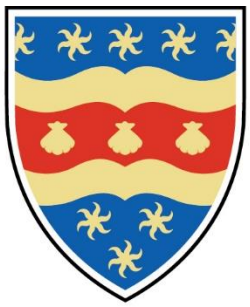
2020

Activity-Based User Authentication Using Smartwatches

AL-Naffakh, Neamah Hasan

Copyright Statement



UNIVERSITY OF PLYMOUTH

Activity-Based User Authentication Using Smartwatches

By

Neamah Hasan Al-Naffakh

A thesis submitted to the University of Plymouth in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

School of Computing, Electronics and Mathematics

July 2020

Acknowledgement

The research project was made possible due to the full scholarship funding provided by the Ministry of Higher Education and Scientific Research (Iraq) and the pleasant facility offered by the Computer Science and Mathematics College, Kufa University, Iraq. I wish to thank both organisations for their support.

I would like to express my sincere gratitude to my advisor Prof. Nathan Clarke for the continuous support of my Ph.D study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor for my Ph.D study.

Besides my supervisor, I would like to thank the rest of my thesis committee: Dr. Fudong Li, and Associate Prof. Paul Haskell-Dowland for their insightful comments and encouragement and spent significant amount of time for proof reading research papers and my thesis. Thanks also go to my fellow researcher within the CSCAN group (especially Mr. Dany Joy) for his support and interesting discussions.

I would also like to express my sincere gratitude to the University of Kufa for their financial support of this research.

Last but not the least, I owe my deepest gratitude to my mother, for her unconditional love and unwavering belief in me, my brother, and sisters for supporting me spiritually throughout writing this thesis and my life in general.

Author's Declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award without prior agreement of the Doctoral College Quality Sub-Committee.

Work submitted for this research degree at the University of Plymouth has not formed part of any other degree either at the University of Plymouth or at another establishment.

This study was financed with the aid of a studentship from the University of Kufa and carried out in collaboration with the University of Plymouth.

'Relevant seminars and conferences were regularly attended at which work was often presented and were published in the course of this research project.

Word count of main body of thesis: 52673 words

Signed ……..……
Date ……24/07/2020…..

Abstract

Activity-Based User Authentication Using Smartwatches

Neamah Hasan Al-Naffakh

Smartwatches, which contain an accelerometer and gyroscope, have recently been used to implement gait and gesture- based biometrics; however, the prior studies have long-established drawbacks. For example, data for both training and evaluation was captured from single sessions (which is not realistic and can lead to overly optimistic performance results), and in cases when the multi-day scenario was considered, the evaluation was often either done improperly or the results are very poor (i.e., greater than 20% of EER). Moreover, limited activities were considered (i.e., gait or gestures), and data captured within a controlled environment which tends to be far less realistic for real world applications. Therefore, this study remedies these past problems by training and evaluating the smartwatch-based biometric system on data from different days, using large dataset that involved the participation of 60 users, and considering different activities (i.e., normal walking (NW), fast walking (FW), typing on a PC keyboard (TypePC), playing mobile game (GameM), and texting on mobile (TypeM)). Unlike the prior art that focussed on simply laboratory controlled data, a more realistic dataset, which was captured within un-constrained environment, is used to evaluate the performance of the proposed system.

Two principal experiments were carried out focusing upon constrained and un-constrained environments. The first experiment included a comprehensive analysis of the aforementioned activities and tested under two different scenarios (i.e., same and cross day). By using all the extracted features (i.e., 88 features) and the same day evaluation, EERs of the acceleration readings were 0.15%, 0.31%, 1.43%, 1.52%, and 1.33% for the NW, FW, TypeM, TypePC, and GameM respectively. The EERs were increased to 0.93%, 3.90%, 5.69%, 6.02%, and 5.61% when the cross-day data was utilized. For comparison, a more selective set of features was used and significantly maximize the system performance under the cross day scenario, at best EERs of 0.29%, 1.31%, 2.66%, 3.83%, and 2.3% for the aforementioned activities respectively.

A realistic methodology was used in the second experiment by using data collected within unconstrained environment. A light activity detection approach was developed to divide the raw signals into gait (i.e., NW and FW) and stationary activities. Competitive results were reported with EERs of 0.60%, 0% and 3.37% for the NW, FW, and stationary activities respectively. The findings suggest that the nature of the signals captured are sufficiently discriminative to be useful in performing transparent and continuous user authentication.

Table of Contents

List of Figures

List of Tables

1 The Need of Better User Authentication for Mobile Devices

This study investigates the feasibility of a novel biometric modality that offers a flexible and robust authentication for the smartphone owners using smartwatches. Given that people use these devices to access sensitive and personal information such as online payment and Internet banking, an enhanced authentication approach that continuously and transparently protects the user's information from unauthorized access is essential. The proposed solution would not ask users to perform certain actions but to wear the smartwatch and data would be collected in the background and used to verify their identity. To this end, chapter one highlights the evolution of mobile devices and their impact in society; it begins with an overview on the rapid development regarding the embedded hardware features and significant number of applications and services that are available on these devices. Security concerns and existing user authentication approaches (including the strength and weakness of each approach) is also highlighted. Finally, the prevalence of smartwatches and vulnerabilities, and suggestion for adding a level of protection against user misuse is also discussed.

1.1 Introduction

Mobile devices have become an irreplaceable part of people's daily life. Over 7.9 billion people currently utilize mobile devices for personal communication, with these devices increasingly having access to sensitive information from financial to health-related and corporate services (Jonsson et al., 2019). Such emerging and diverse applications (apps) encourage consumers to use their mobile devices more frequently than PCs (Enge, E, 2019); smartphones are more susceptible to risk (e.g., lost, misplace or stolen) than other digital devices due to their small

size, portability and ubiquity (Tanviruzzaman and Ahamed, 2014). The use of mobile devices has inherently raised security concerns and there exists a prevalent requirement to secure these devices. Current user authentication approaches (e.g., password and PIN-based authentication) suffer from security and usability issues (as users seek to circumvent or avoid them) (Clarke and Furnell, 2007; Hocking et al., 2013). For example, research conducted by Kaspersky Lab showed that more than 50% of participants disabled their login credentials (i.e., PIN code) because of its intrusive implementation (Kaspersky, 2018). Moreover, a PIN-based authentication technique is susceptible to several types of attacks such as brute force and shoulder surfing (Kim, I. 2012). Therefore, securing information on these devices and continuously checking the user's identity in a more innovative and convenient fashion is pivotal (Al Abdulwahid et al., 2013).

The use of biometric technology in a transparent and continuous manner has been proposed in order to remove the inconvenience of authenticating the user and to improve the overall security of the device (Clarke, N. 2011). However, previous research in the domain still encounters performance caveats due to the increased reliance on behavioural biometrics and their inherent instability (i.e., external environmental factors influencing behavioural authentication approaches) (Saevanee, et al., 2012). Whilst previous research in Transparent Authentication System (TAS) (Clarke, N. 2011) has focused upon its application in computers and mobile devices, little attention has been given to the use of wearable devices – which tend to be sensor-rich highly personal technologies and finding substantial adoption among users.

Wearable computing becomes more prevalent in the market and it is predicted that the trend will continue as the technology improves. A survey showed that

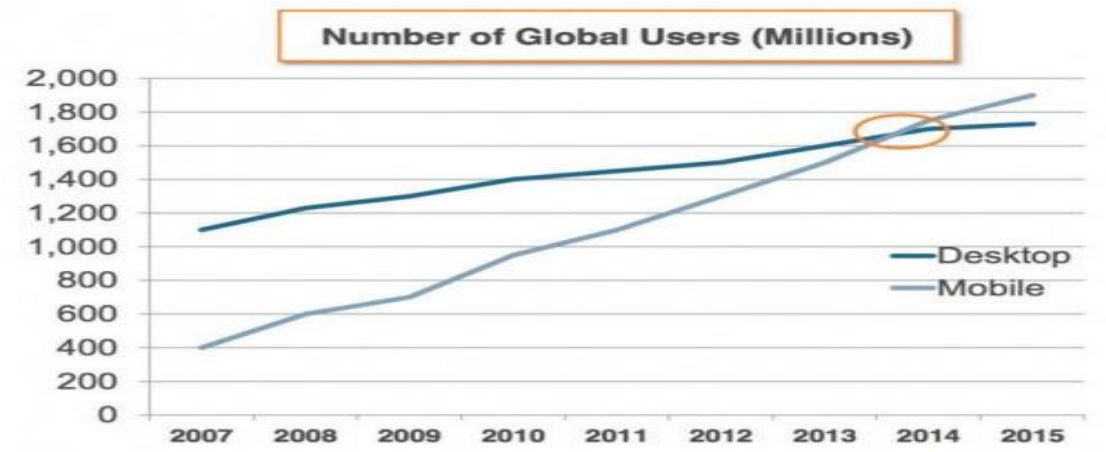
more than 80% of smartwatch consumers said that healthy living and medical care access are major benefits of wearable technology (Phaneuf, A. 2020). Due to their fixed contact with individuals (i.e., either left or right wrist), it is envisaged that smartwatches (e.g., LG and Microsoft Band 2) have the ability to capture more accurate personal data (e.g., acceleration and heart rate) than smartphones do. Therefore, wearables could be used to enhance the mobile security in a more effective way. Most modern smartwatches contain Micro Electro Mechanical System sensors, which are based upon a single chip that offers both tri-axial gyroscope and accelerometer capabilities and can be used on their own for a biometric system (Lau and Tong, 2008). Accelerometer detects acceleration, vibration, and tilt to show the speed of navigation apps and switch the phone's orientation when a user turns it, while gyroscopes provides orientation details (e.g., gyroscope determines where the phone is pointing in three dimensions).

The is a lack of modalities that serve TAS practically well, activity recognition that recognises what a user is doing at a specific point of time is a new approach that attracting an enormous amount of attention. Understanding what the user is doing (e.g., walking, running, or just lying down) can help to better adapt the user's needs; for example, activity recognition can be used in mobile health apps, and identify the user's identity in a transparent and continuous manner. Activity-based user authentication using smartwatches can offer several advantages over traditional authentication techniques. For instance, it is reliable (i.e., nearly impossible to imitate), convenient for a user (i.e., does not require explicit user interaction with a sensor during authentication), and provides transparent and continues user authentication as long the user's hand moves. This research, therefore, proposes to investigate, implement and strengthen the state-of-the-art

in transparent authentication and use wearable computing devices to secure smartphones and smartwatches.

1.2 The Prevalence of Mobile Devices

Nowadays, users are highly dependent upon mobile devices due to their portability and capability (Xu et al., 2017). With the rapid evolution of mobile devices, the sales of mobile phones and tablets dramatically increased and surpassed the PC market (Anthony, S. 2014). According to a study by BrightEdge, more than 55% of the website traffic were being undertaken using mobile devices (Greg, S. 2017) and mobiles become the most popular computing device for Internet access as shown in Figure 1.



Source (Dave, C. 2019)

Figure 1: Internet usage of mobile devices Vs desktop computers

1.3 Sensitive Storage

With the ubiquity of modern smartphones and their enormous capabilities, they now hold a huge amount of private information such as personal photos, emails, and health-related. The smartphone information is often considered more valuable than the cost of the device itself (Lifestylegroup, 2011). Moreover, users access to critical online information by using smartphones such as sending business emails as well as carry out e-commerce activities including making

payments. Email typically contains personal and sensitive user information such as financial data, bills, and business critical information. There are over 3.7 billion email users and approximately 269 billion emails were sent per day during 2017, around half of these emails were browsed using a smartphone (Templafy, 2017). Global users that use mobile devices for banking transactions is predicted to reach 1.75 billion by the end of 2019 compared to 800 million in 2014. Payments for a wide range of services (e.g., bills or online shopping) also take place by the smartphones and in the next coming years VISA/Credit cards might become less relevant. In 2017, around one billion pounds was spent in the British stores via contactless mobile payments and according to Barclaycard the value of mobile and smartwatch payments exceeded 490 billion within one year (Finextra, 2018).

1.4 Mobile Data Security Concerns

Mobile devices become an irreplaceable part of the people's daily life; similar to personal computers (PCs), these devices are also prone to security concerns such as malware. They are more susceptible to risk (e.g., lost, misplace or stolen) than other digital devices because of their portability and ubiquity (Tanviruzzaman and Ahamed, 2014) hence, increase in the vulnerability of sensitive information of these devices. A comprehensive analysis of the NHS health applications (i.e., testing 79 mobile health applications) was conducted by Imperial College London and the study found that the user's privacy can be easily breached as the developers do not use any encryption technique to secure the personal information stored in these applications (Press Association, 2015). Another study showed that news and sports were the most exposed hacked smartphone data that represented 29% of leaks followed by 19%, 11%, and 10% for business and industry services, shopping apps, and travel apps respectively (Porta, 2018). That such a high proportion of leaks (especially for business, shopping, and travel

apps) should be an alarming signal for security leaders. This is because these apps store the user's credit card information for subsequent use hence, misuse would occur when an unauthorised person access this information.

Another mobile security threat is mobile service fraud, an imposter can utilize the available services in the victim's device without paying a charge. For example, buying expensive products due to loss or theft of a mobile device from the proprietor and making international phone calls until the smartphone's user notifies the service provider. With increased use of mobile payments and mobile commerce, mobile payment fraud is on the rise.

1.5 Existing User Authentication Methods on Smartphones

Authentication is a process that verifies and confirms the user's identity. There are a variety of issues that pose a threat to mobile phones such as loss the device and mobile service fraud. Without enabling an authentication mechanism (e.g., PIN or fingerprint) to lock the smartphone, the sensitive information that are stored in the stolen or lost devices could be easily accessed by unauthorized users. Therefore, securing the mobile data in an effective and useable fashion is essential. However, current user authentication approaches on mobile devices are suffering from usability and security issues.

Password and PIN-based authentication methods have become the most popular methods due to a plethora of cost-effective implementations enabled by their low computational overhead (Xiaoyuan Suo et al., 2005; Jesudoss and Subramaniam, 2014). Although traditional passwords do not provide a sufficient level of protection, they are still deployed in many computing services such as ATM machines, Internet services, and smartphones to provide a baseline security (Raza et al., 2012). Conventional password approach is susceptible to several types of attacks such as brute force, spying and phishing, dictionary words (Kim,

I. 2012; Jesudoss and Subramaniam, 2014), smudge (Walters, R. 2012), and shoulder surfing (Luo and Yang, 2015). Other shortcomings of this technique are re-use, infrequent changing, simple to guess, written down, and hard to remember (Wiedenbeck et al., 2005; Chang, et al., 2012).

When it comes to the smartphone's safety, enabling four-digit PIN on the device is important to reduce the chance of disclosing the user's information. Although using a four digits number is not hard to memorize, many mobile users disabled the PIN security due to its intrusive implementation nature (Clarke and Furnell, 2007; Schlöglhofer and Sametinger, 2012). According to a survey that included 1,500 participants, strangely, only 3% of the smartphone's owners used a password to protect their personal data and 15% used PIN authentication, while over 50% did not utilise any authentication method to protect their devices (Bursztein, E. 2014). The reason for this is probably that typing passwords or PINs on touch keyboards is error-prone, time-consuming, and inconvenient. Even if mobile consumers employed this technique, they generally tend to select a simple password as it is easy to remember (Tanvi et al., 2011).

Pattern-based passwords authentication is based upon drawing a sequence of movements on the device touchscreen rather than entering a combination of characters; it is considered easy to remember and use (Khan et al., 2011). However, the available patterns are relatively small compared to PIN and conventional passwords due to the limited number of dots, making them more vulnerable to brute force attacking methods (Lashkari et al., 2009; Jadhao and Dole, 2013). Additionally, some of the issues faced by conventional passwords are also present in this technique such as infrequent changing and sharing with others. From a technical standpoint, Aviv et al., (2010) highlighted the fact that graphical passwords on mobile touch screens can be easily retrieved by

attackers. Given the sensitive of data contained on mobile devices, the desired level of security is arguably not being met.

The conventional mobile based authentication methods (i.e., PIN codes, Password or Patterns) are a single layer of security that can be easily guessed by an intruder or beaten through some social engineering techniques. Apple released iPhone 5s that contains a fingerprint sensor in order to offer the consumers a quick unlock of their devices and provide a better security. Thereafter, the fingerprint scanner was also included on many Android smartphones (e.g., HTC, Samsung, and Huawei). A study conducted by Roy et al., (2017) showed that there is a possibility of generating “MasterPrint” among different smartphones users. The possible explanation of this vulnerability is that the captured sample/s does not contain enough distinctive features due to the limited image size of the user's fingerprint, which is partially captured by using a built-in smartphone fingerprint sensor. Moreover, fingerprint does not provide continuous and transparent user authentication, similar to the traditional user authentication techniques.

Several new Android 4.0 smartphones (e.g., Galaxy S6 and Nexus 6P) supported the facial recognition feature as an alternative authentication solution to passcode (Aune, P. 2011). Nevertheless, Krupp et al., (2013) highlighted that users were dissatisfied in employing face recognition to unlock their mobile devices. This can be attributed to the intrusive implementation of this technique such as user cannot be authenticated in a dark room or keep a particular distance from the sensor (i.e., front camera) in order to obtain the sample (Bursztein, E. 2014; Bhagavatula et al., 2015; Krupp et al., 2013). Moreover, this approach can be easily circumvented if anyone has a good quality picture of the user’s face and poses it in front of the phone (YourSecurityResource, 2013; Moren, D. 2015). This is

caused by the lack of liveness detection implemented on the device. Recently, Apple solution enable facial recognition to securely unlock the user's device. Nevertheless, the proposed technique still suffers from security and usability issues. For example, the face sample cannot be taken if direct sunlight faces the ID camera, the face sample of kids under 13 years old does not contain sufficient biometric characteristics, and there is a possibility that identical twins can deceive the system (James, T. 2017; Leswing, K. 2017).

It is clear that relying solely on the previous authentication methods puts the user's information at risk as intruders seek to circumvent them. Therefore, a sophisticated user authentication approach that does not require explicit user interaction with the device and secure enough to defend against different types of attack is definitely needed.

1.6 The Impact of Wearable Technology in our Society

Wearable technology is a technology that is worn on the user's body and usually connected with a smartphone via Bluetooth. The wrist worn devices have several forms such as health monitoring wristbands and smartwatches that can be used for multi-purposes; for example, Internet of Things and tracking the user's health and fitness activities. About 140 million wearable devices were sold in 2017, most being smartwatches, and is expected that by 2022 the smartwatch users worldwide will be nearly 454 million (Costello, K. 2018). The smartwatch revenue is predicted to exceed 50 billion by 2022 (Steve, S. 2018).

As wearable devices including several embedding sensors (see Figure 2), these devices are capable of capturing various personal based biometrics data such as 3-axis accelerometer, 3-axis gyroscope, temperature, and heart rate. Due to the fixed contact of wearables with individuals (i.e., either left or right wrist), it is envisaged that smartwatches (e.g., LG and Microsoft Band 2) have the ability to

capture more accurate personal data than smartphones do. For example, heart rate, acceleration and gyroscope data could be effectively collected by using smartwatches as data was collected with minimal effort (i.e., data can be collected in a transparent and continuous manner). Therefore, wearables offer the opportunity to get more reliable biometric measurement that can effectively use for authentication-based and activity recognition systems. Figure 2 illustrates that smartwatches are sensor-rich devices that enable a wide variety of personal biometric-based information, potentially more accurately than what can be captured by smartphones.

Figure 2: The existing sensors in smartwatches

In terms of security, the smartwatch needs to be protected and secured just like other computing technologies. The internet security report of Norton highlighted that smartwatch users store sensitive business documents and bank account details on their device (Steve, S. 2018). The sensitive information stored in smartwatches attracting an enormous amount of hacker attention to breach the

user's privacy and security. Given that the smartwatches are usually connected with smartphones, the lack of security means a serious risk of attack on both devices. A comprehensive analysis of 10 smartwatches trademarks (e.g., Apple, and LG watch R) was conducted by Lemos, R. (2019) in order to explore the security concerns of these devices. The study indicated that seven smartwatches do not use any encryption technique for the installed apps, and three watches had vulnerabilities that permit information being misused by unauthorized users. Although the people's life could be much more convenient by using smartwatches (i.e., they can be used for opening door via Near Field Communication, start cars, or paying bills), there is a high risk to misuse the services and the sensitive information stored on the device when the device is lost or stolen. Bluetooth and public WiFi are other pitfalls that can be misused; for instance, if the WIFI traffic is not encrypted, hackers, who are connected to the same network, could access the user's data (Ricci, et al., 2016).

Current user authentication approach on smartwatches (i.e., PIN security) suffers from many issues such as being easy to guess and difficult to enter due to the small screen size of the smartwatch (Junshuang Yang, et al., 2015a). Smartwatches are also vulnerable to different types of attack; for example, brute force attack on Bluetooth passcodes (Karakaya, et al., 2016) or hackers could guess what a smartwatch user is typing through disclosure the motion data produced by the accelerometer or gyroscope sensors of the smartwatch (Kaspersky, 2018 ; Winder, D. 2015; Liszewski, A. 2015). Therefore, such devices need improved mechanisms of user authentication to secure the aforementioned information and continuously check the user's identity in a transparent fashion.

1.7 Aims and objectives of the research

The aim of this study is to explore, propose and evaluate a new biometric modality for smartphones (i.e., an activity-based biometric authentication technique using the smartwatch acceleration and gyroscope data). Such a system would enhance the overall security and offer continuous and transparent user authentication for smartphones and smartwatches users. To achieve this, this research is divided into five distinct stages:

- To evaluate the existing user authentication approaches and the highlight the need for better user authentication techniques.
- To investigate the potential behavioural biometric modalities and their applicability to deploy for smartphones and smartwatches devices, with the aim of increasing the transparent authentication capability available to the device.
- To design and conduct several experiments that provide a robust and reliable authentication system.

1.8 Thesis structure

This thesis describes the research leading to the formulation of a suitable security strategy for smartphone and smartwatch devices. Chapter 2 provides an overview of the current user authentication approaches by highlighting the key issues associated with each approach. It starts by reviewing the popular authentication methods (i.e., secret knowledge-based and token-based authentication approaches). This is then followed by an overview of a generic biometric system, biometrics performance metrics factors, and details of the physiological and behavioural biometric techniques (specifically, techniques that are applicable in the concept of TAS for digital devices.

The state of the art of transparent and continuous biometric-based studies, in particular gait recognition based on accelerometer and gyroscope sensors, is presented in chapter 3. The chapter also provides thoroughly analysis of the key parameters that influence the system performance. For example, scenarios that were used to collect the user's movement data, the extracted feature subset and the process that was used to select the discriminative and unique feature information, and reliable classifier/s that offers high authentication accuracy.

Chapter 4 presents the feasibility of deploying activity-based user authentication using smartwatch. This is achieved via carried out several experimental based upon collecting five different activities (i.e., normal walking, fast walking, typing on a PC keyboard, playing mobile game, and texting on mobile touch screen). These activities were collected under a controlled environment to explore whether the technology is sufficiently capable and the nature of the signals captured sufficiently discriminative to be useful in performing TAS.

To ensure that the proposed technique can be used for real world authentication-based systems, chapter 5 introduces a more realistic experiment by collecting real life data (i.e., uncontrolled data) and evaluates the system performance under unconstrained environment. Several experiments were carried out to provide a robust and reliable authentication system.

Chapter 6 addresses a number of further research questions surrounding the viability of the approach via conducting scientifically valid experiments. It begins by determining the optimal sample size and the amount of data required for the training and validation phase; subsequently, using the majority voting schema in order to improve the classification results.

Finally, Chapter 7 presents the main conclusions from the research, highlighting the key achievements and limitations. The chapter also discuss on the future research and development.

1.9 Conclusion

Billions of mobile devices are being globally used having multiple applications in e-commerce, browsing as well as for storing personal data. The use of smartphones for several purposes (e.g., sending and receiving Emails, and online banking) has inherently raised security concerns and there exists a prevalent requirement to protect these devices. Current conventional technologies for providing device security such as password and PIN, however, fall short of addressing these security concerns due to lack of technical sophistication or simply because of their intrusive implementation. Apart from that password or PIN based authentication approaches demand a high level of memorability from the user to be authenticated (specifically, if the user uses a unique password for each account), these methods fall short of addressing the security concerns due to lack of technical sophistication. Moreover, the intrusive implementation of password or PIN techniques considerably increased the authentication burden and resulted in smartphone users to take no security precautions against unauthorized access.

Although big smartphone brands (e.g., Apple and Samsung) enabled fingerprint and facial based user authentication in order to improve the security level and take the burden away of entering a password or a four-digit PIN, these techniques are one time authentication (i.e., do not continuously verify the user's identity). Apple subscribers might force to retrieve a backup device if their Face ID is not recognized hence, this presents obstacle for applying the aforementioned biometric modality (Palmer, D. 2017). It is therefore, imperative to find a new mechanism while striking a right balance between robust security and ease of

use. Behavioural biometrics technologies have the potential to offer transparent and continuous mode of user authentication promising a greater degree of usability while incorporating resilient security. The next chapter reviews state of the art in user device authentication including conventional knowledge and token based as well as physiological and the recently emerging behavioural biometric techniques.

2 Review of Biometric-Based User Authentication

There are three primary approaches for implementing user authentication: secret knowledge, token, and biometric- based. Details of the he first two approaches were already discussed in the previous chapter. An overview of a user-friendly techniques (i.e., biometric-based user authentication) and the metrics that are frequently used to evaluate the system performance would be presented in this chapter. A brief explanation of various biometric techniques and focuses on the approaches for achieving transparent authentication is also discussed.

2.1 Introduction

The idea of providing security credentials (i.e., a username and password or PIN) before gaining access to a particular service or an account is generally accepted by users (Chiasson and Biddle, 2007; Hocking et al., 2013). However, conventional authentication mechanisms in computing system could be circumvented if not correctly implemented (Al Abdulwahid et al., 2013). Various methods to verify the legitimate user have existed, and each one supplies different levels of security. A thorough review of secret knowledge and token-based user authentication (including their strengths and weaknesses) has presented in the previous chapter, as a result, it is believed that biometrics still offer the greatest potential to solve the security and usability issues for smartphones.

2.2 Biometric -based authentication

Biometrics are used to differentiate between users based on their physiological or behavioural characteristics (e.g., how they look (face) and how they walk (gait)). It is argued that biometrics offer the potential to be the most effective approach to verify the presence of the genuine user not the presence of a device

(i.e., token) nor a pre-set information (i.e., secret). This can be achieved based upon unique features that cannot simply loss and nearly impossible to share (Singh and Singh, 2013).

The traditional user authentication approaches (i.e., KBA and token) cannot prove if the login credentials have been provided by a legitimate user or an imposter. Reason for this is that the aforementioned methods are relying on “what the user has” or “what the user knows”, whereas biometrics techniques are capable of verifying the user’s identity based upon such physical and behavioural characteristics that are linked to a specific user (Kulkarni and Namboodiri, 2014; Jain et al., 2004). Biometrics are considered a user-friendly approach that does not require remembering a password or carrying multiple tokens; it is in the possession of the user all the time (Karnan et al., 2011). Nevertheless, such system also has some drawbacks; For instance, if a hacker obtains access to the user’s biometric samples, it would be difficult to replace or revoke data because of the limited physical features available per user (Ratha et al.,2001). Moreover, biometric-based systems might incorrectly accept unauthorized users or reject a legitimate user. This is because the system decision is based upon measuring the similarity between the reference and test samples rather than exact match between two alphanumeric strings (as in KBA and token-based authentication).

2.2.1 An Introduction of the Biometric System

The modern definition of biometrics by International Biometric Group is:

> “*the automated use of physiological or behavioral characteristics to determine or verify identity”.*
>
> (IBG, 2010).

Biometric systems could utilize two modes: identification or authentication (verification); below is the description of each mode (Smith, R. 2002; Mayhew, S. 2012).

- **Authentication:** is the process of verifying the identity of a user who claims to be. When the user provides a sample (e.g., fingerprint), the system tries to find a match between the presented sample (test sample) and the stored template of that user (reference sample) thus, it is a one-to-one comparison. If the reference and test samples are matched, the access is granted; otherwise, the access is denied.
- **Identification:** is the process of identifying the identity of a person (who is this person?) rather than validating the claimed identity. The system tries to find whether there is a match between anonymous sample and all the reference templates in the database (one-to-many comparison). As a result, the identification mode requires more time than the authentication mode to generate a result. Identification is typically used for surveillance in the airport and in the criminal investigations. Therefore, the feature extraction process in the identification system should be more sophisticated than in an authentication system.

2.2.2 Components of a Biometrics System

A typical biometric system consists of five main components as shown in Figure3 (Li, F. 2012).

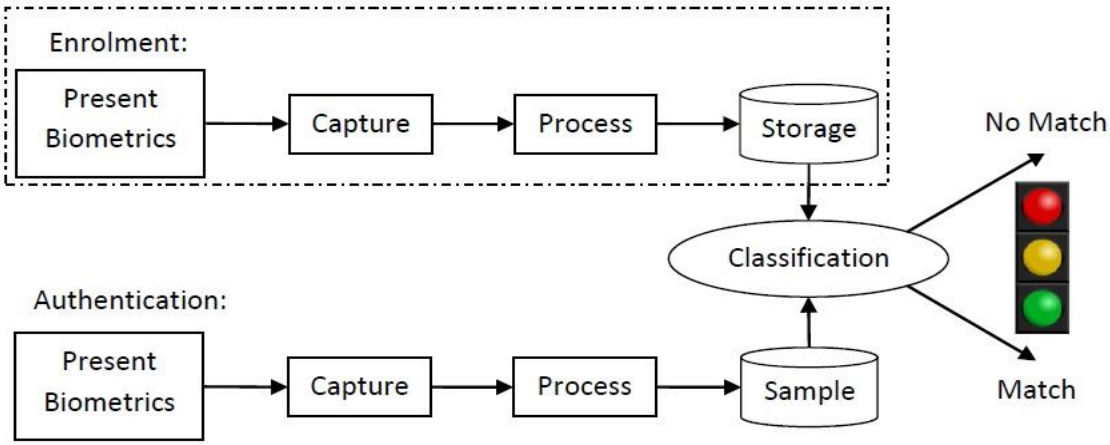
- Sample acquisition: collecting a biometric sample/s from a user using an equipped sensor in the computing devices (e.g., mobile camera, which can be utilized in facial recognition) or specialised sensors (e.g., fingerprint reader).

