error Rates i,e. EERs of as low as 7.16% (walking) and 10.05% (sitting) was achieved when combined with tap, HMOG and keystroke features. It is observed that HMOG features reflect a better performance while walking because of the ability of HMOG features to grasp distinctive and subtle body movements caused by walking besides the hand-movement dynamics caused by taps. With BKG, EERs of 15.1% was achieved while HMOG combined with taps was used. Comparatively, BKG used with key hold, tap, and swipe features had EERs between 25.7% and 34.2%. The energy consumption of HMOG computation and feature extraction was also analyzed. Analysis shows that HMOG features that were extracted at 16 hertz sensor sampling rate had an overhead of 7.9% without affecting accuracy of authentication. By combining HMOG with tap features,8.53% authentication EER during walking and 11.41% authentication ERR during sitting were achieved respectively, which is lower than the EERs achieved individually with HMOG or tap features. So it is demonstrated that HMOG is perfect for continuous authentication. There are two types of HMOG features: resistance features and stability features. Resistance features measure the micro-movements of the phone responding to the forces exerted by a tap; and stability features, give the measurement of how quickly the perturbations in orientation and movement caused by tap forces, diminish. Evaluating HMOG features On a dataset of 100 users who typed on the smartphone led to the following findings: (1) HMOG features that are extracted from gyroscope and accelerometer had better performance over HMOG features extracted from magnetometer; (2) Combining HMOG features with tap characteristics like contact size and tap duration lowered EER i,e. equal error rates from 14.73% to 8.53% for walking and14.34% to 11.41% for

sitting . This indicates that combining HMOG features with tap information considerably improves authentication performance(3) HMOG features add on to tap and keystroke dynamics features, especially for lower authentication latencies at which keystroke dynamics features and tap fare poorly. For example, for 20-second authentication latency, adding HMOG to tap and keystroke dynamics attributes declined the equal error rate from 19.11% to 15.25% for sitting and 17.93% to 11.74% for walking.

2.2. A systematic review on keystroke dynamics (Paulo Henrique Pisani, Ana Carolina Lorena): This paper focuses on keystroke dynamics that recognizes users by their typing mannerism. It provides a systematic review on keystroke dynamics, processes involved in review and the results obtained to recognize the art of keystroke dynamics. The major classifiers, extracted features, performance measures and benchmark datasets used in this area are also discussed in this paper.

a) **Systematic review**: Technique to perform review or survey on Intrusion Detection with the help of keystroke dynamics is referred to as systematic review. It may also be called as a method to conduct bibliographic review in a quite formal manner, following distinctive steps. Systematic review consists of three phases. Planning is the first phase and it defines the review protocol where research questions along with search procedures are specified. In the second phase, that is, Conduction phase, review protocol is executed and information is obtained from the references. At last, in the 3rd phase, that is, the presentation phase the final results are presented.

b) **Extracted Features**: features such as the time at which key was pressed, which key was pressed, when was the same key released etc. are extracted from raw typing data and are used as input to the classification algorithm. Different notations are used to represent the different extracted features.

   Another feature, the pressure, exerted by the mobile user on smartphone touch screen was also acquired and evaluated. The results showed that when pressure was not considered, error rates decreased from 12.2% to 6.9%.

c) **Classification algorithms**: In order to classify the users, various algorithms are used in keystroke dynamics. Usage of numeric keypad was also analyzed. Bright side of numeric keypad utilization is that the implementation of keystroke dynamics is easier to implement in cell phones. The experiment was conducted by using eight character long

password, resulting in 3.6% EER. AAMLP(Auto-associative Multilayer Percetron) and SVM(Support Vector Machine) novelty detectors were tested. Error rates for both of them were same. But for efficient resource usage one-class SVM proved to be better.

2.3. A review on the public benchmark databases for static keystroke dynamics (Romain Giot , Bernadette Dorizzi , Christophe Rosenberger): This paper focuses on the datasets that are used in keystroke dynamics validation systems. There are few public datasets for keystroke dynamics. Usually Researchers use their datasets which frequently suffer from lack of sessions and users. Few keystroke dynamics databases are available publically but none among them provide different login and password for different users. In some datasets users have typed "GREYC-laboratory" during different sessions on two different keyboards on the same computer.100 users provided 60 samples on 5 different sessions spaced of one week. Although this database contains number of users but it lacks in number of sessions and samples to track variability through time. In another dataset that is also available publically, during several sessions users typed the password ".tie5Roanl" on a single computer. 400 samples were provided by 51 users each on 8 different sessions spaced of one day. This database contains a large number of samples, but the time interval is quite small in order to track variability on a long period. These are the only two databases containing large numbers of users and enough samples to provide significant result statistically. Unfortunately, they have not been used by the community but only by their creators. Apart from these two datasets that contain large number of users and samples there are six other datasets that are openly available. Although the two large databases are interesting, but they do not exactly fit for the requirements of realistic studies, which are:

1) There should be different login and password per user because users actually use different logins and passwords in Keystroke Dynamics.

2) In order to grow the variability of the samples it is better to have different computers and keyboards.

3) Web browsers must be used to capture the data because sample collection from different browsers allow to track more variability.

2.4. Biometric Authentication and Identification using Keystroke Dynamics: A Survey (Salil P. Banerjee, Damon L. Woodard): This survey encompasses the research carried out in the area of keystroke dynamics over past three decades. Emphasis in this paper is on the

following areas: Data acquisition, text entry and different approaches in keystroke dynamics. In data acquisition, Spillane was the first to suggest the usage of keyboards in order to measure individuals keystroke dynamics for authentication. During enrollment phase users key pressure and typing pattern is stored along with password .Then the user may make an entry in the system by typing password in his/her own style This information is then compared by the system to the reference template stored already, accordingly user will be recognized.

Two ways to perform keystroke dynamics are: static text or dynamic text. Former is the structured one and its analyses is limited to the fixed expressions while as dynamic text can analyze any free text. Also the former requires less efforts to implement while as latter gives the better performance. This comes under text entry. This paper notes latency as: Release-to-Press (RP), Release-to-Release (RR) and Press-to-Press (PP) latencies. Diagraph is also considered as press to press latency by various researchers. Some researchers call release to press time as flight time. In this paper the classification algorithms used to identify users are categorized into following: Statistical algorithms, Pattern recognition algorithms, neural networks and their combination can be also used. All these approaches provide different FAR and FRR rates.

2.5. Biometric personal authentication using keystroke dynamics: A review (M. Karnan, M. Akila, N. Krishnaraj): The objective of this paper is to brief about the approaches used in keystroke dynamics over last two decades but the center of attention are the classification methods. Classification methods tend to find out the optimal or near optimal patterns. A reference template per user is maintained in the database and the keystroke dynamics captured during login are compared to this predetermined reference to determine whether a user is valid or not. Few classification methods are: statistical approaches, neural network methods, pattern recognition techniques and hybrid techniques. Standard statistical methods use mean and standard deviation to record keystrokes and authenticate the user with 4% FAR(False Acceptance Rate) and 0% IPR(Imposter Pass Rate) for seven users. When mean, standard deviation and diagraph were combinedly used in statistical method, result of 30% FRR and 17% FAR was obtained. Usage of timing information reduced the variation effects resulting in 0.01% IPR and 4% FAR. Neural networks have the capability to parallely explore computing hypothesis. In the area of biometrics neural networks is considered to be of great potential because of the above quality. Parallel learning and testing was made possible by proposing a neural network system that can be placed at each and every processor. There are many other

neural network approaches to identify users. Next is the Techniques of Pattern Recognition. Recognition of patterns is the scientific field whose goal is the classification of patterns into various classes. Number of pattern recognition techniques are used for keystroke dynamics classification and feature selection. Pattern recognition uses patterns and objects and then classifies them into certain categories based on different algorithms like Bayes classifier, minimum distance classifier, Fishers Linear Discriminate (FLD) and many more. Hybrid techniques also come under classification methods. Hybrid techniques combine various neural networks, statistical measures and techniques of pattern recognition.

A set of techniques for password authentication using neural networks, fuzzy approach, statistical approaches and distinctive hybrid combination methods are used and reflect moderate and different performance.

2.6. An Investigation on Touch Biometrics: Behavioral Factors on Screen Size, Physical Context and Application Context (Tao Feng, Xi Zhao, Nick DeSalvo, Tzu-Hua Liu, Zhimin Gao, Xi Wang and Weidong Shi): With the increase in privacy issues and security concerns within mobile devices. behavioral biometric solutions are introduced like touch based user recognition. This paper focuses on Contextual behavior factors (*i.e.,* size of screen and physical context) concept in order to identify users in touch based systems. Moreover, on uncontrolled touch data user recognition method was employed for comparison purpose. The user identity performance was measured with and without this new concept. According to the results obtained it is found that screen size affects user interaction with the particular device and physical activity context plus application context enhance the touch based user recognitions accuracy. In addition to it, other interesting result is that way user holds the device change the screen size of a smartphone device. A larger screen size may result in potential user recognition accuracy.

2.7. Investigating the Discriminative Power of Keystroke Sound (Joseph Roth, Xiaoming Liu, Arun Ross, Dimitris Metaxas): The purpose of this paper is to discover and deduce whether or not keystroke sound can be used to identify a user. The strength of keystroke sound was analyzed with regards to user validation applications. Influenced by the digraph concept used in keystroke dynamics, here from the keystroke sound a virtual alphabet is learned and then digraph latency within virtual letter pairs combined with other statistical features, is used to measure percentage of match. The final scores indicate the similarities and dissimilarities between different sound streams and are used to make final identification decision. Evaluation based on both static text

and free text authentications was carried out based on a database of 50 subjects reflecting both advantages and pitfalls of keystroke sound.

2.8. Biometric Recognition Based on Free-Text Keystroke Dynamics (Ahmed A. Ahmed and Issa Traore): Keystroke dynamics can be performed either using static text or free-text. Due to the scattered and dispersed nature of data, recognition accuracy of latter approach is quite challenging. So, this research presented new approach for analysis of free-text keystroke dynamics. This approach integrated analysis of digraphs and monographs as well as it speculated the missing digraphs depending upon monitored keystroke relations between them, using neural network. This method provides much more accuracy level compared to other best obtained results. This method also reduces the processing time. In heterogeneous environment, 0.0152% of FAR and 4.82% FRR was achieved respectively. Also 2.46% EER was achieved after the evaluation on the dataset of 53 users was performed. Comparatively homogeneous environment yielded FAR, FRR and EER of 0%, 5.01% and 2.13% respectively with 17 users.

