Touch-screen Behavioural Biometrics on Mobile Devices

Elakkiya Ellavarason

A Thesis submitted to the University of Kent
for the Degree of Doctor of Philosophy
in Electronic Engineering

December 2020

Abstract

Robust user verification on mobile devices is one of the top priorities globally from a financial security and privacy viewpoint and has led to biometric verification complementing or replacing PIN and password methods. Research has shown that behavioural biometric methods, with their promise of improved security due to inimitable nature and the lure of unintrusive, implicit, continuous verification, could define the future of privacy and cyber security in an increasingly mobile world. Considering the real-life nature of problems relating to mobility, this study aims to determine the impact of user interaction factors that affect verification performance and usability for behavioural biometric modalities on mobile devices. Building on existing work on biometric performance assessments, it asks: *To what extent does the biometric performance remain stable when faced with movements or change of environment, over time and other device related factors influencing usage of mobile devices in real-life applications*? Further it seeks to provide answers to: *What could further improve the performance for behavioural biometric modalities*?

Based on a review of the literature, a series of experiments were executed to collect a dataset consisting of touch dynamics based behavioural data mirroring various real-life usage scenarios of a mobile device. Responses were analysed using various uni-modal and multi-modal frameworks. Analysis demonstrated that existing verification methods using touch modalities of swipes, signatures and keystroke dynamics adapt poorly when faced with a variety of usage scenarios and have challenges related to time persistence. The results indicate that a multi-modal solution does have a positive impact towards improving the verification performance. On this basis, it is recommended to explore alternatives in the form of dynamic, variable thresholds and smarter template selection strategy which hold promise. We believe that the evaluation results presented in this thesis will streamline development of future solutions for improving the security of behavioural-based modalities on mobile biometrics.

Acknowledgements

This thesis work was supported by AMBER (enhAnced Mobile BiomEtRics) project, a Marie Sklodowska-Curie consortium funded by the European Union's Horizon 2020 Framework for Research and Innovation under Grant Agreement Number 675087.

I would like to express my sincerest gratitude towards those who supported me throughout my PhD. My supervisor, Prof. Richard Guest, has been an invaluable guide to me from the time I applied for the PhD program until today. He has been an unrelenting source of motivation and support. Thanks are due to Prof. Farzin Deravi, my second supervisor, his excellent sense of articulation has given a great source of clarity in the research. His efforts are undeniable in streamlining my research in this topic of mobile biometrics.

I would also like to thank Prof. Raul Sanchez-Reillo and Dr. Ramon Blanco-Gonzalo for hosting me in GUTI, Spain, during my academic secondment. The extensive discussions and their advices on many occasions provided continuous encouragement. I am indebted to my friends at University of Kent who made last three years a great deal of fun and made it worthwhile.

A special thanks to my better half, Deepak, for the love and encouragement. He bailed me out as a copyeditor while proof-reading my scientific papers. Time and again, I was amazed at the errors that eluded me, but that he caught. I am eternally grateful to my dad and mum, who saw a potential in me at an early stage and without whose blessings I would have never endeavoured to pursue this journey. Mum, I hope I am making you proud for you are my guiding light, every day. I also thank my sisters for their support, patience and encouragement.

Table of Contents

List of Acronyms

ANN	Artificial Neural Network
BPNN	Back-propagation Neural Networks
DNN	Deep Neural Network
DTW	Dynamic Time Warping
EER	Equal Error Rate
FAR	False Acceptance Rate
FFMLP	Feed Forward Multi-Layer Perceptron
FMR	False Match Rate
FNMR	False Non-Match Rate
FPR	False Positive Rate
FRR	False Rejection Rate
FTA	Failure to Acquire
GRNN	General Regression Neural Network
GUI	Graphical User Interface
HBSI	Human Biometric Sensor Interaction
HMM	Hidden Markov Model
IEA	International Ergonomics Association
k-NN	k-Nearest Neighbour
MLP	Multi-Layered Perceptron
NB	Naive Bayes
NN	Neural Network
PCA	Principal Component Analysis
PCA	Principal Component Analysis
PNN	Probabilistic Neural Networks
RBF	Radial Basis Function
RBFN	Radial basis functions
SDG	Structural Description Graph
SFS	Sequential Forward Search
SQL	Structured Query Language
SVM	Support Vector Machine
TAR	True Acceptance Rate
UI	User Interface

List of Tables

List of Figures

Contributions of the Thesis

The main objective of this thesis is to enhance the understanding of security and usability of biometric systems used on mobile devices. During this research, three publications were made in peer-reviewed journals and at a conference. The first-author publications are listed below:

- "*A Framework for Assessing Factors Influencing User Interaction for Touch-based Biometrics*", Elakkiya Ellavarason, Richard Guest, and Farzin Deravi, 26th European Signal Processing Conference (EUSIPCO), Rome, 2018, pp. 553-557.

- "*Touch-dynamics based Behavioural Biometrics on Mobile Devices - A Review from a Usability and Performance Perspective*", Elakkiya Ellavarason, Richard Guest, Farzin Deravi, Raul Sanchez-Reillo, Barbara Corsetti, ACM Computing Surveys, Volume 53, Issue 6, Article 120, December 2020.

- "*Evaluation of stability of swipe gesture verification across usage scenarios of mobile device*", Elakkiya Ellavarason, Richard Guest, and Farzin Deravi, Eurasip Journal for Infornation Security, Article number 4, March 2020.

The publication entitled '*Touch-dynamics based Behavioural Biometrics on Mobile Devices - A Review from a Usability and Performance Perspective*' [1] is a state-of-the-art survey on touch-dynamics based modalities on mobile devices. In this paper, building upon the existing reviews, we have examined studies on touch-dynamics based behavioural biometrics based on usability and its impact on verification performance. We emphasize the need for shifting the focus on usability during performance evaluations by presenting a consolidated list of usability and ergonomic-based factors that influence user interaction, and cause performance variations. The paper entitled "*A framework for assessing factors influencing user interaction for touch-based biometric*s" [2] outlines a data collection framework for touch-based behavioural biometric modalities (signature, swipe and keystroke dynamics) that will enable to study the influence of environmental conditions and body movement on the touch-interaction. Our work in the paper titled "*Evaluation of stability of swipe gesture verification across usage scenarios of mobile device*" [3] aims to evaluate the stability of swipe gesture verification across various usage scenarios of a mobile device.

Portions of the above papers appear verbatim within this thesis.

In addition, the following papers have been published as a co-author to support the work documented in this thesis.

- "*Attacking a smartphone biometric fingerprint system: a novice's approach*," Ramon Blanco Gonzalo, Barbara Corsetti, Ines Goicoechea-Telleria, Anas Husseis, Judith Liu-Jimenez, Raul Sanchez-Reillo, Teodors Eglitis, Elakkiya Ellavarason, Richard Guest, Chiara Lunerti, M Azimi, J Khiarak, Salatiel Ezennaya-Gomez, N Whiskerd, R Kuzu, E Okoh, IEEE International Carnahan Conference on Security Technology (ICCST), vol. 675087, no. October, pp. 1–5, 2018.

- "*Biometric Systems Interaction Assessment: The State of the Art*", Ramon Blanco-Gonzalo, Oscar Miguel-Hurtado, Chiara Lunerti, Richard M Guest, Barbara Corsetti, Elakkiya Ellavarason, Raul Sanchez-Reillo, *IEEE Transactions on Human-Machine Systems*, 49 (5), pp. 397-410, 2019.

Chapter 1. Introduction

In recent years, the mobile-commerce industry has witnessed an exponential growth driven by ubiquity of mobile devices, cheaper mobile data and a general market shift towards online purchases [4], [5]. This growth in online transactions has, in turn, created a massive opportunity for the mobile biometrics domain that aids in establishing secure online verification. '*Mobile biometrics'* refers to the implementation of biometric verification on hand-held devices such as smartphones and tablets. The aim of mobile biometrics is to accomplish high security without compromising the convenience and portability of the mobile devices. Acuity's report [6] on mobile biometric market projections until the year 2022 states that 98% of all smart devices in use will be biometrically enabled. It demonstrates immense willingness of the users to adopt alternative verification methodologies to typing a Password or entering a PIN for performing verification on their device. Mainstream biometric modalities such as fingerprint, face and iris are already in widespread commercial use on mobile devices. For instance, Apple's Touch ID and MasterCard's selfie pay have helped users familiarise themselves with biometric technologies.

Despite its wide-spread adoption and a promising market forecast, mobile biometrics has open challenges. Unlike conventional biometric applications such as border control which use biometrics in a supervised environment, the transition of biometric applications to direct consumer-facing mobile devices brings in a key issue of being used in unsupervised operational environments. The biometric systems installed in a supervised environment ensures capturing good quality data with minimal efforts from the user. This is achieved by ensuring favourable conditions during the data capture such as illumination and fixed sensor positioning. The supervised environment enables minimal sensor-to-user interaction errors, which in turn improves the verification performance. On the contrary, a mobile biometric verification is performed in an unsupervised operational environment, with no fixed user interaction protocol with the sensor and in any given environment (indoors or outdoors). These factors may impact the performance of the underlying biometric algorithm.

Mobile biometric verification can be performed using physiological characteristics such as face or fingerprint or behavioural characteristics such as keystroke dynamics. Physiological biometrics on mobile devices have been exposed to a number of vigorously emerging vulnerabilities [7]. In order to augment the security, behavioural biometrics, with secure and robust verification techniques, are rising as an alternative option. Physiological modalities require specific sensors on the device such as fingerprint or iris sensor to capture the data. However, the behavioural biometrics utilises the plethora of sensors available on the mobile device (touch-screen sensor, accelerometer and gyroscope) in order to implicitly or explicitly authenticate a user, hence reducing the overall implementation cost. Although behavioural modalities are an inexpensive choice, in terms of reliability, they are not on par with the conventional modalities such as face or fingerprint. Similar to any verification method implemented on a mobile device, the challenge for behavioural biometrics is attaining high accuracy even in multiple usage scenarios. Therefore, it is crucial to understand the stability of the verification using the in-built sensors across different operational modes and environments.

The major contributions of this thesis are three-fold – firstly, the collection of real-life behavioural biometric data from a mobile device in an unconstrained operational environment such as walking outdoors and travelling on a moving transport. Secondly, extensive analysis of biometric verification performances and identification of the factors affecting the performance (pertaining to the user and the surrounding environment) and understand it is the impact on overall usability. Finally, developing methods to improve recognition accuracy using multi-modal frameworks and their comparison with uni-modal approaches.

This thesis work is dedicated to address these key challenges related to behavioural biometric verification on mobile devices. Detailed descriptions of the research motivation are outlined in Section 1.2.

1.1 Behavioural Biometrics on Mobile Devices

ISO defines biometrics as "*the automated recognition of humans based on their biological and behavioural characteristics*" [8]. Behavioural biometrics quantify behavioural traits exhibited by the users and use resulting feature profiles to successfully verify their identity [9]. Behavioural biometric modalities such as keystroke dynamics and swipe gestures, are used in the context of implicit verification [10], [11] and [12]. The implicit verification techniques non-intrusively verify throughout a session without interrupting the user's actions on the device.

The following list defines how well behavioural biometric systems qualify on the necessary characteristics of standard biometric systems [13]:

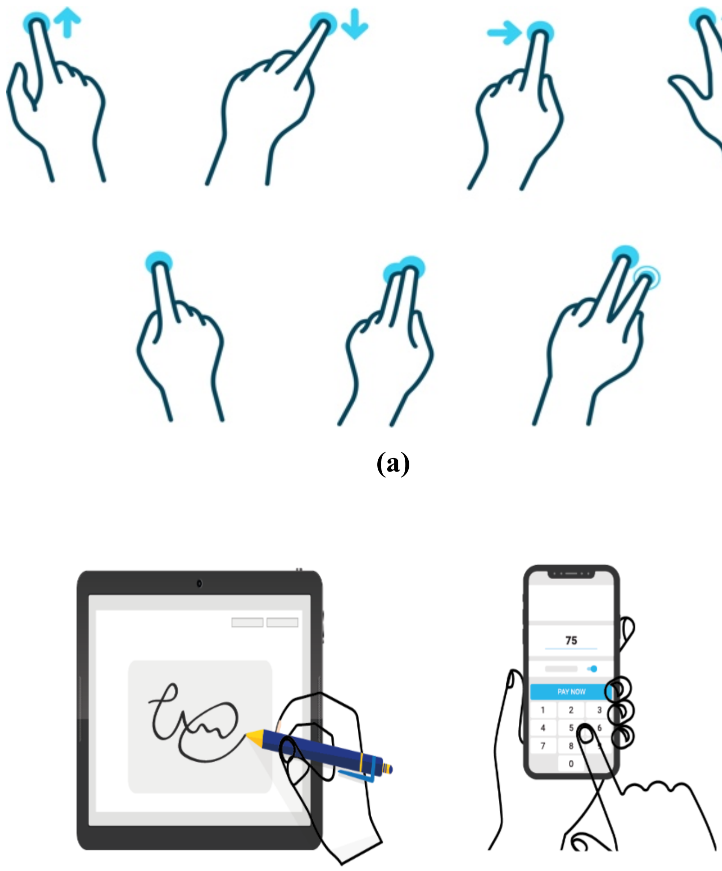
- ***Universality***: Ideally, every individual displays a particular behavioural characteristic. Universality ensures that this characteristic can be obtained from the widest (if not universal) population as possible. Touch characteristics are widely obtainable from the general population.
- ***Uniqueness***: Behavioural biometric properties are distinct for each person [14], [15] and [16].
- ***Usability***: Behavioural biometrics, when used in the context of continuous verification is non-intrusive. Therefore, it does not interrupt the normal flow of user's interaction with the device. Hence, it is perceived as a highly usable verification method.

- ***Acceptance***: Its passive and non-intrusive characteristics ensure a high degree of acceptability from a diverse set of population.
- ***Collectability***: Obtaining behavioural biometrics data is relatively easy as the data can be collected silently at the background during user interaction with the mobile device. Further, data acquisition does not require additional hardware, but the embedded sensors on mobile devices such as touch screen, accelerometer and gyroscope can be utilised. Hence, these methods are cost-effective as well as do not impose delays in user operation as they are implicit in nature.
- ***Invariance of properties/Permanence***: In order to understand the permanence factor of behavioural modalities, further research on the longevity of these modalities needs to be explored with a dataset spanning over years.
- ***Circumvention***: Token-based approaches are vulnerable to duplication or stolen identities, whereas it is relatively difficult to circumvent someone's behaviour. However, newer and affordable methods of attacks to mimic touch-behaviour are evolving such as snooping keystroke latency [17] and robotic-based approaches [18]. Given the amount of emerging vulnerabilities on behavioural biometrics, security measures and data protection techniques are expected to evolve.

1.1.1 Touch-dynamics based Behavioural Biometrics

Behavioural biometrics in context of mobile devices are driven through user's touchscreen behaviour. The touch-dynamics based methods utilise the user's interactions with the touchscreen such as typing rhythm, finger swiping speed and device holding posture. Touchscreen gestures are indicative of muscle behaviour and, hence, provide user discrimination [19], [20] and [21]. Touch sensors embedded within the touchscreen can extract features such as timestamp, touch coordinates (X-coordinate, Y-coordinate), finger pressure and finger touch area. Research efforts are being undertaken to establish such touch-dynamics based behavioural biometric methods as usable and secure methods of verification.

Figure 1.1 illustrates various touch-dynamics based biometric modalities on a mobile device. These are swipe gestures, signatures and keystroke dynamics. The swipe gestures mainly consist of horizontal and vertical swipes with upward, downward, left-to-right, right-to-left and multi-touch features. The signature on a mobile device can be performed using a finger or a stylus pen. Keystroke dynamics is performed for alphabetical or numerical input. All three of these active user-interaction based modalities are analysed in detail in this thesis.



(a)

(b)

(c)

Figure 1.1. Touch-dynamics based biometric modalities on a mobile device - a) swipe gestures b) signature c) keystroke dynamics

1.2 Research Motivation

Based on the literature survey in the domain of behavioural biometrics on mobile devices, a number of key existing problems were identified. In order to address these problems, the main research questions that were formulated and analysed in this thesis are the following:

1. *Is the biometric performance of behavioural modalities consistent and stable in different operational usage scenarios of a mobile device?*

The portability of a mobile device allows unsupervised verification processes to take place anywhere and everywhere. The key challenge for a biometric verification on a mobile device is to be able to work seamlessly under all operational scenarios. As will be demonstrated through an extensive review of the state of the art of behavioural biometric modalities on mobile devices, ensuring consistent performance is a crucial challenge that needs attention. To ensure high accuracy of biometric verification on a mobile device in any given environment, the verification algorithm must be able to adapt to unconstrained scenarios. Therefore, this work presents an in-depth analysis of verification accuracy and consistency amongst real-life scenarios. Obtaining real-life user behaviour on a mobile device requires capturing the data during real-time usage in unconstrained settings. The limitation of publicly available datasets with such real-time dynamic usage scenarios of a mobile device indicates that further research is needed in this area. Based on this, it was inferred that an extensive dataset with diverse usage scenarios from a mobile device can benefit the research in this area.

2. *Which factors in the user interaction process with the mobile device affect the overall biometric performance?*

When using the embedded sensors on the device to capture the biometric data, it is essential to examine the stability of these modalities over time and different environments. Unlike the ceremony-based data acquisition methods adopted in facial biometrics, touch-dynamic based modalities can be affected by multiple factors pertaining to the elements involved in the user interaction - the user, the device and the environment. Understanding whether these factors impact the verification performance, and the degree of impact, can lead to building robust verification methods on mobile devices. Therefore, an extensive study on the impact of these factors on individual touch-dynamic based modalities has been conducted and presented in this thesis.

3. *How can we further improve recognition accuracy using multiple behavioural biometric modalities on a mobile device?*

With the increasing risk of spoofing attacks on behavioural biometric modalities, the demand towards making the underlying security increasingly robust has become paramount. In order to achieve this, understanding the technical feasibility of implementing a multi-modal biometric system and its potential to improve recognition accuracy, in context of mobile device specific challenges, was vital. In this work, a comprehensive analysis and comparison of the biometric performances of uni-modal and multi-modal systems have been presented using a range of conventional classifiers, a commercially available verifier and Deep Neural Network (DNN) methods.

4. *How is the usability of behavioural biometrics affected by the adopted modalities and operational scenarios?*

Another equally crucial factor in mobile biometrics is usability. Users expect simple, yet convenient verification methods on mobile devices depending on the service they intend to use. These usability needs must be addressed when leveraging the in-built sensors such as touchscreen and accelerometer. This work focuses on analysing the impact of operational environments on the overall usability.

1.3 Structure of the Thesis

This thesis is composed of nine chapters in total. Chapter 2 presents the state of the art of individual modalities and multi-modal frameworks using behavioural biometrics on mobile devices. Chapter 3 describes the touch-dynamics based multi-modal dataset collected for conducting this study. This chapter details the data collection protocol - the user group, environment and scenario considerations along with the description of the data collected.

The following three chapters are dedicated to the individual behavioural biometric modalities – Swipe gestures (Chapter 4), Signature (Chapter 5) and Keystroke Dynamics (Chapter 6). Chapter 4 illustrates the performance assessment of swipe gestures-based verification using multiple algorithms – classical classification methods and a DNN method. Chapter 5 describes evaluation of signature verification methods using a commercial system and two widely used classifiers. Analysis on keystroke dynamics has been presented in Chapter 6. Chapter 7 is dedicated to assessing the multi-modal aspect of the touch-dynamics based modalities. Chapter 8 presents the usability evaluations of behavioural biometric modalities. Finally, conclusions and future work has been presented in Chapter 9.

Chapter 2. State of the Art

In the context of mobile biometrics, touch-dynamics based verification methods are distinctly trending due to their implicit and continuous verification techniques that enabled fast adoption in multiple application areas pertaining to fraud detection and cyber security. The sensor-rich touch screens on mobile devices can capture sensitive biometric features such as keystroke typing and finger swiping patterns. These features are utilised to generate the behaviour model of the user at the system level and further used for verification purposes. Extensive research work has been dedicated in recent years on touch-dynamics based verification on mobile devices. This chapter presents the state-of-the-art studies that has utilised the touch-dynamics for verification specifically using mobile devices such as tablets or smartphones.

This chapter focuses mainly on three specific touch-dynamics based modalities - swipe gestures, signature and keystroke dynamics. The reason to have shortlisted these modalities for conducting an in-depth review is because all the three modalities exhibit distinct touch behaviour of a user and require explicit user touch interaction with the mobile device during the data capture. A substantial amount of research work has been performed to improve the verification accuracy using various algorithms and classifiers. Various verification approaches such as a continuous and a one-time verification using the touch-dynamics have been studied extensively. A modality-wise comparative analysis of error rates and accuracy of various studies undertaken on mobile devices using these modalities have been summarised in this chapter.

However, despite its significance, a limited number of studies have focused on studying the external factors that impact the biometric performance on mobile devices. Literature reveals that the user interaction aspect significantly impacts the biometric process as interaction errors result in verification performance variation and subsequently cause poor user satisfaction. A list of user interaction factors causing performance variations has been reviewed and presented. Additionally, along with the review of the uni-modal systems, a review of multi-modal approaches using these modalities has also been presented. While reviewing the performance and usability-oriented studies, a number of unresolved key problems were identified. Based on the identified challenges, the research objectives were developed for this thesis. These objectives are further elaborated at the end of this chapter.

2.1 Introduction

While a majority of existing work on mobile biometrics focuses on the security aspect of touch-dynamics based behavioural biometrics such as improving the recognition accuracy, less attention has been paid to analyse the factors that influence verification performance of behavioural biometrics. Touch-dynamics based behavioural biometrics require active user interaction with a device. A failed user interaction contributes to the overall biometric system performance degradation [22]. Studies such as [23], [24] suggest that the user interaction with the sensor contributes to a higher Failure-to-Acquire rate. Therefore, assessing performance from user interaction perspective becomes important. Additionally, due to lack of standard protocol for making a correct presentation of a given trait to the touch sensor, this challenge becomes all the more crucial to evaluate.

In order to minimize the interaction errors, it is important to identify the factors that generate these errors, for example, ergonomic design of the device, usability factors and human-induced errors (such as cognitive factors). Unlike traditional biometric methodologies, such as fingerprint, which require specific sensor design for data capture, the touch-dynamics based biometric systems simply utilise the existing sensors embedded on mobile devices for data capture and processing. Even though data capture works seamlessly, factors such as ergonomic design of mobile devices can affect usability and performance [25], [23], [24].

Focusing on capturing the interaction between humans and a biometric system, the '*Human Biometric Sensor Interaction (HBSI)*' [23] model was introduced. This model is made up of three components - human, biometric system and the sensor, with overlapping operational regions of usability, sample quality and ergonomics as depicted in Figure 2.1. Designing a usable verification imposes a balanced trade-off between usability and security. These two are often seen as competing factors in design goals of biometric systems. ISO 13407:1999 [18] defines usability as "*the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use*". The term 'ergonomics' is defined by IEA as "*the scientific discipline concerned with the understanding of interactions among humans and other elements of a system, and the profession that applies theory, principles, data and methods to design in order to optimize human well-being and overall system performance*" [26]. Particularly, in biometric sensor design, ergonomics aims to accomplish an optimal relationship between the machines and humans. The central intersection of HBSI model is further expanded and the measurements used to assess the three components – usability, ergonomics and sample quality are presented in Figure 2.2. Additionally, the authors state that environmental conditions such as illumination and noise also influence all these variables. The extensive HBSI model categorically highlights the factors impacting the biometric performance and a number of evaluation methods to assess these factors. In line with this model, assessment of the user, the device and environmental factors affecting the biometric performance on the mobile devices are explored in this chapter.

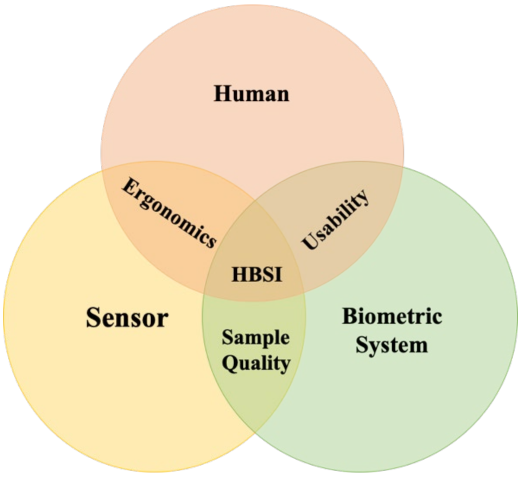


Figure 2.1. HBSI Model [23]

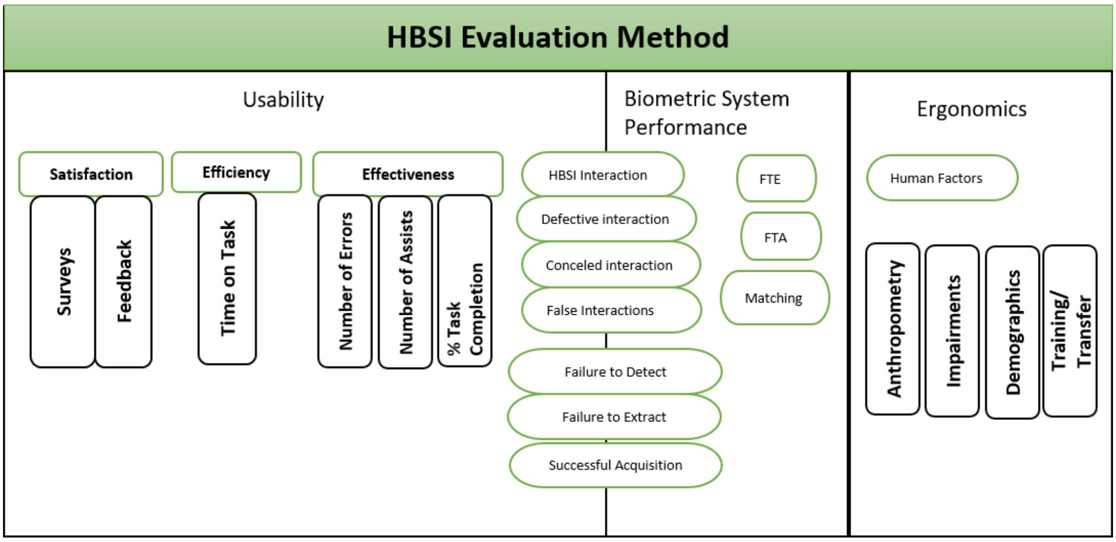


Figure 2.2. HBSI Evaluation Method [23]

This chapter presents an extensive review of performance-based studies focusing on touch-dynamics based modalities on mobile device. We also present factors influencing user interaction for individual modalities - signature, keystroke and swipe gestures. A summary and review of a comprehensive list of research studies focusing on performance-based studies and user interaction in the context of human-sensor interaction factors such as ergonomics (design factors such as screen size and orientation mode) and usability (effectiveness, efficiency and satisfaction) has been presented.

This chapter is structured as follows. Section 2.2 presents uni-modal-based studies on keystroke dynamics, signature and swipe gestures. For each modality, first the introduction of the state-of-the-art techniques used in the verification process consisting of data collection, feature extraction and classification techniques has been presented. Following this, a review of performance-oriented and usability-based studies has been detailed. Further we reviewed studies on ergonomic and human factors impacting the biometric performance. Section 2.3 details the multi-modal approaches used in behavioural biometric modalities on mobile device. In the next section, Section 2.4, we summarise our assessment of open key challenges in this domain, specifically related to the user interaction on mobile device for behavioural biometric verification. Finally, in Section 2.5, we describe the research objectives formulated to analyse in this thesis.

2.2 Uni-modal Behavioural Biometrics

Based on the purpose and the context of usage, users exhibit fundamentally different actions while making a touch interaction on a mobile device. For example, typing a message and browsing through the news content by scrolling on a mobile device are two different set of touch actions. Extensive research has been conducted focusing on the individual categories of touch-dynamics based interactions. We explore keystroke dynamics, signature and swipe gestures in detail in the following sections (Section 2.2.1, Section 2.2.2 and Section 2.2.3). Each of the sub-sections, focusing on a specific touch modality, have been structured as – a) review of the verification process used in the state-of-the-art studies, i.e. their data acquisition, feature extraction and classification techniques, b) review of studies focusing on verification performances, and c) review of studies focusing on ergonomic factors, usability and d) human factors impacting the verification performance.

2.2.1 Keystroke Dynamics

Smart devices have evolved from using a physical keyboard to an embedded soft/on-screen keyboard for typing. The rhythms and patterns exhibited by users while typing have similar neurophysiological mechanism as handwriting or signature [27]; hence, keystroke dynamics are used for verification purposes. The literature reveals that extensive research work has been dedicated towards keystroke dynamics as listed in Table 2.1. In this section, we first discuss the fundamental phases involved in verification using keystroke - data acquisition, feature extraction and classification. Following this, we elaborate on the performance and ergonomic, human-based factors affecting the verification performance.

2.2.1.1 Keystroke Dynamics Verification Process

- *Data Acquisition and Feature Extraction*

The data acquisition phase of the keystroke dynamics captures raw typing data from mobile device such as the key being pressed and the associated timestamp of the keypress event. Further, derived features are extracted from these raw temporal data. The principle concepts behind keystroke dynamic features are key latency/flight-time and hold-time/dwell-time. The time between successive keystrokes is referred as latency and the amount of time between press and release of a key is termed as hold-time.

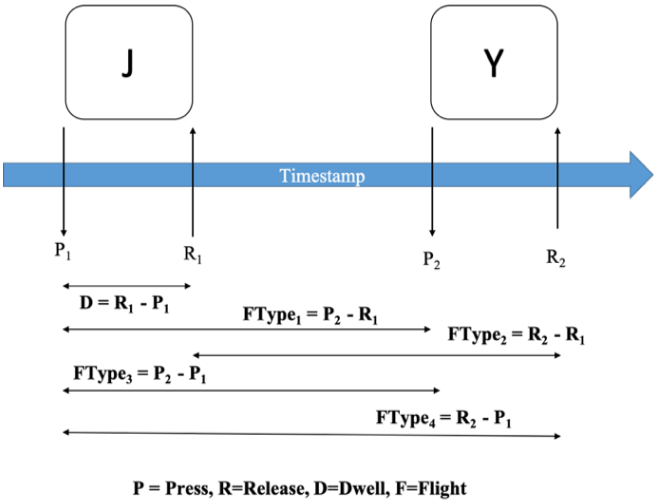


Figure 2.3. Keystroke Features – 'Dwell-time' and 'Flight time' for keys J and Y

Figure 2.3 illustrates feature data of keystrokes with four flight types – $FType_1$, $FType_2$, $FType_3$ and $FType_4$. $FType_1$ is the time between the release of one key (R_I) and press event of next key (P_2). $FType_2$ is the time between the releases of one key (R_1) to release of the consecutive key that has been pressed (R_2). FType$_3$ is the time between two consecutive key presses ($P_2 - P_1$) and FType$_4$ corresponds to time between the first keypress (P_1) to second key release (R_2). The inter-keystroke latency has proven to hold strong discriminative characteristics [28]. Timing information of two consecutive keystrokes, better known as a di-graph, is also the major feature represented in keystroke dynamics domain [28].

- *Classification*

Given the feature vector of a user's typed data, the system has to decide whether the data belongs to a claimed (genuine) user. Two classification methods are commonly used - a one-class classifier and a two-class classifier. A one-class classifier can provide data description based on a positive sample set of the genuine user. Unlike the binary-classifier, for a one-class classification, the information is only available from a single class and any other data is treated as an outlier. A two-class classifier makes use of both positive samples from genuine users and negative samples from imposter users. Once the data is partitioned into training and testing set, a classifier is trained based on the positive samples from a genuine user and negative samples from an imposter. A range of algorithms such as neural networks and statistical learning algorithms including FFMLP, PNN, BPNN, RBFN have been predominantly used in the studies conducted.

2.2.1.2 Performance in Keystroke Dynamics

Publication	Year	Subject Size	Device Information					Performance			
			Input Type	Input Length	Screen Display Size (inch)	Device Used	Keyboard Variant	Features	Classification	EER(%)	FAR(%) or FRR(%)
Saevanee et al. [29]	2008	10	N	10	-	Notebook Touchpad	-	L, HT, P	k-NN	HT-30, L-35, P-1	-
Huang et al. [30]	2012	40	A	40	-	-	-	L & HT	Statistical	7.5	-
Trojahn et al. [31]	2012	35	N & A	7 & 12	4.3	Samsung Galaxy SII	QWERTZ & 12-key	L, HT, P	NN	-	QWERTZ-12.13/8.7 12-key - 9.04/ 6.66
Gascon et al. [32]	2014	300	A	160	-	-	-	L, HT, O & motion	SVM	TPR 92%	-
Xu et al. [21]	2014	30	N (PIN)	6	4.3	Samsung Galaxy SII	QWERTY, QWERTZ, 12-key & Swype	L, HT, P finger size	SVM	28 users in training - 3.33	-
Giuffrida et al. [33]	2014	20	N & A	-	4.0	Samsung Nexus S		L, HT, G & A	SVM, NB, kNN, Mean	OOM> 0.5, TOM> 7%	-
Javed et al. [36]	2017	120	C & A	-	-	-	-	L	Statis-tical	96.5 Accuracy	FRR 3.15
Corpus et al. [37]	2017	30	C, A & S	8 & 16	-	-	Customised QWERTY	L, HT & A	NN	61.11 Accuracy	FAR - 7
Zhang et al. [38]	2017	30	C, A & S	8 & 16	4	iPhone 5s	-	3D magnetic finger motion	NN	85.79- avg. accuracy	-
Lee et al. [39]	2018	22	N (PIN)	6	5.2	Nexus 5X	QWERTY	L, HT & motion sensor data	Euclidean & Manhattan	7.89	-
Lamiche et al. [40]	2019	20	A	39	4.3	Xiaomi 2S	QWERTY	L, HT, motion sensor	MLP	1	-
Kalitha et al. [41]	2020	150 [42], 54 [43], 52 [44]	N & password combination (A, N & special character)	N- 16 & 10 Combination - 11				L, HT	GMM	5.90 (min-max) 6.02 (Z-score)	-

Table 2.1. Studies based on keystroke dynamics on mobile devices

*A-Alphabetical, N-Numeric, S-special character, GT-Gesture typing-Typing a word in single continuous stroke, L-Latency, HT- Hold time, AR-Artificial Rhythms, IL-Input length, OOM-One Order Magnitude, TOM-Two Order Magnitude, DT Decision Tree, LR Linear Regression, MLP-Multi-layered perceptron.

Table 2.1 lists a number of studies focusing on keystroke dynamics, specifically performed on mobile devices. With respect to the impact of the make of keyboard layout on the recognition performance, a comparative analysis between a physical keyboard and soft-keyboard on mobile phones was conducted by Trojahn et al. [45]. They performed the comparison between 12-key layout hardware keyboard model and capacitive screen touch-based keyboard of QWERTZ layout and compared between PIN (numeric) and password-based (alphabetical) input types. For numerical input, a 12-key layout showed average FRR 6.66%, whereas the QWERTZ layout showed an average FRR of 8.75%. For alphabetical input, the 12-key layout showed better FRR rates. Clearly, these results illustrate the impact of change in keyboard design. However, contemporary devices very rarely contain a hardware-based keyboard.

Newer typing techniques such as gesture typing using Swype or SwiftKey feature are emerging as a result of efforts to make typing more usable [46], [47]. Gesture typing enables visually guided and seamless type tracing. Bi et al. [47] compared the traditional touch-typing techniques with the gesture typing. They showed that the recorded EER is 5-10% higher while using gesture typing. The concept of providing verification based on a user's preferred method of typing is appealing. However, preference of users of gesture typing versus traditional keypad method needs to be surveyed to understand the user inclination toward these novel typing techniques.

Smith et al. [46] also explored gesture typing technique and proposed optimizing QWERTY key layout to address the ambiguity issue arising due to identical gestures generated by similar word gestures such as 'for' and 'four'. They argued that layout optimisation is required for accommodating the typing complexities and their solution improves typing speed and accuracy over QWERTY. A challenge with such an approach is that introduction of newer layouts could further confuse users. It should also be noted that different devices also provide keyboard customisation options to its users to improve convenience while typing. Additionally, studies should also take into consideration that some users have auto-correct feature enabled on their mobiles. In such scenarios, the typing instances greatly reduce as a user may just stop typing and wait for auto-correct to complete the word. An analysis of the intra-person variability across auto-correct and non-auto-correct key input scenarios can be valuable to study.

Giuffrida et al. [33] proposed sensor-enhanced keystroke dynamics using gyroscope and accelerometer sensors. Their experimental results prove that compared to standard keystroke dynamics features (e.g. keystroke timings), sensor-enhanced keystroke dynamics can improve the accuracy of gesture-based verification techniques by up to two orders of magnitude (i.e. 0.08% EER vs. 4.97% EER with the best detector/password). The technique to combine keystroke-dynamics with other sensors definitely contributes towards building stronger security. However, a number of factors need consideration. For instance, it is important to evaluate the battery drainage especially when used in the context of continuous verification. Using multiple sensors would also mean recording accurate readings from all the sensors simultaneously. This brings into play factors such as hardware faults that can stop the entire verification process even if it happens from a single sensor. Additionally, the system needs to be certain that the information coming from each one of the sensors is from a real device as opposed to a virtual machine.

2.2.1.3 Ergonomic Factors Impacting the Performance of Keystroke Dynamics

A number of factors related to design of mobile device influence effectiveness of user interaction process.

- Hand Postures - Unlike traditional biometric sensors, keystroke dynamics employed on a mobile device do not prescribe a fixed preferred posture for its usage. Buschek et al. [48] reported on the impact of hand-posture variations on the recognition performance. They propose to have improved usability by adopting an approach that avoids restrictions on the users to use a specific typing posture. They evaluated one-thumb, two-thumbs and index finger typing. They state that “entering a password in a system trained on a different posture increases Equal Error Rate (EER) by up to 86.3% relative to a system assuming a fixed posture” [48]. To address this issue, they developed a probabilistic technique to predict posture and developed posture-specific user model to improve accuracy. They also explored training the model with all possible postures. Their framework achieved an EER of 21.02%. Additionally, they also suggest that performing single session evaluations by choosing training and testing data from same session limits the applicability of the results obtained. They performed across-session evaluations to improve applicability and recommended performing evaluations in at least two sessions. We believe that the idea of developing a posture-specific user model seems promising; however, a thorough assessment of variation in postures needs to be evaluated in cases such as switching between postures and exhibiting multiple hand postures while typing such as index and two thumbs together.

- Keyboard type/Soft-keyboard variants - With the increasing number of mobile phone models and operating systems, a variety of keyboard variants are emerging with a focus on enhancing typing efficiency. Figure 2.4 depicts a number of soft keyboard layout variants on mobile phones. The dominant features for keystroke dynamics are latency and dwell-time, both dependent on time. Both features are influenced by placement of keys in the layout. iOS 11 introduced a one-handed keyboard where the key layout shrinks to the size of a four-inch iPhone and the icons are moved entirely on the left or right in order to make typing with single thumb easier. With the emergence of new layouts, users have to first familiarise themselves to these layouts, leading to a typing behaviour change over time. The underlying learning algorithm should be able to adapt to account for the learnability factor. Cuaresma et al. [36] conducted a study to capture the typing variations using a QWERTY soft keyboard layouts such as Octopus, Curve, and T+ layouts on mobile phones.

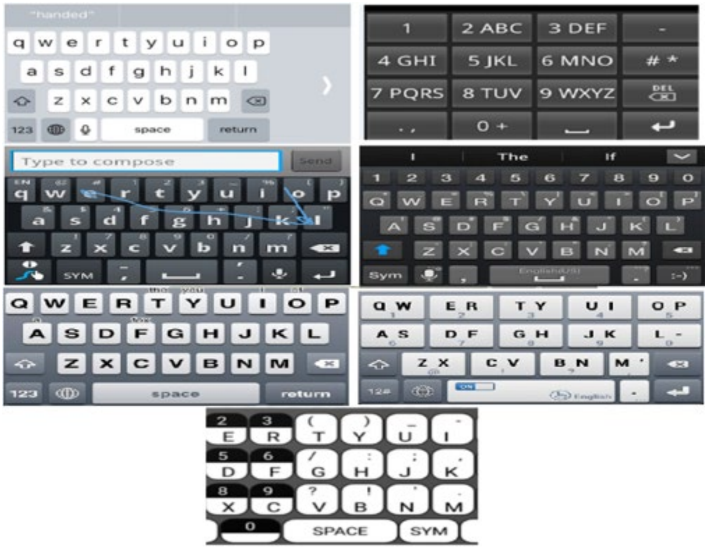


Figure 2.4. Types of soft keyboard layouts a) (Top left) iOS 11 one-handed keyboard [49] **b) (Top right) Android phone numeric keyboard** [50] **c) (Row 2 Left) SWYPE keyboard** [51] **d) (Row 2 Right) Standard QWERTY keyboard e) (Row 3 Left) Octopus keyboard** [51] **f) (Row 3 Right) T+ keyboard** [51] **g) (Bottom) Curve keyboard** [52]

The focus of their study was to evaluate the variation based on speed and error occurrences while typing. Their results show that the Octopus layout performed the best. Although their experimental results suggest that soft-key variants influence typing behaviour with respect to error occurrences and typing speed, their experiments involved a small subject size (12 subjects) and all subjects had prior experience with standard QWERTY keyboard layout, therefore introducing a bias in the experimental results. A deeper investigation into this factor with a larger dataset, to evaluate the extent of influence keyboard layout on typing behaviour is needed.

- Mobile device screen size - The evolution of mobile phone sizes and designs from slider phone to flip phone, smaller screens to larger screens involve physical ergonomic factors. With a larger screen size, it is a natural tendency for the user to use two hands or multiple fingers for typing, for example typing on an iPad with 12.9-inch display screen size. Hence, same user may exhibit different typing behaviour based on the soft-keyboard display size. Developing device inter-operability for keystroke dynamics is still an open challenge. Evaluating the degree of similarity and dissimilarity of typing behaviour of a user across multiple devices having varied screen sizes can shed more light into developing a generic behaviour model or a device-specific model that could be transferable across devices.

2.2.1.4 Human Factors Impacting Performance in Keystroke Dynamics

Users can potentially induce errors to the interaction process itself. A consolidated list of possible reasons for introduction of these interaction errors as found in the literature is presented below.

- Cognitive Factors/Habituation - User's psychological and physiological changes can influence the typing behaviour [53]. Within-user typing variability with time is an ongoing challenge in keystroke dynamics. For instance, habituation and learnability potentially improves the typing speed of a user over time. Multiple methods to cope with this challenge have been developed. Chang et al. [54] introduced a new feature called as cognitive rhythm as they proposed that typing rhythm can be affected by cognitive factors. Their hypothesis was that natural pauses (delays between typing characters in words) are caused by cognitive factors (e.g., spelling an unfamiliar word or pauses after certain syllables). They used a large dataset of 1,977 users to verify the effectiveness of their results. Their best result obtained an FRR of 0.7% and a FAR of 5.5%. Although these techniques focus on addressing the intrinsic typing variability, they have been applied for keystroke dynamics performed using physical keyboards.

- User's Mood - Studies reveal that the keystroke dynamics exhibited can be used to assess the user's emotional state. Bixler et al. [55] conducted a study to discriminate between boredom, engagement, and neutral behaviour of the user using keystroke analysis. They analysed keystroke verbosity, timings and experiment with socially charged issues like abortion, personal emotional experiences, happy/sad experiences and a neutral academic topic. Similarly, Epp et al. [56] conducted a study to identify emotional states based on keystroke analysis. Using typing rhythms to identify the emotion, their results using 2-level classifiers show anger and excitement accuracies of 84%. These studies also suggest that the behaviour exhibited by the users at different emotional states vary. This factor needs a deeper investigation and such studies could enlighten about the persistence factor of the touch-dynamics based behavioural biometrics in general.

- Demographics - Smith et al. [57] evaluated smartphone text input entry preferences amongst young and older adults and measured usability of these input techniques based on number of errors generated per user group. 50 subjects participated in their study and used physical keyboard, on-screen QWERTY key layout, tracing, handwriting, and voice for entering text on the mobile device. They recruited an equal number of young (25) and old adult (25) participants. They also controlled the influence of experience by recruiting only those participants with no prior experience with any of the five text input methods on a smartphone device. Their study revealed that older adults committed more errors while performing on-screen text entry compared to younger ones. However, their result is indicative of the influence of standard QWERTY key layout only and for a small group of adults. An additional factor causing such results could also be the inclusion criteria for participation that was based on having no prior experience with the text input methods used in the experiment. First time use could also have contributed to the overall error generation.

In summary, multiple studies on keystroke dynamics on a mobile device have emerged in the past decade. Despite its rapid development, challenges related to the user, the device and the environment seems to impact the biometric performance. Next, we describe signature modality in detail.

2.2.2 Signature

Handwritten signatures have been the most established means of formal personal verification for centuries. People are familiar with the use of signatures in their daily life [58]. E-commerce and financial institutions have adopted digital signatures as a legally accepted means to verify the identity of a person. As signature is an intensively researched modality, many state-of- the-art surveys already exist. A comprehensive survey by Impedovo et al. [59] addresses automatic signature verification on PDA's, digitizers and mobile devices. Al-Omari et al. [60] presented the state of the art on offline signature verification systems whilst Plamondon et al. [61] presented a comprehensive survey of online and offline handwritten recognition. This chapter includes recent developments and trends in the field of signature verification performed specifically on mobile devices and tablets. The adaptation of signature verification algorithms to smartphones come with its set of advantages and drawbacks. One such advantage is that no external acquisition hardware is required, thereby reducing overhead cost. On the other hand, compact display area of touchscreen limits the user from producing a natural signature contributing to intra-class variability.

Recent developments of implementing the use of e-signatures on multiple platforms illustrate various challenges. One of the dominant challenges of such usage is computational power of a mobile device. Verification algorithms using dynamic signatures need to be compatible with a wide range of low-end devices in terms of accuracy and computational overhead. Promising research has recently been conducted to understand the stability and complexity of signatures on mobile devices. Table 2.2 documents various studies on signature verification on mobile devices. The information about number of participants, signature capturing devices, features extracted, classification algorithms and accuracy rates for individual studies are presented in this table. Numerous studies in the past on signature verification have examined factors that influence signing capabilities of the user, such as device size, angle of stylus and grip of pen. We discuss studies that highlight impact on recognition performance due to various signature acquisition devices, preference of signing tool and variation in placement of mobile device during the data acquisition. We also discuss evolving trends in signature verification using a low number of enrolled signatures.

In this section, we first discuss the fundamental phases involved in verification using signature - data acquisition, pre-processing, feature extraction and classification. Following this, we elaborate on the performance and ergonomic, human-based factors affecting the verification performance.

2.2.2.1 Signature Verification Process

- *Data Acquisition and Pre-processing*

Based on the data acquisition method, handwritten signature verification systems are categorised into two types: static (offline) and dynamic (online). Example signatures are shown in Figure 2.5. Static signatures are drawn on a surface (typically paper), scanned and used for verification. The signature is represented as a grey level image. This is formulated as:

$$\{S(x,y)\}_{(0\leq x\geq X)(0\leq y\geq Y)} \qquad (2.1)$$

where $S\ (x,\ y)$ denotes the grey level at position (x,y). In contrast, dynamic systems produce a signal representative of the signature during the writing process. Here, the signature is represented as a signal sequence:

$$\{S(n)\}_{(n=0,1,2\ldots.N)} \qquad (2.2)$$

S(n) is the signal value samples at time $n\Delta t$ of the signing process (0<=n>=N), Δt represents the sampling period [59].

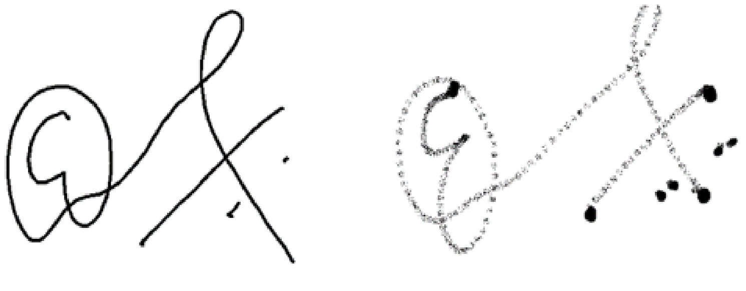


Figure 2.5. a) Static Signature (left) b) Dynamic Signature (black dots represent pen-down and pen-up points) (right)

Traditional online signature acquisition devices such as digitising tablet are used with ceremony-based acquisition processes. Touchscreen devices are used for signature capture, but are, by their very nature, uncontrolled. Pre-processing of online signatures include noise reduction, filtering and smoothing. One of the important steps in pre-processing is segmentation. No two signatures from the same person can be exactly the same. Factors such as local stretching, compression, addition or omission of additional parts [59] affect the signature. Different segmentation approaches used in dynamic signatures are pen-up and pen-down signals, velocity signal analysis, perceptually relevant points [62] and DTW. Various approaches considered end points of pen-down strokes as perpetually relevant points [63] and distance measure based on arc length.

- *Feature Extraction*

Signature features are categorised into two - function features and parameter features [59]. Function features are characterised in terms of time function. Some of the most common function features found in literature are position, acceleration, pressure and velocity. For parameter features, signature is represented as vector of elements. Parameter features are further classified into global and local features. Global features consider the whole signature e.g. number of pen-ups and pen-downs in a single signature pattern and total signature duration. Local features consider the specific features extracted from a specific part of the signature. Local parameters are further classified into component-oriented and pixel-oriented. Features that are extracted at each component level like stroke orientation are component-oriented features. On the other hand, grid-based information, intensity etc. are pixel-oriented features. After the completion of feature extraction phase, feature selection is performed. Feature selection techniques identify the most discriminative features from the given set of features. Algorithms such as PCA, SFS and sequential backward search are used for feature selection.

- *Classification*

The most commonly used verification techniques in the literature are template-matching algorithms like Euclidean distance, DTW [64], displacement functions. Approaches such as Neural Network are widely used due to their learning capabilities. Recent trend shows extensive use of HMM ([65], [66], [67]) for signature verification. Structural analysis methods like SDG analysis, string, graph and tree matching are also used as classification algorithms.

2.2.2.2 Performance in Signature

The performance of signature data captured using a conventional signature-acquiring device and a mobile device was evaluated by Jabin et al. [68]. They analysed signature verification system using DTW and HMM approaches. They also analysed the performance of these algorithms with two different signature datasets SVC 2004 [69] dataset and SG NOTE (using Samsung galaxy note) [66]. They concluded that compared to traditional biometric signature capturing device, the performance of dataset captured using mobile device was low because of the absence of information on pen-tilt angle in the mobile device dataset. In a similar manner, Galbally et al. [66] compared the discriminative power of global and local signature features between mobile devices and pen tablets.

Publication	Year	Subject Size	No. of Signatures	Input Device Information		Performance		
				Input Tool	Device Used	Features	Classification	EER (%)
Mendaza et al. [70]	2011	25	28 Genuine & Forgery, 1400 signatures	-	Samsung Galaxy S, Tab, HTC Tatoo, GeekphoneONE	Capacitive screen - X, Y, Timestamp, Size, Resistive screen - Pressure	DTW and SVM	DTW-2.33 SVM-4.28
Bailador et al. [67]	2011	96	768	In-air	iPhone	Accelerometer sensor	HMM, NB, DTW	DTW- 2.12
Krish et al. [71]	2013	25	500	Stylus	Samsung Galaxy Note	Global features-time, speed & acceleration, direction & geometry.	Mahalanobis distance & DTW	0.525
Galbally at al. [66]	2014	120 - BMDB 25 - SG NOTE, BMDB-70 [40]	-	Stylus	Pen Tablet and PDA and Samsung Galaxy Note	Time, speed and acceleration, direction & geometry, Mahalanobis distance	HMM	SG NOTE & SF -4.2 and EER -6.2 (local)
Sanchez-Reillo et al. [64]	2014	43	20640	Stylus & Finger	Blackberry Play-book, iPad2, Samsung Galaxy Tab & Note, Asus Eee PC touch T101MT	X, Y, Input type (Finger or Stylus), Timestamp	DTW	Best performance - iPad2 0.19
Jabin et al. [68]	2016	25	20 per user	-	Samsung Galaxy Note	X, Y, Timestamp, pen-up, pen-down, pressure, angle	FFNN	0.127
Zalasinski et al. [72]	2016	MCYT-100 users	2500	Stylus	WACOM pen tablet	X, Y, Timestamp, pen-up, pen-down, pressure	Neuro-fuzzy classifier	2.92
Cpaka et al. [73]	2016	MCYT-100 BioSecure 210	2500	Stylus	WACOM pen tablet and Digitizer (BioSecure)	Hybrid partitions	DTW	4.88
Diaz et al. [65]	2017	100	-	Stylus	WACOM pen tablet		DTW-based, HMM & Manhattan Distance	1.96
Yahyatabar et al. [74]	2017	-	-	-	-	X, Y, Timestamp, velocity, pressure	RT and CNN	SVC2004 - 3.0
Vinayak et al. [75]	2018	64	-	Stylus	WACOM Intuos 4 Digitizer	X, Y, Z, Pressure, Azimuth Angle	k-NN	TRR 90.18%
Riesen et al. [76]	2019	SUSIG-V [77]-100, MCYT-100, SIGCOMP-11[78]	-	Stylus	Interlink Electronics ePad-ink tablet, WacomIntuos A6 pen tablet, Wacom Intuos3 A3 Wide USB Pen Tablet	X, Y, Z, Pressure, Angle	String edit distance	EER- 1.63%
Jain et al. [79]	2020	4782	-	Stylus	WacomIntuos A6 pen tablet	X, Y, Z, Pressure, Angle	Shallow CNN	EER – 0.2 (MCYT-100), 04 (MCYT-75)

Table 2.2. Overview of studies on signature on mobile devices

* BMDB- BioSecure Multimodal Database, SG NOTE- ATVS's Samsung Galaxy Note database, HP - Hewlett Packard, SF- Skilled Forgeries, T- Toshiba, RT- Randon Transform, CNN- Convolutional neural network.

Their experimental results show performance degradation in feature discriminative power and a higher verification error rate on hand-held devices. They concluded that one of the main causes for performance degradation was the absence of pen-up trajectory information in hand-held devices.

The signature enrolment process involves the user signing repeatedly to generate the enrolment samples. This enrolment process can be frustrating for the users. In order to improve the user experience during enrolment process, studies have been conducted to utilise only one reference/enrolled signature for verification. Diaz et al. [65] proposed a signature verification methodology using only one real signature sample, instead of five or ten signature specimens to learn interpersonal variability. Their system was assessed against state-of-the-art signature verifiers and multiple databases. Their experimental results show that their system using a single reference signature was able to perform to a similar level as standard verifiers. Their novel system generates intra-personal variability from the synthetic generation of duplicated signatures using only one signature. The idea of generating synthetic signatures was developed to augment situations of limited sample availability.

2.2.2.3 Ergonomic Factors Impacting the Performance in Signatures

- Device Design - Poor ergonomics and small input areas on mobile devices are two key factors that influence signature presentation. Paudel et al. [80] suggested that the area covered by signature and its length depends on screen sizes. As screen size is device dependent, few feature values depend on the screen pixel density [80]. Authors in [81] point out that unfamiliar signing surface may also affect the signing process. The dimensions of signature capture box can influence the signature produced. Prompting users to present a signature in a small box can limit the user from producing a normal signature.

- Signing Tool - Different kind of stylus pens are available with different devices. These styluses have varied thickness influencing the signature production. Blanco-Gonzalo et al. [82] evaluated the usability of signature using three different kinds of styluses (colour coded) on an iPad. Their results show that a stylus with an 11 cm of length and 8 mm of diameter yielded EER of 0.21%, followed by a stylus with 12.6 cm of length and 8 mm of diameter that yielded 0.22% EER and finally a stylus of 0.7 cm length and diameter of 6 mm yielded EER of 0.35%. The results suggest that two styluses with similar diameters yielded similar EER rates and there is a significant rise in EER with reduced diameter. However, the same impact is not observed with change in length of styluses. The clear variation in results produced indicate likelihood of an optimal stylus diameter whereas impact of length is less clear. Sanchez-Reillo et al. [64] conducted a performance evaluation of handwritten signature verification in mobile environments using stylus and finger signatures for 43 subjects. They report based on user experience feedback collected at the end of the study that "*finger-tip-based* *devices are the less preferred by users because of the lack of habituation to make the signature with the fingertip*" [64]. Their paper also raised a relevant question of considering a signature modality into two different ones - stylus-based or fingertip based. Although their study reported better user experience with stylus, it must be noted that a significant majority of new mobile phone models come without a stylus. Hence, future research should focus more on methods to improve the user experience and performance of fingertip-based signatures. In the same experiment, they analysed signatures captured across multiple mobile devices - Blackberry playbook, an STU, an Intuos, an Apple iPad2, an Asus Eee PC touch, a Samsung Galaxy Tab and Note. The results were analysed based on the interoperability, modality tests and visual feedback. Their experiment for intra-device and inter-modality evaluations showed performance variations and iPad yielded the best result of 0.19% EER. They reported that receiving visual feedback was the most important factor for the users as they felt less comfortable with absence of visual feedback from the device. This hypothesis was based on the poor error rate for Intuos device that provided no visual feedback.

- Pen-grasping Posture - The variation in pressure due to different pen grasping postures can influence the pen-down and pen-up positions. A study conducted by Cheng et al. [83] proposed that the pen-grasping posture in itself is a personalised feature. Their experiment used video-based analysis of the pen-grasping posture and signature trajectories through modified motion energy images for identification. Another experiment by Savov et al. [84] analysed hand-pen motion for signature verification. They investigated the dynamics of hand-pen using a web camera. Although aforementioned studies were conducted for different purposes and on small set of users (35 and 10 users respectively), both indicate that the pen-grasping postures could be used as unique feature for identification. It would be interesting to see if the pen-grasping feature is stylus/pen specific. Problem may arise when the same user uses different kinds of pens for enrolment and verification.

2.2.2.4 Human Factors in Signature

- Ageing – Ageing is inherent to human nature. In the context of signature recognition, the changes related to age can bring in changes in behavioural patterns that may result in larger intra-user variability over time. A study conducted by Galbally et al. [85] evaluated a dataset consisting of 29 users acquired over a timespan of 15 months' time difference between the first and second data acquisition sessions. Their results conclude that ageing is user-dependent and both simple and complex signatures revealed same amount of ageing effect. They also suggested a '*ageing detection*' protocol for performing a template update on the enrolled signatures.

- Mood- Signatures depend heavily on the signer's psychophysical state and the condition under which the signature acquisition occurs. Complex theories have been proposed to model the psychophysical features and ink-depository process of the signature. There is a limited understanding on the influence of mood on signature presentation. Blanco-Gonzalo et al. [24] conducted a study to analyse influence of stress on recognition process of dynamic handwritten signatures. In their experiment, they introduced a stress influence test phase. In this phase, they provoked stress on the user by introducing annoyingly loud sounds, a countdown from 5 to 0 and prompting messages on the screen (e.g. you are too slow, go quicker) to speed up the signing process during data collection phase. Their analysis revealed that the stress influenced tests obtained better performance than the verification phase without stress. Therefore, their results indicate that stress factor does not have a major detrimental influence on the performance. However, the result could be biased for certain users who did not feel stressed during the process, since it is hard to conclude whether the stress influence tests did induce stress or not. Also, due to limited studies on this factor, there is a need for further research focus on this topic.

2.2.3 Swipe Gestures

Due to an ongoing trend of moving towards continuous verification ethos, swipe gestures as a biometric modality has gained considerable attention. Swipe is one of the dominant user actions on the touchscreen for content navigation. Studies on swipe gestures suggest that the patterns exhibited by a user are discriminative and, hence, can be employed as a biometric modality for continuous verification [86], [87] and [88]. Unfortunately, as signature, factors such as emotional state can influence the swipe patterns, making behavioural modelling further challenging.