- **Feature extraction**: from the collected sample(s), distinctive characteristics are extracted to construct the reference template.

- **Storage**: the reference template, which has resulted from the feature extraction phase, would be stored in the database. This template is used for comparison in the verification phase subsequently.

- **Classification**: during the classification phase, a matching algorithm is applied to compare between the new biometric template (i.e., probe or test template) and the reference template; accordingly, a similarity score is generated.

- **Decision**: the process of accepting or rejecting a user is based on the comparison between the computed similarity score and the threshold value. If the similarity score meets or is above the threshold value, the user will be granted access to the system; otherwise he/she will be declined.



Source (Li, F. 2012)

Figure 3: The components of a biometrics system

Generally, a biometric system has two main processes, as illustrated in Figure 4 (Sui et al., 2011). Below is the description for each phase:

- **Enrolment phase:** enrolment refers to the stage in which a biometric system extracts a set of features from the user's biometric samples. These features

are subsequently used to generate the reference template of that user. The process of generating the user's reference template should be accurate. From that perspective, the high quality biometric samples must be collected; otherwise, the user is asked to provide the sample again.

- **Verification phase:** in this phase, the system captures a biometric sample from a user, extracts features that would be used to create the probe template, and finally compares that template against the stored template for authentication. If the matching score meets or exceeds the pre-set threshold value, the user will be granted access to the system; otherwise, they will be refused. In general, the accuracy of the biometric system is based upon the selection of threshold value; lack of it makes the system vulnerable to penetration or wrongly rejecting a legitimate user.

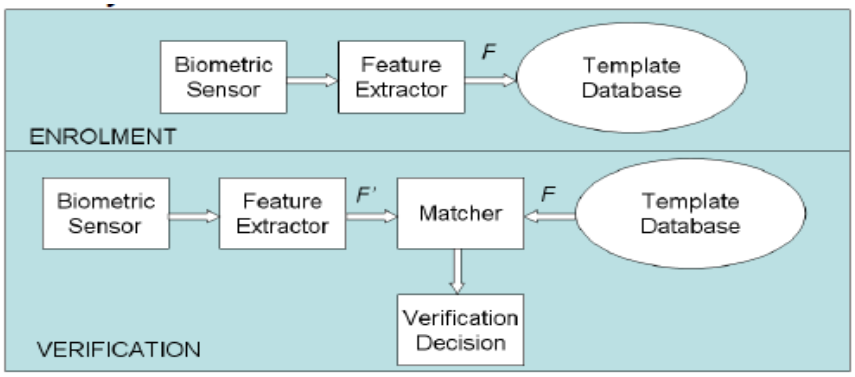


Figure 4: Conventional biometric authentication

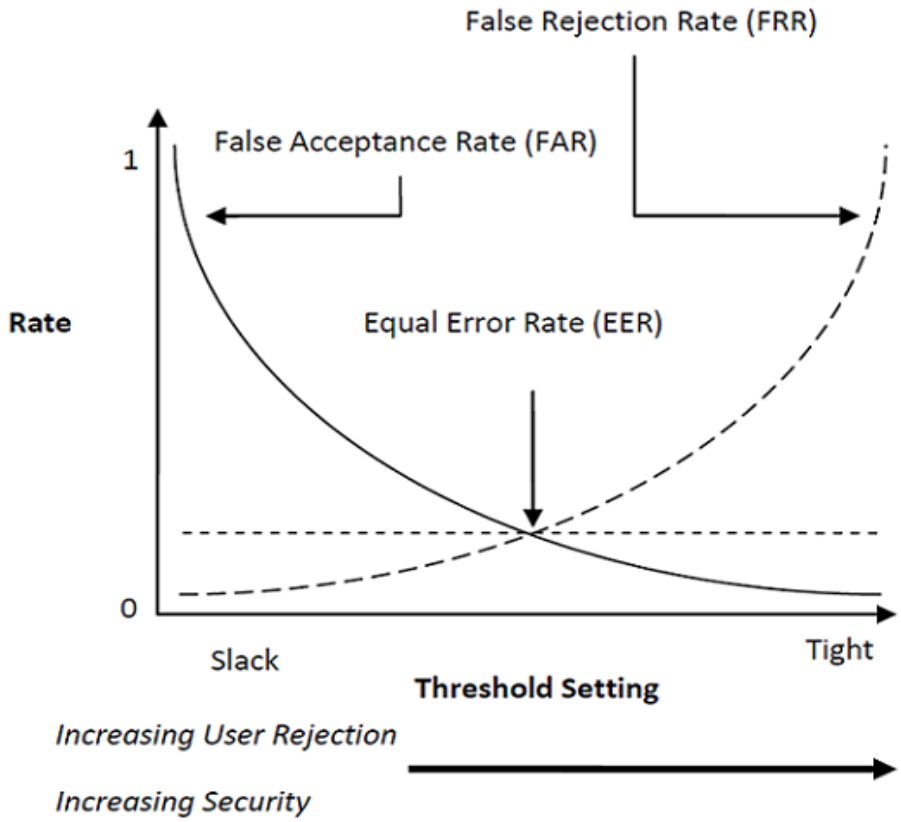
2.2.3 Biometrics Performance Metrics Factors

Having stated that the biometric-based user authentication tends to be more convenient to the users, these systems are susceptible to two basic types of failures: a false acceptance rate (FAR) and a false rejection rate (FRR). The former shows the percentage in which the system incorrectly accepts an imposter as the legitimate user. The latter displays the percentage in which the authorized user is wrongly rejected by the system (Jain et al., 2002). Error rates FAR and FRR are calculated as:

$$FAR = \frac{\text{N accepted impostors}}{\text{total N impostors}}$$

$$FRR = \frac{\text{N rejected genuines}}{\text{total N genuines}}$$

In general, these types of errors, FRR and FAR, result from a variety of issues such as environmental noise and trait variability. The resulting values of both metrics are based on pre-set a threshold value for the biometric system. Figure 5 shows that the two metrics are inversely proportional. Therefore, setting a high threshold value reduces the probability of accepting an imposter by the system (i.e., low FAR), and it may result in high refusal of a legitimate user (i.e., high FRR). Subsequently, genuine users might feel discomfort from repeated authentication failures. So, it is important to take into consideration during design an authentication system to have a balance between security (i.e. FAR) and usability (i.e., FRR).



Source (Clarke and Furnell 2005)

Figure 5: Biometrics performance metrics factors

In addition to the FAR and FRR metrics, the equal error rate (EER) is also widely used to evaluate the performance of authentication systems. The EER is calculated by taking the average of the FAR and FRR. As illustrated in Figure 5, EER represents the intersection point between the FAR and the FRR curves, i.e., FAR equals FRR (Gamassi et al., 2004). There are other performance statistics to evaluate the biometric systems such as failure to enrol rate and failure to acquire rate. The former refers to the error rate that occurs during the enrolment phase. It is typically occurred when the extracted features are not sufficient to form the reference template. The latter is resulted when the system is incapable to capture the user's sample(s) due to a technical failure.

2.2.4 Biometrics System Characteristics

In order to employ a biometric technique for an authentication-based system, there are a number of standard criteria need to be considered (Jain et al., 2004), these include:

- **Universality**: the selected biometric trait/modality should be feasible in every individual; for instance, implementing fingerprint-based system requires each person to have fingers.
- **Uniqueness:** the biometric technique needs to be sufficiently discriminative in order to differentiate between individuals. For example, the user's retina is more distinctive (i.e., unique) than the facial recognition.
- **Permanence:** the biometric characteristics should be stable over time, otherwise, the user would be asked to enroll frequently to the system. For example, unlike gait technique, people's iris contains discriminatory information that would not be affected by mood and age.

- **Collectability:** the process of collecting the biometric samples should be simple and cost-effective (i.e., using embedded or suitable sensor). For example, hand geometry system requires a specialized scanner to obtain the sample (i.e., the sensor size is big and not suitable for kids due to the physiological change of their hand shape over time. In contrast, gait data can be collected in a transparent and continuous manner using the smartphone built-in sensor.

To ensure the biometric-based system is acceptable and can be used in real life scenario, the following criteria should be considered:

- **Performance:** the proposed system should achieve a high recognition rate, speed, and robustness.
- **Acceptability**: an indication whether the end-user is comfortable to use the proposed system (i.e., biometric-based user authentication). For instance, people would prefer to provide a facial scan rather than retina sample as the latter technique is more intrusive.
- **Circumvention**: the authentication system should be sufficiently secure and reliable to defend against different types of attack.

2.2.5 Biometrics Techniques

In general, biometric techniques are classified into two main categories: physiological and behavioural (Nanavati et al., 2002). The former aims to authenticate/identify users based upon their physical characteristics such as face and fingerprint (Wayman et al., 2005). The latter differentiates individuals through utilising unique behavioural feature set such as walk pattern and typing on a keyboard (Woodward, et al., 2003). Given that the physiological biometric characteristics of an individual are nearly stable over time and more resistant to different conditions (e.g., age, body fitness, and mood), they tend to be more

reliable techniques. In addition, the physical features contain high levels of distinguished information (Woodward, et al., 2003).

In contrast, behavioural characteristics tend to be less unique and stable due to the change in mood, health, and environment. However, most of the behavioural biometrics systems are unobtrusive (i.e., do not require explicit interaction from a user) and hence more user-friendly than their physiological counterparts. Table 1 shows the applicability of the physiological and behavioural biometric techniques in smartwatches and highlights their characteristics such as uniqueness, collectability, performance, and acceptability.

	Biometrics approaches	**Universality**	**Uniqueness**	**Permanence**	**Collectability**	**Performance**	**Acceptability**	**Applicability**
Physiological	Ear recognition	Medium	Medium	High	Medium	Medium	High	No
	Face recognition	High	Low	Medium	High	Low	High	No
	Fingerprint recognition	Medium	High	High	Medium	High	Medium	No
	Iris recognition	High	High	High	Medium	High	Low	No
	Retina recognition	High	High	Medium	Low	High	Low	No
Behavioral	Gait	Medium	Low	Low	High	Medium	High	Yes
	Voice verification	Medium	Low	Low	Medium	Low	High	Yes
	Behavioral profiling	Medium	Low	Low	High	Low	High	Yes
	Keystroke dynamics	Low	Low	Low	Medium	Low	Medium	No
	Signature recognition	Low	Low	Low	High	Low	High	No

Table 1: A brief comparison of biometrics approaches

It is shown from Table 1 that none of the physiological biometric approaches are applicable or can be collected from smartwatches to offer a transparent user authentication for smartphones due to the unavailability of data within these devices (i.e., smartwatches). In contrast, the equipped smartwatch sensors (e.g., heart rate, skin temperature, acceleration and gyroscope) enable the collection of a wide variety of behavioural biometric-based information. Based on the presented characteristics in Table 1, none of the behavioural-based techniques outperforms any of the other approaches. Nevertheless, for instance, the success of speaker recognition technique depends completely upon the quality of the input

samples (probe voice samples) that are more likely to be different from the reference samples, which have been collected in a controlled environment. Therefore, this could lead to reducing the performance significantly (Rajasri, et al., 2013). Moreover, some factors such as age, alcohol consumption, emotional state, and health conditions can change the pattern of the person's voice (Sonkamble et al., 2010). On the other hand, gait recognition tends to have a very high acceptability because it is easy to acquire. In addition, the usage of motion sensors (accelerometer and gyroscope) is not limited to capture only gait information but can be extended to collect a wide range of activities such as typing on a PC or mobile keyboard, playing game, gesture, and stationary activities.

Gait Recognition

Gait recognition is a technique that identifies or verifies people using their walk patterns as each individual has a distinctive walk (Arora, P. 2015). It is an unobtrusive mechanism (convenient for a user) that does not require explicit user interaction with a sensor during authentication or identification phase (Derawi et al., 2010a). Recently, researchers showed an increased interest on mobile gait authentication, and performance rates were vary considerably relying on the feature extraction methods and the types of classifiers. In general, the reported EERs ranged from 5% to 19% when training and testing data are collected within the same day (Gafurov, D. 2007a) and in the range of 10% to 33% when multiple days data were used to create the reference and test templates. Although the human gait is visible to monitor, the literature showed that a user's walking style is nearly impossible to imitate (Gafurov et al., 2006a; Gafurov and Snekkenes, 2009; Zhang et al., 2011). Gait is an attractive and cost-effective technique, especially when modern mobile phones and/or smartwatches can be utilized to capture the user's gait data. Moreover, it can be used to provide continuous and transparent user authentication as long as a user is walking. Nevertheless, this

technique relatively suffers from issues such as a shoe type, ground condition, carrying a load, and permanence (Gafurov et al., 2010; Muaaz and Nickel, 2012).

2.3 Background Knowledge on Gait Recognition

Having completed the review of activity recognition literature, there is no research to date has been performed; given the nature of the wearable computing and the sensors, gait recognition using mobile devices is the modality that has the closest link to activity-based user authentication and has been thoroughly explored with only few studies perhaps included the use of smartwatches. In general, gait recognition can be categorized into three main approaches, machine vision based, wearable sensor based, and mobile sensor based. Description of each approach is explained below:

- **Machine Vision based**: in machine vision, the movement of the whole body is captured from a distance using a video-camera (as shown in Figure 6). Thereafter, image/video processing methods are applied in order to extract some unique characteristics such as height and distance between feet (Arora and Gandhi, 2014). It is often utilized for identification purposes such as airport security.

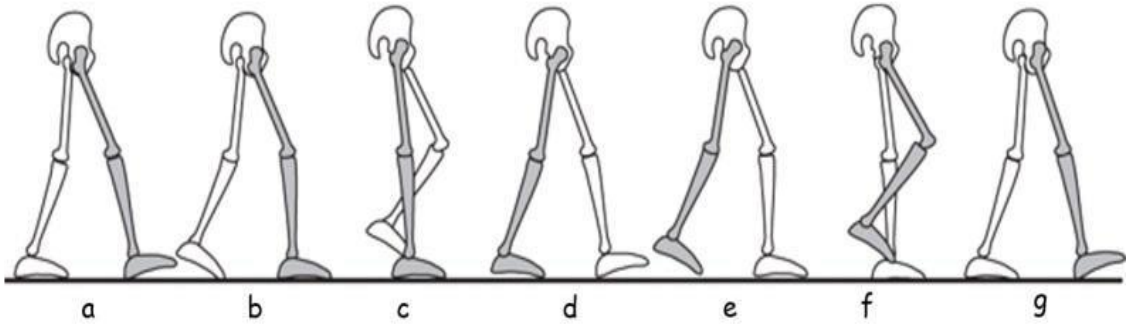
(a) Original image (b) Silhouette

Source (Gafurov, D. 2007a)

Figure 6: An example of machine vision approach

- **Wearable Sensor based:** in this approach, the periodic motion of the legs is captured (see Figure 7) by attaching a wearable recording sensor(s) to different positions around the human body (see Figure 8) such as hip, waist,

pockets, lower leg, and arm (Gafurov and Snekkkenes, 2008; Ngo et al., 2014). The raw time-series accelerometer data of three directions (i.e., x, y, and z) is then segmented into cycles or windows in order to extract discriminative gait information such as average cycle, standard deviation, and the Bark frequency cepstral coefficients (BFCC).



Source (Hoang et al., 013)

Figure 7: Illustration of periodic motion of the legs

Source (Gafurov, D. 2007a)

Figure 8: Different locations of attached wearable sensor

- **Mobile Sensor based:** the third gait approach is mobile sensor based that attempts to utilize the smartphone sensors (i.e., accelerometer and gyroscope) for collecting the gait data. It is cost effective and provides transparent and continuous user authentication (Derawi et al., 2010a). Smartphones, while having the benefit of technological maturity and widespread adoption, suffer from several problems to produce a consistently effective implementation. For example, a survey by Ichikawa et al., (2005) showed that users tend to put their phone in numerous locations around their body wherever there is a pocket

(e.g., inside coat pocket and back pocket). Moreover, the study highlighted that girls mostly keep their phone inside shoulder bags while males put their phone in several locations such as trouser pockets, upper-body pockets and inside a pouch attached to their hip. Therefore, this could make the data collection process less accurate or nearly impossible.

Fundamentally, the majority of the studies applied one of the following two methods to chunk the walking signals, 1) cycle based and 2) segment based. A brief description of each method is described below:

- **Cycle-based Method**

  Cycle-based method can be considered as the most common approach used in gait recognition. Predominantly, studies attempted to detect the periodic steps of the individuals. Cycle-detection methods aim to be invariant to pace by standardizing the number of steps as opposed to the amount of time represented in each instance (Derawi, M. 2010b). In order to extract gait cycles from acceleration signals, two different approaches are often utilized, namely local minima and the salience vector. The former is based upon identifying the initial start of each cycle in the gait signal. After all minima are located, the data points between two consecutive minima are considered as one cycle (Gafurov, et al., 2007a).In the latter approach, cycles are detected by identifying minima and maxima salience vectors. The benefit of detecting the local maxima is to determine the exact start point of each cycle as it typically represents the actual walking pattern (Nickel et al., 2011d).

  According to (Derawi, M. 2010b; Nickel et al., 2011d; Muaaz and Nickel, 2012; Muaaz and Mayrhofer, 2014), the detected cycles in the acceleration signals require further analysis. This has been carried out usually by using

a distance function (e.g., dynamic time warping (DTW) or Manhattan) to remove unusual cycles that are significantly different than other cycles (i.e., high distance to other cycles). Subsequently, the regular cycles are averaged in order to construct the user's reference template, which is further used for comparison against the input template. Finally, the standard classification methods (e.g., Absolute, Euclidean, and DTW distance metrics) were used to recognize the user's walking pattern. This can be achieved by calculating the distances of two feature vectors (i.e., reference and probe templates) through applying one of the aforementioned distance functions in order to obtain a decision. Ideally, distance scores obtained from the user's samples should be as small as possible, an indication that the reference and probe samples have been taken from the same person. Also, when samples of other persons are tested against the user's template, distance scores should be as big as possible, indicating that they were obtained from different persons (Gafurov, et al., 2006b). Below is a brief description of the most common classification algorithms used in cycle extraction approach (Derawi, M. 2012):

Absolute Distance

The Absolute distance is a metric that measures the sum of the absolute values of the differences between all the reference and test samples. However, it requires that the reference and probe templates have the same length as illustrated Equation 1.

$$d_{abs.}(X,Y) = \sum_{i=1}^{k} |x_i - y_i| \quad \text{Equation 1}$$

Euclidean Distance

The Euclidean distance, as shown in Equation2, can be considered as a special case of absolute distance. It measures the square root of the sum of the differences between all the values in the stored template and the corresponding values in the test template.

$$d_{eucl.}(X,Y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad \text{Equation 2}$$

DTW Distance

This algorithm is unlike Absolute and Euclidean distance metrics, it can calculate the optimal distance between two given feature vectors even if the length of these vectors are not equal. The DTW distance function is less sensitive with the variations of the detected cycle features.

In conclusion, the challenge with the cycle extraction method is to find a mechanism for identifying the start and end point of each cycle. Moreover, cycles are not guaranteed to be of the same length (and can vary widely in length depending on the pace of a user) thus, the system does not perform well or fails for unusual (i.e. both slow and fast) paces. This method also requires complicated computations that seem less feasible to implement on the mobile phones due to the limited processing resources in these devices.

- **Segment-based Method:** this method is simplest and easy to implement as the raw motion data are directly divided into fixed size windows (e.g., 5 or 10 seconds) and then extracting set of gait features based upon the acceleration readings in the window (Kwapisz et al., 2010). In general, the calculated gait features from these windows can be categorized into two types: statistical and cepstral coefficient. Although the statistical features

(e.g., standard deviation, and root mean squared) do not require complex measurement (easy to generate), they perform high level of accuracy. These features can be computed for single axis (e.g., vertical, horizontal, and lateral directions) or with the fusion of three axes sensor. Similarly, the cepstral coefficient features, which have been successfully implemented in speaker recognition, can be used alone and still provide extremely strong accuracies, specifically the Bark frequency cepstral coefficients (BFCC) and Mel-frequency cepstral coefficients (MFCC), "which belong to the most widely used spectral representations of audio signals for automatic speech recognition and speaker verification" (Subramanian, H. 2004). Some studies have been successfully used by combination of both features (i.e., statistical and cepstral coefficient) in order to construct more sophisticated feature vectors (Nickel et al., 2011b; Hestbek et al., 2012). Typically, supervised machine learning algorithms, such as Support Vector Machine (SVM), Hidden Markov Model (HMM) and Neural Network, were used to classify the segment-based features. For a given input gait data (training data), the task of supervised learning method is to find a generalized function (e.g., for a given data points of x, y values will be generated). The output of this function is used to predict a class label for each user, which is further used for comparison.

Different machine learning algorithms were utilized in mobile-based gait authentication studies, and the system accuracy was fairly acceptable. Nickel et al., (2011a) and Nickel et al., (2011b) used HMM and SVM respectively to classify the user's gait pattern with EERs of 10.4% and 6.3%. Another research by Kwapisz et al., (2010) applied two learning methods, J48 decision trees and neural networks and reported 85.9% and

95% positive and negative authentication rates respectively. The former (i.e., positive authentication rate) is a rate in which a user is successfully recognized while the latter (negative authentication rate) is a rate in which an imposter is correctly rejected. The classification was performed using data mining suite tool, WEKA. The first attempt to use smartwatches for gait recognition was by Johnston and Weiss (2015). The authors presented a comprehensive test by applying four WEKA approaches, namely Multilayer Perceptron (MLP), Random Forest, Rotation Forest, and Naive Bayes, in order to find the best classification method for gait authentication and identification. There was no clear pattern with respect to which algorithm performs best as the reported EERs were nearly similar.

2.4 Conclusion

Mobile computing and smartwatches have significant security concerns as any other technology. As previously discussed, current user authentication approaches (e.g., secret knowledge and token-based authentication) suffer from usability and security issues. The literature has highlighted that the implementation of these techniques is intrusive (relying upon users to remember something). Also, security issues can result from several factors such as lost or stolen token, using a simple password, and re-using the same password on several websites. Due to these weaknesses, further attention was placed upon using biometrics as they can provide reliable and convenient user authentication and do not require users to carry or remember anything.

With the rapid evolution of smartphones and smartwatches, which tend to be sensor-rich highly personal technologies, a number of biometric techniques can be implemented on these devices such as gait and voice verification. However, it is important to identify appropriate biometric technology that provides a balance

between security and usability and does not require complicated computations. The equipped motion sensors (i.e., accelerometer and gyroscope) on smartphones and smartwatches can be utilized to collect the data transparently (without explicit interaction from the user with the sensor) hence, could be useful to design an effective transparent and continuous user authentication system to secure the both devices in one go.

3 Current State of the Art in Motion-based Biometric Authentication

3.1 Introduction

The present chapter reviews state of the art in transparent and continuous authentication using acceleration and gyroscope sensors technologies. The main sections of the chapter are as follows. Section 3.2 details literature review methodology, gait-based authentication using specialized sensors are highlighted in section 3.3 and mobile gait-based authentication is reviewed in section 3.4; section 3.5 includes the application of these to wearable devices (smartwatches). Final discussion is presented in section 3.6 and conclusions are drawn in section 3.7.

3.2 Review Methodology

A comprehensive overview of the technical and academic disciplines is provided in this chapter. In order to comprehensively review the prior work in the area and identify the limitations of the existing methodologies, the following predefined research questions were highlighted:

- What is the aim of the paper?
- How many samples were collected from each participant and which data collection scenario was applied (i.e., singles or cross day scenario)?
- How many participants were involved in the experiment?
- What feature extraction methods were implemented and what types of features were extracted?
- How well classification methods were performed?
- What are the outstanding questions that were not covered by the literature?

Once the research questions were highlighted; the next task is to identify the search keywords that help to find the most related articles. To this end, finding studies focusing upon sensor-based gait/activity recognition were essential to build concrete background on the state of the art. This step is described by the following:

- Finding all the relevant semantic synonyms and hyponyms to ensure that all papers with a similar problem definition are retrieved (i.e., gait recognition, sensor-based authentication, activity recognition using wearable computing, wearable sensor-based authentication, mobile accelerometers, classifying accelerometer, different approaches to gait recognition, wave to access mobile devices, motion behavior, smartwatch-based authentication, and activity recognition).

- Formation of abstraction and conclusion, at the same time, neglecting any papers that do not meet the predefined search criteria.

In order to find all relevant studies, multiple well-known academic sources (i.e., journals and conference proceeding) and academic online research repositories were explored. Epistemologically, formal websites were used such as the electronic databases of IEEE, ACM, SpringerLink, and Google Scholar. Finally, to ensure the search process yields the best candidate papers, a list of requirements and standards for the selected papers is created, these involve:

- Reviewing a state of the art was focused on the key papers that have been published by the leading scientists in the field of activity and gait recognition.

- Papers that were published since 2005 and forwards were selected due the limited amount of research outputs in gait recognition technology. Furthermore, papers focusing on machine vision or floor sensor-based gait recognition were

excluded as the data collection methodology of this research reflects mobile-based sensor capture employed in this domain is completely different from the proposed system of this research.

- Excluding brief papers that do not include an experimental evaluation.

Database	Number of papers	Final Selected papers	Quality evaluation		
			Conference	Book	Journal
IEEE Xplore	54	28	19	-	9
SpringerLink	7	5	-	5	-
ACM	13	7	3	-	4
Google Scholar	34	11	5	-	6
Total	108	51	27	5	19

Table 2: The total studies and the selected articles with quality evaluation

3.3 Gait Authentication using Attached Sensors

Reviewing of papers was focused on gait authentication using attached sensors, 11 relevant articles were identified and summarized in Table (details of each individual study is presented in Appendix A). The first attempt in wearable- based gait recognition was wearing a dedicated sensor to collect the motion data (dedicated means that the sensor is not a part of mobile/smartwatch devices and physically attached to the user). A variety of studies have been performed in this domain by attaching a recording device to different positions around the human body (i.e., hip, waist, pockets, lower leg, and arm).

	Authors	Year	Type	Citation
1	Mäntyjärvi et al.	2005	Conference	228
2	Gafurov et al.	2006a	Conference	39
3	Gafurov et al.	2006b	Conference	166
4	Okumura et al.	2006	Conference	64
5	Gafurov et al.	2007a	Conference	64
6	Gafurov et al.	2007b	Journal	80
7	Gafurov and Snekkenes	2008a	Conference	16
8	Gafurov and Snekkenes	2008b	Conference	21
9	Gafurov et al.	2010	Conference	30
10	Sangil Choi et al.	2014	Conference	5
11	Cola et al.	2016	Conference	9

Table 3: An overview of the selected gait-based studies using dedicated sensors

Studies by (Gafurov et al., 2006a; Okumura et al. 2006; Gafurov, et al., 2007a; Gafurov et al., 2007b; Gafurov and Snekkkenes, 2008a; Gafurov and Snekkenes 2008; Gafurov et al., 2010 ; Sangil Choi et al., 2014; Cola et al., 2016) were mainly focused on merely gait activities (i.e., normal or fast walking) and utilized the cycle based approach that requires a complex computational processing to detect each cycle from the acceleration signal. Moreover, these studies used the traditional algorithms such as dynamic time warping (DTW) and absolute distance that are not effective for behavioural -based biometric system due to the changes of the human behavioural over time.

The use of wearable dedicated sensors for gait authentication opened a new domain of transparent and continuous user authentication, at best an EER of 2.5% (Cola et al., 2016). However, these studies all relied on extremely limited amounts of gait data from each user (30 to 120 seconds) and required the use of costly specialized devices in order to collect the data. Furthermore, these devices require comprehensive set-up that reduce the usefulness of their performance and increases the cost of implementation into a potential real-world system.

3.4 Mobile Accelerometer-based Gait Authentication

As discussed in the previous section, attaching a dedicated sensor around the human body for gait verification is costly to implement. Therefore, recent studies attempted to utilize the mobile sensors (i.e., accelerometer and gyroscope) for collecting the gait data. A comprehensive analysis of the previous research on mobile gait authentication has been investigated (with the papers included being listed in Table and detailed in Appendix A). There are two main advantages of using mobile sensors in gait verification: the first being that no additional hardware is required, while the second is that users are for the most part accustomed to

carrying the device. Therefore, authentication can be conducted in a transparent manner.