2.9. Keystroke Active Authentications Based on Most Frequently Used Words (Alaa Darabseh, Akbar Siami Namin): This research contributed in making the user recognition techniques using keystroke dynamics more advanced. In this paper the effect and performance of distinctive keystroke features in keystroke dynamics identification systems is analyzed. More emphasis is on the usage of commonly used English words in keystroke dynamics for authenticating the users. Performance analyses of four features: flight time, key duration, latency,and diagraph time latency are analyzed. We can also say that this paper presented the outcome of an experiment to analyze the effect of above four keystroke features. It is observed that best performance is achieved by using the diagraph time, followed by the flight time. So this paper provides the logic behind using latency time between two keystrokes in most researches of keystroke dynamics. This result was obtained on the dataset of 28 users. Future work in this research would be the performance evaluation by combining one or more of the above discussed features in keystroke dynamics.

2.10. Keystroke dynamics in password authentication enhancement(Pin Shen Teh, Andrew Beng Jin Teoh, Connie Tee, Thian Song Ong): Computer is very useful and powerful asset that has tremendously improved the way we live. With the increase in the dependence upon computers and computer networks authentication has become quite important. Authentication is a process of identification of users. Best way to authenticate users is by the use of biometric

technologies. Biometrics is the science that deals with the study and practice of authenticating the individuals based on their behavioral or physiological attributes. One of the biometric techniques used for user authentication is Keystroke dynamics. Keystroke dynamic come under the behavioral characteristics of a person and is used to recognize users by their typing mannerism and rhythm. Contribution of this study is framing an approach that resolves the probability of gradual changes in typing pattern of users. This approach is called as Retraining and this method accommodates latest user typing patterns by continuously updating the reference file or template stored in the database. This work also introduced a new component known as Alternative Authorization Mechanism (AAM). This component is quite useful because it not only acts as a backup for user authentication options but also handles the issues that come forward when a legitimate user is not able to produce his/her frequent typing mannerism because of some untoward factor.

2.11. Supervised classification methods applied to Keystroke Dynamics through Mobile Devices(Ignacio de Mendizabal-Vazquez, Daniel de Santos-Sierra, Javier Guerra-Casanova, Carmen Sanchez-Avila): This paper focuses on relative study of various supervised classification methods in order to use them for biometric system construction, using the already present information in the mobile phones. Some distinguished features like typing speed, finger size, pressure, time, angular and linear acceleration are extracted and processed while the users type the 4 digit PIN. This paper also presents analysis of database with keystroke dynamic patterns in which data was collected in controlled environment in which users where directed to hold phone in a fixed position, with the other movements recorded in an uncontrained environment. Using Linear Discriminant Analysis(LDA) and Principal Component Analysis(PCA), first data manipulation was performed. Pre-processing of acquired data was done using LDA and reduction of information and resource utilization was achieved by using PCA. Since computation power of mobile phones is low, reduced data size should be used .Here data size was considerably reduced using PCA and resulted in easy implementation of supervised classification methods in smartphones. Excellent classification rates were achieved on combination of PCA and multi-layer perceptron classifier. 80% users were correctly identified while using 3 samples per user, this rate increased to 90% with the use of 9 samples per user. Despite of the less samples, Euclidean classifiers were considered to be good in providing

satisfying performance. ERR of 20% was achieved. The best performance under this situation was obtained when samples were directly acquired from smartphone.

2.12 Putting 'Pressure' on Mobile Authentication (Sougata Sen, Kartik Muralidharan):

This paper presents the authentication scheme for mobile phone users. In this scheme the users are made to enter the passcodes. Apart from passcode, the users behaviour of entering passcode is also acquired i,e. features like pressure applied on the screen by a user and the duration of screen press were also captured for authentication purposes. An android application was developed to capture a 4 digit passcode from users. Various classifiers such as j48, Naïve Bayes, K* classifier and multi-layer perceptron were then applied on the collected data in order to evaluate classification accuracy. It was observed that using MLP classifier, highest accuracy result was achieved with FAR(14.06%) and FRR(14.1%).

2.13 Mobile Authentication using Keystroke Dynamics (Sudhir Dhage, Pranav Kund, Anish Kanchan, Pratiksha Kap): This paper presents an authentication method by using the concept of fusion methods. Data was collected from users using a Sony Xperia M smartphone. Features extracted from user inputs are key hold time and key latencies/digraphs(Press-Press time, Release-Press time and Release-release time). Then the statistical method like mean and standard deviation were applied on users input. Two factors used in this research are hits factor and another is deviation ratio. Former is computed by setting a breaking point of qualities for which the elements separated must fall inside. Deviation in terms of mean is measured to calculate Deviation Ratio. For Deviation Ratio's fusion with the first factor i,e. hits factor, its result is multiplied with some constant value. In order to authenticate users threshold for both these factors is set. Feature fusion by the combination of two or more factors significantly improves error rates and efficiency of authentication mechanism for smartphone users. The use of fusion factor led to the improved ERR of 0.806.

2.14 Application of keystroke analysis to mobile text messaging (Nathan Clarke, Steven Furnell & Benn Lines): Mobile users are authenticated on the basis of their behaviour of typing the text messages. Features acquired from the user input were: hold time (the time for which a particular key was pressed) and keystroke latency(duration of time between two consecutive keystroke). In the first phase of this research, accuracy of user authentication was measured upon numeric data entry such as PIN and telephone numbers. Many neural network and pattern recognition algorithms were contrasted and feed-forward multi-layered perceptron neural

network proved to be most efficient resulting in EER of 10.4% and 11.3% with PIN code and telephone number respectively. With the more emphasis on second phase, the users were authenticated by their typing pattern of the text messages. Since there are 26 alpha-numeric characters, 600 diagraph pairs exist, making it more suitable for user authentication. Various algorithms such as FF-MLP, Gradual Training and Early Stopping classification algorithms were used. There results were also compared and it was observed that FF-MLP is the most effective one and provides ERR of 8.8% with text message authentication.

2.15 Touch Gestures Based Biometric Authentication Scheme for Touchscreen Mobile Phones (Yuxin Meng, Duncan S. Wong, Roman Schlegel, and Lam-for Kwok): This paper focuses on authentication mechanism related to touch dynamics, that is, behavioral touch dynamic features such as multi touch, single touch, touch movement were used to verify mobile users identity. With an average of almost 6 sessions per user, data was acquired from 20 android users. Weka was used for evaluation of users input data and for the calculation of FAR and FRR rates per user. Many algorithms were applied over the data acquired from users and it was observed that neural network classifier along with the above mentioned features provided the most satisfactory results. Compared to j48, Naive Bayes and Kstar classifiers which gave average error rates of 23.72%, 20.41% and 15.4% respectively, the two neural network classifiers: Radial Basis Function Network (RBFN) and Back Propagation Neural Network (BPNN) resulted in the significantly efficient performance with average error rates as 7.71% and 11.58% respectively.

2.16 Keystroke Dynamics for Mobile Phones: A Survey (Baljit Singh Saini, Navdeep Kaur and Kamaljit Singh Bhatia): It is a survey paper that encompasses the research carried out in the field of keystroke dynamics for mobile phones. According to this survey, the most prominent keystroke dynamics feature used in various researches is the latency (Press-press, Release-release and Release-press latencies). Other features extracted from the users while typing include Hold time, N-graphs, size of screen that the user touches while typing, pressure etc. Hold time and pressure are the features that marked quite improvement when used combinendly with the latency. When the features, available on android phones such as gyroscope and accelerometer are used, 0.08% of FAR is achieved. Performance evaluation of keystroke dynamics is also enlightened in this paper. The reference template is compared with the features extracted from the particular user at the time of login attempt using various matching algorithms. FAR, FRR and

ERR are the three error rates that are used to measure performance. Lower the error rates, better the performance. This survey also analyzed the various classification methods involved in user authentication. These are: statistical methods, neural network techniques, pattern recognition techniques and hybrid methods. Most used classification methods are neural network and statistical methods. However, it is difficult to judge the better one among them because of different features and testing conditions. In statistical methods, mean and standard deviation achieved best results, 6.9% of EER. Neural network algorithms are slow but they provide quite positive results. Number of Pattern Recognition techniques have been proposed and used. Algorithm k-Nearest Neighbor provides EER of 1% , when applied on features such as inter-key time, hold time and pressure. Also Random Forest achieved accuracy of 93.04%. The other observations made are: combination of pressure and different time features provide more accurate results as compared to their separate analysis, People who do not use mobile phones regularly, keystroke dynamics is not a good option for them. Length of input is quite important. Long input means more accuracy. Another important observation is that the sensor based features surpass the traditional features in accuracy. Sensor based features and traditional keystroke dynamic features achieve EER of 0.08% and 4.97% respectively.

2.17. Keystroke dynamics in password authentication enhancement (Pin Shen Teh, Andrew Beng Jin Teoh, Connie Tee, Thian Song Ong): This paper highlights a technique to strengthen password authentication system by using chunk of keystroke dynamic information under fusion framework. Four latencies were used as keystroke features and also two methods were incorporated to find out the similarity index of the two used latencies. Fusion approach with two layers have been proposed in order to improve the system performance. Also an additional module is presented to improve proposed systems flexibility. This module helps to resolve the problems associated with users gradual typing changes, if any. This module captures the users latest typing pattern by simply updating the reference template. This is named as Restraining approach. Another contribution of this paper is an alternative authorization mechanism component. This component is handy whenever a legitimate user is unable to provide his/her actual typing pattern in case of any injury or any other untoward case. At such time this component will act as backup user authentication. Authors used the static text for the analysis purpose. Total of 10 samples were collected from each user. Among 10 samples, 7 samples were used for training and 3 for testing. The 7 samples of training set constitute user template. In order

to perform FAR test, each users ist testing sample was matched against all the other users' keystroke template and the same procedure was repeated for all other users of the data set resulting in 29,700 fraud attempts. Similarly, for FRR test, 3 testing samples of each user were matched with the reference template of the corresponding user. This resulted into 300 legitimate attempts. The EER of 4.34%, 4.27% and 3.73% was achieved from Sum rule, Produce rule and weighted sum rule respectively. EER of 10.75% was achieved with Max rule.