In this section, we first discuss the fundamental phases involved in verification using swipe are data acquisition, feature extraction and classification. Following this, we elaborate on the performance and ergonomic, human-based factors affecting the verification performance.

2.2.3.1 Swipe Gesture Verification Process

- *Data Acquisition*

A touchscreen is already embedded in most of smart devices. As a result, no special device is required for data acquisition purposes. Unlike conventional biometric modalities, which require users to follow multiple steps for enrolment and verification, a swipe gesture action on smart devices is already familiar to users, thereby enabling a high comfort factor. Raw swipe data acquired from the touchscreen consists of X, Y coordinates, timestamp, finger pressure and finger area in contact with the screen. The acquired raw data is converted into temporal features.

- *Feature Extraction*

The feature extraction of swipe gesture data consists of distance travelled by X, Y coordinates, the first and second differential of distance - velocity and acceleration of the swipe. These features can be categorised as local (from one data point to other) and global features (features calculated from the entire swipe such as *total time of a swipe*).

- *Classification*

A number of classification algorithms are used for swipe gesture-based verification such as SVM using RBF [87], [21], [89] k-Nearest Neighbor [87], Random Forest [90] and Bayes Net [90]. The list of studies using these classification methods are provided in Table 2.3, where a number of studies on swipe modality have been listed. These studies cover topics such as user's preference of app, evaluation of verification performance in unconstrained environment to improve applicability, impact of ergonomic factors such as orientation, evolving methods to generate transferable behaviour model across devices and using swipe data for soft biometrics.

2.2.3.2 Performance in Swipe Gestures

A number of studies have focused on using swipe-based verification (listed in Table 2.3). A study dedicated on understanding the dynamics about how many swipe strokes contributed to better performance rates was conducted by Frank et al. [87]. They proposed that a sequence of swipe strokes performs better compared to a single stroke. Wang et al. [86] conducted an empirical study on transferring a behavioural model from one device onto other devices, and performed continuous verification using cross-device verification. They analysed the app usage of Hacker News Android app by collecting touch features such as X, Y coordinates, timestamp, pressure and finger size. Using SVM and Random Forest classification methodologies, they achieved AUC (Area under a Curve) score of 80% to 96% for determining the authorised user. Miguel-Hurtado et al. [91] presented a work that that predicted user's gender based on the swipe gesture data. They used the SSD [92] dataset with 116 participants and achieved 78% accuracy rates. These emerging studies tapping into soft biometrics using swipe gestures are promising.

Zhang et al. [89] proposed an active user verification approach using touch gestures by building linear and kernalised dictionaries based on sparse representations and associated classifiers. They collected a multi-modal dataset containing face and touch gestures using an iPhone 5s under three illumination settings - well lit, dim-lit and natural daylight. For one swipe, a Radial Basis Function (RBF) SVM performed the best, but as the number of swipes increased, Kernel Dictionary-based Touch Gesture Recognition (KDTGR) showed the best performance. Serwadda et al. [93] conducted a performance evaluation of 10 state-of-the-art touch based verification classification algorithms. They categorized the EERs based on the orientation (portrait and landscape) and horizontal and vertical swipes. Across the ten-classifier algorithms, logistic regression classifier had the lowest mean EER rate of 13.8%.

Publication	Year	Subject Size	Input Device Information				Performance		
			Input	U\C	Device Used	Display Size	Features	Classific ation	EER(%) /FAR or FRR(%)
Frank et al. [87]	2013	41	A	-	-	-	X, Y,Timestamp, orientation	SVM RBF and k-NN	13 strokes- 2-3, 11 to 12 Strokes - 20
Li et al. [94]	2013	75	A	-	Motorola Android phones	480x854 pixels	X, Y, pressure, Timestamp	SVM Gaussian RBF	Portrait - 95.78%, Landscape- 94.20%
Xu et al. [21]	2014	30	A	-	Galaxy SII	4.3	X, Y, timestamp, Finger size, pressure	SVM RBF	30 users- pinch (3.33%) slide (1.3%)
Feng et al. [16]	2014	123	P	U	Samsung Galaxy S-III, S-IV, Nexus 4, S3	4.8 - SIV, Nexus IV- 4.7	X, Y, timestamp, size, pressure, swipe length, curvature	DTW with One Nearest Neighbour	90% accuracy
Zheng et al. [20]	2014	80	A	C	Samsung Galaxy Nexus	4.65	Acceleration, pressure, size, and time	Nearest Neighbour distance	3.65
Bo et al. [95]	2014	100	P	C	HTC EVO 3D and Samsung Galaxy S3	HTC - 4.3, S3 - 4.8	X, Y, timestamp, pressure, vibration, rotation	SVM	Static scenario- FAR Tap- 22,Fling 9,Scroll- 23), Walking - 100% after 12 walking steps
Zhao et al. [96]	2014	78	A	C	Samsung Galaxy S3	4.8	X, Y, Pressure, timestamp	STDI with GTGF	
Saravanan et al. [90]	2014	20	A	C	Nexus 7, Nexus 4	Nexus 4 - 4.7, Nexus 7 - 7	X, Y, Pressure, relative Timestamp	SUA -SVM and RF MUA NB, J48, Random Forests and BayesNet	97.9% accuracy - mobile phones, Tablets -96.79%
Zhang et al. [89]	2015	50	A	C	iPhone 5s	4	X, Y, timestamp	SRC, rbfSVM & KSRC	rbfSVM - 19
Miguel-Hurtado et al. [91]	2016	116	A	C	Galaxy S2	4.3	X, Y, timestamp, pressure, finger size	NB, logistic regression, SVM and decision tree	78 % accuracy
Sharma et al. [97]	2017	42	A	C	Google Nexus 7	7	X, Y, timestamp, pressure, size	SVM	Two class SVM - 7
Ahmad et al. [98]	2017	40	A	C	-	-	Interaction trace map	SVM	All interactions - 80.27
Wang et al. [86]	2017	160 app usag e data	P	U	Nexus S, Nexus 4, Nexus 7-2012, Nexus 7-2013	Nexus S-4, Nexus 4-4.7, 7- 7	X, Y, timestamp, pressure, size	SVM & RF	AUC score of 80% to 96%
Filippov et al. [99]	2018	21	A	C	-	-	X, Y, timestamp, pressure, size	Isolation Forest	FAR – 7.5% FRR – 6.4%
Siirtola et al. [100]	2019	100	A	C	Samsung Galaxy S4	5	X, Y,Timestamp, Accelerometer data	Expectation Maximisation Clustering	EER - 7% (read & walk)
Li et al.	2020	150	A	C	-	-	X, Y, timestamp, pressure, size	SVM, Naïve Bayes, J48, BPNN	SVM - FAR- 3.5, FRR – 4.1

Table 2.3. Overview of studies on swipe gestures on mobile devices

*C-Constrained, U-Unconstrained, A-Active, P-Passive, NB- Naive Bayes, SVM-Support Vector Machine, RF-Random Forest, SUA- Single User Verification, MUA - Multi-user verification, GTGF-Global Touch Gesture Feature, STDI-Statistical Touch Dynamic Images, SRC-Sparse Representation-based Classification, guest users = non-owners of the mobile device, BPNN- Back Propogation Neural Network.

When swipe gestures are used in context of continuous verification on a mobile device, it is expected that the verification performance remains constant across different usage scenarios, without prompting the user to authenticate himself/herself multiple times. Additionally, irrespective of UI context of the mobile device, the verification model must be able to work seamlessly. In order to assess these factors, individual studies have conducted extensive analysis. For evaluating performance in unconstrained environment and different usage scenarios, Bo et al. [95] conducted an experiment involving static and dynamic modes. They developed the '*SilentSense'* framework to transparently authenticate users by utilising their touch pattern and micro movement of the device. Their experiment employed 100 volunteers and used an HTC EVO 3D and a Samsung Galaxy S3 for their experiment. They evaluated the performance under static and dynamic scenarios (walking). Combining walking features (accelerometer and gyroscope) with touch features to establish SVM model for the dynamic scenario, their results stated that the FAR and FRR are reduced to 0% after three steps and after 12 steps the accuracy was 100%.

Similarly, Feng et al. [16] analysed touchscreen gestures in context of running a background application in an uncontrolled environment using a Samsung Galaxy S3, a Galaxy S4 and a Nexus 4. Data was collected from 123 users (23 device owners and 100 guest users) over 3 weeks. Data variations were reported for the same user while using three different applications. Such a conclusion would raise questions related to the adaptability of behaviour model on entire operating system. They also identified that touchscreen data in an uncontrolled environment is noisier compared to a constrained environment. This implies a challenge in selecting the training data set. To tackle this problem, they introduced dynamic template adoption model using multi-stage filtering to prevent a template library growing unbounded as a user's behaviour undergo changes. Their study also evaluated battery consumption for verification process. However, in terms of energy consumption of the recognition framework, it did not exceed 6.2% of the battery usage and, hence, is not very significant. Additionally, it is important to note that app usage may be different for different user groups, for instance older adults may use a limited number of apps compared to a younger population.

A user's touch target on the screen may differ based on the screen context. Common touch gestures such as horizontal swipes, vertical swipes, zoom-in and zoom-out are performed based on the underlying UI element of respective app on the mobile device. Experiments done by Feng et al. [16], [101] propose that there is a significant variation in terms of location and touch pressure of swipe for the same user in different UI contexts. Moreover, the same scroll gesture can differ based on the screen content. Further studies need to analyse how the underlying GUI element influences the swipes produced. For instance, while reading an email, the length of a swipe may differ based on the length of the email and the screen size. How would the verification algorithm handle cases where the swipe strokes produced are short (i.e limited number of data points)? Should the verification algorithm process each incoming swipe or an accumulated set of swipes? The set of criteria to identify a feature-rich stroke and an outlier in such cases becomes challenging. All these factors would substantially affect the design decision of the verification algorithm.

2.2.3.3 Ergonomic Factors Impacting the Performance in Swipe

- Device screen size - Saravanan et al. [90] evaluated touch interaction traces of subjects on different mobile devices with various screen sizes. Their aim was to authenticate users based on the behaviour through user interface elements. Their results revealed an average of 97.9% accuracy with a mobile phone and 96.79% accuracy with a tablet for single user classification. Although the results indicate a difference in accuracy rates amongst two devices, this study involved only 20 participants. Further analysis is needed to investigate this factor.
- Orientation of the phone - The scrolling action on a screen in landscape and portrait mode can be different. Fierrez et al. [102] explored the datasets containing swipes in both landscape and portrait mode. According to their analysis, the results for the inter-session scenario show a better recognition performance for landscape mode than in portrait orientation. This indicates a need to develop matching algorithms that can mitigate this variation caused due to orientation modes.

2.2.3.4 Human Factors in Swipe

- Length of the finger - Predominantly the thumb or the index finger is used for swiping on the touchscreen. Additionally, the finger length may determine the reachable area on the screen. Bevan et al. [103] examined 19,000 swipe gestures captured from 178 volunteers. Their study was to identify the length of a person’s thumb based on how they interact with a smartphone. Their results concluded that people with longer thumbs complete swipe gestures faster than those with shorter thumbs.

- Hand postures - User’s preference for holding the mobile device on the left or right hand can differ in various situations based on their convenience. Especially in case of left-handed users, the phone holding postures can differ compared to right-handed users. Based on the holding posture, the target touch area in contact of the user’s finger can vary. Buschek et al. [104] evaluated the influence of targets and hand postures on touch-dynamics based behavioural biometrics. Their analysis revealed that touch targets near screen edges show the most descriptive targeting pattern.

- Finger used - The finger used on the left or right hand may influence the touch area on the touchscreen. Buschek et al. [104] proposed that thumb touches are more distinctive than the index finger. Does it mean that different enrolment data should be stored for different fingers? If so, an optimum amount of data required for each finger need to be identified. Along with this, which fingers are best suited for enrolment and cross-finger verification also needs to be assessed. The solutions developed so far can be used across different fingers.

2.3 Multi-modal Behavioural Biometrics

Though behavioural biometrics can be obtained in a non-intrusive manner, unimodal behavioural biometric approaches yield lower accuracy rates than practically viable. In order to increase reliability and accuracy, multi-modal behavioural solutions combining two or more biometric modalities are considered.

Additionally, the multi-modality approach safeguards against an intruder attack, as it is difficult for an attacker to simultaneously spoof multiple biometric traits. The study conducted by Fridman et al. [90] characterises performance of the system with respect to intruder detection time. They used four modalities - text entered via soft-keyboard, applications used, websites visited, and physical location of the device as determined from GPS (when outdoors) or WiFi (when indoors). They utilised the parallel-distributed fusion scheme. This system consist of a decision fusion center (DFC), which utilises local decisions u_1, u_2, u_3, ..., u_n made by n local detectors about a binary hypothesis (H_0, H_1). The decision is $u_i = 1$, if the detector decides in favour of H_1 and $u_i = 0$ if it decides in favour of H_0. They applied optimal fusion rule where the local detectors are fixed and local observations are statistically independent conditioned on the hypothesis. This scheme allows each classifier to observe an event, minimize the local risk and make a local decision over the set of hypotheses based on only its own observations. Their results show an EER of 0.05 using 1-minute window to below 0.01 EER using 30-minute window. Their study also quantifies the contribution of each modality to overall performance. According to their results, location contributes the most followed by web browsing. Text entry contributes the least for minor window of 1 minute but improves its contributions for large time windows. They also conclude that app usage is the least predictable contributor. Their analysis show compelling results to prove that the location-based modality contributes most to the performance.

One of the drawbacks of using multimodality approach in verification systems is that each modality happens in certain time window i.e. not always in parallel. For example, the study conducted by Shi et al. [95] utilises multiple sensors on a smartphone - voice, location, multi-touch sensors. While the accelerometer sensor is triggered when the user is walking, location-based sensor data can only be captured when the users' location pattern changes. Similarly, voice-based verification can be activated only on detecting the user's voice. Due to this factor, there needs to be a sliding time window in order to accumulate the scores from each modality to make a final call on the verification.

Regarding the usability aspect of the multi-modal approach, Trewin et al. [105] conducted a study to evaluate the user effort, error and reactions on using multiple modalities. They state, "*the conditions* *that combined two biometric verification modalities were disliked by the participants, had higher FTA and lower performance on the memory recall task*". This suggests that combined sample collection for biometric fusion is not necessarily preferable to collecting individual samples [105]. They suggest that this opinion could have been formed due to server delays resulting in long waiting time for the users. Despite a number of emerging studies showing promising accuracy improvements using the multi-modal behavioural modalities (listed in Table 2.4), one of the paramount concerns with all mobile devices is the battery consumption. The study conducted by Tanviruzzaman et al. [95] compares the modalities with respect to battery usage. They used Google Nexus S phones to collect gait and GPS data from walking of 13 users. Their study suggests that periodic gait computations save battery consumption. On the contrary, obtaining GPS data to check whether the current location is familiar significantly drains the battery.

Studies are focusing on fusing behavioural biometric modalities solely. Saevanee et al. [93] investigated three behavioural biometric modalities - behaviour profiling, keystroke and linguistic profiling based on SMS texting activities. Utilising the dataset from [97], [98] and [99], their results show an increase in classification performance with an overall EER of 8% with matching-level fusion. Crawford et al. [94] proposed a transparent verification framework by integrating multiple behavioural biometrics with conventional verification. Their security and usability evaluations reveal that the legitimate user was asked 67% less often to authenticate compared to implicit verification.

Publication	Year	Subject Size	Device Used	Modalities	EER(%)
Clarke et al. [91]	2009	27	Sony Vaio UX1, HP Mini-Note	Face, voice, keystroke	0.01
Shi et al. [92]	2011	50	Nokia N900	Voice, location, multitouch, accelerometer, GPS, touchscreen, microphone, movements	0.971 for the first user using binary classifier
Saevanee et al. [93]	2012	30		Behavioural profiling, linguistic profiling, keystroke dynamics	BP-20, KS-20, LP-22
Bo et al. [15]	2013	100	HTC EVO 3D and Samsung Galaxy S3	Touch, accelerometer, gyroscope	<1
Tanviruzzaman et al. [95]	2014	13	Google Nexus S	Gait, location tracks	10
Aronowitz et al. [96]	2014	100	iPhone5 and iPad2	Chirography (user writing on multi-touch screens), face, voice	0.1% in Office environment 0.5% in noisy environment
Xu et al. [4]	2014	30	Samsung Galaxy SII	Keystroke + Handwriting + Swipe + Pinch & Slide)	Keystroke-0.88,, Slide-0, Handwriting -13.89, Pinch - 0
Zhang et al. [29]	2015	50	iPhone 5s	Face + Touch-dynamics	
Fridman et al. [90]	2017	200		Linguistic analysis and Behavioural profiling	0.05 for 1 min window

Table 2.4. Overview of studies on multi-modality on mobile devices

Bo et al. [15] built a touch- dynamics based biometrics model of the user by extracting principle features of the touch. They subsequently utilised touch and movement features to build a transparent user verification framework. Their evaluations revealed 99% accuracy. Overall, efforts are being made to combine multiple behavioural modalities at the score level or matching-level fusion to improve the accuracy rates.

2.4 Key Challenges Related to Usability and Performance of Behavioural Biometrics

Building a secure and usable verification system has been an ongoing challenge. Verification process on mobile devices is expected to be fast, convenient and secure. However, considering different use cases, the prioritisation of these factors may vary. Sasse et al. [105] state - *"the core principle of usable security is that security is not the primary goal for regular users of computer systems*" [105]. For example, for a mobile banking app, security becomes more important, while for a phone unlock process, a fast and convenient verification is preferred. Users activate their phone multiple times a day in order to access various services on the phone. Users are considered as the '*weakest link in the security chain*' and user behaviour is identified as one of the major reasons for security failures [105]. In the same study, the authors show how undesirable user behaviour with password-based solution are triggered due to lack of support, training and unattainable task demands. However, as discussed earlier in this paper, user interaction is the most significant part of behavioural biometrics and thus, needs detailed evaluation based on each modality. Summarily, it is equally important to incorporate the user factor along with the performance considerations during design and development of verification process.

With respect to interaction errors arising due to device ergonomic factors, Feng et al. [101] revealed that screen size of a smartphone could change the touch and device-holding behaviour of a user. Clearly, the diversity in emerging mobile phone models, with varied screen sizes, could make this observation challenging. Especially in case of users owning multiple devices, considerations have to be made to create device-specific behaviour model. Bo et al. [95] analysed the effect of motion (walking) on swipe gestures. Their studies reveal a considerably negative impact of motion on overall performance of the verification algorithm. The False Reject Rate goes as high as 18% after two walking steps in the walking scenario. These studies confirm that owing to the mobile device design specifications, user interaction errors can arise that result in performance degradation and, thereby, affect the user preference. A detailed discussion on modality-based usability evaluations on various device specific ergonomic factors are provided.

Kukula et al. [23] define three categories that affect the biometric system performance – the users, the environment and the algorithm. As illustrated in Section 2.2.1, Section 2.2.2 and 2.2.3 a number of ergonomic and human-based factors affecting the overall verification performance have been researched for the touch-dynamics based modalities on mobile devices. However, a limited number of studies have considered varied usage scenarios of a mobile device. Table 2.5 presents a list of studies conducted on evaluating influence of posture, screen-size, distance from the device, input tool used (stylus/finger) and environmental variations. Bo et al. [2] analysed the effect of motion (walking) on swipe gestures. Their studies reveal a considerably negative impact of motion on overall performance of the verification algorithm.

Zhao et al. [96] studied how different hand-holding positions impact the performance. These studies confirm that owing to the mobile device design specifications, user interaction errors can arise that result in performance degradation and, thereby, affect the user preference.

Publication	**Year**	**Subject Size**	**Device Used**	**Modality**	**Scenarios Considered**
Blanco-Gonzalo et al. [82]	2014	21	iPad with different styluses	Signature	Sitting and standing positions, device holding positions (table, hand)
Zhao et al. [96]	2014	78	Samsung Galaxy S3	Swipe gestures	Gestures Sitting, different device hand holding position
Blanco-Gonzalo et al. [24]	2014	56	Samsung Galaxy Note with stylus	Signature	User Training, HBSI implemented in the design
Buschek et al. [48]	2015	28	Nexus 5	Keystroke	Sitting position, Portrait and landscape positions, different hand postures
Bo et al. [95]	2015	100	HTC EVO 3D and Samsung Galaxy S3	Swipe	Sitting and walking
Zhang et al. [89]	2015	50	iPhone 5s	Face and Touch gestures	Lighting variation - in a well-lit room, in a dim-lit room, and in a room with natural daytime illumination
Miguel-Hurtado et al. [106]	2017	27	iPhone 5s	Voice and face	Variation in noise within a noisy office environment

Table 2.5. Overview of research that considered influencing factors in behavioural biometrics on mobile device

Given the number of emerging studies that focus on acquiring data from the mobile device under different scenarios and their results documenting the impact on the performance, we identified this as one of the important factors to consider in our study. Most of these studies have been conducted indoors with a fixed set-up for the experiment.

Although several data collections have been performed using the mobile devices, the datasets show behavioural biometric datasets collected in laboratory set-ups. None of the studies have considered external environmental factors in their usage scenarios. This indicates that a dataset that considers these factors must be explored to depict the real-life scenarios of a mobile device usage.

While several studies on multi-modality are existing, studies combining the active user interaction based modalities seem to be limited. An extensive analysis of combination of modalities along with their stability over multiple usage scenarios need to be analysed in detail. While multi-modal framework with behavioural modalities augments the verification, it is important to assess the challenges related to implementation of such frameworks.

As shown in the literature, limited number of studies have been conducted on evaluating the usability for signature and keystroke dynamics. However, considering a number of external factors that impact the data donation, it is important to evaluate how the usability is affected.

2.5 Research Objectives

Literature reveals the significance of user interaction factors in the biometric verification process. Section 2.4 highlights a few existing challenges pertaining to the user interaction in the domain of behavioural biometrics on mobile devices. These existing challenges indicate that an in-depth analysis focused towards the factors impacting the user interaction can highly benefit this domain. Owing to this, the main goal of this research was to analyse user interaction on the mobile device using touch-dynamics based behavioural biometrics under real-life usage scenarios and to evaluate the stability of the verification performances using multiple methods of verification across these unconstrained usage scenarios.

Firstly, this research focuses on addressing the challenge of capturing data with real-life scenarios. In order to do so, a data collection experiment had been conducted in which the user had the flexibility to use the mobile device in different environments and in different usage contexts. The scenarios involved using a mobile phone indoors while seated (office scenario), whilst walking outdoors, whilst exercising (treadmill) and whilst seated in a moving transport. These usage scenarios were chosen to replicate some of the real-life scenarios where a mobile device is used.

Secondly, the goal of this research was to capture the data in unconstrained scenarios. No restrictions were implied on the device holding posture and user's walking speed. With respect to the environment, the walking path was chosen where people movement was in place. This way, the user was required to be aware of the environment while using the mobile device in the outdoors scenario.

Finally, this study also focuses on performing a rigorous analysis of the stability of the verification performances of the individual modalities (swipe, signature and keystroke) across different real-life scenarios. Multiple methods, namely, conventional, deep neural network and commercially used verification system have been adopted to analyse the impact of these factors and to evaluate multiple research questions that emerged during this thesis. Following this, a multi-modal verification framework had been developed to analyse the applicability and usage of different modalities.

2.6 Conclusion

In recent decade, touch-dynamics based verification using behavioural biometrics is trending due to its fast adoption in multiple application areas pertaining to fraud detection and cyber security. This chapter is a structured review of studies done on behavioural biometric modalities, specifically on mobile devices, with two perspectives – performance and usability. A great deal of work has been dedicated towards optimising recognition performance at the algorithmic level, with limited focus given to impact of user interaction facet. We have reviewed performance aspects throughout this chapter, but also emphasise the growing need to consider user interaction factors that impact the verification performance. On the positive side, we see a change in user perception towards adopting newer verification technique – continuous verification using behavioural biometric modalities. Although the concept of being continuously monitored haunts the users but its non-intrusiveness makes it an attractive alternative option. A number of factors as discussed throughout this paper needs deeper insights to deliver this attractive verification. Towards this end, we do witness an evolving interest in studies based on usability evaluations and understanding of the impact of ergonomic factors such as device screen size on the recognition performance. With regards to user interaction, deeper analysis and modelling of the device interaction complexities can help design better feedback mechanisms for the verification process which, in turn, would improve the user experience. It would also help to find empirical evidence to the perceived superiority of usability provided by the implicit nature of continuous verification using behavioural biometrics.

Given that behavioural biometrics is a fairly new domain, we believe further work is required to investigate the persistence of individual behavioural biometric modality across usage scenarios, user's emotional states and modalities. Additionally, it is also vital to understand the time invariance of the behaviour-based modalities, since that would immensely impact the long term adoption for verification.

At the same time, the number of gadgets owned by an individual is rising, and thus we see a greater need to integrate user's behavioural model across multiple devices. Hence, we witness efforts enable interoperability as a research topic to is gaining focus. To add to the challenge, mobile devices are ever emerging and have diverse hardware specifications. This means performance evaluations need to be made across devices, reporting on the battery drainage, total time taken for verification and user experience during verification. As the number of gadgets owned by individuals rise, efforts to enable interoperability is gaining focus and there is a greater need to integrate user's behavioural model across multiple devices.

The growing vulnerabilities can result in compromising sensitive private data of a user; however, we notice that considerable efforts are being taken to fuse multiple modalities to improve security using behavioural biometrics. This would enhance in building accurate user model, which can be further across multiple fields such as gaming.

In conclusion, the behavioural biometrics are certainly emerging as a new form of silent and transparent verification that can complement password-based verification. In addition to well-documented performance improvement initiatives, there is a growing need to consider user-interaction factors that have the potential to significantly improve verification usability. At the same time, this is also a concern, which may induce errors to interaction process during verification due to multitude of factors which still need in-depth research. It is our belief that some of the biggest challenges of this domain lie with replicating the real-life user interaction scenarios in training dataset and this factor would fundamentally define the adoption of behavioural biometrics as a mainstream verification method for mobile devices alongside performance improvements.

Chapter 3. Touch-dynamics based Multi-modal Dataset

3.1 Introduction

As highlighted in Chapter 2, this thesis work focuses on multiple key issues existing in the domain of behavioural biometrics on mobile devices. This chapter addresses one of the primary problems concerning a lack of real-life mobile device-based dataset by developing a multi-modal behavioural biometric dataset, captured under various operational usage scenarios of a mobile device. A need for such a dataset arose because compared to the traditional biometric application areas such as in Airports, mobile biometric solutions are used in a variety of usage scenarios. All mobile devices, by their nature, are portable and designed to be hand-held. This characteristic enables their use in far more diverse scenarios compared to personal computers and laptops. While this provides users utmost flexibility, it brings new challenges for mobile biometric solutions that are expected to work seamlessly anywhere and everywhere a mobile device is put to use. In the context of behavioural biometric verification, one of the concerns is to ascertain consistency of verification accuracy across diverse usage scenarios of a mobile device in both indoor (office, home, at the gym, etc.) and outdoor (walking, running, etc.) environments.

Literature survey revealed that most of the existing studies on touch-dynamics based behavioural biometrics have analysed touch data obtained in a laboratory setting [10], [87], [107]. Due to the unavailability of mobile device-based touch-dynamic datasets with multiple usage scenarios, for this research work the first experimental step was to conduct a data collection and develop a dataset that captures real-life scenarios of a mobile device. For the data collection process, the experimental scenarios followed specifications as stated in ISO/IEC-19795-2 [108] and were designed to simulate typical mobile phone usage. A careful consideration of including few external factors such as the environment (indoors and outdoors) and body movement variations were carried out. The experimental scenarios were designed to be ceremony-based, where the data donation method was specified to the user such as prompting the user to type a sentence or to have them signing within in a given box on the mobile device. However, the mobile device usage was unrestricted.

In order to simulate real-life usage scenarios, the data acquisition sessions were designed to collect data from the users performing tasks such as interacting with the phone in constrained (laboratory) and in-the-wild/unconstrained environments. The constrained scenarios consisted of the user performing the touch tasks on the mobile device while seated on a chair indoors and the unconstrained scenarios consisted of the user performing the tasks while walking outdoors, while walking at a fixed pace on a treadmill or whilst travelling on a bus. This dataset has been utilised in this study to analyse verification performance stability for different touch-dynamics based modalities. Additionally, multiple research questions developed during this study have been answered based on this dataset.

3.2 Related Work

There are a number of publicly available datasets on touch-dynamics based modalities acquired on mobile devices. Table 3.1 lists these datasets for individual behavioural biometric modalities – swipe gestures, keystroke dynamics and signature. Each of these datasets are composed of a different number of participants belonging to various user groups such as students and faculty of a University. The details of the subject size, device used, and the input types of individual datasets are provided Table 3.1.

In order to capture the real-world scenarios, the most recent dataset was released in 2019 by Papamichail et al. [109]. The dataset was crowdsourced and was collected by introducing a gaming app in Google's Play Store and Apple's App Store. This app was downloaded and used by 2000 participants. This app captured touch gestures such as taps and swipes. Around 2418 different devices were used, where Redmi Note was the dominant device used by maximum users in the dataset. Despite having a large number of participants in the dataset, the data capturing parameters were not fixed for all the participants and multiple devices having multiple operating systems were used. Hence, a bias, either by the device or the scenario has been introduced. To avoid the data capture using a device of individual participant's choice, Mahbub et al. [110] came up with a technique to deliver a research phone to participants and asked them to use this phone as their primary phone for a period of two months. Although this guarantees that the same device is used for data capture by all the participants, however, with such an experimental set-up, it is still difficult to account for the actual usage scenario in which the data has been captured.

The constrained data collection techniques establish control over participants' adopted data donation method. A constrained data acquisition method was adopted by the remainder of the studies listed in Table 3.1. These studies established a set of predefined tasks and actions for the participants. For some of these experiments, the data collection processes were divided into different sessions with varying number of days between sessions. For instance [111] had four sessions within a time span spread across 4 months. The number of biometric samples donated per session for each user and study also varied. Datasets such as [112] focused on using multiple devices for data collection, whereas other datasets such as [113] and [114] used a single device. The context of usage for swipe gesture data during these experiments was mostly image navigation and reading activities. The participants were allowed to use landscape and portrait device orientations during data acquisition, but the majority of data was acquired for portrait orientation [87]. The keystroke dynamic datasets consisted of alphabetical or numerical inputs in soft-keyboards. The length of keystroke inputs varies from four digits (PIN) to full sentences. The standard signature datasets contained finger-based and stylus-based signature capturing methods. In [115], multiple signature capturing devices such as Wacom STU-500, 501, DTU1031, ATIV 7 & Galaxy Note 10 were used.

Publication	Modality	Year	Subje-ct Size	Device Used	Input Type
Papamichail et al. [109]	Swipe gestures	2019	2000	2418 devices, Redmi Note (majority)	Tap and swipes
Sitova et al. [10]	Swipe gestures	2016	100	Android smartphone	Tap, swing, scale, scroll and key press gesture
Mahbub et al. [110]	Swipe gestures	2016	48	Nexus 5	Swipes, data collected over two months, free use of the device
Serwadda et al. [18]	Swipe gestures	2013	190	Google Nexus S	One finger swiping, no multi-touch gestures, two sessions, at-least one day apart
Frank et al. [87]	Swipe gestures	2013	41	Droid Incredible, Nexus One, Nexus S and Samsung Galaxy S	One finger swiping, no multi-touch gestures, two sessions, one week apart
Antal et al. [116]	Swipe gestures	2015	71	Eight Android devices-tablets & smartphone. Only Device ID provided in dataset description	Not specified
A. Morale [111] (KBOC DB)	Keystroke dynamics	2016	300	-	Alphabetical, four sessions in four months' time span
Antal et al. [117]	Keystroke dynamics	2015	42	Nexus 7 & LG Optimus	Alphanumeric password
El-Abed et al. [113]	Keystroke dynamics	2014	51	Nokia Lumia 920	Alphabetical
Tasia et al. [114]	Keystroke dynamics	2014	100	Motorola Milestone	PIN
e-BioSign-DS1-Signature DB [115]	Signature	2017	65	Wacom STU-500, 501, DTU1031, ATIV 7 & Galaxy, Note 10	Two Sessions three weeks apart
BiosecurID [118]	Signature	2015	132	Intuos3 A4/Inking pen tablet	Real and synthetic signatures
ATVS_SG_NOTE_DB [66]	Signature	2014	25	SG NOTE	Two sessions with five days gap between them
ATVS-SLT DB [66]	Signature	2014	29	Wacom Intuos 3	Signatures captured over a time span of fifteen months

Table 3.1. Publicly available datasets on touch-dynamics based behavioural biometrics on mobile devices

Despite a number of emerging publicly available datasets on behavioural biometric modalities on mobile devices, it can be noted that usage scenarios of the device have not been taken into account during the data capture. However, a limited number of studies that did not publish their datasets publicly, started including varied scenarios in their data collection. Bo et al. [14] captured the touch-dynamics data in stationary (static) and in motion (walking) modes. They intended to capture the tiny perturbation of a mobile device when a user interacts with the phone and subsequently utilised those features for verification. For the walking scenario, they reported that after two walking steps, the False Acceptance Rate reduced to 0% and after four walking steps, the False Reject Rate was 18%. This indicates that walking has an impact on the verification performance. Additionally, multiple studies [19], [20] have indicated a need to further investigate the influence of external contexts, such as sitting and walking, on the verification performance.

Considering the relative lack of studies focusing on this aspect and the extent of benefit such a dataset would bring to the behavioural biometric domain, the decision was taken to evaluate these factors facilitated by a data collection. In this data collection, multiple factors such as environment, device parameters and user-based factors along with the usage scenarios were taken into account. Data collection scenarios consisted of variation in a user's walking behaviour with controlled and uncontrolled speeds and variation in environmental setup – indoors and outdoors. A detailed explanation of the data collection set-up has been presented in Section 3.3.

3.3 Data Collection Framework

A data collection framework was developed to capture the multi-modal dataset using a smartphone. This section describes the data collection setup details.

3.3.1 Data Collection Setup

The data collection set-up process consisted of making a series of design decisions with respect to the choice of the device, usage scenarios to include, modalities to capture, participant's tasks, application development process and finding the data channels to capture. A detailed description of how each of these data collection design related questions were answered are described in the following sections.

3.3.1.1 Usage Scenarios

This data collection experiment was conducted in two sessions (Session 1 and Session 2). Both sessions were separated by a week and each session typically lasted for 45 minutes. During these sessions, three behavioural biometric modalities - swipe gestures, signatures (finger-based and stylus-based) and keystroke dynamics were captured. The reason to have focused on these three modalities is because these modalities require *active* touch interaction from the user with the device.

Each session comprised of three different usage scenarios. A usage scenario is defined as "*a real-world example of how an individual interacts with a system in a given environment*". Three elements involved in a usage scenario are the user, the system and the surrounding environment. Accordingly, variations were introduced in the experiment on basis of user's body movement, environment location and environment movement (as shown in Figure 3.1).

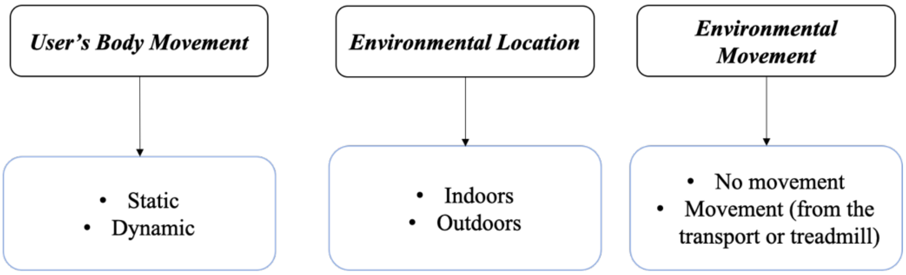


Figure 3.1. Variables considered for the data collection

User's body movements considered for the experiment were static (no movement) and dynamic (walking at controlled and uncontrolled speed). We assumed that while the user was static, the user movement was zero; however, device movements were present, which was mainly caused by the device holding method of the user. In order to capture the natural mobile device usage, hand movement of the user while holding the device was not restricted at any point of time during the experiment. Additionally, throughout the experiment, the mobile device was hand-held by the users whilst performing the experiment. Users were allowed to handle the mobile device freely (not constraining to a specific device-holding posture) while carrying out different scenarios. Next, the environmental location variations were indoors and outdoors. Finally, the environmental movements considered were from a moving transport (bus) and the treadmill.

Based on these parameters, the overall usage scenarios were designed, and they were holistically categorised into static and dynamic categories (as shown in Figure 3.2).During static scenario (*Scenario 1 – Sitting Indoors*), the participant performed the experiment on the mobile device while seated on a chair and holding the phone in their hand. Dynamic scenarios consisted of three different categories: a) the user as well as the environment are moving (*Scenario 2 - Treadmill*) , b) the user is moving, and the environment is static (*Scenario 3 – Walking Outdoors*) and c) the user is static, and the environment is moving (*Scenario 4- Travelling on a Moving Bus*).

The data collection was performed on the acquisition device – a smartphone (described in 3.3.1.6) and the participants of this study carried out various experimental tasks (described in Table 3.2) on that smartphone while the biometric data was being collected continuously in the background by the device sensors.

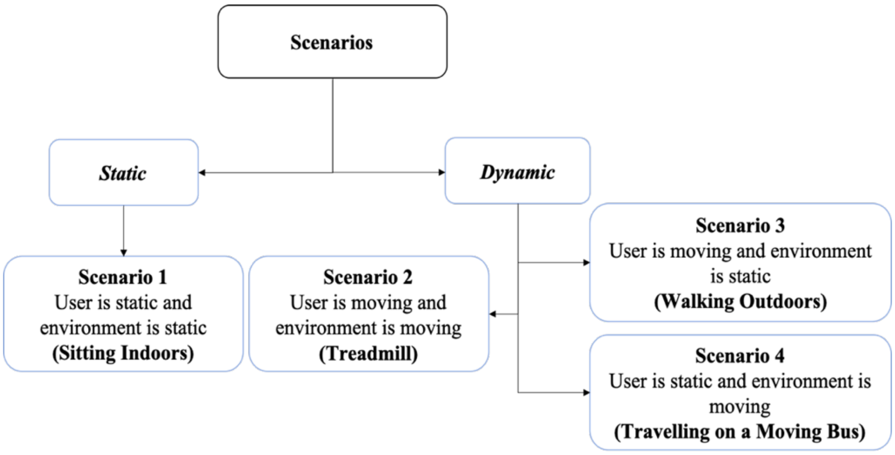

Scenarios
Static
Dynamic
Scenario 1
User is static and environment is static
(Sitting Indoors)
Scenario 2
User is moving and environment is moving
(Treadmill)
Scenario 3
User is moving and environment is static
(Walking Outdoors)
Scenario 4
User is static and environment is moving
(Travelling on a Moving Bus)

Figure 3.2. Usage scenarios

The experimental tasks were in the form of a general knowledge quiz on a mobile app installed on the smartphone. The quiz consisted of multiple types of exercises where the participants had to perform simple actions such as typing a sentence, swiping through images and signing using finger and stylus on a mobile device.

3.3.1.2 Environment

The environmental locations considered for the study were both indoors and outdoors. The indoor configuration was designed to imitate an office environment and the outdoor configurations were unconstrained. The details of these locations are provided below:

- Indoor - The indoor set-up was arranged in an experimental room in the School of Engineering and Digital Arts department of University of Kent. The floorplan of the room has been provided in Figure 3.3. The user was asked to be seated on a chair while carrying out the experiment. To have the entire indoors experiments video recorded, two cameras (one on the left-side of the chair and another on the right-side of the chair) were installed in order to record the user's hand movements during the experiment. These cameras were specifically focused towards the hand-posture of the user instead of the participant's face to avoid privacy concerns.

  In the same experimental room, Scenario 2 (*Treadmill*) was performed. For this scenario, one camera was set-up facing towards the front of the treadmill in order to capture the device holding posture of the user while performing the tasks on the treadmill.

- Outdoor - The outdoor scenarios were Scenario 3 and Scenario 4, which involved walking around the campus and travelling on a bus inside the university campus. In the outdoors scenarios, the participant was expected to be aware of the surroundings while using the phone. Two different paths were chosen for Session 1 and Session 2 (shown in Figure 3.4).

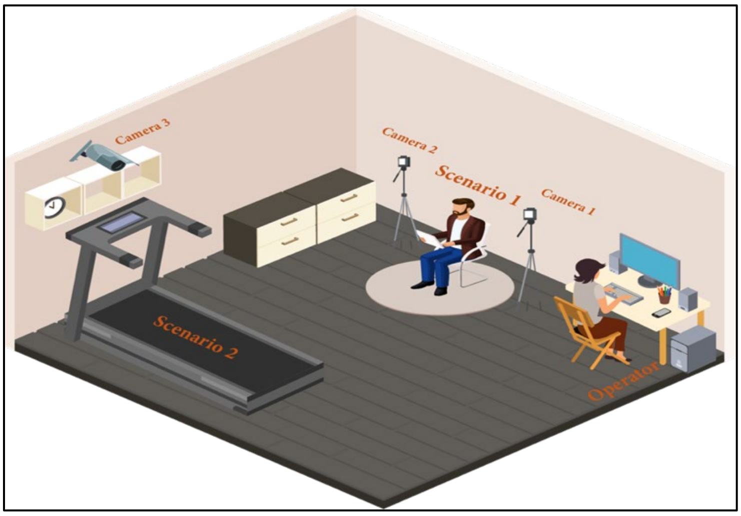


Figure 3.3. Experimental room layout for Scenario 1 and Scenario 2 conducted indoors

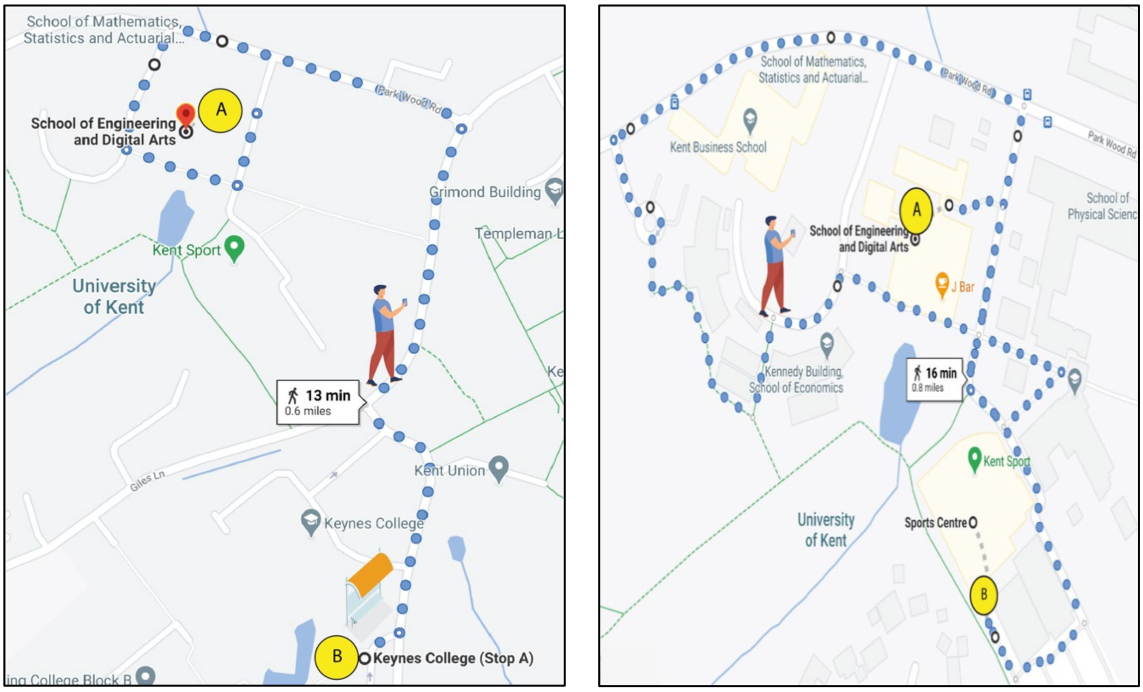


Figure 3.4. Walking path for Scenario 3 in Session 1 (left) and Session 2 (right). '*A'* is the starting point and '*B*' the end point of walking

At particular junction of the pathway in Session 2, the user had to cross the road and, thus, needed to pause performing the experiment and resume once the road had been crossed.

The bus route taken during Scenario 4 is shown in Figure 3.5. Route 1 and Route 2 took place as two-way journeys and the total time spent by the participant on the bus was around 10 minutes.

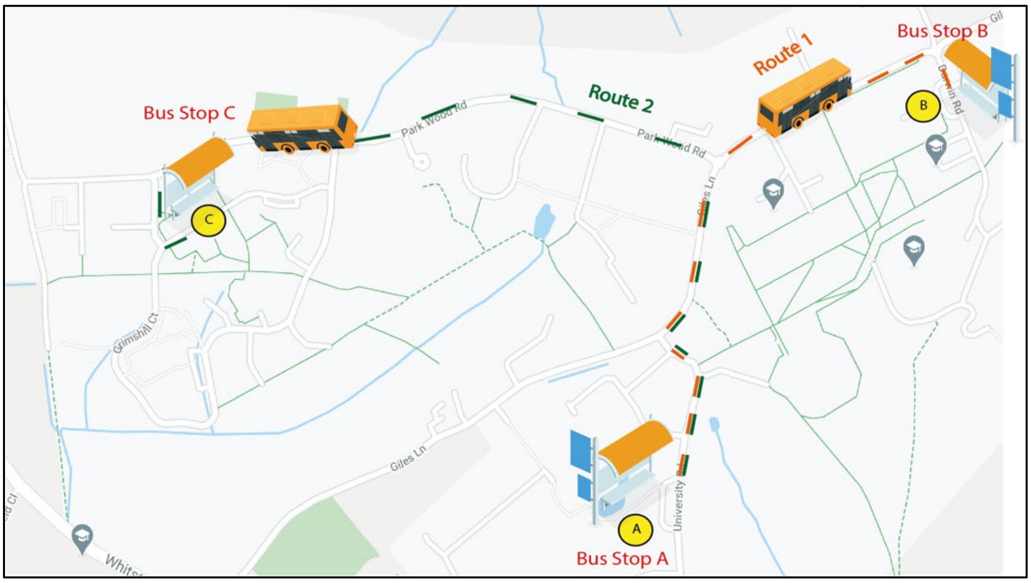


Figure 3.5. Bus routes followed during Scenario 4 of data collection

3.3.1.3 Sequence of the Experiment

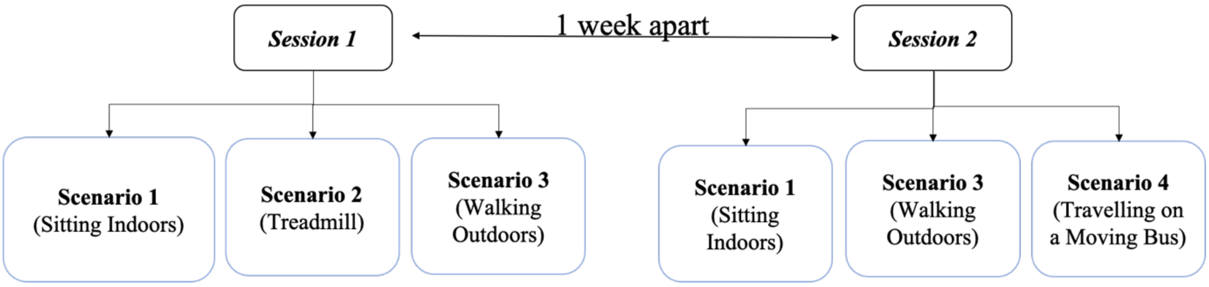

Session 1
1 week apart
Session 2
Scenario 1 (Sitting Indoors)
Scenario 2 (Treadmill)
Scenario 3 (Walking Outdoors)
Scenario 1 (Sitting Indoors)
Scenario 3 (Walking Outdoors)
Scenario 4 (Travelling on a Moving Bus)

Figure 3.6. Experiment details

Figure 3.6 shows the scenarios belonging to Session 1 and Session 2. At the beginning of the experiment, the participants were provided with the information sheets (Appendix B) and consent forms for participating in the experiment. Following this, the participants were briefed about the sequence of the tasks during the experiment. Once the consent form was signed by the participant, the acquisition device (a smartphone) was provided to the participant for carrying out the experiment. The user was instructed about the placement of the stylus pen in the smartphone in order for them to perform the signature tasks. The description of each scenario is as follows:

- *Scenario 1 (Sitting Indoors)* - This was the first scenario of the experiment and it took place indoors in an experimental room at School of Engineering and Digital Arts. The participants were asked to be seated on a chair whilst performing this scenario. It typically lasted for 15 minutes. During the session, the participant was allowed to ask questions to the instructor (person facilitating the experiment). Individual tasks of this scenario are provided in Table 3.2.

- *Scenario 2 (Treadmill)* - Scenario 1 was immediately followed by Scenario 2. This scenario took place indoors in the same experimental room as Scenario 1 and typically lasted for 15 minutes. In this scenario, the participants were asked to perform the experiment while walking on a treadmill that was installed in the experimental room. Before starting the experiment, the participants were given five minutes to perform a trial to walk on the treadmill while using the phone simultaneously for practice. During this time, the participants were allowed to set a comfortable speed on the treadmill for walking (which was fixed throughout this scenario). Participants had a choice to set this pace for themselves based on their comfort level. Once the speed was set, they were handed over the acquisition device to carry out the tasks.

- *Scenario 3 (Walking Outdoors)* - Following Scenario 2, Scenario 3 was conducted. This scenario took place outdoors, inside the University of Kent campus. In this scenario, the participants were asked to perform the experiment on the mobile device while walking. During this scenario, the participants covered 0.6-0.8 miles distance and this scenario lasted around 15 minutes. For some participants, the walking pace was slower than the others, therefore in those cases, the scenario completion time exceeded 15 minutes. The walking paths chosen for Session 1 and Session 2 were different (as shown in Figure 3.4). For Session 1, the walking path was close to the School of Engineering and Digital Arts building. For Session 2, the walking path led to Keynes bus station, following which Scenario 4 was carried out by the participant.

- *Scenario 4 (Travelling on a Moving Bus)* - This scenario was only a part of Session 2. This scenario took place outdoors, inside a moving public transport (bus). Canterbury's regularly running Stagecoach's UNI buses (UNI1 and UNI2) were utilised for carrying out this scenario. These UNI buses have two routes going through various bus stops inside the university campus. The participants were asked to board a UNI bus and carry out the experiment on the mobile device while being seated on the bus. The total duration of the bus ride inside the campus was typically 10 minutes. An illustration of the scenarios carried out by a participant in the study is presented in Figure 3.7.

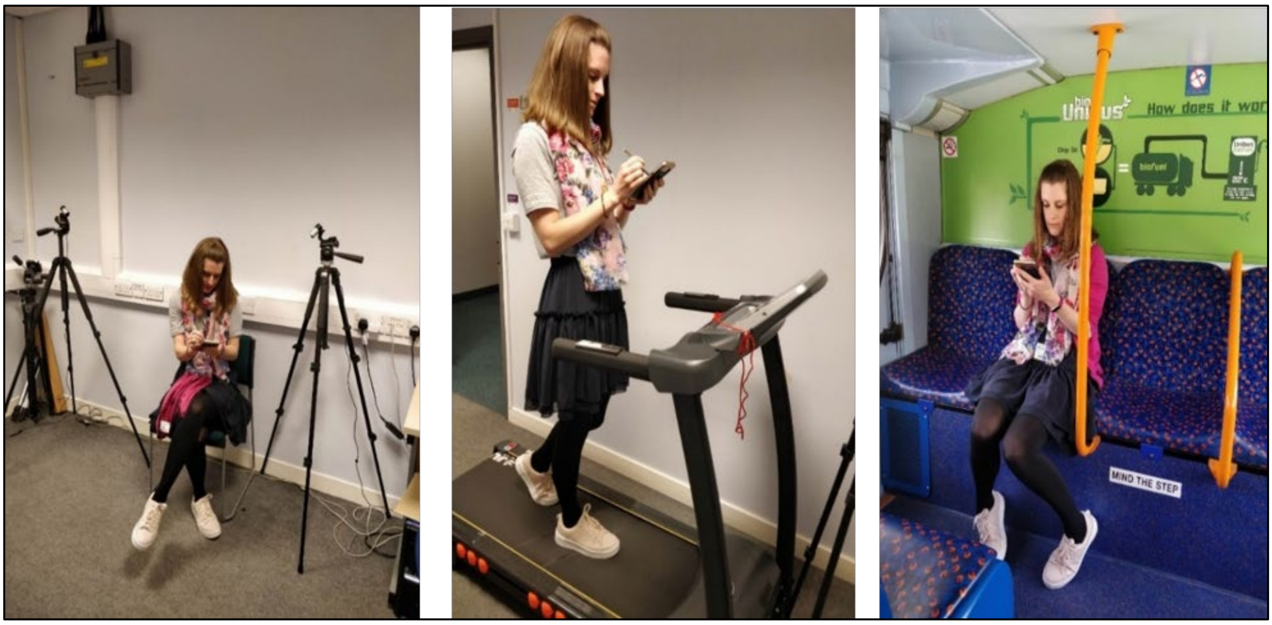


Figure 3.7. Participant performing data collection scenarios – a) Sitting indoors b) Treadmill and c) Travelling on a moving bus

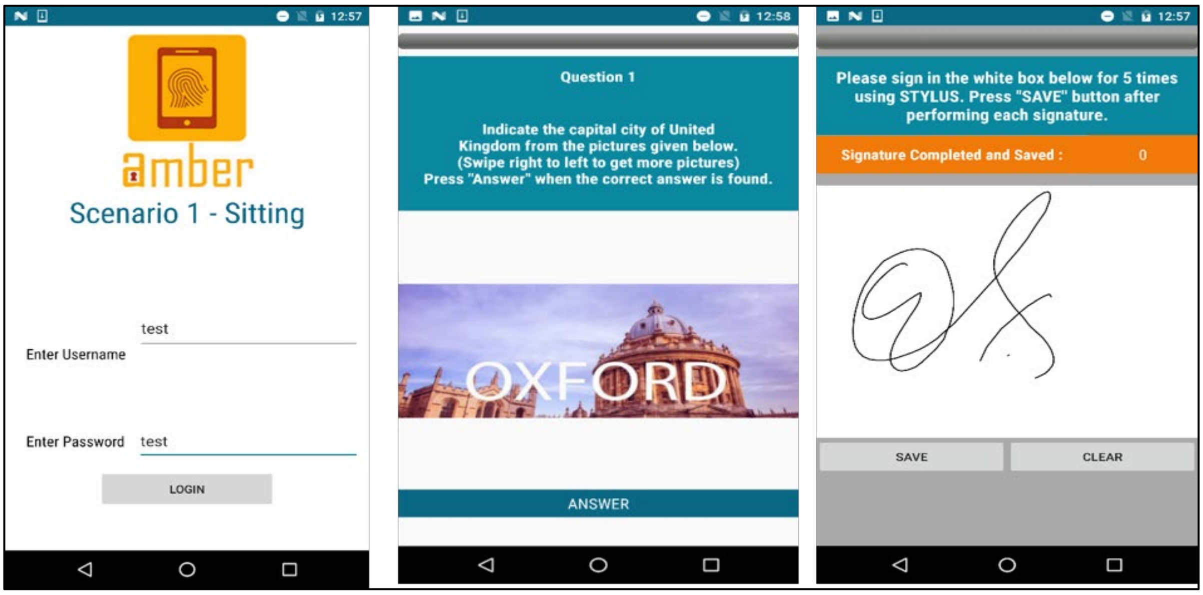
Session Number	Scenario Number	Scenario Description	Tasks		
			Swipe	Signature	Keystroke Dynamics
Session 1	1	Seated on a chair indoors	12 questions with multiple choices: 82 Horizontal swipe samples 72 Vertical swipe samples	10 Finger-based and 5 stylus-based signatures	10 typing tasks: 5 Alphabetical sentences and 5 numerical phone numbers
	2	Walking on a treadmill indoors	10 questions, with multiple choices 53 Horizontal swipe samples 43 Vertical swipe samples	6 Finger-based signatures	3 Alphabetical sentences and 3 numerical phone numbers
	3	Walking outdoors	10 questions, with multiple choices 40 Horizontal swipe samples 40 Vertical swipe samples	5 Finger-based and 5 stylus-based signatures	3 Alphabetical sentences and 3 numerical phone numbers
Session 2	1	Seated on a chair indoors	12 questions, with multiple choices 77 Horizontal swipe samples 53 Vertical swipe samples	10 Finger-based and 5 stylus-based signatures	5 Alphabetical sentences and 5 numerical phone numbers
	3	Walking outdoors	10 questions, with multiple choices 38 Horizontal swipe samples 39 Vertical swipe samples	5 Finger-based and 5 stylus-based signatures	3 Alphabetical sentences and 3 numerical phone numbers
	4	Travelling on a bus outdoors	10 questions, with multiple choices 40 Horizontal swipe samples 40 Vertical swipe samples	3 Finger-based signatures	3 Alphabetical sentences

Table 3.2. Task details during each scenario of the data collection

3.3.1.4 Application Development

In order to collect the touch-dynamics based data using different modalities, an Android-based app named *‘Touch Logger’* was developed using Android studio. The user interface (UI) of the app is depicted in Figure 3.8.

Once the participant was given the acquisition device to initiate the experiment, the participant was provided with a username and password to login and start the experiment as shown in the home screen of the *‘Touch logger’* app. Login button press triggered the recording of the sensor data in the background. The *Touch Logger* mobile application collected the touch data from the touchscreen and motion sensor metadata from the mobile device. As the user performed common touch manoeuvres such as swiping left-to-right, scrolling up-down, key typing and signing in the UI of the device, the app recorded timestamped touch parameters (X and Y coordinates and finger pressure) continuously in the background. We used the embedded SQL database engine – ‘SQLite’ to store the touch and other sensor data on the mobile device within a CSV file. At the end of the experiment, individual sensor data was saved in separate CSV files in the smartphone folder and extracted once the participant completed the experiment. The device orientation was fixed to portrait mode in all the scenarios. Therefore, the participants could not switch to a landscape mode of the UI while undertaking the experiment. The reason for fixing the orientation was to ascertain that all users followed the same phone orientation and to avoid variability in the orientation context.



Enter Username: test

Enter Password: test

Question 1

Indicate the capital city of United Kingdom from the pictures given below. (Swipe right to left to get more pictures) Press "Answer" when the correct answer is found.

Please sign in the white box below for 5 times using STYLUS. Press "SAVE" button after performing each signature.

Signature Completed and Saved : 0

Figure 3.8. User interface of *Touch Logger* app a) home screen b) swiping task screen c) signature task screen

The various touch actions that were captured during the experiment are as follows:

- **Horizontal swiping task –** left-to-right swiping, an action, that is usually performed for sliding through images or flipping to the next page of a document. We designed a photo-flipping activity in the app in the form of a quiz. For example, the user was asked to select the capital city of the United Kingdom from a list of images of various cities.
- **Vertical scrolling task –** (down-up scrolling) is an action usually undertaken whilst reading through a large document such as a news article on a mobile device. In order to imitate the vertical scrolling actions usually performed on a phone, we included vertical scrolling of images in the app.
- **Alphabetical keystroke typing task –** The keystroke entry tasks involved participants typing a sentence on a soft-keyboard. For this task, we used common sentences to obtain the key input action for commonly used characters.
- **Numerical input typing task –** For the numerical task, the participants were required to type a sequence of given numbers in a text field. The numbers appear in the form of a phone number (UK format).
- **Signature task using stylus and finger –** For the signature task, a boxed signing area of 49.21mm (height) and 76.1mm (width) was assigned on the screen. The participants were asked to sign multiple times using a stylus or a finger in the assigned box.