	Authors	Year	Type	Citation
1	Derawi et al.	2010a	Conference	136
2	Kwapisz et al.	2010	Conference	102
3	Nickel et al.	2011a	Conference	48
4	Nickel et al.	2011b	Conference	20
5	Nickel et al.	2011c	Conference	7
6	Nickel et al.	2011d	Conference	25
7	Nickel and Busch	2011e	Journal	15
8	Hestbek et al.	2012	Conference	4
9	Busch and Nickel	2012	Conference	1
10	Wirtl et al.	2012	Conference	17
11	Muaaz and Nickel	2012	Conference	9
12	Ho et al.	2012	Conference	5
13	Shrestha, et al.	2013	Conference	6
14	Muaaz and Mayrhofer	2013	Conference	11
15	Ross, A	2013	Conference	3
16	Hoang et al.	2013	Journal	15
17	Muaaz andMayrhofer	2014	Conference	4
18	Watanabe, Y	2014	Conference	3
19	Gascon et al.	2014	Conference	58
20	Watanabe, Y	2015	Conference	1
21	Damaševičius et al.	2016	Journal	15
22	Ehatisham-ul-Haq, et al	2017a	Journal	1
23	Kumar, et al.	2017	Conference	1
24	Kumar, et al.	2017	Conference	1
25	Ehatisham-ul-haq, et al	2017b	Conference	1
26	Shen, et al.	2017	Conference	13
27	Lee et al.	2017	Journal	6

Table 4: A summary of mobile-based gait authentication studies

Although mobile-based gait authentication provides an unobtrusive and user-friendly method for authentication, the majority of previous studies collected the motion data by placing a mobile phone in a fixed position (i.e., in the trouser pocket or on the hip). However, users can put their phone in numerous locations around their body wherever there is a pocket (i.e., inside coat pocket and back pocket). Moreover, the collected signals by smartphones are too noisy that require extensive pre-processing, which add extra cost in terms of the required resources.

3.5 Smartwatch Accelerometer-based Gait Authentication

The increased popularity of smartwatches, which tend to be sensor-rich highly personal technologies (e.g., accelerometer, gyroscope, and heart rate), attract an enormous amount of interests. So far, however, little attention has been given to the use of wearable devices for the authentication purposes. Given that the smartwatches are usually worn in a fixed position (i.e., on either the right or left wrist), they offer more accurate and reliable personal biometric data than smartphone do.

While this research was in the progress, there were only three articles (Mare et al., 2014; Johnston and Weiss, 2015; Junshuang Yang et al., 2015) published in the area of activity about gait-based user authentication using smartwatches. Since then, 12 studies have been published and identified covering issues related to gait and gesture-based user authentication. However, these publications have not influenced the direction of this research, in regard to the data collection methodology and the broad spectrum of the results that were collected in this thesis. Moreover, the majority of these papers still suffered from several shortcomings; for example, using limited dataset and samples, data collected on the same day, using unrealistic methodology to collect the user's motion data (i.e., controlled environment), and a limited range of activities were considered. Table 5 displays an overview of the selected smartwatch-based user authentication studies.

	Authors	Year	Type	Citation
1	Mare et al.	2014	Conference	49
2	Johnston and Weiss	2015	Conference	38
3	Yang et al.	2015	Conference	24
4	Kumar et al.	2016	Journal	18
5	Davidson et al.	2016	Journal	7
6	Shrestha et al.	2016	Journal	8
7	Lewis et al.	2016	Conference	3
8	Dong and Cai	2016	Conference	1
9	Lee and Lee	2017	Conference	15
10	Griswold et al.	2017	Conference	1
11	Liang et al.	2017	Conference	1
12	Wang et al.	2017	Conference	3
13	Xu et al.	2017	Conference	9
14	Ahmad et al.	2018	Journal	1
15	Acar et al.	2018	Journal	1

Table 5: An overview of the selected smartwatch-based authentication studies

Using smartwatches for collecting the user's movement data have several advantages over smartphones that are summarized below:

- The captured signals from the wearables are less noisy due to the consistent placement of the device (i.e., on the left or right wrist).
- Unlike smartphones that capture limited activities (e.g., gait and typing activities), a wide variety of personal data could be collected from smartwatches such as eating, typing on PC, dribbling, clapping, brushing teeth, drinking, and several arm gestures (for example, punch gesture or drawing a circle)
- Smartwatches can be used to capture more accurate and personal biometric data (e.g., acceleration, heart rate, and skin temperature) than smartphones do.

3.6 Discussion

Despite a large body of research, the problem of sensor based- authentication is far from a solved problem. Table 6 displays a comprehensive analysis of the prior studies on gait authentication which has been discussed in this literature. The commentary that follows describes the key achievements and milestones that have taken place.

	Authors	FA	Features type	Classification methods	Performance %	Number of	Scenario
1	Mätyjärvi et al.(2005)	C	FD	SC	EER= 7& 10	36	CD
2	Gafurov et al.(2006a)	C	TD	ABS	EER= 5 & 9	21	SD
3	Gafurov et al. (2006b)	C	TD	EUC	EER= 16	22	SD
4	Okumura et al. (2006)	C	TD	DPI	EER=5	22	SD
5	Snekkenes et al.(2007)	C	TD	ABS & Correlation	EER= 7.3 & 9.3	50	SD
6	Gafurov et al. (2007)	C	TD	EUC	EER= 13	100	SD
7	Gafurov and Snekkkenes (2008a)	S	FD	EUC	EER=13	30	SD
8	Gafurov and Snekkenes (2008b)	C	TD	EUC	EER= 5.6	30	SD
9	Gafurov et al. (2010)	C	TD	EUC	EER= 1.6	30	SD
10	Sangil Choi et al. (2014)	C	TD	K-NN	CCR =100	10	SD
11	Cola et al. (2016)	C	TD	k-NN	EER=2.5	15	SD
12	Derawi et al. (2010a)	C	TD	DTW	EER=20.1	51	CD
13	Kwapisz et al.(2010)	S	TD	J48 & FFMLP	CCR=100	36	SD
14	Nickel et al. (2011a)	S	FD	HMM & MV	FRR=10.42 FAR=10.29	48	CD
15	Nickel et al. (2011b)	S	TD&FD	SVM &MV	FRR =6.3 FAR=5.9	48	CD
16	Nickel et al. (2011c)	S	FD	SVMs, HMMs & QV	EER= 10 and 12.63	36	CD
17	Nickel et al. (2011d)	C	TD	Manhattan, DTW	EER= 21.7 and 28	48	CD
18	Nickel& Busch (2011e)	S	FD	HMM& QV	EER= 6.15	48	CD
19	Hestbek et al. (2012)	S	TD& FD	SVM &QV	FAR = 9.82 FRR=10.45	36	CD
20	Busch and Nickel (2012)	C	FD	HMM & QV	EER=15.46 & 13.89	36	CD
21	Wirtl et al. 2012	S	FD	K-NN & QV	HTER=8.4	36	CD
22	Muaaz and Nickel (2012)	C	TD	DTW & MV	EER=29.39	48	CD

23	Muaaz and Mayrhofer (2013)	C	TD	DTW & MV	EER= 33.3	51	CD
24	Shrestha, et al. (2013)	S	TD	-	FRR=10 FAR =1%.	20	SD
25	Ho et al. (2012)	C	TD	SVM	CCR =100	32	SD
26	Ross, A. (2013)	S	TD	J48 & FFMLP	CCR= 90.3 & 70.5	9	SD
27	Hoang et al. (2013)	S	TD	SVM	CCR=91.33	14	SD
28	Muaaz and Mayrhofer (2014)	C	TD	DTW	EER=19	35	CD
29	Watanabe, Y. (2014)	S	TD	FFMLP	FAR =1.30 FRR =2.34	4	SD
30	Gascon et al. (2014)	S	TD	SVM	TP= 92 FAR= 1	315	SD
31	Watanabe, Y. (2015)	S	TD	FFMLP	CCR=97.92	8	CD
32	Damaševičius et al.(2016)	S	TD	Jacc	EER=5.7	14	SD
33	Ehatisham-ul-Haq, et al. (2017a)	S	TD&FD	SVM, BN, DT & k-NN	CCR= 99, 97.4, 97& 93	10	SD
34	Kumar, et al. (2017)	S	TD&FD	K-NN, SVM, & RF	EER=12.1, 10.7, 5.6	57	CD
35	Ehatisham-ul-haq, et al. (2017b)	S	TD&FD	K-NN, BN& SVM	CCR=89.7 94.5 & 94.2	10	SD
36	Shen, et al. (2017)	S	TD&FD	HMM	EER=4.93	102	CD
37	Lee et al. (2017)	S	FD	DTW	FAR =0 FRR= 7.6	24	CD
38	Johnston and Weiss (2015)	S	TD	RF, FFMLP, & NB	EER= 1.4, 2, & 4.5	59	SD
39	Yang et al. (2015)	S	TD	DTW	EER= 5%	26	CD
40	Kumar, et al. (2016)	S	TD&FD	k-NN	CCR=86.8	13	CD
41	Davidson et al. (2016)	S	FD	k-NN	TP =88.4 FP =1.3	10	SD
42	Shrestha et al. (2016)	S	TD&FD	RF	EER=2.6	18	CD
43	Lewis et al. (2016)	C	TD	DTW	FRR= 30 FAR=15	5	SD
44	Dong and Cai (2016)	S	TD	SVM	EER=0.65	20	SD
45	Lee and Lee (2017)	S	TD&FD	KRR	FRR=22.3 FAR=13.4	20	CD
46	Griswold-Steiner et al. (2017)	S	TD&FD	SVM	EER=7, 10, & 15	20	CD
47	Liang et al. (2017)	S	TD	SVM	EER=4	20	CD
48	Wang, et al. (2017)	S	TD	Manhattan	EER=4.3	10	SD
49	Xu et al. (2017)	C	TD	k-NN	CCR=96	20	CD
50	Ahmad et al. (2018)	S	TD&FD	DT, K-NN, SVM, & NB	CCR=90.4, 90.2, & 77	6	CD
51	Acar et al. (2018)	S	TD&FD	FFMLP	EER=1	34	SD

Table 6: Comprehensive analysis of the prior studies on gait authentication

(C: Cycle-based; S: Segment-based; TD: Time Domain Features; FD: Frequency Domain Features; DTW: Dynamic Time Warping; EUC: Euclidean distance; ABS: absolute distance; MV: majority voting; QV: Quorum voting; Jacc: Jaccard distance; BN: Bayesian network; SC: Signal Correlation; DT: Decision trees; LR: Logistic Regression; RF: Random Forest; KRR: Kernel Ridge Regression; k-NN: k-Nearest Neighbors; HMM: Hidden Markov Model; SVM: Support Vector Machine; EER: Equal Error Rate; CCR: Correct Classification Rate; SD: Same Day; CD: Cross Days).

In most evaluations, a relatively small data set was used and frequently was obtained on the same day. This contradicts the notion that the only more reliable test comes from multi-day testing. This maxim holds because performance on single day datasets does little to test how resistant the system is to the variability of the human gait over the time (Nickel et al., 2011b; Muaaz and Mayrhofer, 2014). Most studies claiming a system resilient to the CD problem either trains on mixed data from both days (thus not making it a true CD test as a user will be required to enroll in the system every day) or has an error rate so high that the system would not be practical. Notably, the lack of realistic data underpins a significant barrier in applying these systems in practice (in both mobile and smartwatch contexts). In cases when multi-day scenario was considered, the error rates were significantly increased but this is more realistic evaluation scenario as it avoids training the user's model every day.

The use of smartwatches for capturing the user's activity data have several advantages over smartphones. It is envisaged that smartwatches have the ability to capture more accurate personal data (e.g., acceleration and heart rate) than smartphones do due to their fixed contact with individuals (i.e., on either the left or right wrist). The majority of previous studies collected the user's movement data by placing a smartphone in a fixed position (e.g., in the trouser pocket or on the hip). It is widely understood that smartphones suffer from several issues to produce consistent and reliable data collection in real life; for example, carrying the device in a handbag makes the data collection process less accurate or nearly

impossible. In contrast, smartwatches provide a more consistent data collection of the user's motion as it is almost fixed to the user regardless of their clothing choices. Smartwatches can provide a consistent orientation (i.e., it is worn in such a way that the text on screen is easily readable to the user). As a result, smartwatches offer the opportunity to collect the user's motion data in a more effective and reliable fashion than smartphones could. Several activities (e.g., eating, PC browsing, and hand gestures) would not be recognized when a smartphone is used to collect the movement data. However, smartwatches tend to capture a wide variety of personal activities.

Although sensor based-authentication systems could be implemented using accelerometers or gyroscopes as the source triaxial (three axes) sensor, the literature seems to overwhelmingly support the use of the accelerometer alone. Intuitively, both sensors should offer similar information and thus similar levels of predictive power, but in practice only few studies (Johnston and Weiss, 2015; Lee et al, 2017; Ahmad et al., 2018) that test systems using both sensors independently overwhelming show that accelerometers offer better accuracies and error rates. This constraint should not present a realistic problem; both sensors are almost ubiquitous on all smartwatches. It is possible that the fusion of data from both sensors would offer a greater level of accuracy than either sensor alone (Damaševičius et al. 2016; Lee et al, 2017); however, there is little research on the subject. This is presumably the result of Android Wear (a popular if not dominant operating system for smartwatches), which does not allow the two sensors to be sampled simultaneously (rather, they must be sampled successively), making fusion difficult to perform in a precise manner.

So far, all the early studies have focused upon using data that has been collected within a controlled environment (i.e., all users were asked to do exactly the same

type of activity such as walking on a flat floor in an indoor environment). This experimental approach, whilst standard in assessing the feasibility of a biometric in the early stages of research, is arguably not reflective of real-world use (i.e., tends to be far less realistic for real world applications). This is because capturing labelled data for training and testing a classifier under different operation conditions (e.g., carrying a load, hands in a jacket or trouser pocket, and various walking speeds and surfaces) is challenging or nearly impossible. In reality, the process of labelling the motion data for the reference and test templates is quite intrusive and unlikely to be implemented by industry and/or accepted by common smartphone users. Therefore, a more realistic experiments should be investigated by collecting real life data to make sure the captured signals can be used for practical authentication system.

As outlined previously, there are fundamentally two different approaches used to pre-process the raw acceleration data, cycle extraction and segmentation. Cycle extraction purportedly offers a precise manner of generating instances from the testing data by detecting steps and splitting the data accordingly. This offers an exciting opportunity where if such a system is implemented effectively, a system may be able to be trained in just a manner of steps. Nevertheless, based on the recent mobile-based gait studies (Derawi et al., 2010a; Nickel et al., 2011d; Muaaz and Nickel, 2012; Muaaz and Mayrhofer, 2013), the performance of using cycle extraction method was low. At best, cycle extraction methods can operate at 15.46% of EER (Nickel and Busch, 2012). The high error rate of using this approach is most likely the result of the complicated and unclear nature of cycle extraction, as gait is only semi-periodic and the signals originating from these devices are noisy due to a confluence of factors (e.g., the device not being securely fastened to the user, cheap sensors, rounding errors, etc.). Furthermore,

cycles are not guaranteed to be of the same length (and can vary widely in length depending on the pace of a user); cycle extraction must be paired with a system that normalizes the length of each step, which adds yet another parameter to be tested and refined. In contrast, the segmentation-based methods focus on fixed-length blocks of data. These methods, while not guaranteeing the number of steps (in the case of short windows, there may be no full steps at all) or that the completeness of all steps within the window, is simple to implement. Despite the simplicity of segmentation based method, it appear to be more effective in most implementations with an EER of 10% in the worst scenario (Nickel et al., 2011a; Nickel et al., 2011b; Nickel and Busch, 2011; Watanabe, Y. 2014; Johnston and Weiss, 2015; Yang et al., 2015).

With respect to features, there have been several studies in literature that suggested generating the statistical and cepstral coefficient features from a fixed segment size could produce better performance scenario (Nickel et al., 2011a; Nickel et al., 2011b; Nickel and Busch, 2011; Watanabe, Y. 2014; Watanabe, Y. 2015; Johnston and Weiss, 2015; Nickel and Busch, 2012; Hestbek et al., 2012; Ho et al., 2012; Hoang et al., 2013). These studies used statistical features such as AAD, RMS, BD, TBP, Max, Min, Mean, and Std. Likewise, more recent features have borrowed from signal processing or speaker recognition areas by using features derived from the Fourier transform of the signals. Specifically, MFCCs and BFCCs were used in some papers to great success. In addition, some studies have relied on a combination of MFCCs and BFCCs alone and still managed to produce strong results (Nickel et al., 2011c; Nickel and Busch, 2011; Nickel and Busch, 2012).

The majority of researches in this literature do not seem to be an overwhelming concern with the length of feature vectors. Unless a specific need for the biometric

system to reside entirely on the smart device (thus severely limiting the amount of available processing power and memory) arises, it is likely that feature vectors will continue to expand as long as the additional features provide a greater level of accuracy. Nevertheless, an advance feature selection approach is required, especially for smartwatches/smartphones-based user authentication system in order to reduce the potentially large dimensionality of input data and to maximize the system performance.

Various feature selection approaches were proposed in the prior gait/ activity-based user authentication systems (Nickel et al., 2011b; Nickel and Busch, 2011; Nickel and Busch, 2012; Hestbek et al., 2012; Hoang et al., 2013). Nevertheless, these studies were based upon evaluating the performance of individual feature and then pick out a subset that achieved the lowest EER under some classification system. This could be useful if the proposed system consists of few features (e.g., 5, 10 or 15 features), otherwise the implementation of a such method would be worthless as the extracted features are relatively correlated to each other. In comparison, several biometric-based authentication systems created the user's reference and test templates based upon selecting the most common features (e.g., features that have the smallest standard deviation for all the population. This could result in making the system vulnerable to accepting illegitimate user (i.e., high FAR). However, a balance between security and usability needs to be taken for TASs (i.e., low FAR and low FRR). Most recent smartwatch-based gait recognition study by Kumar et al., (2017) utilized two feature selection algorithms, namely Information Gain Based Feature Ranking and Correlation Feature Selection. However, the prediction accuracy was relatively low (i.e., 86.8% correct classification rate). Therefore, a novel feature selection strategy is required to offer a delicate balance between usability and

security. In term of classifiers, using the standard classification methods (e.g., Absolute, Euclidean, and DTW distance metrics) for training biometric systems is another subject that is still debated within the literature.

Many researchers prefer a more traditional (to the area of biometrics) approach where a single template is generated (much as a system that relies on fingerprints or facial recognition would) and is later tested by finding the template most similar to the test data. While this approach works well for certain domains, it does not seem to be the most effective type of system for activity recognition or other behavioural biometric techniques. This is due to the fact that the user's behaviour changes over the time. Hence, applying these methods resulted in high EERs ranged between 19% (Muaaz and Mayrhofer, 2014) and 33.3% (Muaaz and Mayrhofer, 2013). Therefore, it is more reasonable to collect multiple instances from each individual on multiple days and utilising more complex algorithms than have been tried in earlier studies.

It is interesting to note that the majority of findings of the aforementioned investigations were based upon applying majority and quorum voting schemas in order to make a decision. Although quorum voting usually yielded greater performance (Nickel et al., 2011b; Nickel et a, 2011c; Nickel and Busch, 2011), the majority voting appears to be more resilient to error given the higher threshold for classification (Kwapisz et al., 2010; Nickel et al., 2011a). Quorum, while lowering the level of accuracy required to verifying a user, may result in a high false acceptance rate. This failure to identify imposters can be explained by the extremely low proportions of correct classifications required to accepting a user as genuine. Although this may be acceptable for systems more concerned with usability, such permissiveness will most likely render the system impractical for most uses. Majority voting, while requiring the system to be more discriminative,

offers a greater level of security and thus is more likely to offer a suitable balance between usability and security. Ultimately, conscious decisions must be made to create a system that does not appear to the end user as too demanding without compromising too much security.

3.7 Conclusion

The literature on sensor based- biometric authentication demonstrates increasing levels of promise. Initial experiments conducted 10 years ago barely obtained an EER of 19% to more modern systems nearing to an EER of 2.6%. This drastic improvement can be attributed to more intricate feature vectors that utilize more complex features and a departure from purely statistical methods to more artificial algorithms.

It is apparent that smartwatches are the most effective hardware option to collect the motion data Johnston and Weiss, (2015). Smartphones, while having the benefit of technological maturity and widespread adoption, suffer from too many problems to produce a consistently effective implementation. Namely, the problems of orientation and off-body carry (i.e., when the device is not carried in a pocket or somewhere else close to the body) make obtaining consistent accuracy nearly impossible. Smartwatches, by virtue of being watches, guarantee consistent placement on the body regardless of clothing choices of an individual user. Similarly, since the smartwatches do not rotate their screen based on orientation, the smartwatch is worn in a consistent orientation at all times (i.e., it is worn in such a way that the text on screen is easily readable to the user). These advantages make the possibility to design an effective transparent and continuous user authentication system for both mobile/smartwatch, as the need to develop orientation and placement independent features is negated.

The majority of prior studies in the domain collected data within a controlled environment (i.e., users were asked to perform specific activities or gestures in an indoor environment) and subsequently utilize this data in order to verify the user's identity in a transparent and continuous manner. However, such dataset tends to be far less realistic for real world applications. Moreover, these studies have relied upon limited activities (i.e., gait or gestures). Collecting real life motion data is a big challenge as the user's arm pattern could be vary depending on the activity type. As the process of obtaining labelled samples in the real-life scenario is unexpected or quite intrusive, developing an approach that automatically identifies the activity type for each context might significantly improve the authentication decisions. Further influencing factors on the biometric system performance is the selected feature subset; selecting unique features for each user would improve the results and reduce the complex computations on the smart devices which have limited processing resources. Therefore, a feature selection approach of any mobile/smartwatch-based biometric system needs to be sophisticated enough before the classification phase takes place.

4 Feasibility Study into the Capture & Analysis of Smartwatch-based Activity Recognition

Chapter 3 has identified the possibility of using the smartphone and smartwatch acceleration and gyroscope data for TAS. It has been highlighted that smartphones suffer from providing consistent and reliable movement data. Although smartwatches offer the opportunity to capture rich and personal biometric-based user information, only few studies utilized these devices and were based upon limited activities (i.e., gait or unrealistic gesture). The aim of this chapter is to present a feasibility study to use the smartwatch movement sensors (i.e., accelerometer and gyroscope) in order to capture multiple activities (not merely gait or gesture). It presents a comprehensive evaluation on wearable technology, details of the collected dataset, feature extraction, a novel feature selection method, and comprehensive results to determine whether the proposed system can be applied to protect the sensitive information on both devices (i.e., smartphones and smartwatches).

4.1 Introduction

The earlier discussion has identified that the intrusive implementations of the current user authentication approaches (i.e., PIN and passwords) spur smartphone users to take no security precautions against unauthorized access. Entering PIN code adds loads of burden to the smartwatch users due to the small touch screen of these devices. It is widely recognized that those methods are considered an unreliable basis for user authentication hence, they are an attractive target for attackers to misuse the user's personal data.

As long as the current wearables are connected to smartphones via Bluetooth, a permanent access would be provided to the smartwatch users. This is because

the identity check happens only during the pairing process (i.e., the smartwatch will automatically be connected to the smartphone without requiring user credentials). Therefore, securing information on these devices from unauthorized access in an effective and usable fashion is crucial. Several TAS for smartphones were proposed such as the user's typing rhythm (Banerjee and Woodard, 2012), behaviour profiling Li et al., (2011), ear and face recognition techniques (Ali Fahmi et al., 2012; Clarke et al., 2008). Nevertheless, one of the key challenges for using TAS is the lack of appropriate biometric modalities. In addition, previous research in this domain also encounters performance issues due to the reliability of behavioural biometrics (i.e., the performance can be influenced by external environmental factors such as mood) (Saevanee et al., 2012).

In recent studies, biometric measurements based on motion signals (e.g., the accelerometer and gyroscope readings) were collected by utilizing mobile phone sensors for transparent and continuous user authentication. Nowadays, wearable devices have become increasingly prevalent among users and equipped with rich sensors that are capable of holding versatile and quite frequently highly sensitive user data. This data was employed to develop several applications such as health-related, conducting financial transactions, and capturing physical activities. The possibility of collecting the motion data from a dedicated sensor and/or smartphone technology for implementing a transparent and continuous user authentication system is highlighted in the previous chapter. However, little attention is given to the use of wearable devices – which tend to be sensor-rich, highly personal technologies.

Wearables could be used to enhance mobile security in a more effective way. Few studies have demonstrated that smartwatches can provide continuous and transparent biometric authentication service by using the accelerometer and/ or

gyroscope data (Mare et al., 2014; Johnston and Weiss, 2015; Junshuang Yang et al., 2015; Kumar, et al., 2016). However, the prior research either used a limited dataset or trained and tested the system on data that was collected on the same day (which is not a realistic model for a real-world application as the user would be required to enrol on the system every day). Moreover, early smartwatch-based user authentication studies focussed upon merely a limited range of activities (i.e., gait activities only).

To this end, the present chapter examines the possibility of using smartwatch technology for acquiring the desired motion signals for the TAS based upon the user's daily activities. The main contributions of this study are demonstrated as follows:

- Based upon prior art, this is the biggest dataset for activity -based user authentication using smartwatch, which contains data of 60 users over multiple days.
- To provide an evaluation of the approach against a number of activities (rather than a single activity). Five popular daily life activities were captured (i.e., normal walking, fast walking, playing a mobile game, typing on a PC keyboard and texting on a mobile touch screen).
- To explore a comprehensive feature set that was extracted in the time and frequency domains to highlight their usability and the impact on system performance.
- To investigate and propose a novel feature selection method that was based upon generating a dynamic feature vector for each user and successfully reduced the feature vector size with better performance.
- To evaluate and compare the optimal source sensor for the authentication task.

The rest of this chapter is organized as follows: section 5.2 provides a comprehensive evaluation of wearables technology. The methodology for data collection, pre-processing, feature extraction and reduction is presented in section 5.3. Experimentation along with corresponding discussion of the results are presented in sections 5.4 and 5.5 respectively. Finally, conclusions are detailed in section 5.6

4.2 Technology Evaluation

Several wearable technologies are launched in the market that contain a wide variety of sensors. In order to select the suitable technology for capturing the motion-based signals that fit the research aims and objectives, a comprehensive analysis needs to be conducted. These include, what are the available sensors in the wearable technology?, what smartphones can be connected with the wearable technology?, what are the existing application(s) in order to obtain data from the device?, how good is the sensor precision readings?, and how expensive is the smartwatch . Based on the answers to these questions, the optimal device will be selected for obtaining the motion data that can be used for a transparent and continuous biometric authentication system.

- **Microsoft Band 2:** the Microsoft Band 2 is an advanced fitness tracker that can be paired with smartphones running Android 4.1 or above, iOS 8.2 and later, and Windows 8.1. This provides users with the benefit of being able to use the smartwatch regardless of preferred smartphone ecosystem. It includes 13 sensors (i.e., optical heart rate, 3-axis accelerometer and gyroscope, GPS, pedometer, skin temperature, ambient light, galvanometer, magnetometer, altimeter, thermometer, ultraviolet, Galvanic skin response and microphone). This means the Band 2 can collect several biometric-based data that would be useful for any behavioural-based biometric systems.

The readings of a galvanic skin response sensor, which checks if the smartwatch user is wearing the band or not, can be effectively used for activity- based user authentication using smartwatches. For example, if a user takes off the band, the collected data for this period would be neglected. Moreover, the Band 2 screen is always active allowing collection of motion data while the user on the go. Collecting samples from the smartwatch accelerometer and gyroscope sensors would require a custom application that runs on both the phone (for storage and transmission purposes) and the smartwatch itself. However, there is an open source application available that performs this function.

Figure 9: List of smartwatches

- **LG Urbane Watch:** the LG Urbane is a small and light watch (66.5g) having three triple-axis inertial motion sensors (compass, accelerometer, and gyroscope), heart rate monitor, and barometer. The accelerometer and gyroscope are present as a single microelectromechanical system chip (i.e., a MEMS chip), which provides the device with information about instantaneous acceleration and rotational velocity. The watch display is a 1.3-inch PO led screen that is protected by a Gorilla Glass 3 holding up against scratches, and for some added protection the glass is slightly recessed into the body of the watch to help against accidents. The LG Urbane works with smartphones and all smartwatch applications that run on Android Wear; this includes many popular applications such as Runtastic Running and Fitness and Cloud Magic.
- **Samsung Gear Live:** the Samsung Gear is slightly lighter than LG G Watch (59g) and offers dust and water resistance. In comparison with LG Urbane, the Samsung Gear contains only four sensors (i.e., accelerometer, gyroscope, heart rate monitor, and compass). It is optimised for mobiles running on Android 4.3 and onwards but does not support devices running iOS of any version.
- **Sony Smartwatch 3:** it has 5 in-built sensors (ambient light sensors, GPS, compass, accelerometer and gyroscope). Sony Smartwatch 3 is compatible with smartphones running Android 4.3 and later. The watch weight is 45g. Although the Sony company claims that the battery life is up to 2 days of normal use, testing the design to reflect usage over an average day suggests the battery is completely drained within a few hours (Summerson, C. 2015). Furthermore, this problem has been publicly acknowledged by Google support (Hayden, 2015), with the support team stating that a conflict between Android

Wear 5.1 (the most current version of the Android Wear subsystem) conflicts with the firmware of specific watches and results in massive battery drain and overheating (Hayden, 2015).

- **Apple Watch:** the Apple Watch is Apple's sole offering in the smartwatch field and is compatible only with Apple devices. There are several different sensors built into the Apple Watch (e.g., heart rate, ambient light sensors, pulse oximeter, accelerometer and gyroscope) to measure steps taken, calories burned, pulse rate and a variety of other metrics. In terms of battery performance the Apple Watch has the same issues of Android Watches, its battery life is about 18 hours of normal use (Stables, J. 2015). In comparison with the other dominant offerings in the smartwatch market, the Apple Watch is 78% more expensive than the most expensive Android Wear offering (see Table 7).