2.18. Mobile User Identification through Authentication using Keystroke Dynamics and Accelerometer Biometrics (Kyle R. Corpus, Ralph Joseph DL. Gonzales, Alvin Scott Morada): This paper enlightens the process of using keystroke dynamics with the accelerometer biometrics. Keystroke dynamics deals with the typing rhythm of a person while as accelerometer reflects how a particular person holds his/her mobile phone. Authors used homemade dataset of 30 users. These users were made to type a particular password of 8-16 characters 8 times using customized tool incorporated in their mobile phones. Among the 8, first 6 samples constituted training set and the remaining 2 samples were kept aside for test set. Then a tool written in java was used to process and acquire accelerometer biometric features and keystroke dynamic features. With the help of RapidMiner data mining tool some well-known classifiers were trained to develop models. These include Neural Networks, j48, Naïve Bayes and Decision Tree and for model validation, 10-fold cross-validation was used. Then these different classifiers were applied on keystroke dynamic features alone, then on accelerometer biometrics alone and finally on the combination of both. It was found that Neural network classifier provided the best results when applied on the combination of both features. Results were further made efficient by the use of filters. Removal of highly correlated and low ranking features added greatly to the efficiency. It was found that the combination of keystroke dynamics and accelerometer features improved performance from 49.44% to 61.11%. It can be said that accelerometer biometrics play quite important role in user authentication via biometrics. FAR of 7.0% was achieved and FAR of 40% was achieved. The lower rate of FAR reflects that this particular model is good at blocking the illegitimate users access. But the results also reflect that this model can't be used to identify mobile users accurately but can be used as a contributing factor to traditional password based user authentication system by setting some threshold.

2.19. User Authentication Through Typing Biometrics Features (Livia C. F. Araujo, Luiz H. R. Sucupira Jr. , Miguel G. Lizárraga, Lee L. Ling, and Joao B. T. Yabu-Uti): This paper discusses the use of static keystroke for user authentication purpose. Inputs used are the Key-up time, key-down time and the key ASCII code i.e., the ASCII code of a particular key is also captured while the user is typing the string. Total of four features were used, which are: ASCII code of a key, key duration and the latencies of two keystrokes. These features were then analyzed and seven experiments were carried out on the different combinations of these features. Evaluation of results was carried out with three different user types: the imposter, the observer imposter and the genuine user. Several experiments were carried out and the best result was obtained from basedon distance statistical classifier with the incorporation of all the four features i,e., ASCII code of particular key, Down-Down time, Up-Down time , and Down-Up time. FAR of 1.89% and FRR of 1.45% was obtained. The methodology provided by this paper enhances the efficiency of password based user authentication when password is no more secret. These rates achieved are quite competitive with the previous ones that only used one string and 10 samples during the enrollment phase. Also using the four distinctive features added to the efficiency, since most of the previous researchers used one or two features only. This paper also reflects the impact of several real practical aspects that were observed and tested. It was found that these factors influence the performance of a user. These different aspects observed are: how much a user is familiar with the target string, timing accuracy, sample number in the enrollment phase, adaptation mechanism and the two-trial authentication.

2.20. Keystroke Template Update with Adapted Thresholds (Abir Mhenni, Christophe Rosenberger and Estelle Cherrier, Najoua Essoukri Ben Amara): There are still many open and demanding areas of research for biometric keystroke recognition. With time the relevance of the acquired features to the personal typing rhythm is becoming less representative. This can cause rise in failure rate of biometric authentication. Due to the changing nature of such features, periodic updation of the representative model is needed. This paper discusses the use of sliding and growing window as methods to update the template which in turn are based on statistical classifier. This paper also demonstrates that the error rates are reduced if the user specific thresholds are used as compared to the usage of fixed thresholds. User-specific thresholds vary from one update session to another. To study this, the statistical recognition method was used. This method is considered to be quick and efficient as compared to other methods. Standard

deviation and Mean of the training samples are used to represent the reference in this particular method. In order to update the decision criteria double thresholding mechanism is used. First is the verification threshold and other one is the update threshold. Former is used to accept or reject new query while as latter is used to determine whether new query can be used to update the reference template or not. This mechanism excludes imposter samples in reference biometric. Authors presented the method for updating the template via variable thresholds. Variation of thresholds from one session to another considerably reduces the update error rates. Compared to use of individual or single threshold, the performance gets better over time. Authors validated this mechanism on two datasets and this approach provides more accurate results than the classical ones (EER of 2% less is obtained).

2.21. Improving Performance and Usability in Mobile Keystroke Dynamic Biometric Authentication (Faisal Alshanketi, Issa Traore, Ahmed Awad E. A): Mobile phones have become the most important part of our lives. So far the protection mechanism used on these devices is the pattern or Personal Identification Number (PIN) but these mechanisms are weak form of authentication. Keystroke dynamics is one of variations of authentication mechanism and is considered to be quite strong and secure. This paper shows that Random forest classifier provides the improved results and performance. Authors also designed their own algorithm in order to handle the typos, which is quite important for improving usability. The authors studied pressure-based features as well as timing features and the combination of both. Evaluation was performed on total of 3 datasets, two publically available and one homemade dataset. Authors used Random Forest in order to categorize the legitimate users and find out the imposters. The 10-fold cross validation was used with 10% dataset used for testing and 90% being used for training. Authors in their own evaluation process used under-sampling and cost sensitive learning. To accomplish the cost sensitive training imposter category was assigned a weight corresponding to the ratio of total legitimate samples and total imposter samples. The best results with EER of 5.8% was achieved when the dataset 1 with 51 users was evaluated with the weight set equal to 0.000515 and when cost sensitive learning was applied on it. Authors then applied their own algorithm to dataset 2. Best results with EER of 2.3% was obtained when the combination of both timing features (i.e., flight and dwell times) and pressure features (finger area and pressure) were used for 42 subjects.

2.22. Keystroke dynamics for authentication in smartphone (Jong-hyuk Roh, Sung-Hun Lee , Soohyung Kim): In this paper, authors collected the data from 15 subjects. The users were made to type the PIN 7-6-6-4-2-0 in three different postures: while walking, while holding the mobile phone with comfort and while putting the mobile phone on table. The features that were extracted from the users data are: Key-stamp, size of the tab, acceleration, coordination and gyroscope. Key-stamp includes Keydown-keydown time and Keyup-keyup time, leading to total of 16 key-stamp features. The Accelerometer sensor is used to measure the force applied by the user on the mobile phone. It is used to identify the movement along all the three axis: x, y and z. Four different ways are used to capture this feature set. Gyroscope sensor is used to calculate the rotational velocity along the three axis: roll, pitch and yaw axis. Like accelerometer features, gyroscopic features are also captured in four different ways. Evaluation and experimentation was performed on various feature combinations. Since each feature is further divided into different features, the total number of feature combinations was calculated to be 103. To find the accuracy rate classifiers used were Euclidean and Manhattan distance methods. Also pre-processing with standardization and scaling reflected quite improved results. The order of best results among different postures was calculated to be: walking, holding mobile phone and putting phone on table with EER of 6.39%, 7.35% and 10.81% respectively.

2.23. Authentication System using Behavioral Biometrics through Keystroke Dynamics (D. D. Alves, G. Cruz Jr and C. Vinhal): This paper presents the work for improving the user authentication via keystroke dynamics. Authors proposed the algorithm that observes the users typing rhythm in real time, capture the keystroke times at which the key was released and pressed. This proposed process is non-intrusive and low cost. Total of five features were acquired from the user data: ASCII code, PS, SP, SS and PP. The duration of time for which key is pressed is named PS. SP is the time during which user doesn't press any key or in other words it is the split time between keys. SS is the time duration between two consecutive keys release and PP is the time interval between pressing of two consecutive keys. This data was used to trace the metrics that in turn was used to identify a particular user. The proposed algorithm results in proficient accuracy when applied on equal size strings or the strings that have more than five characters. First only the statistical methods were applied on the data. The statistical classifiers also provided the good results. However, optimization of the proposed method using differential evolution provided FAR and ERR rates below 4%. This method was applied on a dataset of 150

subjects with distinctive typing experiences via some software created by authors. Integrating the target string and all the 5 extracted features provided the best results as compared to other combinations. Choosing the target string is very important task because it makes users well versed with the password and thus making acceptance of imposters in the system less. But the limitation while choosing password is that it should be at least 5 characters long to reduce FAR. The training set should consist of large number of samples so that standard deviation and mean will be more authentic and reliable regarding typing characteristics. Smaller string length leads to compulsion of larger number of samples in order to compose this model.

2.24. User Authentication with Keystroke Dynamics in Long-Text Data (Hayreddin Çeker and Shambhu Upadhyaya): In traditional authentication mechanisms users provide their login credentials at a single point of entry. But in some cases where session remains active for a longer period of time systems need to verify whether the user who is still at terminal is the one who was originally authenticated or not. So there is a trade-off between usability and security and this can be tackled via behavioral biometrics based on active authentication. This paper discusses the use and importance of active authentication paradigm. Active authentication directs that the system must validate and recognize the user in continuous manner. Since Support Vector Machine is quite handy tool to fulfill such criteria to classify and analyze users working in the background. Authors also extended usability of Support Vector Machine for constant authentication of long text data instead of using short text to authenticate users at single time only. Keystroke Dynamics is also one of the efficient behavioral biometric through which users can be authenticated. This method can be used to validate users in the background while they are working actively at the terminal. This paper reflects use of SVM as classifying tool for active authentication because of its systematic processing and sharp recognition rate. By using the timing features such as dwell time and flight time of commonly used diagraphs over 34 subjects who were registered for experiment with RBF kernel in one-class SVM, were all distinguished correctly. The final results show that all imposters were denied the access and all legitimate users were validated correctly i,e., EER of almost 0% was achieved. This was possible by setting one-class SVM for each user using a dataset collected in controlled environment. Results also reflect that by correctly setting the kernel scale and by standardization of the input, one-class SVM can be used to authenticate users on constant basis and also identify keystroke dynamics with significantly high accuracy.