3.3.1.5 Data Captured

This dataset consists of a range of modalities: swipe gestures, signature (finger and stylus), and keystroke dynamics alongside device accelerometer and gyroscope data. Although all of these biometric modalities exhibit touch-dynamics, each modality differs in terms of data donation method and the data being captured. Swipe gestures consist of single or multi-finger movement on the touchscreen such as drag and flick motions. Signatures mainly focus on the stylus and finger-based inputs. For finger-based signatures, participants used their fingertips to draw the signatures on the touchscreen. Lastly, keystroke dynamics data was captured whilst the user tapped on the soft-keyboard of the mobile device using a finger.

Each modality contained different data attributes as listed in Table 3.3. There are some common attributes across swipe gestures and signature (in contrast to attributes of keystroke dynamics) such as finger X-coordinate, Y-coordinate and pressure. As keystroke data was generated by the keypress, the attributes are associated with the virtual key being pressed and any deletions made.

Modality	Attributes	Format / Input types
Swipe Gesture	Timestamp	YY:MM:DD:SS:sss
	Touch Action	- ACTION_DOWN (when a pointer (finger or stylus) touches the screen) - ACTION_UP - ACTION_MOVE (when the pointer moves on the screen))
	X-coordinate	Value may be from 0 (the first pointer that is down) to Maximum X value of the screen
	Y-coordinate	Value may be from 0 (the first pointer that is down) to Maximum Y value of the screen
	Multi-touch-pointer X-coordinate	Value may be from 0 (the first pointer that is down) to Maximum X value of the screen
	Multi-touch pointer Y-coordinate	Value may be from 0 (the first pointer that is down) to Maximum Y value of the screen
	Pressure	Ranges from 0 (no pressure at all) to 1 (normal pressure)
	Orientation	0 radians - indicates that the major axis of contact is oriented upwards, positive angle - indicates that the major axis of contact is oriented to the right, negative angle - indicates that the major axis of contact is oriented to the left. The full range is from -PI/2 radians (finger pointing fully left) to PI/2 radians (finger pointing fully right).
	Touch Size	Ranges from 0 to 1
	Tool Type	0-Finger or 1-Stylus
Signature	Timestamp	YY:MM:DD:SS:sss
	Touch Action	ACTION_DOWN (when a pointer (finger or stylus) touches the screen), ACTION_UP, ACTION_MOVE (when the pointer moves on the screen))
	X-coordinate	Value may be from 0 (the first pointer that is down) to Maximum X value of the screen
	Y-coordinate	Value may be from 0 (the first pointer that is down) to Maximum Y value of the screen
	Multi-touch-pointer X-coordinate	Value may be from 0 (the first pointer that is down) to Maximum X value of the screen
	Multi-touch pointer Y-coordinate	Value may be from 0 (the first pointer that is down) to Maximum Y value of the screen
	Pressure	Ranges from 0 (no pressure at all) to 1 (normal pressure)
	Orientation	Same as swipe gesture orientation
	Size	Ranges from 0 to 1
	Tool Type	0-Finger or 1-Stylus
Keystroke Dynamics	Timestamp	YY:MM:DD:SS:sss
	Key pressed	Character (A-Z, a-z) or Numeric (0-9)
	Key released	Character (A-Z, a-z) or Numeric (0-9)
	Deleted position	Numerical value
Location	Timestamp	YY:MM:DD:SS:sss
	Longitude and Latitude	a latitude or longitude in the form [+-] DDD.DDDDD, where D indicates degrees

Table 3.3. Data captured for each modality

As an example, Figure 3.9 shows few data attributes of swipe gesture with a series of sample swipe gestures captured during the experiment. Each swipe gesture input comprises multiple sample points consisting of X-coordinate and Y-coordinate pairs (highlighted in red, blue and yellow dots in the figure). The X and Y coordinates represent the display screen dimensions. The number of sample points in an input swipe can be different; for instance, the red coloured swipe gesture consists of 11 sample points, whereas yellow coloured swipe gesture contains 9 and blue coloured swipe gesture consists of 11 sample points.

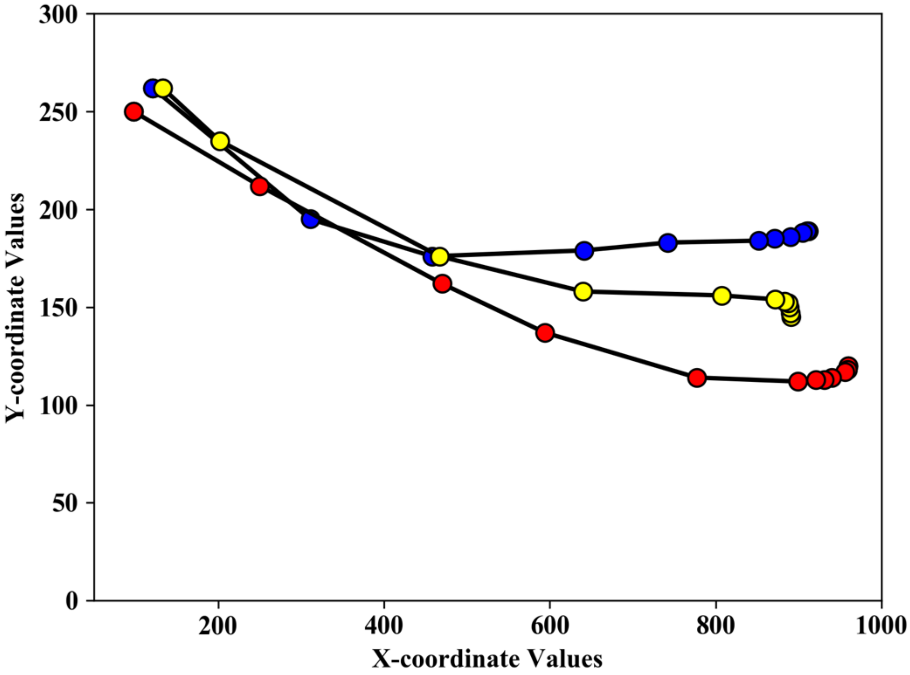


Figure 3.9. Sample horizontal swipe gestures

Table 3.4 provides the total number of samples donated for each modality (for all the 50 participants) involved in the data collection experiment.

	Swipe Gestures		Signature		Keystroke Dynamics	
	Horizontal Swipes	**Vertical Swipes**	**Finger-based**	**Stylus-based**	**Alphabetical**	**Numerical**
Session 1	8750	7750	1050	500	650	550
Session 2	7750	6600	900	500	550	400

Table 3.4. Total data collected for 50 participants

For swipe gestures, the reason for the variation in the number of samples captured in Session 1 and Session 2 is because during the vertical scroll action, the participants exhibited this action in different ways. A few users swiped faster than usual to get to the bottom of the screen easily, where the button to move to the next page content was placed. Hence, reducing the number of swipes captured in that particular task. For keystroke dynamics, an entire sentence entered by a user is considered as one sample. A total of 650 samples were present in Session 1 and 550 for Session 2.

3.3.1.6 Acquisition Device

One of the considerations for this data collection was to select a suitable mobile device for the experiment. Since smartphones are more commonly owned by users compared to a tablet or a smartwatch coupled with the fact that they are also widely used in outdoor scenarios, a smartphone was selected as the acquisition device for the experiment. Based on the statistics contained in [21], the sales trend of the global market share held by different smartphone operating systems revealed that Android-based devices are growing strongly over many years, which has led it to become the most popular operating system on mobile devices. Considering its widespread use, for this experiment an Android-based smartphone was chosen; a Samsung Galaxy Note 5 (Figure 3.10) to capture the data. The design specifications of this smartphone are provided in Table 3.5 :

Figure 3.10. Galaxy Note 5 with stylus pen [119]

Specification	Details
Dimension	153.2 x 76.1 x 7.6 mm (6.03 x 3.00 x 0.30 in)
Screen size	5.7 inches, 88.5 cm2 (~75.9% screen-to-body ratio)
Resolution	1440 x 2560 pixels, 16:9 ratio (~ 518-ppi density)

Table 3.5. Design specifications of Samsung Galaxy Note 5

Another reason for choosing this device for the experiment is that this smartphone has a built-in stylus pen (dimensions of 7.1 x 17.8 x 2 cm). As the data collection includes signature (stylus-based) as a modality, it was considered necessary to select a smartphone that comes with a stylus for signature data acquisition. The '*S-pen*' that was provided with the Galaxy Note 5 was utilised for tasks involving signature capture using the stylus pen alongside finger-based signature capture.

In order to minimise differences caused by varying device specifications in the experiment, the same mobile phone specification was used for all the experimental sessions. The participant group for the experiment included owners of iOS-device as well as Android-devices. However, in order to avoid bias for previous experience on Android-devices, the iOS-device users were given extra time before the experiment to familiarise with the Android-device.

3.3.1.7 Ethics

As this experiment involved the collection of sensitive biometric data from humans, it was conducted according to appropriate ethical, legal and professional frameworks. Approval was obtained from the University-level Research Ethics & Governance Committee of the University of Kent. As AMBER is a European Union funded project, ethics approval from the AMBER project Ethics Advisor was also acquired.

Description of hazard	**Risk Controls**
Task involves walking on the treadmill: a) Participants might fall from the treadmill while using the mobile phone on the treadmill b) Participants might feel dizzy while doing the experiment	The emergency stop button on the treadmill will be used in such cases. Instructions/demo on how to use the emergency stop will be given to the participant before the experiment.
Participants with health issues related to heart, blood pressure etc.	We ask for medical conditions of the participants in terms of heart related health problems, dizziness etc. We will not allow participants with health issues to take part in the experiment.
Tasks involve walking outdoors. a) Might get hit by a vehicle b) May get injured while walking and using phone	The experiment involves users to use the mobile phone only on a dedicated and safe walkway. Instructions shall be provided to the participants on using the mobile phone carefully while they are performing the experiments outdoors.

Table 3.6. Risk assessment form of the experiment

Additionally, departmental risk assessment procedures for conducting this experiment were also assessed. The risk assessment procedure designed by the School of Engineering and Digital Arts at the University of Kent was carried out and steps were taken to ensure mitigation of associated risks. Table 3.6 shows the risk assessment form used for the experiment. As biometric data is classified as personal under GDPR regulations, we did not obtain explicit authorisation from the participants to distribute the collected samples publicly, therefore we were unable to make these data available to the research community.

3.4 Dataset Description

In this section, we describe the dataset in detail. We report on the demographic details of the participants, operating systems used and handedness features of the participants along with the data description.

3.4.1 Demographics

This dataset contains behavioural biometric data from 50 participants. The participant crew for the experiment were recruited based on inclusion criteria such as age (above 18 years old), familiarity with using smartphones and ability to be physically mobile (without any visual or walking impairment). These criteria were included considering the walking scenarios as the participants were required to be able to walk and get onto public transport during these scenarios. We tried to include participants across a range of ethnic backgrounds and age groups. In the first step of the data collection process, the participants were asked to fill a form containing questions about their demographics such as age, ethnicity and the mobile device they currently own and use. With respect to the gender distribution in the dataset, the male participants were 60% and the female participants were 40% (shown in Figure 3.11). The age distribution of the participants is shown in Figure 3.12. As the experiment was conducted on university premises, the majority of participants were of postgraduate student age.

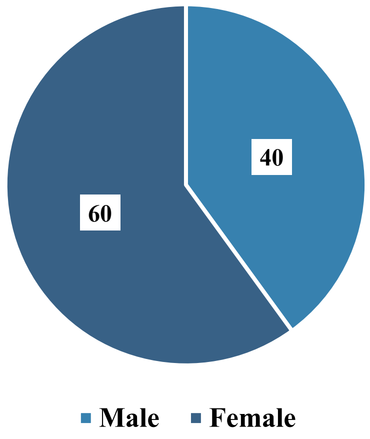


Figure 3.11. Gender distribution

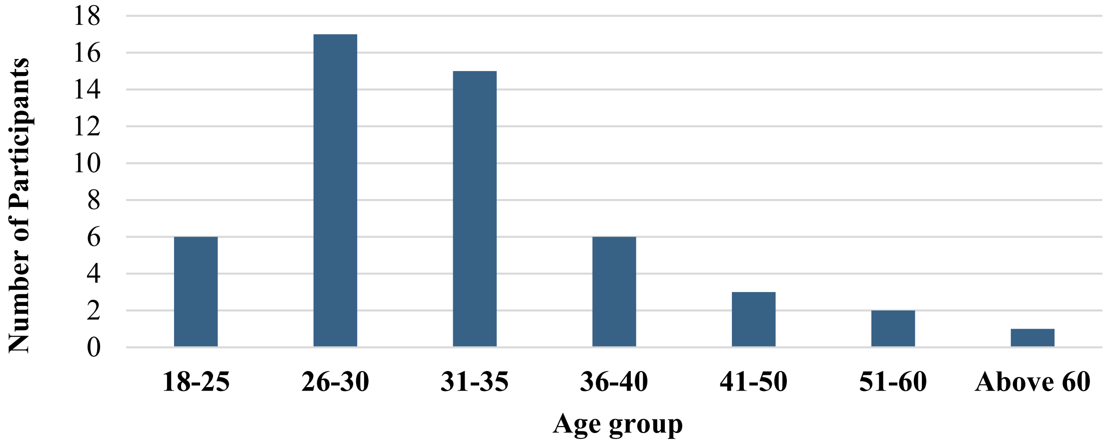


Figure 3.12. Age distribution

SL no.	Ethnic Group	Description	Number of Participants
1	American Indian or Alaska native	A person having origins in any of the original peoples of North and South America	0
2	Asian	Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam	12
3	Black or African American	A person having origins in any of the black racial groups of Africa	1
4	Hispanic or Latino	A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race	0
5	Native Hawaiian or other Pacific Islands	A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.	1
6	Caucasian	A person having origins in any of the original peoples of Europe, the Middle East, or North Africa	37

Table 3.7. Ethnic group distribution

Participants of the study were asked to indicate their ethnic background by specifying their country of origin. In order to define the ethnic categories, the reporting of National Institutes of Health (NIH) [22] has been utilised. Table 3.7 shows the ethnic group distribution of the participants in the study.

After the demographic information had been collected, the participants were asked to fill the questionnaire related to the mobile device they own, their handedness, followed by previous experiences using our test modalities in their daily life.

3.4.2 Operating System Usage

Based on the device related information, participants owned diverse smartphone models such as iPhone, Samsung, Motorola, Huawei, Sony, HTC, Nokia and LG. Based on this information, we categorised iOS-based and Android-based device owners (depicted in Figure 3.13).

Both operating systems have core differences with respect to the user experience. They engage differently with the mobile content. The UI content works differently for both with different kinds of app drawers, icons, controls, and notifications. Along with these elements, a smooth app navigation process contributes to a better overall user experience. Most of the participants in this dataset were Android-based users.

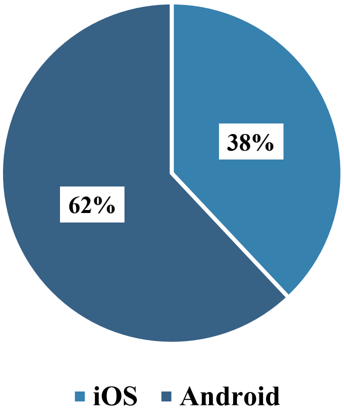


Figure 3.13. Percentage of operating systems used by the participants

3.4.3 Handedness

As this research focuses on touch-dynamics based modalities, it was important to record the handedness of the participants. Figure 3.14 shows the handedness characteristics of the participants; with the majority of the participants being right-handed. Six users were left-handed and one user indicated an ambidextrous nature for performing touch-based activities on mobile devices.

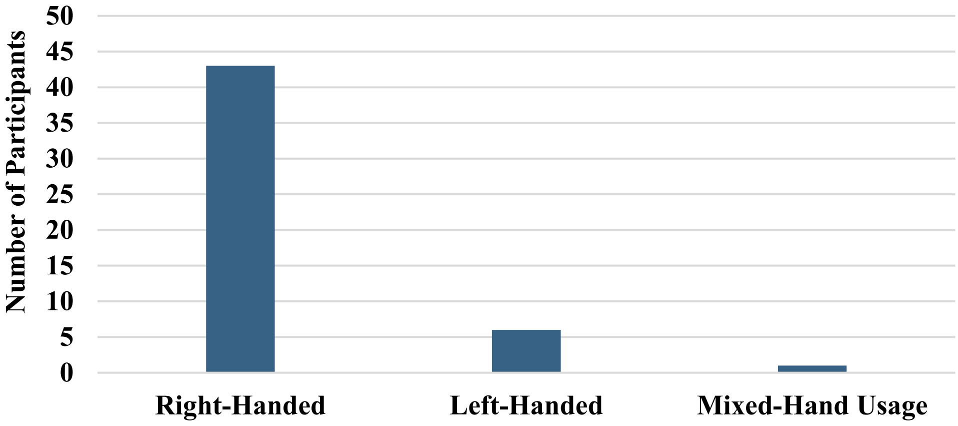


Figure 3.14. Handedness of the participants

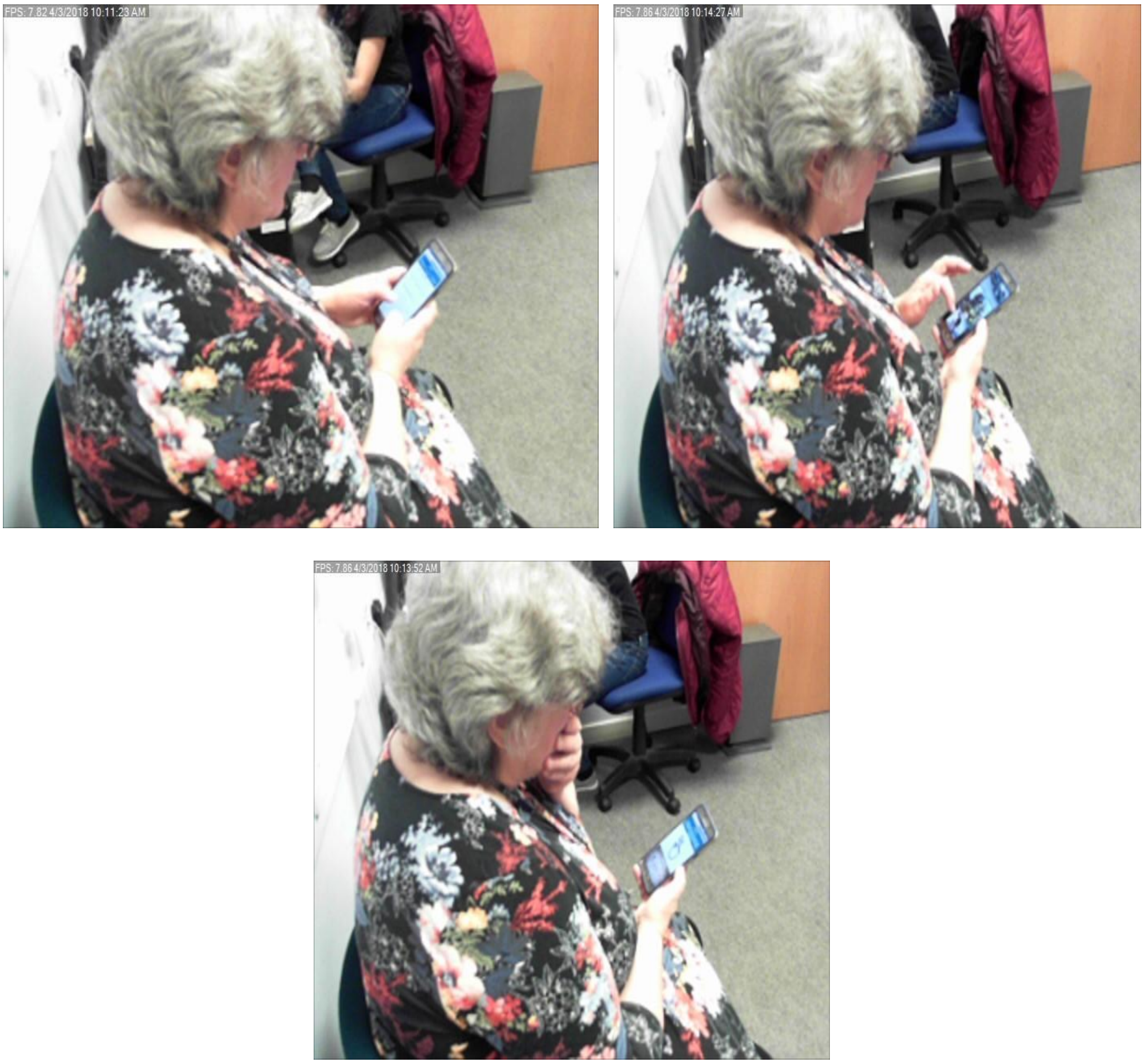

FPS: 7.82 4/3/2018 10:11:23 AM
FPS: 7.86 4/3/2018 10:14:27 AM
FPS: 7.86 4/3/2018 10:13:52 AM

Figure 3.15. Left-handed participant's illustrative hand postures – a) Both hands holding the smartphone (top left), b) One hand holding the smartphone and one hand used to perform touch actions (top right) and c) One hand used for holding the smartphone (bottom center)

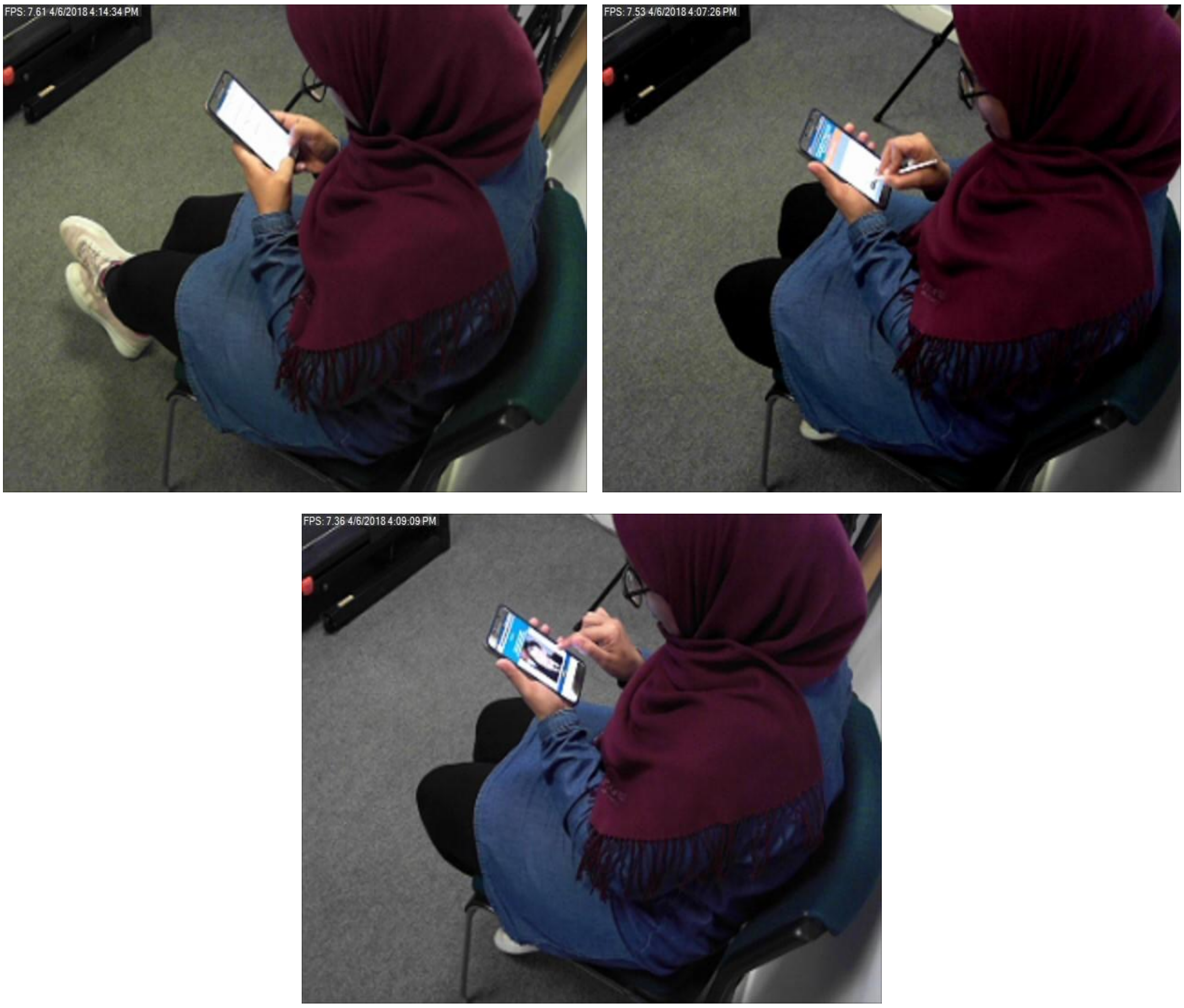

FPS: 7.61 4/6/2018 4:14:34 PM
FPS: 7.53 4/6/2018 4:07:26 PM
FPS: 7.36 4/6/2018 4:09:09 PM

Figure 3.16. Right-handed participant's illustrative hand-postures during - a) typing task (top left), b) signature task (top right) and c) swiping task (bottom center)

This information was acquired in order to assess if the touch behaviour of the left and right-handed participants were different. In order to make this evaluation, the participants were observed during the task executions using video recording. After the completion of tasks, the participants were interviewed to have further understanding on their feedback and experience. Figure 3.15 and Figure 3.16 illustrate examples of images taken from video recording of the hand-postures maintained of a left-handed and a right-handed participant performing the typing and swiping tasks.

Conducting a visual analysis of the video recordings of the device holding postures of 50 participants, the typical behaviour of right-handed users was to switch between different hands based on the requirements of the modality. For instance, a typing task can be performed by either holding the device with both hands and typing with the thumb (as shown in Figure 3.16 a) or holding the device in one hand and typing with the other hand.

3.4.4 Other Parameters

A range of user-based, device-based and environment-oriented characteristics that can potentially impact user verification performance and usability were captured during the data collection.

3.4.4.1 User

The aspect of the user that was controlled during data collection was the walking speed on the treadmill.

- Walking Speed

  In Scenario 2 of Session 1, the participant was asked to perform the experiment while walking on a treadmill. The walking speed was fixed based on the comfort level of the participant and this speed was recorded for every participant. For each participant ID, the speed that was chosen is presented in Figure 3.17. The least speed that the participant chose was 0.3 km/h and the highest speed that the participant chose was 2.8 km/h. The average speed maintained was 1.38 km/h. Figure 3.18 shows a sample image of a participant performing tasks on a treadmill.

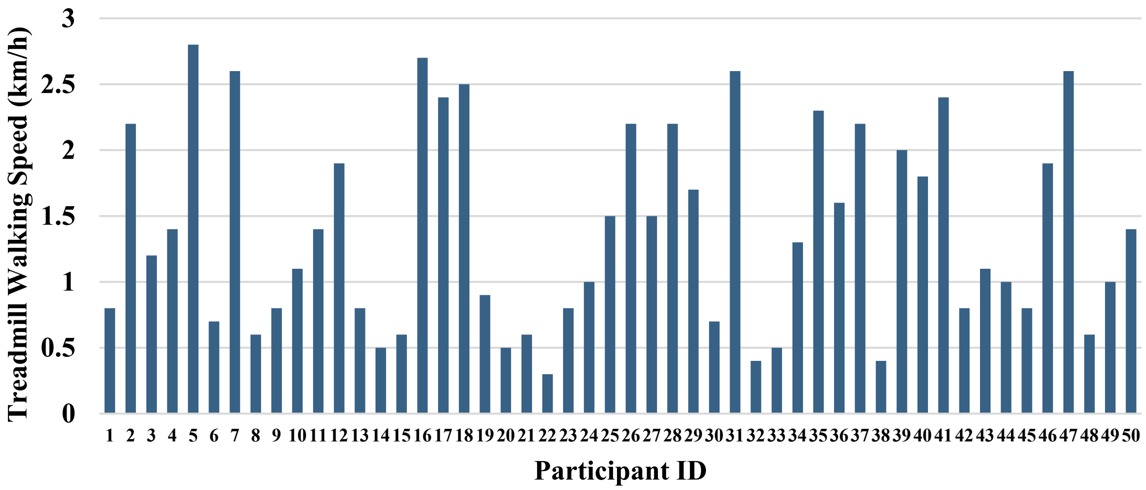


Figure 3.17. Treadmill Speed per Participant ID

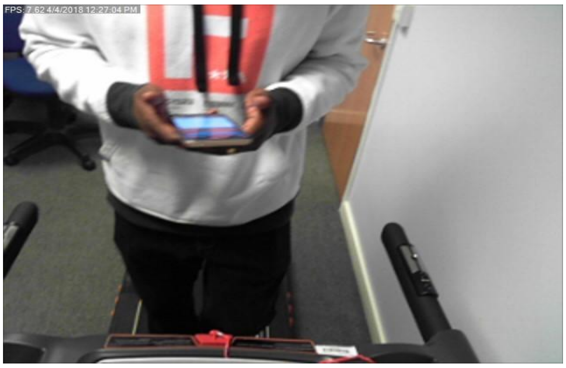


Figure 3.18. Image of a participant performing tasks on a treadmill

3.4.4.2 Device

Location characteristics from the device were captured alongside user interaction during the data collection.

- Location / GPS Sensor

Along with Scenario 1 and 2 which were carried out indoors, Scenario 3 and 4 also recorded the participants location data whilst they were outdoors. As the participant moved around the university campus in Session 1 and Session 2, the in-built GPS sensor was activated and was used to confirm the location and movement of the participant. The GPS location tracker recorded the longitude and latitudinal coordinates. Figure 3.19 shows the GPS coordinates of a participant during Scenario 1, Scenario 2 and Scenario 3 of Session 1 as an example.

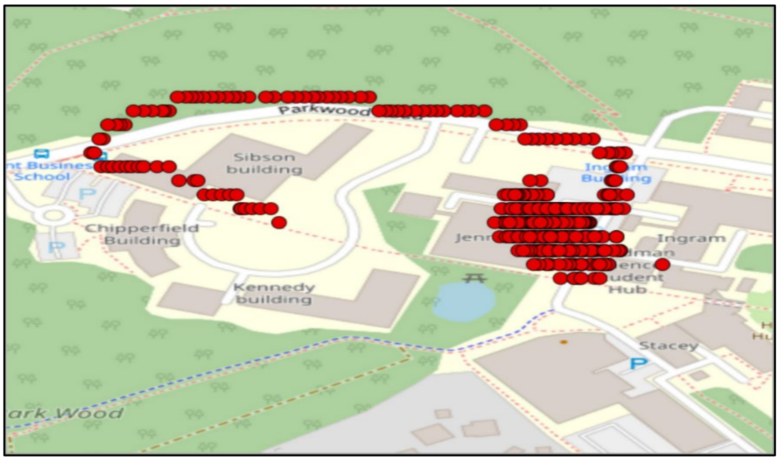


Figure 3.19. Map generated from GPS coordinates of Scenario 1, Scenario 2 and Scenario 3 of a participant from the dataset. Red dots represent the longitude and latitude data points recorded during the experiment

3.5 Conclusion

This chapter details the data collection process adopted for collecting a multi-modal dataset of touch-dynamics based behavioural biometric modalities using a smartphone. The novelty of this work lies in the collection of a real-life dataset containing behavioural biometric data in diverse usage scenarios of a mobile device such as travelling on a bus and walking on a treadmill using a smartphone. The experimental scenarios and protocols were designed to evaluate the factors influencing the user interaction that may cause performance variation during the verification phase.

The dataset consists of behavioural biometric data captured from 50 participants using a Samsung Galaxy Note 5. The dataset extracted data from the touchscreen sensor of the smartphone. The participants had varied ethnic and educational backgrounds. The dataset consists of a combination of right-handed and left-handed participants. The participants in the dataset were owners of iOS and Android devices.

The data collection was performed in two sessions (Session 1 and Session 2) separated by a week. Each session consisted of three scenarios. The scenarios were categorised as static (baseline scenario) and dynamic scenarios. The dynamic scenarios had movements caused either by the user (by way of walking) or the environment (caused by movement of the bus). The scenarios had environmental variations – indoors and outdoors. During the data collection experiment, three different modalities were collected – swipe gestures, signature (stylus and finger-based), keystroke dynamics (alphabetical and numerical). The data capturing method was ceremony-based, as the user was prompted to type a sentence or to sign in the area provided on the device screen.

This dataset was developed to address a number of existing research questions in behavioural biometric domain as highlighted in Chapter 1. This multi-modal and multi-scenario-based dataset has helped in evaluating the robustness of the behavioural biometric verification methods under various usage scenarios of a mobile device. This dataset has been extensively utilised in this thesis work to perform a number of evaluations using multiple classifiers and verification techniques (one-time and continuous verification). A thorough analysis on the impact of user interaction-based factors on the verification accuracy has been performed using this dataset.

Chapter 4. Swipe Gesture Verification

4.1 Introduction

Swipe gesture dynamics exhibited by a user on a mobile device can be utilised to build a behavioural model which can subsequently be used for user verification. There has been an increased number of research studies focusing on swipe gestures based verification in the recent decade [97], [102], [35] and [18]. As described in Chapter 2, one of the key challenges of swipe gesture verification is to ascertain that it can be performed with high verification accuracy across different usage scenarios of a smartphone. After collecting behavioural biometric data from 50 participants through our experiment simulating various real-life scenarios (described in detail in Chapter 3), the next focus of study was to analyse if the verification performance is consistent across these scenarios of a mobile device.

The main goal of this experiment detailed in this chapter is to ascertain the robustness of the swipe gesture verification using a dataset consisting of diverse usage scenarios and using verification performed using different algorithms. During the data acquisition sessions, users were asked to perform the tasks such as interacting with the phone in constrained (laboratory set-up) and in the wild. In order to assess the robustness of swipe gesture-based verification across usage scenarios, an evaluation was conducted using two models - conventional classification methods and DNN. The conventional methods chosen for the analysis were widely used classification algorithms such as SVM, k-NN and Naïve Bayes along with a newer technique using Feed Forward Deep Neural Network architecture. We primarily focused on assessing three important research questions, which were aligned to the overall research objectives provided in Chapter 1, as listed below:

- Analysis of minimum number of swipes required to accurately verify a user
- Evaluation of verification performance using swipe gestures under different usage scenarios of a mobile device – conducted by intra-session comparison
- Evaluation of time persistence of the verification performance – conducted by inter-session comparison

The results of the above mentioned individual research questions are presented in Section 4.5.2 and Section 4.6.3. The outcome of this evaluation raises relevant questions that are detailed in Section 4.7 - Discussion.

4.2 Related Work

Multiple studies have used swipe gesture-based verification in the context of continuous verification and one-time verification methods. These studies have been extensively detailed in Chapter 2 – State of the Art. Considering that this research was focused on assessing the robustness of the verification algorithm across the usage scenarios, studies specifically focusing on this factor have been reviewed in this chapter.

There are limited studies that have considered a variety of usage scenarios of a mobile device. The experiment conducted by Bo et al. [14] explored touch-dynamics for a user in stationary (static) and in motion (walking) scenarios. They intended to capture the tiny perturbation of a mobile device when a user touches it and utilised those features for verification. For the walking scenario, they reported that after two walking steps, the FAR reduced to 0% and after four walking steps, the FRR was 18%. This indicates that body movement has an impact on the verification performance. These results indicate that the movement induced either by the user or the environment can cause performance deterioration. Building upon their study, we decided to not only include a walking scenario but also other scenarios that involve body movement, caused by the user or the environment.

Literature reveals that a number of classifiers have been utilised to conduct swipe-gesture based verification. Amongst emerging studies [120], [121], DNN architectures are being used extensively. Therefore, a decision was made to perform this evaluation using multiple classifiers using the conventional and DNN methods. The obtained results are presented in this chapter.

4.3 Metrics Used

In order to assess the performance of the biometric frameworks, the metrics described in ISO/IEC 30136:2018 [122] have been utilised. Following are the list of metrics used in this thesis:

- False Acceptance Rate - It is defined as the proportion of verification transactions with wrongful claims of identity that are incorrectly confirmed [122].
- False Rejection Rate – It is defined as the proportion of verification transactions with truthful claims of identity that are incorrectly denied [122].
- Equal Error Rate – The point at which the false match rate is same as the false non-match rate.

4.4 Methodology

This section describes the dataset used, pre-processing and feature extraction methods adopted for this analysis. These data processing phases were common for both – the conventional classifiers and the DNN method.

4.4.1 Dataset

Swipe gesture data acquired from 50 users under different usage scenarios has been utilised. This dataset has been described in Chapter 3. Each stroke of swipe generated a list of data points with parameters: timestamp, X-coordinate value, Y-coordinate value, touch action, finger pressure and finger touch area. The touch action parameter consisted of – ACTION-DOWN, ACTION-MOVE and ACTION-UP (in the same order for every swipe). Each swipe stroke was categorised as a horizontal or a vertical swipe. The swipe stroke that has an x major axis deviation higher than y major axis was considered as a horizontal swipe and swipe stroke with deviation of y major axis higher than x major axis was categorised as a vertical swipe.

4.4.2 Pre-Processing

Pre-processing was performed for both the categories of swipes - horizontal and vertical swipes. The pre-processing phase was focused on identifying outliers in terms of lower number of data points and invalid swipe inputs. Swipes containing less than three data points were discarded. We observed that mostly these short swipes were mostly generated through button press. We also identified swipes with no ACTION-DOWN action but having the other touch actions, that is, ACTION-MOVE and ACTION-UP. A majority of such swipes belonged to the vertical swipe category. One possible explanation could be a hardware limitation where the sensor failed to capture ACTION_DOWN for vertical swipes performed very fast and close to each other. These swipes were considered invalid and disregarded from further analysis.

4.4.3 Feature Extraction

For every swipe stroke, a set of 28 features were computed, which are listed in Table 4.1. The features taken into consideration for this analysis were global features, calculated for the entire swipe, such as average velocity and total stroke time. The features ‘Start X Position’ and ‘Start Y Position’ refer to the corresponding X and Y coordinate of ACTION-DOWN action of the swipe. In similar manner, the X and Y coordinate of ACTION-UP were assigned as ‘End X’ and ‘End Y’ features respectively.

A single swipe stroke consisted of multiple ACTION-MOVEs; the distance travelled from one ACTION-MOVE to the next was calculated for the X and Y positions as Delta X and Delta Y. ‘Average Delta X Position’ and ‘Average Delta Y Position’ were calculated by obtaining the average of the calculated Delta X and Delta Y. The ‘width’ of the swipe was calculated as the Euclidean distance from its first touch point (ACTION-DOWN) to the last touch point (ACTION-UP). The slope was calculated as the change in Y position with respect to X position for each data point of a swipe. The maximum and minimum slope values obtained from swipe were assigned as ‘Maximum Slope’ and ‘Minimum Slope’. First and second derivatives of the distance (Delta X and Delta Y) were calculated as Velocity and

Acceleration. From these values, 'Average Velocity' and 'Average Acceleration' were obtained. 'Attack Angle' was calculated as gradient between the first and the second touch data point. Similarly, 'Leaving Angle' was calculated as gradient of the second last and the last data point. The touch finger-size was captured at every touch action point, and 'Average Finger Size' was calculated. These set of global features calculated for every single swipe were fed into the classifiers as inputs.

Identifier	Feature name	Description
1	Start X Position	X-coordinate value of the first position sample value of the swipe
2	End X Position	X-coordinate value of the last position sample value of the swipe
3	Start Y Position	Y-coordinate value of the first position sample value of the swipe
4	End Y Position	Y-coordinate value of the last position sample value of the swipe
5	Average Delta X Position	Average X-position variation. Calculated by finding the difference of X-coordinate move from one sample to the next, until the end Y position of the swipe.
6	Average Delta Y Position	Average Y-position variation. Calculated by finding the difference of Y-coordinate move from one sample to the next, until the end Y position of the swipe.
7	Swipe Height	Height of the swipe
8	Swipe Width	Width of the swipe. Calculated by finding the Euclidean distance from its first touch point (ACTION-DOWN) to the last touch point (ACTION-UP).
9	Total Length	Length of the swipe
10	Mid-Location	Mid-location value of the swipe
11	Minimum Slope	Minimum slope value
12	Maximum Slope	Maximum slope value
13	Total Stroke Time	Total time of the entire swipe
14	Number of data points	Total number of sample data points present in the swipe
15	Average Acceleration	Average acceleration value of the swipe
16	Standard-Deviation	Standard deviation of the swipe
17	Average Finger Pressure	Average finger pressure value
18	Mid Action Pressure	Pressure value of the mid-data point of the swipe
19	Finger Size Finger Down	Finger size captured during ACTION_DOWN
20	Finger Size Finger Up	Finger size captured during ACTION_UP
21	Average Finger Size	Average finger touch area on the screen during the entire swipe
22	Stroke Area Outer	Outer Area of the swipe. Calculated by multiplying the height and width of the swipe
23	Attack Angle	Gradient between the first and the second data point of swipe
24	Leaving Angle	Gradient between the second last and the last data point of swipe
25	Average Velocity	Average value of the velocity of the entire swipe
26	Peak Velocity Value	Peak velocity captured in the entire swipe
27	Mean velocity in first half of swipe	Mean velocity value of the first half of the swipe
28	Mean velocity in second half of stroke	Mean velocity value of the second half of the swipe

Table 4.1. Swipe feature set

4.5 Conventional Classifiers

Three conventional discriminative classifiers - SVM, k-NN and Naive Bayes have been used for the analysis. Python's Scikit-learn library has been utilised for code implementation. The verification method was selected as verification. This is because typically a mobile device is owned by a user and the biometric algorithm is expected to successfully verify the device owner and reject the imposters. Therefore, the context of verification is befitting for the mobile device scenario. This section presents a detailed description of the methodology adopted using the conventional classifiers.

The decision to use these three classifiers has been motivated by multiple reasons. Firstly, this study is aimed at showing the impact of usage scenarios of a mobile device on verification performance using the most proven and widely used classifiers. As shown in the literature (Table 2.3), SVM classifier has been proven widely effective in multiple studies using swipe-based verification [87], [95] and [86]. Secondly, SVM is a powerful classifier used for supervised binary classification problems such as verification, where the model has to accurately classify the genuine and imposter class. The SVM algorithm finds the optimal hyperplane in the N-dimensional feature space that can distinctly classify the data. Additionally, instead of a one-class SVM classifier that utilises only positive samples, we have used a two-class classifier as it is more appropriate for the verification problem under investigation. Regarding the attack model, a *'zero-effort'* attack model has been adopted, where one randomly chosen user from the dataset is considered as an attacker. This attack scenario can be interpreted as an individual attacker getting hold of a mobile device of a genuine user and trying to gain access to app services by forging the genuine user samples.

The model parameters for SVM, k-NN and Naïve Bayes have been carefully chosen to generate a high classification accuracy for the genuine and imposter class. A SVM classifier with a linear kernel has been used with multiple *C* (regularization parameter) values. The '*C'* value is a parameter that controls the trade-off between the decision boundary and the misclassification rate. In order to identify the optimal parameters for the classifier, parameter-tuning using a grid search method has been performed. This has been done using increasing sequences of *C* values (0.01, 1, 10, 100 and 1000). The parameter tuning has been performed for every user model using the baseline scenario *(Scenario 1)* as the enrolment samples always belonged to this scenario. Based on this search, the best value of C has been assigned as 1 for the entire evaluation. A RBF kernel has also been utilised for evaluation. However, as the obtained accuracy rates were not in acceptable range, we have not reported it in this chapter. A possible reason for the poor performance could be that the data is linearly separable and using a RBF kernel forces the classification of data in the mapped hyperplane.

The reasons for choosing k-NN classifier are its fast computation and robustness characteristics. This algorithm is based on the concept that similar features exist in close proximity. Therefore, every incoming swipe stroke is first located in the feature space with respect to the training swipes and, based on the majority of class labels of the k neighbouring training samples, a class is assigned to the incoming stroke. In order to select the k value, multiple runs (minimum of five) with different k values with randomly selected training and validation sets have been performed for every user verification model for all scenarios. Based on the outcome of this analysis, the final k value has been picked as five as it produced the lowest classification/estimation error across different validation sets.

The third classifier used was Naive Bayes which "*assigns the most likely class to a given example described by its feature vector*" [123]. Naive Bayes was chosen, as it is a simple probabilistic model used for classification purposes.

4.5.1 Verification Framework

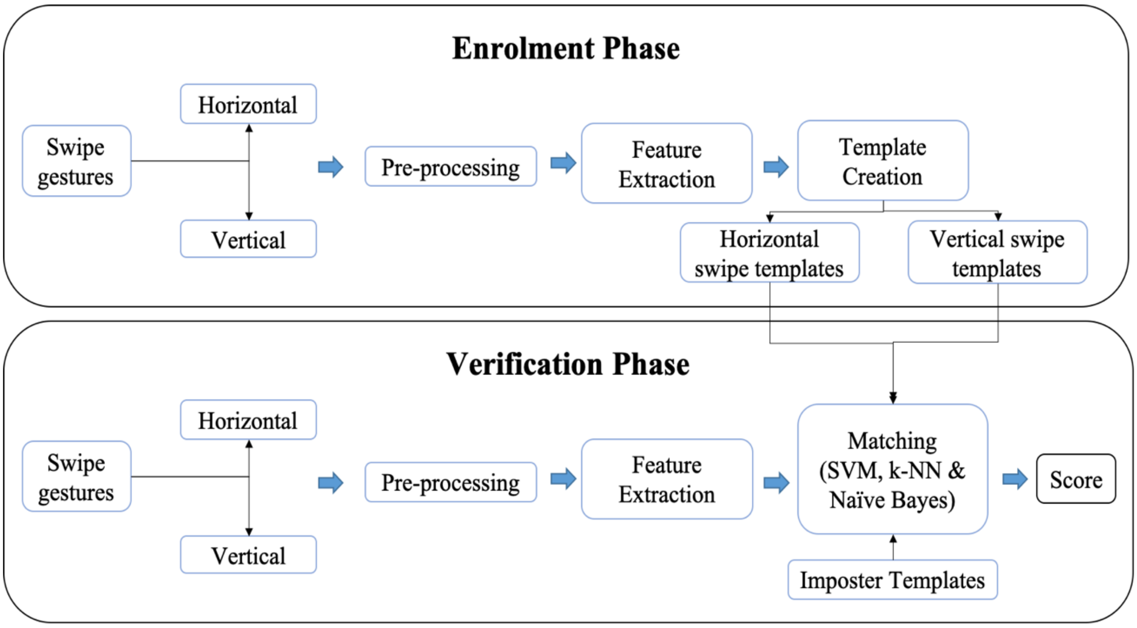


Figure 4.1. Enrolment and verification phases

The data acquisition of the swipe gestures was performed on Samsung Galaxy Note 5 and the verification process was performed off-device. This analysis was based on the hypothesis that there is only one primary user (owner) of a mobile device; therefore, the verification process was designed to verify templates of the primary user. From the dataset consisting of 50 users, a verification model was built for each user. A genuine and an imposter user were selected for each verification model. The imposter user was chosen from the dataset based on the random forgery method. All the imposter samples belonged to that chosen user. The same genuine and imposter user samples were then used in the verification phase.

The model had two phases - enrolment and verification phase. depicts the enrolment and verification phases. The swipe samples used in the enrolment phase were not used in the verification phase. The total number of swipe samples belonging to the genuine and imposter user were split for the enrolment

(25%) and verification phase (75%). From the enrolment sample set, only a few swipe samples were randomly chosen for enrolment. To avoid class bias, an equal number of genuine and imposter samples were chosen for training the user model (for example, if five swipes were chosen from the genuine class, then five samples from the imposter class were chosen for enrolment). The number of swipe samples chosen for enrolment were different based on the research question considered for the analysis (detailed in Section 4.5.2.1).

4.5.1.1 Enrolment

Swipe gesture samples belonging to the genuine and imposter user were enrolled in the model. Both the categories of swipes (horizontal and vertical) were enrolled separately and the same category of swipes were compared during the verification phase. The enrolment strategies employed to carry out the analysis for each of the research questions are outlined below.

- In order to find an optimum number of enrolment samples, the model was enrolled with 2, 4, 6, 8, 10 and 12 randomly chosen swipe samples belonging to the genuine from the baseline scenario (Scenario 1) from Session 1. The results obtained using a range of enrolment samples are explained in Section 4.5.2.1 – '*Analysis of minimum number of swipes required in enrolment to verify accurately'*.
- For intra-session evaluations (comparison between scenarios within each session), the user verification model was enrolled with the swipe gestures captured in the baseline scenario (Scenario 1) and was verified against the swipe samples captured on different scenarios Scenario 2 and Scenario 3. The number of swipe samples enrolled were 2, 4, 6, 8, 10 and 12.
- For inter-session evaluations (between Session 1 and Session 2), swipe samples acquired in Session 1 were enrolled and verified against swipe samples 2 belonging to the same usage scenario from Session. The number of swipe samples enrolled were 2, 4, 6, and 10.

4.5.1.2 Verification

For verification, the genuine and the imposter samples belonged to the same genuine and imposter user respectively which were used in the enrolment phase. The number of swipes used in the verification phase were around 50 for each comparison. All of these swipes belonged to the pool of 75% swipe samples earmarked for verification at the beginning. During the verification process, the incoming swipe stroke was first classified as a horizontal or a vertical swipe. Following this, pre-processing and feature extraction steps were conducted. Based on the swipe category, a user template (horizontal or vertical) was selected and compared with the incoming swipe. In the matching phase, probability similarity scores were generated by comparing it to the identified template class. This process was carried out for swipes from genuine as well as imposter users. Based on the generated scores, FAR and

FRR were calculated for different thresholds. Further, the EER for both types of swipes, horizontal and vertical, were obtained.

4.5.2 Results

The performance evaluations were carried out for four different purposes: a) to analyse the minimum number of swipes required in enrolment to accurately verify a user, b) to analyse stability of swipe gestures across different usage scenarios, c) to analyse stability of swipe verification over time and d) to evaluate stable features across usage scenarios. The results obtained for each of these research questions are explained in detail in this section.

4.5.2.1 Analysis of minimum number of swipes required in enrolment for accurate verification

In order to investigate this factor, the user model was enrolled with a different number of swipe samples during the enrolment phase, while remaining swipe samples were used for verification. It was expected that the verification accuracy would improve with an increase in number of enrolled swipes. Figure 4.2, Figure 4.3 and Figure 4.4 show the mean EER obtained from the verification model of 50 users using different enrolment samples for horizontal and vertical swipes separately using SVM, k-NN and Naïve Bayes algorithms.

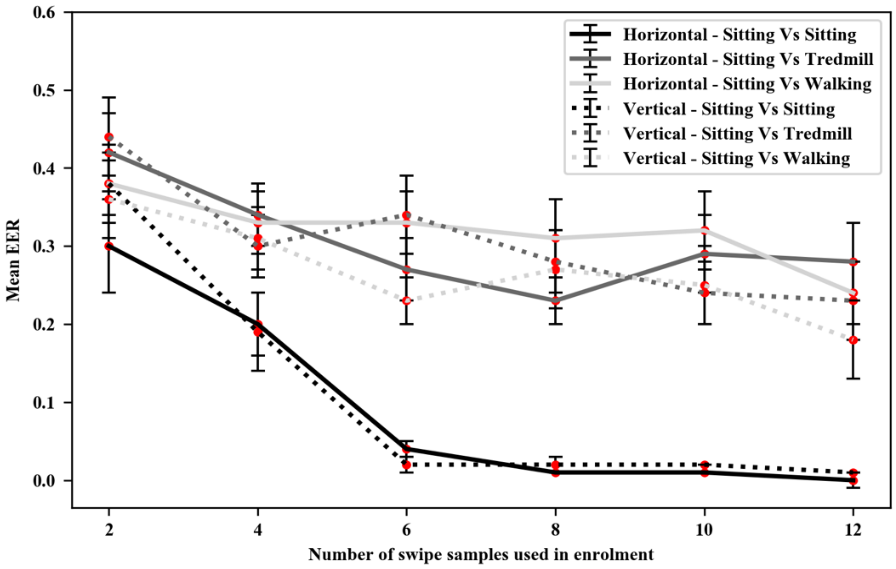


Figure 4.2. Impact of varying number of enroment samples on mean EER using SVM (Intra-session comparison results of Session 1)

It can be observed that with an increase in the number of enrolled swipes, the mean EER becomes significantly low. The trend of decrease in mean EER rate with increased enrolment samples can be seen across different usage scenarios and classification algorithms. For example, using the SVM classifier, with 12 swipes in enrolment, the mean EER attained is 1% for horizontal and 2% for vertical swipes. In going from four enrolled swipe strokes to six swipe samples in the enrolment, the mean EER value drops from 20% to 2% for the horizontal and 19% to 1% for the vertical swipes. The inter-session analysis (Figure 4.5) also reveals a similar trend of decrease in EER rate with increased enrolment samples. According to the results acquired, it can be concluded that a minimum of 6 swipes are required to attain an acceptable verification accuracy.

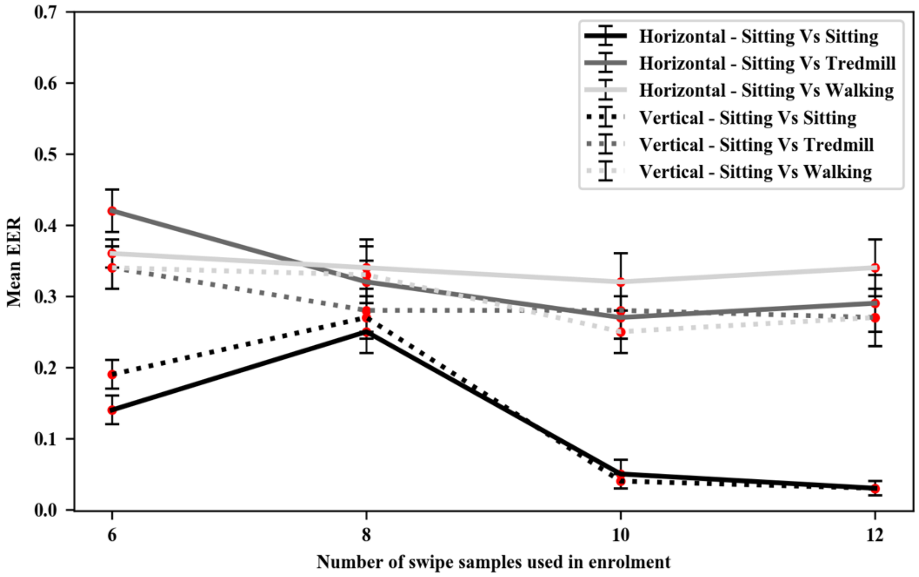


Figure 4.3. Impact of varying number of enroment samples on mean EER using k-NN (Intra-session comparison results of Session 1)

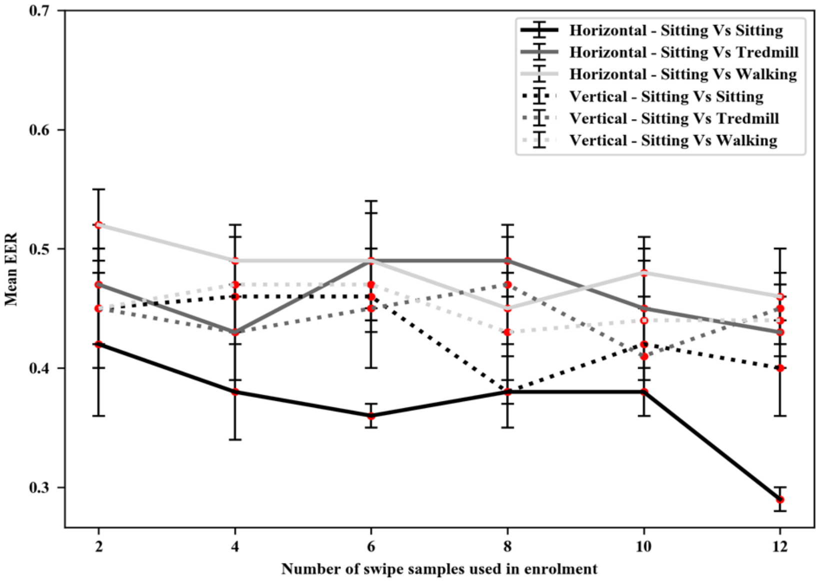


Figure 4.4. Impact of varying number of enroment samples on mean EER using Naive Bayes (Intra-session comparison results of Session 1)

4.5.2.2 Performance analysis across different usage scenarios (Intra-session Analysis)

The intra-session analysis was performed on swipe gestures captured on the same day but using different usage scenarios. The results were obtained individually for Session 1 and Session 2. For this analysis, a number of comparison strategies with respect to enrolment and verification were implemented which are detailed in Table 4.2. Each of the scenario refers the following:

- Scenario 1 – Seated on a chair, indoors
- Scenario 2 – Walking on a treadmill, indoors
- Scenario 3 – Walking outdoors
- Scenario 4 – Travelling on a moving bus

Session 1		**Session 2**	
Enrolment	**Verification**	**Enrolment**	**Verification**
Scenario 1	Scenario 1	Scenario 1	Scenario 1
Scenario 1	Scenario 2	Scenario 1	Scenario 3
Scenario 1	Scenario 3	Scenario 1	Scenario 4

Table 4.2. Details of enrolment and verification dataset for intra-session evaluations

For Session 1, the user model was enrolled with swipe samples captured during Scenario 1, and the verification swipe samples were taken from Scenario 1, Scenario 2 and Scenario 3 (captured on the same day during Session 1). For Session 2, the user model was enrolled with data from the Scenario 1 and verified against the Scenario 1, Scenario 3 and Scenario 4 (captured on the same day during Session 2).

Scenario		**SVM**		**k-NN**		**Naïve Bayes**	
Enrolment	*Verification*	*Session 1*	*Session 2*	*Session 1*	*Session 2*	*Session 1*	*Session 2*
Horizontal Swipes							
Sitting	Sitting	1(3.0)	0(0.0)	25(19.0)	21(14.0)	38(2.0)	41(30.0)
Sitting	Treadmill	23(25.0)	N/A	32(23.0)	N/A	49(21.0)	N/A
Sitting	Walking	31(31.0)	27(30.0)	34(24.0)	33(23.0)	45(19.0)	50(21.0)
Sitting	Bus	N/A	30(30.0)	N/A	34(27.0)	N/A	44(19.0)
Vertical Swipes							
Sitting	Sitting	2(4.0)	1(2.0)	27(17.0)	29(14.0)	38(2.0)	33(18.0)
Sitting	Treadmill	28(27.0)	N/A	28(23.0)	N/A	47(19.0)	N/A
Sitting	Walking	27(31.0)	23(30.0)	33(25.0)	27(27.0)	43(21.0)	46(16.0)
Sitting	Bus	N/A	26(28.0)	N/A	25(25.0)	N/A	49(21.0)

Table 4.3. Performance of the intra-session evaluation - mean equal error rate % (standard deviation) across users with 8 genuine swipe samples used in the enrolment dataset

The results of individual comparisons for Session 1 and Session 2 using Linear SVM classifier, k-NN and Naïve Bayes classifiers are given in Table 4.3. It can be observed that the SVM algorithm produced lowest EERs, followed by k-NN and Naive Bayes algorithms for all the scenarios. It can also be seen that Naive Bayes algorithm shows the worst performance with mean EERs ranging from 33% and above, even for the Scenario 1 versus Scenario 1 comparison. The important factor to notice is that using SVM, the static scenario – Scenario 1 Vs Scenario 1 (having no body movement) in Session 1 showed an EER of 1% and 2% for horizontal and vertical swipes respectively. On the contrary, the mean EER's obtained for the dynamic scenarios - Scenario 1 Vs Scenario 2 (Session 1), Scenario 1 Vs Scenario 3 (Session 1 and 2) and Scenario 1 Vs Scenario 4 (Session 2) are significantly higher at 23%, 31%, 27% and 30% respectively for horizontal swipes and 28%, 27%, 23% and 26% for the vertical swipes. A similar trend of increased EERs can be seen across k-NN and Naive Bayes for scenarios involving any body movement - caused by either users or environmental factors. The Scenario 1 Vs Scenario 4 (Session 2) and Scenario 1 Vs Scenario 3 acquired similar EER's. Using k-NN, EER's acquired for horizontal swipes were 33% and 34%, and using Naive Bayes were 50% and 44% for the Scenario 1 vs Scenario 3 and Scenario 1 Vs Scenario 4 comparisons respectively.