  As shown in Table 7, all of the selected smartwatches offer the basic sensors: accelerometer and gyroscope. Given that there is no particular advantage in opting for a more expensive device (at least for research purposes), it is reasonable to suggest that an optimal device for data collection is the one that is most cost effective. It is apparent that Microsoft Band 2 has more sensors (e.g., GPS and Skin temperature) and the cheapest compared to other smartwatches. These sensors offer the opportunity to capture various personal, biometric-based data, which can be useful for a transparent and continuous biometric system. Also, it can be connected to multiple mobile platforms (i.e., Android, iPhone and Windows Phone); therefore, there are no restrictions in order to collect data from a large pool of participants who have different types of smartphones. In addition, unlike other smartwatch technologies, Microsoft Band 2 offers the opportunity to collect data in a

continuous manner for at least 4 hours without recharging and thus offers the potential to collect a huge amount of real life data.

Features	Microsoft Band 2	LG Urbane	Samsung Gear	Sony 3	Apple Watch
Sensors	Accelerometer Gyroscope Compass Heart rate Ambient light GPS Skin temperature Pedometer Microphone Magnetometer Altimeter Ambient Light	Accelerometer Compass Gyroscope PPG Barometer	Accelerometer Compass Gyroscope ECG	Accelerometer Compass Gyroscope Ambient light sensors	Accelerometer Gyroscope, Heart rate, Ambient light sensors Pulse oximeter
Bluetooth	✓	✓	✓	✓	✓
Smartphone Compatibility	Android 4.3 and later , iOS 8.2 or newer, Windows 8.1 or later	Android 4.3 and later , mobiles running iOS 8.2 or above	Android 4.3 and above	Android 4.3 and above	iPhone5 and newer
Battery life	two days	two days	One days	Two days	One day
Operating system	Android Wear	Android Wear	Android Wear	Android Wear	IOS
Price (in £)	125	165	190	190	340

Table 7: Comprehensive evaluation of wearable technology

4.3 Experimental Methodology

With the aim of investigating the feasibility of using wearable computing for transparent user authentication, extensive experiments were conducted to capture and analyse the user's movement data. In order to overcome some of the shortcomings of the prior work, this section will explore the following research questions:

1. What is the impact of the time and frequency domain features on the system performance?

2. Which sensor can provide a more consistent and reliable motion data for recognizing individuals?

3. What is the most effective classification strategy – generic or activity-based authentication model?

4. Can the captured activities use for identifying the user's arm pattern?

5. How does the performance vary across same and cross-day evaluation methodologies?

6. Does the proposed feature selection approach have a positive effect on the proposed system performance?

To address these questions, the following experiments were conducted:

- Time and Frequency Feature Analysis, accelerometer vs gyroscope sensor (research questions 1 and 2).

- Evaluating Generic vs Activity-based authentication model, different activities, (research questions 3, 4, and 5).

- Single and cross day scenario, all features against selective feature subset (research questions 5 and 6).

4.3.1 Data collection

To determine and evaluate the feasibility of the proposed activity-based user authentication system, it is important to ensure the population sample being used as large and significantly reliable as much as possible. Therefore, this experiment aims to capture sufficient number of samples from each individual to effectively train the user’s reference template with a variety of possible instances of the same

activity; as a result, the recognition rate could be increased. In order to collect user's movement data, the Microsoft band 2 is utilized due to its wide range of built-in sensors.

A third-party application called Companion for Band, which is compatible with all android smartwatches and smartphones that run versions of Android 4.3 and later, was utilized to capture the accelerometer and gyroscope signals (see Figure 10). The application contains three different sampling rates (i.e., 16 Hz, 32Hz and 128Hz as shown in Figure 11) and data was captured at 32 Hz. Reasons for selecting 32hz sampling rate was to capture enough accelerometer and gyroscope readings and to avoid repetition of the axis values; as a result, less signal noise and better performance can be obtained. In addition, more power will be consumed, and more storage space will be required if a higher sampling frequency is applied. As soon as the data was acquired by the smartwatch, it was sent to a smartphone residing in the user's pocket via Bluetooth (in the rare event if the user did not have a pocket he/she was told to hold the phone in their dominant hand).

Figure 10: View life data streams of all sensors

Figure 11: Sensors sampling rate

As highlighted earlier, the prior art merely focused on gait or gesture data that are limited activities; therefore, a wide range of activities were considered in this study that are non-intrusive, frequently used, contain unique arm pattern, and more natural. These included five physical activities (i.e., normal walk (NW), fast walk (FW), typing on PC (TypePC), playing mobile game (GameM), and typing on mobile phone (TypeM)). Based on the previous studies (Nickel et al., 2011d; Nickel et al., 2012b; Lee et al, 2017; Shen et al., 2018; Kumar et al., 2017) that were able to capture acceptable number of samples (in the range of 36 to 100 samples) and achieved stronger accuracies than other prior art, this study aims

to have at least the same amount of data but also a trade-off between the time required and cost involved for getting people participate in the experiment. Therefore, it was considered that a total of 60 hours of the movement data from 60 users (26 males and 34 females) was appropriate and, hopefully, shows better data than any other prior art.

For each activity, 72 samples were obtained from each user (in total, each participant provided 360 samples for the all activities over two days), The age of the participants was ranging from 18-55 years old as shown in Figure 12; most participants in the data collection methodology were university students (80%) while the rest were university staff or faculty (each participant received compensation of £25). Once ethical approval was sought and obtained, accordance with the guidelines provided by University of Plymouth, written informed consent was obtained from each test subject prior to data collection.

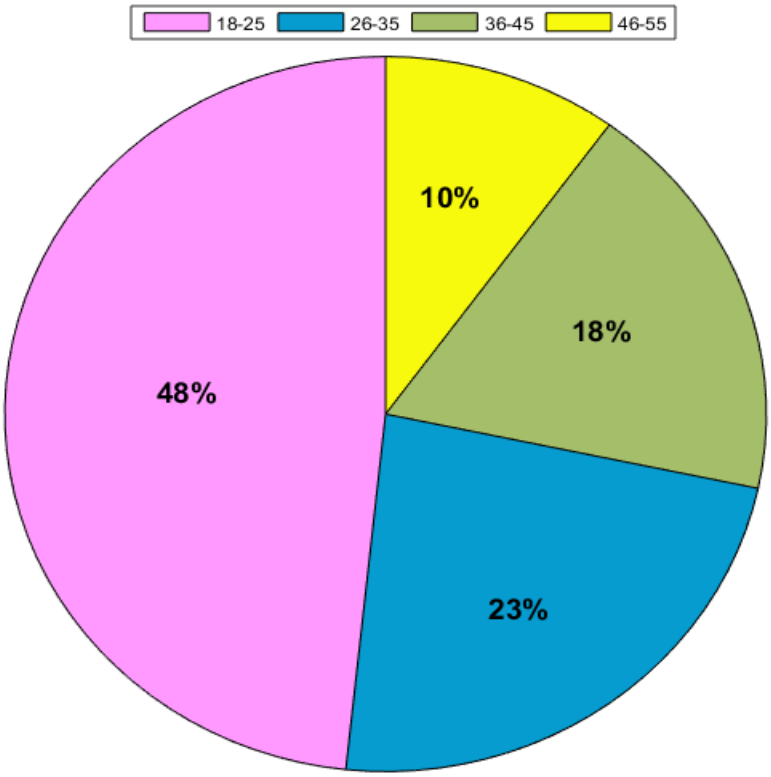


Figure 12: The age ranges across the participants for the controlled dataset

In order to be able to perform both same-day and cross-day analysis (as pre the prior art), two sessions were obtained per participant for each activity occurring on two separate days within a time frame of 3 weeks. The reason for capturing the controlled data over two days only is to ensure that each participant follow up and complete the requested sessions. Data for each activity and each session was carried out in three phases (i.e., phase 1, phase 2, and phase 3) separated by at least 15 minutes time interval. The single phase contained two minutes of the user's motion data; the reason for capturing data of one session in three different phases is to get more reprehensive data rather than repeating the same activity.

The raw gait signals (i.e., NW and FW) were collected by asking users to walk on a predefined route and encouraged to walk on flat ground in their own natural and comfortable manner. For consistency, the gait data was collected on the second day in a manner similar to the first, with the user walking over the same route. For a more realistic scenario, the user had to stop in order to open a door and take multiple turns. Moreover, no other variables, such as type of footwear or clothing, were controlled.

In addition to gait data, typing activities (typing on PC and smartphones keyboard) were also considered in this study. Although the prior studies of keystroke-based behavioural biometrics showed the possibility of verifying users based upon their typing rhythm, keystroke technique requires plenty of data to train the classifier. Moreover, the discriminative characteristics of this modality are based upon the inter-key latency and hold time that effected by external factors such as change of keyboard. The applicability of such system was limited on computers or smartphones and to the best of the author's knowledge this is the first study that

investigated the usage of smartwatch motion sensors for recognizing the user's identity based upon their typing rhythm.

In the typing activities, users were asked to sit and continuously type a short and predefined text on the touch screen of their smartphone and on a PC keyboard. Regarding of the game activity, users were asked to sit and playing Candy Crush Saga on their smartphone. The criterion for selecting this game was based on many factors; for example, it was the top mobile games by downloads, free application to install, simple to play, and contains enough touch gestures on the mobile touch screen to obtain a unique pattern for each individual.

4.3.2 Data Pre-Processing

This section describes the procedure of collecting and transforming the data into a form suitable for traditional machine learning classification algorithms. Pre-processing provides a mechanism to remove unnecessary noise from the signal data; once the data collection was completed, the signal processing phase was undertaken- a brief description of the steps is described below:

- **Time interpolation:** Due to the limited accuracy of sensors in Android devices, the smartwatch only outputs whenever there is a change in acceleration and gyroscope values Therefore, time interpolation was applied to ensure that the time period between two successive data points was always equal.
- **Filtering:** several studies identified that the application of a low pass filter could be useful in reducing the unwanted/non-discriminative information from the signal hence enhance a better performance can be achieved. Therefore, this study carried out with several settings (i.e., 10, 20, and 30) and through experimentation the cut-off frequency of 20Hz achieved the best accuracy (examples of the filtering are shown in Figure 13).

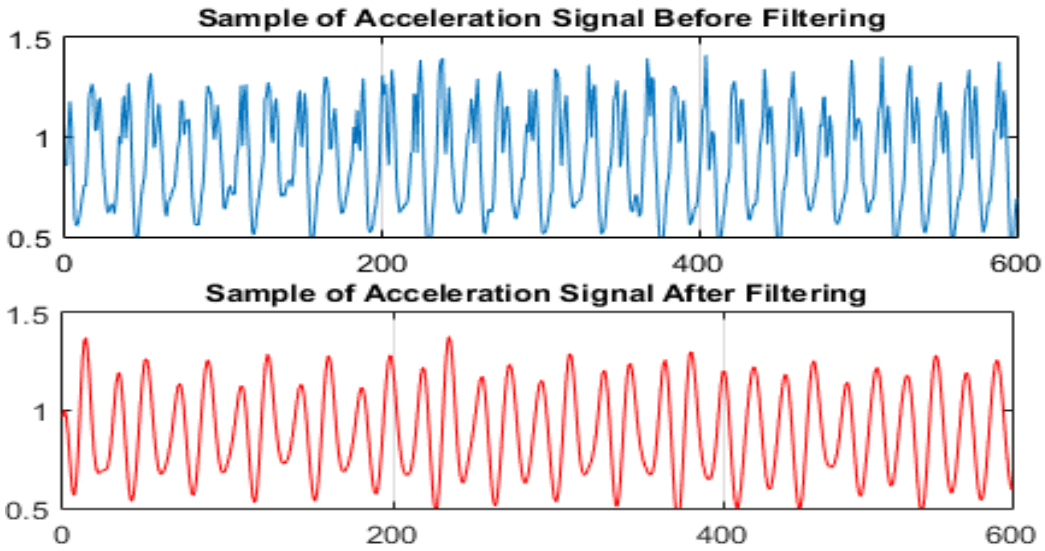


Figure 13: The acceleration signal before and after filtering

- Segmentation: most classification approaches do not directly operate on time-series data and require the data to be represented as a set of samples. As discussed early in chapter 3, there are two main approaches to segment the raw movement data, namely cycle-based and segment-based. The literature shows that the performance varies significantly by using these two methods. The error rate of using cycle-based is considered as high with the EER is ranging from 19% to 33.3%.

  In contrast, the performance of the segment-based method appears to be more effective and stable, with studies reporting EERs between 1.4% and 10%. Therefore, the tri-axial raw format for both accelerometer and gyroscope signals were segmented into 10-second segments, which ensures that each sample includes several movement data and any brief period of non-movement signal (e.g., a pause) will not dominate the sample. This was achieved by using a sliding window approach with no overlapping. Therefore, in total 72 samples were collected for each activity and each user over two different days. Examples of the accelerometer and gyroscope data along the x, y, and z axes of two users are illustrated in Figures 14 and 15 respectively. Discriminating patterns can be clearly observed between the accelerometer

and gyroscope data of the selected two users across the x, y and z axes. Preliminary analysis suggests users do have distinctive movements that can be used to transparently and continuously authenticate individuals.

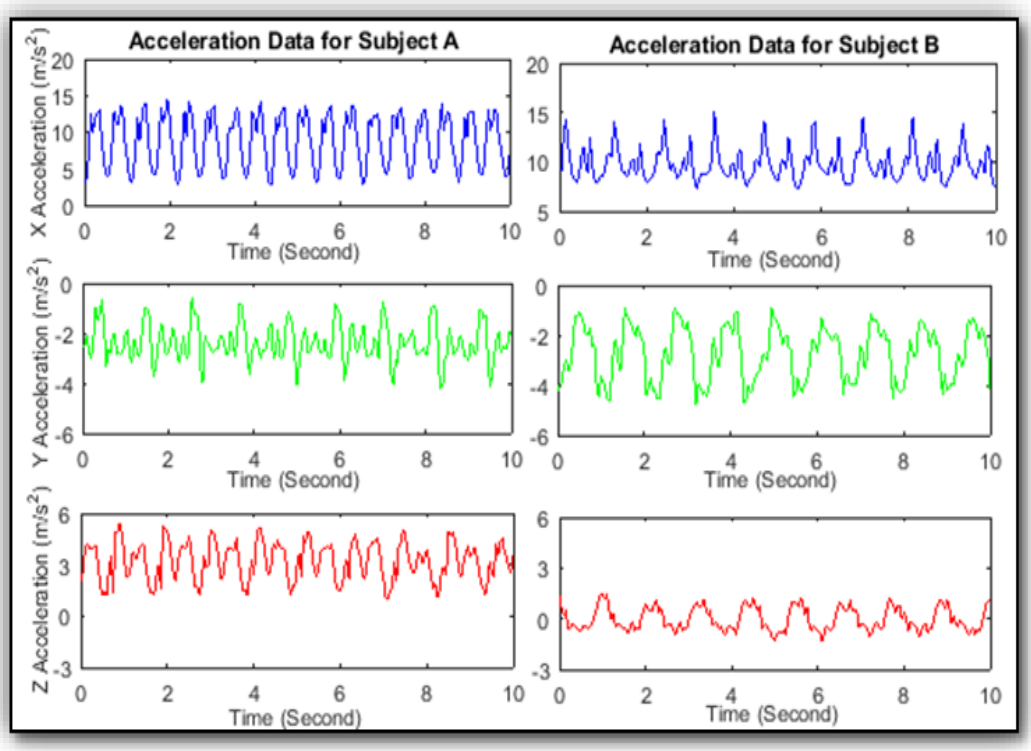


Figure 14: Acceleration sample of three axes for subject A and B

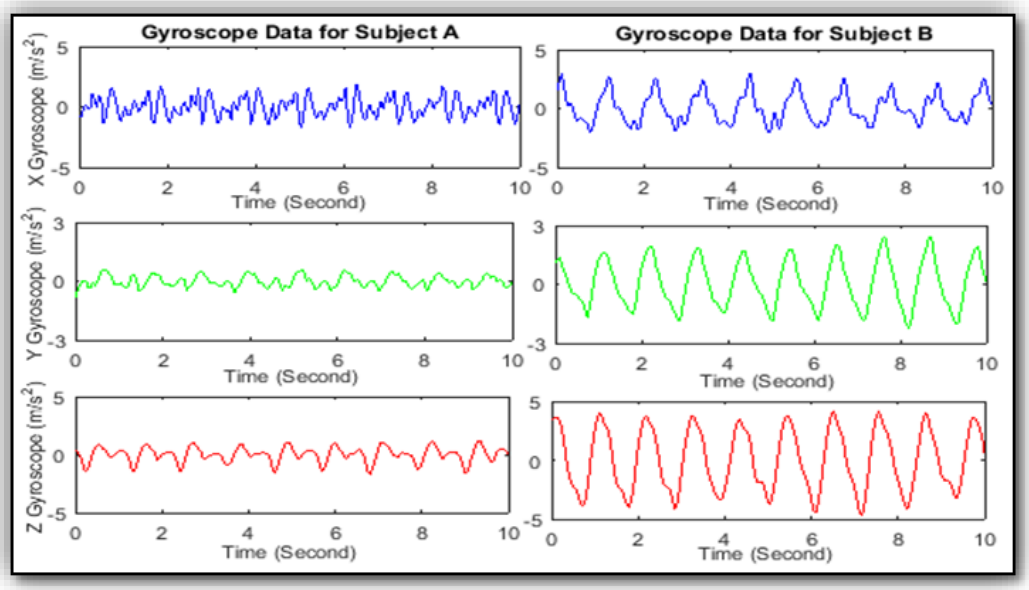


Figure 15: Gyroscope sample of three axes for subject A and B

Further investigation was conducted to show the inter and intra-variance of collected sensor data between users. Ideally, the acquired data from a genuine user is quite similar (i.e., low intra-variance) and different enough from other users (i.e., high inter-variance) to be used for authentication purposes. In order to check the similarity of the user's movement pattern, three gait samples are randomly selected from each user and represented in the following Figures. Each Figure contains the accelerometer signals in three orthogonal directions (x, y, and z).

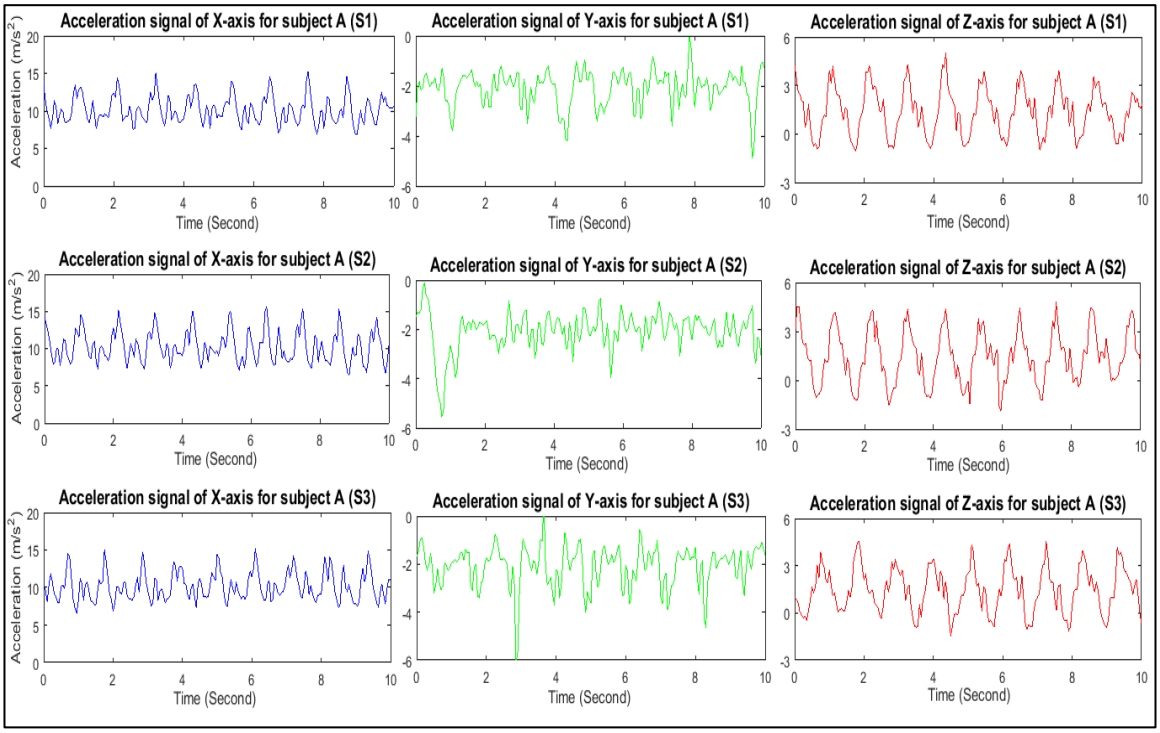


Figure 16: Three acceleration gait samples of three axes for Subject A

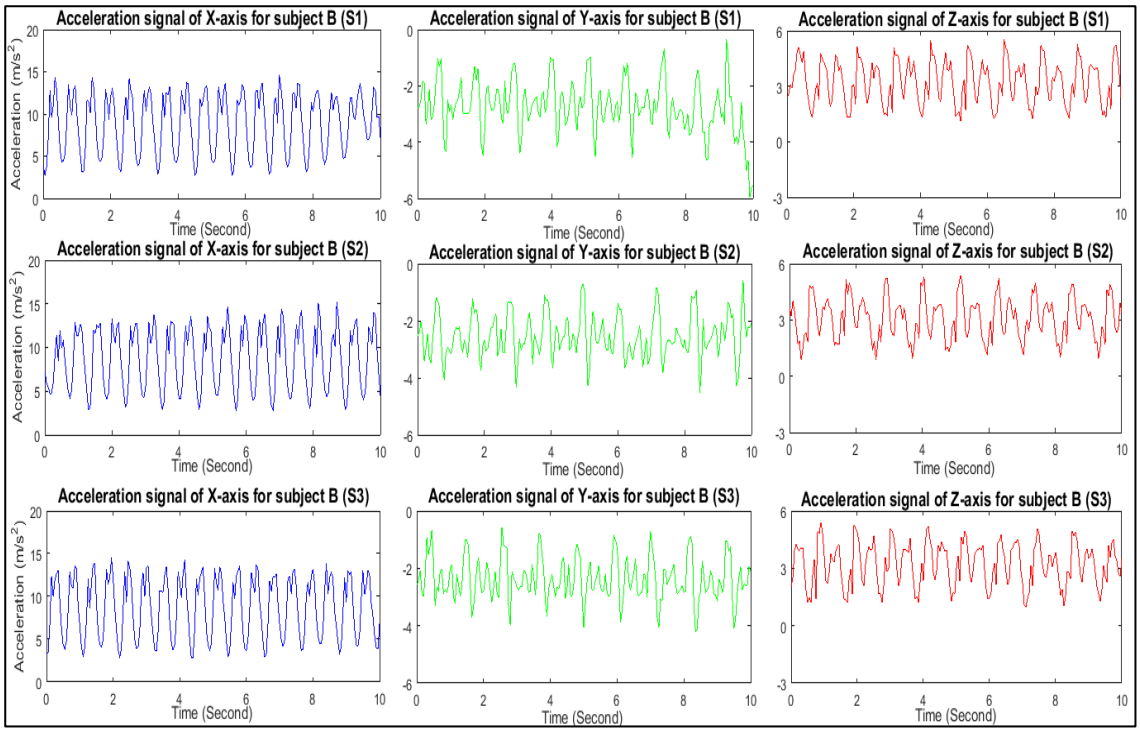


Figure 17: Three acceleration gait samples of three axes for Subject B

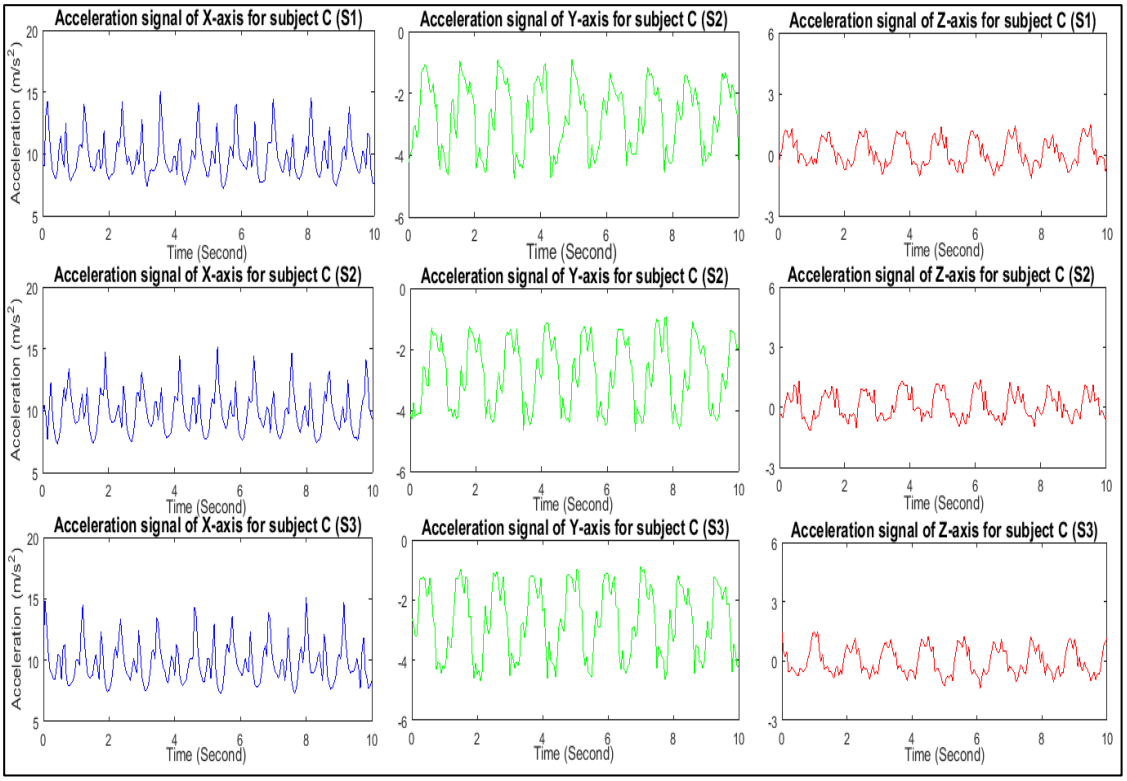


Figure 18: Three acceleration gait samples of three axes for subject C

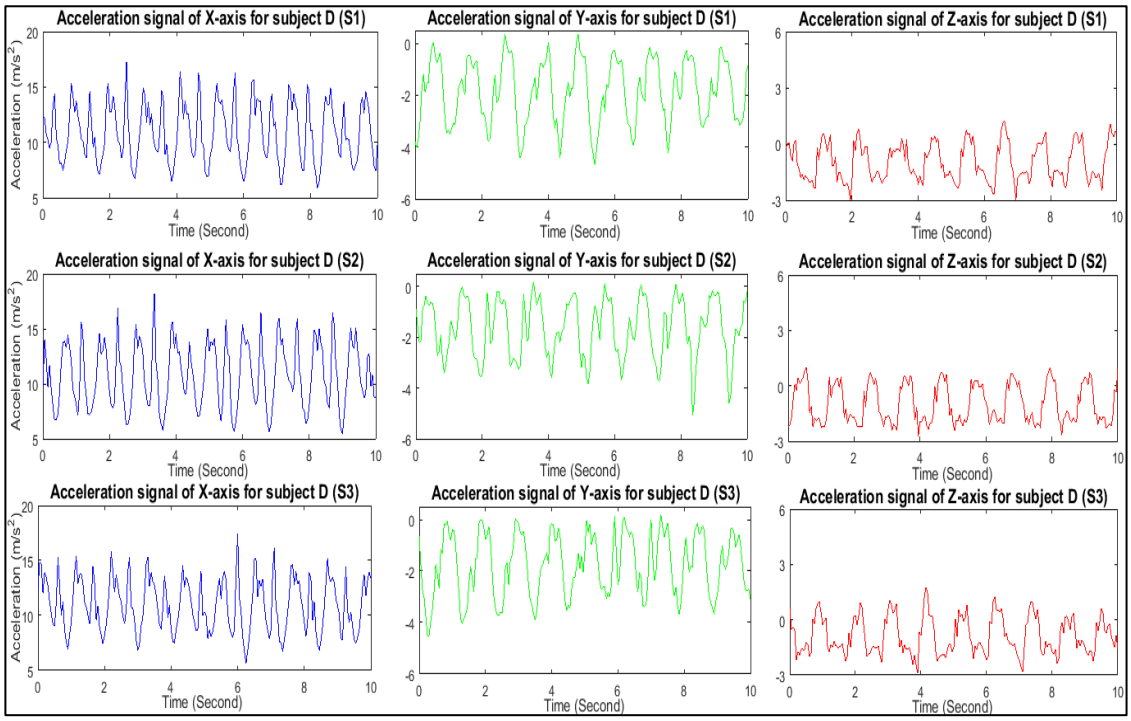


Figure 19: Three acceleration gait samples of three axes for subject D

The previous Figures (i.e., 16, 17, 18, and 19) show that of the three accelerometer parameters measured along each axis (x, y, and z), the variance along y-axes is even for the same user (i.e., high intra-variance) and less distinctive between different users (i.e., less inter-variance). The other accelerometer values along x and z axis do appear to be quite unique among users and nearly similar for the genuine user.