2.25. Authentication through Keystrokes: What You Type and How You Type (Md. Asraful Haque, Namra Zia Khan, Gulnar Khatoon): This paper encompasses the development of authentication mechanism that integrates secure and simple authentication method called keystroke dynamics with the conventional username/password method. Authors aim to validate the users based on integration of user's pattern corresponding to their typing mannerism and text password. For this reason statistical approach was deployed. Mean and standard deviation are computed in Statistical Method. In statistical method there are number of distance measures and algorithms that can be used for keystroke dynamics. The policy proposed in this paper consists of two phases in order to authenticate the user: registration phase and login phase. In registration phase, the user is directed to enter his/her login credentials along with the password. After this the training session of user starts in which he/she is asked to type password 10 times in order to gain understanding with his/her typing pattern. During this time features like flight time, dwell time and total time are captured. These features are stored for correct password only. In case of typing error user is directed to enter the password again. Next is the login phase. In this phase apart from checking the correctness of the username and password provided by the user, typing verification also takes place. Since every user has distinct typing characteristics, this phase authenticates the user by checking his/her typing characteristics along with correct login credentials. Then comes the comparison and decision. During this process users test sample is compared with the reference samples. Then the difference between them is calculated. Standard deviation for each of the acquired timing feature is also calculated. Based on the standard deviation calculated and the difference determined, final score is derived. Finally, final score is compared with the threshold value in order to accept or reject a particular user. It is concluded that this mechanism is quite simple as it is based on statistical method. This method reflects very interesting results with 93% of accuracy rate. This approach will combat various attacks that the traditional password based authentication mechanism fail to vanquish.

CHAPTER 3
PRESENT WORK

3.1 PROBLEM FORMULATION

Keystroke dynamics provide a completely transparent solution to the users. Keystroke dynamics can be used to authenticate mobile phone users as well as desktop/laptop users. Here we put more emphasis on mobile phone user authentication. De Luca et al [24] proposed a solution to authenticate smartphone users by proposing a mechanism that captured users way to draw a pattern as input and the pressure applied while drawing pattern. S. Sen and K. Muralidharan[26] authenticated the mobile users by making users to type the passcodes and simultaneously capturing their behavior while entering passcodes. Features monitored were time for which screen was pressed, pressure applied while typing a passcode, Inter key time and key hold time, which was not possible using patterns only .This mechanism resulted to be more usable and secure method of authentication. Zdenka Sitova et al [1] used Hand Movement, Orientation, and Grasp as a set of behavioral features to authenticate the users of smartphone under two different conditions: walking and sitting. They used different datasets to authenticate users in different positions. This increased the overhead because of maintaining different profiles for different positions of same user. Also the numbers of samples captured from the different subjects are not quite adequate to encompass their typing proficiency and mannerisms completely. To this end, we put forward a different scheme to authenticate mobile phone users. In this scheme we tend to undertake following steps:

i. Keystroke data will be collected and analyzed from various mobile phone users under two different conditions: walking and sitting. Along with this we will use four distinguished features for user authentication: Dwell time or hold time, press-press time, Release-release time and Release-press time/ flight time.

ii. No additional hardware is needed for this scheme. All that we need to do is to install a software application to record keystroke events, in a mobile phone that will act as data collection apparatus.

iii. Afterwards we will model a mechanism that will accept valid users under both conditions i,e. walking and sitting, using a single dataset.

iv. Data will be collected in an uncontrolled environment i,e. we will collect data from uncongested and un-crowded/spacious area. Also the touch dynamic features such as single touch, touch movement and multi-touch would not be considered.

v. In this analysis we would not consider the mood or state of mind of a user while taking inputs from them, making this future scope of a proposed work.

3.2 OBJECTIVES OF THE STUDY

The objectives of this study are as follows:

1. Analysis of typing pattern of users while sitting and walking while taking into account four distinguished keystroke dynamic features such as Dwell time or hold time, press-press time, Release-release time and Release-press time/ flight time using two widely used algorithms: Random Forest and Naïve Bayes.

2. Analyze the difference in typing pattern of users while sitting and walking using the above mentioned features and algorithms

3. Create a single Model to authenticate users in two different postures: walking and sitting, thereby decreasing the overhead.

3.3 RESEARCH METHODOLOGY

The research will be carried out in the following steps:

1. Creation of new dataset with 40 subjects.
2. Creation of single profile for users in different postures and extraction of different features in these different positions
3. Installation of Weka and study the concepts
4. Apply different algorithms on the user data.
5. Find out the difference between FAR, FRR and EER rates of users in different positions and also find out the best suited position for user verification/authentication.
6. Find three different error rates and compare them with existing results

3.3.1 Tool to be used:

Weka: Weka stands for Waikato Environment for Knowledge Analysis. It is an open source software and acts as a workbench that comprises of various algorithms and visualization tools for predictive modeling and analysis of data. It has Graphical User Interface (GUI) in order to make access to the provided functions simple and easy for users. Weka consists of set of machine

learning algorithms that are used to handle various data mining tasks. These different algorithms can be directly applied on the available datasets or can be called from the Java code. Weka contains collection of tools for data classification, clustering, data pre-processing, association rules, regression and data visualization. Weka is also considered to be well-suited for development of new machine learning schemes.

3.3.2 Algorithm steps:

Step 1: Data collection: In the first and foremost step of the work put forward, we employed 40 users for data collection. New dataset of 40 users was created. Data collection was carried out in 5 sessions. Each user was directed to type a strong password "**tie5Raonl**" 80 times per session, 20 times in each posture: walking and sitting and in two different screen orientations: portrait and landscape. For each user and situation, profile was created by using training set that consists of positive samples or the samples that are of a genuine user and imposter or negative samples from other users and for model validation, 10-fold cross-validation was used The increase in number of samples give rise to the variability which in turn pushes accuracy towards the higher rates. This methodology does not deal with typographic errors. If a user makes some typographical error that sample is not considered. Data collection mechanism was carried out in uncontrolled environment. In this way authors created a single profile for a user under two different postures: sitting and walking.

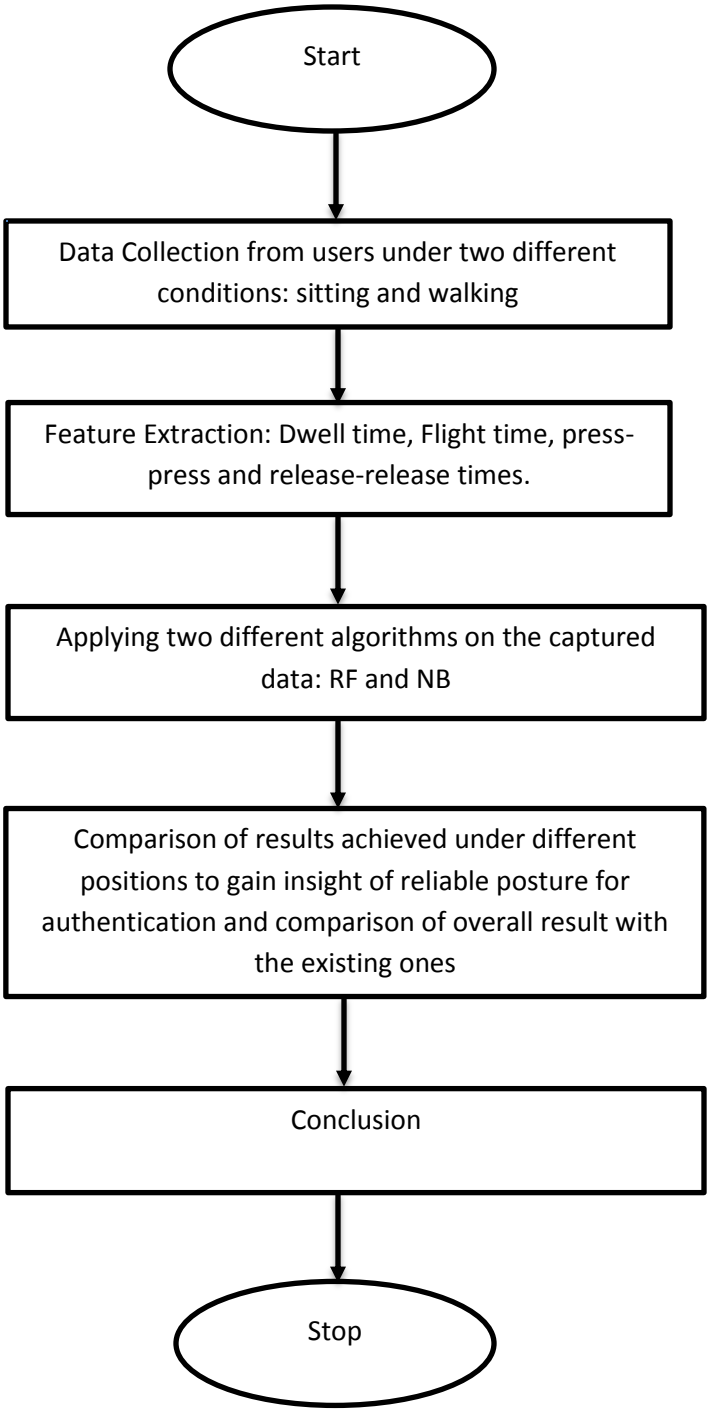
Step 2: Feature Extraction: No extra hardware was used for the data collection and feature extraction. Different smartphones with android O.S were used for the data collection to gain variability. An application was installed on the mobile phones which captured the features of data given by users via default sensors present in mobile phones. Thus the mobile phone itself acted as the data collection apparatus. The features that were acquired are: Dwell time, Flight time, press-press and release-release time. Extracted features will be then used for classification. Whenever a subject presses or releases a key, the software application will record the events such as key-down, key-up time etc. Data collected will be stored in a database as a reference template. This constitutes the enrollment phase of a biometric system. Later on users typing traits will be compared to their particular reference templates already stored in the database. Based on the matching percentage users will be either rejected or authenticated. This is the authentication phase**.**

Step 3: Applying different algorithms: After all the above procedure is carried out, the two most widely used algorithms: Random Forest and Naïve Bayes are applied on the captured user data. Random Forest, a supervised machine learning algorithm is an ensemble approach for regression and classification that work by creating number of decision trees during training time and resulting into the class that is mode of the classes yielded by individual trees. Random forest algorithm works as a comprehensive collection of decision trees that are decorrelated. On the basis of random selection of variables and data, Random forest develops lot of decision trees. Resultant trees are considered to be random trees because of their nature of random data and variable selection, leading to the construction of random forest. Among the current algorithms available, Random Forest is considered to be the unexcelled and unsurpassed algorithm in accuracy. It works quite efficiently for large datasets and also provides insight of the important variables in the classification. Another algorithm, Naïve Bayes, a family of simple probabilistic classifiers is based on Bayes theorem with the assumptions that are independent amongst features. Naive Bayes classifier is quicker in term of convergence. Thus needs less training data. Naïve Bayes gives good results even if the Naïve Bayes assumptions don't hold. These algorithms are used to evaluate the FAR, FRR and EER rates under the two different conditions: sitting and walking. These error rates are then compared in order to gain insight of the more reliable position or condition for authentication. We also applied the algorithms on data before and after outlier detection to find out the better way to achieve efficient results.