Further, to verify that above results were not just a chance occurrence, two-tailed statistical significance tests were conducted. The purpose of performing these tests was to evaluate the hypothesis that acquired mean EER's for static and dynamic scenarios are different. The null hypothesis and alternative hypothesis were considered for each of these intra-session comparisons. The alternative hypothesis chosen was a two-sided hypothesis claim as shown in the equation given below:

$$Null\ Hypothesis \Longrightarrow [\, H_0 \,:\, \mu_{Scenario1} = \mu_{Scenario3}] \tag{4.1}$$

$$Alternative\ Hypothesis \Longrightarrow [\, H_1 \,:\, \mu_{Scenario1} \neq \mu_{Scenario3}] \tag{4.2}$$

Scenario	P-value
Sitting Vs Walking (Session 1)	0.008
Sitting Vs Treadmill	0.038
Sitting Vs Walking (Session 2)	0.040
Sitting Vs Bus	0.021

Table 4.4. P-value of the statistical significance tests

In order to perform these significance tests, EER's obtained for the static (population group $\mu_{Scenario1}$) and dynamic scenarios (population group $\mu_{Scenario\ 3}$) were randomly chosen. For a given hypothesis test, 'α' denotes significance level. For these tests, the 'α' value was set as 0.05. A P-value was calculated for individual statistical significance tests performed between different scenarios. As shown in , all calculated P-values were below the significance level α, therefore, the null hypothesis was rejected. These results once again ascertain the hypothesis that there is a difference in performance of swipe gesture verification for a static scenario versus dynamic scenarios.

4.5.2.3 Performance analysis for time-separated swipe gestures (Inter-session Analysis)

For this analysis, swipe gestures obtained on different days/sessions (separated by a week) but under the same usage scenario were compared. The user model was enrolled with swipe gestures taken from the Scenario 1 (Sitting) of Session 1 and verified against the Scenario 1 of Session 2. Similarly, the comparison was made for Scenario 3 (walking). As seen in Figure 4.5, with ten swipe samples in enrolment, an EER of 44% was obtained for horizontal swipes and 39% for vertical swipe for Scenario 1 comparison. On the other hand, an EER of 32% for horizontal swipes and 16% for vertical swipes were obtained for inter-session comparison for Scenario 3 (walking outdoors). This highlights that the EER's gets considerably worse for inter-session comparisons, which raises questions about the stability of the swipes over time. Considering that the same mobile device was used in both sessions, there is merit in further investigating if this variability in the EER's could be associated with the users' behaviour. From the data, it was noted that some users had large variance in feature set acquired from Session 1 and Session 2 compared to others. Therefore, it is possible that the user behaviour for some users are similar over time compared to others. This would need further research with a larger sample size and well-defined used behaviour scenarios to ascertain impact of this factor.

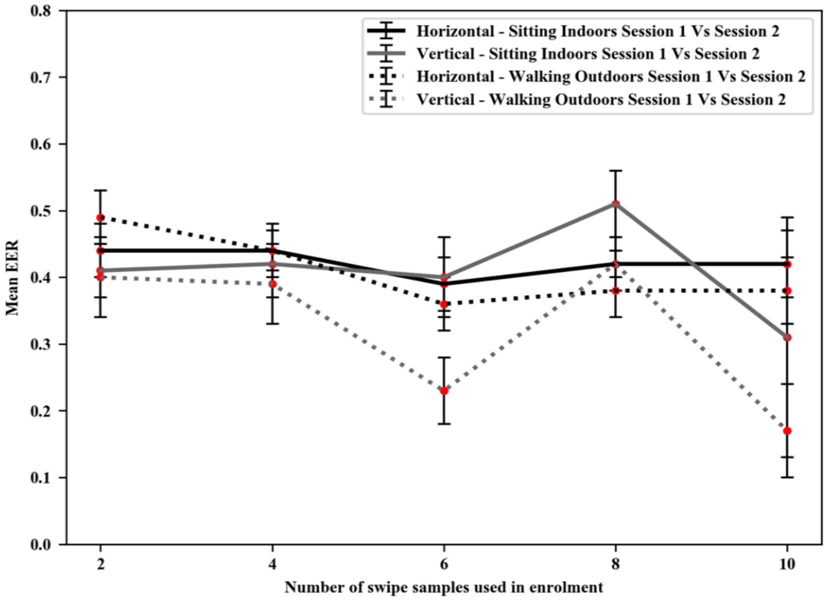


Figure 4.5. Performances in inter-session analysis using SVM

4.6 Deep Neural Network

To complement the performance assessment of the swipe gesture-based verification using the conventional techniques, a deep neural network model was developed and utilised for the same analysis. As described in Section 4.1, one of the goals of this analysis is to assess the performance of swipe gesture verification on diverse usage scenarios of a mobile device, therefore, in this section the intra-session and inter-session comparisons performed on the dataset have been presented in detail. The methodology and model architecture adopted along with the outcomes of the evaluation are also presented in this section.

4.6.1 Architecture

In order to perform this analysis, a feed-forward deep neural network model was built and utilised for the analysis. The next subsections explain the data used, the network architecture, enrolment and verification process and the results obtained.

The input data used for this analysis was the same as the data used for the conventional techniques. The input data consisted of swipe strokes of genuine and imposter users. For each swipe stroke, a 28-set feature was generated and fed into the feed-forward deep neural network. These features were global features generated for both the categories of swipes - horizontal and vertical swipes (feature-set is described in Section 4.4.3). The reason for using the feature-set as an input instead of raw data (timestamp, X and Y coordinates) was to conduct the evaluation with the same input data as for the conventional method.

A deep neural network is defined as an artificial neural network consisting of multiple layers between the input and output layer. A deep neural network usually contains two or more layers between the input and output layers. The idea behind a feed-forward neural network is that the information or computation progression is in the forward direction only and the data flow does not form a circle, that is, data flows from the input nodes, through the hidden nodes and finally to the output nodes. The reasons to have chosen a feed-forward neural network are simpler training and shorter convergence time. For this analysis, a multi-layered perceptron has been utilised in order to deal with data separation in supervised learning. MLP’s have at least one hidden layer that consists of multiple perceptrons. The architecture built for this analysis is shown in Figure 4.6.

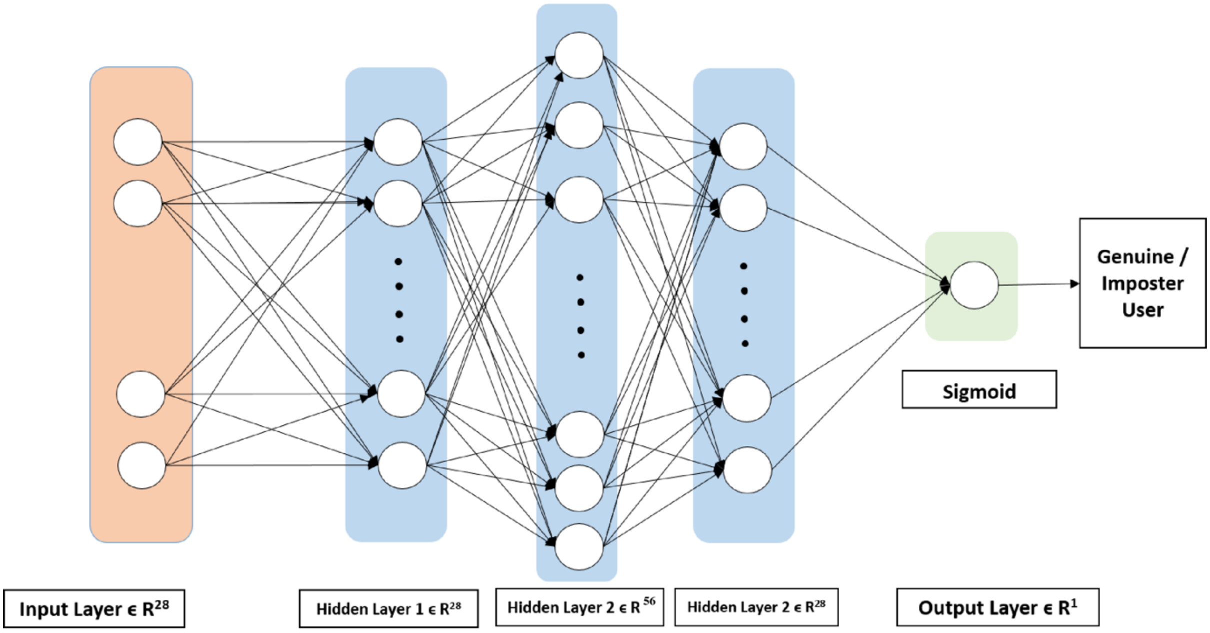


Figure 4.6. Feed-forward Deep Neural Network architecture

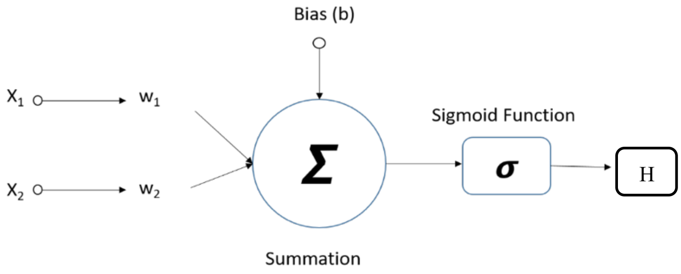


Figure 4.7. Sigmoidal unit with two inputs (X_1 and X_2), weight vectors W_1 and W_2 and b as bias

It is a fully connected MLP on 28 input perceptrons with three hidden layers, each with 28, 56 and 28 perceptrons respectively. The output of each perceptron is 0 or 1. Every perceptron has an input layer and output layer. Each sigmoidal unit/perceptron has a structure as shown in Figure 4.7. A neuron-based model is a linear classifier that splits the input data into two with a linear boundary. The *weight* function is defined as the slope of the linear boundary and the *bias* is the intercept. The linear boundary of the sigmoid function can be adjusted using the weight (w and b) and the bias values. The output of a perceptron is calculated as the sum of product of weight with corresponding input unit plus the bias.

$$H = (w * x + b) \tag{4.5}$$

The value of H is determined as following:

$$H(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \tag{4.6}$$

4.6.1.1 Error Function

The error function (*E*) needs to be defined in the learning process of the network. This function quantifies the true value versus the computed value of the output. The goal of the network is to keep the computed error value close to zero. Therefore, the error function is minimised using the weights and biases in order to get a good classification boundary in the network.

In order to minimise the error function, typically, '*Gradient Descent'* function is used. It is an iterative function that converges the value of the error to a local minimum. $\vec{w}$ and $\vec{b}$ values are randomly set and updated using gradient descent. The initializations of $\vec{w}$ and $\vec{b}$ are denoted $\overrightarrow{w_0}$ and $\overrightarrow{b_0}$, and the gradient descent updates of $\vec{w}$ and $\vec{b}$ are calculated as:

$$\mathrm{W}_{\mathrm{i}+1} = \mathrm{w}_{\mathrm{i}} - \alpha \frac{\partial \mathrm{E(x)}}{\partial \mathrm{w}_{\mathrm{i}}} \tag{4.7}$$

$$\mathrm{b}_{\mathrm{i}+1} = \mathrm{b}_{\mathrm{i}} - \alpha \frac{\partial \mathrm{E(x)}}{\partial \mathrm{b}_{\mathrm{i}}} \tag{4.8}$$

Here, W_i and b_i are the values of weight and bias after the i^{th} iteration of the gradient descent and the learning rate is represented as α. The 'α' value is usually chosen as a small value such as 0.001. It is the value of the step size that gradient descent takes in each iteration.

In order to compute the output, the algorithm undergoes multiple iterations. Each iteration is divided into two steps:

Step 1. Calculate the forward values. For a given input $X = \{(x_1, y_1),\ldots\ldots(x_N, y_N)\}$, calculate the h value (product sum plus bias for perceptron) and o value (output of node)

Step 2. According to the gradient descent, calculate the error function and update the weights and biases.

Figure 4.8. Algorithm steps for each iteration in DNN model

Each layer calculates the output in the following steps:

```
Step 1. Input layer initialisation. The outputs of the input vector $x = \{x_1, x_2.....x_n\}$ for each
    node is set as $o_i^0 = x_i$.
Step 2. Calculate product sum 'h' and output 'o' for every layer ($l_1 – l_{m-1}$).

        For k from 1 to m-1:
        a. compute $h_i^k = w_i^k \bullet o^{k-1} + b_i^k$

        b. compute $o_i^k = g(h_i^k)$, for $i = 1, ...., r_k$

Step 3. Compute output y for output layer.
```

Figure 4.9. Algorithm steps carried out at each layer of the DNN model

4.6.1.2 Network Parameters

The implementation was carried out using Python's TensorFlow library [124]. The parameters that were set for the network are provided below.

- ***Depth of the network*** - The depth is defined in the case of feed-forward neural networks as having multiple nonlinear layers between input and output. The network consists of three hidden layers. The first layer has one-to-many mapping and contains 28 nodes. Next, we have a fully connected layer with 56 nodes, followed by another fully connected layer of 28 nodes. Lastly, an output layer utilises the sigmoid function to produce a binary output.
- ***Learning Rate of the network -*** The learning rate was set as 0.001.
- ***Batch size*** - The batch size of the input was chosen as 25 for the Deep Neural Network (DNN) framework**.**
- ***Epochs*** – The training epochs were set to 100.

4.6.2 Verification Framework

The deep neural network model was designed for every individual user in the dataset, i.e., 50 models pertaining to 50 users. The verification model consisted of one genuine user and one imposter user. The genuine swipe samples belonged to the genuine user (owner of the phone) and the imposter samples belonged to a randomly chosen imposter user from the dataset. The same random-forgery method utilised in the conventional technique has been adopted in the deep neural network method as well.

4.6.2.1 Enrolment

The enrolment of the model was performed on horizontal and vertical swipes separately. From the total number of swipe strokes of a given user, a proportion of 50% were utilised for enrolment purposes and the remaining 50% of the swipes were utilised for the verification purposes. In order to simulate a real-life scenario, the enrolment samples were taken from the first few sample donations. The first sample batch (acquired in the beginning of the experiment) were chosen for enrolment and the second batch (swipe samples acquired after the first batch) were chosen as the verification set. Similar selection was performed for the swipe strokes of the imposter user. Python's '*train_test_split'* method was utilised to split the dataset into enrolment and verification subsets with stratified shuffle method.

The number of samples chosen for enrolment were different for each research question that was being addressed (listed as following).

- In order to find an optimum number of enrolment samples, the number of swipe samples set for the enrolment were - 6, 8, 10 and 12. As the network required a minimum number of swipe samples for training, the number of swipe samples started from 6.
- Intra-session analysis – Swipe samples from Scenario 1, Scenario 2 and Scenario 3 were chosen for enrolment.
- Inter-session analysis – Swipe samples from Scenario 1 (Session 1) and Scenario 3 (Session 1) were used for enrolment.

4.6.2.2 Verification

The verification data belonged to the verification subset that was separated before the enrolment phase. In this phase, the incoming swipe sample was pre-processed, and a 28-feature set was generated for the incoming swipe. This feature set of the swipe sample was fed into the deep neural network and the incoming swipe sample was compared with the enrolled swipe samples to generate a predicted class and a probability score. The predicted class is compared against the actual class label to generate the FNMR, FMR and EER.

The number of verification swipe samples chosen for the verification phase were selected based on the research question at hand:

- In order to find an optimum number of training samples, the verification samples were randomly chosen from the remaining swipe samples of the genuine user from the verification set.
- Intra-session analysis – The swipe samples from *Scenario 1*, *Scenario 2* and *Scenario 3* scenarios belonging to the verification set were chosen for verification.

- Inter-session analysis – The swipe samples from *Scenario 1* (*Session* 2) and *Scenario 3* (*Session* 2) were used for verification.

4.6.3 Results

The results obtained using the DNN method for individual research question have been detailed in this section.

4.6.3.1 Number of swipe gestures required to attain optimal accuracy rate

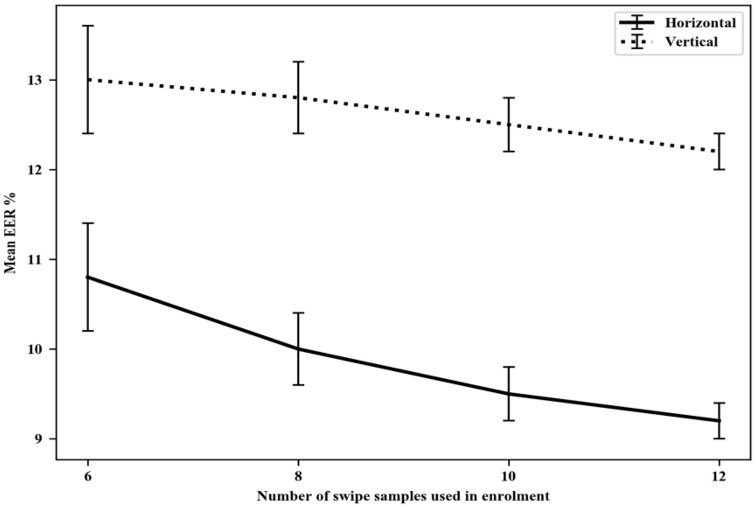


Figure 4.10. Impact of enrolment samples on EER using DNN

Figure 4.10 shows the impact of enrolment samples on the equal error rate. The number of swipe samples chosen for the enrolment were 6, 8, 10 and 12 for every single user. The figure shows the mean EER obtained from the verification model of 50 users from the dataset. The mean EER % has been calculated for the horizontal and vertical swipes separately. Based on the acquired result, the mean EERs for horizontal swipes were lower compared to the vertical swipes. It can be observed in the figure that with the increase in number of swipes in the enrolment, the mean EER value decreases. This indicates that the more the enrolment data, better the performance. The least mean EER attained using 12 samples was 9.2% for horizontal swipes and 12.6% for vertical swipe.

4.6.3.2 Intra-session Comparison

The intra session comparison was performed within-session scenarios for Session 1 and Session 2 individually. For Session 1, the scenarios were Sitting Indoors (Scenario 1), Treadmill (Scenario 2) and Walking Outdoors (Scenario 3). The verification was performed using different enrolment and verification combinations (provided in Table 4.5).

	Enrolment	Verification Scenario
Session 1		
1	Sitting Indoors (Scenario 1)	Sitting Indoors (Scenario 1)
2	Treadmill (Scenario 2)	Treadmill (Scenario 2)
3	Walking Outdoors (Scenario 3)	Walking Outdoors (Scenario 3)
4	Mixed Samples from all scenarios	Mixed Samples from all scenarios
5	Sitting Indoors (Scenario 1)	Treadmill (Scenario 2)
6	Sitting Indoors (Scenario 1)	Walking Outdoors (Scenario 3)
Session 2		
7	Sitting Indoors (Scenario 1)	Sitting Indoors (Scenario 1)
8	Walking Outdoors (Scenario 3)	Walking Outdoors (Scenario 3)
9	Travelling on a Moving Bus (Scenario 4)	Travelling on a Moving Bus (Scenario 4)
10	Mixed Samples from all scenarios	Mixed Samples from all scenarios
11	Sitting Indoors (Scenario 1)	Walking Outdoors (Scenario 3)
12	Sitting Indoors (Scenario 1)	Travelling on a Moving Bus (Scenario 4)

Table 4.5. Enrolment and verification strategy for scenarios

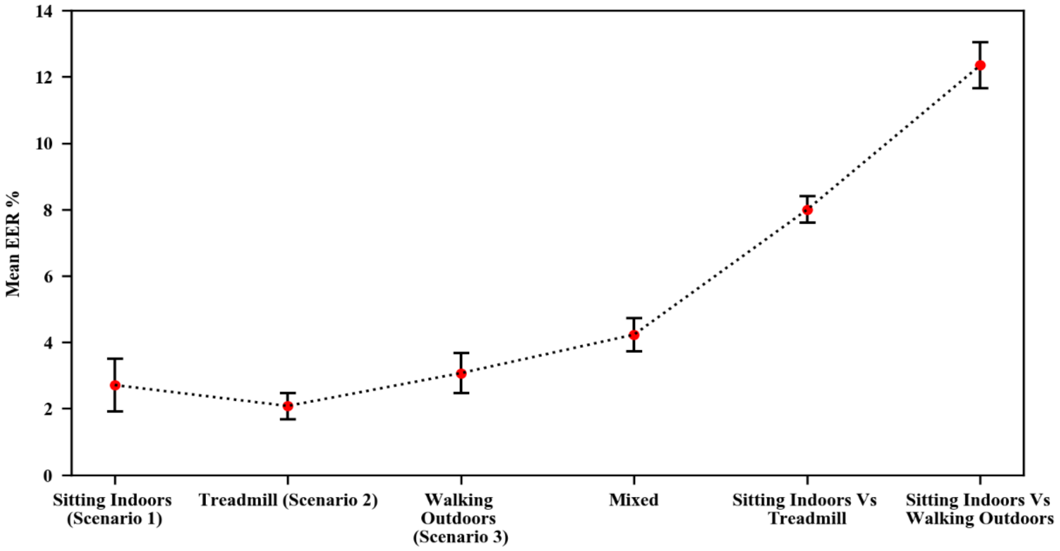


Figure 4.11. Intra-session comparison results for Session 1

The results attained using these strategies are provided for Session 1 and Session 2 in Figure 4.11 and Figure 4.12 respectively. These figures show that if the swipe gesture samples belonging to the same scenario were used during the enrolment and verification phases, the average EER obtained is comparatively lower when compared to the samples coming from different scenarios in enrolment and verification phases. For instance in Figure 4.11, for the Sitting Indoors, Treadmill and Walking Outdoors scenarios, the mean EERs obtained are 2.71%, 2.08% and 3.07% respectively compared to Sitting Vs Treadmill (8%) and Sitting Vs Walking (12.35%).

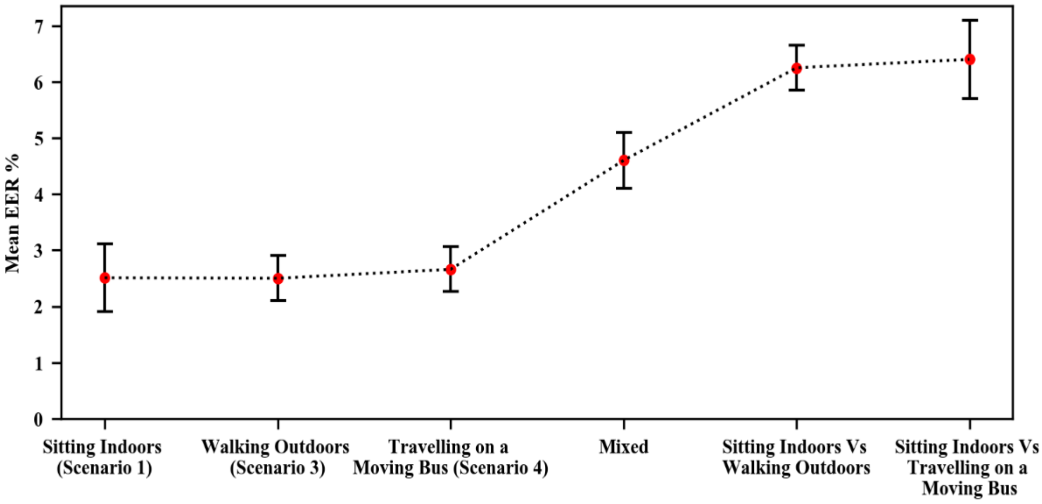


Figure 4.12. Intra-session comparison - Session 2

It can also be noted from Session 1 results, that for the swipe samples coming from a mixed sample set from all the three scenarios, the mean EER obtained was 4.23%, while the Sitting Vs Treadmill was almost double and Sitting Vs Walking EER was three times higher. A similar trend of performance deterioration can be seen for Session 2 results when swipe samples belonging to different scenarios were compared. When the enrolment and verification samples belonged to the same usage scenario, the obtained EER's were in similar range for all the three scenarios of Session 2 - Sitting Indoors (2.51%), Walking Outdoors (2.5%) and Travelling on a Moving Bus (2.66%). However, the performance deteriorates when the enrolment and verification samples belonged to different scenarios - Sitting Indoors Vs Walking Outdoors (6.25) and Sitting Indoors Vs Travelling on a Moving Bus (6.4). For both the sessions, the intra-session comparison shows that when the swipe data come from the same scenario, the verification performance is better compared to the inter-scenario comparison.

4.6.3.3 Inter-session Comparison

In order to conduct the inter-session comparison, the horizontal swipe samples from Scenario 1 of Session 1 were compared against the Scenario 1 of Session 2 and Scenario 3 of Session 1 were compared against Scenario 3 of Session 2 (Figure 4.13). Based on the results obtained, the average EER attained for Scenario 1 was 11.96% and for the Scenario 3 was 7.8%. The verification performance of the intra-session analysis showed better results compared to inter-session comparison. However, the Walking Outdoors scenario obtained a better EER compared to the Sitting Indoors scenario in this inter-session analysis.

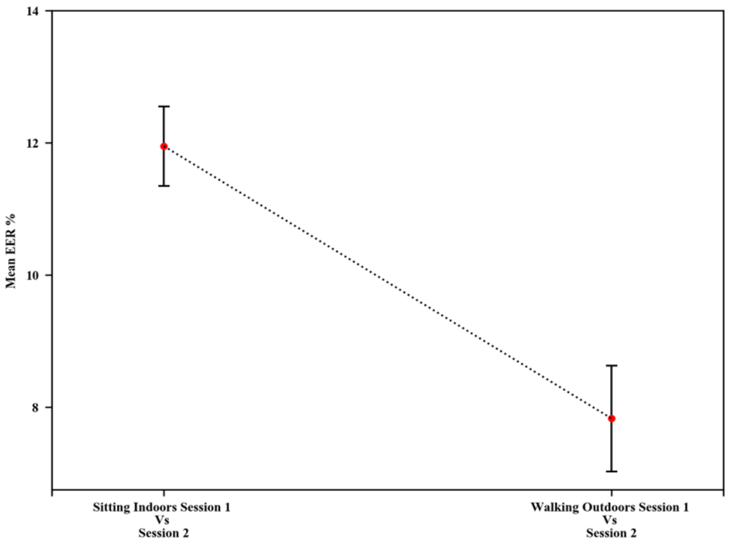


Figure 4.13. Inter-session comparison of horizontal swipes of Session 1 versus Session 2

4.7 Discussion

Unlike traditional biometrics, mobile biometrics provide flexibility for the users to carry out verification on devices anywhere and everywhere, thereby also introducing additional challenges to address. One of the key challenges is to provide stable verification across usage scenarios. The outcome of performance evaluations across usage scenarios for swipe gestures-based verification reveal a significant difference in verification accuracy for a stationary scenario (Sitting Indoors) compared to scenarios with body movement.

One possible reason for the variation in verification accuracy could be the selection of enrolment swipe samples. The enrolment swipes for all evaluations were taken from baseline scenario (Sitting Indoors) and were captured under controlled settings. However, the verifications were done on swipe gesture samples that came from uncontrolled data captured using scenarios with or without body movement. Given the promising verification rate of Scenario 1 Vs Scenario 1 using the SVM and the DNN methods, it is likely that the verification accuracy may improve if the enrolment and verification samples are always coming from the same usage scenario. This suggests that there is a need to choose appropriate templates for swipes to improve verification accuracy. A template selection strategy to dynamically choose appropriate enrolment swipe gestures based on movement and non-movement scenarios can be developed. However, such a technique would mean that the enrolment process would require the user to provide swipes for different scenarios, thereby implying additional efforts from the user. In addition, such a strategy would raise additional concerns such as validity of the enrolled swipes (do the enrolled swipes need replacement? If so, how frequently the enrolment swipes have to be replaced) and the extent to which usage scenarios need to be defined.

For swipe gesture verification, evaluating the verification persistence is another key challenge. The experimental results obtained using conventional classifiers and DNN method show that intra-session comparisons (performed on the same day) are more acceptable than inter-session comparisons (performed on different days of a week). A possible reason could be that the user behaviour is more stable on the same day compared to over a whole week. Nonetheless, a significant difference in the verification performance for inter-session comparisons raises questions about the longevity of this behavioural feature. Hence, the concept of 'one-time enrolment' may have to be investigated for swipe-based verification.

Another key aspect of swipe-based verification is usability. From this perspective, the idea of attaining higher accuracy rates based on minimal data is highly attractive. The results of this analysis show that there is a need to have at least six swipes in enrolment using the conventional classifiers and the DNN method. However, a standardised method to qualitatively select these enrolment swipe samples needs to be established. Considering that the swipe gesture based verification is silent and non-intrusive to the user, there is a need to identify if the enrolment data would consist of the first few swipes exhibited by the user or would be qualitatively selected from a sequence of swipes acquired over a period of time.

One of the limitations of this study is that the analysis has been done on swipe data acquired only in portrait mode and using one device model (Samsung Galaxy Note 5). It would be interesting to conduct a similar analysis across multiple devices with different screen-sizes and including landscape mode which could highlight further challenges with regards to interoperability. Future work would include development of template selection strategies, studying the impact of usage scenarios across multiple devices, orientation types (portrait and landscape) and while using multiple fingers.

Another noteworthy aspect of this study is that the EER's acquired for the dynamic scenarios using the state-of-the-art methods are not in an acceptable range to be adopted as a usable application. However, this work is only a proof-of-concept that demonstrates the impact of dynamic scenarios on the verification accuracy. Based on the results, we are concluding that the verification performance is negatively impacted by movement of either the subject or the environment (compared to static scenarios). Thus, to achieve an acceptable EER in dynamic scenarios, developing a multi-modality approach that combines data from other sensors on the mobile device could possibly improve the verification accuracy. Additionally, we would like to point out that these results have been achieved with a limited dataset of only 50 users. Carrying out similar analysis with more data points and a greater number of users would further confirm the impact of usage scenarios on the verification performance.

4.8 Conclusion

This analysis aimed at evaluating the swipe gesture verification across various usage scenarios of a mobile device. The evaluation was performed on a touch-dynamics based dataset captured under four scenarios - the user seated on a chair, the user walking on a treadmill, the user walking outdoors and the user sitting on a bus.

Three conventional classification algorithms - SVM, kNN and Naive Bayes and a feed-forward neural network were used for the analysis. The intra-session evaluation results obtained using a linear SVM classifier showed the best performance with a mean EER of 1% for horizontal swipes and 2% for vertical swipes when the enrolment and verification swipes were belonging to the static scenario (Sitting Indoors). However, the mean EER grew significantly when the enrolment and verification samples were belonging to different scenarios. The results obtained a mean EER of 23% (Sitting Indoors Vs Treadmill) and 31% (Sitting Indoors Vs Walking Outdoors) in Session 1 and 23% (Sitting Indoors Vs Walking Outdoors), 26% (Sitting Indoors Vs Travelling on a Moving Bus) in Session 2. This significant rise in mean EER values for dynamic scenarios were seen across all three classification algorithms and DNN method. These results show the extent of impact of the usage scenarios on the verification accuracy, especially the scenarios involving body movement such as walking and travelling on a bus.

The results raise questions about the stability of swipe gesture verification when faced with multiple usage scenarios encountered on a mobile device. Further, the inter-session results using the conventional classifiers and DNN technique show that the swipes performed on the same day yielded better EER compared to the swipes acquired on different days. This aspect puts into question the time persistence of swipe gestures produced, particularly, considering the fact that a typical user could use a mobile device over several years.

Chapter 5. Signature Verification

5.1 Introduction

Signatures are a widely accepted trait for verification. Its extensive use over centuries in legal documents such as contracts, makes it a popular personal verification attribute. Signature verification methods can be broadly classified into two categories – offline and online. In an offline verification, a signature is produced on paper and scanned as an image, whilst for an online signature the signature is captured and processed on a device or digitising tablet enabled with pressure reporting features. Currently, in the domain of signatures, e-signatures have substantially overtaken the physical static signature productions in various application areas. These e-signatures are performed on mobile devices such as a tablet or a smartphone. From a hardware perspective, most of the mobile devices can capture fingertip movement on the touchscreen and therefore finger-based signatures are a feasible option for mobile device-based signatures. However, compared to a signature produced on a standard signature-capturing device such as those manufactured by Wacom, signatures captured using mobile devices can show different verification performance due to varying sampling rates and device properties.

One of the key challenges of signature verification in mobile biometrics is the intra-personal variability. No two signatures produced by a person can be exactly the same and this challenge of intra-personal variability can become even more apparent when using signature verification on a mobile device. Another factor that adds complexity is that there is no defined method for signature donation on a mobile device such as position of the device or input tool (finger/stylus) to be used.

Current work on signature-based verification largely utilises publicly available datasets [115], [118] and [125] that have acquired signatures from users in controlled/stable conditions such as an indoor capture environment and have human supervision during signature donation. However, in order to capture the real-life signature production, it is necessary to acquire signatures in an unconstrained environment. The novelty of this chapter is that it presents the performance evaluation of the signatures captured under various usage scenarios of a mobile device in an uncontrolled environment. The main contributions of this chapter are fourfold: intra-session analysis, input tool analysis (finger and stylus-based signature), inter-session analysis (Session 1 & Session 2 separated by a week) and intra-algorithm analysis using a function-based, a feature-based signature verification algorithms and a commercially available signature verification solution.

Based on the above criteria, we assess a number of research questions that are aligned to the overall research objectives provided in Chapter 1. They are:

- Evaluate the influence of input-type on intra-person variability
- Evaluate the influence of usage scenarios of a mobile device on verification performance
- Evaluate the influence of time-separated signatures on verification performance

5.2 Related Work

An in-depth review of existing studies on signature verification performed using mobile devices has been presented in Chapter 2. A list of publicly available signature databases containing signatures acquired on mobile devices are presented in Table 3.1. In this chapter, studies highlighting mobile device related factors that impact the overall verification performance are described.

There are a number of studies that present work on dynamic signature verification using mobile devices [70], [67], [71], [66] and [64]. The study conducted by Galbally et al. [66] compares the verification performance of mobile devices (a PDA and Samsung Galaxy Note) and pen tablets. Their results show a decrease in discriminative power and higher biometric error rates on hand-held devices. They conclude, "*it has been observed that mobile conditions negatively affect feature discriminative power, especially when local features are considered*". Although their results indicate variation in performance of hand-held devices and PDAs, there is clearly a need to make a deeper analysis of individual factors causing the impact that resulted in poor performance for mobile devices.

With respect to user interaction, the impact of factors such as user posture during the signature donation (for example device placement) and input-tool on the verification performance has been studied. A study conducted by Blanco-Gonzalo et al. [82] analysed the impact of user posture on signature verification. They performed a series of usability evaluations on dynamic signatures using an iPad. They assessed the impact of using - a) different styluses; b) varied postures of users such as sitting, standing etc; and c) having a device placed on a table and held in the hand. They evaluated usability based on three metrics – effectiveness, efficiency and satisfaction. Based on user feedback received after completion of the experiment, the authors claim that *"the users consider staying seated and having the paper/device over a table as the most common and comfortable way to sign"*. Their experimental results also ascertain that having the device placed on a surface leads to best performance as it avoids negative effects of pressure changes. Although their study indicates this preference towards a comfortable setting, in practice, when signing using a mobile device, the user holds the device in one hand and grasps the stylus pen in another hand and there is no dominant posture for attaining the most stable signature.

Sanchez-Reillo et al. [64] conducted a performance evaluation of handwritten signature verification in mobile environments using stylus and finger signatures for 43 subjects. They report, based on user experience feedback collected at the end of the study, that *"fingertip-based devices are the less preferred by users because of the lack of habituation to make the signature with the fingertip"* [64]. Their paper also raised a relevant question of considering signature as two different modalities - stylus-based and finger-based. Although their study reported better user experience with stylus, it must be noted that a significant majority of new mobile phone models come without a stylus. Hence, future research should focus more on methods to improve the user experience and performance of fingertip-based signatures. In the same experiment, they analysed signatures captured across multiple mobile devices – a Blackberry playbook, a Wacom STU signature pad, a Wacom Intuos tablet, an Apple iPad2, an Asus Eee PC touch, a Samsung Galaxy Tab and Samsung Note. The results were analysed based on interoperability, modality tests and visual feedback. Their experiment for intra-device and inter-modality evaluations using DTW algorithm showed that compared to other devices, iPad yielded the best result of 0.19% EER. They reported that receiving visual feedback was the most important factor for the users as they felt less comfortable with the absence of visual feedback from the device. This hypothesis was based on poor error rates for the Intuos device that provided no visual feedback.

As concluded by these studies, performance variation in signature verification is caused by multiple factors such as user posture and input tool. Conducting an in depth analysis using multiple signature verification algorithms and understanding the impact of - a) input-tool (stylus and finger-based signatures) from same device, b) user's body movement and a range of unconstrained environment on the verification performance, and c) time-separated signatures, can help in improving the verification efficiency. This chapter presents this analysis in detail.

5.3 Experimental Framework

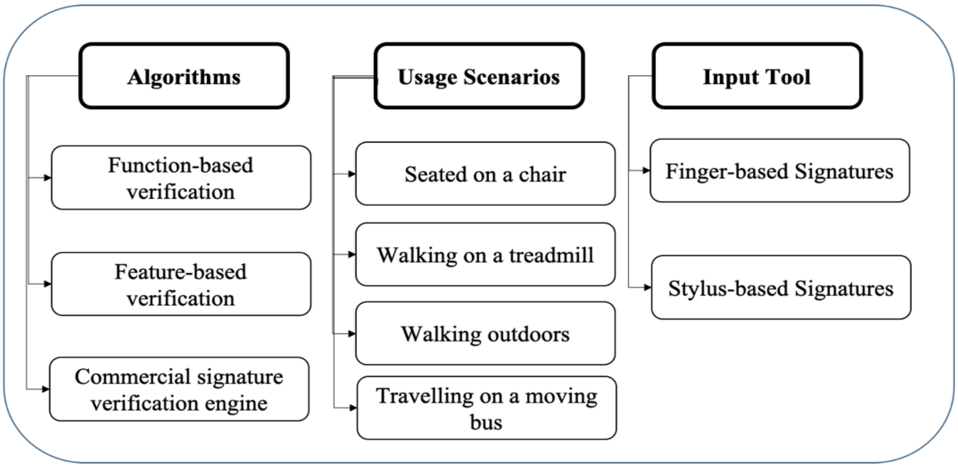


Figure 5.1. Experimental framework parameters

In order to conduct this analysis, a signature dataset with multiple usage scenarios, input-tools and time-separated sessions has been utilised. The holistic view of the experimental framework parameters considered for this analysis is shown in Figure 5.1. Dynamic signature verification systems are classified into two categories - feature-based and function-based. The feature-based methods make use of a global multi-dimensional feature vector that takes into account the entire signature, for example the total time of a signature and average pressure of a signature. A function-based system makes use of the discrete time functions, otherwise known as local features, such as velocity between two consecutive sample points in a signature. The analysis has been conducted with three different signature verification set-ups/algorithms. The first framework utilises the raw signature data - DTW algorithm, the second is a feature-based algorithm – SVM and the third framework is a commercial signature verification framework that uses Levenshtein distance algorithm. The reason for using three different frameworks was to validate the results using a black-box commercial and open-access signature verification algorithm. Although the verification frameworks are different, the same input data were used for all three analyses. The input tools used for signature production were finger and stylus. Signatures were captured in four different usage scenarios of a smartphone – whilst seated on a chair indoors, whilst walking on a treadmill, whilst walking outdoors and whilst seated on a moving bus. The signature processing techniques and verification criteria varied in all the three cases. These are explained in detail in the following sections.

5.3.1 Dataset

The signature dataset used for this experiment has been described extensively in Chapter 3. Both finger-based and stylus-based signatures were captured for 50 participants in different usage scenarios. The dataset consisted of left-handed (6), right-handed (43), and ambidextrous (1) users. The signature data was captured using a smartphone – a Galaxy Note 5. The sampling rate of the touchscreen was 240 Hz and the number of pen pressure levels reported by the screen were 2048. Table 5.1 shows the number of signatures captured per scenario in Session 1 and Session 2.

	Scenario 1	**Scenario 2**	**Scenario 3**	**Scenario 4**
Session 1				
Finger-based	10	3	5	-
Stylus-based	5	3	5	-
Session 2				
Finger-based	10	-	5	3
Stylus-based	5	-	5	3

Table 5.1. Number of finger and stylus-based signatures captured per scenario in Session 1 and Session 2

Two participants indicated that they had previously owned a smartphone with a stylus. 96% of the participants revealed that they have used a stylus on Wacom devices whilst signing for a home delivery. 30% of the participants revealed that they have used stylus devices such as tablets for taking notes. All the signatures were captured in portrait mode of the smartphone to maintain consistency.

The performance analysis was conducted using three different algorithms – DTW, SVM and a commercial system based on Levenshtein distance. All frameworks are described in the following sections.

5.3.2 DTW (Function-based Signature Verification)

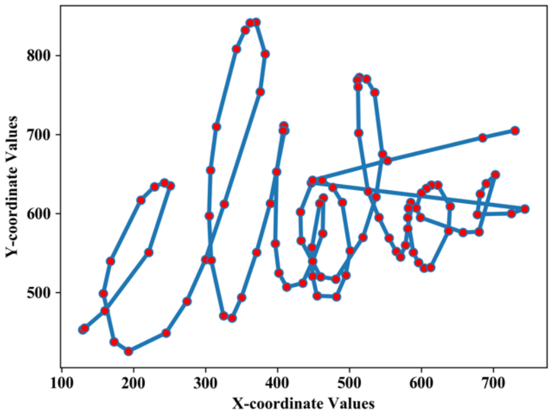


Figure 5.2. Signature sample 1 of User ID 20

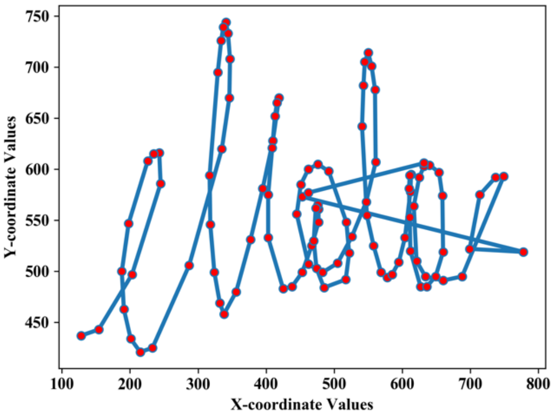


Figure 5.3. Signature sample 2 of User ID 20

In this analysis, a function-based signature verification using DTW algorithm has been applied to find the similarity between signatures. DTW is a widely used algorithm to compute distances from time-based functions [66], [70], [126]. It is primarily used to calculate distances between two signatures of varying sample lengths. A signature consists of a number of sample points of X and Y-coordinate pairs captured from the pen/finger down to pen/finger up action on the touchscreen. No two signatures from a given user are the same. They can vary in terms of number of sample points (represented as black dots) as shown in Figure 5.2 and Figure 5.3.

5.3.2.1 Pre-processing

Signature pre-processing is performed to reduce signal noise and to identify and remove outliers of data points from the signature data. The pre-processing steps undertaken are as follows:

- Equi-spacing – In order to eliminate inconsistency due to sampling rate variation, linear interpolation was performed. This generated samples at an equi-spaced sequence of time. All signature input sequences were interpolated to a length of 256.

- Location normalisation – In order to avoid the variation due to the location on the touchscreen, the X and Y axis coordinates of the sample points were normalised based on their mean value:

    - $X' = X(t) - \bar{X_t}$ (5.1)
    - $Y' = Y(t) - \bar{Y_t}$ (5.2)

Here X' denotes the normalised X-coordinate, $X(t)$ is the X-axis value at a given time *'t'* and $\bar{X_t}$ is the mean of the X-axis coordinates for the entire signature sample points. Similarly, Y' is the normalised Y-coordinate, $Y(t)$ is the Y-axis value at a given time *'t'* and $\bar{Y_t}$ is the mean of the Y-axis coordinates for the entire signature sample points.

5.3.2.2 Temporal Sequences for DTW Comparison

After applying the pre-processing step to the raw signature data, seven types of temporal sequences were derived. Table 5.2 provides a description of individual sequences along with their unit of measurement and the formula used to calculate it. These discrete-time sequences were categorised as raw (X-coordinate, Y-coordinate, pressure) and derived sequences (velocity, acceleration, sine, cosine and angle between consecutive sample points). The derived sequences were calculated for every sample point of the signature, from one sample point to the next sample point.

Sequence Number	Feature	Description	Metric
S1	X Coordinate	X coordinate value	Pixels
S2	Y Coordinate	Y coordinate value	Pixels
S3	Pressure	Finger / Stylus Pressure (variable only valid for stylus-based signatures)	Discrete pressure value ranging from 0 (no pressure at all) to 1 (normal pressure)
S4	Velocity	Velocity from one sample point to the next sample point. $V_n = \frac{\sqrt{Y_n^2 - X_n^2}}{t_{(n-1)} - t}$	Pixels/Milliseconds
S5	Acceleration	$A_n = \frac{V_{n-1} - V_n}{t_{(n-1)} - t_n}$	Pixels/Milliseconds
S6	Sine	$Sin(n) = Sin\ (\propto_n)$	-
S7	Cosine	$Cos(n) = Cos\ (\propto_n)$	-
S8	Angle of consecutive samples	$\propto_n = \arctan(\frac{Y_n - Y_{(n-1)}}{X_n - X_{n-1}})$	radian

Table 5.2. Temporal sequences for DTW

5.3.2.3 DTW Algorithm

The dynamic time warping algorithm has been chosen for this analysis as it has been successfully used in online signature verification in multiple studies [66], [70], [126] and it efficiently computes the distance between temporal signals. In our study, the discrete time signals for which DTW distance has been calculated are for raw signals - X and Y coordinates, pressure and derived sequences – velocity, acceleration, sine, cosine and angle between the consecutive sample points in a signature.

With respect to the algorithm, DTW takes two independent time signals into consideration – a sample 1 (S1) of length 'n' and a sample 2 (S2) of length 'm',

$$S1(x) = x_1, x_2, x_3, \dots \dots .. x_n \tag{5.3}$$

$$S2(x') = \ x'_1, x'_2, x'_3, \dots \dots .. x'_m \tag{5.4}$$

The DTW algorithm calculated the warping path '*W*', which indicates the similarity measure between two signals S1 and S2.

$$W = w_1, w_2, \dots .. w_k \qquad \max(|n|, |m|) <= k < |n| + |m| \tag{5.5}$$

The variable 'k' indicates the warping path length. Given *'i'* as the index of the sample points of S1 and j as the index of the sample points of S2, 'w' is calculated for each element in S1 and S2 as shown in equation 5.6. A two-dimensional cost matrix is constructed with dimensions of $|\mathrm{n}| * |\mathrm{m}|$, containing the distance value in each cell for the corresponding elements in S1 and S2. The warping path is calculated from the starting point $W_1 = (1,1)$.

$$W_k = (i, j) \tag{5.6}$$

Every cell of the matrix is filled with value of $D(i, j)$. A warping path is found from D (1,1) to D (|n|, |m|) by applying a greedy search algorithm to obtain the lowest distance value from the adjacent cells as shown in Figure 5.4. This is calculated as follows:

$$\begin{aligned} D(i, j) = Dist(i, j) \\ + \min\big(D(i-1, j), D(i, j \\ - 1), D(i-1, j-1)\big) \end{aligned} \tag{5.7}$$

For example, in Figure 5.4, D (6,4) = Dist (6,4) + min (D(5,4),D(6,3), D(5,3). The value *'D (|n|, |m|)'* will contain the minimum warping path. The minimum distance warp path is calculated as:

$$Dist(W) = \sum_{k=1}^{k=K} Dist(w_{ki}, w_{kj}) \tag{5.8}$$

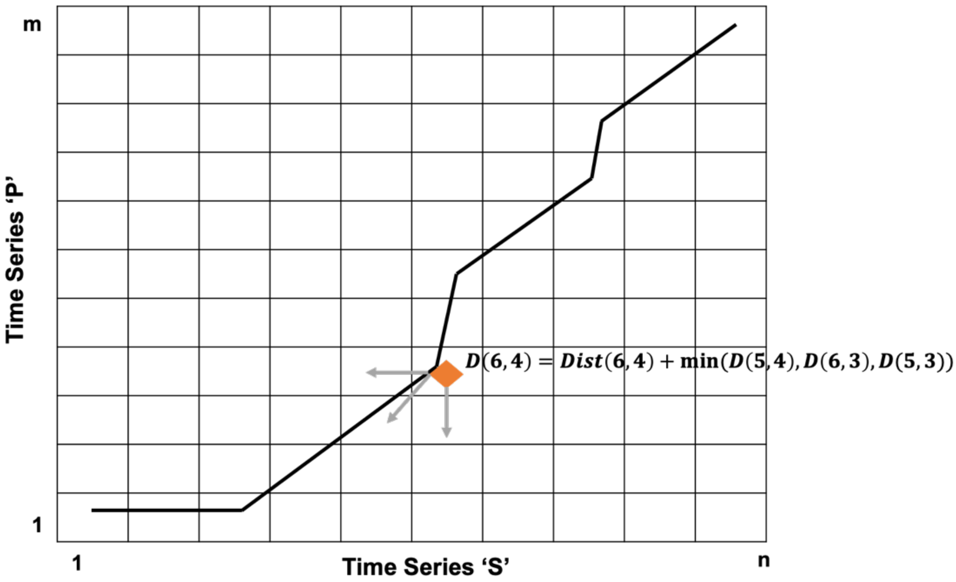


Figure 5.4. Cost matrix with warping path

5.3.2.4 Enrolment and Verification

The verification configuration is based on the concept of verification. For each user in the dataset, a verification configuration was built. Each model had one genuine and one imposter user. The owner of the mobile device was assumed as the genuine user and the imposter user was chosen based on a random forgery method. A skilled forgery method was also considered for choosing the imposter user, but not used due to a lack of expertise in signature mimicking skills. Equal number of samples were used for the genuine and imposter comparisons.

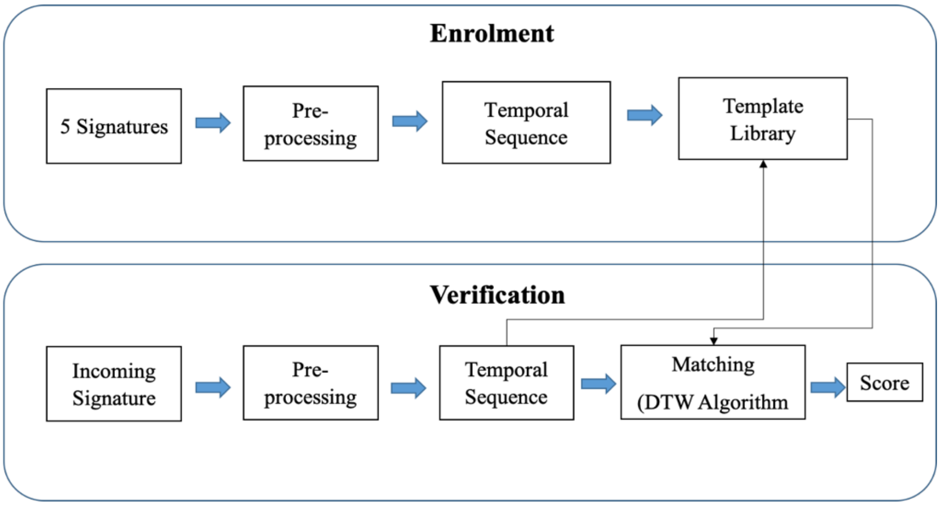


Figure 5.5. Enrolment and verification process overview

The enrolment and verification processes are depicted in Figure 5.5. The enrolment process adopted is as follows:

- *Step 1: Choosing enrolment samples* – The enrolment process uses the first five signatures acquired from the genuine user during data collection. Choosing the first five signatures as enrolment samples allowed us to account for a scenario when the user obtains a new mobile device and has to provide their signature samples.

- *Step 2: Temporal Sequence* – After setting apart the first five signatures as an ‘enrolment-set’, raw and derived sequences were extracted from each of the enrolment signatures. An example signature sequence for individual signature is presented in equation 5.9 These sequences were stored in a template library, to be used during the verification phase.

$$Sig_1 = [\, S1^1_x, S2^1_y, S3^1_{pressure}, S4^1_{velocity}, S5^1_{acceleration}, S6^1_{sin}, S7^1_{cos}, S8^1_{angle}\,] \qquad (5.9)$$

- *Step 3: Cost Matrix calculation* – Along with the extracted feature set, the cost matrices obtained from the individual features from the five enrolment signatures and their associated variables ($avgNearest_f$ and $avgFarthest_j$) are stored in the template library. In order to calculate the cost matrix for each feature, DTW distance was calculated by comparing each signature with all the signatures present in the enrolment set. This step resulted in generating a 5 x 5 DTW distance matrix for each feature (as shown in Figure 5.6). A total of 8 DTW distance matrices were generated. As an example, Figure 5.6 represents the DTW distance matrix calculated for a single user for the velocity feature.

	Sig 1	**Sig 2**	**Sig 3**	**Sig 4**	**Sig 5**
Sig 1	0	36	56	114	69
Sig 2	36	0	89	158	167
Sig 3	56	89	0	93	88
Sig 4	114	158	93	0	187
Sig 5	69	167	88	187	0

Figure 5.6. DTW distance matrix for velocity

- *Step 4: Calculation of Average Nearest and Average Farthest variables* – Once the DTW distance matrix was calculated, the nearest and farthest signature pair along with their average distance value were determined for each sequence. From each DTW distance matrix, the 'Average distance of nearest pair' ($avgNearest_f$) and 'Average distance of farthest pair' ($avgFarthest_j$) were calculated by finding an average of the minimum and the maximum distances from the matrix. These values were stored in the template library to be used during the verification process.

The verification process adopted is as follows:

- *Step 1: Pre-processing and Feature Extraction* - During the verification process, the incoming signature was pre-processed, followed by the temporal sequence extraction phase to generate a feature vector (x-coordinate, y-coordinate, velocity, acceleration, sine, cosine, angle of consecutive samples).

- *Step 2: Comparison with enrolment set* - Each individual sequence (f) from the signature temporal sequence vector of the incoming signature was compared against the corresponding feature of the signatures from the enrolment set. This generated a matrix of distances as shown in .

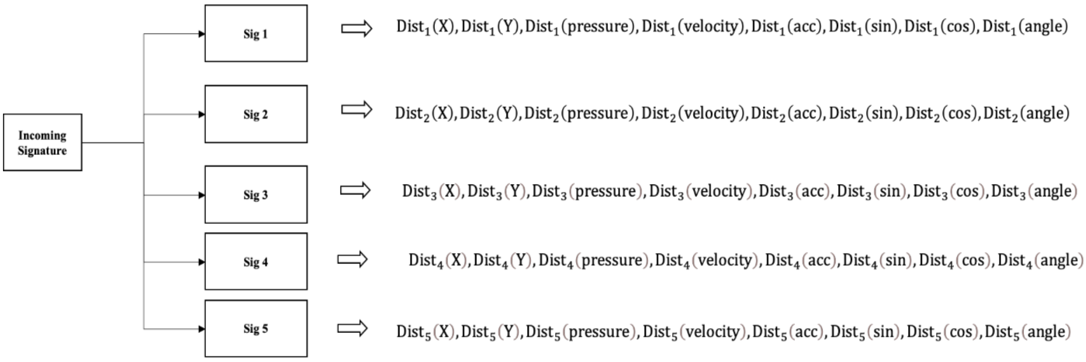

Incoming Signature
Sig 1
Sig 2
Sig 3
Sig 4
Sig 5
$Dist_1(X), Dist_1(Y), Dist_1(pressure), Dist_1(velocity), Dist_1(acc), Dist_1(sin), Dist_1(cos), Dist_1(angle)$
$Dist_2(X), Dist_2(Y), Dist_2(pressure), Dist_2(velocity), Dist_2(acc), Dist_2(sin), Dist_2(cos), Dist_2(angle)$
$Dist_3(X), Dist_3(Y), Dist_3(pressure), Dist_3(velocity), Dist_3(acc), Dist_3(sin), Dist_3(cos), Dist_3(angle)$
$Dist_4(X), Dist_4(Y), Dist_4(pressure), Dist_4(velocity), Dist_4(acc), Dist_4(sin), Dist_4(cos), Dist_4(angle)$
$Dist_5(X), Dist_5(Y), Dist_5(pressure), Dist_5(velocity), Dist_5(acc), Dist_5(sin), Dist_5(cos), Dist_5(angle)$

Figure 5.7. Processing of incoming signatures with the enrolled signatures

- *Step 3: Calculation of minimum and maximum distances* – From the distance matrix generated from step 2, the minimum distance and maximum distance were calculated for each feature from all the 5 distances. After acquiring the minimum and maximum distances per feature, 'Diff (Minf)' and 'Diff (Maxf)' was calculated. These variables were calculated to maintain stability of the enrolment samples and were calculated based on the following formulas:

$$Diff(Min_f) = \frac{minimum\ distance_f - avgNearest_f}{avgNearest_f} \quad (5.10)$$

$$Diff(Max_f) = \frac{maximum\ distance_f - avgFarthest_f}{avgFarthest_f} \quad (5.11)$$

$$Diff = \sum Diff(Min_f) - Diff(Max_f) \quad (5.12)$$

If Diff <= Threshold, then Accept as genuine signature

Else Reject as imposter signature

Finally, the '*Diff'* value is compared with the threshold set for the user verification configuration. The threshold selected for every user model in the dataset was different, they were generated based on the scores generated. The performance of the DTW model has been presented in Section 5.4.

5.3.3 Support Vector Machine (Feature-based Signature Verification)

Identifier	Feature	Description
1	Mean Velocity	Average pen velocity maintained in the signature
2	Signature Height	Height of the signature
3	Signature Width	Width of the signature
4	Signature Area	Area of the signature
5	Mean Acceleration	Mean acceleration
6	Mean Pressure	Mean pressure
7	Mean X Acceleration	Mean pen x-axis acceleration
8	Mean Y Acceleration	Mean pen y-axis acceleration
9	Mean X Velocity	Mean pen x-axis velocity
10	Mean Y Velocity	Mean pen y-axis velocity
11	Peak Acceleration	Peak/maximum acceleration
12	Peak Velocity	Peak/maximum velocity
13	Peak X Acceleration	Peak/maximum X-axis acceleration
14	Peak Y Acceleration	Peak/maximum Y-axis acceleration
15	Peak X Velocity	Peak/maximum X-axis Velocity
16	Peak Y Velocity	Peak/maximum Y-axis Velocity
17	Total Duration	Total time of the entire signature
18	Number of Pen-ups	Total number of pen-ups of the signature
19	Average Jerk	Average jerk maintained in the signature
20	Standard Deviation X-axis velocity	Standard deviation of X-axis velocity
21	Standard Deviation Y-axis velocity	Standard deviation of Y-axis velocity
22	Standard Deviation Acceleration	Standard deviation of Acceleration

Table 5.3. Global Features extracted from the PenTools application

To perform a feature-based signature verification, an SVM algorithm has been used. From the raw time-based signature data, a number of global features were derived such as average velocity of the signature and total time of the signature. These features take the entire signature data into account instead of the local sample points that form the signature. The PenTools application [127] has been utilised in order to generate these features. The features used for this analysis are presented in Table 5.3.