4.3.3 Feature extraction

Feature extraction is a key component of any biometric system and needs to contain the user discriminative information necessary for classification. Therefore, a comprehensive feature extraction process was carried out on both the accelerometer and gyroscope sensor data. Based upon the prior art, features were extracted in both the time and frequency domains and resulted in 140 features (Kwapisz et al., 2010; Nickel et al., 2011b; Nickel et al., 2011c; Nickel et al., 2011d; Ross, A. 2013; Watanabe, Y. 2014; Johnston and Weiss, 2015). These features are the same regardless of whether the sample is being generated from

accelerometer and gyroscope signals. Since most features are generated on a per-axis basis and each sensor has 3 axes, most features are represented by a vector of three values. Several statistical features were also extracted; some of these features are new and have not been included in the prior studies (i.e., interquartile range, skewness, kurtosis, percentile, correlation coefficients). Details of these features (e.g., what they are and how they are calculated) are presented in Table 8, and the digit in brackets specifies the number of generated features for each feature type.

Features	NF	TD	FD	Description
Interquartile range	3	✓	✓	The range in the middle of the data. It is the difference between the upper and lower quartiles in the segment.
Skewness	3	✓	✓	A measure of the symmetry of distributions around the mean value of the segment.
Kurtosis	3	✓	✓	A measure of the shape of the curve for the segment data
Percentile 25,50	6	✓	✓	The percentile rank is measured using the following formula: **R** = (P/100) * (N+1). Where **R** represents the rank order of the values, **P :** percentile rank, and **N** is the total number of data points.
Correlation Coefficients	3	✓	✓	The relationship between two axes is calculated. The Correlation Coefficients is measured between X and Y axes, X and Z axes, and Y and Z axes.
Difference	3	✓	✓	The difference between the maximum and minimum of the values in the segment.
Median	3	✓	✓	The median values of the data points in the segment.
Root Mean square	3			The square root of the mean squared.
Maximum	3	✓	✓	The largest 4 values are calculated and averaged.
Minimum	3	✓	✓	The smallest 4 values are calculated and averaged
Average	3	✓	✓	The mean value of the values in the segment for each axis
Standard Deviation	3	✓	✓	The standard deviation is a measure of how spread the data points from the mean. It is calculated for each axis.
Average Absolute Difference	3	✓	✓	The average absolute distance of all values in the segment from the mean value over the number of data point in the segment (for each axis).
Time Between Peaks	3	✓	-	During the user's walking, repetitive peaks are generated in the gait signal. Thus, the time

				between consecutive peaks was calculated and averaged (for each axis).
Peaks Occurrence	3	✓	-	Determines how many peaks are in the segment.
Variance	3	✓	✓	Average of the sum of the squared differences of each value in the segment from the mean over the segment size (for each axis).
Cosine Similarity	3	✓	-	All pairwise cosine similarity measurements between axes.
Covariance	3	✓	-	All pairwise covariances between axes.
Energy	3	-	✓	The summation of the mean square of each frequency component multiplied by time interval of the signal
Entropy	3	-	✓	Spectral entropy describes the complexity of the signal based on the Shannon entropy
Binned Distribution	30	✓	-	Relative histogram distribution in linear spaced bins between the minimum and the maximum acceleration in the segment. Ten bins were used for each axis
Average Resultant Acceleration	1	✓	✓	For each value in the segment of x, y, and z axes, take the square roots of the sum of the values of each axis squared over the segment size (i.e., 10 seconds)

Table 8: List of the extracted time domain (TD) and frequency domain (FD) features

The process of extracting frequency domain features is somewhat different from the time domain. Before extracting a frequency domain feature, a Fourier transform is applied to the data. A set of frequency domain features are calculated which might be useful to create a discriminative feature vector for each individual.

4.3.4 Feature selection

Feature selection plays a central role in the pattern recognition system, which takes place after extraction and prior to classification. Prior work has highlighted some features are more useful than others and the discriminative ability of features can vary between users Feature selection is used to select feature subset from the entire extracted features through identifying the most optimal and remarkable features for the machine learning algorithms in order to reduce potentially large dimensionality of input data (Hoang et al., 2013; Kumar et al.,

2017). When the feature set size is relatively large, feeding all features to a classifier without selecting of a distinguish feature subset might negatively affect the system performance. Therefore, the feature selection step has become the focus of many research studies in the area of authentication (Nickel et al., 2011b; Nickel et al., 2011c; Hoang et al., 2013; Kumar et al., 2017) with the resultant effect of enhancing performance and reducing the computational complexity of the classifier. Subsequently making it easier to manipulate and calculate feature vectors on processing and battery limited digital devices.

Previous authors have identified that creating a dynamic feature vector for each individual could be beneficial, however, there are limited studies that explored the approach in more details. Moreover, the feature selection approaches that have been presented so far do not seem to accel in terms of performance. To this end, this study carried out an exhaustive exploration of data using descriptive statistics for better understanding the nature of features and to explore the relationship between inter and intra variance that might exist. Thereafter, the output of this exploration process was used to develop a novel dynamic feature vector algorithm to see how that would impact the system performance. For each individual, a unique feature subset was generated (i.e., creating a dynamic feature vector that contains distinctive features for each user). This is achieved by calculating the mean and standard deviation (STD) for each feature individually for all users. Thereafter, comparison the authorized user's results against impostors to select the feature set with the minimal overlap. In other words, for each feature, a score is calculated based upon the following condition:

- If the mean of imposter's feature is not within the range of the mean +/- STD of genuine, add 1 to the total score.

- Dynamically select the features according to their score order from high to low. The highest means less overlap between imposters and the genuine user as shown in Figure 20.

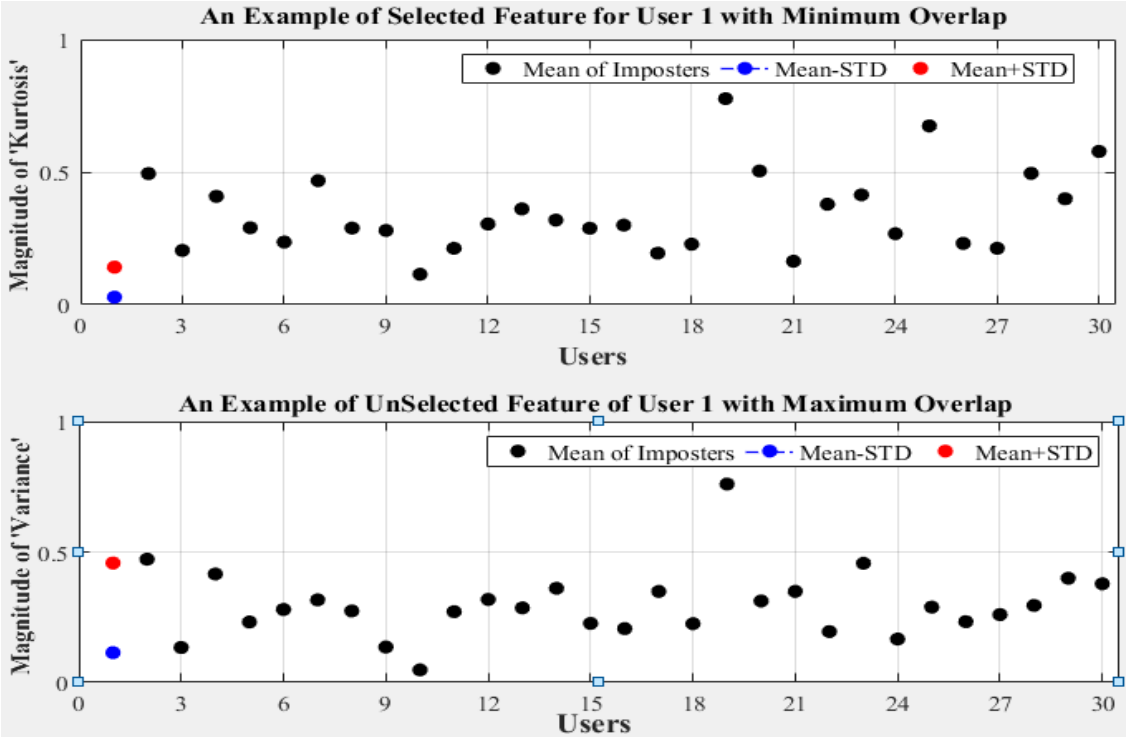


Figure 20: The effect of the dynamic feature selection approach

Figure 20 show an example of the automatic feature selection approach that was utilized in this study; this procedure was carried out for all the time domain features and retained the most predictive feature subset for each individual. From the presented information in Figure 20, it is apparent that the Kurtosis feature for user1 has less overlap than Variance feature, which means this feature would be used to create the reference template for user1 as it shows low intra-variance and high inter-variance. Although the proposed dynamic feature vector approach successfully maximized the system performance and reduced the feature vector size, it is not a definitive solution for the problem and a comprehensive evaluation for different feature selection approaches is required to find out the optimal method for TAS.

4.3.5 Experimental Procedure

The aim of the biometric-based authentication or verification is to determine if a system can classify a user correctly (a “genuine” user) or as an imposter. This study utilized two approaches namely, generic and activity-based models. Both models, require separate training data and applying different mechanisms. The former used the whole collected data (i.e., NW, FW, TypePC, GameM, and TypeM) without considering the user’s activity type hence, one classifier was created for each individual.

In contrast, multi- classifier/algorithmic (i.e., the more realistic and novel technique) was used in the latter; this was achieved by generating a separate model for each of the aforementioned activities hence, five models were created for each user. It is argued given the variability of the signal data, creating specialized models based upon activity will exhibit better recognition performance than a generic model. Therefore, it was necessary to design and develop a comprehensive experiment that confirms this assumption.

Once these models were prepared, the reference and testing templates were created under two different scenarios (i.e., SD, and CD). In the SD scenario, data set was divided into two parts: 60% was utilised to train the classifier while the remaining 40% was used to evaluate the performance. The reason for selecting this ratio (i.e., 60% versus 40% for the training and testing respectively) is to ensure that the classifier is trained with sufficient representative samples and evaluate the robustness of the proposed system by using fairly acceptable testing samples. To test the system under the Cross Day (CD) scenario, the data of the first day was used for training and the evaluation was carried out by employing the second day data.

To train a classifier, a reference template needs to be created for each individual; the user's reference template consisted of samples from the user itself and from other users (i.e., imposters). To distinguish between the genuine and imposter feature vectors, the genuine samples were labelled as 1 and 0 was used to label the remaining samples. To this end, 21 samples of the genuine user samples was selected and 295 random samples (i.e., 5*59) from the imposter group were used to build the user's profile under the SD scenario. The criteria for selecting this proportion is that it showed low EER (compared to 200, 400, 500 of imposter samples). For the CD scenario, the same proportion of the imposter's samples (i.e., 300 samples) were used and the only difference was the amount of the legitimate user samples (i.e., 36 samples were utilized). This procedure was repeated for all users (so in total 60 tests), and different legitimate user was selected for each test.

Once the user's templates were created, a Feedforward Multi-Layer Perceptron (FF MLP) neural network was used as the default classifier; this is because neural network is less sensitive with the variation of the user's arm pattern and require less training data compared to SVM, HMM, and K-NN (Nickel et al., 2011c; Nickel et al., 2012b). Moreover, it showed the possibility to build a high level of distinctive reference template for each individual and hence reliable performance for the proposed system (Kwapisz et al., 2010; Watanabe, Y. 2014; Johnston and Weiss, 2015). For each experiment, four different FF MLP neural network training sizes were evaluated (i.e., 10, 15, 20, and 25) with each being repeated 10 times in order to account for errors that could occur due to the random setting of the neural network weights. In order to complete these results, 259, 200 tests were carried out (i.e., 4320 tests for each individual including the variation of the network and feature subset* 60 users). Nevertheless, the presented results in this chapter are

the key findings for the most important part of the conducted experiments. The results presented in this study were based on using FF MLP neural network of size 10 as it showed the lower EER for all the collected activities.

4.4 Results

After research questions of the prior work were already identified in the previous section, several extensive experiments were conducted. Details of the results and deep analysis of the conducted experiments are described in the following subsection.

4.4.1 Time VS Frequency Domain Features and Sensor Selection

Selecting a set of features that are unique and distinguish can result in a better classification and easier to manipulate small feature subsets on digital devices. However, the majority of the prior acceleration-based biometric studies have not considered the effect of time domain (TD) and frequency domain (FD) on the system accuracy. Therefore, to avoid negative effects on the system performance, the EERs of both features (i.e., TD and FD) were calculated and presented in Table 9 (using the SD scenario, the acceleration (Acc) and gyroscope (Gyr) of the NW activity).

Feature type	NF	EER (%)	
		Acc	Gyr
All Features	132	0.18	3.37
Time domain	88	0.15	3.73
Frequency domain	44	3.09	12.69

Table 9: The EERs of using all features, time and frequency domains

The previous studies have already demonstrated that more features that are incorporated usually degraded the classifier's accuracy. This is because some of these features could be irrelevant and/or redundant. It is clear that good

performances were achieved by using the TD features and all feature sets (i.e., little difference in results is observed between the two sets). By using the FD features alone, reasonable performance is obtained; however, its performance is far less promising in comparison with the results of using TD features alone, suggesting that FD features add little contribution towards the classification process and even negative impact on both sensors (i.e., the EERs were significantly increased to 3.09% and 12.69% for accelerometer and gyroscope respectively comparted to 0.15% and 3.73% when the time domain features were used). Moreover, it is difficult for the system to compute these features in real time on the smartphones and/or smartwatches due to their complicated calculation and the limited resources of these devices. Given the fact that detecting redundancies features makes the system more efficient, therefore, only the TD features (i.e., 88 features) were used in the subsequent experiments.

Although sensor based-authentication systems could be implemented using accelerometer and/or gyroscope as the source triaxial sensor, further analysis was carried out to select the best sensor that offers lower error rates. As shown in Table 10, the evaluation results overwhelmingly support the use of the accelerometer sensor alone for smartwatch-based user authentication systems (with EERs of 0.15% and 0.93% for the SD and CD scenarios respectively). These errors are increased into 3.73% and 8.29 % of EER by using the gyroscope data of both scenarios respectively. Another analysis was conducted to reflect the EER spread within the population and the findings are presented in Figure 21. It can be seen in Figure 21, the EERs of using the gyroscope signal were significantly increased for the majority of users (or nearly similar) compared to the acceleration data. As a result, all the subsequent results are based on the use of the acceleration data only.

Evaluation Scenario	Sensors	TD Features
SD	Acc	0.15
CD	Acc	0.93
SD	Gyr	3.73
CD	Gyr	8.29

Table 10: The EERs of using the Acc and Gyr

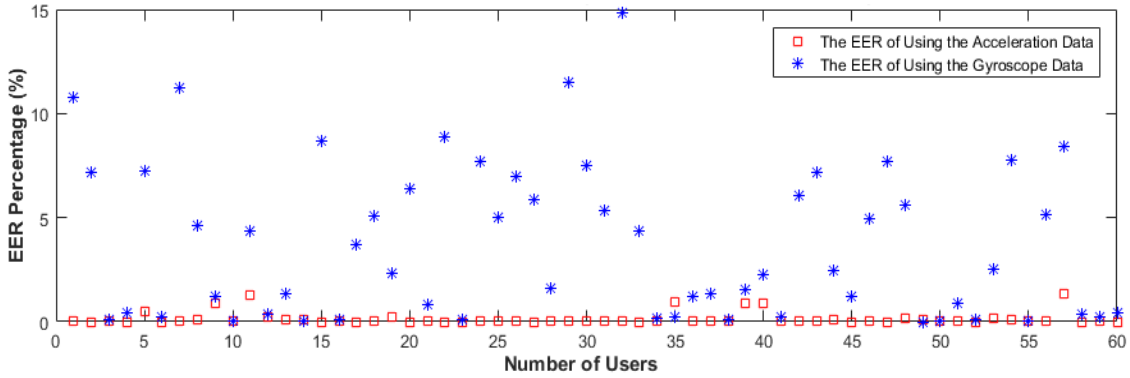


Figure 21: The EERs of the Acc versus Gyr sensors separated by users

4.4.2 Single classifier versus multi-classifier/algorithmic

To evaluate the efficiency of the generated reference templates of individuals, two experiments were carried out, namely, single-and multi algorithmic. The first experiment (i.e., single classifier) utilized the generic model, which contains samples from all five activities but with the activity label removed. The second experiment (i.e., multi-classifier) evaluated by using the activity-based model, which contains five subsets, each contains the user's movement data of single activity only (i.e., NW, FW, TypePC, GameM, and TypeM).

So far, all the conducted analysis was based upon using all the extracted (i.e., time and/or frequency domain features). However, it is important to optimize the user's authentication model by selecting the most discriminative feature subset. Moreover, as mentioned earlier that the CD scenario is the most reliable test for any behavioral biometric-based user authentication systems. Therefore, Figure

22 depicts the EERs of using single classifier approach, CD scenario, and the effect of the proposed dynamic feature vector approach.

As demonstrated in Figure 22, the EERs become flat (ranging from 7.1% to 7.6%) between 50 to 88 features suggesting little additional value over a feature length of 50. Shorter feature lengths do have a significant impact on the performance – possibly due to the noisier feature vector based upon all activities. The experimental set up of this approach (i.e., single classifier approach) bears a close resemblance to the prior work by Kwapisz et al., (2010). Nevertheless, they have reported low accuracy (i.e., about an EER of 19% compared to 7.1% in this study). The significant improvement could be the result of creating a complex and discriminative feature vector for each individual independently and to the selection of appropriate classifier (i.e., FF MLP neural network). It is clear that the reported results of utilizing the single algorithmic in this study greatly outperforms the prior work. Moreover, these findings are more reliable due to the utilization of the CD test (which is realistic evaluation) compared to the prior work that used the same day scenario (i.e., SD test).

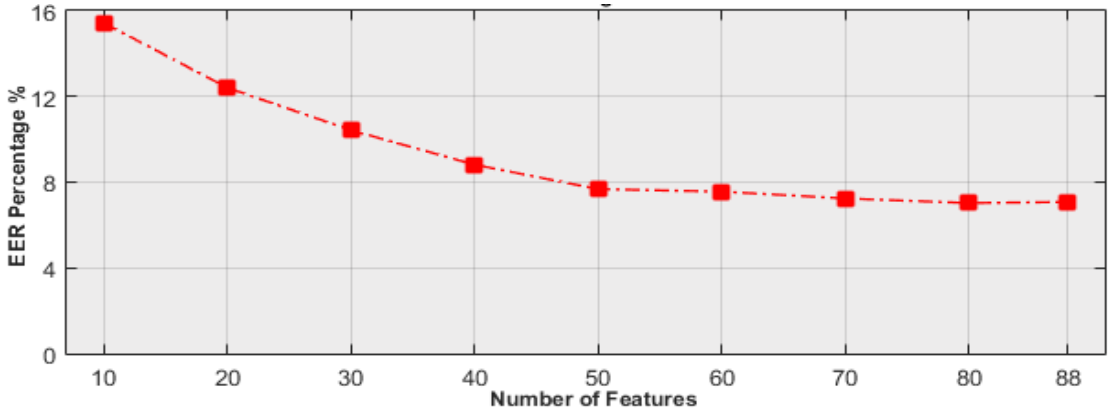


Figure 22: The EERs of using generic authentication model

Although, the findings appear to be good enough to identify unique arm pattern for each individual, noting the system performance might worsen when using real life data due to a higher degree of variability in the signal data. Therefore, the

most realistic or practical experiment is to detect the user's activity and creating multiple models, each trained with data of a specific activity. Table 11 shows the benefit of generating the activity- based authentication model and its leverage on the system accuracy. As hypothesized, the results demonstrated that activity-based authentication model greatly enhanced the recognition rate (i.e., the EER dropped drastically from 7.03% to the range of 0.69%- 5.81%). This substantiates previous findings in the literature that showed the accuracy of using the activity-based authentication model could improve the system accuracy. Nevertheless, the evaluation results of this study are greatly surpassed the prior art that reported EERs in the range of 5.7% to 33.3% (as shown in Table 12); moreover, these studies were based upon collecting limited activities (specifically normal walking and gesture activities).

In contrast, only two studies utilized the fast walking activity for the authentication purpose and obtained poor results, at best an EER of 17.4%, while the proposed system of this study achieved promising performance (i.e., 3.16% EER). The findings showed that some activities performed better than others (especially the gait activities that reported EERs of 0.69% and 3.16% for the NW and FW accordingly). Nevertheless, the typing and game activities still highly recommended for the use of TAS due to their high classification performance (i.e., at best EERs of 4.94%, 5.81%, and 4.54% for the TypeM, TypePC, and GameM respectively). The performances of TypeM and GameM are slightly superior to TypePC. This could be due to the position of the user's hand being not fixed during the TypeM and GameM activities compared to TypePC where the hand position was fairly static. Thus, more differential movement data can be observed from the typing or interacting on a smartphone touch screen. The obtained results

offer vital evidence that the collected activities of this study have the potential to accurately recognize the legitimate user in a transparent and continuous fashion.

Activity	10 Features	20 Features	30 Features	40 Features	50 Features	60 Features	70 Features	80 Features	88 Features
NW	4.68	2.39	1.43	0.9	0.84	0.83	**0.69**	0.77	0.93
FW	5.42	3.92	3.63	4.17	3.56	3.32	**3.16**	3.40	3.90
TypeM	5.97	5.92	5.93	5.69	5.04	**4.94**	5.57	5.60	5.69
TypePC	8.12	7.21	6.98	6.45	**5.81**	5.85	5.92	5.91	6.02
GameM	4.97	4.82	4.83	4.79	4.62	**4.54**	5.17	5.80	5.61

Table 11: EERs of using activity-based user authentication model for different activities

Study	EER (%)	Users	Activity	Device
Derawi et al. (2010a)	20	51	NW	Mob
Nickel et al. (2011a)	10.4	51	NW	Mob
Nickel et al. (2011c)	10	36	NW	Mob
Nickel et al. (2011c)	17.4	36	FW	Mob
Nickel et al. (2011d)	21.7	48	NW	Mob
Nickel and Busch (2011e)	6.2	48	NW	Mob
Hestbek et al. (2012)	10	36	NW	Mob
Nickel et al. (2012b)	8.8	36	NW	Mob
Muaaz and Nickel (2012)	29.4	48	NW	Mob
Muaaz and Nickel (2012)	33.8	48	FW	Mob
Muaaz and Mayrhofer (2013)	33.3	51	NW	Mob
Muaaz and Mayrhofer (2014)	19	35	NW	Mob
Damaševičius et al.(2016)	5.7	14	NW	Mob
Shen, et al. (2017)	4.9	102	Ges	Mob
Junshuang Yang et al. (2015)	3.3	26	Ges	SW
Lewis et al. (2016)	22	5	Ges	SW
Kumar et al. (2016)	86.8 CCR	13	NW	SW
Shrestha et al. (2016).	8.7	18	NW	SW
Xu et al. (2017)	96 CCR	20	NW	SW
Liang et al. (2017)	4	20	Ges	SW
Griswold-Steiner et al. (2017)	12.5	20	Ges	SW

Table 12: Comprehensive analysis on gait authentication using mobile and smartwatch sensors

Regarding to the classification accuracy of the previous gesture-based authentication studies, remarkable recognition rates were achieved ranging between 3.3% and 4.93% of EERs (apart from Lewis et al., (2016)) that showed a high EER of 22%). Results from comprehensive evaluations for the TypeM, TypePC, and GameM activities were consistent with the prior findings that utilized different gestures (at best EERs of 4.94%, 5.81, and 4.54% for the aforementioned activities respectively). However, a fair comparison is required as this study utilized certain activities that are non-intrusive, frequently used, and more natural.

In contrast, serious criticisms of the literature are capturing gestures that tend to be intrusive, do not offer continuous authentication, not realistic (i.e., complicated gesture such as a punch), and/ or not robust against imitation attack scenario. To show the efficiency of individual users' performance, a comprehensive analysis for each activity was conducted (as shown in Figure 23). Figure 23 proves that each individual has a distinctive arm pattern, thus one third of the users reported an EER of around 0% for all the collected activities (e.g., 3, 4, 7, 12, 17, 19, 21, and 32), while the rest of the users reported an average of low EER in the range of 0-10%, apart from users 6, 8, 22, 23, 27, 46, and 50 that reported high EERs for particular activities (i.e., typing on a smartphone touch screen and/or PC).

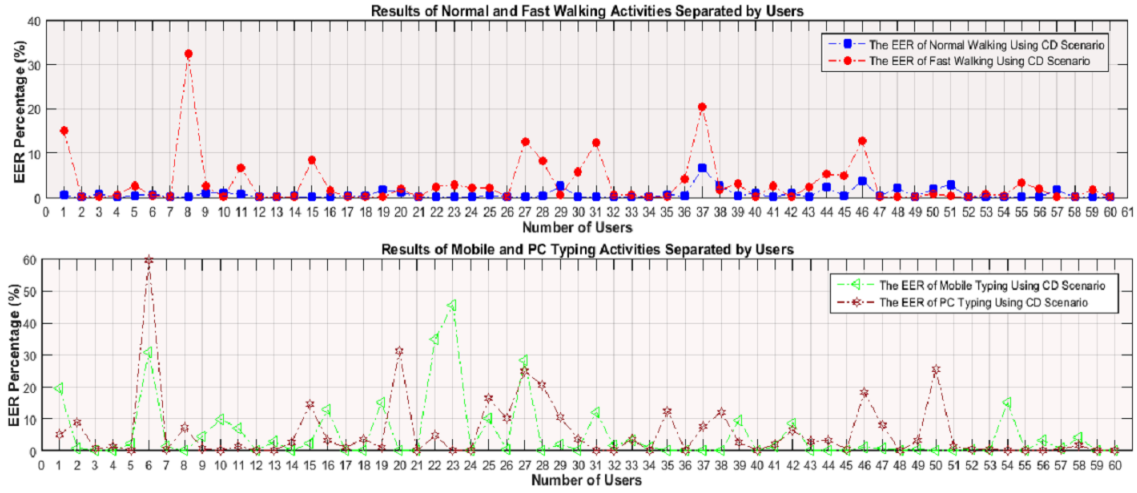


Figure 23: The EER of all activities separated by users

4.4.3 Single Vs Cross Day Scenario

While this study was in progress, the majority of prior work relied upon data from single session for both training and testing, which is not realistic and can lead to overly optimistic performance results. Much of the other acceleration-based recognition studies suffer from the same limitation and in cases when the CD scenario is considered, the evaluation is often either completed improperly or the results are poor. To overcome some of these past problems, a comprehensive experiment was carried out by training and evaluating the proposed system on data from across different days. Moreover, this section highlights the benefit of identifying the optimal features by creating a dynamic reference template for each individual. Two experiments are presented under two different scenarios; namely, Same-Day (SD) and Cross-Day (CD). The first experiment used all the extracted features (i.e., 88 unique features) while for comparison a more selective set of minimal features are used in the second experiment.

Activity	**Evaluation Scenario**	**10 Features**	**20 Features**	**30 Features**	**40 Features**	**50 Features**	**60 Features**	**70 Features**	**80 Features**	**88 Features**
NW	SD	1.13	0.78	0.24	0.26	0.27	**0.13**	0.20	0.16	0.15
FW	SD	1.55	0.80	0.62	0.36	0.35	0.32	**0.28**	0.32	0.31
TypeM	SD	2.40	1.76	1.38	1.18	**0.99**	1.21	1.24	1.39	1.43
TypePC	SD	2.28	1.36	1.38	**1.15**	1.15	1.30	1.39	1.33	1.52
GameM	SD	2.40	1.76	1.38	1.18	**0.89**	1.20	1.14	1.20	1.33
NW	CD	4.68	2.39	1.43	0.9	0.84	0.83	**0.69**	0.77	0.93
FW	CD	5.42	3.92	3.63	4.17	3.56	3.32	**3.16**	3.40	3.90
TypeM	CD	5.97	5.92	5.93	5.69	5.04	**4.94**	5.57	5.60	5.69
TypePC	CD	8.12	7.21	6.98	6.45	**5.81**	5.85	5.92	5.91	6.02
GameM	CD	4.97	4.82	4.83	4.79	4.62	**4.54**	5.17	5.80	5.61

Table 13: The Impact of the SD, CD scenarios, dynamic feature selection technique on the performance in details

The results in Table 13 show if 10 seconds interval of the accelerometer data is evaluated properly, the proposed authentication system was able to achieve very high accuracy for both scenarios. For the SD scenario, EERs ranged between 0.13% (for NW) 1.15% (for TypePC) by using 60 and 40 features respectively. Improvement on performance is obvious when comparing the outputs to the previous mobile-based acceleration studies under the SD scenario (i.e., EERs ranging from 5.7% to 1.4%). Moreover, the system is still able to effectively recognize the user's arm pattern with low EERs of 0.78%, 0.80%, 1.76%, 1.36%. and 1.76% by using only 20 features for the NW, FW, TypeM, TypePC, and GameM activities respectively. These results suggest that the selected feature subset was highly discriminative, which was based upon automatic selection of the most relevant or optimal attributes.

In addition to the SD, the more realistic test (i.e., the CD scenario) was also applied. As expected the system performance decreased under the CD methodology; this is because the behavioural biometric can be affected by several factors such as mood, clothes, tiredness, and permanence. Nonetheless, the reported CD results are still promising in comparison with the prior work that reported EERs in the range of 5.7% - 33.3% (for the gait data) and 12.5% (for the hand writing activity). Moreover, the proposed feature reduction method has further strengthened the author's confidence by minimizing the number of features and maximizing the discriminative information. The best findings of the captured activities were obtained by utilizing feature subset size ranged between 50 to 70 features. With respect to the feature subset size, the findings in Table 13 show that the SD test for the all activities, apart from the FW, requires less features than the CD (i.e., 60, 50, 40, and 50 features for the NW, TypeM, TypePC, and GameM respectively). This could be explained because the user's

arm pattern could vary or be inconsistent over the time, hence more features are required for individual to be identified for the CD scenario.