Step 4: Comparison: These error rates are also compared with the existing results to find the more reliable mechanism for the authentication.

Step 5: Conclusion: Overhead is decreased by creating a single profile of users under different postures. Also the more reliable posture for authentication is highlighted. The comparison also reflects that the combination of features that the authors undertook and capturing quite large number of samples is very handy in providing better results than most of the available mechanisms.

Data Flow Diagram:



CHAPTER 4
RESULTS AND DISCUSSIONS

4.3 RESULTS

The results are shown on the basis of three error rates: FAR, FRR and EER. Lower the error rates, better the performance. The parameters that we considered are: Dwell time, Flight time, press-press and release-release times.

The results are discussed as follows:

Table 5: Error rates while considering single model for each user using Random Forest algorithm

Keystroke timing-features	Algorithm	Posture	Screen Orientation	FAR (%)	FRR (%)	EER (%)
Dwell time, Flight time, PP and RR times	Random Forest	Both(Sitting and walking)	Both(Landscape and portrait)	8.231	3.85	4.122

Table 6: Error rates while considering single model for each user using modified Random Forest algorithm

Keystroke timing-features	Algorithm	Posture	Screen Orientation	FAR (%)	FRR (%)	EER (%)
Dwell time, Flight time, PP and RR times	Modified Random Forest	Both(Sitting and walking)	Both(Landscape and portrait)	5.554	3.016	3.868

From the results shown in above two tables, it is quite evident that modified Random forest provides far better results than the original algorithm in case of single model. Single model refers to the model that is combination of samples provided by the user in both sitting and walking postures and in both screen orientations: landscape and portrait. The original random forest

algorithm was modified by changing the seed i.e. the number used to start a pseudorandom number generator and the number of trees to be created.

Table 7: Error rates achieved by applying NB on the dataset while the users have given the samples in sitting posture under both screen orientations (landscape and portrait) without outlier removal.

Keystroke timing-features	**Algorithm**	**Posture**					
Dwell time, Flight time, PP and RR times	Naïve Bayes	**Sitting (landscape)**			**Sitting(Portrait)**		
		FAR (%)	**FRR (%)**	**EER (%)**	**FAR (%)**	**FRR (%)**	**EER (%)**
		5.775	42.843	20.624	6.951	48.443	20.072

Table 8: Error rates while applying NB on the dataset while the users have given the samples in sitting posture under both screen orientations (landscape and portrait) with outlier removal.

Keystroke timing-features	**Algorithm**	**Posture**					
Dwell time, Flight time, PP and RR times	Naïve Bayes	**Sitting (landscape)**			**Sitting(Portrait)**		
		FAR (%)	**FRR (%)**	**EER (%)**	**FAR (%)**	**FRR (%)**	**EER (%)**
		5.802	38.075	20.562	9.6	38.270	19.786

Table 9: Error rates while applying NB on the dataset while the users have given the samples in walking posture under both screen orientations (landscape and portrait) without outlier removal.

Keystroke timing-features	Algorithm	Posture					
		Walking (landscape)			Walking (Portrait)		
Dwell time, Flight time, PP and RR times	Naïve Bayes	FAR (%)	FRR (%)	EER (%)	FAR (%)	FRR (%)	EER (%)
		9.6	55.113	32.22	6.805	51.721	27.051

Table 10: Error rates while applying NB on the dataset while the users have given the samples in walking posture under both screen orientations (landscape and portrait) with outlier removal.

Keystroke timing-features	Algorithm	Posture					
		Walking (landscape)			Walking (Portrait)		
Dwell time, Flight time, PP and RR times	Naïve Bayes	FAR (%)	FRR (%)	EER (%)	FAR (%)	FRR (%)	EER (%)
		13.645	44.589	21.094	7.040	43.270	18.281

Table 11: Error rates while applying RF on the dataset while the users have given the samples while walking under both screen orientations (landscape and portrait) without outlier removal.

Keystroke timing-features	Algorithm	Posture					
Dwell time, Flight time, PP and RR times	Random Forest	Walking (landscape)			Walking(Portrait)		
		FAR (%)	FRR (%)	EER (%)	FAR (%)	FRR (%)	EER (%)
		3.497	4.718	4.684	3.232	4.740	4.489

Table 12: Error rates while applying RF on the dataset while the users have given the samples while walking under both screen orientations (landscape and portrait) with outlier removal.

Keystroke timing-features	Algorithm	Posture					
Dwell time, Flight time, PP and RR times	Random Forest	Walking (landscape)			Walking(Portrait)		
		FAR (%)	FRR (%)	EER (%)	FAR (%)	FRR (%)	EER (%)
		3.632	4.670	3.697	3.329	4.659	4.486

Table 13: Error rates while applying RF on the dataset while the users have given the samples while sitting under both screen orientations (landscape and portrait) without outlier removal.

Keystroke timing-features	**Algorithm**	**Posture**					
Dwell time, Flight time, PP and RR times	Random Forest	**Sitting (landscape)**			**Sitting (Portrait)**		
		FAR (%)	**FRR (%)**	**EER (%)**	**FAR (%)**	**FRR (%)**	**EER (%)**
		2.437	4.521	4.459	3.397	5.516	5.432

Table 14: Error rates while applying RF on the dataset while the users have given the samples while sitting under both screen orientations (landscape and portrait) with outlier removal.

Keystroke timing-features	**Algorithm**	**Posture**					
Dwell time, Flight time, PP and RR times	Random Forest	**Sitting (landscape)**			**Sitting (Portrait)**		
		FAR (%)	**FRR (%)**	**EER (%)**	**FAR (%)**	**FRR (%)**	**EER (%)**
		3.459	4.5	3.991	4.013	5.3	5.244

Results in Table 7 to Table 14 reflect that among the two algorithms used i.e. Random forest and Naïve Bayes, Random Forest proved to be the better one. Naïve Bayes resulted in very high error rates that is not acceptable in biometric authentication. The results also enlighten the fact that without removing the outliers from the user data FAR rates are quite good compared to the results achieved after removing the outliers in all the above cases whilst better FRR and EER rates are achieved after the outliers and extreme values are removed from the data.

Considering the results achieved in all the above scenarios it is quite apparent that the best and most suited posture for authentication is sitting. Best FAR rates are achieved when user provided samples while sitting and when Random forest classifier was applied on the data without removing outliers. However, best FRR and EER rates are achieved in the same posture, applying the same classifier (Random forest) but with outlier and extreme value removal.

Results via Graphs:

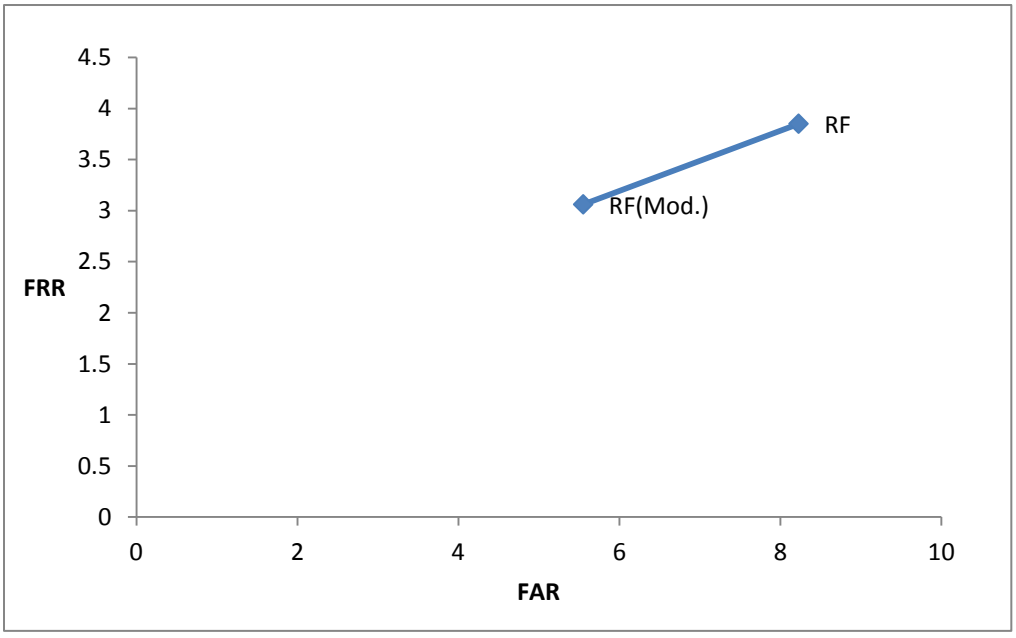


Fig. 4 Results of RF and Modified RF

The above graph shows the FAR and FRR rates of dataset while using Random Forest and Modified Random Forest. It clearly reflects that by modifying some values of original algorithm, the error rates are decreased. Thus increasing the efficiency.