These features are divided as:

- *Time-based Features* – these features relate to the duration of the signature and total number of pen-ups in the signature.
- *Position-based Features* – these features relate to the characteristics derived from the position change of the signature, such as first and second order derivative of X and Y-axis positions, mean velocity, peak velocity.
- *Pressure-based Features* – pressure associated feature – mean pressure of the signature.
- *Geometry-based Features* – Height, width and area of the signature.

Separately, an SVM with a linear and RBF kernels was utilised to analyse the data. The RBF kernel utilises the C and gamma parameters. The C parameter trades off correct classification of training samples against maximization of the decision function's margin. In order to attain the best cross-validation score, the parameter optimisation method Grid Search [128] has been applied, to estimate the suitable hyper-parameters. The final C value was chosen as 1 and the gamma value was set as 0.01.

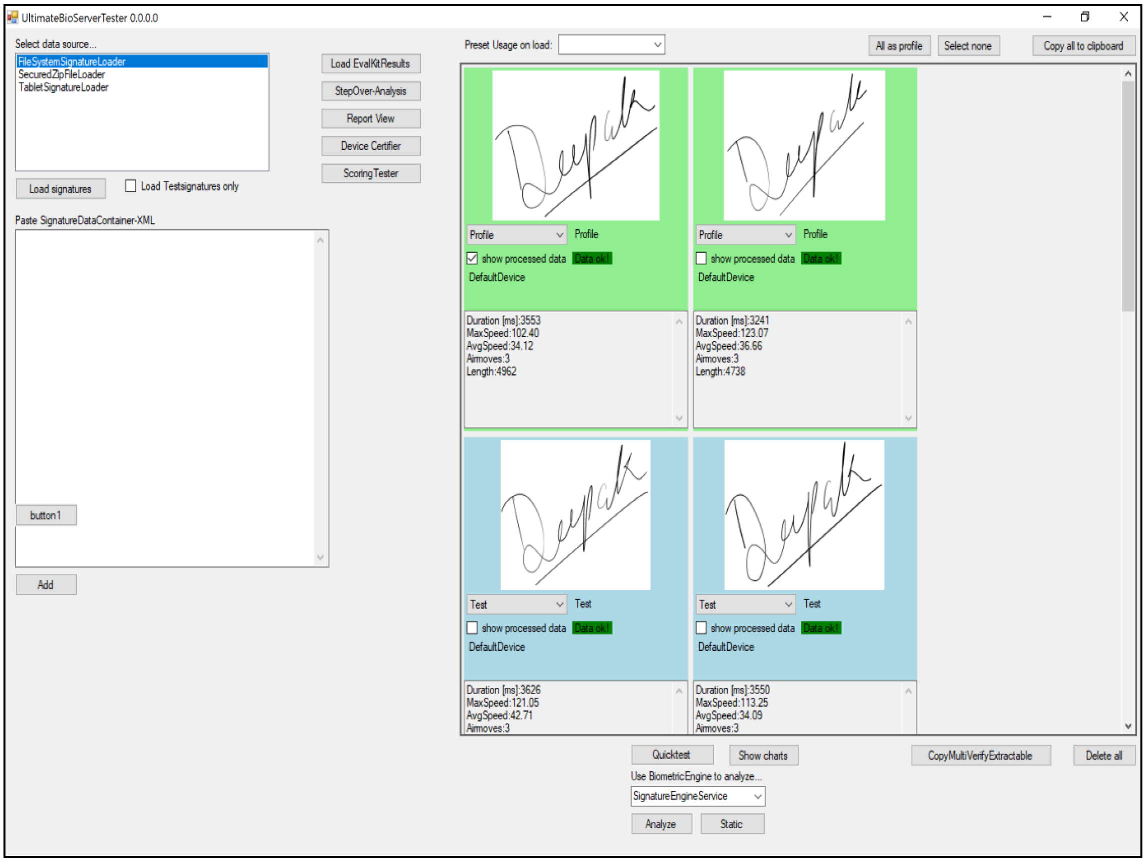
5.3.4 Commercial Signature Verification Engine

A commercial signature verification engine has also been utilised to conduct the biometric performance analysis of the mobile-based signatures. This system is used to process legally compliant electronic signatures, manage and track the flow of documents, conduct secure transactions and ensures secure storage of data. The platform supports all types of e-signatures capture mechanisms such as Click-to-sign, Draw-to-sign and Type-to-sign mechanisms. Its biometric verification platform provides real time verification by comparing the signature against a pre-enrolled signature present in the profile database. An overview of the biometric verification engine architecture, adopted signature processing techniques such as pre-processing, feature extraction and classification methods has been presented in this section.

5.3.4.1 Architecture

- *Signature Data Container* - The raw signature data consists of multiple sample points with parameters such as timestamp, X-coordinate, Y-coordinate and pressure. In order to use our data from the data collection, the data needed to be converted into a 'Signature Data Container', which was in XML format. In order to make this conversion, a converter program was developed in C# using the Visual Studio platform. The signature data container comprised a number of parameters regarding the signature and the capturing device. The capturing device information such as device name, version number, pressure information - minimum and maximum pressure, whether air moves are supported and sampling rate points per second / fixed sampling rate were stored in this file. The resulting XML file generated from the program had sample points with parameters - X, Y, pressure, timestamp in milliseconds, pen-up/pen-down flag along with the device information.

- *Signature Verification Engine* - For each signature present in the dataset, individual signature data container files were generated. These XML signature files were fed into the '*Signature Verification Engine*'. This engine used *'UltimateBioServer Test'* platform (non-real-time signature verification) to perform the signature verification. The first step of processing was the creation of a user and their associated profiles in a database. In order to conduct the experiment, 50 individual users were selected, with each having two different profiles in the database – one with a finger-based signature and another with a stylus-based signature. A screenshot of the UI of UltimateBioServer is provided in Figure 5.8.



Profile — ☑ show processed data — DefaultDevice
Duration [ms]:3553
MaxSpeed:102.40
AvgSpeed:34.12
Airmoves:3
Length:4962

Profile — □ show processed data — DefaultDevice
Duration [ms]:3241
MaxSpeed:123.07
AvgSpeed:36.66
Airmoves:3
Length:4738

Test — □ show processed data — DefaultDevice
Duration [ms]:3626
MaxSpeed:121.05
AvgSpeed:42.71
Airmoves:3

Test — □ show processed data — DefaultDevice
Duration [ms]:3550
MaxSpeed:113.25
AvgSpeed:34.09
Airmoves:3

Figure 5.8 User interface of UltimateBioServer Test

The platform consists of an in-built signature chart visualiser that shows the charts of the discrete time signature signal data such as X-coordinates, Y-coordinates and pressure (as shown in Figure 5.9) and the derived features (Figure 5.10). Each of the signatures of individual users were uploaded into the platform using the '*File System Signature Loader'* option (as shown in Figure 5.8). These signatures had the option of being selected as a 'Profile' – enrolment, 'Test' – verification, 'Fake' – forgery, ‘Random Fake’ – random forgery signature type. The profile signatures are shown in green colour and test signatures in blue (as depicted in Figure 5.8).

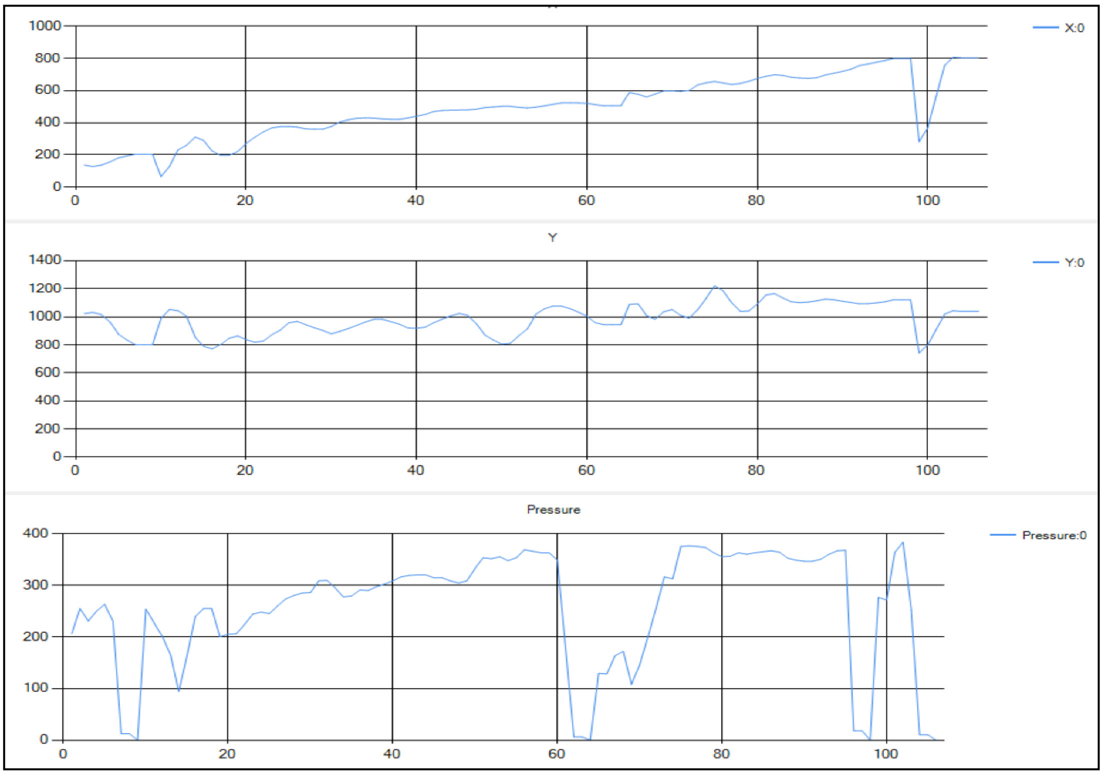
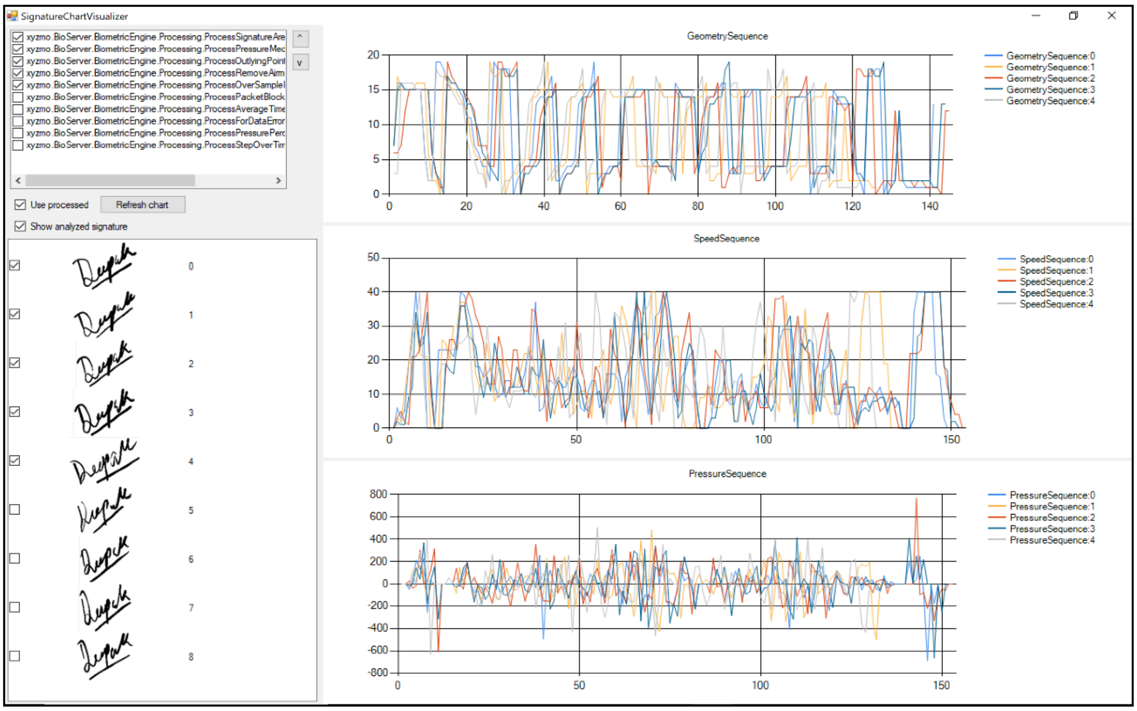


Figure 5.9. Visualiser showing discrete time signature data of X-coordinates, Y-coordinates and pressure


SignatureChartVisualizer

- ☑ xyzmo.BioServer.BiometricEngine.Processing.ProcessSignatureAre
- ☑ xyzmo.BioServer.BiometricEngine.Processing.ProcessPressureMec
- ☑ xyzmo.BioServer.BiometricEngine.Processing.ProcessOutlyingPoint
- ☑ xyzmo.BioServer.BiometricEngine.Processing.ProcessRemoveAim
- ☑ xyzmo.BioServer.BiometricEngine.Processing.ProcessOverSample
- ☐ xyzmo.BioServer.BiometricEngine.Processing.ProcessPacketBlock
- ☐ xyzmo.BioServer.BiometricEngine.Processing.ProcessAverageTime
- ☐ xyzmo.BioServer.BiometricEngine.Processing.ProcessForDataError
- ☐ xyzmo.BioServer.BiometricEngine.Processing.ProcessPressurePerc
- ☐ xyzmo.BioServer.BiometricEngine.Processing.ProcessStepOverTim

☑ Use processed

☑ Show analyzed signature

- ☑ 0
- ☑ 1
- ☑ 2
- ☑ 3
- ☑ 4
- ☐ 5
- ☐ 6
- ☐ 7
- ☐ 8

Figure 5.10. Derived features in the visualiser

For this analysis, a user's finger-based signatures were selected as profiles, while the test signatures were considered from both stylus-based and finger-based samples obtained in Scenario 2, 3 and 4 from Session 1 and Session 2 of the data collection. Once the verification mode was active, the profile signatures were compared against the test signatures and similarity scores were generated. The signature processing phases in the biometric engine are described in the next section.

5.3.4.2 Signature Verification Engine

The pre-processing of the signature involved linear interpolation of the raw signature data. Using this method, all the raw signatures are converted into fixed length signature data and normalisation of time and size were performed as well.

Although we have treated the algorithm as a black box, the underlying algorithm used on the commercial signature verification was based on Levenshtein distance. Levenshtein distance [129] is a string-based algorithm used for calculating difference between two string sequences. This textual pattern recognition method is based on *'edit-distance'*, which represents the least number of edit operations required to modify one string to obtain another. This distance is calculated by transforming one string into another by performing a number of insertions, deletion or substitutions on individual characters. The minimum number of edit operations required to convert one string as another is known as *edit-distance*. The edit-distance can be calculated using a matrix method. For example, for two strings String 1 (FLOMAX) and String 2 (VOLMAX), $n=6$ and $m=6$ are their respective lengths, a |n| x |m| matrix is created. Each element in the matrix is calculated as equation 5.13 - 5.16 and w_i, w_d and w_r are calculated as the scores for edit operations insert, delete and replace respectively.

$$D(i,j) = \min \begin{matrix} [\,D(i-1,j) + w_d, \\ D(i,j-1) + w_i, \\ D(i-1,j-1) + w_r] \end{matrix} \tag{5.13}$$

$$D(i,0) = D(i-1,0) + w_d \tag{5.14}$$

$$D(0,j) = D(0,j-1) + w_i \tag{5.15}$$

$$D(0,0) = 0 \tag{5.16}$$

In order to adapt this string-based algorithm to signature verification, extracted data were encoded. During the verification phase, the incoming signature feature set was also encoded, and the edit-distance was calculated. A final similarity score based on the calculated edit-distance was generated. Lower edit-distance values represented similar signatures.

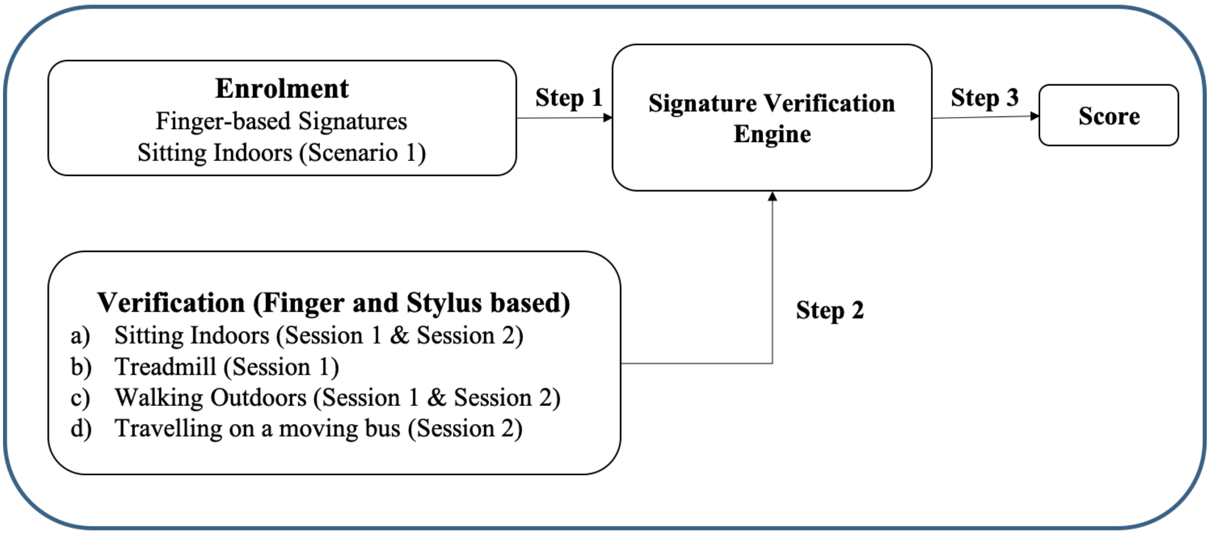

Enrolment
Finger-based Signatures
Sitting Indoors (Scenario 1)
Step 1
Signature Verification Engine
Step 3
Score
Step 2
Verification (Finger and Stylus based)
a) Sitting Indoors (Session 1 & Session 2)
b) Treadmill (Session 1)
c) Walking Outdoors (Session 1 & Session 2)
d) Travelling on a moving bus (Session 2)

Figure 5.11. Verification process using the commercial system

Figure 5.11 shows the overview of the signature verification process used in the commercial system. The verification process was split into enrolment and verification phases. The enrolment was performed for each user present in the dataset individually, after which genuine and imposter comparisons were made with the enrolled samples.

The enrolment phase involved creating a profile for the genuine user. In order to create a profile for a user, a minimum of 5 signatures were required by the signature verification engine. For this, finger-based signatures belonging to the baseline scenario – Sitting Indoors (Scenario 1) were used because they are devoid of any impact due to movement. From the first ten signatures acquired from the user in Scenario 1, five signatures that had minimum distance between them were chosen as the enrolment set by the system. The reason to select the most similar signatures was to maintain internal stability of the enrolled signatures. The dissimilar signatures were discarded to avoid scope for forgery.

The verification signatures belonged to a combination of finger and stylus-based signatures belonging to Sitting Indoors, Treadmill, Walking Outdoors and Travelling on a moving bus of Session 1 and Session 2. The cut-off threshold was set by the system as 80%. The scores attained below 80% were termed as *'no match'* and 80% and above as *'match'*. The results obtained using these evaluation tests are provided in Section 5.4.

5.4 Results

In this section, the analysis results of individual research questions (provided in Section 5.1) have been presented. The evaluation results utilise all three signature verification frameworks – commercial signature verification engine, DTW-based and SVM-based system. Firstly, Section 5.4.1 details the overall performance evaluations for the data obtained from Session 1 and Session 2. Secondly, Section

5.4.2 presents the evaluation results of influence of input-type on the verification performance. Following this, Section 5.4.3 presents a comparison of verification performances under different usage scenarios of a mobile device. Finally, Section 5.4.4 presents the influence of time-separated signature verification capture.

5.4.1 Performance Evaluation

For conducting this evaluation, the enrolment was performed using the finger-signatures belong to the baseline scenario (Sitting Indoors – Scenario 1) and Session 1 and Session 2 respectively. The verification signatures belonged to all the scenarios and input types (finger-based and the stylus-based) present in the dataset. The reason for conducting such an analysis was to understand the overall performance of the system irrespective of the input tool type and scenario variation.

5.4.1.1 Commercial Signature Verification Engine

In order to conduct the performance evaluation using the commercial signature verification engine, genuine and imposter comparisons of the signatures were carried out for each individual user (50 users) present in the dataset. The genuine signature comparison was carried out by comparing the signatures of the genuine user with their own signatures. An imposter user was randomly chosen from the dataset. The imposter signature comparison was carried out by comparing a genuine user's signature with an imposter user's signature.

For performing each user verification, each session contained 21 genuine comparisons with 8 finger-based signatures and 13 stylus-based signatures and 21 imposter comparisons (8 finger-based and 13 stylus-based). For 50 users, a total of 1050 genuine comparisons were made for each session. It was noted that all the imposter comparisons (from different scenarios and sessions) resulted in a '*no match*' status, giving a 100% True Rejection Rate (TRR). A possible reason for this could be the rejection threshold value set by the system. On the contrary, genuine comparisons showed varying performance of acceptance. Therefore, in this section, only the results obtained from the genuine comparisons are presented. Figure 5.12 presents the true acceptance and false rejection rates for the genuine comparisons of Session 1 and Session 2 respectively.

The results reveal that the mean True Acceptance Rate % of Session 1 is lower than Session 2 and the mean False Rejection Rate % attained for Session 1 is higher compared to Session 2. The reduction in percentage of false rejects in Session 2 can be due to user's familiarity with the UI of the app, leading to the production of stable/similar signatures.

On further analysis, it was noted that most of the falsely rejected signatures in both the sessions belonged to signatures acquired from different usage scenarios and different input-tool type (stylus-based signatures). The amount of failures caused for finger-based and stylus-based signatures and the different usage scenarios are detailed in Section 5.4.2 and Section 5.4.3.

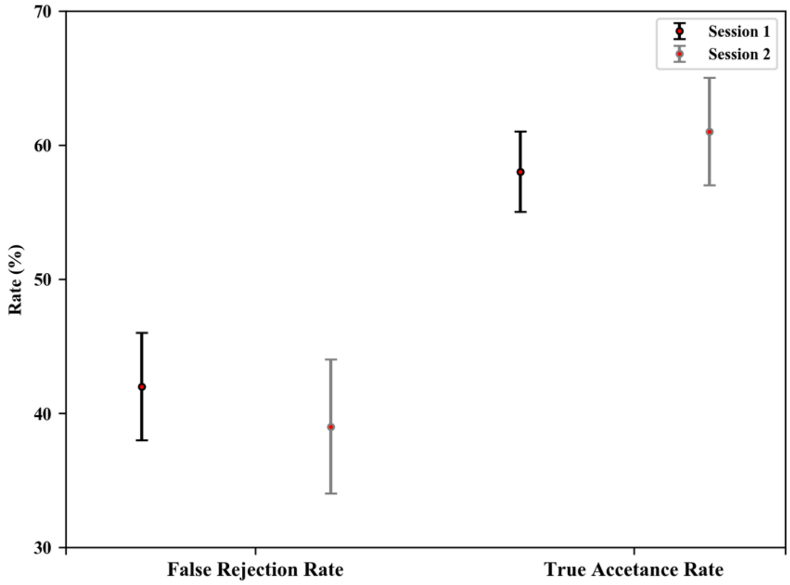


Figure 5.12. Genuine signature comparisons for Session 1 and Session 2 using the commercial signature verification system

5.4.1.2 DTW

As the DTW algorithm was applied for multiple time-based sequences of the signature (namely X-coordinate, y-coordinate, velocity, acceleration, sine, cos, angle of consecutive points), DTW-based performance evaluation based on individual time-sequences have been presented in this section. Figure 5.13 shows the mean EER % obtained for the finger-based signature comparisons conducted for Session 1 and Session 2. Equal number of genuine and imposter comparisons were conducted for this analysis to avoid class bias. It can be seen that sequences such as y-coordinate and angle of the consecutive sample points acquired the highest mean EER rate, while velocity, acceleration and x-axis yielded lowest mean EERs. The cumulative score sequence was calculated by finding an average of the accumulated scores generated by all the sequences. The cumulative score sequence's mean EER % obtained from both Session 1 and Session 2 was the lowest, compared to other sequences.

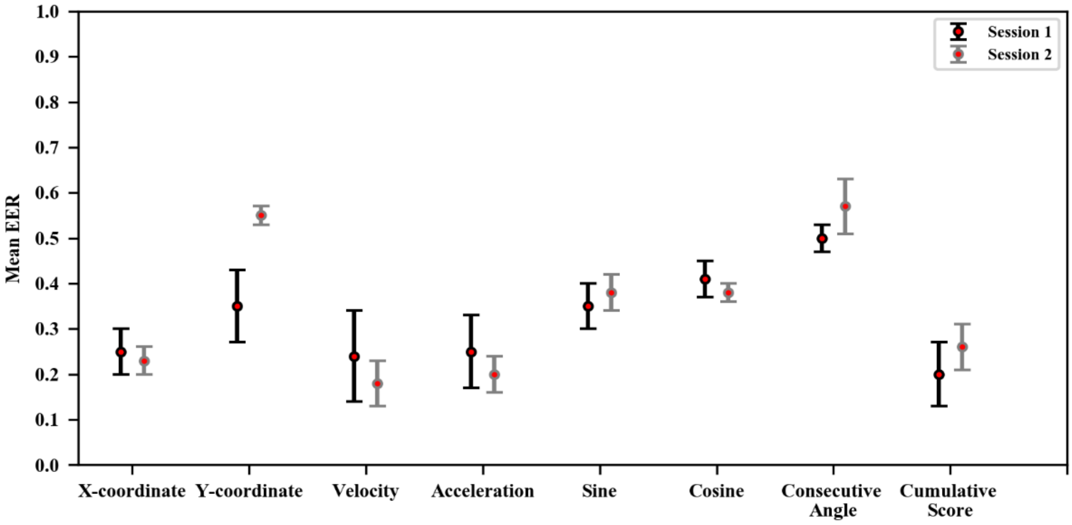


Figure 5.13. Mean EER acquired for finger-based signature comparison using DTW algorithm for Session 1 and Session 2

5.4.1.3 SVM

The linear and RBF kernels were used for the analysis using the SVM algorithm. As used in the other two methods, the enrolled signatures belonged to finger-based signatures acquired from the Sitting Indoors scenario from Session 1 and Session 2 respectively. The verification signatures were from Sitting, Treadmill and Walking of Session 1 and Sitting, Walking and Bus from Session 2.

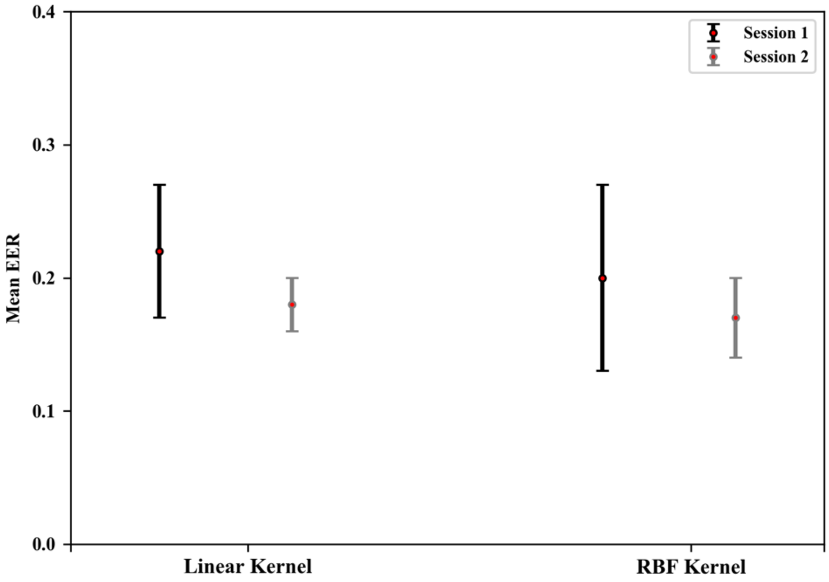


Figure 5.14. Mean EER acquired for signature verification using SVM's linear and RBF kernel for Session 1 and Session 2

Figure 5.14 shows the results obtained for Session 1 and Session 2 respectively using the linear and RBF kernels. It can be seen that the RBF kernel yielded lower mean EER % compared to linear for both sessions. Therefore, RBF kernel configuration was used for the intra-session and inter-session analysis.

5.4.2 Influence of input-type on signature verification

In order to analyse the influence of the input-tool used during signature capture, all three signature verification systems followed the same format of enrolment and verification framework. The systems were enrolled with finger-based signatures acquired from Sitting scenario of Session 1 and Session 2 separately. The verification signatures belonged to both finger and signature-based signatures from Sitting Indoors, Treadmill, Walking Outdoors and Travelling on a moving bus scenario of Session 1 and Session 2. The reason for considering signatures acquired from all the scenarios was to understand the influence of the tool-type on the verification accuracy irrespective of the capturing scenario.

5.4.2.1 Commercial Signature Verification Engine

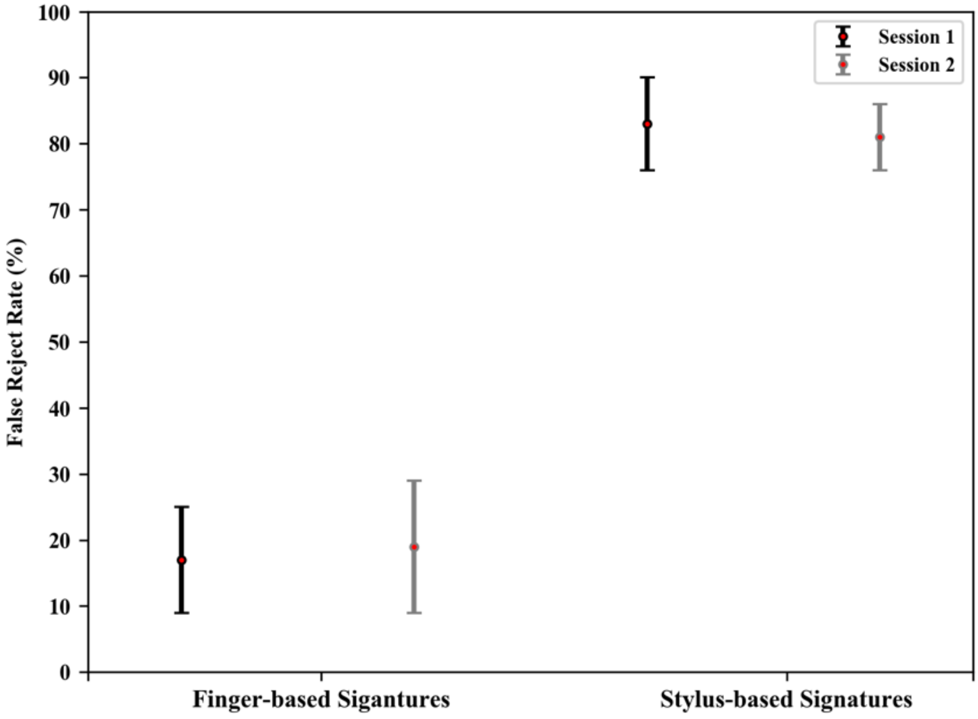


Figure 5.15. FRR percentage acquired based on different input-tools for commercial signature verification system for Session 1 and Session 2

Figure 5.15 presents the results obtained using the commercial system. It shows that the false rejection rates for the stylus-based signatures are much higher compared to finger-based signatures when using the commercial signature verification system. This is because the current system extensively supports only stylus-based signatures from a variety of commercial electronic signature pads such as Wacom tablets. The commercial verification engine is device dependent as it works with a broad range of signature pads of basic and advanced capturing devices from various manufacturers. Some of the devices the system supports are – Wacom STU 430, 530, iPad and Android-powered tablets. All these devices support stylus-based signatures. However, we used finger-based signature in the enrolment for our analysis, therefore the results for comparing finger-based with stylus-based signatures were expectedly poor. Additionally, most of the finger-based signatures that were falsely rejected belonged to different usage scenarios (as shown in Figure 5.18).

5.4.2.2 DTW

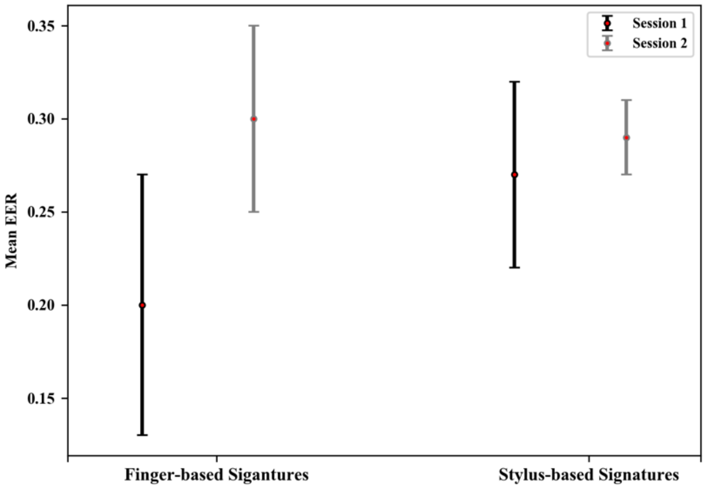
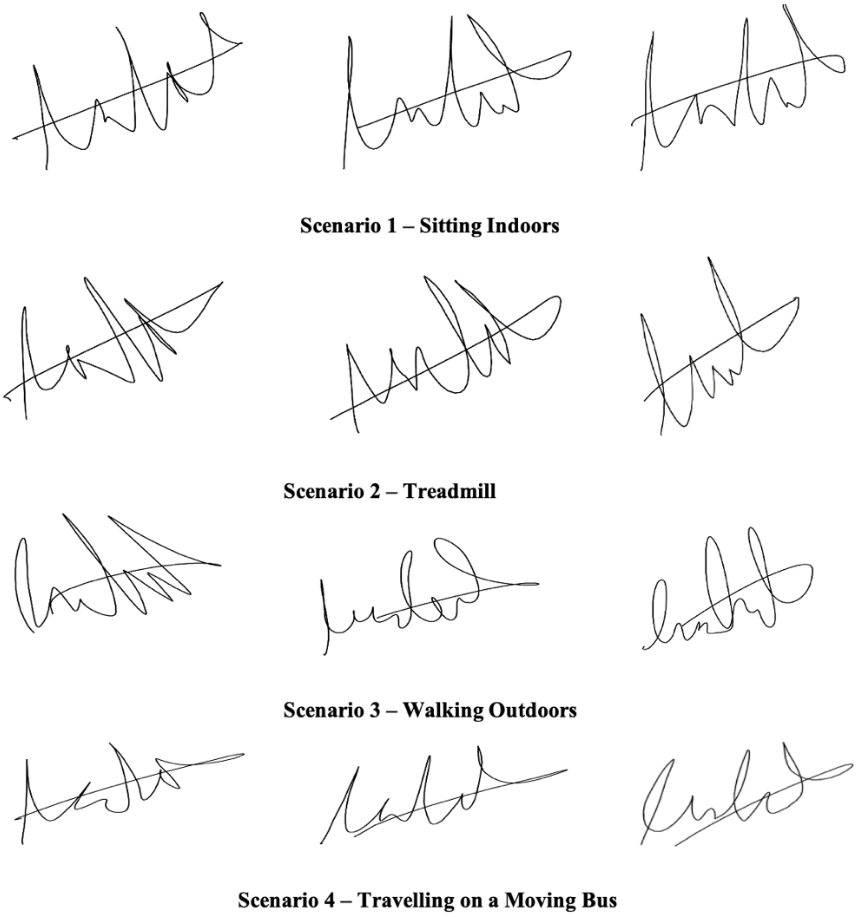


Figure 5.16. Input-tool based analysis using cumulative DTW score method for finger-based and stylus-based signatures for finger-based enrolment

Figure 5.16 shows the influence of input tool type on the verification performance using the cumulative score sequence based DTW comparison. Both, the finger-based and the stylus-based signatures obtained better performance in Session 1 compared to Session 2. This suggests that the signatures acquired in Session 1 were more consistent with each other. Considering Session 2, two of the scenarios (Scenario 3 and 4) present in this session had movements caused either by walking or the transport during the experiment. Hence, the performance of Session 2 was worse compared to Session 1.

5.4.3 Influence of Usage Scenarios of a Mobile Device

In order to evaluate the influence of various usage scenarios on the verification accuracy, signatures acquired from different scenarios within sessions were compared. Figure 5.17 shows various signatures of a user as captured in in different usage scenarios during data collection. The next sub-sections present the results obtained from different signature verification systems. The evaluations using all three systems had finger-based signatures in the enrolment set, belonging to the Sitting Indoors scenario of Session 1 and Session 2 respectively.



Scenario 1 – Sitting Indoors

Scenario 2 – Treadmill

Scenario 3 – Walking Outdoors

Scenario 4 – Travelling on a Moving Bus

Figure 5.17. Signatures captured in different usage scenarios of a mobile device

5.4.3.1 Commercial signature verification system

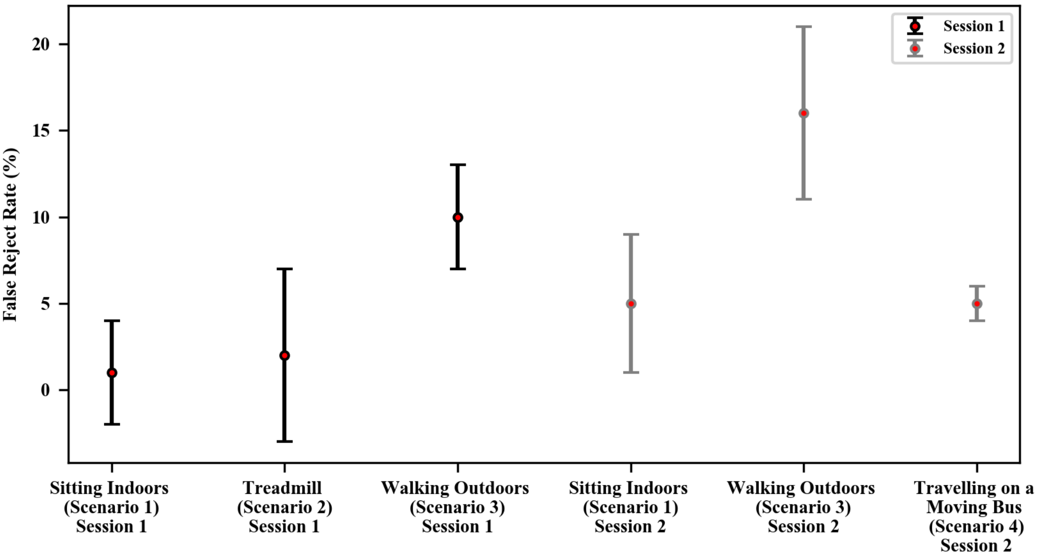


Figure 5.18. False rejection rates of finger-based signatures based on different usage scenarios of a mobile device for Session 1 and Session 2

Figure 5.18 shows the impact of usage scenarios on the verification performance using the commercial system. Based on the genuine comparisons, the false rejection rates from each category of usage scenario were calculated. It can be observed from the figure that the lowest false rejection rate was from the Sitting Indoors and Treadmill scenario of Session 1. The Walking Outdoors scenario from Session 1 and Session 2 yielded maximum false reject rates. Traveling on a Moving Bus scenario showed better FRR compared to the Walking Outdoors scenario of both Session 1 and Session 2.

5.4.3.2 DTW

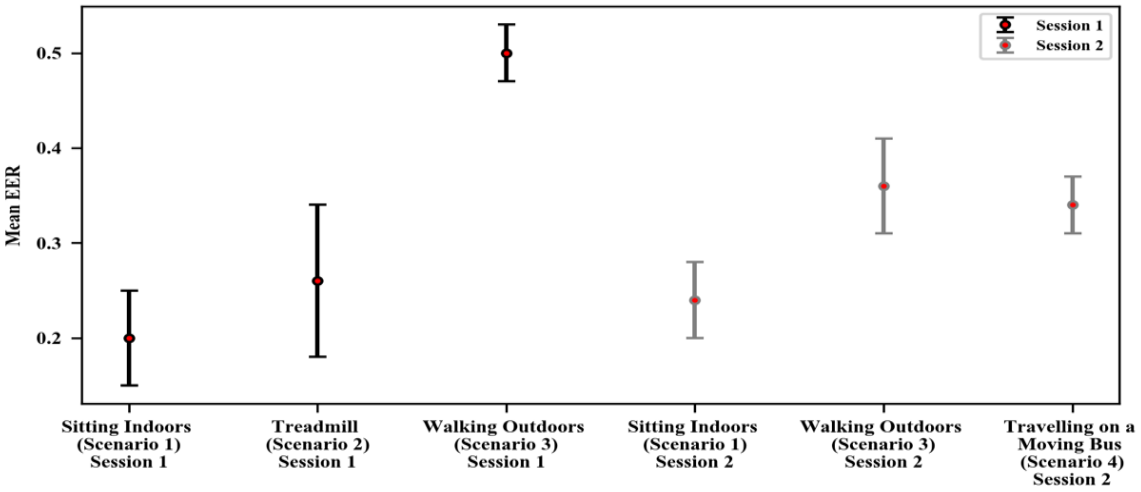


Figure 5.19. Intra-session comparison results using DTW

Figure 5.19 shows the intra-session results obtained using DTW algorithm. The lowest mean EER of 20% and 25% was obtained for the Sitting Indoors scenario in Session 1 and Session 2 respectively. Following this, the signatures in the Walking Outdoors scenario obtained the highest mean EER% both in Session 1 and Session 2. The Treadmill scenario for Session 1 obtained better performance compared to the Sitting Indoors scenario. Travelling on a Moving Bus scenario obtained lower mean EER % than the Walking Outdoors scenario of Session 1 and Session 2.

5.4.3.3 SVM

The results obtained using the feature-based signature verification method utilising SVM classifier are presented in Figure 5.20. The Sitting Indoors scenario yielded the lowest mean EER% for Session 1 (13%) and Session 2 (12%), whilst the Walking scenario of Session 1 and Session 2 yielded highest mean EERs. It can be seen that Travelling on a moving bus showed 17% mean EER, whereas treadmill scenario for Session 1 yielded better mean EER compared to the Walking Outdoors scenario from both the sessions.

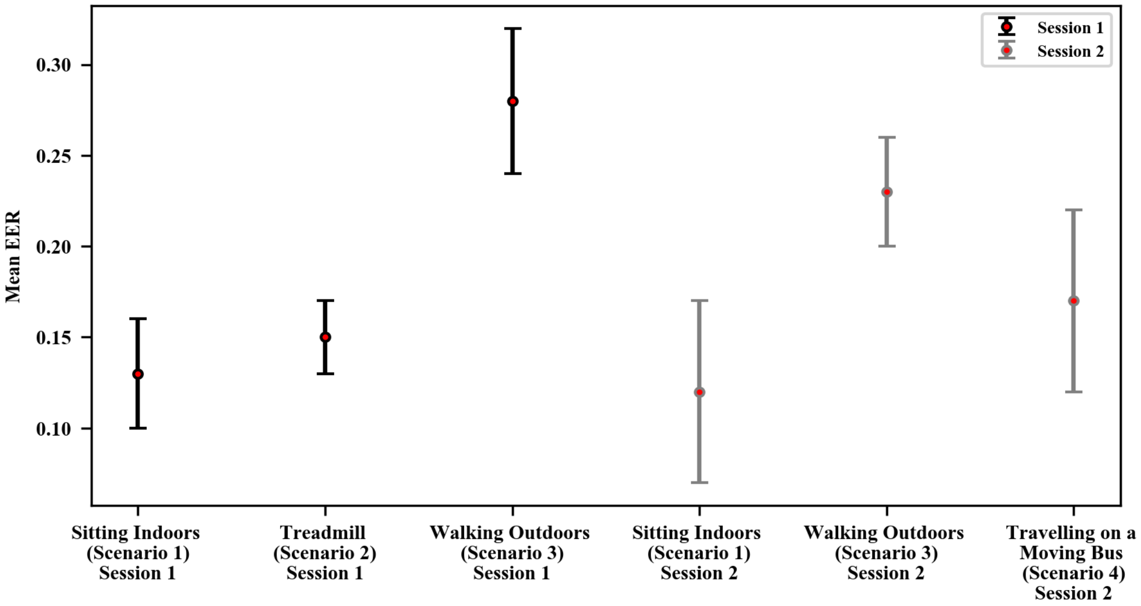


Figure 5.20. Mean EER of usage scenario-based comparison using SVM for Session 1 and Session 2

5.4.4 Influence of time-separated signatures on verification performance

For this analysis, the enrolment was conducted separately for finger and stylus-based signatures. To achieve inter-session comparisons, the Session 1 signatures were compared against signatures acquired during Session 2. For instance, the finger and stylus-based signatures belonging to the Sitting Indoors scenario in Session 1 were used for enrolment and corresponding verification signatures belonged to the Sitting Indoors scenario of Session 2. Similarly, the signatures acquired in the Walking Outdoors scenario of Session 1 were enrolled and compared against the signatures acquired from the Walking Outdoors scenario of Session 2. The results obtained are shown in Figure 5.24.

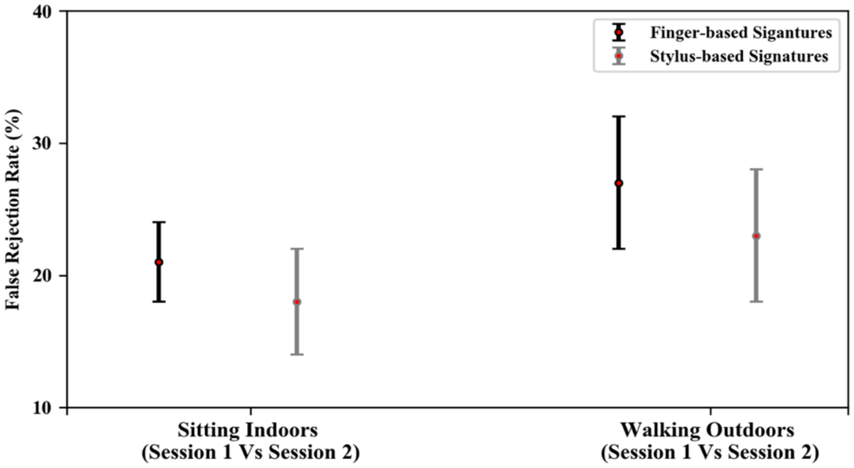


Figure 5.21. Inter-session comparisons of finger and stylus-based signatures using the commercial signature verification engine

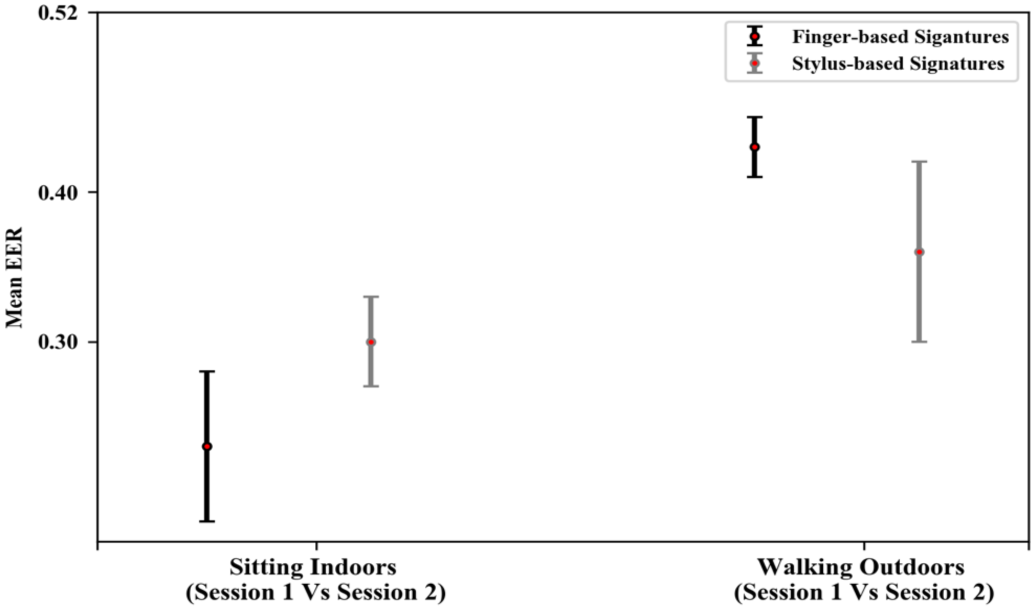


Figure 5.22. Inter-session comparison of finger and stylus-based signatures using SVM

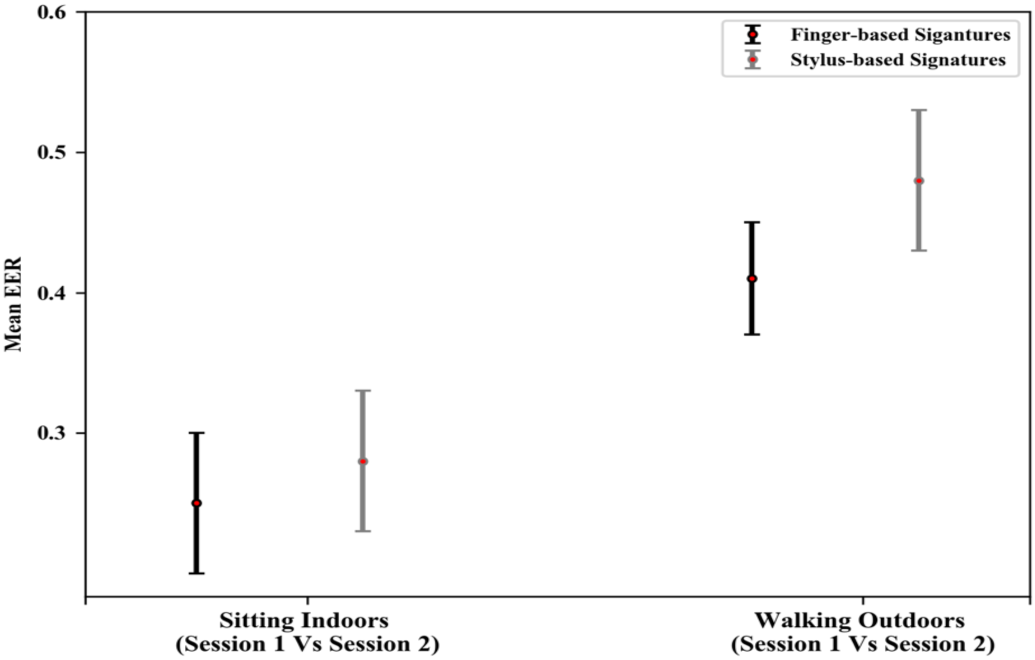


Figure 5.23. Inter-session comparison between Session 1 and Session 2 using DTW

It can be observed from the results obtained using all the three signature verification methods in Figure 5.21, Figure 5.22 and Figure 5.23 that the inter-session results for the Walking Outdoors scenario were worse compared to the Sitting Indoors scenario using both finger and stylus-based signature comparisons. Therefore, it can be concluded that the inter-session results of the scenario having user movements showed poor performance compared to the static scenario. Additionally, it can be observed from the results that the commercial signature verification engine showed better performance for stylus-based signature comparisons compared to finger-based as expected. However, using the DTW (Figure 5.23), the finger-based inter-session comparisons yielded better performance compared to the stylus-based comparisons.

5.5 Conclusion

Online signature verification performed on mobile devices comes with an additional set of challenges compared to the traditional signature verification systems. These challenges arise due to acquiring signatures in dynamic scenarios and uncontrolled environment, which can impact the overall verification performance. In such scenarios, the impact of the motion affects the consistency of signature samples. Similarly, the user interaction factors while donating the signature sample on the device can vary based on the surrounding environment. This chapter focuses on analysing a number of factors using three signature verification methods – a black box commercial signature verification engine, a function-based signature verification method using open access DTW algorithm and a feature-based verification technique using SVM.

First, a performance assessment was conducted using the signatures captured under all usage scenarios and different input-tools. The main motivation to conduct such an analysis was to evaluate the performance of three verification methods utilised. For the evaluations using DTW and SVM algorithms, first five finger-based signatures donated by the user were used for enrolment, whereas for the commercial system, out of the first ten finger-based signatures, five were chosen by the system. The commercial system attained 100% TRR, however, the FRR% results showed a performance deterioration due to signature comparisons of signatures belonging to different scenarios. The mean EERs attained using the DTW and SVM showed similar performances. The performance deterioration was mainly contributed by failure to verify genuine signatures acquired under different usage scenarios or while using a different input tool, compared to enrolled signature data. These results indicate the importance of choosing the appropriate enrolment signatures.

Secondly, the results obtained to assess the influence of input-type on signature verification using DTW algorithm showed acceptable performance variation between the two input types. However, assessment performed using the commercial verification system showed a poor performance when comparing finger-based signatures with stylus-based signatures. However, the commercial system supports only stylus-based devices such as Wacom models, therefore, the input-type comparison results are not conclusive.

Thirdly, the results obtained to assess the influence of usage scenarios on the verification performance showed that the scenario where the user was seated indoors had minimum mean EER% both in Session 1 and Session 2. This showed a stable performance for signatures captured with no movement (bodily and environmental). Analysis of the impact of dynamic usage scenarios using DTW and SVM algorithms show that scenarios having movements from a vehicle or from the user itself have higher mean EER%, thereby yielding a poor performance. Similarly, the commercial signature verification system showed high FRR for Walking Outdoors scenario.

Finally, the results for the inter-session analysis using DTW and SVM algorithms showed performance deterioration for both Sitting Indoors and the Walking Outdoors scenario compared to the intra-session results (for both finger-based and stylus-based comparisons). The results obtained using the commercial system show similar performance when time-separated signatures were compared. This indicates that the verification performance using signatures on mobile devices has challenges with regards to time persistence.

Similar to our results on swipe gestures, the results obtained in this chapter challenge the adaptability of the signature verification algorithms for unconstrained, dynamic usage of a mobile device. This has been observed for both within sessions as well as inter-session comparisons. On the other hand, input-tool type did not show a significant impact on performance. Although more work could be done on that aspect, we also observe that most upcoming smartphones do not feature a stylus, and thus a deeper research into the impact of input tool type might have limited practical value.

Chapter 6. Keystroke Dynamics Verification

6.1 Introduction

Keystroke dynamics is a behavioural biometric modality that utilises the typing behaviour of an individual to establish their identity. Like swipe gestures and signature, keystroke dynamics data can be captured using the touchscreen of a mobile device using the soft-keyboard and, like signature, the data acquisition requires active user interaction with the device. Attributes such as timestamp and the typed key are captured by the operating system of the device. The keystroke events that the OS records are the key-press and the key-release timings. These characteristics are then utilised to build the behavioural model of a user and further used for verification purposes.

Different verification models using keystroke dynamics can be built based on the type of input being captured. The type of input can be categorised as either static fixed input (such as username and password) or free-text input. The main difference between the static and the free-text method is that in order to verify a static input type, an incoming keystroke input is compared against the same fixed-text during the verification process. However, the enrolment samples used in free-text method may be different from subsequent verification samples. When a free-text input is utilised, the concept of continuous verification is applied. Unlike one-time static verification such as a password or a PIN, a continuous verification model authenticates the user throughout the entire session of user interaction. To assume that that the user who logged onto the session is the same user throughout the entire session is rather naive; therefore, continuously authenticating the user becomes important. As the user continues to use the keyboard after logging in, for browsing the internet or chat, it becomes easier to implement continuous verification even after the login process.

In this chapter, a detailed description of a continuous verification model using keystroke dynamics is presented. One of the challenges of using keystroke dynamics in a continuous verification model is that the current techniques require significant amount of keystroke timing information to build a reliable model [28], [130]. In order to address this issue, a continuous verification model using minimal keystroke dynamics data has been designed in this work. This idea was developed to specifically address the issue of data storage on the mobile device, where recording every event data can put load on the device storage and can potentially increase the processing time. As mentioned earlier, traditional models utilise the *key-press* and the *key-release* timing information, however, in this study we developed a distance-based verification configuration that utilises only the *'key-press'* event of the keystroke. In contrast to previous studies that utilised a considerable amount of typing information to generate a verification model, in this chapter, a model that uses a limited amount of keystroke information, acquired using a ceremony-based data collection, in different usage scenarios of a mobile device has been presented. This evaluation is a proof-of-concept that demonstrates a continuous verification configuration, where the enrolment and the verification samples were acquired from different scenarios. The concept of continuous authentication in this study is applied to continuously ensure the genuineness of the user in every key input entered on the device instead of a block of keystrokes (consisting on N key inputs) and the data utilised to evaluate this configuration is based on static keystroke inputs where the user is asked to type a given sentence. However, the enrolment and verification samples were different. The enrolment of the system was performed using a set of sentences acquired at the beginning of the experiment, and the verification samples belonged to new samples. This experiment covers typical alphabetical mobile device interactions of typing text messages.

The main goal of this chapter is to outline a continuous verification model that makes use of keystroke data collected in an unconstrained environment and under unsupervised scenarios. The assessment of the biometric performance of the model under various usage scenarios of the device such as typing whilst seated on a chair, whilst walking on a treadmill machine, whilst walking outdoors and whilst seated on a bus has been utilised for this study. The verification configuration represents a trust-model built on a distance-based algorithm, which calculates a verification score on pairs of key input, and if the score is below an assigned threshold (trust-score), the user is logged out of the session. This configuration was built and evaluated off-device. The configuration uses a sequence of sentences, typed at the beginning of the experiment by the user to generate the enrolment samples. The enrolment samples were stored in the form of digraphs, formed by combination of pairs of keys from the given sentence. The verification process initiated when the user entered at least two characters using the soft keyboard on the touchscreen. The results obtained using this configuration discriminated the genuine from the imposter user.

A number of research questions aligned to the overall research objectives of the thesis have been identified and addressed in this chapter. These questions are listed below:

- Is it possible to build a continuous verification model using limited amount of keystroke information and using only *'key-press'* data?
- Evaluate the impact of amount of enrolment data on the verification accuracy
- Perform intra-session and inter-session comparison of verification accuracy based on keystroke inputs acquired under different usage scenarios and time-separated sessions.

The remainder of the chapter is structured as follows. Section 6.2 details the state-of-the-art studies on keystroke dynamics. The experimental methodology has been presented in Section 6.3. Following this, the performance evaluation results are detailed in Section 6.4 and finally, the conclusions are presented in Section 6.5.

6.2 Related Work

Keystroke dynamics, as a behavioural biometric modality has been widely researched on devices such as computers and laptops, however, using keystroke dynamics for user verification on mobile devices have recently gained traction [131], [31], [132]. A detailed literature review on keystroke dynamics has been presented in Chapter 2. In this section, a number of keystroke dynamics studies focusing on the continuous verification technique has been reviewed and presented. Table 6.1 presents a list of studies using keystroke dynamics and their data capture characteristics along with their overall performance.