With the aim to understand how individual user performed, a most common activity was selected (i.e., NW activity) and results on each user's acceleration for both SD and CD scenario are presented in Figure 24. As shown in Figure 24, high level of performances (i.e., in the range of 0-2% EER) were obtained for 90% users (apart from users 31, 37, 38, 42, 48, and 51). More than 25% of the participants reported 0% of EERs such as users 2, 4, 13, 15, 17, 21, and 27; this suggests that users have consistent and distinctive set of acceleration pattern characteristics.

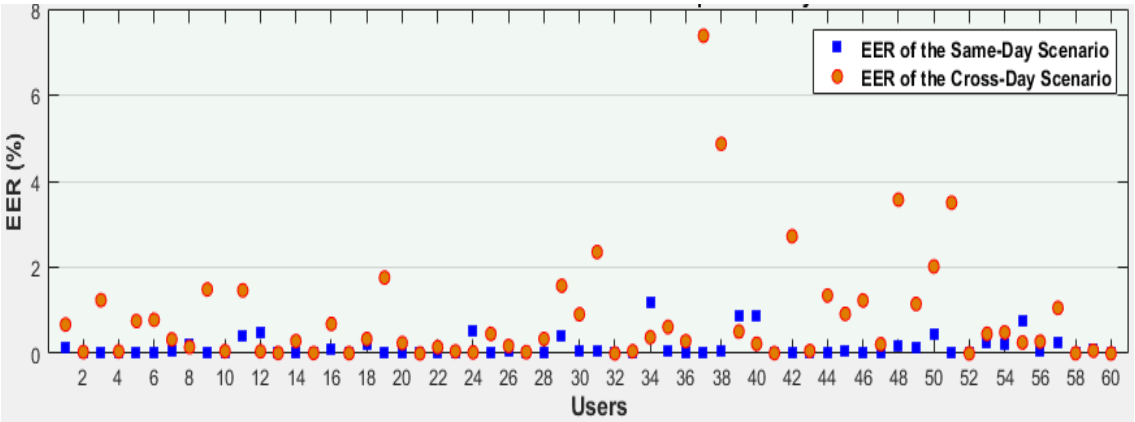


Figure 24: The EERs of the SD and CD scenarios for each user individually

So far, the presented results in Table 13 were based upon static feature vector size for all users (i.e., the user's reference template size was fixed for all users such as 70 features for the NW) although the composition of the feature vector was dynamic. Therefore, further analysis was carried out to optimize the feature vector for each user independently. For example, the reference template of user1 might contain 40 features while 20 or 30 accelerometer features utilize to form the reference template of user 2. The aim of this investigation is to determine whether the optimized feature vector for each user independently can further improve the system accuracy. Moreover, to find out the requisite number of

features that build a robust reference template for each user independently. Table 14 shows a comparison of using static and optimized feature vectors for each activity.

Activity	Scenario Evaluation	EER (%) for SFV	EER (%) for OFV
NW	SD	0.13	0.05
FW	SD	0.28	0.14
TypeM	SD	0.99	0.5
TypePC	SD	1.15	0.3
GameM	SD	0.89	0.25
NW	CD	0.69	0.29
FW	CD	3.16	1.31
TypeM	CD	4.94	2.66
TypePC	CD	5.81	3.85
GameM	CD	4.54	2.3

Table 14: The system performance using the static feature vector (SFV) and optimized feature vector (OFV)

The finding in Table 14 confirms the hypothesis that creating optimized feature vector for each user independently might greatly reduce the EER; the proposed technique (i.e., optimized feature vector) clearly has an advantage over the static feature vector. As expected the EERs of the SD evaluation were decreased for all activities. Similarly, applying the more realistic test (i.e., CD scenario) revealed a significant improvement (i.e., a minimal of 34% and up to nearly 65%) over the classification performance, at best EERs of 0.29%, 1.31%, 2.66%, 3.85%, 2.3% for the aforementioned activities (compared to 0.69%, 3.16%, 4.94%, 5.81%, 4.54% of using static feature vector method). The possible explanation for the significant improvement on the system accuracy is that the movement pattern of some users requires fewer features to produce the lowest EER and vice versa (i.e., the user's arm movement for particular users is inconsistent hence, more features are required to obtain the optimal or lowest EER). For example, fixing the size of the reference template for all users (60 features) might negatively affect the overall system accuracy. To support the above assumption, further

tests were undertaken for each activity and the evidence presented in Figures 25, 26, 27, 28, and 29 for the NW, FW, TypeM, TypePC, and GameM respectively.

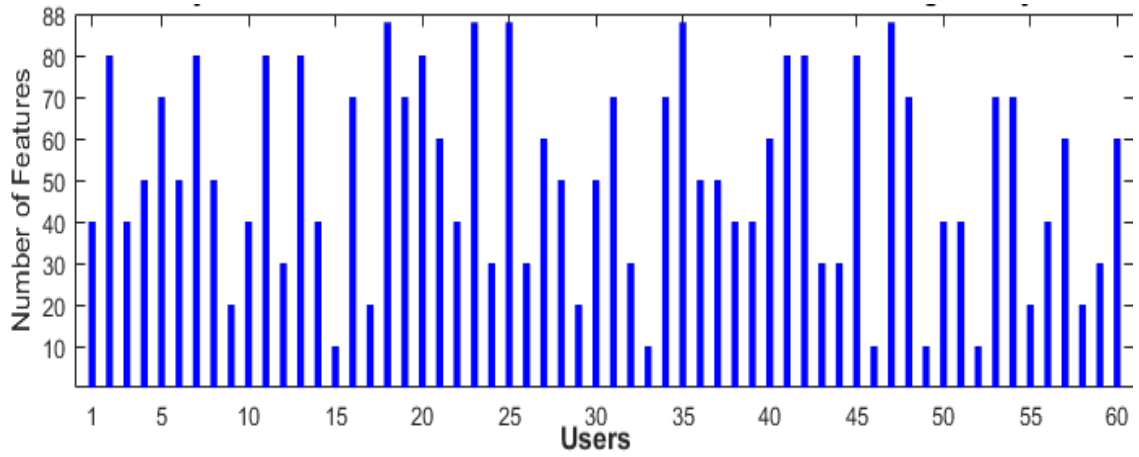


Figure 25: The optimal feature vector size of each user for the NW activity

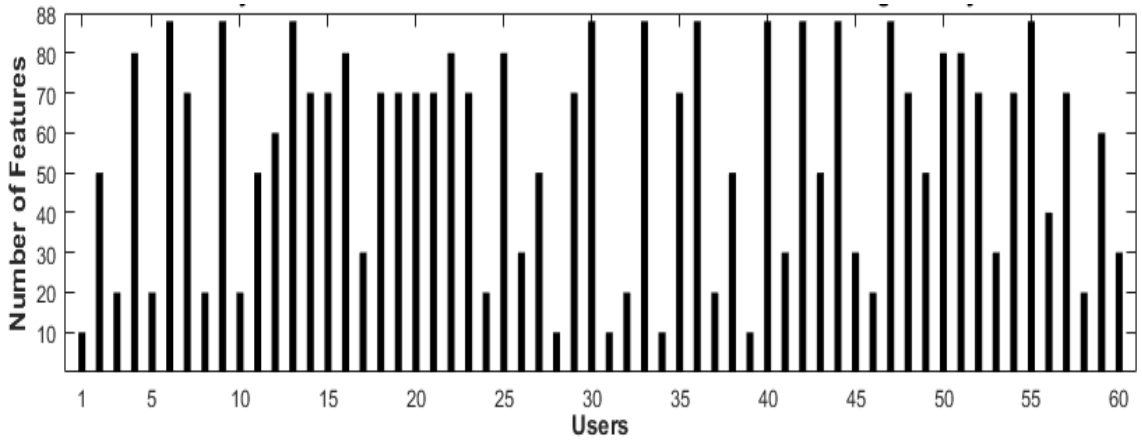


Figure 26: The optimal feature vector size of each user for the FW activity

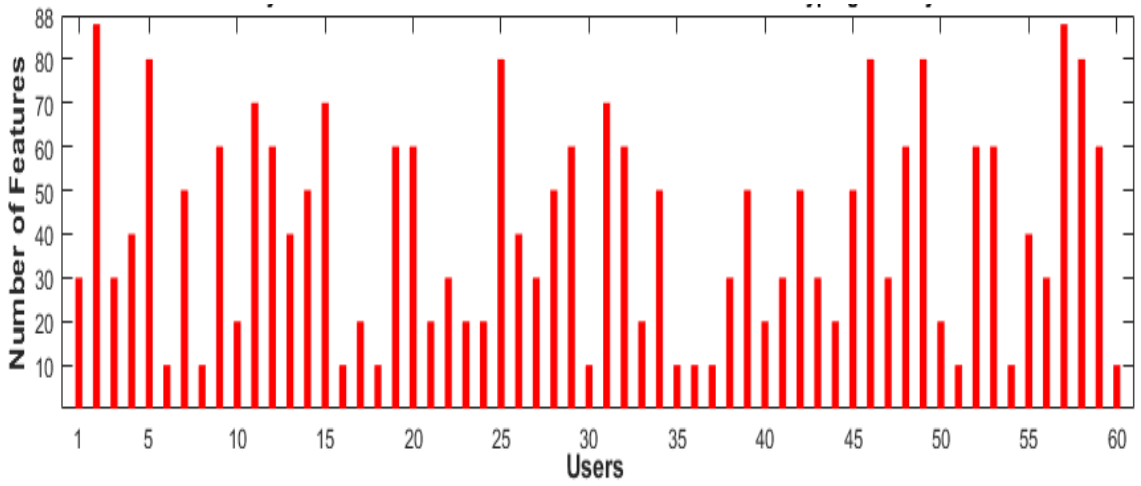


Figure 27: The optimal feature vector size of each user for the TypeM activity

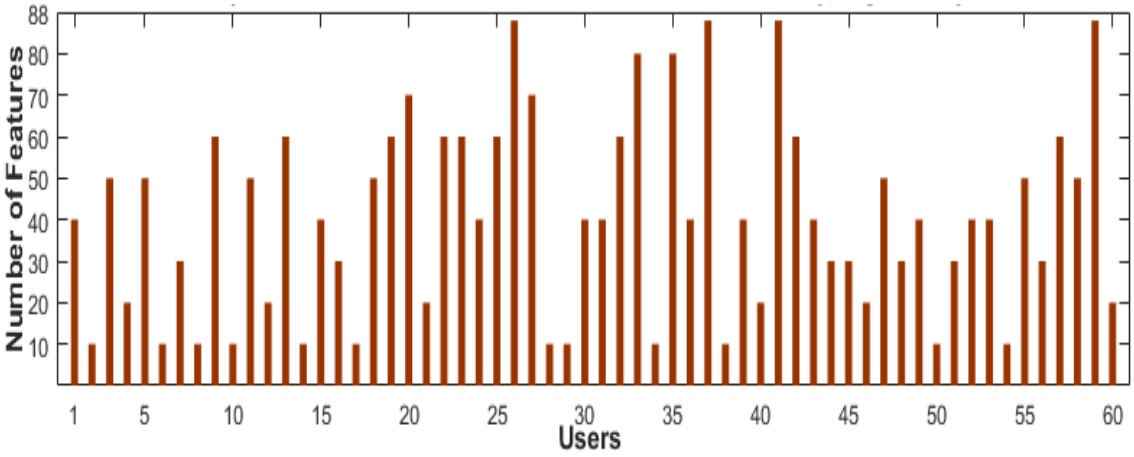


Figure 28: The optimal feature vector size of each user for the TypePC activity

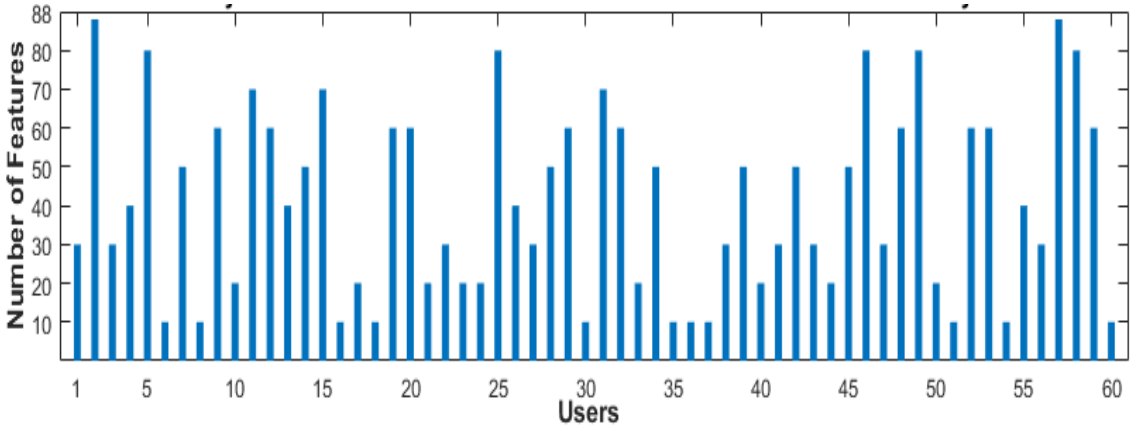


Figure 29: The optimal feature vector size of each user for the GameM activity

Apart from the improvement in the system performance of using the optimization technique, Figures 25 and 26 show the gait templates size was reduced for more than half of users. For example, FW-based model of users 3, 5, 8, 10, 24, 32, 37, 46, and 58 was created by utilizing only 20 prioritized features and even less features were used for users 1, 28, 31, 34, and 39 (i.e., 10 features). In contrast, other users such as 4, 6, 9, 13, 16, and 22 required more features (i.e., 80 to 88 features) to produce the lowest EER. The possible explanation is the walking pattern of these participants was varied or inconsistent over the time. Therefore, more features are required to generate a reference template that is robust to impersonation attacks and effectively identify the user's identity. For the remaining activities (i.e., TypePC, TypeM, and GameM), Figures 27, 28, and 29 show a clear trend that the proposed DFVS technique was successfully reduced

the vector size of more than two thirds of users. This could be due to that less movement data could be obtained for these activities compared to the walking activities hence, less feature subset was able to form the user's reference template.

4.5 Discussion

The conducted analysis seeks to address the following questions:

- Can smartwatches provide a more reliable and consistent user's motion signal than smartphones could?
- What is the impact of the time and frequency domain features on the system performance?
- Can activity- based user authentication model have a positive effect on the verification accuracy?
- What is the influence of applying CD scenario on the system accuracy?
- Does the proposed feature selection approach have a positive effect on the gait biometric performance?
- Does the optimized feature vector have further improvement on the system accuracy?

The obtained results suggest that smartwatches have the ability to capture more accurate personal data than smartphone could. Moreover, the experimental analysis reveals that activity- based user authentication is a highly efficient and recommended to be used for verifying the user in a transparent and continuous manner. Although features were extracted in both time and frequency domains, the findings in Table 9 supports the use of TD features alone due to their high correlated and distinctive characteristics for sensor-based user authentication

systems. Moreover, the complexity of calculating FD features make them less practical especially for mobile devices and smartwatches that have limited resources.

When it comes to build the feature vector of individuals, a comprehensive experiment was carried out to determine the necessity of creating multiple reference templates for each user, each contains feature subset of specific activity. The results in section 5.4.2 support the usage of activity-based user authentication model rather than utilizing a generic model for all data. For example, the best EERs were 0.69%, 3.16%, 4.94%, 5.81%, and 4.54% for the NW, FW, TypeM, TypePC, and GameM respectively (compared to 7.02% when a generic-based model was applied). These errors were significantly reduced into 0.29%, 1.31%, 2.66%, 3.85%, and 2.3% for the aforementioned activities by utilizing DFVS technique.

As expected, the results demonstrate that biometric performance is degraded under the more realistic evaluation scenario (i.e., CD scenario). However, the levels of performance being achieved are excellent in comparison to other research on behavioral biometrics and transparent authentication. Using the CD scenario resulted in EERs of 0.29%, 1.31%, 2.66%, 3.85%, and 2.3% for the NW, FW, TypeM, TypePC and GameM respectively against 0.05%, 0.14%, 0.5%, 0.3%, and 0.25% utilizing the SD test for the above activities. Compare with the prior art, this study utilized the biggest dataset in the domain and achieved overwhelming results.

Further influencing factors on the biometric system performance is the selected feature subset; selecting unique features for each user would improve the results and reduce the complex computations on the smart devices, which has limited processing resources. Therefore, a feature selection approach of any

mobile/smartwatch-based biometric system needs to be sophisticated enough before the classification phase takes place. As expected, the proposed feature selection approach in this study, which was based on creating a dynamic feature vector for each user, successfully reduced user's feature vector size. The reported EERs of using static feature vector for the CD evaluation were 0.69%, 3.16%, 4.94%, 5.81%, and 4.54% for the NW, FW, TypeM, TypePC and GameM respectively (compared to 0.93%, 3.90%. 5.69%, 6.02% and 5.61% when the whole features were used). These errors were dramatically decreased into 0.29%, 1.31%, 2.66%, 3.85%, and 2.3% for the aforementioned activities by optimizing the user's template (i.e., optimized feature vector technique). In general, the proposed dynamic feature selection approach achieved a significant improvement on the system performance; this is because the most distinctive and unique features were selected to generate the dynamic feature vector for each individual hence, better recognition rates were obtained. Nevertheless, the effectiveness of the proposed feature selection approach should be examined by collecting data over weeks or maybe months to find out the robustness of the user's reference template versus the changes of the human pattern.

The experimental methodology (specifically, the collected activities) and the findings show that of the proposed system is user-convenient and secure to authenticate users on their smart devices; moreover, the technology (i.e., smartwatches) is sufficiently capable and the nature of the signals captured sufficiently discriminative to be useful in performing activity recognition. Although the amount of the extracted samples from each user and for each activity was fairly acceptable (at least for a research purpose), a generalized activity-based acceleration dataset (i.e., dataset that contains movement data from a large number of users over long period of time as well as involves several human

activities such as jogging, eating, gestures, and driving a car) is necessary to claim the proposed system is robust to impersonation attacks. Moreover, a thorough evaluation is required for interpreting these results; this can be achieved by developing an application to test the efficiency of the system, using different artificial machine algorithms, and evaluating different segment size.

Research has tended to focus on collecting data under constrained environment rather than capturing real life data. This scenario is only practical if a user declares/ labels the performed activity (which is unexpected in real-world implementation as users tend to do several activities during their daily life). As a result, capturing real life data is essential to make sure that the collected data can be used for real practical authentication system.

Although the proposed system achieved high accuracy (i.e., as low as EER of 0.29% and up to 3.85%), smoothing functions such as majority voting could reduce the error rates and offer a user-friendly environment by reducing the rejected user's samples as well as monitoring the system from being misused. One of the major drawbacks to adopting a smoothing function is the time required by the system to predict the user's identity (i.e., more time is required to make decision by the system). As a result, there is an increase chance for the imposter to abuse the system.

4.6 Conclusion

The experimental research has shown the effectiveness of using smartwatch-based activity recognition system to identify the legitimate user based upon five different activities. The aim of this study is to strike a right balance between robust security and ease of use. The study also examined the effect of using the CD scenario on the system performance and presents a novel feature selection approach that effectively reduced the feature vector size without overtly affecting

performance. Moreover, the advantage of creating an activity-based user authentication model is highlighted in order to decrease the EER. Further investigation was carried out to present an analysis of the optimal feature vector size for each individual, which has resulted in lower EERs for the proposed system. The proposed system was evaluated by collecting the motion data from 60 users and analysed the feature set to determine its uniqueness. However, more experimental work should be carried out to explore different feature reduction approaches.

The next chapter will aim to remove the one factor that is explicitly controlled in all previous studies – the nature of the controlled data collection and instead look to understand what the performance of the approach is with real life data over a prolonged period of time (weeks). As the nature of the real-life signals is likely to be noisy, activity-recognition approach will be used in order useful to predict the user's activity.

5 Continuous Smartwatch-Based User Authentication Using Unlabeled Motion Data

There is no doubt that the reported findings in Chapter 5 were very competitive and outperformed the previous art, however, the majority of sensor-based studies (as well as the presented experiments in the previous chapter) were implemented within a controlled environment. This means that all participants were asked to do exactly the same type of activity such as walking on flat floor in an indoor environment. However, the more realistic test comes by capturing uncontrolled data (i.e., real-life signals where users will not ask to perform certain activities, but merely wear the smartwatch). To this end, this chapter aims to

- Collect uncontrolled data to find whether the system still capable to achieve relatively good accuracy.
- Investigate the effectiveness of the dynamic feature selection approach and its impact upon real-life data.
- Demonstrate whether the fusion of the acceleration and gyroscope data has a positive impact on the system accuracy.

5.1 Introduction

The use of motion signals for TAS requires a scientifically valid experiment to collect, analyses and evaluate the feasibility of wearable computing. Therefore, the purpose of this research is to investigate and look at an experiment that provides the empirical basis for understanding how well this technique will offer by using data under unconstrained environment. When it comes to behavioural biometric systems, the majority of previous acceleration-based studies were based upon using data that have been collected within a controlled environment.

Nevertheless, the laboratory experiment tends to be far less realistic for real world applications; moreover, this might be very difficult to any classifier to discriminate between individuals. The big challenge in behavioural biometrics is implementing an experiment in a real-world scenario. This is because the random arm movements in the real scenario such as making a phone call and hand shaking. These factors can significantly affect the nature of the captured signal. Therefore, this study aims to collect real time accelerometer and gyroscope data within uncontrolled environment in order to setup a real practical authentication system. Collecting real data could be useful to enhance the performance of TAS; this could be due the fact that users are not doing exactly the same thing in the normal practical scenario, which might greatly help the classifier to differentiate between them.

Although sensor- based user authentication systems could be implemented by using the fusion of both sensors (i.e., accelerometer and gyroscope), the majority of the previous studies utilized the accelerometer sensor alone. The variation of the real-life signals for both sensors could result in a unique and distinctive pattern for individuals. Therefore, this study utilized the fusion of both sensors to explore if the proposed technique could improve the system performance. The nature of the real-life signals is likely to be very noisy thus, it is very difficult to predict the activity type of individuals. It was already highlighted in the previous chapter that detecting the activity type would significantly reduce the EER for the sensor-based user authentication system. Therefore, this study proposed a lightweight activity detection method based upon the frequency component of the acceleration signals (more details can be found in section 6.2.2). As such, this study has sought to improve upon the prior art in the following manner:

- To provide a robust and realistic dataset that was captured in a completely unconstrained environment, involving 30 users with up to 10 days of real-life data collection. This is potentially the largest dataset for activity-based user authentication using commercial smartwatches.
- To evaluate the performance of the proposed system based upon real-life data not simply laboratory controlled.
- To propose a light activity detection approach to identify the activity type before classifying the user's identity that significantly improved the system performance.
- To investigate the effect of the fusion feature level on the classification accuracy.

The present chapter describes the methodology of capturing the acceleration and gyroscope signals, pre-processing, feature extraction and selection process, and the classification performance. Section 6.2 details the experimental setup that are used for designing the activity-based user authentication system. The results are explained in Section 6.3; Section 6.4 and 6.5 present the discussion and the conclusions respectively.

5.2 Experimental Methodology

This section describes the process of collecting the movement signals, divide the dataset into groups based upon the activity type, and feature extraction. In order to overcome some of the shortcomings of prior work, this study explores the following research questions:

- To what degree can activity-based user authentication be successfully achieved in an uncontrolled environment (i.e., real-life)?

- Does the fusion approach (i.e., combination of both accelerometer and gyroscope data) for the real-life signal enhance the authentication performance?
- Does the proposed activity detection and the majority voting method maximize the classification accuracy?

5.2.1 Data Collection

In order to evaluate the activity-based user authentication under unconstrained environment, the acceleration and gyroscope data streams were collected from 30 users. The acceleration and gyroscope data streams were collected from a subset of 30 users from the original controlled experiment (which involved 60 users). Both genders were included in the data collection process (17 males and 13 females) with a range between 18-55 years old as shown in Figure 30.

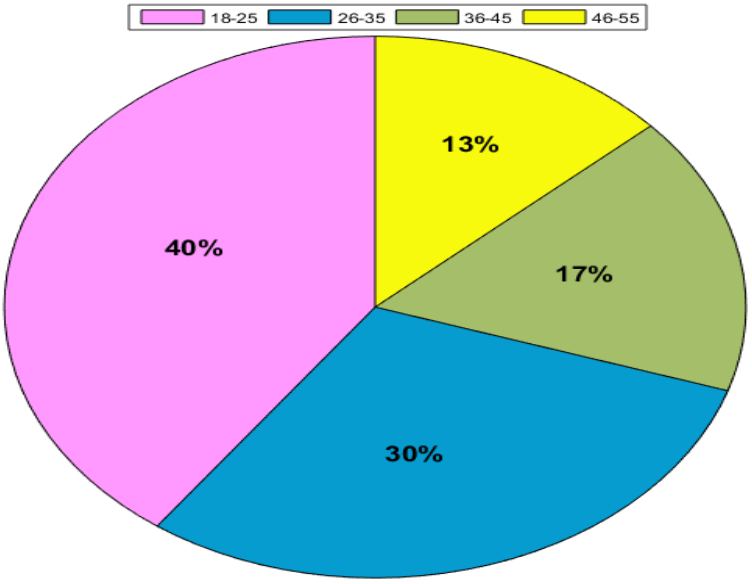


Figure 30: The age ranges across the participants for the real life data

Once the smartphone and Microsoft band 2 were turned on, the user's arm signal was captured in a continuous and transparent manner by running an android application in the background. For consistency, the sampling rate was fixed (i.e., 32 Hz) for all participants; users were not asked to provide predefined activities,

but merely to wear the smartwatch for 10 days to enable a real-world evaluation of the proposed approach (i.e., users were encouraged to freely undertake their daily routine in order to make sure the collected data represented the user's actual and typical behaviour). Each user was asked to wear the watch for at least 4 hours per day (or until the smartwatch battery was drained). The total collected data per user was approximately 40 hours (4 hours * 10 days) and 1200 hours for all users. To the author's best knowledge, this is the largest smartwatch-based sensor data in the domain.

The total extracted samples per user of the uncontrolled experiment (over 10 days) for each activity are presented in Table 15. The amount of the collected gait samples (i.e., for both normal and fast walking) were a total of 32327 (compared the prior accelerometer-based studies art that collected limited dataset ranging between 900 and 1000 samples). For the non-walking signal, 93637 samples were obtained which is the first acceleration/gyroscope-based smartwatch study that used the stationary signal for the TAS.

User ID	NW	FW	Non-W	User ID	NW	FW	Non-W
1	1314	1763	2813	16	1243	381	9081
2	276	199	1329	17	390	173	2810
3	978	747	2270	18	1179	564	2250
4	898	336	3461	19	847	145	3640
5	897	246	3418	20	618	159	1025
6	447	213	2929	21	758	185	5640
7	416	135	1089	22	375	238	3400
8	427	296	1797	23	209	107	2186
9	1160	281	3066	24	276	155	2151
10	832	425	2880	25	120	93	1484
11	551	102	2865	26	629	528	1990
12	844	333	2749	27	192	384	6910
13	245	173	1062	28	970	750	2371
14	391	152	8070	29	899	352	3161
15	840	430	2418	30	997	264	3322

Table 15: The total samples of the uncontrolled data separated by user

5.2.2 Data Pre-Processing

Several sensor-based studies have identified that the collected motion signal contained some noise and errors (Gafurov, et al., 2007a; Derawi et al., 2010a; Hestbek et al., 2012; Hoang et al., 2013; Muaaz and Nickel, 2012; Sangil Choi et al., 2014). It is complicated/ challenging to get the signal in the symmetric form, specifically for the real-life data as it contains several user's activities such as running, walking, typing that could produce completely different signal shapes (i.e., waves), hence result in a very noisy data (which is nearly impossible to classify). As mentioned earlier, the noise could be resulted by shaking the hand, provide a quick gesture (e.g., suddenly raising the user's hand) or changing clothes. Therefore, it is important to train multiple reference templates; each contains the data of specific activity and an activity recognition should be applied to distinguish the performed activity and select the correct authentication template. The following steps were adopted to the original acceleration and gyroscope signals to eliminate or reduce the noise.

- **Removing unworn signal**: as long as the smartwatch is on, the application would keep running in the background and capture the movement data in a continuous manner. Therefore, the information of galvanic skin sensor was used to remove the signals in the case of the smartwatch user takes off the band. Figure 31 shows the original signal and the highlighted red part was removed as it represents the signal when the watch was not worn.

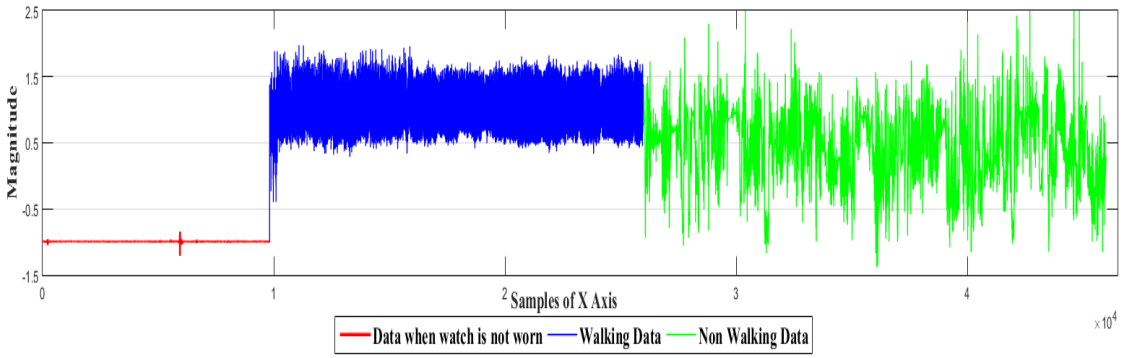


Figure 31: An example of real-life data of user 1

- **Gait detection:** without identifying what a user is doing at a specific point of time, recognizing or predicting the user's identity is difficult to achieve (i.e., the error rates would be increased significantly). The real-life signal is more likely to greatly fluctuate (i.e., very noisy) as shown in Figure 31; this is because users are more likely performing different activities during their daily routine. Other factors that negatively affect the regularity of the captured signal could be carrying a load, hand in a jacket or trouser pocket. As a result, dividing the signal into subsets could result in distinctive arm pattern among the population (each subset contains data of specific activity).