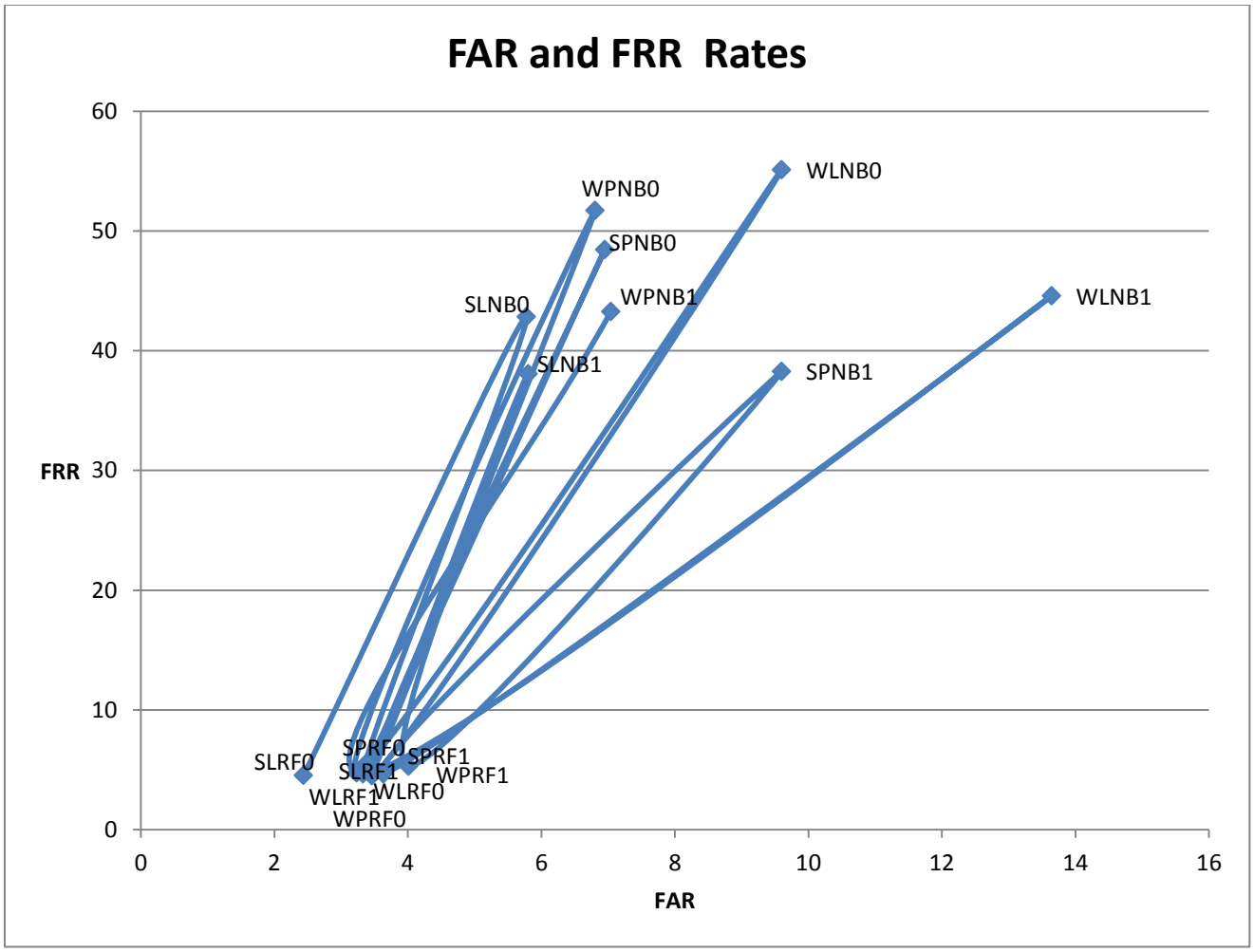


Fig. 5 Result of different algorithms in different positions with different screen orientations with and without outlier removal.

The above graph shows the FAR and FRR rates by applying two different algorithms on the data samples achieved from users under two different positions: walking and sitting with two different screen orientations: landscape and portrait with and without removing the outliers and extreme values. First and second letter of each label in the graph represent the posture and screen orientation respectively. Next two letters represent the algorithm used i.e. either Random Forest (RF) or Naive Bayes (NB). The 0's and 1's in the graph reveal whether outliers are removed or not. Value 0 means outliers are not removed while as 1 means that the outliers are removed from the data.

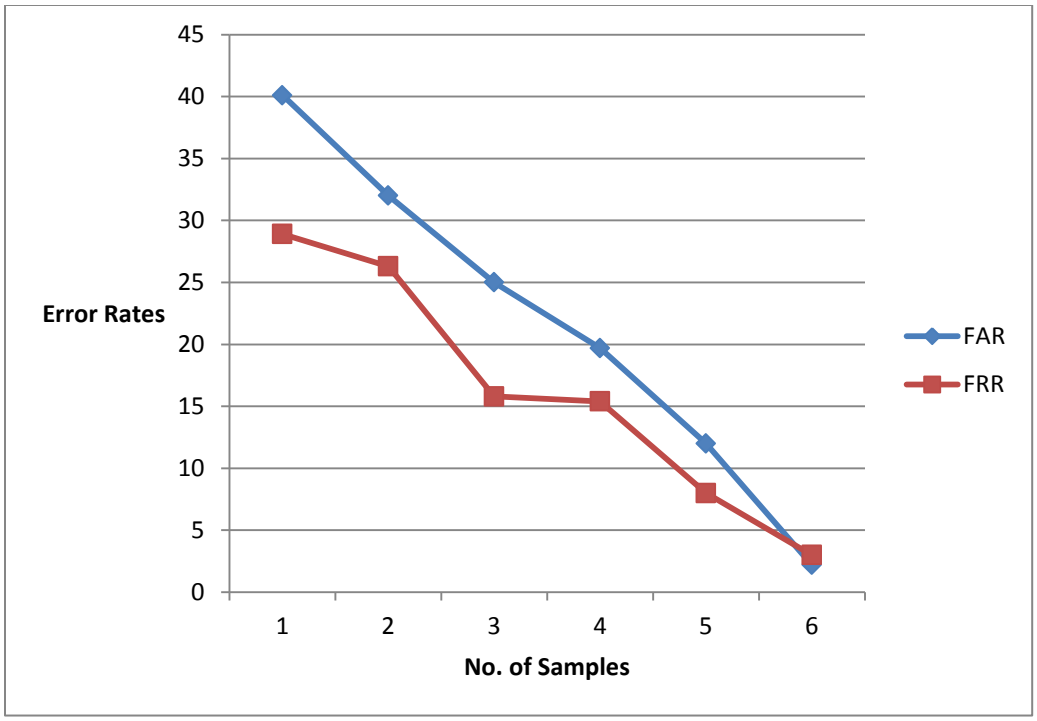


Fig. 6 Graph depicts the variation of FAR and FRR rates with the number of samples