Publication	Number of Subjects	Environment	Task	Device	Classifier	EER %
Clarke et al. [133]	32	Constrained	Fixed (4 and 11 Digits)	Nokia 5100	Counter-Propagation Artificial Neural Network (CPANN) and Artificial Neural Network (ANN) And SVM	5%
Clarke et al. [134]	32	Constrained	Fixed (4 and 11 Digits)	Nokia 5100	Neural Network	12.8%
Gascon et al. [32]	300	Constrained	Free text	Smartphone with soft keyboard prototype for the Android OS	SVM	TPR-92%
Wu et al. [135]	10	N/A	Free text	Smartphone	SVM	Average accuracy – 98.6%

Table 6.1. Previous studies on continuous verification using keystroke dynamics

On a mobile device, the keystroke dynamics can be captured using a soft-keyboard. However, the design of the soft-keyboard can differ based on the device model, operating system and operating language. The soft-keyboard format can differ primarily due the difference in the placement of keys in the keypad. Multiple studies have researched using key input from multiple devices having physical in-built keypad and soft keyboard of diverse keypad formats [51], [136], [47] . Newer soft-keyboard layout such as standard QWERTY, Octopus and Swype are emerging continuously. However, the most dominant layouts in Android devices is QWERTY.

Clarke et al. [133] presented results of evaluation of numerical and alphabetical inputs that were captured as entry of both telephone numbers and text messages on a mobile phone. They observed that users were verified with an average EER of below 5%. Following this study, Clarke et al. [134] in 2007, utilised neural network classifiers to authenticate the users using keystroke dynamics data. This model performed robust and transparent verification by maximising the security and minimising user inconvenience. They used a feed forward multi-layered perceptron framework to build the model and, using this framework, their evaluation results yielded an average EER of 12.8%. However, these results are based on the data collected from primitive mobile handsets having physical keyboards built into them.

A few studies have focused on utilising data from other sensors of a mobile device along with keystroke dynamics to build a continuous verification model. Gascon et al. [32] developed a model that captures motion sensor data along with typing to build a unique typing motion profile of the genuine user. In order to conduct the data collection, they built a soft-keyboard prototype of Android OS that recorded user's motion and typing sensor data. They used accelerometer, gyroscope and orientation sensors to measure acceleration, torque and relative position of the mobile device respectively. They utilised free-text data collected from 300 participants. Their results show an FPR of 1%, however, a certain number of users could not be verified properly.

Similarly, Wu et al. [135] used keystroke dynamics along with touch gesture information to demonstrate a transparent and continuous verification model that combines the extracted finger pointing and sliding features using a SVM classifier. They used input from 10 participants with 150 data samples. The average accuracy reported in their experiment is around 98.6% for ten-runs.

This study contains evaluations conducted with real-life data captured under an unconstrained environment. The data used in this study contains keystroke data captured under various usage scenarios of a mobile device. This study also presents performance assessment conducted based on the intra-session and inter-session comparison of the verification. The next sections detail the architecture of the configuration and the evaluations.

6.3 Experimental Methodology

Unlike above studies, the novelty of this work is, in reducing the data load on the model. Limited keystroke data parameter has been used to build this discriminative verification model. Only the *'key-press'* characteristic has been captured and features such as '*digraph'* have been utilised to build the verification model. The next sub-sections provide description of the data, the verification algorithm and the architecture of the verification framework.

6.3.1 Dataset

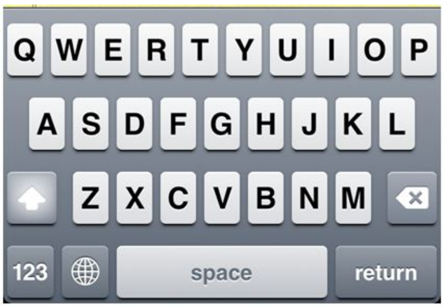


Figure 6.1. Alphabetical input QWERTY layout used in data collection [51]

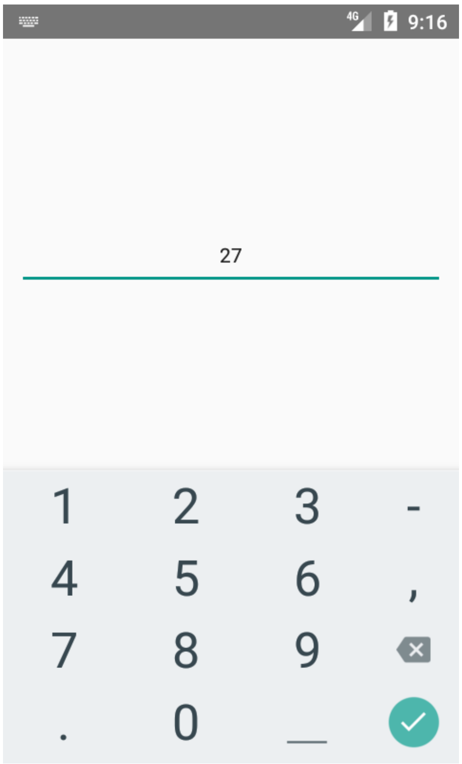


Figure 6.2. Numerical input QWERTY layout used in data collection [137]

The keystroke dynamics data used for this experiment was acquired from the key entry tasks of the data collection described in Chapter 3. The keystroke dynamics data was captured using the touchscreen of a smartphone (Samsung Galaxy Note 5) for 50 participants under various usage scenarios. The keystroke dynamics data was of two types - alphabetical and numerical. The keyboard layout used for the keystroke input was standard QWERTY, as shown in Figure 6.1(alphabetical) and Figure 6.2 (numerical).

The keystroke data entry task was in the form of typing a fixed sentence and a phone number on the mobile device. The reason to obtain fixed sentences instead of a password-like word during the data collection was to generate the inputs that could be used for developing a continuous verification model. The alphabetical data consisted of typing a few commonly used sentences as shown in Table 6.2.

Alphabetical Input	Numerical Input
Dear all, I am working from home today.	0044 7900412433
Please find attached the document along with this mail.	0044 7854632190
I will call you soon. Please wait for my call.	0044 7985642316
Hi, How are you doing? I am fine. Are we meeting today?	0044 7195682453
Happy birthday dear.	0044 6572839450

Table 6.2. Alphabetical and numerical input during the typing tasks of the data collection

The sentences had a combination of upper and lower-case letters with spaces between words. The most commonly used special characters - full stop, comma and question mark were also used. The numerical data was in the form of a ten-digit phone number along with the dial code of the United Kingdom (e.g. 0044 7985412436).

Each key entry event on the touchscreen, whether alphabetical or numerical, generated a number of parameters – timestamp, character being typed, scenario id, deleted character and the position of the character being inserted/deleted (as highlighted in Table 6.3). This information was saved for every key entry. The timestamp refers to the key-press time, as only the key-press time was recorded. A sample event recorded in key-press and key-delete are given in Figure 6.3.

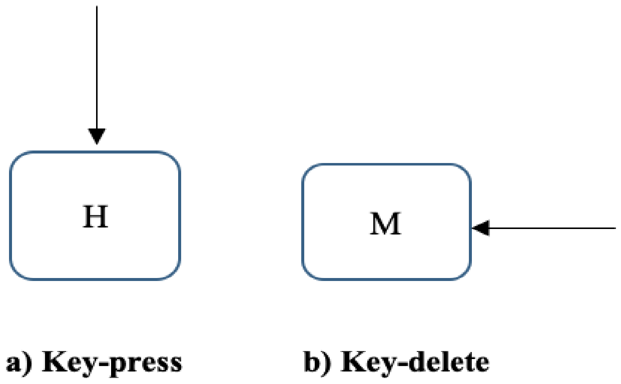


Figure 6.3. Illustrative key-press and key-delete events for keys H and M

Keypress Event	Timestamp	Character	Scenario ID	Position
Keypress H	18/04/03 13:49:23:794	H	1	1
Keypress M	18/04/03 13:49:23:996	M	1	2
Keypress Backspace	18/04/03 13:49:24:003	Deleted	1	2

Table 6.3. Keypress events data generation

Data saved in these key presses are – '*Keypress H'* - Timestamp, H, 1, 1 and '*Keypress 2'* - Timestamp, M, 1, 2. For a deleted character keypress, the character parameter is saved as '*Deleted'* and the position parameter captures the position of the deleted character. The position is an incremental number assigned to every key entry. The key entry position always initiates at 1 and therefore, the deleted character position was calculated based on this.

As described in detail in Chapter 3, the keystroke dynamics data was collected under various usage scenarios of a mobile device. The scenarios *Scenario 1* involved typing while being seated on a chair indoors; *Scenario 2* involved typing while walking on a treadmill, *Scenario 3* involved typing while walking outdoors and *Scenario 4* involved typing while seated on a moving transport. Details of the number of sample donations for alphabetical and numerical data for an individual user from the dataset have been provided in Table 6.4. The data from all 50 participants were used for the analysis. Each of the samples was further divided into words and number of characters per sentence.

Input Type	Scenario 1		Scenario 2	Scenario 3		Scenario 4
	Session 1	Session 2	Session 1	Session 1	Session 2	Session 2
Number of sentences	5	5	3	5	5	3

Table 6.4. Keystroke dynamic data collected per individual user from the dataset

6.3.2 Feature Extraction

As described in Section 6.3.1, the raw key touch event generated a number of parameters – timestamp, key being pressed and position of the key entry. These parameters were stored in the mobile device during the experiment as the user performed typing tasks on the app. On the completion of the experiment, these data were extracted from the device and used for creating the verification framework. From these raw parameters, a number of digraphs were generated. A digraph is defined as the key-value pair of two consecutively typed keys. It consisted of the digraph key and flight time value. The digraphs can be of different categories (as shown in Table 6.5). These digraphs differ based on the characters being typed. The reason for utilising only the digraphs (and not trigraphs) was to reduce the complexity of computation in the configuration. For a digraph, the number of comparisons for the similarity score generation is limited to the number of enrolled digraphs. However, these comparisons would increase with the increase in number of latency graphs included in the verification configuration, hence increase in time and complexity.

Digraph Types	**Digraph identifier value** **Illustrative sentence – Hi, how are you?**
Letter to Letter	Hi, ho, ow, ar, re, yo, ou
Letter to Space	w_, e_
Space to Letter	_h, _a, _y
Space to special character (Special characters -'?', ',', '.')	None
Special character to space	,_ , ?_

Table 6.5 Digraph types for the input 'Hi, how are you?'

As the keypress data contained only the keypress time, the flight time was calculated as total time between one keypress to the next keypress. Therefore, only two basic features are used to build this verification model –

- Digraph key
- Flight time of digraph

Every incoming keystroke data was used to generate digraphs. These digraphs were stored in the template library during the enrolment phase and then compared with incoming digraphs.

6.3.3 Verification Architecture

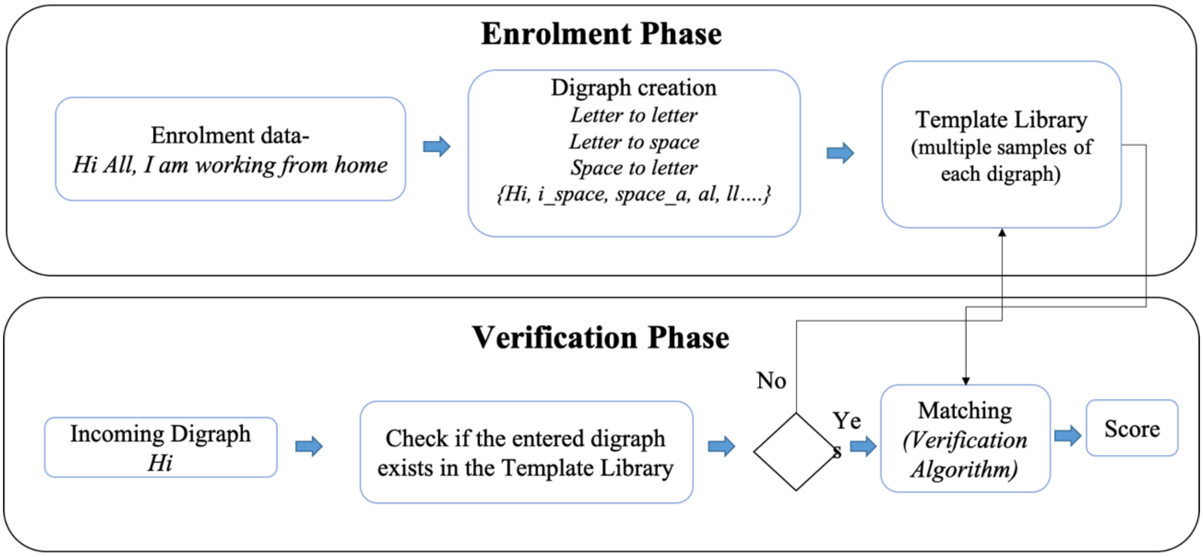


Figure 6.4. Verification configuration

Unlike signature verification where a static method was used to compare a fixed input to the templates stored in the template library, a continuous verification-based verification method was chosen for this analysis. This was motivated by the fact that behavioural biometric modalities are extensively used for continuous verification as well. Since the collected keystroke dynamics data was in form of a number of sentences and phone numbers instead of fixed words such as password or PIN pattern, it was apt to be utilised for a continuous verification model.

The continuous verification configuration which was developed as an initial proof of concept has been depicted in Figure 6.4. As shown in the figure, the configuration was divided into two phases – enrolment and verification. In the enrolment phase, the input samples were used to build the template library for individual users. Every input sample was used for generating different digraphs of various categories such as letter-to-letter and letter to space. Once these digraphs were created, associated data of the digraphs were stored in the template library to be used in the verification phase. Multiple digraph samples can exist for the same category of digraph key with different flight times. In the template library, the features saved were – *digraph identifier and flight time*. For multiple samples of the same digraph, each was saved with an *id*. For example, 'om' was stored with om_1 and the next occurrence of 'om' was saved as om_2.

In the verification phase, digraphs were created from incoming key input. The digraph creation process was initiated when at least a minimum of two key press events were generated. During the verification process, the first step was to check if the digraph was present in the template library. If the digraph was not present in the template library, a new sample of the digraph was created in the template library. On the contrary, if the digraph was present in the template library, the matching process using the verification algorithm was initiated. This generated a match score using a distance-based algorithm (described in next section) for that particular digraph. If the generated score was below a defined threshold, the user was logged out of the session.

6.3.4 Verification Algorithm

Figure 6.5 depicts the algorithm used for keystroke verification. The verification algorithm consisted of two steps – a) distance calculation and b) distance to score conversion. After checking if the incoming digraph existed in the template library, the distance calculation step was carried out. This step involved calculation of Euclidean distance of the flight time of the incoming digraph with every existing digraph in the template library for the same digraph keys. Once the distances were calculated, they were normalised using min-max normalisation method and they were stored as $D_1, D_2 ,.., D_N$, where *'N'* denotes the total number of samples in the template library of a particular digraph.

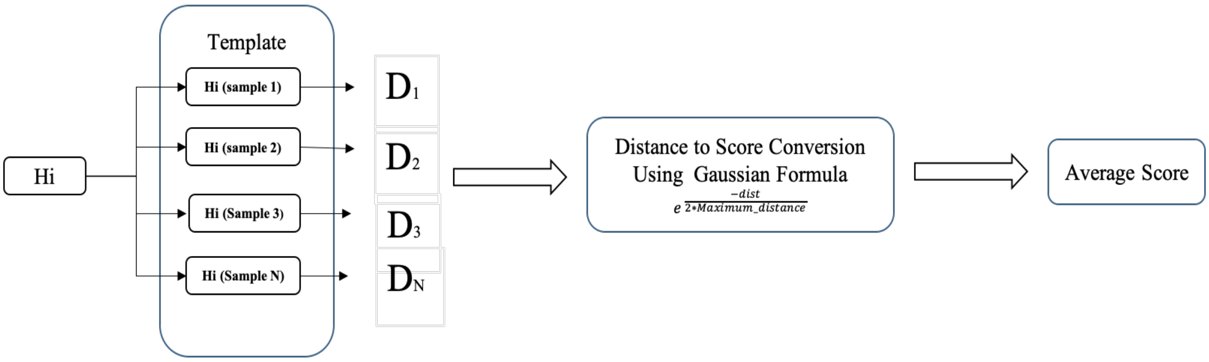


Figure 6.5. Verification algorithm steps

The total number of distances equals the total number of samples of a given digraph present in the template library. For each distance, a score was generated using the Gaussian formula (as shown in Figure 6.5). In the final step, these generated scores were averaged to arrive at the final match score for given input digraph.

As described in earlier sections, the data collection was part of a planned experiment and thus, real-time key typing data could not be gathered for verification. The analysis was conducted off-device and multiple strategies were applied with regards to number of enrolment samples. The description of the enrolment and verification data has been provided in the next section.

In this chapter, the results from only the alphabetical input type have been reported. This is because enrolment and verification process required a minimum of 50 digraphs to attain reasonable performance and the numerical data for each user was insufficient.

6.3.5 Enrolment and Verification

The continuous verification model was built based on a binary classification problem for classifying a genuine and an imposter user. Separate models were enrolled with genuine samples after which genuine to genuine and genuine to imposter comparisons were made. The genuine user was considered as the owner of the mobile device and the imposter user was randomly chosen from the dataset. The genuine user was given a user id as ‘0’ and imposter user as ‘1’. Each user from the dataset of 50 participants generated a verification model where they were enrolled and verified using their alphabetical and numerical data respectively.

The keystroke data used for enrolment was taken from the baseline scenario (Sitting Indoors - Scenario 1) of Session 1 and Session 2. Out of the five sentences captured in Scenario 1, three sentences were used for enrolment and the remaining two were used for verification. As the users entered multiple numbers of key entries, the total number of characters in the sentences and digraphs generated per user for different scenarios are provided in Table 6.6.

Scenario	Session	Total number of characters	Total number of digraphs
Scenario 1	Session 1	215	213
	Session 2	215	213
Scenario 2	Session 1	140	137
Scenario 3	Session 1	215	213
	Session 2	215	213
Scenario 4	Session 2	105	102

Table 6.6. Scenario based digraph information for individual user

The digraphs generated above were divided in the enrolment and verification sets. Different enrolment strategies were implemented based on the research questions being addressed.

- For evaluating the impact of enrolment data on the verification accuracy, 20, 30, 40 and 50 digraphs were used in the enrolment set.
- For the intra-session comparison, when within-scenario comparison of Scenario 1 Vs Scenario 1 (Session 1) was carried out, first 50 digraphs were enrolled, and the rest of the digraphs were present in the verification set. When different scenarios were compared, the enrolment set contained data from all the digraphs from Scenario 1 (Session 1) and the verification set belonged to all the digraphs from Scenario 2 and Scenario 3 of Session 1. Similarly, for Session 2, when within-scenario comparison of Scenario 1 Vs Scenario 1 (Session 2) was carried out, first 50 digraphs were enrolled, and the rest of the digraphs were present in the verification set. When different scenarios were compared, all the digraphs from Scenario 1 were present in the enrolment set and all the digraphs belonging to Scenario 3 and Scenario 4 were present in the verification set.
- For the inter-session comparison, the enrolment was performed with all the digraphs present in Scenario1 and Scenario 3 of Session 1 and verified against all the digraphs belonging to Scenario 1 and Scenario 2 of Session 2 respectively.

Based on these enrolment and verification strategies, multiple results were generated, which are presented in the next section.

6.4 Results

This section presents the results acquired from the evaluations. Research questions described in Section 6.1 have been addressed individually in this section. The metric used to evaluate each research question is based on EER.

6.4.1 Build continuous verification model using only key-press data

The continuous verification model using only the '*key-press*' parameter has been developed based on the description provided in Section 6.3.3 and 6.3.4. For this evaluation, the alphabetical data from the baseline scenario (Sitting Indoors - Scenario 1) was utilised. For every genuine user, 50 digraphs were randomly chosen from user's entire digraph set belonging to that scenario and classified as enrolment set. The remaining digraphs were separated in different batches of verification sets in order to make multiple rounds of verification runs. Once the enrolment process was performed using the samples from the genuine user, an imposter user was randomly chosen from the dataset in order to make the imposter comparisons. Equal number of genuine and imposter comparisons were carried out in order to avoid bias due to class. A total of five runs were made with different samples in the verification set for both genuine and imposter comparisons. The match scores generated for two verification runs using this method for the genuine and imposter user are provided in Figure 6.6 and Figure 6.7.

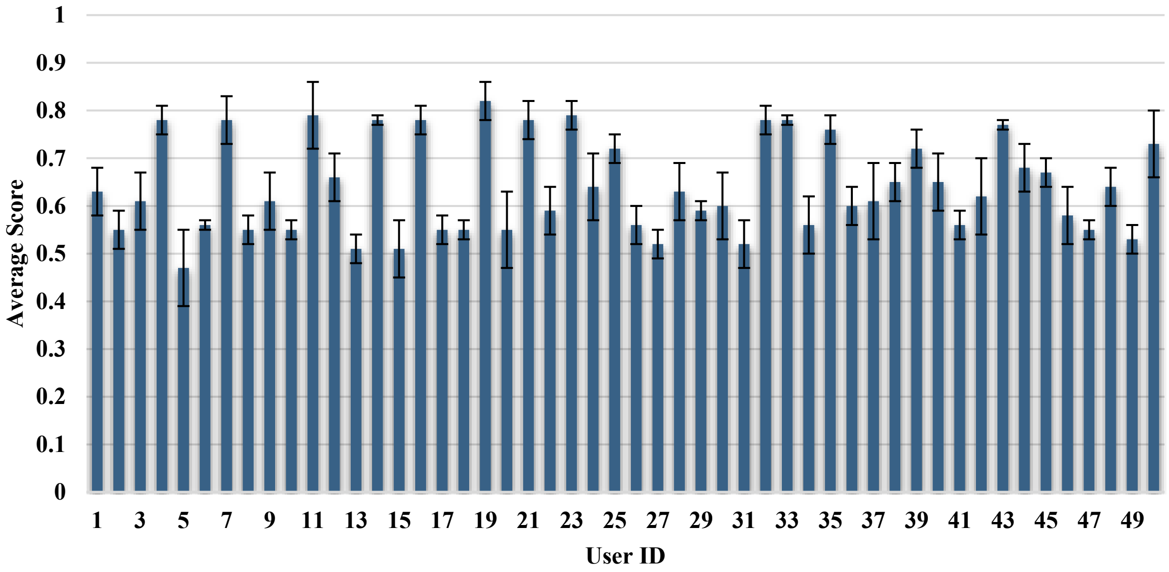


Figure 6.6. Genuine score distribution of 50 users from the dataset

Figure 6.6 shows the average match score acquired for genuine comparisons. For certain users such as user ID 4, 7, 32, 33, 43 and 50, the results show that the obtained average score is above 0.7. User ID 5 obtained the lowest average match score. In the similar manner, the average match scores acquired from the imposter comparisons are shown in Figure 6.7. The imposter scores for three users are below 0.06.

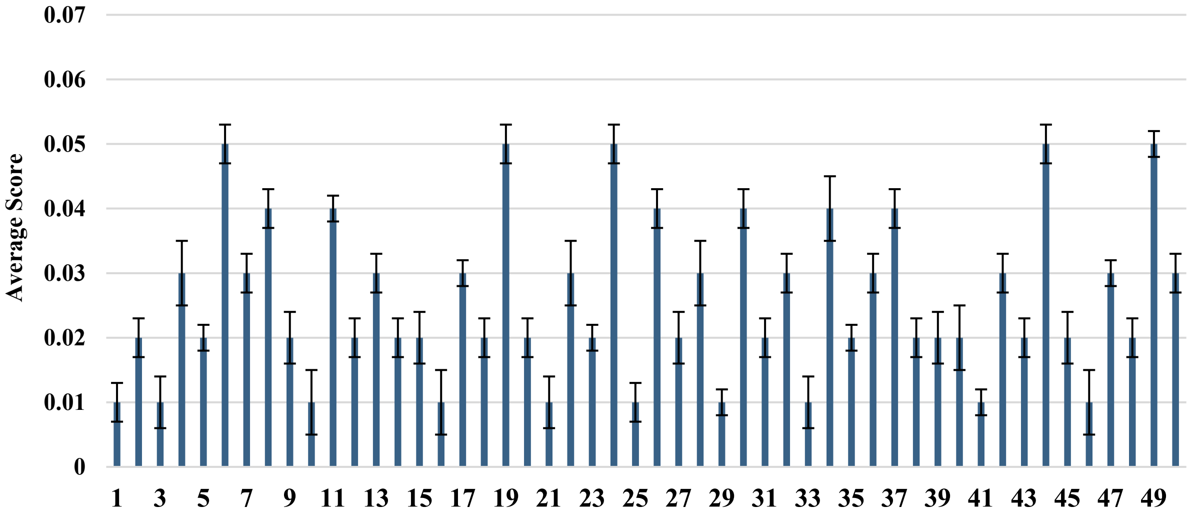


Figure 6.7. Imposter score distribution of 50 users from the dataset

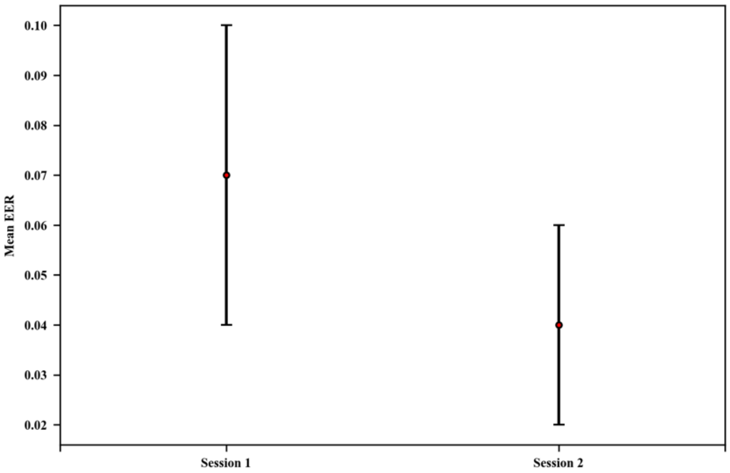


Figure 6.8. Mean EER % acquired for Session 1 and Session 2 of enrolment and verification belonging to Scenario 1

Using this method, the mean EER's attained were 7% for Session 1 and 4% for Session 2 (Figure 6.8). The reason for a higher mean EER in Session 1 compared to Session 2 could be due to the user's lack of familiarity with the typing tasks during the experiment. The acquired results are only for the baseline scenario. The reason for choosing the baseline scenario for evaluating the performance was to avoid the impact from factors such as body movement on the performance. The motivation to select random samples for enrolment instead of first 50 digraphs in the dataset was to assess the performance when the enrolment set consisted of samples captured at different time frames during the data collection. Based on the acquired results, the mean EER % attained for both Session 1 and Session 2 are in acceptable range.

6.4.2 Evaluate the impact of enrolment data on verification accuracy

The second research question focuses on evaluating the impact of the amount of enrolment data on the verification performance. In order to do this, the number of enrolment data considered in the enrolment set were taken from the first 20, 30, 40 and 50 digraphs of each user. For each of the sessions, the enrolment data belonged to Sitting Indoors (Scenario 1) of Session 1 and Session 2 respectively. The verification set contained the remaining digraphs from Sitting Indoors (Scenario 1), both for Session 1 and Session 2 separately.

As shown in Figure 6.9, when only 20 digraphs were used in the enrolment set, a mean EER of 48% was obtained. This mean EER % successively improved with more digraphs in enrolment data. With 50 digraphs in the enrolment set, a mean EER of 2% was obtained for Session 1. This outcome indicates that with the increase in number of digraphs in the enrolment set, better performance is obtained.

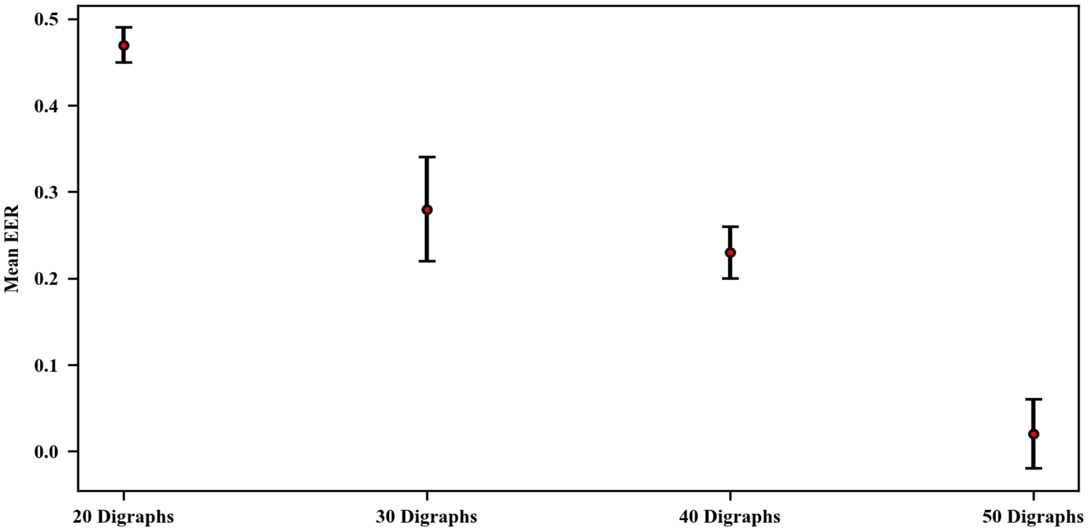


Figure 6.9. Mean EER obtained for different numbers of samples in the enrolment set for Session 1

6.4.3 Intra-session and Inter-session comparison

The intra-session comparison was performed to compare verification performance under various scenarios to evaluate the impact of user interaction factors on verification. The enrolment data belonged to the baseline Sitting scenario of Session 1 and Session 2. The verification data belonged to Scenario 2 and Scenario 3 of Session 1 and Scenario 3 and Scenario 4 of Session 2. Since previous results showed best performance with 50 digraphs as enrolment data, thus in order to conduct this evaluation first 50 digraphs from Sitting Indoors (Scenario 1) were used for enrolment. The verification set consisted of all digraphs belonging to Scenario 2 and Scenario 3. When Scenario 1 was compared with Scenario 1, the remaining digraphs (excluding the enrolment set) were present in the verification set.

6.4.3.1 Intra-session comparison

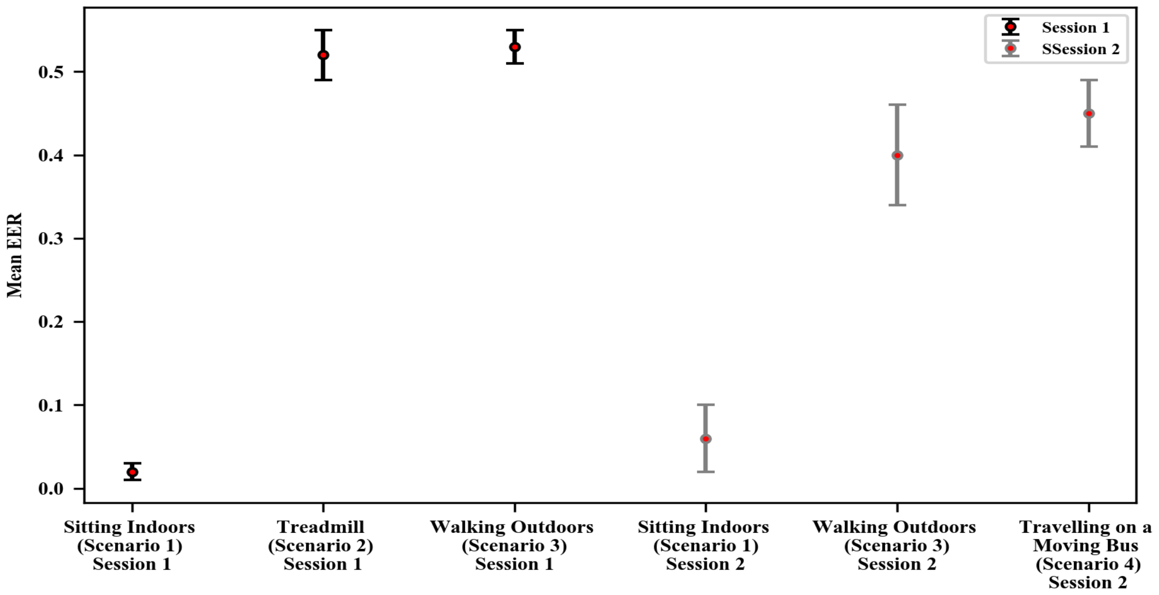


Figure 6.10. Mean EER obtained for intra-session comparison for Session 1 and Session 2

Figure 6.10 shows the intra-session comparison results for *Session 1* and *Session 2* respectively. When the enrolment and the verification data belonged to the same scenario, the mean EER obtained was the lowest with 2% Session 1 and 6% for Session 2. However, when the verification data was from Scenario 2 and Scenario 3, the mean EERs acquired were 50%. A similar trend can be seen for Session 2 comparisons – Scenario 1 Vs Scenario 2 and Scenario 1 Vs Scenario 3. However, the mean EER % for Scenario 3 and Scenario 4 of Session 2 were lower than that of Scenario 2 and Scenario 3 of Session 1.

6.4.3.2 Inter-session Comparison

The inter-session comparison was conducted by comparing keystroke data belonging to Session 1 with Session 2. The idea behind conducting an inter-session comparison is to verify if the model can verify the data having time difference between them. Session 1 and Session 2 had a one-week time difference between them.

Figure 6.11 shows the results of the inter-session comparisons. The mean EER obtained for the walking outdoors scenario was higher compared to the Sitting scenario. Even for sitting scenario, inter session comparison has a far higher mean EER value compared to the intra-session results. This signifies that when time-separated data are compared, the model does not perform as expected.

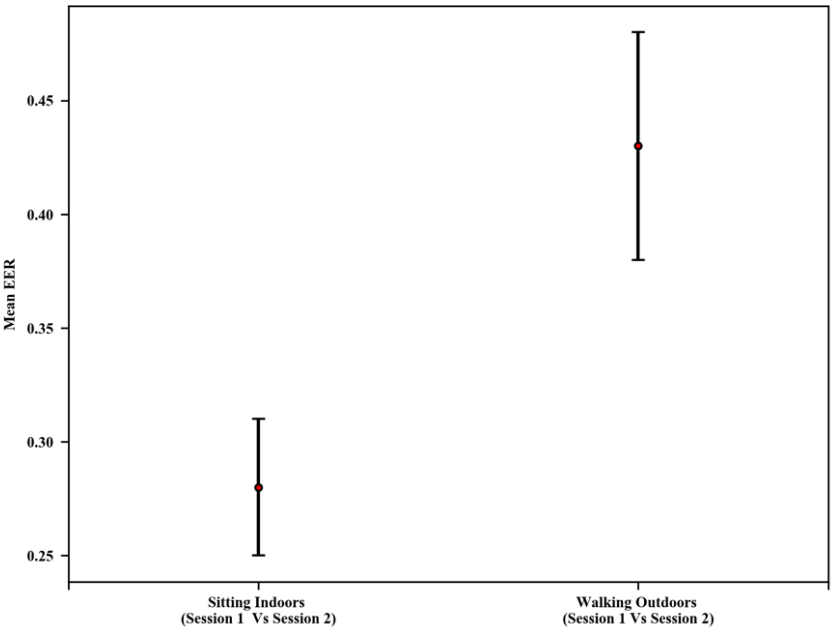


Figure 6.11. Mean EER attained from inter-session evaluations

6.5 Conclusion

The main goal of this chapter was to develop a verification model using minimal keystroke dynamics data and to further evaluate the reliability of the model with increased enrolment data, variation in verification performance owing to different usage scenarios and time-separated data inputs. This idea was developed to address the issue of data storage on the mobile device. Recording every event data can put load on the device storage and can potentially increase the processing time. Given that the data processing capabilities of a mobile device is limited compared to a normal computer, it was useful to explore the options to reduce the data load on the verification model. With this motivation, we designed a proof-of-concept experimental verification algorithm that only utilises the key-press feature of the key event from the device. The key event parameters are converted into digraphs and the verification is performed using these digraphs. The results reveal a clear distinction of the matching scores generated for the genuine and imposter comparisons and a mean EER in acceptable range with sufficient enrolment data.

The results obtained on the impact of enrolment data on the continuous verification model reveal variation in mean EER percentage with different digraphs used in enrolment. With the increase in the enrolment data, the performance got better. Based on the results obtained in this research, at least 50 digraphs need to be enrolled in order to get acceptable error rate for the alphabetical input type comparison.

Finally, the evaluation results of the inter-session and intra-session comparisons reveal that using only the *'key-press'* data is not sufficient to build a reliable model that can deliver on challenges of adapting to multiple usage scenarios and be time persistent. This is because, the obtained verification performance when the enrolment and verification data belonged to different usage scenarios showed poor results. However, the Sitting Indoors Vs Sitting Indoors in Session 1 and Session 2 showed better performance.

Similar performance deterioration can be seen for inter-session comparisons. Results show that the scenario with user movement (Walking Outdoors) performed worse than the static scenario. Additionally, compared to the intra-session results, the inter-session results for both static and dynamic scenarios showed performance deterioration. These results signify that the key-press data varies significantly for a person collected on different days. Therefore, the reliability of the continuous verification model based only on the keypress characteristics becomes questionable. Our recommendation would be to adopt a variable threshold-based model when comparing keystroke data with time-difference. This can possibly reduce the false rejections caused due to intra-personal variations over time.

Chapter 7. Multi-modal Verification

7.1 Introduction

Previous chapters, Chapter 4, Chapter 5 and Chapter 6 focused on uni-modal based verification techniques using swipe gestures, signatures and keystroke dynamics respectively. The results showed the impact of user interaction, as captured in various usage scenarios, on the biometric performance of the uni-modal systems. Literature also shows that a biometric verification system using a single trait is susceptible to issues such as spoof-attacks and intra-class variability [17], [18], [138]. Owing to these existing challenges, our research direction shifted towards exploring methods to enhance the reliability and security of the touch-dynamics based biometric verification. In view of this, a multi-modal verification framework was developed that utilises characteristics from multiple touch modalities for performing the verification. Unlike a uni-modal system, these systems depend on multiple sources of information to establish the identity of a person, hence enhancing security as it gets difficult for an attacker to spoof multiple traits simultaneously of a genuine user.

A multi-modal system can be built based on physiological or behavioural modalities, or a combination of both. However, literature reveals limited work on integrating exclusively touch-dynamics based behavioural biometric modalities in a multi-modal verification system specifically performed on mobile devices. Based on this identified need, in this chapter an in-depth investigation of fusing multiple touch-dynamic based behavioural biometric modalities has been explored. Three modalities – swipe gestures, signature and keystroke dynamics have been integrated, multiple evaluations have been conducted, and the results are presented in this chapter. The novelty of this work is, in this chapter, the impact of multiple usage scenarios across the three simultaneous modalities has been evaluated in order to establish the reliability and accuracy.

In this chapter, performance assessment of different combinations of touch-dynamics based modalities has been reported. In order to conduct the assessment, a multi-modal framework has been developed using a feature-fusion method which combines two or more biometric modalities. This system makes use of multiple traits obtained from a single sensor of the mobile device. Only the touch screen sensor of the mobile device has been utilised to acquire the data as using multiple sensors can introduce additional noise. The acquired data was presented to multiple classifiers (SVM, k-NN and Naïve Bayes) for verification. Additionally, a score-fusion method was evaluated for signature and swipe gestures using the commercial signature verification system.

Based on the identified problems in the domain of multi-modality, various research questions aligning to the overall research objectives (described in Chapter 1) were developed. They are as follows:

- Does a multi-modal verification system using touch-dynamics based behavioural biometrics improve verification accuracy compared to a uni-modal solution?
- Does the impact from usage scenarios be seen using the multi-modal solutions?
- Combine different touch-dynamics based modalities and identify which combination performs the best.

The remainder of the chapter is organised as follows - Section 7.2 details the related work on this topic, Section 7.3 presents the description of the experimental framework. The detail of the dataset used, modes of operation, integration strategy and the multi-modal framework have been provided in this section along with the feature fusion and verification phase information. Following this, Section 7.4 presents the results obtained based on different evaluations. Section 7.5 presents conclusions drawn based on the results.

7.2 Related Work

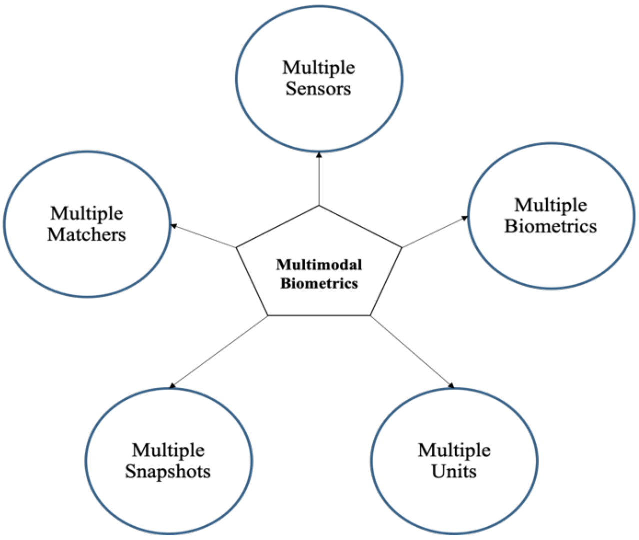


Figure 7.1. Multimodal biometric scenarios [128]

Typically, a multi-modal system adopts a specific fusion topology. The categories of the fusion topology taken from [139] are presented in Figure 7.1. Choosing the fusion topology depends on the number of modalities, feature sets and sensors. Based on these factors, the types of fusion topology are categorised by - multiple biometrics/modalities, units, snapshots, matcher and sensors. The multiple biometric/modalities method is used when different types of biometric modalities are combined. Multiple unit fusion involves utilising two or more units of the same biometric modality, for example, finger and stylus-based signatures. When two verification attempts or enrolment templates of the same biometric trait are used, it is known as multiple snapshot fusion. When multiple matchers are used on the same biometric trait, it is called as multiple matcher fusion. These classifiers can use different feature sets of the same modality or use the same feature set for processing. In case where a single biometric modality is captured through multiple sensors and the data from individual sensors are used for fusion, this method is known as the multiple sensors' fusion method.

Literature reveals that the multi-modal systems on mobile devices adopt one of the above-mentioned fusion topologies. The most common method of fusion is multiple biometric fusion. With respect to behavioural biometric modalities, two categories of biometric fusion are seen in the literature - fusing only behavioural traits [140], [141], [142] or fusing physiological and behavioural biometric modalities together [89]. Chapter 2 presents the review of a number of studies using behavioural modalities for multi-modal based biometric systems.

Saevanee et al. [140] used behavioural profiling, linguistic profiling and keystroke dynamics fusion. Their experimental results showed that via text-entry method, the users can be authenticated with an average EER of 3.3%. They also report 91% reduction in the number of intrusive verification requests using this system. Tanviruzzaman et al. [141] used location tracts and gait signals to generate a multi-modal system of 13 users. The data was captured using Google Nexus S device and acquired an average EER of 10%. Xu et al. [21]'s study combined touch-dynamics based modalities on mobile devices such as keystroke dynamics, handwriting, swipe, pinch and slide. Their experiment was conducted on the data collected of 30 users on a Samsung Galaxy SII device. They reported an average EER of 0% for slide and pinch when 3 to 5 consecutive operations are combined. However, keystroke and handwriting did not show a stable performance over time.

In conclusion, efforts are being taken to develop multi-modal systems with optimised accuracy and sensor data availability. This is an emerging domain, especially in terms of using behavioural biometric modalities in a multi-modal framework. Our work aims at exploring fusion of touch-dynamics based modalities acquired from a mobile device. The novelty of this experiment is that we have investigated the stability of fusing behavioural data in different usage scenarios and time-separated data , in context of a multi-modal application on mobile device. In the context of mobile biometrics, a feasible system utilising active-user touch interaction on a mobile device needs to be explored. Therefore, in this chapter, a *single sensor* based multi-modal system utilising swipe gestures, signature and keystroke dynamics has been developed and discussed. A number of evaluations were carried out based on this system and the results are presented in Section 7.4.

7.3 Experimental Framework

This section describes in detail the dataset used, modes of operation, integration strategy, classifiers and the framework design.

7.3.1 Dataset

The dataset used for this experiment is the multi-modal dataset described in Chapter 3. A total of 50 participants donated three different touch-dynamics based behavioural biometric modalities - swipe gestures, signature and keystroke dynamics in two sessions, separated by a week. The data was acquired under various usage scenarios as well. Different UI contexts were used to capture these data. For example, for a keystroke task, the UI context was an alphabetical key entry task using a soft keyboard on the mobile device and for swipe gesture capturing task, it was image navigation for vertical and horizontal swipes. As the participants used fingers to perform the swipe gesture and keystroke dynamics tasks, therefore, only the finger-based signature data were used in this multi-modal experiment. The stylus-based signature data were excluded.

7.3.2 Mode of Operation

A multi-modal system can work in different operational modes – serial or parallel. In the serial mode of operation, the outcome of one modality is used to verify the identity before using the next modality. When the outcome of multiple modalities is used at the same time in the verification process, it is known as parallel mode of operation.

With regards to the data capturing method, the modalities were captured in serial mode, one after another. In this study, only the touch sensor of the mobile device has been utilised for data acquisition, therefore, the data inconsistency from other sensors was avoided. However, in the data collection setup, the touch operation of all three modalities were independent of each other. That is, only one specific modality was captured at a given time and therefore the modalities could not be captured in parallel.

With regards to the data processing method, parallel mode of operation was chosen. The experimental evaluation was performed off-device. Therefore, one of the hypotheses of this analysis was availability of data from all three modalities. Considering the practical implementation, in order to conduct a feature-level fusion of two or three modalities, it is important to have data from all the modalities available in real-time.

7.3.3 Integration Strategy

The integration strategy adopted in this study was feature-level fusion. One of the objectives of the current analysis is to show the advantages of using feature-fusion techniques on behavioural biometric modalities and the subsequent verification accuracy improvement. Compared to the information available when using a match score outcome, the feature set from a biometric modality contains richer information from the raw data. Hence, fusion at this level can enhance the verification accuracy. However, feature-level integration has inherent issues such as incompatibility of scaled feature sets. This is due to an increased dimensionality of a single feature vector due to concatenation of features from different modalities. In this experiment, we use a feature selection method to deal with the *'curse of dimensionality'* problem. All three modalities generate fixed length temporal feature sets and these feature sets were normalised, hence achieving feature compatibility.

7.3.4 Multi-Modal Framework

Based on the above-mentioned mode of operation and feature fusion strategy, a multi-modal framework was developed. This multi-modal system worked on two different phases – feature fusion phase and verification phase. The feature fusion phase is depicted in Figure 7.2 and the verification phase is depicted in Figure 7.3, which is further divided into enrolment and verification steps.

7.3.4.1 Feature Fusion Phase

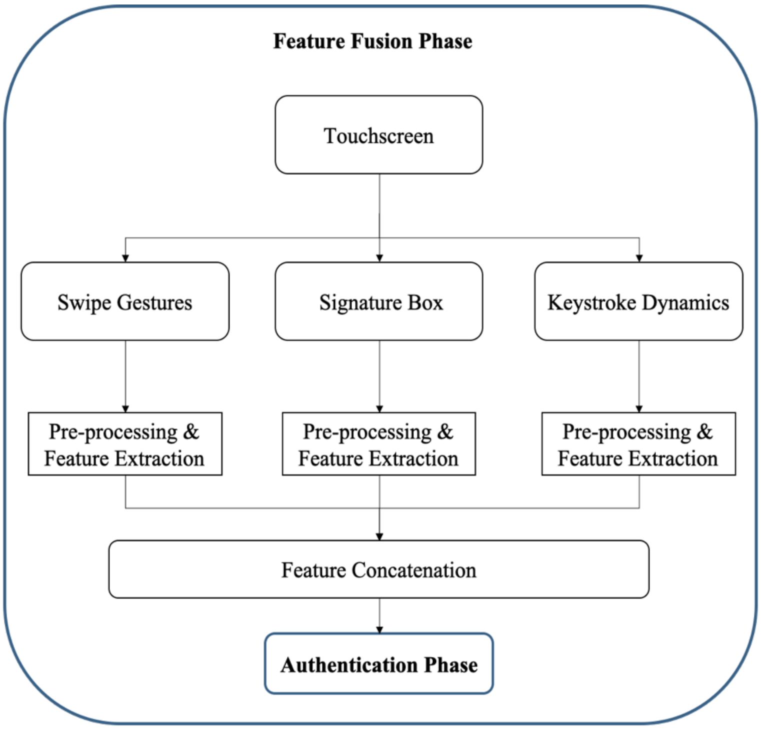


Figure 7.2. Feature fusion phase

This phase consisted of pre-processing, feature extraction and feature concatenation of the touch input.

- *Step 1. Pre-processing* – Same pre-processing methods as described in Chapter 4 (swipe gesture), Chapter 5 (signature) and Chapter 6 (keystroke dynamics) were utilised for these modalities. For swipe gestures, the pre-processing steps consisted of first separating the horizontal and vertical swipes. The next step was to identify outliers in terms of low number of data points and invalid swipe inputs. Swipes containing less than three data points were discarded. The swipes which had ACTION_UP value missing from the TOUCH ACTION parameter were also considered invalid. The pre-processing of signatures involved removal of incomplete signatures, normalisation to avoid inconsistency due to screen location and generation of fixed length signatures. For keystroke dynamics, the pre-processing step involved removal of incomplete data entry samples, such as incomplete phone number entry.

- *Step 2. Feature Extraction* – The feature extraction step for all the three modalities were performed individually. The number of features and the type of features extracted for each modality were different. The list of features along with the description of each feature is provided below.
  - Swipe gestures - For every swipe stroke, a set of 28 global features were computed (listed and detailed in Table 4.1).
  - Signature - The pre-processed inputs for a signature were used to extract global features listed in Table 5.3.
  - Keystroke Dynamics - The input sample of a keystroke dynamics data consisted of the entire sentence/phone number entry typed by the user. Based on this input sample, the global features extracted are described in Table 7.1.

Feature	Description
Total time	Total time spent on typing the input sample
Number of errors	Total number of errors committed while typing the input sample
Average flight time	Average flight time of different digraphs extracted from the input sample

Table 7.1. Keystroke Dynamics feature set

- *Step 3. Feature Concatenation* - In this step, all the feature sets from individual modalities were concatenated together to form a single feature vector. 28 features from swipe gestures, 22 features from signatures and 3 features from keystroke dynamics were concatenated. A total of 53 features were used. This feature vector was fed into the verification phase.

7.3.4.2 Verification Phase

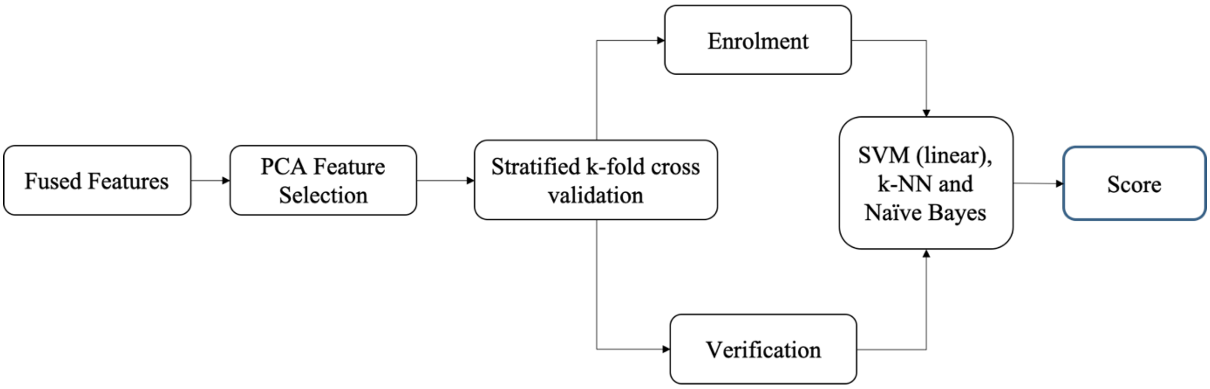

Enrolment
Fused Features
PCA Feature Selection
Stratified k-fold cross validation
SVM (linear), k-NN and Naïve Bayes
Score
Verification

Figure 7.3. Verification phase

After the completion of the feature fusion phase, the verification phase received the fused feature vector as an input. This fused feature vector underwent feature normalisation and selection process. In order to have a dimensionally-reduced feature set and to optimise the computational time and prediction performance, Principal Component Analysis [143] feature selection technique has been applied. The dimensionally-reduced features were then split into enrolment and verification sets from the input data. The verification was performed based on three different classifiers – SVM, k-NN and Naïve Bayes and a final match score was generated for the incoming input sample.

Step 1. Feature Normalisation - Individual feature values of two feature vectors X and Y may have different range and distribution. Feature normalisation is performed to adjust the mean and variance of each individual feature value to normalise and compare the contribution of each feature to the final match score. A min-max normalisation technique was adopted in this phase. The formula used to find the normalised feature *x*' of every individual feature (F_x) is provided below:

$$x' = \frac{x - \min(F_x)}{max(F_x) - min(F_x)} \tag{7.1}$$

Each feature in the feature set was normalised based on this formula. After this, feature selection was performed on the normalised feature set.

Step 2 - Feature Selection - Concatenating two feature vectors - x' and y', results in a new vector – z' = {x'$_1$, x'$_2$, ... x'$_n$, y'$_1$, y'$_2$, ... y'$_m$}, where n and m are the total number of features in x' and y' respectively. The idea behind the feature selection process is to choose a minimal feature set of size k, where k < (n+m) contains maximum characteristics, hence improving the classification performance. PCA has been used to perform the feature selection. The process in PCA involves the calculation of a matrix that defines the relation of different variables with each other. This matrix is then divided into two components – direction and magnitude. Finally, transformation of the original feature set data to align with the direction is obtained by deriving the principal components 1 and 2, which contains the maximum explained variance ratio. Figure 7.4 shows that first two principal components (PC1 and PC2) accounted for maximum variance ratio of 30% and 27% respectively. Therefore, these two principal components were considered, and the discriminative features were selected based on this.

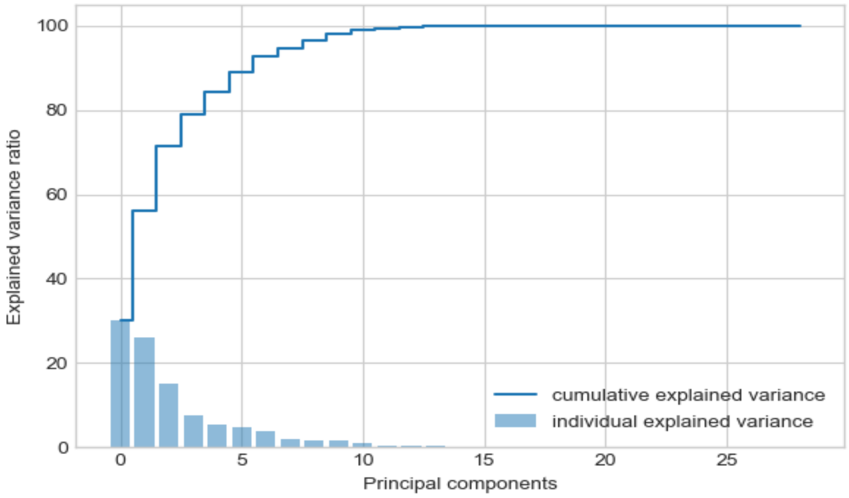


Figure 7.4 Principal components explained variance

Step 3. Stratified K-fold Cross validation - Once the normalised and dimensionally-reduced feature set were acquired, input samples were split into enrolment and verification sets. In order to do this, stratified k-fold cross validation method was adopted.

Cross-validation is a resampling procedure used for evaluating the models on limited data samples such as the one used in our experiment. Parameter '*k*' refers to the number of groups that a given data sample is to be split into. These are the steps carried out using the stratified k-fold cross validation:

- Feature set samples were shuffled randomly.
- Feature sets were split into 5 folds.
- For each unique group:
    - One group was set aside as test data set
    - The remaining groups were taken as a training data set
    - The model was fit to the training set and evaluation was performed on the test set
    - An evaluation score was obtained
- Finally, the evaluation scores were accumulated and average EER was obtained.

The number of folds was chosen as five owing to the limited number of samples. The shuffle parameter was set to true, for shuffling the indices of each class samples before the class split.

Step 4. Enrolment - The verification configuration for individual user present in the dataset was built. The owner of the mobile device was considered as the genuine user and the imposter user was chosen using the random forgery method. The verification model was enrolled with the concatenated features of the genuine user.

Modality	Session 1	Session 2
Swipe gestures	175	155
Signature	18 Finger-based	18 Finger-based
Keystroke dynamics	13 alphabetical, 15 numerical	13 alphabetical, 15 numerical

Table 7.2. Total number of samples from each modality per individual user

The samples were divided based on the usage scenarios (Scenario 1, 2, 3 and 4) and sessions (Session 1 and Session 2). The swipe gesture data were divided as horizontal and vertical swipes. A total of 175 horizontal swipe and 155 vertical swipe samples were used from Session 1 and Session 2 respectively. Each user had 18 finger-based signature samples and the keystroke dynamics had 13 alphabetical samples and 15 numerical samples were used.

As the number of samples acquired from each of the modality varied, a different combination of the enrolment samples set was formed with the existing number of samples from each modality. For swipe gestures, each user had around 60 samples from each usage scenario of Session 1. These 60 samples were split into 50% in training and 50% in testing class. The finger signatures were limited to 10 samples per scenario; therefore, the same samples were repeated on different swipe data / keystroke data to form the input sample. Similarly, with the keystroke dynamics data, same data were used with different swipe gestures to increase the number of input samples used.

The enrolment strategy differed based on the research question under investigation. For example, horizontal or vertical swipe with a combination of keystroke dynamic data and signature data from the baseline scenario were used in enrolment. Additionally, the enrolment was performed for each combination of fused modality as listed below:

- Enrolment combination 1 – Swipe gesture features + Signature features
- Enrolment combination 2 – Keystroke Dynamics features + Swipe gesture features
- Enrolment combination 3 – Keystroke Dynamics features + Signature features
- Enrolment combination 4 – Swipe Gesture features + Keystroke Dynamics features + Signature features

All the enrolment samples belonged to Sitting Indoors scenario (Scenario 1) of Session 1 when Session 2 evaluation was performed and Sitting Indoors scenario (Scenario 1) of Session 2 when Session 2 evaluation was performed.

Step 5. Verification- Every incoming sample was verified against the existing template generated in the enrolment phase. The classification algorithms used for this analysis were SVM, k-NN and Naïve Bayes (described in Chapter 4). The verification sample was compared against the enrolment samples and a probability score was generated based on the classifier's output. This score was used to generate the false acceptance rate and false rejection rate. Further, the equal error rate was generated using these parameters to analyse the performance of the system. Different thresholds on the match scores were used to classify between genuine and imposter samples and the analysed results based on these thresholds are provided in the results section.

7.3.4.3 Score Fusion

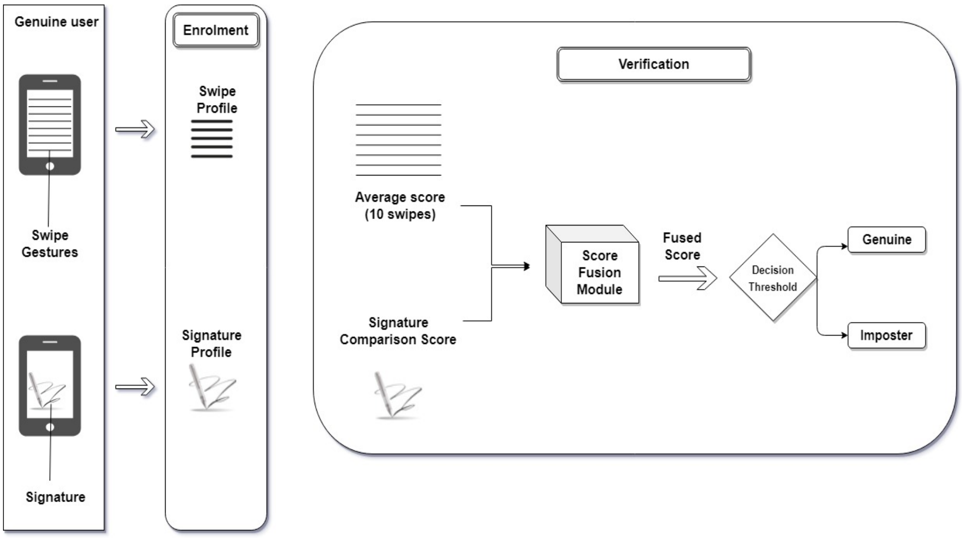


Figure 7.5. Score fusion method adopted using the commercial signature verification system

Score fusion method was applied for combining signature and swipe gesture modalities using the commercial signature verification system. The context of the application that was considered for using such a method was a contract signing scenario, where the user first browses through the document browsing (hence producing swipe gestures during this process) and finally, signs contract.