  This study proposed a lightweight approach that automatically detects the repetitive cycles of the user's walking pattern from the original signal. This was achieved by analysis the horizontal (x) acceleration signal of different users due to the high discrimination power compared to the vertical (y) and sideways (z) motions. However, data of other axes (i.e., y and z), as well as x axis, were further utilized to create the reference and test templates. Based on the above observations, it is hypothesized that the detected cycles represent the user's gait pattern while the rest of data is considered as non- walking (Non-W) samples. Primarily, it was important to determine the start point of the actual walking pattern that was identified of about 1.3m/s (Gafurov et al., 2007a) and thereafter detecting

the repetitive peaks based upon the initial gait sample. This was carried out by identifying the initial minima that is found at the following equation:

$$Mi1 = \min(Aw-d1, ..., Aw+d2)$$

Where (d1 = 50, d2 = 150, i =32 acceleration values and Aw was the first accelerometer value greater than 1.3m/s). The first minima was considered the start point of the first cycle, and the second local minima was selected as the terminus of the cycle. To compute the second minima, the following equation was used:

$$Mi2 = \min(Mi1+D-d,..., Mi1+D+d), \text{ where } D=32 \text{ and } d=20$$

This procedure was repeated until all remaining minima were found in the signal as shown in Figure 32. The end point of one cycle was considered as a start point for the next cycle and so on. Once the gait data was extracted, it was split up into two groups (i.e., normal and fast walking) in order to improve the classification accuracy. This was achieved by detecting the local maxima peaks for each segment and average the values. Segments that have high peak values were considered as fast walking (FW) samples while segments with low peak values reflect the normal walking (NW) data (see Figure 33). In total, the original movement data was segmented into 3 groups: NW, FW, and Non-W.

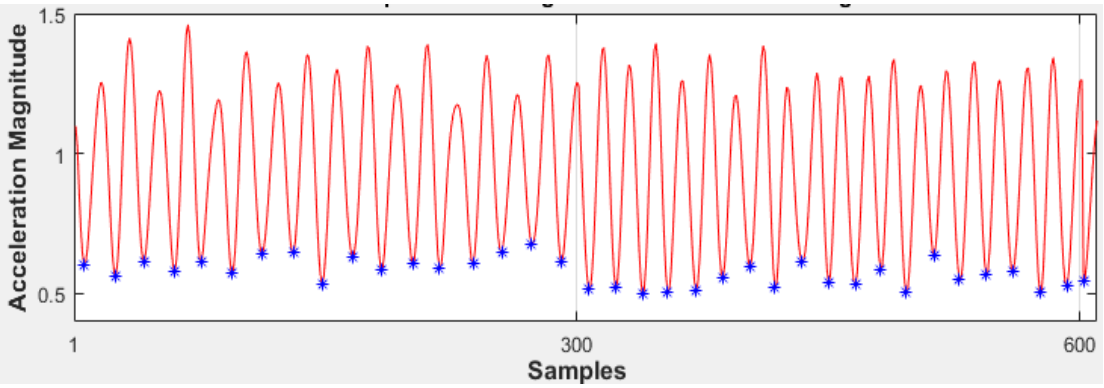


Figure 32: An example of the detected local minima from the Acc signal

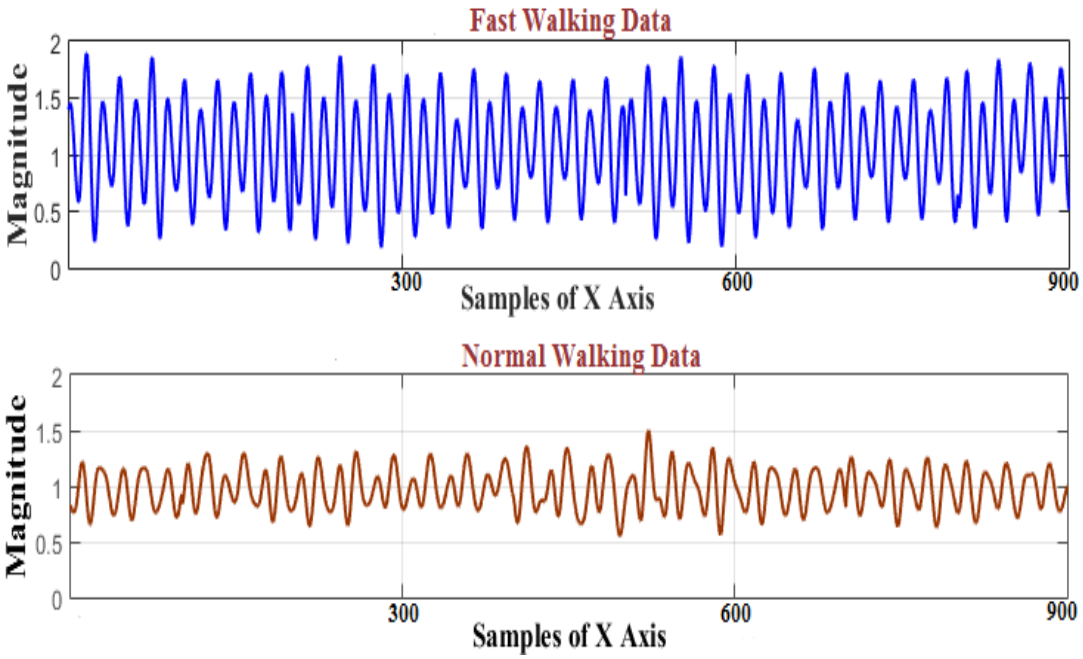


Figure 33: An example of filtering out the real-life data and fast walking VS normal walking

- **Filtering:** similar to the conducted experiment (i.e., the controlled experiment) in the previous chapter, a low pass filter and the cut-off frequency of 20 Hz was used in order to minimize the unwanted accelerometer and gyroscope signals. Apart from eliminating irregular walk steps, the aim of applying the filtering technique was to reduce fake gait samples (i.e., when a user moves their hand in a symmetric way). Figure 33 presents the efficiency of the proposed filter that resulted in consistent walking style; moreover, it shows an example for the detected NW and FW samples by using a simple and lightweight gait detecting technique. It can be clearly observed from Figure 33 that the proposed gait detection approach was able to detect series of the same peaks range for the FW and NW respectively. For example, the magnitude range of the fast walking peaks were between 0 and 2, while it ranged from 0.4 to 1.2 for the normal walking peaks.

- **Segment size:** as the segment-based showed high level accuracy rather than the cycle-based method, the acceleration and gyroscope signals were segmented by using sliding window approach. Another important parameter that may affect the system performance is the window size. Several studies highlighted that choosing a short segment interval of data (e.g., 3, 4, and 5 seconds) has dramatically increased the EER as it does not contain enough information to recognize the user's pattern (Nickel et al., 2011b; Shen et al., 2018; Mare et al., 2014; Kumar et al., 2016). In contrast, utilizing a bigger segment size (e.g., 15 or 20 seconds) requires more processing time and gives a high chance to attacker to misuse the proposed system as well as offers less number of samples for training the classifier. Therefore, selecting a suitable window size that offers a balance between the security and usability is required. This study divided the raw movement data into 10 seconds due to its high performance in the controlled experiment, which is presented in the previous chapter.

5.2.3 Feature Extraction and Feature Selection

The 10 seconds of time-series data is transformed into a single example via the use of a number of summary features. The transformation process and summary features used in this study are identical to the ones used in the prior activity-based user authentication study (i.e., the controlled experiment in the previous chapter) due to their high performance. These features are the same regardless of whether the example is being generated from accelerometer or gyroscope data. Details of these features can be found in Section 5.3.3. The discrimination capabilities of the most relevant features that are invariant to changes were investigated. This was achieved by applying the feature selection approach, which was based upon creating a dynamic feature vector for each user.

5.2.4 Experimental settings

In order to train the authentication model, 40% was utilized (which is equivalent to 4 days data), while the remaining 60% samples were used to test the classifier. As mentioned earlier, the SD evaluation does not represent a realistic test for any behavioural -based biometric system. Therefore, all the findings in this study were based upon applying the most reliable scenario (i.e., CD test). After preparing the user's templates, a FF MLP neural network was used as the default classifier due to its reliable performance shown in the previous chapter. For each experiment, two different FF MLP neural network training size were evaluated (i.e., 15 and 20) with each being repeated 10 times in order to account for errors that occur due to the random setting of the neural network weights. The experimental setup for the real-life experiment included a total 43,200 tests (i.e., 1440 test per user), the following section presents the key findings of this study. The results presented in this study were based on using FF MLP neural network of size 10 as it showed the lower EER.

5.3 Results

5.3.1 The impact of gait detection method on the system accuracy

Having devised and applied the gait detection method, data was divided into NW, FW and Non-W activities. To permit a comparison and discuss the results of the gait detection, four models were created for each individual. The first model (i.e., generic model) contains all data without predicting the activity type; on the other hand, each of the remaining three models (called as activity-based authentication models) trained with data of specific activity (i.e., NW, FW, and Non-W activities). Once these models were generated, two experiments were conducted; the first experiment utilized the generic model and reported EERs of 24.54% and 26.11%

for the accelerometer and gyroscope respectively. As expected, high EERs were obtained by using the generic authentication model due to the high degree of variability that exists within the captured signal. The variability of the real-life signal caused by the free arm movement, which was captured within uncontrolled environment. To find out if the proposed activity detection approach would improve the system performance, the second experiment focused upon identifying the activity type and then classifying the user's identity. Table 16 shows the EERs for each detected activity by using the CD scenario.

Activity	Sensor	10 Features	20 Features	30 Features	40 Features	50 Features	60 Features	70 Features	80 Features	88 Features
NW	**Acc**	10.45	6.41	5.47	5.21	5.12	**4.35**	4.60	4.77	4.9
	Gyr	13.16	11.15	9.90	9.59	9.38	**8.96**	9.33	9.17	9.46
FW	**Acc**	5.05	2.37	2.00	1.62	1.44	**1.24**	1.25	1.44	1.74
	Gyr	9.50	7.08	6.34	5.87	5.81	5.74	**5.66**	5.88	6.01
Non-W	**Acc**	7.27	7.22	**7.04**	7.40	7.46	7.20	7.69	8.68	8.74
	Gyr	12.24	**11.25**	11.51	11.29	11.30	11.29	11.48	11.37	12.05

Table 16: The EERs of using activity-based model

The presented results in Table 16 showed that the accelerometer sensor achieved high level of performance, at best EERs of 4.35% 1.24% and 7.04% for the NW, FW and Non-W activities respectively (compared to an EER of 24.54% of utilizing a generic-based authentication model for the accelerometer data). Using the gyroscope -based signal for TAS reported EERs of 8.96%, 5.66%, and 11.25% for the above-mentioned activities. These findings were consistent with prior art that highlighted the gyroscope sensor is less effective than accelerometer. However, these results are considered quite impressive if they are compared to the accuracy of generic-based authentication model (i.e., 26.11% of EER).

Since the goal of the proposed feature selection approach is to reduce the user's reference template and improve the system accuracy, Table 16 clearly shows an advantage over the verification performance and generating smaller feature vectors. For example, the feature reduction method was effective by decreasing the EERs from 4.9%, 1.74%, and 8.74% (using the all features of the accelerometer sensor) to 4.35%, 1.24% and 7.04% for the NW, FW and Non-W activities respectively. Although there is not a significant improvement on the system accuracy, the best EERs were achieved by curtailing the number of features from 88 to 60, 60, and 30 features for the aforementioned activities. This implies that the proposed method successfully discarded about 32% of the gait features and nearly 66% of the Non-W features. Clearly, the proposed activity detection method has a significant positive impact on the authentication accuracy and suggest that commercial smartwatches can be effectively utilized to design a robust, secure, and user-friendly authentication system (i.e., TAS).

With the aim to understand how individual user performed for each activity, results for each user's acceleration and gyroscope are presented in Figures 34, 35, and 36. As shown in Figure 34, the acceleration EERs for the NW activity were relatively small (i.e. in the range of 0-5%) for 90% of users, while users 10, 11, and 17 achieved EERs ranged between 10% and 15%. This suggests that users have a consistent and distinctive set of acceleration pattern characteristics. Although the individual performance for the gyroscope sensor was less promising in comparison with the acceleration findings, the EERs for two third of the participants were fairly acceptable (ranging from 0-10%). The individual user performance for the FW activity using the acceleration signal was vastly good, where the majority of users achieved an EER of less than 1% (see Figure 35). These findings are in line with the controlled experiment (which is presented in

the previous chapter) that showed the fast walking signal of individual contains more distinctive information compared to other activities hence, the user's identity can be identified even in a noisy environment with lower EER. For the Non-walking activity, Figure 36 showed that one third of the users reported an EER in the range of 0-5% by utilizing the acceleration data, while the rest of the users resulted in EERs between 5% and 10%, apart from users 3, 4, 5, and 7. In contrast, the gyroscope performances were varied and less effective than the acceleration sensor; for instance, high EERs were achieved for some users (i.e., more than 20% EER such as users 1, 3, 7, and 20) while an EER of around 0% was achieved for others (e.g., 6, 16, and 19).

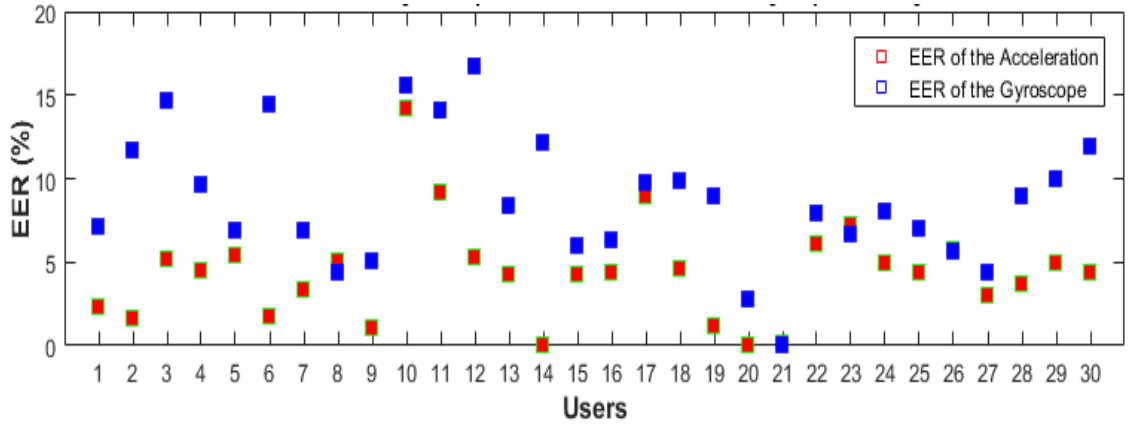


Figure 34: The Acc vs Gyr EERs separated by users using the NW activity

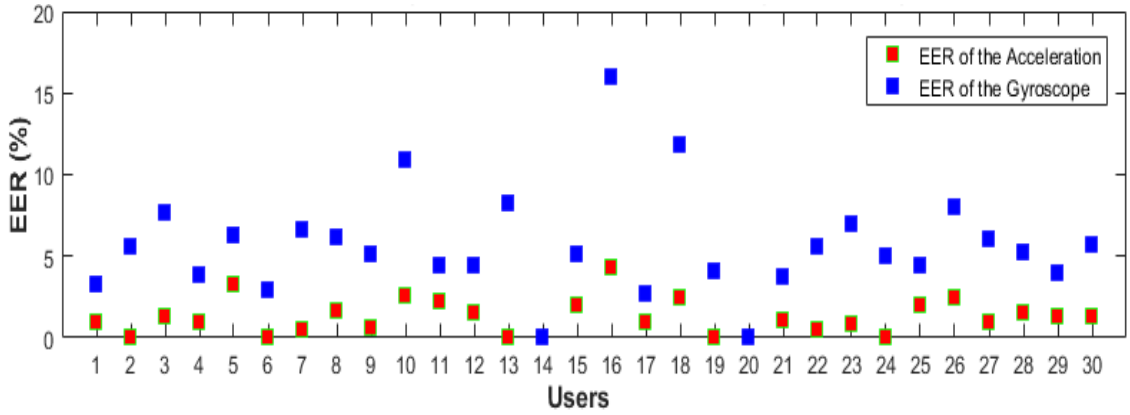


Figure 35: The Acc vs Gyr EERs separated by users using the FW activity

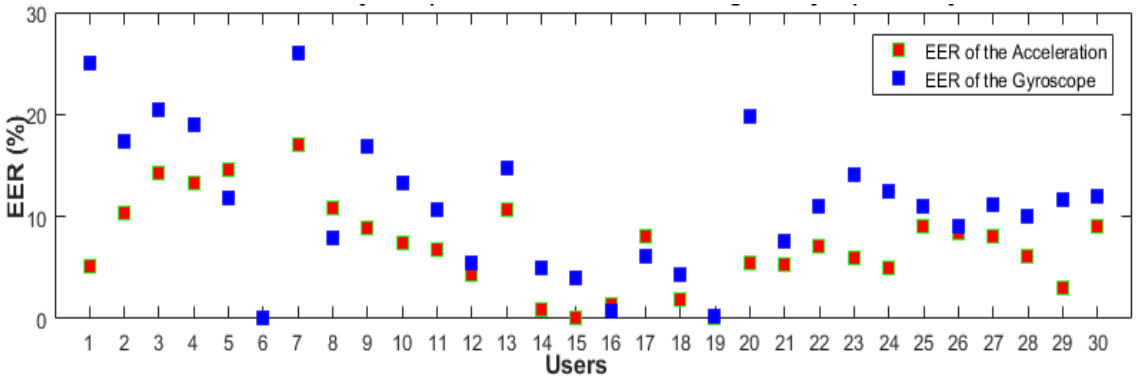


Figure 36: The Acc vs Gyr EERs separated by users using the Non-W activity

5.3.2 The effectiveness of using the fusion of both sensors on the biometric performance

Although the findings in Table 16 suggest that accelerometer sensor resulted in lower EERs (i.e., better performance) than gyroscope for all activities, prior studies (Kumar et al., 2016; Kumar et al., 2017) have highlighted that sensor-based authentication systems might be susceptible to attacks if a single sensor (e.g., accelerometer or gyroscope) is used. Moreover, other authors Johnston and Weiss, (2015) suggested that the system accuracy might be improved by using fusion features of both sensors. The fusion schema in biometric-based systems can be implemented at three different levels: sensor level, feature level, and score level. In the sensor level fusion, data of single modality or multiple biometric traits are used together; for example, capturing face samples from different cameras and different angles (in case of unibiometric system) or collecting multi-biometric modalities such as face and voice.

When it comes to the feature level fusion, features that are extracted from different sensors readings are fused in order to generate a resultant reference template. Finally, the score level is an approach that measures the similarity scores between the reference and test templates and combines the resultant scores of each modality together. This study investigated whether the feature

level fusion can offer a better verification rates. As mentioned earlier, 88 time domain features were extracted for each sensor, so the fusion approach resulted in 172 features for accelerometer and gyroscope sensors (88 features * 2 sensors). Table 17 displays the EERs of using the fusion approach of both signals for the detected activities.

Activity	10 Features	20 Features	30 Features	40 Features	50 Features	60 Features	70 Features	80 Features	90 Features	100 Features	110 Features	120 Features	130 Features	140 Features	150 Features	160 Features	All Features
NW	9.32	5.51	5.51	4.26	3.98	3.91	3.66	3.84	3.96	3.86	**3.55**	4.04	3.82	3.56	3.98	3.84	4.45
FW	5.01	2.70	1.89	1.68	1.61	1.36	1.36	1.28	1.23	1.23	1.16	**0.92**	1.45	1.03	1.22	1.38	1.41
Non-W	7.36	5.58	**5.31**	5.59	5.39	5.39	5.39	5.36	5.44	5.40	5.47	5.55	5.44	5.68	6.52	6.75	6.84

Table 17: the EERs of applying feature level fusion separated by activities

Obviously, Table 17 shows that the authentication performance is significantly improved for all activities. Using the fusion approach decreasing the EERs from 4.35%, 1.24%, and 7.04% (for accelerometer) to 3.55%, 0.92% and 5.31% for the NW, FW, and Non-W activities respectively. This positive effect of using the fusion technique is more noticeable if it is compared to the gyroscope findings (i.e., EERs of 8.96%, 5.66% and 11.25%). Meanwhile, the feature reduction approach greatly reduced the user's templates, specifically for the Non-W data that reported an EER of 5.31% by using 30 features only. For the gait activities (i.e., NW and FW), neglecting about 40% features led to obtaining lower EERs.

To find out whether the individual user performance is better with or without the fusion approach, further analysis for each user separately was conducted (as shown in Figures 37, 38, and 39). This was achieved by comparing the findings of the fusion approach against the acceleration user's performance (as it was better than the gyroscope results). Figure 37 proves that using the fusion approach successfully minimize the EERs for half of the users while the individual

user performance was nearly similar for the rest of users (e.g., users 9, 11, 12, 13, and 14). A possible explanation for not maximizing the accuracy for 50% of the users may be that their gyroscope features were not sufficiently discriminative to add a noticeable contribution to the individual user performance. Another possible justification for this is that the majority of the selected feature subset for generating the user's reference template was acceleration-based feature, hence the EER was nearly similar.

Similarly, using the fusion approach for the FW activity improved the individual user accuracy for around 50% of the users (e.g., the EERs were significantly reduced for users 5, 11, and 15); the rest of the users reported nearly similar performance (as shown in Figure 38), apart from users 12 and 21 where their EERs were slightly increased by using the fusion approach. For the Non-Walking activity, a significant reduction for the reported EERs was achieved for more 65% of the users (apart from user 7) by utilizing the fusion method while no impact was noticed for the remaining users.

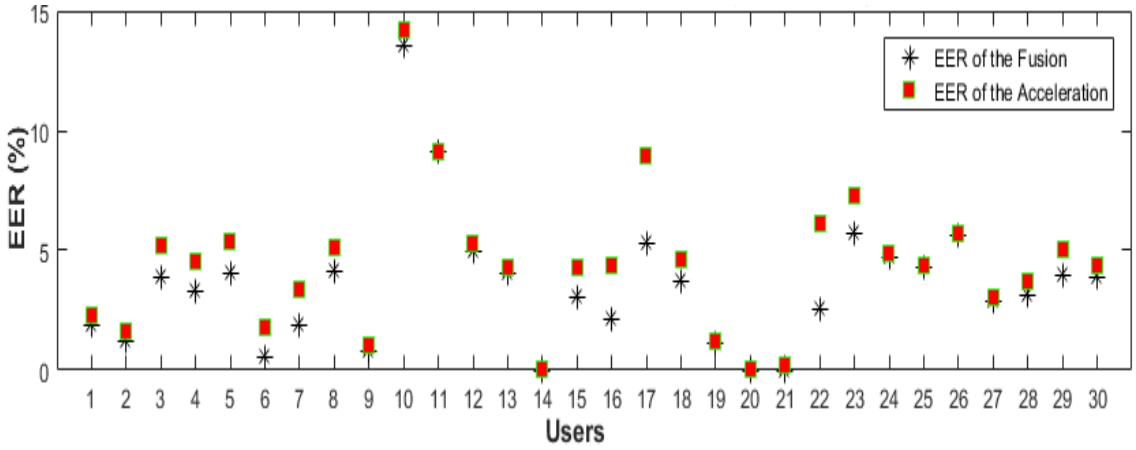


Figure 37: The fusion vs Acc EERs separated by users using the NW activity

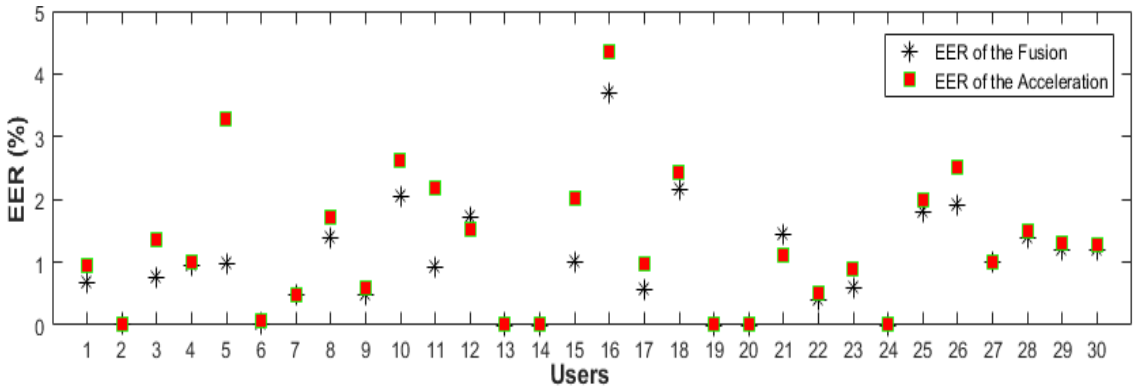


Figure 38: The fusion vs Acc EERs separated by users using the FW activity

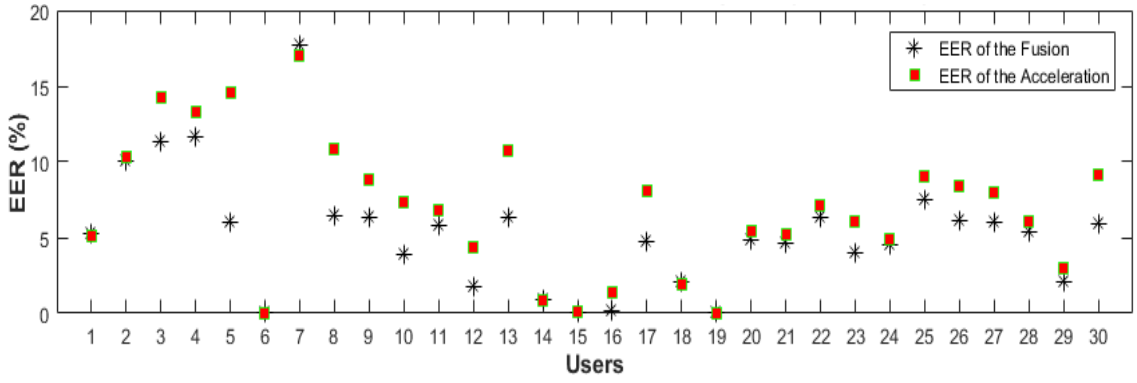


Figure 39: The fusion vs Acc EERs separated by users using the Non-W activity

5.3.3 The influence of the optimized feature vector upon performance

The presented findings in Tables 16 and 17 were based upon creating dynamic reference template for each user but the feature vector size was fixed for all users; for example, the best EER for the NW activity was 3.55% by using 110 features for each individual. Nevertheless, the findings of the controlled experiment demonstrated that optimized feature vector could be useful to maximize the system performance. For instance, EERs of 0.29%, 1.31%, 2.66%, 3.85%, and 2.3% were achieved for the NW, FW, TMob, TPC, MobG respectively (compared to 0.29%, 1.31%, 2.66%, 3.85%, 2.3% when the static feature vector was used). Therefore, further investigation was carried out to identify the optimal feature subset size for each user independently. For instance, some users might require few features to accurately recognize their pattern, while increasing the feature size may offer better accuracy/error rates for other users. Table 18 displays the

results of the best EERs for the all activities using static and optimized feature vector.

Activity	Sensor	EER (%) of SFV	EER (%) of OFV
NW	Acc	4.35	3.93
	Gyr	8.96	8
	Fusion	3.55	2.18
FW	Acc	1.24	0.73
	Gyr	5.66	5.15
	Fusion	0.92	0.70
Non-W	Acc	7.04	6.51
	Gyr	11.25	10.47
	Fusion	5.31	4.77

Table 18: The system performance of using static and optimized feature vector

As can be seen from Table 18, the authentication accuracies were enhanced by employing the optimized feature vector. The EERs of the NW were reduced from 4.35%, 8.96%, and 3.55% to 3.93%, 8.10% and 2.18% by using the accelerometer, gyroscope, and the fusion data respectively. Similarly, the optimized feature vector offered lower EERs for the FW data (i.e., 0.73%, 5.15%, and 0.70% for the accelerometer, gyroscope, and fusion data sequentially compared to 1.24%, 5.66% and 0.92% when the static feature vector was utilized). For the Non-W activity, little difference in findings was noticed between the two approaches (i.e., the static feature vector and optimized feature vector). For example, the EERs decreased from 7.04% to 6.51% (for accelerometer) and from 11.25% to 10.47% (for gyroscope). The possible explanation for this outcome could be the reference template for the majority of users was nearly optimized. For example, the best EERs for the Non-W data were obtained by using small feature subset such as 30 and 20 for the accelerometer and gyroscope signals respectively as shown in Table 16.

Apart from the improvement in the system performance of using optimized feature vector, Figure 40 shows that the user's reference template size was decreased for half of the users. For example, the feature vector of users 2, 19, and 29 was created by utilizing only 20 prioritized features and even less features were used for users 13, 14, 24, 25 (i.e., 10 features). In contrast, other users such as 8, 9, 11, 8, 15, 16, and 18 required more features (i.e., 70 to 88 features) to produce low EER. This can be explained that the gait pattern of some users is more inconsistent over time hence, more features are required to recognize their pattern.