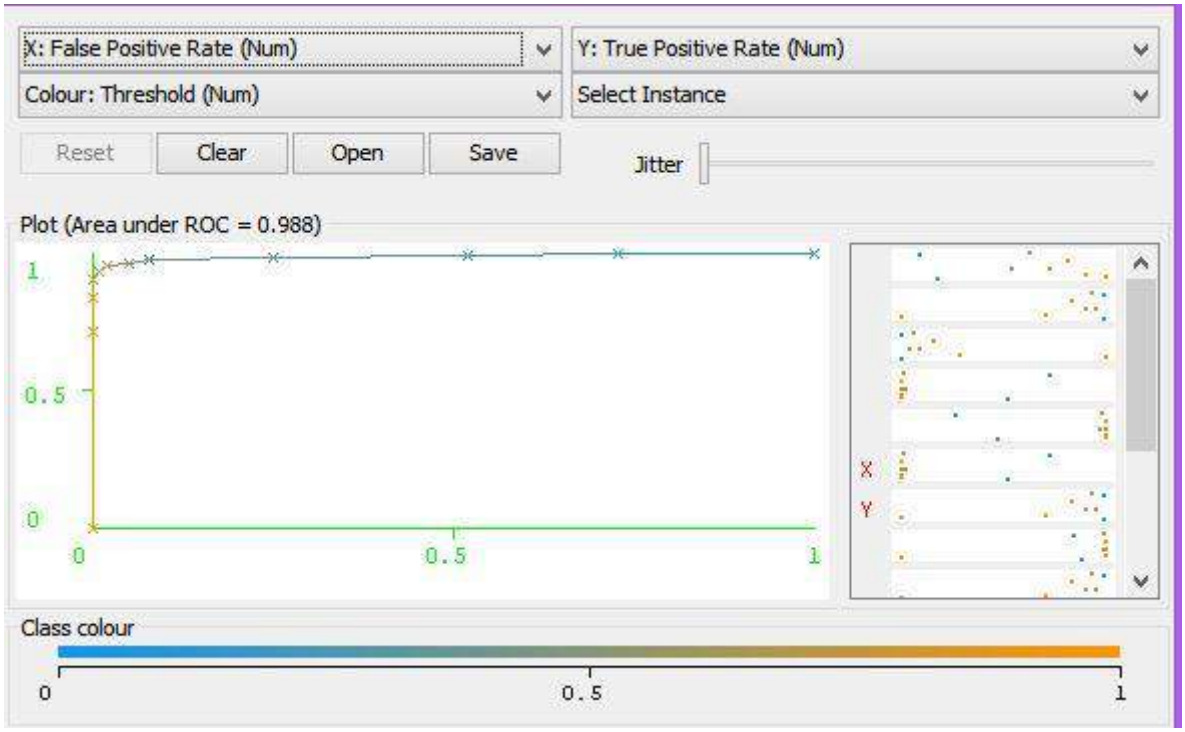


Fig. 7 ROC curve of RF

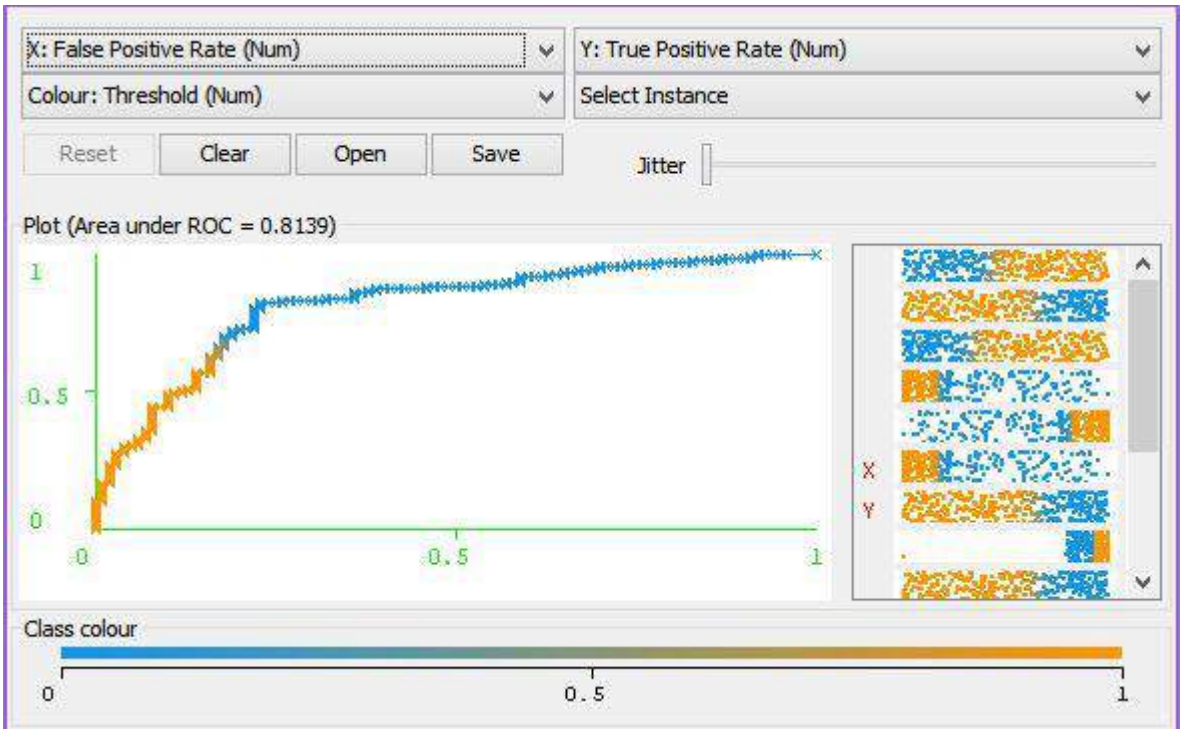


Fig. 8 ROC curve of NB

Since ROC(Receiver Operating Characteristic) curve, a visualization tool is used to efficiently find out whether the classifier that we use is appropriate or not. It in a way provides a path for us to understand which classifier is better for our work. More the ROC curve is towards the left upper side; more appropriate is the classifier for that particular work. Above two figures clearly reflect Random Forest is the better classifier than Naïve Bayes in case work undertaken by us.

4.4 COMPARISON OF RESULTS

The following table shows the comparison of existing research works with our work:

Table 25: Summary of various approaches and their results and their comparison with our work.

Paper	Author	Approach	Features	Outlier Removal	Posture and Screen Orientation	Results (%)
[1]	Z. Sitova et al.	Statistical	HMOG, tap, Hold time and swipe features	–	Sitting	EER: 10.05%

Paper	Author	Approach	Features	Outlier Removal	Posture and Screen Orientation	Results (%)
[1]	Z. Sitova et al.	Statistical	HMOG, tap, Hold time and swipe features	–	Walking	EER:7.16%
[16]	Clarke et al	Neural Network	Flight time	–	–	EER:5
[17]	Cho et al.	Neural Network	Flight time and Hold time	–	–	FAR:0, FRR:1
[37]	Douhou et al.	Statistical	Flight time, Dwell time	–	–	FAR:16, FRR:1
[38]	Antal et al.	Machine learning	Flight time, Dwell time, Pressure and Finger area	–	–	EER:12.9
[39]	Rybnik et al.	Statistical	Dwell time	–	–	EER:6.1
[40]	Grabham et al.	Statistical	Dwell time, Flight time and Pressure	–	–	FAR:15, FRR:0
[41]	Karnan et al.	Machine learning	Flight time, Dwell time	–	–	Accuracy:92.8
[42]	Trojahn et al.	Statistical	Flight time, Dwell time, Pressure and Finger area	–	–	FAR:4.19, FRR:4.59

Paper	Author	Approach	Features	Outlier Removal	Posture and Screen Orientation	Results (%)
[43]	Giuffrida et al.	_	N-graph	_	_	EER:4.97
[45]	Joyce et al.	Statistical	Flight Time	_	_	FAR:0.25, FRR:16.36
_	Our work	Random Forest	Dwell time, Flight time, press-press and release-release time	No	Both (Sitting and Walking)	FAR:8.231,FRR:3.85, EER:4.122
_	Our work	Modified Random Forest	Dwell time, Flight time, press-press and release-release time	No	Both (Sitting and Walking)	FAR:5.554,FRR:3.016, EER:3.868
_	Our work	Random Forest	Dwell time, Flight time, press-press and release-release time	No	Sitting (portrait)	FAR:3.397, FRR:5.516, EER:5.432
_	Our work	Random Forest	Dwell time, Flight time, press-press and release-release time	Yes	Sitting (portrait)	FAR:4.013, FRR:5.3, EER:5.244
_	Our work	Naïve Bayes	Dwell time, Flight time, press-press and release-release time	No	Sitting (portrait)	FAR:6.951, FRR:48.443, EER:20.072
_	Our work	Naïve Bayes	Dwell time, Flight time, press-press and release-release time	Yes	Sitting(portrait)	FAR:9.6, FRR:38.270, EER:19.786

Paper	Author	Approach	Features	Outlier Removal	Posture and Screen Orientation	Results (%)
_	Our work	Random Forest	Dwell time, Flight time, press-press and release-release time	No	Walking(landscape)	FAR:3.497, FRR:4.718, EER:4.684
_	Our work	Random Forest	Dwell time, Flight time, press-press and release-release time	Yes	Walking (landscape)	FAR:3.632, FRR:4.670, EER:3.697
_	Our work	Naïve Bayes	Dwell time, Flight time, press-press and release-release time	No	Walking (landscape)	FAR:9.6, FRR:55.113, EER:32.22
_	Our work	Naïve Bayes	Dwell time, Flight time, press-press and release-release time	Yes	Walking (landscape)	FAR:13.645, FRR:44.589, EER:21.094
_	Our work	Random Forest	Dwell time, Flight time, press-press and release-release time	No	Sitting (landscape)	FAR:2.437, FRR:4.521, EER:4.459
_	Our work	Random Forest	Dwell time, Flight time, press-press and release-release time	Yes	Sitting (landscape)	FAR:3.459, FRR:4.5, EER:3.991
_	Our work	Naïve Bayes	Dwell time, Flight time, press-press and release-release time	No	Sitting (landscape)	FAR:5.775, FRR:42.843, EER:20.624
_	Our work	Naïve Bayes	Dwell time, Flight time, press-press and release-release time	Yes	Sitting (landscape)	FAR:5.802, FRR:38.075, EER:20.562

Paper	Author	Approach	Features	Outlier Removal	Posture and Screen Orientation	Results (%)
_	Our work	Random Forest	Dwell time, Flight time, press-press and release-release time	No	Walking (portrait)	FAR:3.232, FRR:4.740, EER:4.489
_	Our work	Random Forest	Dwell time, Flight time, press-press and release-release time	Yes	Walking (portrait)	FAR:3.329, FRR:4.659, EER:4.486
_	Our work	Naïve Bayes	Dwell time, Flight time, press-press and release-release time	No	Walking (portrait)	FAR:6.805, FRR:51.721, EER:27.051
_	Our work	Naïve Bayes	Dwell time, Flight time, press-press and release-release time	Yes	Walking (portrait)	FAR:7.040, FRR:43.270, EER:18.281

The above comparison reveals that we achieved better results compared to the existing results for both postures: sitting and walking while applying Random forest. It also reflects that Naïve Bayes algorithm when applied with the features that we acquired under two different positions is not efficient because it results into quite high error rates that is not acceptable in keystroke dynamic authentication.

CHAPTER 5
CONCLUSION AND FUTURE SCOPE

5.1 CONCLUSION

Biometric authentication is a method of identifying or authenticating users based on their behavioral traits (handwriting, voice, signature, keystroke dynamics etc.) or physiological attributes (palm, face, iris etc.). Biometrics is an excellent way of identity verification because biometric traits cannot be overheard, stolen or lost. Keystroke dynamics, a behavioral biometric technique aims to identify the users based on their typing characteristics: key hold time or duration of a keystroke, inter-keystroke times i,e. latency of keystrokes, force of keystrokes, typing error etc. Keystroke dynamics can be used to authenticate laptop as well as mobile phone users. This report presents an overview of research on keystroke dynamics over past few decades with emphasis on how keystroke dynamics can be used for authenticating mobile phone users. Till now only one research has been carried out on authenticating the mobile phone users under two different conditions: walking and sitting but there was no provision of authenticating a user under different positions using a single dataset. Separate datasets have been created for a single user for different positions, thereby increasing overhead. To overcome this problem we proposed a method to authenticate mobile phone users under different positions by creating a single profile of a user. This avoids the overhead included in using different datasets for same user in walking and sitting postures. We also strived to achieve the insight of reliable posture for authentication and it was found that sitting is the better position for authenticating a user. We also compared the results of two different algorithms: Random forest and Naïve Bayes. Random forest proved to be better algorithm than Naïve Bayes.

5.2 FUTURE SCOPE

In our work we used the dataset of 40 users only and only acquired some of the timing features from the user data. Data collection from diverse and large number of users will increase the authentication efficiency. Also the combination of wide features like sensor based features, flight time, pressure, size, hold time and other keystroke dynamic features can be evaluated to add to the accuracy rate. Also the relaxing posture is missed in our work. We will further try to extend our work by considering relaxing posture as well. Also we look forward to combine the timing

features with those of sensor based features to add to efficiency of keystroke dynamics. However, there are still open and demanding areas of research that need to be addressed to make keystroke dynamics as productive biometrics.

CHAPTER 6
REFERENCES

[1] Z. Sitova, J. Sedenka, Q. Yang, G. Peng, G. Zhou, P. Gasti and K. Balagani, "HMOG: New Behavioral Biometric Features for Continuous Authentication of Smartphone Users", *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 5, pp. 877-892, 2016.