The method adopted to perform the score fusion has been presented in Figure 7.5. The swipe gestures and signatures of the individual user from the dataset were enrolled separately under two different profiles of the same user. Only finger-based signatures were used for this analysis. It was assumed that the user performed a number of swipe gestures while reading the document on the device before signing at the end of the document. Therefore, instead of combining individual swipe gestures to the signature, the match scores generated from ten swipe gestures were combined to generate an average score of swipe gestures. This score was combined with the match score from the signature.

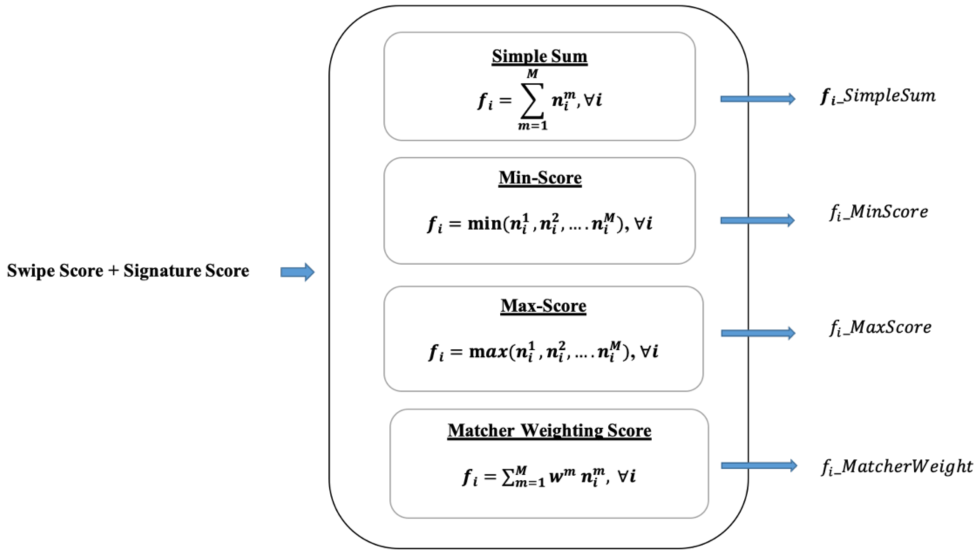

Swipe Score + Signature Score
Simple Sum
$f_i = \sum_{m=1}^{M} n_i^m, \forall i$
Min-Score
$f_i = \min(n_i^1, n_i^2, \ldots . n_i^M), \forall i$
Max-Score
$f_i = max(n_i^1, n_i^2, \ldots . n_i^M), \forall i$
Matcher Weighting Score
$f_i = \sum_{m=1}^{M} w^m \, n_i^m, \forall i$
f_i_SimpleSum
f_i_MinScore
f_i_MaxScore
f_i_MatcherWeight

Figure 7.6. Score fusion module (*f*-fused score, n-score, m-matcher, M - total number of matchers)

The score fusion module combined the individual scores from signature and swipe gestures using four different methods – simple sum, min-score, max-score and matcher weighted score. The simple sum method simply added the scores from both the modalities. The min-score and max-score methods chose the minimum and maximum scores respectively, obtained from both the modalities. Finally, the matcher weighting method assigned weight to each modality. The weights given for each modality were calculated using the formula

$$w^m = \frac{\dfrac{1}{\sum_{m=1}^{M} \frac{1}{r^m}}}{r^m} \tag{7.2}$$

Where w^m is the weight of a matcher and r^m is the equal error rate of a matcher. Based on this, the signature matcher was given 0.8 weight and swipe gestures matcher were given 0.2 weight. The results obtained from this method have been presented in Section 7.4.1.2.

7.4 Results

In this section, the results obtained for different evaluations are presented. The metric used for this study is EER. Different combinations of modalities: a) swipe gestures and signature, b) swipe gestures and keystroke, c) signature and keystroke dynamics, and d) all three modalities are analysed. The results acquired from each of the individual combinations are explained in detail in the following sub-sections.

The results are presented as a mean EER percentage, acquired from 50 verification configurations having varied thresholds. The verification configurations belonged to the individual users present in the dataset. All the modalities were combined using the feature fusion method except for the swipe gesture and signature combination, which has been analysed using both -feature and score fusion methods.

7.4.1 Swipe gestures and Signature

Two techniques were used for combining the swipe gestures and the signature – a) feature-fusion and b) score-fusion method. The feature-fusion method utilised the global features obtained for the signatures using the PenTools (listed in Table 5.3) and the swipe gesture features described in Section 4.4.3. The score-fusion method was applied on the commercial signature verification system. Here, the match scores generated from the swipe gesture verification and signature verification of the same user were combined. Verification using both the modalities were performed on the commercial signature verification system (described in Section 5.3.4). The match scores were fused using different score fusion techniques and the results obtained using all these techniques are presented in this section.

7.4.1.1 Feature-Fusion Method

In order to combine the swipe gesture and signature modalities, the features extracted from each modality were concatenated separately for the genuine and the imposter user. First, the swipe gesture data for the genuine user was categorised into horizontal and vertical swipes. Once they were separated, feature extraction was performed, and swipe gesture feature set was obtained. The finger-based signatures belonging to the same user was used for generating signature feature set. The swipe gesture and signature features of the same user were then concatenated together to form the input to the multi-modal framework. Here a combination of fused data was formed as following:

- Horizontal swipes and Finger signatures (Scenario 1, 2, 3 of Session 1 and Scenario 1,3 & 4 of Session 2)
- Vertical swipes and Finger signatures (Scenario 1, 2, 3 of Session 1 and Scenario 1,3 & 4 of Session 2)

Each of these combinations were used to evaluate the performance of the multi-modal system. First, horizontal swipes and finger signatures obtained in the baseline scenario (Sitting Indoors - Scenario 1) were combined. Both the enrolment and the verification samples belonged to this scenario.

Next, the enrolment samples belonging to Sitting Indoors (Scenario 1) were compared with samples from Treadmill (Scenario 2) and Walking Outdoors (Scenario 3) of Session 1 for both horizontal and vertical swipes respectively. Similarly, the Session 2 enrolled samples from Sitting Indoors (Scenario 1) were compared with Walking Outdoors (Scenario 3) and Travelling on a Moving Bus (Scenario 4). The mean EER% obtained using different classification algorithms are presented in Figure 7.7 and Figure 7.8.

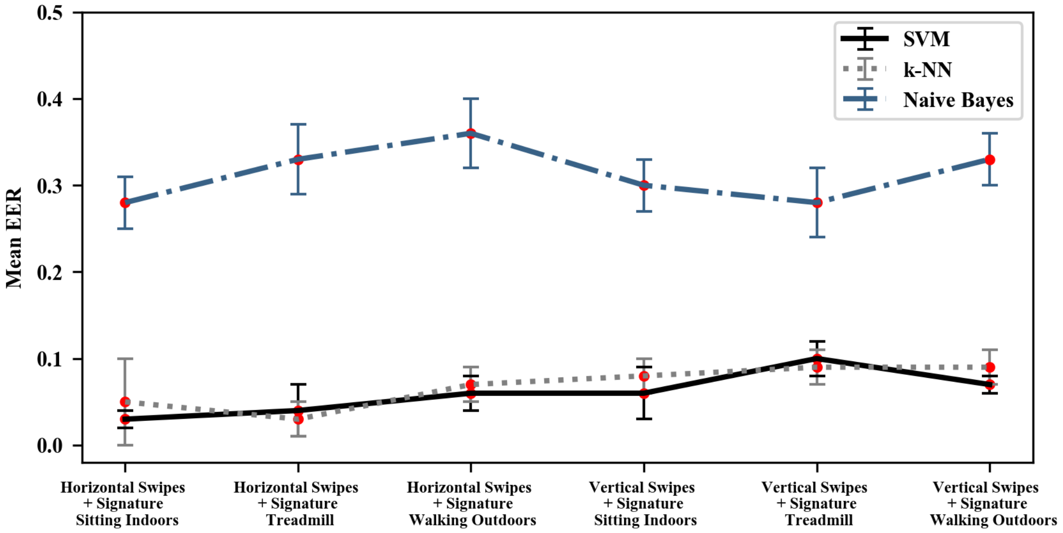


Figure 7.7. Mean EER attained from intra-session comparison of Session 1 for swipe gestures and finger-based signature fusion

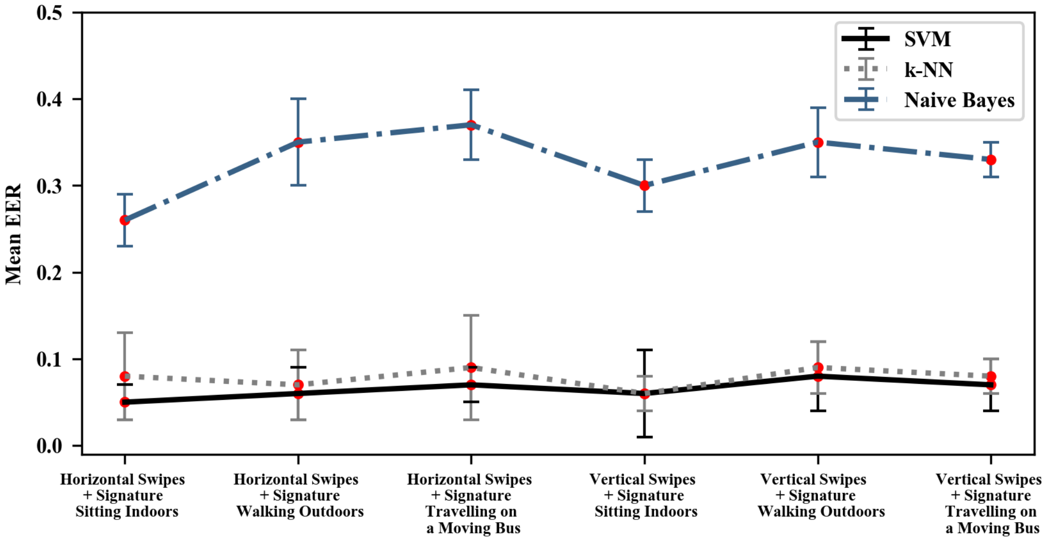


Figure 7.8. Mean EER attained for intra-session comparison of Session 2 for swipe gestures and finger-based signature fusion

As shown in the above figures, the acquired mean EER % for Sitting Scenario of horizontal swipes with finger-based signature showed that SVM attained best performance with the lowest EER of 3% for Session 1 and 5% for Session 2. It can be observed that the performance using the Naïve Bayes algorithm is the worst compared to SVM and k-NN for all the evaluations in Session 1 and Session 2 as the mean EER% obtained were significantly higher for each evaluation. The SVM algorithm performs the best with lowest mean EER% in case of both horizontal and vertical swipe gesture combinations with signature.

On comparing between horizontal and vertical swipe combinations, the combination of vertical swipes with the signature obtained relatively higher mean EER % using SVM and k-NN for Session 1. However, for Session 2, vertical and horizontal swipe gesture combinations with signature in walking Outdoors and Travelling on Moving Bus attained performances in similar ranges. This may suggest that for both the swipe category (horizontal and vertical swipes), the combination with signature show similar performances. The results also show that the intra-session EERs, even for usage scenarios involving movement, are in acceptable ranges for SVM or k-NN algorithms. A comparison of results acquired using the uni-modal approach of swipe gestures and signatures has been presented in Table 7.3.

7.4.1.2 Score Fusion

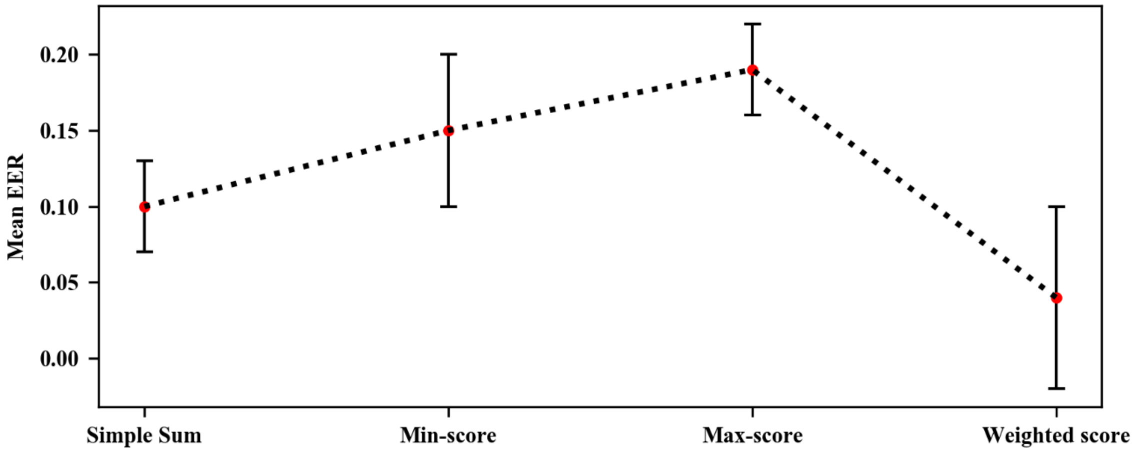


Figure 7.9. Mean EER from score fusion methods for swipe gestures and signature combination using the commercial signature verification system from scenario 1 of Session 1

Using the score fusion method, the mean EERs obtained from different score fusion techniques are shown in Figure 7.9. The enrolment and verification samples were horizontal swipes and finger-based signatures belonging to Scenario 1 of Session 1. Based on the obtained results, the weighted score method obtained the lowest mean EER of 5% and the max-score method obtained the highest mean EER of 18%. Therefore, it was concluded that the weighted score method performed the best compared to all the other score fusion methods.

7.4.2 Keystroke Dynamics and Swipe Gestures

Similar to the fusion method chosen for combining the swipe gestures and signature, the keystroke and swipe combinations were also divided into,

- horizontal swipes and keystroke dynamics features
- the vertical swipes and the keystroke dynamics features.

Keystroke dynamics consisted of both alphabetical and numerical inputs. The results obtained for performing intra-session comparison of samples belonging to Session 1 has been presented in Figure 7.10.

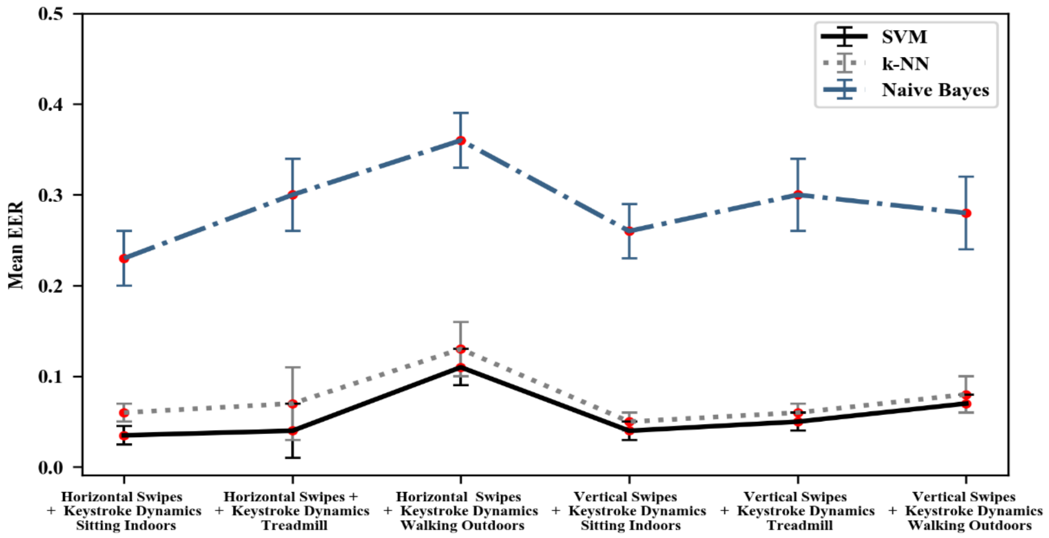


Figure 7.10. Mean EER attained for combining keystroke dynamics and swipe gestures belonging to Session 1

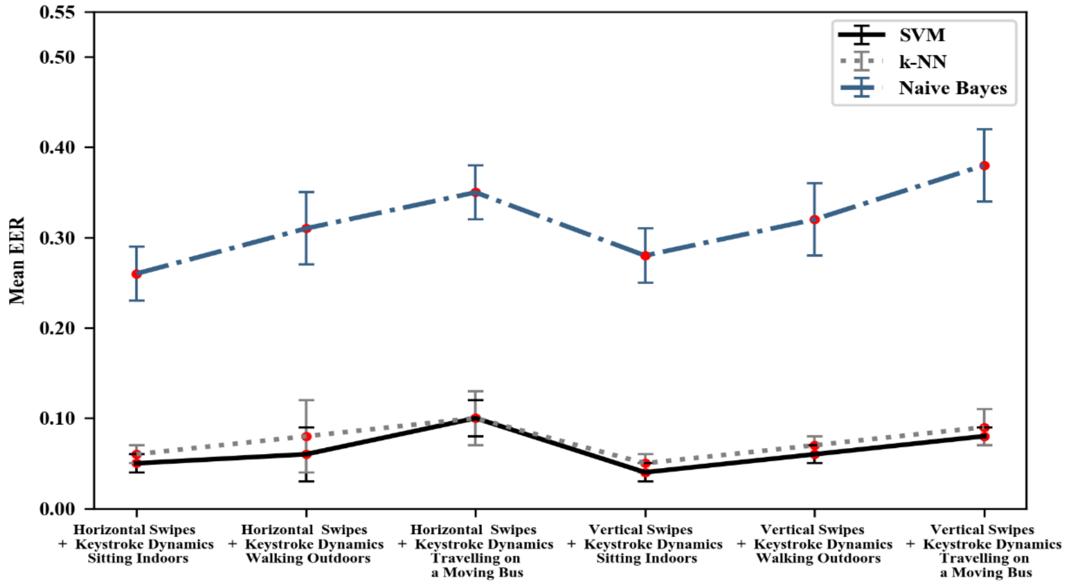


Figure 7.11. Mean EER obtained for combining keystroke dynamics and swipe gestures belonging to Session 2

It can be observed from Figure 7.10 and Figure 7.11 that SVM algorithm performed the best compared to k-NN and Naïve Bayes for both sessions as it obtained the lowest mean EER across all scenarios and combinations. Across all the comparisons, the performance obtained using the Naïve Bayes algorithm is the worst. Even with the best performing algorithm (SVM), the results from Session 1 and Session 2 show that horizontal swipes and keystroke dynamics combination belonging to the Walking Outdoors (Scenario 3) scenario performed the worse with a mean EER of 11% (Session 1) and 10% (Session 2). The combination of horizontal or vertical swipes with keystroke dynamics belonging to Sitting Indoors (Scenario 1) scenario yielded the lowest mean EERs of 5.2% (Session 1) and 5% (Session 2).

7.4.3 Signature and Keystroke Dynamics

The enrolment samples for this evaluation belonged to the Sitting Indoors (Scenario 1) of Session 1 and Sitting Indoors (Scenario 1) of Session 2 separately. The intra-session results obtained for Session 1 and Session 2 have been presented in Figure 7.12 and Figure 7.13. It can be observed that for Sitting Indoors (Scenario 1) from Session 1, the SVM algorithm attained the lowest mean EER of 8%, followed by k-NN with 12%. k-NN performed worse in Walking Outdoors (Scenario 3) of Session 1 and Travelling on a moving bus (Scenario 4) of Session 2. On comparing with the results of other combinations of modalities, the lowest EER attained by signature and keystroke combination was 8%, even on comparing the samples from the same scenario, Sitting Indoors (Scenario 1). Therefore, the performance of signature and keystroke dynamics combination the worse out of all the three combinations.

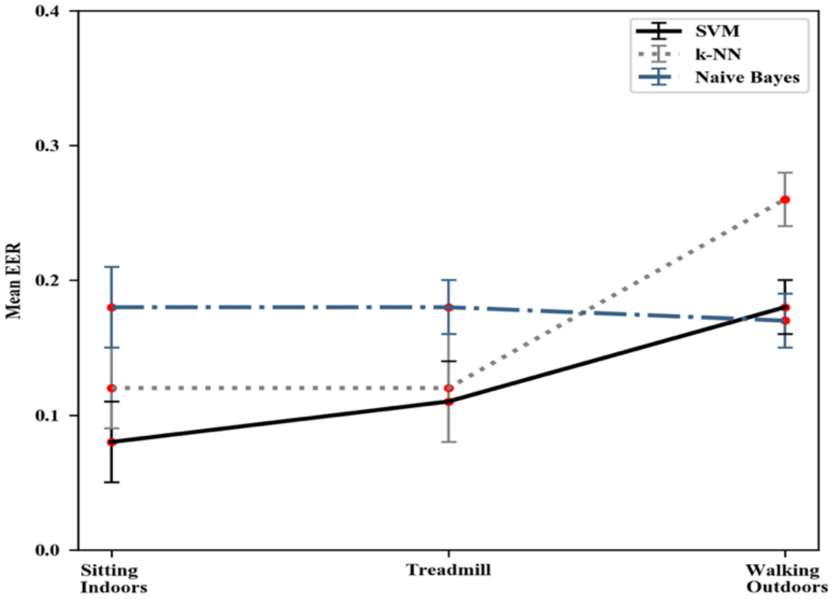


Figure 7.12. Mean EER attained for combining signature and keystroke dynamics data of Session 1

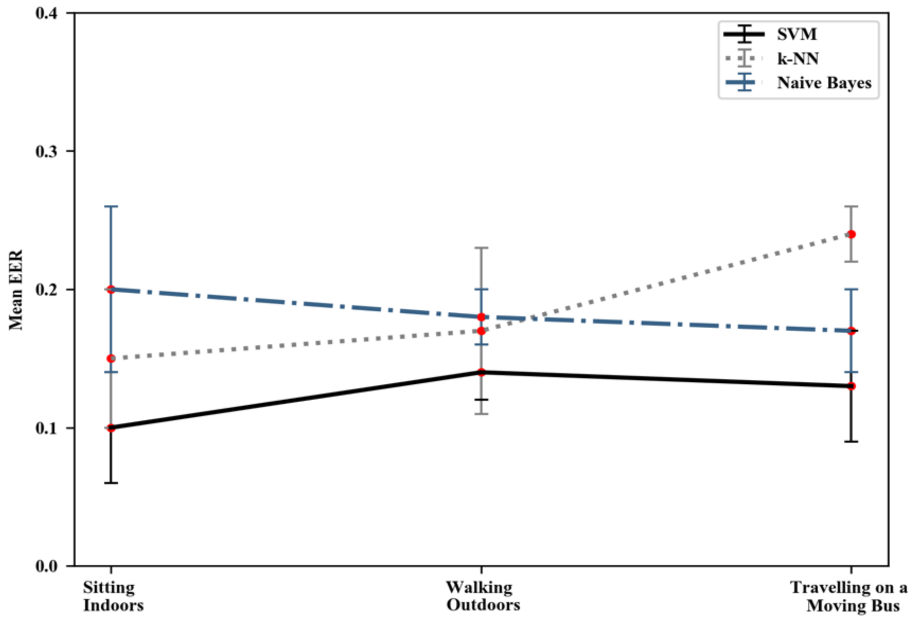


Figure 7.13. Mean EER attained for combining signature and keystroke dynamics data of Session 2

7.4.4 Swipe Gesture, Signature and Keystroke Dynamics

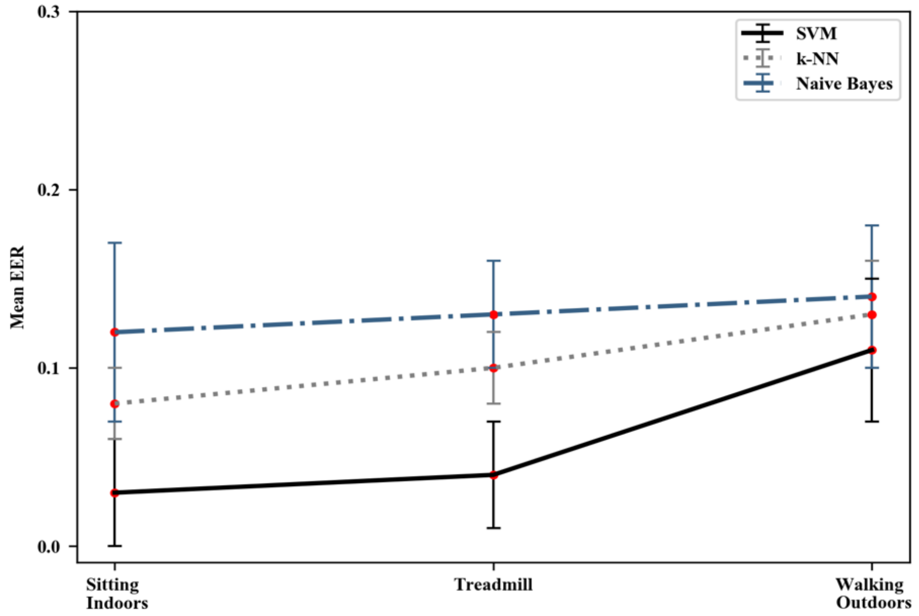


Figure 7.14. Mean EER attained on combining swipe gestures, signature and keystroke dynamics for Session 1.

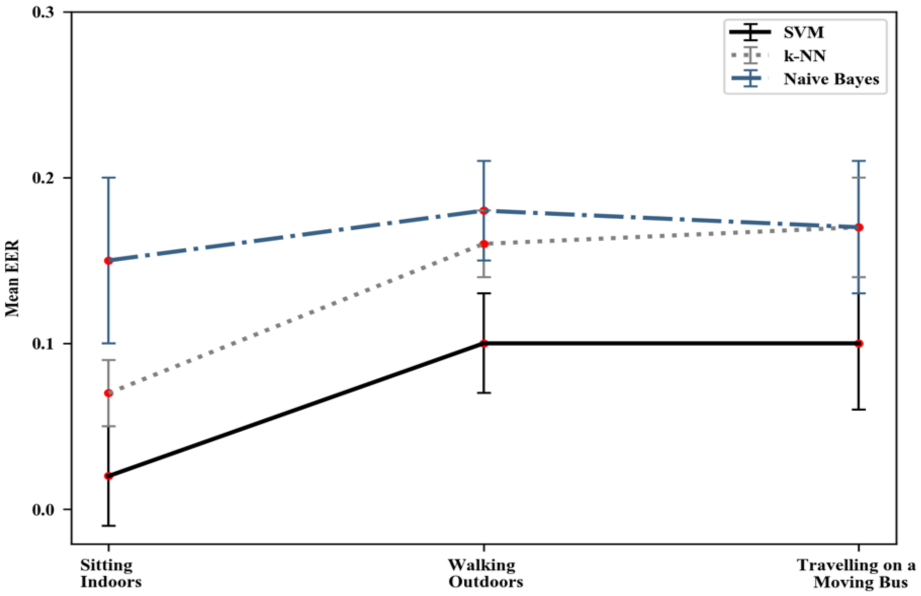


Figure 7.15. Mean EER attained on combining swipe gestures, signature and keystroke dynamics for Session 2

For this evaluation, the enrolment samples belonged to Sitting Indoors (Scenario 1) of Session 1 for Session 1 evaluations and Sitting Indoors (Scenario 1) of Session 2 for Session 2 evaluations. The verification samples belonged to Sitting Indoors, Treadmill and Walking Outdoors for Session 1 and Sitting Indoors, Walking Outdoors and Travelling on a Moving Bus in Session 2. Figure 7.14 (Session 1) and Figure 7.15 (Session 2) show the mean EERs obtained using the SVM, k-NN and Naïve Bayes algorithms when all the three modalities were combined. It can be observed that the SVM algorithm performed the best with the lowest mean EER of 2% (Session 2) and 3% (Session 1) in Sitting Indoors scenario of Session 2. Compared to the results obtained in combination of two modalities, Naïve Bayes algorithm performed better when all three modalities were combined. However, the mean EER % for k-NN are better than Naïve Bayes algorithm in all the scenarios for Session 1 and Session 2. Walking Outdoors (Scenario 3) of Session 1 as well as Session 2 obtained the highest mean EERs, followed by Travelling in a Moving Bus of Session 2.

7.4.5 Comparison of Uni-modal and Multi-modal Verification

Table 7.3 shows the best mean EER percentage values attained using different classifiers for both - uni-modal and multi-modal verification systems used for the evaluations in this thesis. The uni-modal based verification attained best mean EERs when the enrolment and the verification samples belonged to the same usage scenario of the data collection. However, there was performance deterioration when the verification samples belonged to a different usage scenario than the enrolled ones. For uni-modal systems, out of all the usage scenarios, the baseline scenario (Sitting Indoors) with no variation in environment or user's body motion acquired best mean EER.

Scenario	**Uni-modal (Mean EER (%))**			**Multi-modal (Mean EER (%))**			
	Swipe gestures	**Signature**	**Keystroke Dynamics**	**Swipe + Finger-based signatures (SVM)**	**Swipe + Keystroke Dynamics**	**Keystroke + Finger-based signatures**	**Swipe + Keystroke Dynamics + Signatures**
Scenario 1	Horizontal - 1%	13%	2%	3%	3.5%	8%	3%
	Vertical –2 %	-	-	6%	4%	-	-
Scenario 2	Horizontal - 23	15%	52%	4%	4%	11%	4%
	Vertical –28%	-	-	10%	5%	-	-
Scenario 3	Horizontal –27%	28%	53%	6%	11%	18%	11%
	Vertical -31%	-	-	7%	7%	-	-
Scenario 4	Horizontal –30%	17%	45%	8%	11%	13%	10%
	Vertical -26%	-	-	10%	8%	-	-

Table 7.3. Mean EER percentages of uni-modal and multi-modal verification

On the contrary, a multi-modal verification showed better performance even when the enrolment and the verification samples do not belong to the same usage scenario. When all the three modalities were combined, the best mean EER attained was for the scenarios that were performed indoors – 3% and 4%. The scenarios having body movement either whilst the user was walking or due to the transport, the performance deteriorated attaining a mean EER of 11% (Scenario 3) and 10% (Scenario 4). However, the error rates are in acceptable range compared to the uni-modal results.

On comparison, the multi-modal solution certainly obtained better performance compared to the uni-modal methods. However, the need for combining such modalities together depends on the requirement of the biometric application.

7.5 Conclusion

In this chapter, the effectiveness of a multi-modal framework using solely the behavioural biometric modalities has been demonstrated. Multiple evaluations with different combinations of modalities have been carried out and the results have been presented. The multi-modal dataset of 50 participants, acquired using a Samsung Galaxy Note 5 smartphone over two sessions separated by one week with multiple usage scenarios, has been used in the experiment. Feature fusion and score-fusion methods had been applied on combining the swipe gesture, signature and keystroke dynamics data. The classification algorithms such as SVM, k-NN, Naïve Bayes and commercial signature verification system were used for the evaluation.

In terms of compatibility of combination of behavioural biometric modalities, swipe gesture and signature feature fusion method attained an average EER of 3% with SVM and swipe gesture and keystroke dynamics obtained 3.5% and signature and keystroke dynamics feature fusion obtained 8% average EER using SVM classifier. This shows that fusing signature and swipe features gives best results in terms of the average EER. However, fusing signature and keystroke modalities showed comparatively poor performance. The best combination using the feature-fusion method is swipe gestures with signature. The score-fusion method using the commercial signature verification system showed that the weighted score method performed the best for combining swipe gestures and signature modalities.

One of the reasons for fusing different modalities was to enhance the mobile device security. The experimental results obtained reveal a boost in the performance with 3% mean EER using SVM classifier when all three modalities were fused. This is because different modalities possess distinct characteristics corresponding to that data source. When these features are integrated together, a valuable and distinctive feature set is constructed that aids in performance enhancement. The results of the multi-modal verification system yielded better results compared to the uni-modal verification.

One of the challenges faced while conducting this study was the limitation of the number of samples from individual modality. There were unequal number of samples from every modality per scenario. For example, Session 1 contained 175 swipe gesture samples, 18 signature and 28 keystroke dynamics samples. Therefore, 18 unique samples of swipe, signature and keystroke dynamic data could be formed. However, in order to generate more samples to be used for enrolment and verification, samples from signature and keystrokes were repeatedly combined with the remaining swipe samples. Hence, although the swipe gesture samples were new, the signature and keystroke data were repeated to conduct the analysis.

A challenge regarding the practical implementation of such a multi-modal system is that the system would work only when data from all the three modalities are available. Hence, introducing a wait in the verification score calculation process until data from all three modalities were acquired. The system is designed to combine the data acquired from one sensor – the touchscreen, hence, it is not possible to attain all the three-modality data at the same time.

In conclusion, this chapter details a multi-modal framework that is built on behavioural biometric modalities such as swipe gesture, signature and keystroke dynamics. Using this framework, an evaluation of different combinations of modalities has been explored. Additionally, intra-session evaluation results are also presented that reflect the stability of this framework across different usage scenarios of a mobile device.

Chapter 8. Usability Evaluation

8.1 Introduction

The implementation of conventional biometric techniques to mobile environments brought in a number of open challenges. Usability is likely one of the biggest amongst these challenges. Along with recognising users with high accuracy, mobile biometric solutions are expected to deliver better user experience. A satisfied user can continue using the biometric solution in an expected manner; however, a dissatisfied user can end up making unexpected interactions with the device. Such interactions would lead to biometric verification causing further user annoyance. As a result, an unsatisfied user is likely to reject the biometric solution altogether. Therefore, it is important to assess the usability aspect of the biometric application in a mobile solution.

Usability has been defined by ISO 13407:1999 [144] as "*the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use*". Usability can be evaluated in a qualitative and a quantitative manner. A qualitative evaluation indicates the user attitude and acceptability. In order to measure the quantitative aspect, three key parameters (defined in ISO 13407:1999 [144]) namely effectiveness, efficiency and satisfaction are commonly used. Effectiveness is measured based on the number of successful/unsuccessful user interactions. Efficiency is calculated based on the task completion time and user satisfaction is measured based on the user feedback regarding the user experience.

In this chapter, the usability assessment of two different behavioural modalities - keystroke dynamics and dynamic signatures are presented. The reason to have excluded swipe gestures for the usability analysis is because there is no standard method to measure a successful/correct swipe gesture interaction with the device. However, for the keystroke dynamics modality, this can be measured based on correct/incorrect typing of alphabetical or numerical data during the data collection. Similarly, incorrect signature presentation can be measured based on deleted and incomplete signatures during the data acquisition.

The assessments presented in this chapter are based on the metrics defined by ISO [144]. The novelty of this usability study is that we assess these behavioural modalities in different usage scenarios of the device and in various environmental contexts – indoors and outdoors. The main focus of this evaluation was to obtain conclusions regarding the usability factor on keystroke and signature modalities, which can be used to improve the biometric design and implementation on a mobile device.

The following section presents related work on usability, specifically using the behavioural biometric modalities on mobile devices. The next section, Section 8.3, details the methodology used for calculating the usability metrics for individual modalities, following this, Section Chapter 1 describes the results obtained for keystroke dynamics and Section Chapter 1 presents the results obtained for signature modality. The final section, Section Chapter 1, presents the conclusions drawn based on the obtained results.

8.2 Related Works

This section presents the qualitative and quantitative usability analysis performed on behavioural biometric modalities on mobile devices.

8.2.1 Qualitative Analysis of Usability

User perception can influence willingness of the user to accept and ultimately utilise a particular biometric modality. Unlike point-of-entry verification methods like PIN or password, touch-dynamics based behavioural biometrics are also used in context of continuous verification, where the verification is performed in the background, throughout the entire session of the user interaction with a device. As these are fairly new techniques adopted for verification, a number of surveys have been conducted to understand the acceptability of these methods.

Alhussain et al. [145] conducted a usability survey to investigate users' perception on the physiological and behavioural verification techniques, both requiring active user interaction on a mobile device. Their questionnaire collected information from 331 participants and compared different verification methods: a) physiological - PIN or password, fingerprint and iris b) behavioural - signature and voice. Their results show that, overall, 87.3% of participants agreed that biometric verification would be effective for security of their mobile devices. 66.2% of participants preferred fingerprint for verification, whilst voice recognition and signature verification were preferred by only 6.43% and 5.47% participants respectively. A possible reason for considerably low degree of preference towards behavioural biometric methods could be convenience factor. For instance, the data capture process for a voice recognition would not be favoured in all scenarios. For example, in a movie theatre, it is hard to capture voice data with a background noise. Additionally, the user preference is subjective to the experience during the experiments.

Furnell et al. [146] surveyed 175 users to determine the user acceptance of continuous verification. 43% of participants showed a positive response towards continuous monitoring whilst 40% of participants stated that they consider continuous monitoring as invasion of privacy. 85% participants stated that they must be made aware of the monitoring. Karatzouni et al. [147] recorded views and attitudes of a focus group towards transparent and continuous verification versus traditional point-of-entry methods on mobile devices. The participants offered varying responses. Negative views were mainly due to fear of inconvenience caused through false rejections. Participants also showed more concern for privacy in terms of where the biometric data is stored – on the mobile device or the cloud. The fear of phones being stolen discouraged the idea of storing the biometric data on devices. However, storing it in a cloud network would compromise control over data.

Khan et al. [148] evaluated the usability and security perceptions of implicit verification with a group of 37 users. They report that 91% of the participants found implicit verification to be convenient and 81% indicated that the level of security provided is satisfactory. Detection delays and amount of false accepts were the main reasons for security concerns. Additionally, the false rejects were causing annoyance.

Rasnayaka et al. [149] surveyed roughly 500 respondents. They suggest that adoption of continuous verification was dependent on security awareness factor. They show with statistical significance tests that device operating systems, gender, educational level, occupation and age can have an impact on the awareness of security aspects on a mobile device. Their results propose that more security-aware users are unconvinced to use continuous verification methods while, on the other hand, participants who were less aware about security were positive towards the adoption of continuous verification.

It is observed based on the above surveys that the respondents are still conservative about newer methods of verification. However, it should be noted that the participants may not have practically experienced the techniques that they were asked to comment upon. It may therefore be possible that they change their opinion if they practically experience these verification techniques. Further, significant concerns on acceptance of continuous supervision are due to ethical considerations directed towards invasion of privacy. Certain safeguards should be considered to reassure users that continuous verification is a safe, reliable and secure verification method. In particular, users should be informed about intended use of information collected. Overall, improved awareness of the vulnerabilities of existing verification methods combined with a practical experience of using non-invasive continuous verification techniques would possibly improve perceived acceptability.

8.2.2 Quantitative Analysis of Usability

A number of studies have applied the HBSI model to evaluate usability in terms of effectiveness, efficiency and satisfaction. Brockly et al. [150] validated HBSI model for dynamic signature verification using two different digitizers. Based on complexity of potential interactions with a device, they suggested revision to the HBSI model and proposed five different categories of interaction errors. These errors arose from a bad or good presentation. The incorrect signature presentation was categorised as defective, concealed and false interactions. For correct presentations, the errors were divided as failure to detect and failure to process. They suggested that with the change in ceremony and digitizers, the potential interactions change. Such a conclusion suggests increased complexity in possible interaction models and errors for touch-dynamics based modalities such as swipe. Unlike dynamic signatures, unconstrained swipes used in the context of continuous verification have no specific pattern to be repeated for enrolment and verification and it does not have a defined ceremony-based data acquisition system in place, and this would result in increased complexity in user interaction modelling.

Based on the conclusions obtained from the qualitative studies, continuous verification using behavioural biometric modalities have mixed responses. A comparison of user experience, specifically using the behavioural biometric modalities on various scenarios of a mobile device is missing in the literature. This chapter presents user responses regarding factors that limits data donation process while presenting the behavioural data on a mobile device. The qualitative usability assessment of modalities - keystroke dynamics and signature, has been presented in various usage scenarios of the device. The results based on effectiveness and efficiency reveal the impact of external factors such as environment and user's body movement on usability.

8.3 Methodology

This section presents the methodology adopted to assess the usability for keystroke dynamics and signature modalities. Table 8.1 highlights the metrics used to measure usability for keystroke dynamics and signature modalities.

Usability Parameter	**Metrics Used for Keystroke Dynamics**	**Metrics Used for Signature**
Efficiency	Total time taken to type one complete sentence per scenario	Total time taken to complete a signature per scenario
	$\text{Efficiency Error }\% = \frac{\text{No. of typed sentences with errors}}{\text{Total number of typed sentences per scenario}}\%$	$\text{Efficiency Error }\% = \frac{\text{No. of times 'Clear Signature' button was pressed}}{\text{Total number of signatures per scenario}}\%$
Effectiveness	Total number of error occurrences for each input sample	number of wrong signatures
	$\text{Effectiveness Error }\% = \frac{\text{No. of typed sentences with errors correctly accepted}}{\text{Total number of typed sentences per scenario}}\%$	$\text{Effectiveness Error }\% = \frac{\text{No. of wrong signatures}}{\text{Total number of signatures per scenario}}\%$
Satisfaction	User feedback	User feedback

Table 8.1. Usability metrics used for measuring efficiency, effectiveness and satisfaction

8.3.1 Keystroke Dynamics

The keystroke dynamics data contains both alphabetical and numerical inputs from 50 participants in different usage scenarios. The alphabetical input was acquired by asking the users to type a given sentence on the phone screen and the numerical input was acquired based on phone number entry. The entire sentence entered on the smartphone by the user was considered as one sample. Similarly, for a numeric input, entry of the complete phone number was considered as a sample. A total of 1200 alphabetical and 950 numerical samples were entered by 50 users in both the sessions as provided in Table 3.4. The user generated typing errors, which were also logged during the keystroke entry tasks in the experiment.

Efficiency is defined as the completion of a task on time. The metrics used for measuring this were - a) the total time taken to type the entire sentence was calculated for every scenario in the experiment, and b) the ratio of number of sentences having typing errors per scenario divided by the total amount of sentences typed signified the error percentage. The metrics used for measuring the effectiveness was a) total number of error occurrences for each input sample (a sentence), and b) the ratio of typing errors accepted as correct divided by the total amount of typed inputs signified the effectiveness error percentage. The errors considered were based on these factors: wrong character typed, numerical character entered in alphabetical input or alphabetical character entered in numerical input, no space character entered when required, trailing and leading white spaces.

Finally, user satisfaction was assessed based on the feedback acquired at the end of the first session of the experiment. The feedback form has been provided in the (Appendix A). Users were also asked if they used auto correct feature in their own device and if they usually performed typing while walking. They were also asked to indicate the factors limiting the key entry during the typing tasks.

8.3.2 Signature

With regards to the signature data, multiple finger-based and stylus-based signatures were captured in every scenario of the experiment. Each session contained 15 finger-based signatures and around 10 stylus-based signatures. For the finger-based signatures, the pressure parameter acquired was 0, however, for the stylus-based signatures, a valid pressure value between the ranges of 0 to 1 was recorded.

Efficiency was calculated based on the total time taken to complete a signature. A signature donation process was considered complete when the user pressed the '*Save Signature*' button on the UI screen. The efficiency error rate was calculated as the ratio of number of the times the '*Clear Signature*' button was pressed divided by the total amount of signatures captured.

Effectiveness was calculated based on the number of wrong signatures. A signature was tagged as wrong signature if, a) the user clicked on the *'Redo Signature'* button on the screen, b) empty or incomplete signature was found when *'Save Signature'* button was pressed, c) the user performed all five signature samples on the same window by mistake, d) using a finger for stylus-based signature donation or vice-versa and e) a significant part of the signature pattern was missing from the signature. Effectiveness error rate was calculated as total number of wrong signatures divided by total amount of signatures per scenario.

Finally, the user satisfaction factor was assessed based on the feedback acquired from the users regarding the signature tasks. This captured users' preferred signing tool - finger or stylus, comfortable usage scenario for signature donation and the factors limiting the signature presentation during the experiment. The results obtained based on all of these factors are presented in the next section.

8.4 Results

This section presents the usability evaluation results obtained for the keystroke dynamics and signature.

8.4.1 Keystroke Dynamics

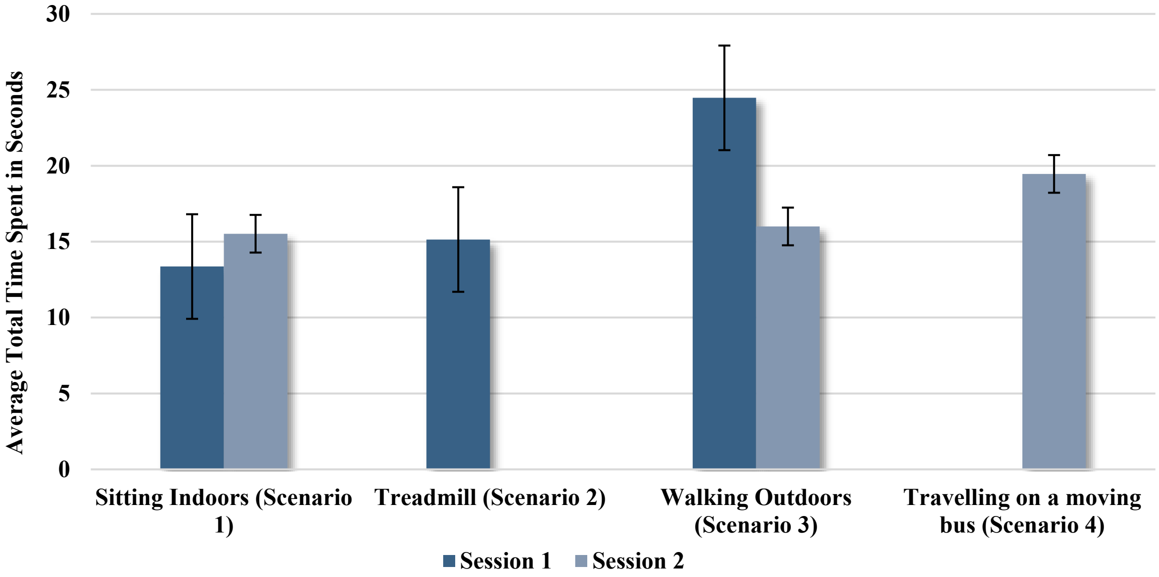


Figure 8.1. Average total time acquired for different scenarios in Session 1 and Session 2

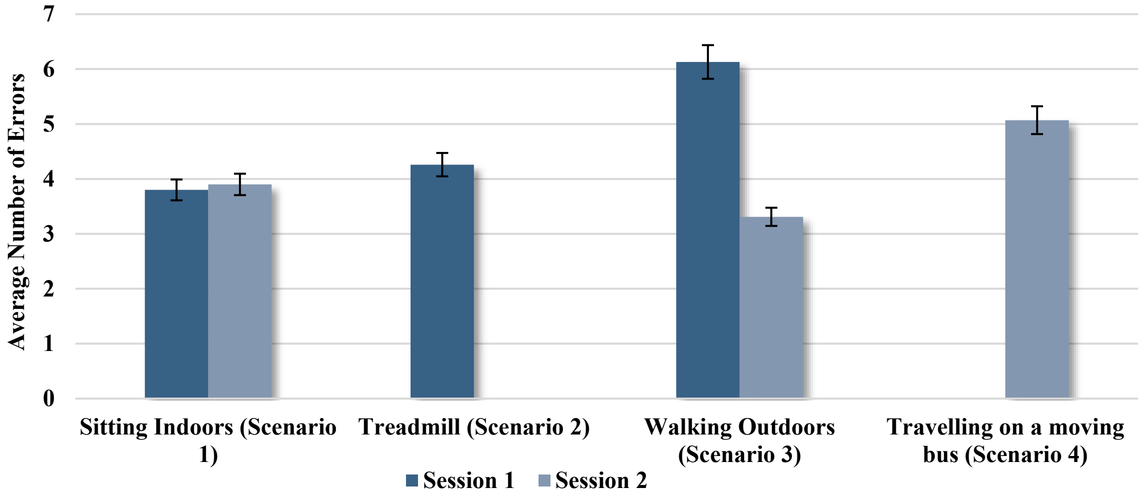


Figure 8.2. Average number of errors in different scenarios of Session 1 and Session 2

The average of total time spent on typing the keystroke inputs respectively for 50 participants were calculated and depicted in Figure 8.1. The average total time spent during the indoors set up in Scenario 1 and Scenario 2 are lower compared to the outdoors setup (Scenario 3 of Session 1. and Scenario 4 of Session 2). Similarly, the number of error occurrences in Scenario 3 and Scenario 4 are higher than Scenario 1 and 2. However, for Scenario 3, in Session 2, the number of error occurrences reduced to almost half compared to Session 1. This could be due to the familiarity with the typing task. Therefore, the users made comparatively lesser typing mistakes in Session 2 Walking Outdoors scenario.

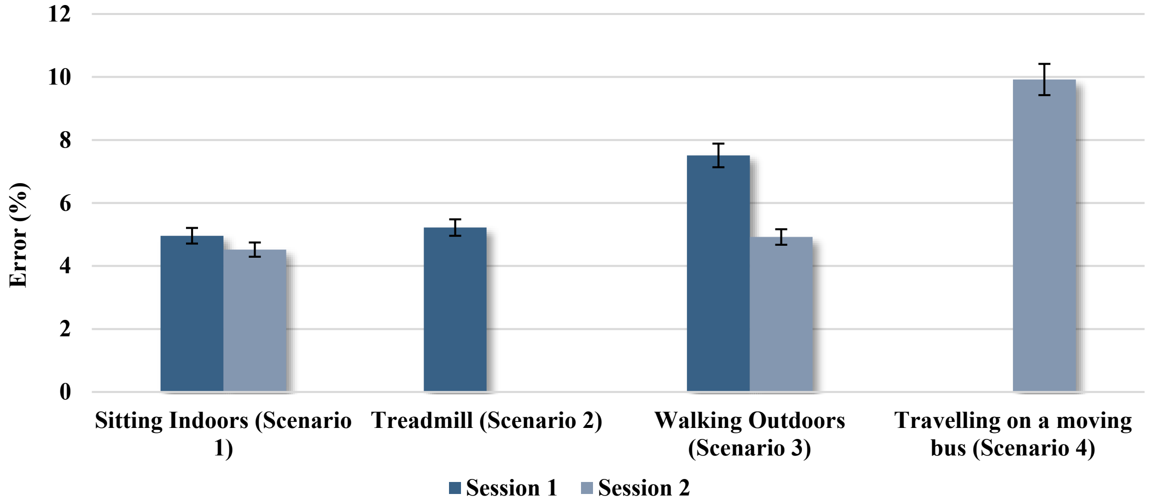


Figure 8.3. Efficiency error rates of keystroke dynamics across different scenarios

As shown in , for Scenario 1 and Scenario 3, results from both the sessions – Session 1 and Session 2 are provided. It can be seen that the error rate for Session 1 is higher than Session 2 for Scenario 1 and Scenario 3. A possible reason leading to a lower error rate in the second session could be due to the learnability factor. The users might have become familiar with the process of the data donation in Session 2. With regards to the comparison of environmental location, the error rates outdoors are considerably higher than indoors. However, comparing the different scenarios conducted indoors, Scenario 1 and Scenario 2 do not show a significant difference, but, for the outdoor scenarios, Scenario 4's error rate is higher compared to Scenario 3. The possible reason can be associated with the unexpected movements, abrupt stops, etc., caused by the transport in the uncontrolled environmental condition.

These results were assessed against the feedback acquired from all the users at the end of the experiment. As shown in , 68% of the users were comfortable typing with the smartphone, while 22% users felt that the soft keyboard size was causing a problem during the typing tasks. 10% of users revealed that the screen-size of the mobile phone used in the experiment was bigger than the mobile device they own and use daily, therefore causing a discomfort for holding the device during the typing task.

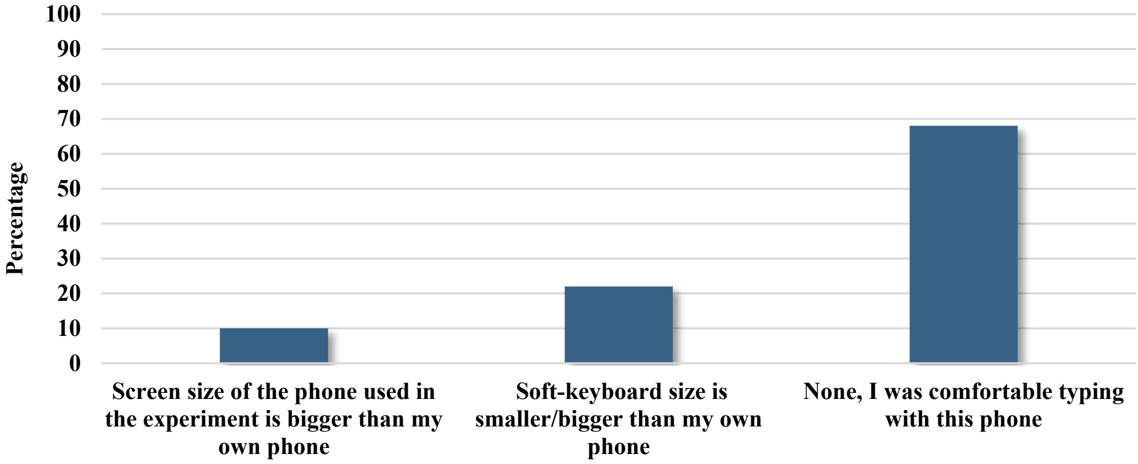


Figure 8.4. User feedback on the factors restricting the keystroke input during the experiment

Comparing these results to user data, we note that 62% participants in the experiment owned Android-based smartphone. Therefore, the QWERTY keypad format was familiar to the users. This is the most likely reason for a similar ratio of participants indicating comfort typing with the mobile device used during the experiment.

8.4.2 Signature

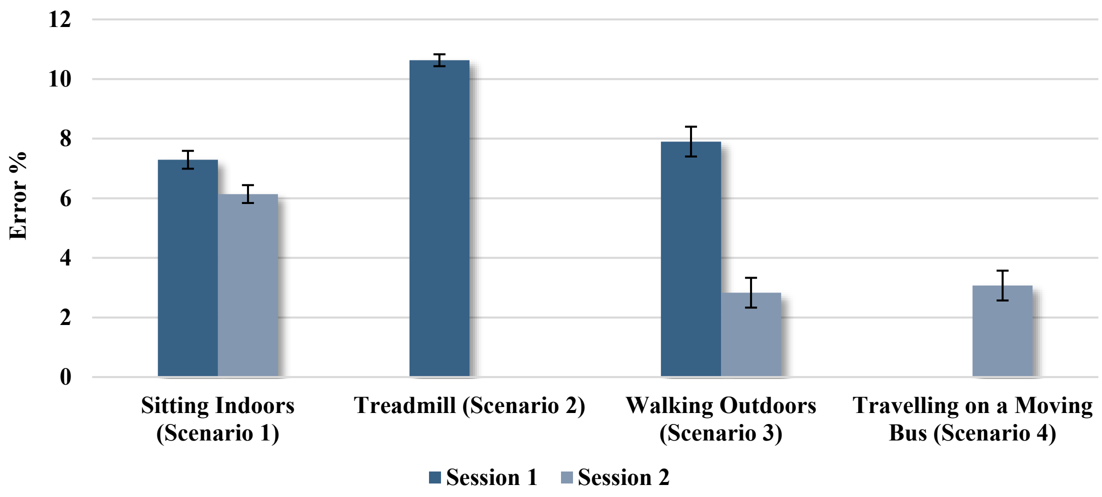


Figure 8.5. Efficiency error rate of dynamic signatures across different scenarios of Session 1 and Session 2

shows the efficiency rate acquired for different scenarios for signature modality. It can be observed from the figure that the error rate acquired for Scenario 1 in Session 2 is lower compared to Session 1. Similarly, Scenario 3 in Session 2 is lower compared to Session 1. Scenario 4 in Session 2 showed 3.07% error rate. It should be noted that the number of signatures collected per participant is lower (only three signatures) for Scenario 4. For Scenario 3, walking outdoors, Session 2 error rate is reduced from 7.9% to 2.83%. A possible reason could be associated with the familiarity in the process of signature donation while walking.

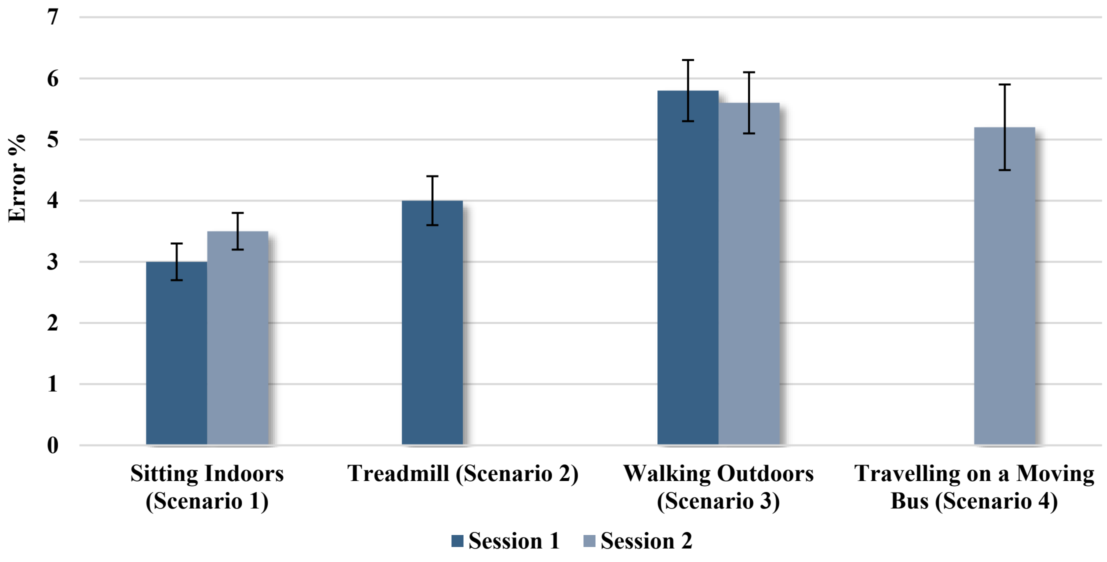


Figure 8.6. Effectiveness error rate for different scenarios in Session 1 and Session 2

shows the effectiveness rate for different scenarios in Session 1 and Session 2. Scenario 1 attained 3% effectiveness rate in Session 1 and around 3.58% in Session 2. Scenario 3, walking outdoors attained the highest effectiveness rate of 5.8% in Session 1. Travelling on a moving bus scenario attained slightly lower effectiveness rate compared to Scenario 3 in Session 1.

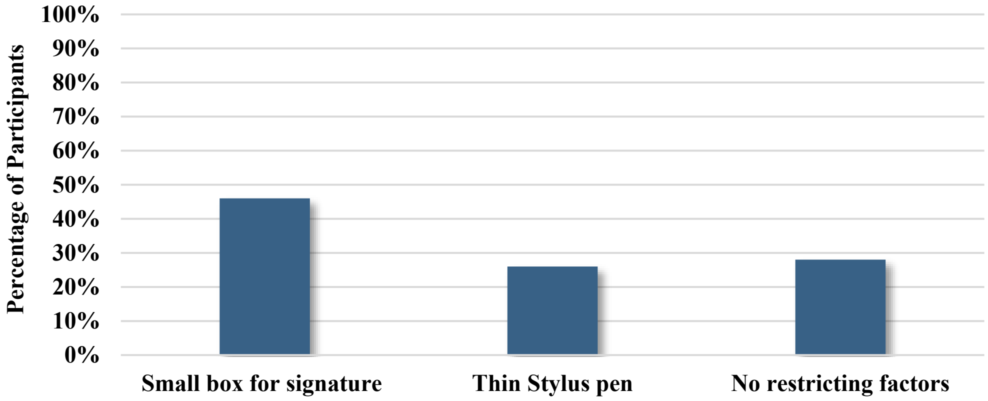


Figure 8.7. Dynamic Signatures – Satisfaction Factor

46% of the participants indicated small signature production area as a reason limiting the signature presentation during the experiment. Another 26% participants indicated that it is due to the thin stylus pen and they had problems holding the pen. 28% participants indicated that neither of the factors restricted them from the signing process. 98% of the participants indicated that Sitting was a comfortable position for the signing process, 2% indicated that Treadmill and Walking scenario were also comfortable for them to sign.

8.5 Conclusion

The work presented in this chapter highlights the impact of mobile device usage scenarios and environmental characteristics on usability. Results cover two different environmental location types and the usage scenarios were selected as representative of realistic daily life situations. The results are presented in terms of effectiveness, efficiency and user satisfaction for keystroke dynamics and signatures.

For keystroke dynamics, the average number of errors and average time spent during the data donation by the participants were higher for scenarios having movements either from the user or the environment. Therefore, the efficiency error rate for Travelling on a Moving Bus and Walking Outdoors scenarios obtained worse error rate compared to the Sitting Indoors and the Treadmill scenarios.

For signature modality, the efficiency error rate for Session 1 (for all the three scenarios) were higher compared to Session 2. This indicates that the participants became familiar with the signature donation process in Session 2 and hence, spent less time and committed lesser errors in the second session. Compared to all the scenarios, the efficiency error rate of the Treadmill scenario was the highest. The reason for a performance deterioration in the Treadmill scenario could be due to the simultaneous movement of the user and treadmill during the signature donation process. Hence, it can be stated that signature presentation whilst on a Treadmill is not an unfavourable signature donation scenario. The treadmill scenario can be associated with user performing signature on a moving walkway in airports. In such practical situations, a poor efficiency can challenge the overall user experience of using the signature on a mobile device. Furthermore, the feedback received from the users after the completion of the experiment revealed that sitting was the most comfortable posture while performing the keystroke entry and signature donation tasks. The user feedback also revealed that the device factors such as screen size and soft-keyboard size and format were some of the limiting factors during the data donation.

Chapter 9. Conclusions and Future Work

9.1 Introduction

Biometric verification presents a number of challenges when implemented on a mobile device. Modality performance and user experience are considered to be the two key dimensions for measuring a successful implementation of a mobile biometric solution. Both of these factors are directly impacted by user interaction with the device. Whilst a majority of the work on behavioural biometrics on mobile devices have been dedicated towards improving the performance at the algorithmic level, less attention has been paid to understanding the user interaction aspect during the biometric process. Through this thesis, we have provided a comprehensive analysis of the user interaction with the mobile device, specifically while using behavioural modalities such as swipe gestures, signature and keystroke dynamics. The reason for choosing these three modalities is that all three require active user interaction with the device and are acquired from a single sensor, the touchscreen.

The work demonstrates the influence of factors related to the user and the surrounding environment on biometric performance and usability. The results obtained and the conclusions show the extent of the impact of a user's body movement and environmental changes on uni-modal and multi-modal biometric systems. The novelty of this work is the extensive evaluation over a real-life behavioural dataset that was collected in an unconstrained environment. The data collection was performed in an unsupervised manner to capture natural user actions on the mobile device. The multi-modal dataset was then used to a) establish how data from the embedded sensors on the mobile device can be leveraged to accurately verify a person using individual touch-based modalities; b) evaluate the stability of behavioural biometric modalities over different usage scenarios of a mobile device and understand the longevity of performance; c) perform usability analysis and d) evaluate how a multi-modal solution can improve the consistency of the modalities.

In the next section, our conclusions based on the evaluation results obtained from the experiments for individual research questions are presented. Following this, the lessons learned and the direction of future research in this area is provided.

9.2 Research Findings

This work was dedicated to assessing the performance stability and usability consistency of behavioural biometrics on mobile devices. Based on the literature review, four main research questions were identified (listed in Chapter 1) and evaluated throughout this research work. Based on the experimental results, a summary of our assessments (Table 9.1) and conclusions for individual research questions have been provided in this section.

Modality	Research Questions	Best Algorithm	Results (Metric Used – Mean Equal Error Rate)
Swipe Gestures	Optimum number of swipes required in the enrolment	SVM	6 swipes – 1%
		FF DNN	12 swipes – 9.2%
	Impact of usage scenarios	SVM	Intra-session – 1% (Sitting), 25% (Treadmill), 31% (Walking), 30% (Bus)
			Inter-session – 39% (Sitting) and 16% (Walking)
		FF DNN	Intra-session – 2.5% (Sitting), 8% (Treadmill), 6.25% (Walking), 6.4% (Bus)
			Inter-session – 11.96% (Sitting) and 7.8%(Walking)
Signature	Influence of input-type	DTW	Finger - 20%, Stylus - 27%
	Impact of usage scenarios	SVM	Finger-based - 13% (Sitting), 15% (Treadmill), 24% (Walking), 17% (Bus)
	Influence of time	SVM DTW	Finger-based – (Sitting) - 25% Stylus-based – (Walking Outdoors) - 35%
Keystroke Dynamics	Continuous authentication	Euclidean Distance	4%
	Amount of keystroke data required	Euclidean Distance	50 Digraphs – 2%
	Intra-session	Euclidean Distance-based	2% (Sitting), 50% (Treadmill), 51% (Walking), 44% (Bus)
	Inter-session comparison	Euclidean Distance-based	28%(Sitting), 42% (Walking)
Multi-modal	Swipe Gestures & Signature -Feature Fusion and Score Fusion	SVM, Commercial	3% (Sitting), 4%(Treadmill) 8%(Walking), 5% (bus) And 5% (Weighted score)
	Keystroke Dynamics & Swipe Gestures	SVM	5%(Sitting), 6%(Treadmill), 10%(Walking), 8%(Bus)
	Keystroke Dynamics & Signature	SVM	8% (Sitting), 10 (Treadmill), 18% (Walking), 13 (Bus)
	Swipe Gesture & Keystroke Dynamics & Signature	SVM	2% (Sitting), 3% (Treadmill), 8% (Walking), 9% (Bus)

Table 9.1. Overall summary of key results

Research Question 1: *Is the biometric performance of behavioural modalities consistent and stable in different operational usage scenarios of a mobile device?*

With an aim to assess the reliability of the biometric performance under different operational usage scenarios, we started with collection of an experimental dataset with diverse usage scenarios and environmental variation. The dataset contained a multi-modal (swipe gestures, signature and keystroke dynamics), multi-scenario and time-separated behavioural data acquired in an unconstrained environment across two sessions, separated by a week. Such a dataset enabled analysis of a variety of factors that impact the stability of the biometric performance.

Unlike other state-of-the art studies, the evaluation results demonstrated the impact of using real-life data on the verification performance. The evaluation for the uni-modal-based biometric verification methods were first performed for individual touch-based modalities, followed by a multi-modal approach. As employed across the state-of-the-art studies, multiple conventional and deep neural network-based classifiers were used to build the verification model and carry out evaluations.

The results obtained for swipe-gesture based verification showed that users can be verified with high accuracy when the enrolment and verification swipe samples belonged to the usage scenario with no body movement and no environmental variations. The conventional classifiers attained 1% and 2% mean EERs for horizontal and vertical swipes respectively. The deep neural network model's performance, although inferior to the conventional model, was in an acceptable range as well, with a mean EER of 2.7%. One of the reasons for the difference in performance was due to a lower number of samples present for training the deep neural network model. In contrast, the results demonstrate that when the verification samples belonged to the scenarios with body movement from the treadmill, walking outdoors or transport movement, verification performance deteriorated. The inter-session results obtained by comparing the swipe gestures over two time-separated sessions revealed performance deterioration. A possible reason could be that the user behaviour is more stable on the same day compared to multiple days.

The verification performance for the second touch-based modality, signature, was analysed based on three different methods – a black-box commercial system, a function-based signature verification system and a feature-based signature verification system. The results showed lower verification accuracy when the signatures belonged to the scenarios having movements either from the user or the environment. The commercial signature verification engine showed higher true rejection rate, however, a high false rejection rate of the genuine signatures was also reported. Further analysis showed that the false rejections happened for genuine signatures captured in usage scenarios having movements, while the enrolment signatures were taken from static scenario. These results, once again, pointed to performance deterioration owing to movement. It was observed that the variation in the signature input-tool was also detrimental to verification performance.

Performance for third touch-based modality, keystroke-dynamics, were obtained through a distance-based algorithm comparing digraphs information of the user. Analysis of keystroke data over different usage scenarios showed deterioration in the mean EER values. Similar performance impact was observed when time-separated keystroke data were compared.

When all three modalities were fused together, the verification performances across different usage scenarios improved compared to the uni-modal based results. Results from this combination also showed performance deterioration when static scenario was compared with the dynamic scenarios having influence from the body movement and environmental factors. However, the variation was lesser and in acceptable range compared to uni-modal systems. Therefore, it can be concluded that using the multi-modal solution, the impact from the usage scenarios were observed to be less.

Research Question 2: *Which factors in the user interaction process with the mobile device affect the overall biometric performance?*

From the verification perspective, seamless user interaction with the biometric sensor on the device throughout the entire verification process is ideal. However, when mobile devices are used for behavioural biometric verification with an unconstrained and unsupervised interaction process, it is crucial to evaluate the errors caused due to the components associated with the user interaction. The main components involved in the interaction process are the user, the environment and the device. This work assessed both the user and the environmental factors impacting the biometric performance.

Users can interact with their devices while walking. The walking speed of the user can impact the overall interaction. This work evaluated the impact of a user’s body movement under controlled and uncontrolled speeds. The performance assessment of the swipe gesture-based study showed a deterioration in performance using the conventional and DNN methods for the data acquired from the walking outdoors and walking on the treadmill indoors scenarios. The mean EER % of the controlled speed using the treadmill showed better performance compared to the uncontrolled speed. The signature and keystroke-based verification methods indicated similar performance variation.

The environmental variations considered were location - indoors and outdoors and the degree of movement by the environment (treadmill and moving transport). The evaluation results of the indoors scenarios yielded lower mean EER%, compared to the outdoors scenarios using swipe gestures, signature and keystroke dynamics verifiers. Considering the degree of movement by the environment, the scenario of the user seated on a moving transport showed poor performance compared to a user seated indoors and walking on a treadmill indoors. However, compared to the outdoor walking scenario, the moving transport scenario showed a better performance by attaining a lower mean EER%.

Based on the above findings, it can be concluded that user's body movement, environmental uncertainty such as an outdoor walking set-up, and movements caused by the environment negatively impact the biometric performance. However, when movements are controlled by the user (e.g. treadmill scenario with fixed speed), verification performance results are relatively better compared to scenarios with continuous, non-regular movements as expected in outdoor walking or while being on a moving transport.

Research Question 3: *How can we further improve recognition accuracy using multiple behavioural biometric modalities on a mobile device?*

The evaluation results of the uni-modal approaches for swipe gestures, signature and keystroke dynamics modalities showed the adverse impact of various user interaction factors on the verification performance. Hence, a multi-modal approach was designed with different combinations of modalities, and evaluations were performed to assess the impact of the user interaction factors to compare the uni-modal and multi-modal approaches. A feature-fusion method was applied for combining different modalities.

Fusion of the signature and the swipe gestures were performed using feature-fusion and score-fusion methods. With this approach, there was an improvement in the mean EER% compared to the uni-modal approach, however, we noticed that the performance for horizontal and vertical swipes showed similar performances for Session 1 and Session 2. The score fusion method revealed that using a weighted average technique yielded the best performance.

A combination of keystroke dynamics features with the horizontal and vertical swipe gestures separately showed similar results using SVM and k-NN algorithms for both the categories of swipes. We noticed the deterioration in the error rates when other usage scenario data were compared.

The results when the signature and the keystroke dynamics were combined using a feature-fusion method showed the worse performance of the minimum EER as 8%. Of all the three combinations, this combination resulted in the highest error rate even for the Sitting Indoors scenario and, hence, displayed poorer performance compared to other two combinations of modalities.

Finally, a combination of all the three modalities showed an improved performance across all the usage scenarios. Therefore, our recommendation is to use a multi-modal system to enhance performance and security.

Research Question 3: *How is the usability of behavioural biometrics affected by the adopted modalities and operational scenarios?*

In this study, we assessed the usability using ISO-defined metrics such as effectiveness, efficiency and user satisfaction. The evaluation was performed for keystroke dynamics and signature. The results show that the number of errors and average time spent on during the data donation were high for the scenarios with user movement and environmental changes. The feedback recorded during the data collection process revealed that the outdoors scenarios were the most difficult for the users to perform data donation.

This backs our conclusion for research question 2 that verification performance suffers the most for outdoor operational scenarios that not only bring the movement, but also uncertainty to the user interaction factors.

Based on the experimental results and overall observations, our recommendation is that when dynamic movements (such as the ones encountered in scenarios of outdoor walking or travelling on a moving bus) are detected on the mobile device, verification methodology should adopt one of the following solutions:

a) switch to a variable threshold-based verification model for different scenarios

b) build the enrolment set of the verification model comprising of data from different scenarios and consecutively use the appropriate enrolment set with movements for comparison during the verification process, or

c) adopt a multi-modal solution by combining either the implicit or the explicit behavioural data from the sensors available in the mobile device.

The study also observes that while a multi-modal solution may be a way forward to reduce the impact of usage scenarios on the verification accuracy, the data acquisition using multiple modalities is time consuming and tiring for the users. Each modality requires multiple samples to be provided. The same problem would exist when considering the solution of building the enrolment data based on different usage scenarios. Therefore, finding the right trade-off between capturing sufficient amounts of data from the users with lesser inconvenience whilst attaining high accuracy with the limited data remains a challenge.

9.3 Lessons Learned and Future Work

Through conducting extensive research on touch-dynamics based verification on mobile devices, there were a number of lessons learnt during the experimental phase. These are detailed below.

As the data collection was focused on capturing diverse usage scenarios of the device and variation in environmental surroundings, only one device was used for the data collection experiment. The UI of the device was fixed to portrait mode during the data collection. While these protocols helped to keep the number of variables low, they also potentially limited the users from donating data in a set-up that is completely natural to them. The future work can include different mobile device models and different UI modes for capturing data.

A ceremony-based data acquisition method was adopted for keystroke dynamics and signatures. For swipe gestures, although the app UI was designed to acquire horizontal and vertical swipes, the user had the option to perform any touch-action on the screen. In this study, we focused on specific types of swipe gestures – horizontal and vertical. However, when the swipe data was analysed, different categories of touch-actions such as zoom-in and zoom-out and a single touch such as button touches were observed. Methods to categorise these swipe gestures and to handle different swipes needs to be assessed. Another factor is related to handling of the multi-touch gestures. In this work, we assessed them as separate swipe gestures. However, a method to compare multi-touches with other multi-touch gestures needs to be developed.

While collecting the data, a few participants were reluctant to provide their signatures due to privacy concerns. However, none of the participants showed any hesitation while providing swipe-gestures and keystroke tasks. Future work can capture user perception and awareness of using behavioural biometrics for verification.

The multi-modal approach demonstrated an improved biometric verification performance across variations of usage scenarios of the device, however, the practical implementation of such a system would require sequential data collection from the touch sensor as the data capture occurs one after another for each modality from the same touch-sensor. This implementation is well suited for a ceremony-based verification. The future work should investigate how a multi-modal approach could be applied to a continuous verification framework.

References

[1] E. Ellavarason, R. Guest, F. Deravi, R. Sanchez-Reillo, and B. Corsetti, “Touch-dynamics based Behavioural Biometrics on Mobile Devices - A Review from a Usability and Performance Perspective,” *ACM Comput. Surv.*, vol. 53, no. 6, pp. 1–36, May 2020, doi: 10.1145/3394713.

[2] E. Ellavarason, R. Guest, and F. Deravi, “A framework for assessing factors influencing user interaction for touch-based biometrics,” in *26th European Signal Processing Conference (EUSIPCO), Rome*, 2018, pp. 553–557, doi: 10.23919/EUSIPCO.2018.8553537.

[3] E. Ellavarason, R. Guest, and F. Deravi, “Evaluation of stability of swipe gesture authentication across usage scenarios of mobile device,” *Eurasip J. Inf. Secur.*, no. 4, 2020, doi: 10.1186/s13635-020-00103-0.

[4] Arrepim, “Mobile Phone Market Forecast | 2017,” 2018. https://stats.areppim.com/stats/stats_mobilex2017.htm (accessed Sep. 11, 2020).

[5] Gartner, “Gartner Says Worldwide Smartphone Sales Will Grow 3% in 2020,” *www.gartner.com*, 2020. https://www.gartner.com/en/newsroom/press-releases/2020-01-28-gartner-says-worldwide-smartphone-sales-will-grow-3-- (accessed Sep. 11, 2020).

[6] Acquity, “Mobile Biometrics,” 2020. https://www.acuitymi.com/mobile-biometrics (accessed Oct. 28, 2020).

[7] I. Goicoechea-Telleria, A. Garcia-Peral, A. Husseis, and R. Sanchez-Reillo, “Presentation Attack Detection Evaluation on Mobile Devices: Simplest Approach for Capturing and Lifting a Latent Fingerprint,” in *2018 International Carnahan Conference on Security Technology (ICCST)*, Oct. 2018, pp. 1–5, doi: 10.1109/CCST.2018.8585605.

[8] International Organization for Standardization., “ISO/IEC 2382-37:2017(en), Information technology — Vocabulary — Part 37: Biometrics,” 2017. https://www.iso.org/standard/66693.html (accessed Oct. 23, 2020).

[9] R. V. Yampolskiy and V. Govindaraju, “Behavioural biometrics: a survey and classification,” *Int. J. Biom.*, vol. 1, no. 1, p. 81, 2008, doi: 10.1504/IJBM.2008.018665.

[10] Z. Sitova *et al.*, “HMOG: New Behavioral Biometric Features for Continuous Authentication of Smartphone Users,” *IEEE Trans. Inf. Forensics Secur.*, vol. 11, no. 5, pp. 877–892, 2016, doi: 10.1109/TIFS.2015.2506542.

[11] S. Alotaibi, S. Furnell, and N. Clarke, “Transparent authentication systems for mobile device security: A review,” in *2015 10th International Conference for Internet Technology and*

Secured Transactions, ICITST 2015, 2016, pp. 406–413, doi: 10.1109/ICITST.2015.7412131.

[12] H. Saevanee and P. Bhattarakosol, “Authenticating user using keystroke dynamics and finger pressure,” in *6th IEEE Consumer Communications and Networking Conference*, 2009, pp. 1–2, doi: 10.1109/CCNC.2009.4784783.

[13] A. K. Jain, A. Ross, and S. Prabhakar, “An Introduction to Biometric Recognition,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 4–20, 2004, doi: 10.1109/TCSVT.2003.818349.

[14] C. Bo, L. Zhang, X.-Y. Li, Q. Huang, and Y. Wang, “SilentSense: Silent User Identification via Dynamics of Touch and Movement Behavioral Biometrics,” *MobiCom ’13*, p. 187, 2013, doi: 10.1145/2500423.2504572.

[15] E. Shi, Y. Niu, M. Jakobsson, and R. Chow, “Implicit authentication through learning user behavior,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2011, vol. 6531 LNCS, pp. 99–113, doi: 10.1007/978-3-642-18178-8-9.

[16] T. Feng and Y. Jun Yang Zhixian, “TIPS: Context-Aware Implicit User Identification using Touch Screen in Uncontrolled Environments,” in *In Proceedings of the 15th Workshop on Mobile Computing Systems and Applications (HotMobile ’14). Association for Computing Machinery, New York, NY, USA*, pp. 1–6.

[17] K. A. Rahman, K. S. Balagani, and V. V. Phoha, “Snoop-forge-replay attacks on continuous verification with keystrokes,” *IEEE Trans. Inf. Forensics Secur.*, vol. 8, no. 3, pp. 528–541, 2013, doi: 10.1109/TIFS.2013.2244091.

[18] A. Serwadda, V. V. Phoha, Z. Wang, R. Kumar, and D. Shukla, “Toward robotic robbery on the touch screen,” *ACM Trans. Inf. Syst. Secur.*, vol. 18, no. 4, pp. 1–25, 2016, doi: 10.1145/2898353.

[19] T. Feng *et al.*, “Continuous mobile authentication using touchscreen gestures,” in *2012 IEEE International Conference on Technologies for Homeland Security, HST 2012*, 2012, pp. 451–456, doi: 10.1109/THS.2012.6459891.

[20] N. Zheng, K. Bai, H. Huang, and H. Wang, “You are how you touch: User verification on smartphones via tapping behaviors,” in *Proceedings - International Conference on Network Protocols, ICNP*, 2014, pp. 221–232, doi: 10.1109/ICNP.2014.43.

[21] H. Xu, Y. Zhou, and M. R. Lyu, “Towards Continuous and Passive Authentication via Touch Biometrics: An Experimental Study on Smartphones,” *10th Symp. Usable Priv. Secur.*, pp. 187–198, 2014, doi: 10.1145/2559206.2581364.

[22] M. Brockly, R. Guest, S. Elliott, and J. Scott, “Dynamic signature verification and the human

biometric sensor interaction model," in *International Carnahan Conference on Security Technology*, 2011, pp. 1–6, doi: 10.1109/CCST.2011.6095937.

[23] E. P. Kukula, M. J. Sutton, and S. J. Elliott, "The humanbiometric-sensor interaction evaluation method: Biometric performance and usability measurements," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 4, pp. 1–8, 2010, doi: 10.1109/TIM.2009.2037878.

[24] R. Blanco-Gonzalo, R. Sanchez-Reillo, O. Miguel-Hurtado, and E. Bella-Pulgarin, "Automatic usability and stress analysis in mobile biometrics," *Image Vis. Comput.*, vol. 32, no. 12, pp. 1173–1180, 2014, doi: 10.1016/j.imavis.2014.09.003.

[25] R. Blanco-Gonzalo, R. Sanchez-reillo, O. Miguel-Hurtado, and J. Liu-jimenez, "Usability analysis of dynamic signature verification in mobile environments," *Biometrics Spec. Interes. Gr. (BIOSIG), 2013 Int. Conf.*, pp. 1–9, 2013.

[26] W. Karwowski, "The Discipline of Ergonomics and Human Factors," in *Handbook of Human Factors and Ergonomics*, 2006.

[27] Y. Zhong and Y. Deng, "A Survey on Keystroke Dynamics Biometrics: Approaches, Advances, and Evaluations," in *Recent Advances in User Authentication Using Keystroke Dynamics Biometrics*, 2015, pp. 1–22.

[28] T. Shimshon, R. Moskovitch, L. Rokach, and Y. Elovici, "Clustering di-graphs for continuously verifying users according to their typing patterns," in *2010 IEEE 26th Convention of Electrical and Electronics Engineers in Israel, IEEEI 2010*, 2010, pp. 445–449, doi: 10.1109/EEEI.2010.5662182.

[29] H. Saevanee and P. Bhatarakosol, "User authentication using combination of behavioral biometrics over the touchpad acting like touch screen of mobile device," in *Proceedings of the 2008 International Conference on Computer and Electrical Engineering, ICCEE 2008*, 2008, pp. 82–86, doi: 10.1109/ICCEE.2008.157.

[30] X. Huang, G. Lund, and A. Sapeluk, "Development of a typing behaviour recognition mechanism on android," *Proc. 11th IEEE Int. Conf. Trust. Secur. Priv. Comput. Commun. Trust. - 11th IEEE Int. Conf. Ubiquitous Comput. Commun. IUCC-2012*, pp. 1342–1347, 2012, doi: 10.1109/TrustCom.2012.127.

[31] F. O. Trojahn, Matthias, "Biometric Authentification Through a Virtual Keyboard for Smarthphones," *Int. J. Comput. Sci. Inf. Technol.*, vol. 4, no. 5, pp. 1–12, 2012, doi: 10.5121/ijcsit.2012.4501.

[32] H. Gascon, S. Uellenbeck, C. Wolf, and K. Rieck, "Continuous authentication on mobile devices by analysis of typing motion behavior," in *Sicherheit 2014–Sicherheit, Schutz und Zuverlässigkeit*, 2014, pp. 1–12.

[33] C. Giuffrida, K. Majdanik, M. Conti, and H. Bos, "I sensed it was you: Authenticating mobile users with sensor-enhanced keystroke dynamics," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014, vol. 8550 LNCS, pp. 92–111, doi: 10.1007/978-3-319-08509-8_6.

[34] U. Burgbacher and K. Hinrichs, "An Implicit Author Verification System for Text Messages Based on Gesture Typing Biometrics," *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, pp. 2951–2954, 2014, doi: 10.1145/2556288.2557346.

[35] H. Crawford and E. Ahmadzadeh, "Authentication on the Go: Assessing the Effect of Movement on Mobile Device Keystroke Dynamics," *Thirteen. Symp. Usable Priv. Secur.*, no. Soups, pp. 163–173, 2017.

[36] J. Miya, M. Bhatt, M. Gupta, M. Anas, G. College, and G. Noida, "A Two Factor Authentication System for Touchscreen Mobile Devices Using Static Keystroke Dynamics and Password," *Int. Res. J. Eng. Technol.*, no. 2395–0072, pp. 1829–1832, 2017.

[37] K. R. Corpus, R. J. D. Gonzales, A. S. Morada, and L. A. Vea, "Mobile user identification through authentication using keystroke dynamics and accelerometer biometrics," in *Proceedings of the International Workshop on Mobile Software Engineering and Systems - MOBILESoft '16*, 2016, pp. 11–12, doi: 10.1145/2897073.2897111.

[38] Y. Zhang, M. Yang, Z. Ling, Y. Liu, and W. Wu, "FingerAuth: 3D magnetic finger motion pattern based implicit authentication for mobile devices," *Futur. Gener. Comput. Syst.*, pp. 325–330, 2018, doi: 10.1016/j.future.2018.02.006.

[39] H. Lee, J. Y. Hwang, D. I. Kim, S. Lee, S. H. Lee, and J. S. Shin, "Understanding Keystroke Dynamics for Smartphone Users Authentication and Keystroke Dynamics on Smartphones Built-In Motion Sensors," *Secur. Commun. Networks*, 2018, doi: 10.1155/2018/2567463.

[40] I. Lamiche, G. Bin, Y. Jing, Z. Yu, and A. Hadid, "A continuous smartphone authentication method based on gait patterns and keystroke dynamics," *J. Ambient Intell. Humaniz. Comput.*, 2019, doi: 10.1007/s12652-018-1123-6.

[41] H. Kalita, E. Maiorana, and P. Campisi, "Keystroke Dynamics for Biometric Recognition in Handheld Devices," 2020, doi: 10.1109/TSP49548.2020.9163524.

[42] P. S. The, N. Zhang, A. B. J. Teoh, and K. Chen, "Recognizing Your Touch: Towards Strengthening Mobile Device Authentication via Touch Dynamics Integration," 2015, doi: 10.1145/2837126.2837127.

[43] M. Antal and L. Nemes, "The MOBIKEY keystroke dynamics password database: Benchmark results," 2016, doi: 10.1007/978-3-319-33622-0_4.

[44] M. J. Coakley, J. V. Monaco, and C. C. Tappert, "Keystroke biometric studies with short

numeric input on smartphones," 2016, doi: 10.1109/BTAS.2016.7791181.

[45] M. Trojahn and F. Ortmeier, "Toward mobile authentication with keystroke dynamics on mobile phones and tablets," in *Proceedings - 27th International Conference on Advanced Information Networking and Applications Workshops, WAINA 2013*, 2013, pp. 697–702, doi: 10.1109/WAINA.2013.36.

[46] B. A. Smith, X. Bi, and S. Zhai, "Optimizing touchscreen keyboards for gesture typing," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.*, 2015, pp. 3365–337, doi: 10.1145/2702123.2702357.

[47] X. Bi, S. Azenkot, K. Partridge, and S. Zhai, "Octopus: Evaluating touchscreen keyboard correction and recognition algorithms via 'Remulation,'" in *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2013, pp. 543–552, doi: 10.1145/2470654.2470732.

[48] D. Buschek, A. De Luca, and F. Alt, "Improving Accuracy, Applicability and Usability of Keystroke Biometrics on Mobile Touchscreen Devices," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, 2015, pp. 1393–1402, doi: 10.1145/2702123.2702252.

[49] "Inside iOS 11: Apple's new one-handed keyboard allows for easier typing | Appleinsider." https://appleinsider.com/articles/17/09/21/inside-ios-11-apples-new-one-handed-keyboard-allows-for-easier-typing (accessed May 13, 2020).

[50] "[K9 keyboard problem]Finding the words by ranking - updating." http://petercrushcode.blogspot.com/2016/04/finding-words-by-ranking-code-complete.html (accessed May 13, 2020).

[51] J. Cuaresma and I. MacKenzie, "A study of variations of Qwerty soft keyboards for mobile phones," *Proc. Int. Conf. Multimed. Human-Computer Interact. - MHCI 2013*, no. 126, pp. 1–8, 2013, [Online]. Available: http://www.yorku.ca/mack/mhci2013g.html.

[52] "Blackberry Curve Keyboard 1.1 Apk | APK Tools." https://apk.tools/details-blackberry-curve-keyboard-apk/ (accessed May 13, 2020).

[53] F. Bergadano, D. Gunetti, and C. Picardi, "User Authentication through Keystroke Dynamics," *ACM Trans. Inf. Syst. Secur.*, vol. 5, no. 4, pp. 367–397, 2002, doi: 10.1145/581271.581272.

[54] J. M. Chang *et al.*, "Capturing cognitive fingerprints from keystroke dynamics," *IT Prof.*, vol. 15, no. 4, pp. 24–28, 2013, doi: 10.1109/MITP.2013.52.

[55] R. Bixler and S. D'Mello, "Detecting boredom and engagement during writing with keystroke analysis, task appraisals, and stable traits," in *Proceedings of the 2013 international conference on Intelligent user interfaces - IUI '13*, 2013, p. 225, doi:

10.1145/2449396.2449426.

[56] C. Epp, M. Lippold, and R. L. Mandryk, “Identifying emotional states using keystroke dynamics,” in *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI ’11*, 2011, p. 715, doi: 10.1145/1978942.1979046.

[57] A. L. Smith and B. S. Chaparro, “Smartphone Text Input Method Performance, Usability, and Preference with Younger and Older Adults,” *Hum. Factors*, vol. 57, no. 6, pp. 1015–1028, 2015, doi: 10.1177/0018720815575644.

[58] W. S. Wijesoma, K. W. Yue, K. L. Chien, and T. K. Chow, “Online Handwritten Signature Verification for Electronic Commerce over the Internet,” in *Lecture Notes in Computer Science*, 2001.

[59] D. Impedovo and G. Pirlo, “Automatic signature verification: The state of the art,” *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 38, no. 5, pp. 609–635, 2008, doi: 10.1109/TSMCC.2008.923866.

[60] Y. M. Al-Omari, S. N. H. S. Abdullah, and K. Omar, “State-of-the-art in offline signature verification system,” *2011 Int. Conf. Pattern Anal. Intell. Robot.*, vol. 1, no. June, pp. 59–64, 2011, doi: 10.1109/ICPAIR.2011.5976912.

[61] R. Plamondon and S. N. Srihari, “Online and off-line handwriting recognition: a comprehensive survey,” *Pattern Analysis and Machine …*, vol. 22, no. 1. pp. 63–84, 2000, doi: 10.1109/34.824821.

[62] C. Schmidt and K. Kraiss, “Establishment of personalized templates for automatic signature verification,” *Proc. Fourth Int. Conf. Doc. Anal. Recognit.*, vol. 1, pp. 263–267, 1997, doi: 10.1109/ICDAR.1997.619853.

[63] M. Shafiei and H. Rabiee, “A new online signature verification algorithm using variable length segmentation and hidden Markov models,” *Seventh Int. Conf. Doc. Anal. Recognition, 2003. Proceedings.*, pp. 1–4, 2003, doi: 10.1109/ICDAR.2003.1227706.

[64] R. Sanchez-Reillo, J. Liu-Jimenez, R. Blanco-Gonzalo, and O. Miguel-Hurtado, “Performance evaluation of handwritten signature recognition in mobile environments,” *IET Biometrics*, vol. 3, no. 3, pp. 139–146, 2014, doi: 10.1049/iet-bmt.2013.0044.

[65] M. Diaz, A. Fischer, M. A. Ferrer, and R. Plamondon, “Dynamic Signature Verification System Based on One Real Signature,” *IEEE Trans. Cybern.*, vol. PP, no. 99, pp. 1–12, 2017, doi: 10.1109/TCYB.2016.2630419.

[66] J. Galbally, R. P. Krish, J. Fierrez, and M. Martinez-Diaz, “Mobile signature verification: feature robustness and performance comparison,” *IET Biometrics*, vol. 3, no. 4, pp. 267–277, 2014, doi: 10.1049/iet-bmt.2013.0081.

[67] G. Bailador, C. Sanchez-Avila, J. Guerra-Casanova, and A. De Santos Sierra, “Analysis of pattern recognition techniques for in-air signature biometrics,” *Pattern Recognit.*, vol. 44, no. 10–11, pp. 2468–2478, 2011, doi: 10.1016/j.patcog.2011.04.010.

[68] S. Jabin and F. J. Zareen, “Authentic mobile-biometric signature verification system,” *IET Biometrics*, vol. 5, no. 1, pp. 13–19, 2016, doi: 10.1049/iet-bmt.2015.0017.

[69] D. Y. Yeung *et al.*, “SVC2004: First international signature verification competition,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 2004, doi: 10.1007/978-3-540-25948-0_3.

[70] A. Mendaza-Ormaza, O. Miguel-Hurtado, R. Blanco-Gonzalo, and F. J. Diez-Jimeno, “Analysis of handwritten signature performances using mobile devices,” in *IEEE Carnahan Conference on Security Technology*, 2011, pp. 1–6, doi: 10.1109/CCST.2011.6095930.

[71] R. P. Krish, J. Fierrez, J. Galbally, and M. Martinez-Diaz, “Dynamic signature verification on smart phones,” in *Communications in Computer and Information Science*, 2013, vol. 365, pp. 213–222, doi: 10.1007/978-3-642-38061-7_21.

[72] M. Zalasiński, K. Cpałka, and Y. Hayashi, “A new approach to the dynamic signature verification aimed at minimizing the number of global features,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, vol. 9693, pp. 218–231, doi: 10.1007/978-3-319-39384-1_20.

[73] K. Cpałka, M. Zalasiński, and L. Rutkowski, “A new algorithm for identity verification based on the analysis of a handwritten dynamic signature,” *Appl. Soft Comput. J.*, vol. 43, pp. 47–56, 2016, doi: 10.1016/j.asoc.2016.02.017.

[74] M. E. Yahyatabar and J. Ghasemi, “Online signature verification using double-stage feature extraction modelled by dynamic feature stability experiment,” *IET Biometrics*, vol. 6, no. 6, 2017, doi: 10.1049/iet-bmt.2016.0103.

[75] V. Bharadi, B. Pandva, and G. Cosma, “Multi-Modal Biometric Recognition Using Human Iris and Dynamic Pressure Variation of Handwritten Signatures,” in *Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS).*, 2018, pp. 233–238, doi: 10.1109/SNAMS.2018.8554960.

[76] K. Riesen and R. Schmidt, “Online signature verification based on string edit distance,” *Int. J. Doc. Anal. Recognit.*, vol. 22, pp. 41–54, 2019, doi: 10.1007/s10032-019-00316-1.

[77] A. Kholmatov and B. Yanikoglu, “SUSIG: an on-line signature database, associated protocols and benchmark results,” *Pattern Anal. Appl. 2008 123*, vol. 12, no. 3, pp. 227–236, Apr. 2008, doi: 10.1007/S10044-008-0118-X.

[78] M. Liwicki *et al.*, “Signature Verification Competition for Online and Offline Skilled

Forgeries (SigComp2011),” 2011, doi: 10.1109/ICDAR.2011.294.

[79] A. Jain, · Satish, K. Singh, K. P. Singh, and S. K. Singh, “Handwritten signature verification using shallow convolutional neural network,” doi: 10.1007/s11042-020-08728-6.

[80] N. Paudel, M. Querini, and G. F. Italiano, “Handwritten Signature Verification for Mobile Phones.,” *Icissp*, no. May, pp. 46–52, 2016, doi: 10.5220/0005675200460052.

[81] U. A. M. Dıaz, Marcos Martınez. Doctoral dissertation, “Dynamic signature verification for portable devices,” 2008.

[82] R. Blanco-Gonzalo, L. Diaz-Fernandez, O. Miguel-Hurtado, and R. Sanchez-Reillo, “Usability Evaluation of Biometrics in Mobile Environments,” *Adv. Intell. Syst. Comput.*, vol. 300, pp. 289–300, 2014, doi: 10.1007/978-3-319-08491-6_24.

[83] C.-C. Yu, H.-Y. Cheng, V. Gau, and C.-L. Lin, “Video-based signature verification and pen-grasping posture analysis for user-dependent identification authentication,” *IET Comput. Vis.*, vol. 6, no. 5, pp. 388–396, 2012, doi: 10.1049/iet-cvi.2010.0136.

[84] M. Savov and G. Gluhchev, “Signature Verification via ‘ Hand-Pen ’ Motion Investigation,” *Cybern. Inf. Technol.*, vol. 6, no. 1, 2006.

[85] J. Galbally, M. Martinez-Diaz, and J. Fierrez, “Aging in Biometrics: An Experimental Analysis on On-Line Signature,” *PLoS One*, 2013, doi: 10.1371/journal.pone.0069897.

[86] X. Wang, T. Yu, O. Mengshoel, and P. Tague, “Towards Continuous and Passive Authentication Across Mobile Devices: An Empirical Study,” *WiSec ’17 (ACM conference on Security and privacy in wireless & mobile networks)*, vol. 11. 2017, doi: 10.1145/3098243.3098244.

[87] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song, “Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication,” *IEEE Trans. Inf. Forensics Secur.*, vol. 8, no. 1, pp. 136–148, 2013, doi: 10.1109/TIFS.2012.2225048.

[88] C. Shen, Y. Zhang, X. Guan, and R. A. Maxion, “Performance Analysis of Touch-Interaction Behavior for Active Smartphone Authentication,” *IEEE Trans. Inf. Forensics Secur.*, vol. 11, no. 3, pp. 1–1, 2015, doi: 10.1109/TIFS.2015.2503258.

[89] H. Zhang, V. M. Patel, M. Fathy, and R. Chellappa, “Touch gesture-based active user authentication using dictionaries,” in *Proceedings - 2015 IEEE Winter Conference on Applications of Computer Vision, WACV 2015*, 2015, pp. 207–214, doi: 10.1109/WACV.2015.35.

[90] P. Saravanan, S. Clarke, D. H. (Polo) Chau, and H. Zha, “LatentGesture: Active User Authentication through Background Touch Analysis,” *Proc. Second Int. Symp. Chinese CHI - Chinese CHI ’14*, pp. 110–113, 2014, doi: 10.1145/2592235.2592252.

[91] O. Miguel-Hurtado, S. V. Stevenage, C. Bevan, and R. Guest, “Predicting sex as a soft-biometrics from device interaction swipe gestures,” *Pattern Recognit. Lett.*, vol. 79, pp. 44–51, 2016, doi: 10.1016/j.patrec.2016.04.024.

[92] R. Guest, O. Miguel-Hurtado, S. V. Stevenage, G. J. Neil, and S. Black, “Biometrics within the SuperIdentity project: A new approach to spanning multiple identity domains,” in *Proceedings - International Carnahan Conference on Security Technology*, 2014, vol. 2014-Octob, no. October, doi: 10.1109/CCST.2014.6986992.

[93] A. Serwadda, V. V. Phoha, and Z. Wang, “Which verifiers work?: A benchmark evaluation of touch-based authentication algorithms,” *IEEE 6th Int. Conf. Biometrics Theory, Appl. Syst. BTAS 2013*, 2013, doi: 10.1109/BTAS.2013.6712758.

[94] L. Li, X. Zhao, and G. Xue, “Unobservable Re-authentication for Smartphones.,” in *NDSS - Network and Distributed System Security Symposium*, 2013, pp. 1–16, [Online]. Available: http://internetsociety.org/doc/unobservable-re-authentication-smartphones.

[95] C. Bo, L. Zhang, T. Jung, J. Han, X. Y. Li, and Y. Wang, “Continuous user identification via touch and movement behavioral biometrics,” in *IEEE 33rd International Performance Computing and Communications Conference (IPCCC).*, 2014, pp. 1–8, doi: 10.1109/PCCC.2014.7017067.

[96] X. Zhao, T. Feng, W. Shi, and I. A. Kakadiaris, “Mobile user authentication using statistical touch dynamics images,” *IEEE Trans. Inf. Forensics Secur.*, vol. 9, no. 11, pp. 1780–1789, 2014, doi: 10.1109/TIFS.2014.2350916.

[97] V. Sharma and R. Enbody, “User authentication and identification from user interface interactions on touch-enabled devices,” in *Proceedings of the 10th ACM Conference on Security and Privacy in Wireless and Mobile Networks - WiSec ’17*, 2017, pp. 1–11, doi: 10.1145/3098243.3098262.

[98] J. Ahmad, M. Sajjad, Z. Jan, I. Mehmood, S. Rho, and S. W. Baik, “Analysis of interaction trace maps for active authentication on smart devices,” *Multimed. Tools Appl.*, vol. 76, no. 3, pp. 4069–4087, 2017, doi: 10.1007/s11042-016-3450-y.

[99] A. I. Filippov, A. V. Iuzbashev, and A. S. Kurnev, “User authentication via touch pattern recognition based on isolation forest,” in *Proceedings of the 2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, ElConRus 2018*, 2018, vol. 2018-Janua, pp. 1485–1489, doi: 10.1109/EIConRus.2018.8317378.

[100] P. Siirtola, J. Komulainen, and V. Kellokumpu, “Effect of context in swipe gesture-based continuous authentication on smartphones,” 2019. [Online]. Available: http://www.oulu.fi/bisg/node/40364.

[101] T. Feng *et al.*, “An investigation on touch biometrics: Behavioral factors on screen size,

physical context and application context," in *IEEE International Symposium on Technologies for Homeland Security (HST)*, 2015, pp. 1–6, doi: 10.1109/THS.2015.7225318.

[102] J. Fierrez, A. Pozo, M. Martinez-Diaz, J. Galbally, and A. Morales, "Benchmarking Touchscreen Biometrics for Mobile Authentication," *IEEE Trans. Inf. Forensics Secur.*, 2018, doi: 10.1109/TIFS.2018.2833042.

[103] C. Bevan and D. S. Fraser, "Different strokes for different folks? Revealing the physical characteristics of smartphone users from their swipe gestures," *Int. J. Hum. Comput. Stud.*, vol. 88, pp. 51–61, 2016, doi: 10.1016/j.ijhcs.2016.01.001.

[104] D. Buschek, A. De Luca, and F. Alt, "Evaluating the Influence of Targets and Hand Postures on Touch-based Behavioural Biometrics," *Proc. 2016 CHI Conf. Hum. Factors Comput. Syst. - CHI '16*, pp. 1349–1361, 2016, doi: 10.1145/2858036.2858165.

[105] M. A. Sasse, S. Brostoff, and D. Weirich, "Transforming the 'weakest link' - A human/computer interaction approach to usable and effective security," *BT Technol. J.*, 2001, doi: 10.1023/A:1011902718709.

[106] O. Miguel-Hurtado, R. Blanco-Gonzalo, R. Guest, and C. Lunerti, "Interaction evaluation of a mobile voice authentication system," in *In 2016 IEEE International Carnahan Conference on Security Technology (ICCST)*, 2017, pp. 1–8, doi: 10.1109/CCST.2016.7815697.

[107] S. Mondal and P. Bours, "Swipe gesture based Continuous Authentication for mobile devices," *2015 International Conference on Biometrics (ICB)*. pp. 458–465, 2015, doi: 10.1109/ICB.2015.7139110.

[108] I. Jtc, "ISO/IEC JTC 1/SC 37 N 1768 Text of FDIS 19795-2, Information technology – Biometric performance testing and reporting – Part 2: Testing methodologies for technology and scenario evaluation," 2006.

[109] M. D. Papamichail, K. C. Chatzidimitriou, T. Karanikiotis, N.-C. I. Oikonomou, A. L. Symeonidis, and S. K. Saripalle, "BrainRun: A Behavioral Biometrics Dataset towards Continuous Implicit Authentication," *Data*, 2019, doi: 10.3390/data4020060.

[110] U. Mahbub, S. Sarkar, V. M. Patel, and R. Chellappa, "Active user authentication for smartphones: A challenge data set and benchmark results," in *IEEE 8th international conference on biometrics theory, applications and systems (BTAS)*, 2016, pp. 1–8, doi: 10.1109/BTAS.2016.7791155.

[111] A. Morales *et al.*, "KBOC: Keystroke biometrics OnGoing competition," in *IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 2016, pp. 1–6, doi: 10.1109/BTAS.2016.7791180.

[112] M. Antal and L. Z. Szabó, "Biometric Authentication Based on Touchscreen Swipe Patterns,"

Procedia Technol., vol. 22, no. October 2015, pp. 862–869, 2016, doi: 10.1016/j.protcy.2016.01.061.

[113] M. El-Abed, M. Dafer, and R. El Khayat, “RHU Keystroke: A mobile-based benchmark for keystroke dynamics systems,” in *Proceedings - International Carnahan Conference on Security Technology*, 2014, vol. 2014-Octob, no. October, doi: 10.1109/CCST.2014.6986984.

[114] C. J. Tasia, T. Y. Chang, P. C. Cheng, and J. H. Lin, “Two novel biometric features in keystroke dynamics authentication systems for touch screen devices,” *Secur. Commun. Networks*, vol. 7, no. 4, pp. 750–758, 2014, doi: 10.1002/sec.776.

[115] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, “Benchmarking desktop and mobile handwriting across COTS devices: The e-BioSign biometric database,” *PLoS One*, 2017, doi: 10.1371/journal.pone.0176792.

[116] M. Antal, Z. Bokor, and L. Z. Szabó, “Information revealed from scrolling interactions on mobile devices,” *Pattern Recognit. Lett.*, vol. 56, pp. 7–13, 2015, doi: 10.1016/j.patrec.2015.01.011.

[117] M. Antal, L. Z. Szabó, and I. László, “Keystroke Dynamics on Android Platform,” *Procedia Technol.*, vol. 19, pp. 820–826, 2015, doi: 10.1016/j.protcy.2015.02.118.

[118] J. Fierrez *et al.*, “BiosecurID: A multimodal biometric database,” *Pattern Anal. Appl.*, vol. 13, no. 2, pp. 235–246, 2010, doi: 10.1007/s10044-009-0151-4.

[119] “Galaxy Note5 | Samsung Support LEVANT.” https://www.samsung.com/levant/support/model/SM-N920CZKAMID/ (accessed Dec. 09, 2020).

[120] M. Santopietro, R. Vera-Rodriguez, R. Guest, A. Morales, and A. Acien, “Assessing the Quality of Swipe Interactions for Mobile Biometric Systems,” 2020.

[121] N. C. A. and T. S. U. Montgomery, “Touch-Based Continuous Authentication Using Deep Neural Net and Genetic Algorithm,” 2019.

[122] ISO/IEC, “ISO/IEC 19795-1:2006(en), Information technology — Biometric performance testing and reporting — Part 1: Principles and framework,” *19795-1:2006*, 2006. .

[123] W. Shi, J. Yang, Y. Jiang, F. Yang, and Y. Xiong, “SenGuard: Passive user identification on smartphones using multiple sensors,” in *International Conference on Wireless and Mobile Computing, Networking and Communications*, 2011, pp. 141–148, doi: 10.1109/WiMOB.2011.6085412.

[124] “TensorFlow.” https://www.tensorflow.org/ (accessed Dec. 10, 2020).

[125] J. Ortega-Garcia *et al.*, “The multiscenario multienvironment biosecure multimodal database (BMDB),” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 6, pp. 1097–1111, 2010, doi:

10.1109/TPAMI.2009.76.

[126] O. Miguel-Hurtado, L. Mengibar-Pozo, M. G. Lorenz, and J. Liu-Jimenez, “On-line signature verification by dynamic time warping and Gaussian mixture models,” in *41st annual IEEE international Carnahan conference on security technology*, 2007, pp. 23–29, doi: 10.1109/CCST.2007.4373463.

[127] “PENTOOLS - A MATLAB Toolkit for On-line Pen-Based Data Experimentation,” in *10th International Conference on Document Analysis and Recognition. IEEE*, 2009, pp. 1121–1125, Accessed: Nov. 16, 2020. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5277641.

[128] “3.2. Tuning the hyper-parameters of an estimator — scikit-learn 0.23.2 documentation.” https://scikit-learn.org/stable/modules/grid_search.html (accessed Dec. 11, 2020).

[129] V. Levenshtein, “Binary codes capable of correcting deletions, insertions, and reversals,” *Sov. Phys. Dokl.*, vol. 10, no. 8, pp. 707–710, 1966.

[130] B. Draffin, J. Zhu, and J. Zhang, “KeySens : Passive User Authentication through Micro-behavior Modeling of Soft Keyboard Interaction,” *Mob. Comput. Appl. Serv.*, vol. 130, pp. 184–201, 2014, doi: 10.1007/978-3-319-05452-0_14.

[131] N. L. Clarke and S. M. Furnell, “Advanced user authentication for mobile devices,” *Comput. Secur.*, vol. 26, no. 2, pp. 109–119, Mar. 2007, doi: 10.1016/j.cose.2006.08.008.

[132] H. Saevanee, N. Clarke, S. Furnell, and V. Biscione, “Continuous user authentication using multi-modal biometrics,” *Comput. Secur.*, vol. 53, pp. 234–246, 2015, doi: 10.1016/j.cose.2015.06.001.

[133] N. L. Clarke and S. M. Furnell, “Advanced user authentication for mobile devices,” *Comput. Secur.*, vol. 26, no. 2, pp. 109–119, 2007, doi: 10.1016/j.cose.2006.08.008.

[134] N. L. Clarke and S. M. Furnell, “Authenticating mobile phone users using keystroke analysis,” *Int. J. Inf. Secur.*, vol. 6, no. 1, pp. 1–14, 2007, doi: 10.1007/s10207-006-0006-6.

[135] T.-E. W. Jain-Shing Wu, Chih-Ta Lin,Wan-Ching Lin, “Smartphone Continuous Authentication based on Keystroke and Gesture Profiling,” in *International Carnahan Conference on Security Technology (ICCST), Taipei*, 2015, pp. 191–197, Accessed: Jun. 23, 2020. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7389681.

[136] H. Saevanee and P. Bhattarakosol, “Authenticating user using keystroke dynamics and finger pressure,” in *6th IEEE Consumer Communications and Networking Conference, CCNC 2009*, 2009, pp. 1–2, doi: 10.1109/CCNC.2009.4784783.

[137] “How to Get Only Numeric Value From TextInput in React Native.”

https://reactnativecode.com/get-only-numeric-value-textinput/ (accessed Dec. 11, 2020).

[138] R. Blanco Gonzalo *et al.*, “Attacking a Smartphone Biometric Fingerprint System: A Novice’s Approach,” in *International Carnahan Conference on Security Technology (ICCST). IEEE*, 2018, pp. 1–5, doi: 10.1109/CCST.2018.8585726.

[139] A. Ross and A. K. Jain, “Multimodal biometrics: An overview,” in *12th European Signal Processing Conference. IEEE*, 2004, pp. 1221–1224.

[140] H. Saevanee, N. L. Clarke, and S. M. Furnell, “Multi-modal behavioural biometric authentication for mobile devices,” in *IFIP Advances in Information and Communication Technology*, 2012, vol. 376 AICT, pp. 465–474, doi: 10.1007/978-3-642-30436-1_38.

[141] M. Tanviruzzaman and S. I. Ahamed, “Your phone knows you: Almost transparent authentication for smartphones,” in *Proceedings - International Computer Software and Applications Conference*, 2014, pp. 374–383, doi: 10.1109/COMPSAC.2014.60.

[142] H. Crawford, K. Renaud, and T. Storer, “A framework for continuous, transparent mobile device authentication,” *Comput. Secur.*, vol. 39, no. PART B, pp. 127–136, 2013, doi: 10.1016/j.cose.2013.05.005.

[143] S. Wold, K. Esbensen, and P. Geladi, “Principal component analysis,” *Chemom. Intell. Lab. Syst.*, 1987, doi: 10.1016/0169-7439(87)80084-9.

[144] G. and S. International Standards for Business, “ISO 13407:1999, Human-centred design processes for interactive systems,” 1999, [Online]. Available: https://www.iso.org/standard/21197.html.

[145] T. Alhussain, R. AlGhamdi, S. Alkhalaf, and O. Alfarraj, “Users’ Perceptions of Mobile Phone Security: A Survey Study in the Kingdom of Saudi Arabia,” *Int. J. Comput. Theory Eng.*, vol. 5, no. 5, pp. 793–796, 2013, doi: 10.7763/IJCTE.2013.V5.798.

[146] S. M. Furnell, P. S. Dowland, H. M. Illingworth, and P. L. Reynolds, “Authentication and Supervision: A Survey of User Attitudes,” in *Computers & Security*, vol. 19, no. 6, 2000, pp. 529–539.

[147] S. Karatzouni, S. M. Furnell, N. L. Clarke, and R. A. Botha, “Perceptions of user authentication on mobile devices,” *Proc. ISOneWorld Conf.*, pp. 11–13, 2007.

[148] H. Khan, U. Hengartner, and D. Vogel, “Usability and Security Perceptions of Implicit Authentication : Convenient , Secure , Sometimes Annoying,” *Elev. Symp. Usable Priv. Secur. (SOUPS 2015)*, 2015.

[149] S. Rasnayaka and T. Sim, “Who wants continuous authentication on mobile devices?,” in *IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS), Redondo Beach, CA, USA*, 2019, pp. 1–9, doi: 10.1109/BTAS.2018.8698599.

[150] R. Guest, M. Brockly, S. Elliott, and J. Scott, “An assessment of the usability of biometric signature systems using the human-biometric sensor interaction model,” *Int. J. Comput. Appl. Technol.*, vol. 53, no. 4, p. 336, 2016, doi: 10.1504/IJCAT.2016.076810.

Appendix

A. <u>Feedback Form</u>

1. Which phone do you own?

………

2. Have you performed a digital signature using your phone before?

………

Which body posture was comfortable while signing

- Sitting
- Treadmill
- Walking

3. Which signing process was easier to perform?
   - Signing with finger
   - Signing with stylus
4. Which factors limited the signature presentation during the experiment?
   - Small signing box
   - Thin stylus pen, had problem holding the pen
5. During typing activities, which factors limited your key entry?
   - Soft-keyboard size is smaller/bigger than my own phone
   - Screen size of the phone used in the experiment is bigger than my own phone
   - None, I was comfortable typing with this phone
6. Do you use auto-correct / dictionary in your phone always while typing?
   - Yes
   - No
   - Not always, but sometimes
7. Have you used stylus before in any mobile device?
   - Yes
   - No
8. Have you ever performed typing on the mobile phone while walking?
   - Yes
   - No

B. **Participant Information Sheet** (provided to the participants during the data collection)

Performance Assessment of touchscreen-based Behavioural Biometrics

Thank you for considering participating in this study.

You are being invited to take part in this research project. Before you decide to do so, it is important you understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with research supervisor if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. If you feel comfortable to participate after reading through this sheet, you will be asked, when you come to take part, to sign a simple consent form. Your data will be collected and then, for the purposes of our analysis, anonymously associated with a unique ID number.

Thank you for reading this.

The purpose of the study:

The aim of this study is to understand the influence of user's touch interaction on mobile device under different environmental conditions and body movements. In this experiment, we evaluate touch-dynamic based behavioural biometric modalities such as signature (using finger and stylus), swipe and keystroke dynamics. We will gather the touch data of the user on a mobile device (Galaxy Note 5) under different scenarios.

The experiment is divided into two sessions – indoor and outdoor. The indoor session includes three scenarios – using the mobile phone with a stationary body posture (sitting at a desk), walking (on a treadmill at a controlled and comfortable walking speed), walking (on a dedicated walking trail provided by the research team). The outdoor session involves three scenarios – using the mobile device whilst seated at a desk, walking and on a moving vehicle (seated in a bus). During these sessions, the participant will be asked to perform a number of tasks such as typing a sentence, swiping through images and signing using finger and stylus.

Why have I been chosen?

You have been invited to take part in this experiment as an individual aged 18 and over. Participation in any part of the collection process is entirely voluntary, and you are permitted to withdraw at any time, without giving any reason. You may also withdraw retrospectively and ask that all data relating to you is destroyed. Please be aware that once the analysis is completed it will not be possible to delete the contribution of your data in the analysis since all the data is anonymised.

We will be publishing the collected data online for academic research purposes. The release will be controlled by the research team collecting the data. You can ask for withdrawal of your data before publishing online by contacting the researcher using the contact details provided below. After completion of both the sessions of the experiment, participants will receive the stated monetary incentive.

What do you want me to do?

The experiment includes two different sessions. Each session will typically last around thirty minutes. At the beginning of the first session, the researcher will provide you with a mobile phone containing "Touchlogger" Android app and will explain the tasks you will undertake during the experiment. You will be asked to perform the experiment under two different scenarios: the first session (indoor) – answering quiz questions in the mobile phone app while seated at a desk, walking on a treadmill and walking in outdoors.

The second session (outdoors) involves participants to performing touch activities on the app whilst seated at a desk, walking on a dedicated walking trail provided by the research team and whilst seated on a bus. You will be provided with the bus timings and valid travel ticket. After each session you will be prompted to fill the feedback form on the app. Once you have completed the tasks for session two, you will return to the lab to hand over the mobile device.

We will require your contact details such as name, student ID number, phone number, address as we will be providing you with a mobile device during these experimental sessions. This personal information would not be a part of the saved data.

What will happen to the samples I provide?

Raw touch data (x and y coordinates, finger pressure, timestamp, phone orientation, tool used etc.) will be recorded along with background data (such as output gyroscope, accelerometer, etc.) on the mobile device handed to the you. The stored data will be saved in a secure server. All the anonymised data collected will be stored on a secure server linked to a reference number rather than to your name. Only the research team collecting the data will be able to link your samples with you personally, and this information will be kept strictly confidential within the research team.

What will happen to the results of the evaluations using your data?

Results of the evaluation will be documented and may be published in the scientific literature to help others benefit in the future from the knowledge we have gained. However, no participant will be identified individually, and no samples will appear in any publication or report which is published. Copies of any publication will be available via the contact point noted below.

Are there any risks involved?

There are no significant risks involved in this study. However, we strongly recommend you to be always aware of your surrounding and only to use the mobile phone while walking on the dedicated walking trail given to you at the beginning of the experimental session.

Are there any benefits in my taking part?

We hope that you will find the research interesting and that you will have the satisfaction of knowing that your contribution will help to develop the knowledge on biometric systems. Also, you will receive a £15 Amazon voucher once you completed the two sessions.

Faculty of Sciences Research Ethics Advisory Group at the University of Kent has approved this study.