[2] S. Banerjee and D. Woodard, "Biometric Authentication and Identification Using Keystroke Dynamics: A Survey", *Journal of Pattern Recognition Research*, vol. 7, no. 1, pp. 116-139, 2012.

[3] P. Pisani and A. Lorena, "A systematic review on keystroke dynamics", *Journal of the Brazilian Computer Society*, vol. 19, no. 4, pp. 573-587, 2013.

[4] M. Karnan, M. Akila and N. Krishnaraj, "Biometric personal authentication using keystroke dynamics: A review", *Applied Soft Computing*, vol. 11, no. 2, pp. 1565-1573, 2011.

[5] R. Giot, B. Dorizzi and C. Rosenberger, "A review on the public benchmark databases for static keystroke dynamics", *Computers & Security*, vol. 55, pp. 46-61, 2015.

[6] T. Feng, X. Zhao, N. Desalvo, T.-H. Liu, Z. Gao, X. Wang, and W. Shi, "An investigation on touch biometrics: Behavioral factors on screen size, physical context and application context," *2015 IEEE International Symposium on Technologies for Homeland Security (HST)*, 2015

[7] J. Roth, Xiaoming Liu, A. Ross and D. Metaxas, "Investigating the Discriminative Power of Keystroke Sound", *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 2, pp. 333-345, 2015.

[8] A. Ahmed and I. Traore, "Biometric Recognition Based on Free-Text Keystroke Dynamics", *IEEE Transactions on Cybernetics*, vol. 44, no. 4, pp. 458-472, 2014.

[9] A. Darabseh and A. S. Namin, “Keystroke Active Authentications Based on Most Frequently Used Words,” *Proceedings of the 2015 ACM International Workshop on International Workshop on Security and Privacy Analytics - IWSPA '15*, 2015

[10] P. Teh, A. Teoh, C. Tee and T. Ong, "Keystroke dynamics in password authentication enhancement", *Expert Systems with Applications*, vol. 37, no. 12, pp. 8618-8627, 2010.

[11] D. Woodard and P. Flynn, "Finger surface as a biometric identifier", *Computer Vision and Image Understanding*, vol. 100, no. 3, pp. 357-384, 2005.

[12] A. Guven and I. Sogukpinar, "Understanding users' keystroke patterns for computer access security", *Computers & Security*, vol. 22, no. 8, pp. 695-706, 2003.

[13] R. Joyce and G. Gupta, "Identity authentication based on keystroke latencies", *Communications of the ACM*, vol. 33, no. 2, pp. 168-176, 1990.

[14] M. Obaidat and D. Macchiarolo, "An online neural network system for computer access security", *IEEE Transactions on Industrial Electronics*, vol. 40, no. 2, pp. 235-242, 1993.

[15] S. Yong, W. K. Lai, and G. Goghill, “Weightless Neural Networks for Typing Biometrics Authentication,” *Lecture Notes in Computer Science Knowledge-Based Intelligent Information and Engineering Systems*, pp. 284–293, 2004

[16] N. L. Clarke and S. M. Furnell. "Authenticating mobile phone users using keystroke analysis." *International journal of information security, vol.* 6, no.1, pp.1-14, 2007.

[17] E. Yu and S. Cho, "Keystroke dynamics identity verification—its problems and practical solutions", *Computers & Security*, vol. 23, no. 5, pp. 428-440, 2004.

[18] R. Giot, M. El-Abed, and C. Rosenberger, “GREYC keystroke: A benchmark for keystroke dynamics biometric systems,” *2009 IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems*, 2009

[19] G. L. F. B. G Azevedo, G. D. C. Cavalcanti, and E. C. B. C. Filho, “Hybrid Solution for the Feature Selection in Personal Identification Problems through Keystroke Dynamics,” *2007 International Joint Conference on Neural Networks*, 2007.

[20] K. Revett, “A Bioinformatics Based Approach to Behavioural Biometrics,” *2007 Frontiers in the Convergence of Bioscience and Information Technologies*, 2007

[21] J. Montalvao, C. A. S. Almeida, and E. O. Freire, “Equalization of keystroke timing histograms for improved identification performance,” *2006 International Telecommunications Symposium*, 2006.

[22] F. Monrose and A. Rubin, “Authentication via keystroke dynamics,” *Proceedings of the 4th ACM conference on Computer and communications security - CCS '97*, 1997

[23] D. Gunetti and C. Picardi, "Keystroke analysis of free text", *A32CM Transactions on Information and System Security*, vol. 8, no. 3, pp. 312-347, 2005.

[24] A. D. Luca, A. Hang, F. Brudy, C. Lindner, and H. Hussmann, “Touch me once and i know it's you!,” *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, 2012.

[25] S. Cho, C. Han, D. H. Han and H. I. Kim. "Web-based keystroke dynamics identity verification using neural network." *Journal of organizational computing and electronic commerce*, vol.10, no. 4, pp. 295-307, 2000.

[26] S. Sen, and K. Muralidharan. "Putting ‘pressure ’on mobile authentication." *Mobile Computing and Ubiquitous Networking (ICMU), 2014 Seventh International Conference on*. IEEE, 2014.

[27] B. S. Saini, N. Kaur, and K. S. Bhatia, “Keystroke Dynamics for Mobile Phones: A Survey.” *Indian Journal of Science and Technology*, vol. 9, no. 6, 2016

[28] P. S. Teh, A. B. Jin Teoh, C Tee, and T. S. Ong, "Keystroke dynamics in password authentication enhancement." *Expert Systems with Applications*, vol. 37, no. 12, 2010.

[29] K.R. Corpus, R. J. DL. Gonzales, A. S. Morada, "Mobile user identification through authentication using keystroke dynamics and accelerometer biometrics." *Proceedings of the International Workshop on Mobile Software Engineering and Systems*. ACM, pp. 11-12, 2016.

[30] L. Araujo, L. H. R. Sucupira, M. G. Lizárraga, L. L. Ling and J. B. T. Yabu-Uti, "User authentication through typing biometrics features." *IEEE transactions on signal processing, vol.* 53, no. 2, pp. 851-855, 2005.

[31] A. Mhenni, C. Rosenberger, E. Cherrier, and N. E. Ben Amara, "Keystroke template update with adapted thresholds." *Advanced Technologies for Signal and Image Processing (ATSIP) 2nd International Conference*, IEEE, pp. 483-488, 2016.

[32] F. Alshanketi, I. Traore, and A. A. Ahmed, "Improving Performance and Usability in Mobile Keystroke Dynamic Biometric Authentication." *Security and Privacy Workshops (SPW), IEEE,* pp. 66-73, 2016.

[33] J. H. Roh, S. H. Lee and S. Kim, “Keystroke dynamics for authentication in smartphone.” *Information and Communication Technology Convergence (ICTC), 2016 International Conference, IEEE,* pp. 1155-1159, 2016

[34] D. D. Alves, G. Cruz, and C. Vinhal, "Authentication system using behavioral biometrics through keystroke dynamics." *Computational Intelligence in Biometrics and Identity Management (CIBIM), IEEE Symposium*, IEEE, pp. 1-6, 2014.

[35] H. Çeker and S. Upadhyaya, "User authentication with keystroke dynamics in long-text data." *Biometrics Theory, Applications and Systems (BTAS), IEEE 8th International Conference*, pp. 1-6, IEEE, 2016.

[36] M. A. Haque, N. Z. Khan, and G. Khatoon. "Authentication through keystrokes: What you type and how you type." *Research in Computational Intelligence and Communication Networks (ICRCICN) IEEE International Conference*, pp. 251-261, IEEE, 2015.

[37] S. Douhou and J. R. Magnus. "The reliability of user authentication through keystroke dynamics." *Statistica Neerlandica*, vol. 63, no. 4, pp. 432-449, 2009

[38] M. Antal, L. Z. Szabó, and I. László. "Keystroke dynamics on android platform." *Procedia Technology* 19, pp. 820-826, 2015

[39] M. Rybnik, M. Tabedzki, M. Adamski and K. Saeed. "An exploration of keystroke dynamics authentication using non-fixed text of various length." In *Biometrics and Kansei Engineering (ICBAKE) International Conference*, IEEE, pp. 245-250, 2013.

[40] N. G. Grabham, and N. M. White. "Use of a novel keypad biometric for enhanced user identity verification." *Instrumentation and Measurement Technology Conference Proceedings, IMTC 2008, IEEE*, pp. 12-16, 2008.

[41] M. Karnan and M. Akila "Personal authentication based on keystroke dynamics using soft computing techniques." *Communication Software and Networks ICCSN'10, Second International Conference*, IEEE, pp. 334-338, 2010.

[42] M. Trojahn, F. Arndt and F. Ortmeier. "Authentication with keystroke dynamics on touchscreen keypads-effect of different n-graph combinations." *Third International Conference on Mobile Services, Resources and Users (MOBILITY).*, pp. 114-119, 2013.

[43] C. Giuffrida, K. Majdanik, M. Conti and H. Bos. "I sensed it was you: authenticating mobile users with sensor-enhanced keystroke dynamics." *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*. Springer International Publishing, pp. 92-111, 2014.

[44] S. Theodoridis, and K. Koutroumbas, "Pattern recognition," New York publications, 2003

[45] J.D. Marsters, " *Keystroke dynamics as a biometric."* Doctoral dissertation, University of Southampton, 2009.

CHAPTER 7
APPENDIX

ABBREVIATIONS

IA	Information Assurance
FAR	False Acceptance Rate
FRR	False Rejection Rate
PIN	Personal Identification Number
EER	Equal Error Rate
CER	Cross Error Rate
RP	Release-to-Press
PP	Press-to-Press
RR	Release-to-Release
HT	Hold Time
FT	Flight Time
SOP	Sum of Products
SVM	Support Vector Machine
HMOG	Hand Movement, Orientation, and Grasp
MLP	Multi-layer Perceptron
GUI	Graphical User Interface
FF-MLP	Feed Forward-Multi-layer Perceptron
LDA	Linear Discriminate Analysis
PCA	Principal Component Analysis
WEKA	Waikato Environment for Knowledge Analysis

RF	Random Forest
NB	Naïve Bayes