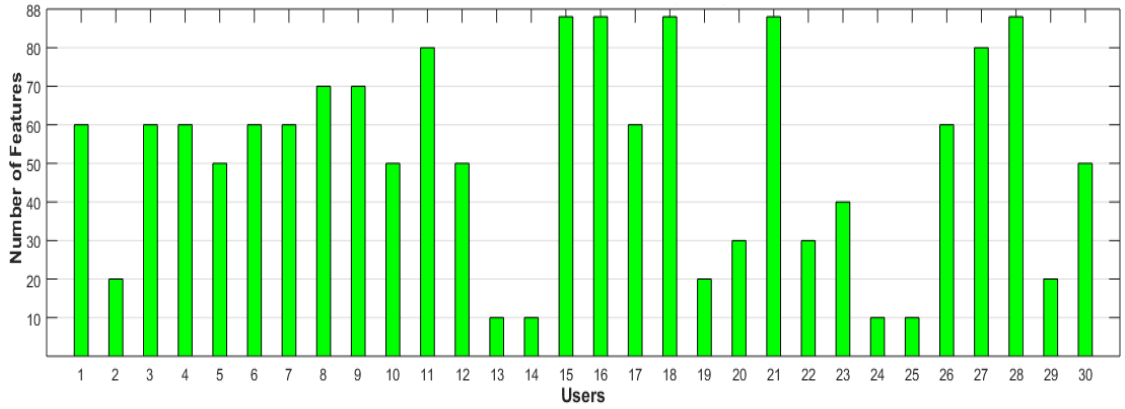


Figure 40: Applying optimized feature vector of each user for the FW activity

5.4 Discussion

The discussion would be formed around four of the following core questions:

- To what degree smartwatch-based user authentication can achieve with uncontrolled environment?
- What is the effect of the feature level fusion on the verification accuracy?
- Does the classifier performance improve by using the proposed activity detection method?
- Does the optimization of the feature vector maximize the classification recognition rate?

The main goal of this study is to evaluate the scalability of smartwatch-based user authentication system within the uncontrolled environment. To achieve that, real life signals (i.e., unlabelled data) was captured from 30 users over 10 days with a minimum of 4 hours per day from each user. The findings in Table 18 strongly suggest that smartwatch-based user authentication can be used to replace, or at least supplement, password-based authentication systems. The proposed system has an advantage over password-based user authentication, in which impersonation is much more difficult to accomplish and even video footage of the arm movement (to match the victim's arm pattern) is not sufficient to mimic a user Gafurov et al., (2007b).

Although the evaluation of any behavioural-based biometric system is a big challenge (due to the noisy signals), competitive results with EERs of 2.18%, 0.70%, and 4.77% were achieved for the NW, FW, and Non-W data respectively. These results show that the proposed system highly efficient in identifying the legitimate user in a transparent and continuous manner; the most closely related work was conducted by Lee and Lee (2017), which was based upon capturing uncontrolled data. The stated authentication accuracy of their experimental work (i.e., an EER of 8%) is considerably lower than the findings of this study (i.e., an EER of 2.18%, 0.70%, and 4.77% for the NW, FW, Non-W respectively). Moreover, the gait results still better than the prior gait studies that were based upon controlled data and reported EERs in the range of 5.7% to 33.3% (Muaaz and Nickel, 2012; Damaševičius et al. 2016).

When it comes to non-walking activities such as gesture or typing activities, the prior art reported EERs ranged between 22% and 4.9% compared to 4.77% of EER in this study (Lewis et al., 2016; Shen et al., 2018). Nevertheless, these studies based upon a dataset collected from a controlled laboratory environment

(i.e., unrealistic setup for real practical systems). Although the authors in Yang et al., (2015) and Liang et al., (2017) slightly outperform the Non-W performance with EERs of 3.3% and 4%, their systems suffer from a number of pitfalls. For example, the collected gestures (i.e., punch or drawing a 3D circle) were unrealistic for the authentication purpose and cannot offer transparent and continuous authentication. Moreover, their system explicitly required labelled data (i.e., constrained environment) and the authentication phase was based upon limited amount of test samples (in the range of 10 to 30 samples).

In order to explore the impact of the selected sensor on the system performance, three fundamental experiments were carried out; the first experiment utilized the acceleration data, while the gyroscope signal evaluated in the second and finally the feature level fusion of both sensor was employed in the third experiment. The authentication performance of using the acceleration signal was quite impressive when one considers that the system evaluation was based upon utilizing realistic unconstrained real time data; at best EERs of 3.93%, 0.73%, and 6.51% for the NW, FW, and Non-W activities respectively. On the other hand, the gyroscope sensor was less effective and reported EERs of 2.18%, 5.15%, and 10.47%.

Although the above findings show that the nature of the captured signals sufficiently discriminative to be useful in performing TAS, Table 18 elaborates the benefit of the fusion approach on the recognition accuracy. For example, the EERs of using the accelerometer data were further reduced from 3.93%, 0.73%, and 6.51% into 2.18%, 0.70%, and 4.77% (by using the fusion method). Apart from the improvement on the system performance, the combination features of both sensors could add an extra layer of security for the user authentication-based biometric system.

As the nature of motion-based real-life signals is very noisy, building robust and applicable authentication system without identifying the activity type nearly impossible or would lead to overly undesirable performance results (i.e., above 20% of EER). The reported authentication rates of using the generic- based authentication model confirm the above hypotheses by reporting EERs of 24.54% (for accelerometer) and 26.11% (for gyroscope). It is highlighted earlier, this model was created by using unlabelled data (i.e., without predicting the activity type). Therefore, the proposed activity detection method (which created a separate model for each individual activity) greatly reduced the EERs to 2.18%, 0.70%, and 4.77% for the NW, FW, Non-W respectively. Nevertheless, developing a context aware approach might give better understanding to the user's daily activities (rather than dividing the data into walking and non-walking activities) (Benzekki et al., 2018; Feng et al., 2014; Habib and Leister, 2015; Primo et al., 2014; Witte et al., 2013). This can be achieved by obtaining information from other smartwatch sensors (e.g., GPS, and ambient temperature) that could be used as a basis for making a more intelligent decision and improve the system accuracy still further.

To highlight the positive effect of the feature reduction, all the evaluations of the proposed smartwatch-based user authentication system were conducted with and without the feature selection. Based upon the results in Table 16, it is obvious the proposed approach alleviated effectively the computation overhead by neglecting about 33% of the total gait features, and even more when the Non-walking activity is considered (i.e., nearly 66%).

In the case of the fusion-based method (i.e., 176 features), the best EERs were achieved by curtailment the subset size into 110, 120, and 30 features for the NW, FW, and Non-W activities respectively (see Table 17). That means the

feature reduction method was able to minimize the user's gait templates of around 30% and more than 80% for the Non-W feature vector, which is a significant achievement. In comparison with walking templates, it can be seen that the user's reference template of the Non-W activity requires few features (i.e., 30 features by using the accelerometer data alone and the fusion approach). This could be explained if the user was not active (i.e., there is not a significant arm movement), few observation and features are sufficient to make the optimal authentication decision.

In contrast, the human gait varies and could be influenced by several factors in the real scenario such as clothes and carrying a load. As a result, more features are required to recognize the user's pattern in an effective fashion. Apart from the significant reduction on the user's templates, the proposed feature selection approach successfully improved the system performance with EERs of 2.18%, 0.70%, and 4.77% compared to 4.45%, 1.41% and 6.84% for the NW, FW, and Non-W respectively. This is an improvement of around 50% over the classification performance of the gait activities and the EER of the Non-W data was nearly reduced by 40%. Nevertheless, more investigations are necessary to explore and implement different feature selection strategies that further improve the classification decisions.

5.5 Conclusion

The investigation undertaken in this research has positively demonstrated that smartwatch- based biometric is a feasible approach in achieving reliable transparent user authentication. Although data was collected under unconstrained environment, the findings showed that the proposed system can effectively verifies the user's identity with low EERs for all the detected activities. A preliminary study was conducted to examine the effectiveness of using the

fusion-approach and its impact on the system performance; a novel feature selection approach was also proposed to effectively reduce the feature vector size without overtly affecting performance. Based upon the findings of the conducted investigations, the aforementioned approaches (i.e., the fusion and feature selection) have significantly reduced the EERs for a subset of 30 users. Employing the optimized feature vector has further strengthen the system performance, at best, EERs of 2.18%, 0.70%, and 4.77% for the NW, FW, and Non-Walking respectively (compared to 3.55%, 0.92%, and 5.31% when a static feature vector for each individual was used). Experimental results also demonstrate the advantage of predicating the activity type in order to enhance the system performance. The proposed detection method has been proved that the gait activities contain distinctive information (rather than the non-walking data) as more movement data can be captured while a user is walking, which is further positively contributed towards the classification results.

6 Evaluation of the Activity-based User Authentication System Using Smartwatches

The prior two experimental chapters have provided a solid foundation to demonstrate the feasibility of performing activity recognition using a smartwatch. There are however, a number of additional research questions that present themselves when considering whether this approach is practically feasible. This chapter will investigate the principle variables that poses serious concerns to the proposed system and impact the authentication accuracy. These include, evaluate the optimal segment size that offer a trade-off between security and usability, determine the amount of training samples, and propose a smoothing function (i.e., the majority schema) in order to maximize the system accuracy. Details are determined in the following sections.

6.1 Introduction

Whilst competitive experimental results are obtained in Chapters 4 and 5 that show the effectiveness of the proposed system, it is important to evaluate the system performance to show how smartwatch-based acceleration/gyroscope data can provide transparent and continuous protection and would be used in a practical context. Therefore, further practical evaluations were undertaken to ensure the system is both secure and maximise a user's convenience. Moreover, to explore the improvement of certain operations against the base experimental results as well as to determine the optimal settings that are required for a practical perspective. Therefore, this chapter aims to provide a comprehensive analysis in the following manner:

- Examine the authentication accuracy by testing different segment sizes and find out the optimal sample size that offers secure and usable authentication-based system.
- Avoid the training overfitting in order to maximize the authentication accuracy and reduce the dimension of the training set size.
- Highlight the necessity of having a more balanced situation (i.e., system should be adjusted at a certain level of security and provide a high degree of transparency to the end-use) by applying the majority voting schema.
- Propose a context aware approach that could be useful to predict a wider variety of activities rather than dividing data into gait and non-gait samples hence, improving the recognition rates.

6.2 Investigation into segment size and recognition performance

Although the experimental setup in the previous chapters shows competitive results that outperform the majority of the prior art, it is unclear how the sample size can affect the classification performance. So far, all the findings were obtained by utilizing a sliding window approach in order to chunk the raw motion data into equal windows of 10 seconds interval, with no overlap between the extracted segments. Nevertheless, deep analysis is required to investigate whether increasing or decreasing the window size would influence the authentication performance. Various segment sizes (2.5, 5, 7.5, 10, 12.5, and 15 seconds) were evaluated to select the optimal segment length that offers secure and usable biometric-based authentication system.

The same database that was used in the previous chapter (i.e., real life data of 30 participants over 10 days) and the cross-day evaluation were employed for this investigation. For consistency, data pre-processing, the amount of samples

for each setting, the split of the training and testing data (i.e., 40% and 60% were utilized to form the reference and test templates respectively), and the classifier configuration (i.e., FF MLP neural network of size 10) were exactly same to the uncontrolled experiment setup. Finally, the conducted analysis of this investigation was based upon analysis only one activity (i.e., the FW activity) due to the time constraint and the EERs are presented in Table 19.

Segment Length in Seconds	EER%
2.5	1.50
5	1.32
7.5	1.22
10	0.92
12.5	0.90
15	0.82

Table 19: Evaluation results for different segments sizes

Table 19 shows that the system performance enhanced by increasing the segment size from 2.5 to 15 seconds, at best an EER of 0.82% was reported. This substantiates previous findings in the literature that showed a decrease in the EERs by using sufficient data points in the window (Nickel et al., 2011b; Shen et al., 2018; Mare et al., 2014; Kumar et al., 2016). This improvement could be explained that employing larger segment sizes helps to accurately extract unique and distinctive features hence, constructing a robust and effective reference template for the proposed system. Nevertheless, the main downsides of employing a large segment size are implementation costs (i.e., more time and resources are required to process the segmented data) and reduce the usability aspect of the authentication system. Moreover, the user's pattern might be inconsistent, which negatively affects the system performance. Although the EER was decreased from 0.92% to 0.82% for the segment size of 10 and 15 seconds respectively, using 10-second sample of data could provide a high level of security and user convenience and contain sufficient predictive features for TAS.

6.3 Overfitting in machine learning and the negative impact on the classification performance

The number of samples required to train the user's model is an important system parameter. Overfitting is "*the production of an analysis that corresponds too closely or exactly to a particular set of data, and may therefore fail to fit additional data or predict future observations reliably*" (López, X. 2018). An example of the overfitting problem is shown in Figure 41. There are two possible explanations for the overfitting; firstly, the training data may contain noise hence, the machine learning algorithms may fit the noise into the model and therefore poor performance would be obtained. Secondly, limited training samples (i.e., small dataset) that are not sufficient to train the model. As this study captured fairly acceptable number of samples from each subject, specifically for the NW and Non-W activities, it is essential to analysis and investigate the proper training dataset size that avoids the overfitting issue.

To determine required sample sizes, a comprehensive experiment was carried out by using real life data (i.e., NW and Non-W activities). Users that have limited samples were excluded from the experiment (i.e., users that have less 400 samples). In order to train the user's reference template, an equal number of samples were chosen from each user (i.e., 85 samples per day for the selected activity) and the remaining samples for that particular day were neglected to make sure the evaluation scenario is equivalent to the most realistic test (i.e., the CD test). For example, if the acceleration signal of one day contained 250 samples for user1, random 85 samples were utilized to train the user's model and, thereafter, data of different days was utilized to evaluate the system accuracy. Details of the results for each activity and the required number of samples to train the FFMLP classifier are presented in Table 20.

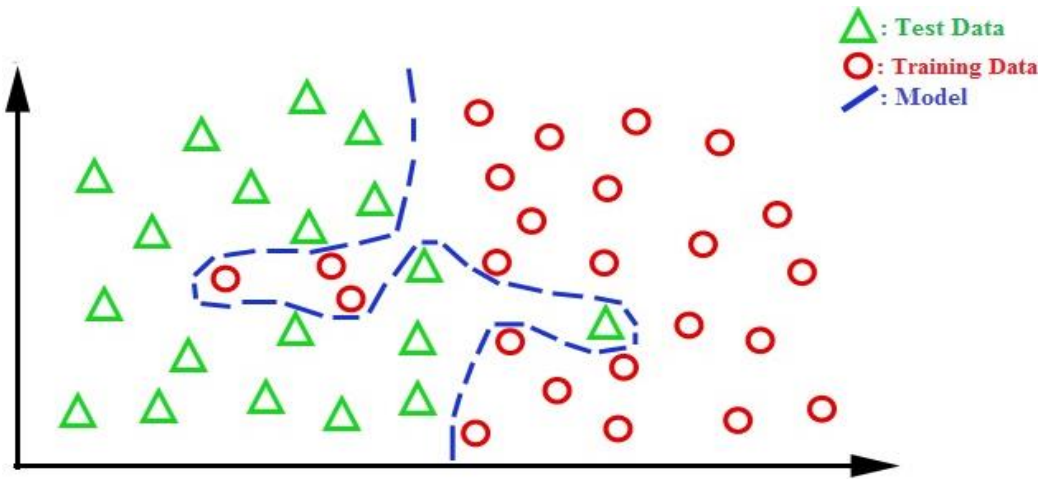


Figure 41: Model training and overfitting problem in machine learning

Activity Type	Training Scenario	Samples	EER (%)
NW	One Day	85	11.19
NW	Two Days	170	7.60
NW	Three Days	255	5.75
NW	Four Days	340	3.92
NW	Five Days	425	3.05
Non-W	One Day	85	8.51
Non-W	Two Days	170	6.15
Non-W	Three Days	255	5.67
Non-W	Four Days	340	5.41
Non-W	Five Days	425	5.09

Table 20: The effect of training size on the authentication performance

As can be seen in Table 20, increasing the sample size has a positive effect on the system accuracy. For example, using a single day training data (i.e., 85 samples) reported EERs of 11.19 and 8.51% for the NW and Non-W activities respectively (compared to EERs of 5.75% and 5.67% when 255 samples were utilized for the training purpose, which is a significant improvement on the authentication recognition rate). This amelioration was slightly lower by increasing the number of samples into 340 and 425 (e.g., EERs of 5.41% and 5.09% for the Non-W activity); this is not particularly surprising given the fact that the classifier was trained with sufficient representative samples. However, more experimental work is required to find out whether feed up more data to the

machine learning algorithms could further reduce the EERs. This can be done by collecting real life data over a longitudinal basis.

6.4 Investigation into majority schema and trade-off between the system security and usability

Based upon the classification result, a decision on whether to accept or reject the output is made by the system. According to the literature, two standard schemas are used: majority or quorum voting. The former scheme accepts a user as genuine if a half or more of the user's test samples are positive. The biometric decision is then based upon merging multiple classification outputs to a single one and it either represents a genuine user or an impostor. The latter authenticates a user as genuine if a requisite number of the user's test samples are positive. A better performance is normally obtained by using the quorum voting technical while the system is more resilient to error when the majority voting is applied. Under the quorum voting scheme, a small number of correct classification outputs are required to accept a user. While this will improve the user convenience (i.e., the user will be highly likely to accept the deployment of such system), it will result a high false acceptance rate (i.e., there is a high chance for the imposter to abuse the system).

In contrast, more discriminative user behaviour is required when utilizing the majority voting technique; otherwise, a high false rejection rate will be produced by the system. It is understood that the system will provide a better security when using the majority voting method; at the same time, the system is more intrusive (i.e., less user friendly). As a result, it is important that a proper decision logic that can balance the system security and user convenience is applied for the gait authentication system.

So far, all the presented results were based upon classifying single sample in order to calculate the EER. In order to reduce FAR and FRR (i.e., low EER), majority voting was used. Two parameters need to be identified: number of samples (*#S*) and the number of votes (*#V)*. Several *#S* (i.e., 5, 7, 9, and 11 samples) were tested to select the best experiment configurations that offer a balance between the both errors. The evaluation process was carried out on the real-life dataset by using NW activity (as it contains enough samples for the evaluation purpose) and using the fusion approach as it showed the best performance. Results of the voting investigation are presented in Figure 42. If the *#V* of each experiment is equal or more than half of the selected samples, the whole votes are considered for genuine. For example, if the *#S* is chosen to be 9 examples and the proposed system recognized 5 only, the FRR in this case would be zero.

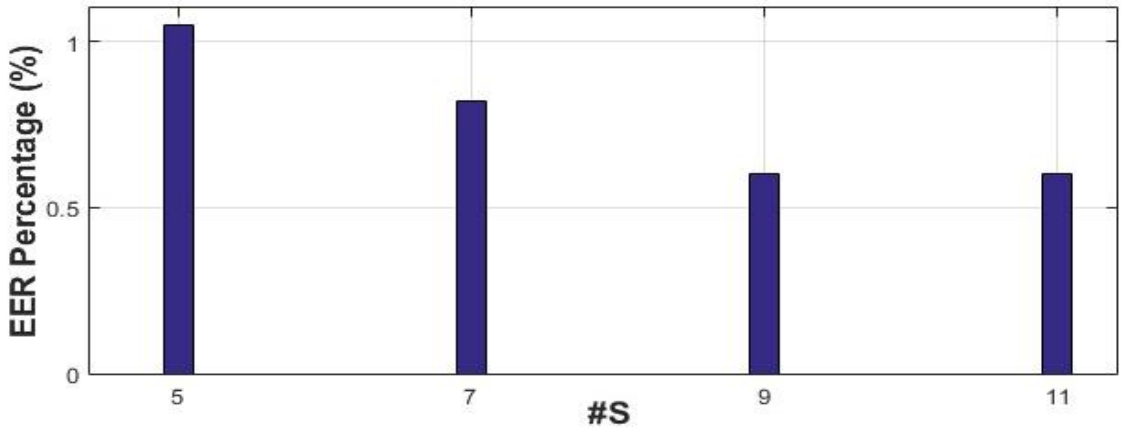


Figure 42: Voting results using different number of samples

Figure 42 clearly demonstrates that the lowest EER was 0.60% for #S=9 (i.e., 90 seconds of real movement data) and nearly similar when #S=11 compared to EERs of 1.06% and 0.82%, for #S of 5 and 7 respectively. By employing the majority schema, the reported EERs were much less compared with the single sample mode (i.e., 10 seconds data). For instance, at best 3.55% of EER was reported by utilizing 10 seconds real life NW signal while this error was

significantly dropped down into 0.60% by using the aforementioned schema. Therefore, a series of experiments were carried out to examine the effect of the majority voting on the system performance (using #S=9) and the findings are displayed on Table 21.

Activity type	Sensor type	EER (%) of using majority voting	NF	EER (%) of using single sample	NF
NW	Acc	0.73	60	4.35	60
NW	Gyr	1.07	60	8.96	60
NW	Fusion	0.60	70	3.55	110
FW	Acc	0	50	1.24	60
FW	Gyr	0.34	50	5.66	70
FW	Fusion	0	50	0.92	120
Non-W	Acc	4.95	30	7.04	30
Non-W	Gyr	7.20	30	11.25	20
Non-W	Fusion	3.37	30	5.31	30

Table 21: The best EERs with and without using the majority voting.

As expected, the majority voting scheme greatly improved the authentication performances for the captured activities. It also showed that a single sensor can effectively recognize the legitimate user and rejecting imposter with invaluable recognition rates. For instance, accelerometer-based authentication reported low EERs of 0.73%, 0%, and 4.95% for the NW, FW, and Non-W activities respectively (compared to 4.35%, 1.24%, and 7.4% when the system decision was based upon 10 seconds duration). Further improvements were achieved with EERs of 0.60% and 3.37% for the NW and Non-W respectively by using the fusion sensor approach (and the error rate of the FW was not affected and remained 0%). In addition to the significant improvement on the authentication accuracy, using the majority voting schema required less features to attain a low EER (e.g., the feature subset size of NW and FW were reduced from 110 and 120 into 70 and 50 features respectively). As a result, the proposed system can be implemented in in a more efficient and less time-consuming manner.

To investigate the stability and reliability of the results over all users, the individual user performance was further analysed to find out if particular users reported high

EER. The conducted analysis was based upon selecting the NW and Non-Walking activity as the FW data achieved an EER of 0% for all users. Figures 43 and 44 display the EERs for each of the 30 users and compare the user's performance with and without utilizing the majority voting (and the results of the fusion approach were presented in the Figures as it showed better accuracies compared to the acceleration and gyroscope sensors).

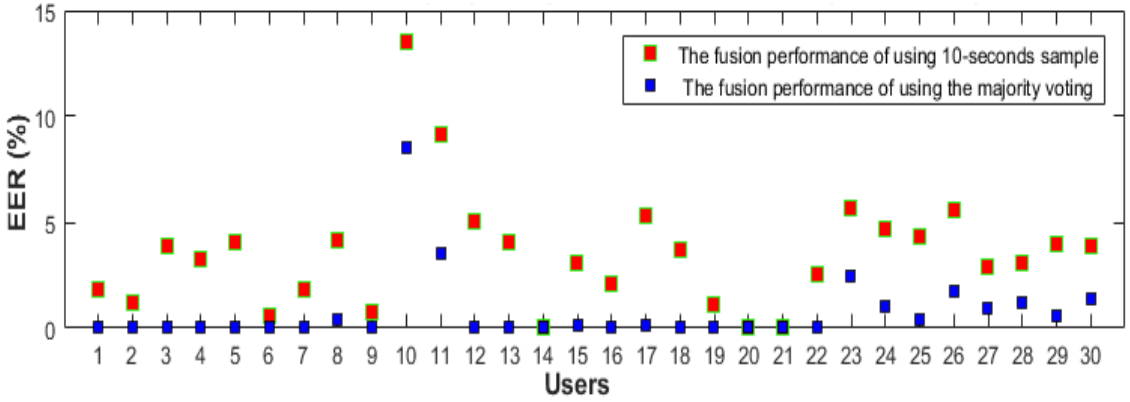


Figure 43: The single sample mode vs majority voting results separated by users using the NW data

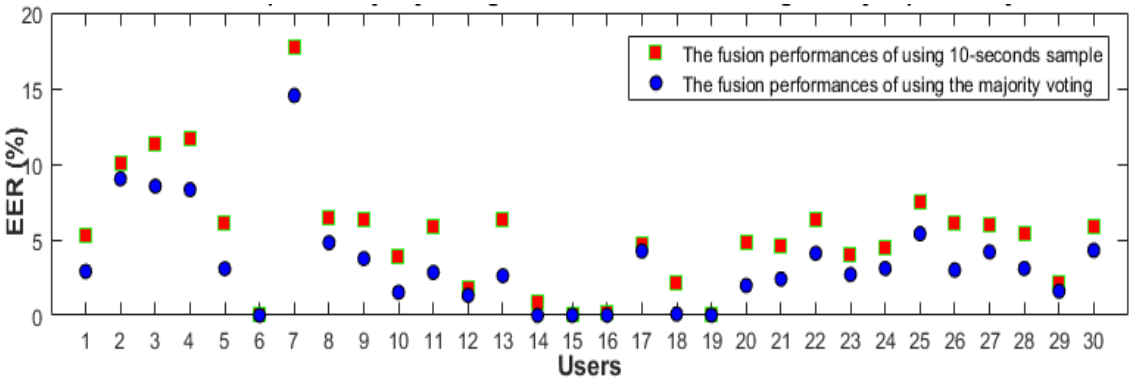


Figure 44: The single sample mode vs majority voting results separated by users using the Non-Walking data

It can be seen from Figure 43 that employing the majority voting for the NW activity resulted in an EER of 0% for two third of the users (e.g., the users 1, 2, 3, 4, 5, 7, 9, and 12) and low EERs ranging between 1% and 2% for the rest of users (apart from the users 10 and 11). Even for the users 10 and 11, the majority voting significantly reduced the EERs from 10% and 14.5% to 4% and 9% respectively

(which is an improvement of more than 60% and 40% over the individual user performance). The positive impact of the majority voting is also highlighted and shown in Figure 44 for the Non-Walking activity, where the performance for all users were improved (apart from the users 6, 15, and 16 as their EERs were already 0% before employing the majority schema). It is apparent from Figure 44 that even after using the majority voting, the ERRs of some users (in particular, users 2, 3, 4, and 7) were not good as other user' performances (ranging from 9% to 15%). The primary cause of this outcome is that the proposed activity detection method is heavily reliant on the user's gait information and less effective to identify the actual activity when the user is not walking. Therefore, more advanced method is required to predict the activity type rather than dividing the user's movement data into gait and non-gait information.

6.5 Conclusion

This study shows that activity-based user authentication is a viable means for verifying the user's identity by evaluating the system under the most realistic dataset (i.e., real-life data was collected over multiple days). It does show that the system performance could be improved significantly by using the majority voting approach. It is argued that using the majority approach would require more time to make a decision by the system (as it depends on a number of results rather than each individual result). Nevertheless, the best results of aforementioned approach were achieved by using 50 and 70 features only for the FW and NW data respectively (compared to 110 and 120 features when the single sample mode was considered). The segment size and the amount of the training samples were also investigated. By performing in-depth analysis for the suitable segment size and the required training samples, it was found that that 10 seconds of data is sufficient for performing TAS and training the classifier with samples that were

captured over four to five days could be sufficient to avoid the overfitting problem and construct a robust reference template.

7 Conclusions and Future Work

This chapter presents a brief discussion about the main accomplished contributions and highlights the shortcomings of this study. Subsequently, suggestions and scope for future work to secure smart devices in a transparent and continuous manner are also highlighted.

7.1 Achievements of the research

Considerable progress has been made in order to offer a robust and useable biometric-based user authentication system for smartwatch devices. The reported findings attained the overall objectives of this research, which were highlighted in Chapter 1 and the full achievements are described below:

- Having understood the feasibility of activity-based user authentication using smartwatches; this was achieved by conducting a comprehensive analysis of the prior art on gait and gesture authentication using dedicated, mobile, and smartwatch sensors (this is highlighted in Chapter 4).

- Evaluating the recognition performance across a range of activities and examination of the most effective classification strategy (i.e., single or multi classifier approach). The single and cross day evaluation methodologies were also explored. By conducting extensive experiments, several time and frequency domain were extracted from the acceleration and gyroscope data and the impact of these features was highlighted (Chapter 5).

- Exploring the use of static and dynamic feature vectors and has proposed a new feature vector mechanism that maximize the system performance and successfully reduce the user's reference template size (this is highlighted in Chapter 5 and 6).

- Capturing a real-life data over multiple days- rather than using test data collected under laboratory conditions to ensure the captured signals can be used for real practical authentication system (Chapter 6).

- Proposing a light activity detection approach that in order to predict the user's activity for better training practice hence, the system can effectively verify the user's pattern. Individual sensor performance and the fusion of both sensors were also explored, and the findings outperform the prior accelerometer -based studies that used unrealistic setup (i.e., laboratory dataset captured within controlled environment) (Chapter 6).

- Conducting a comprehensive analysis of three important parameters (i.e., the segment size, the training sample size, and the majority voting) with the aim of showing the best system configurations that could enhance the authentication decisions and determine the requirement for practical system implementation. The aforementioned parameters are critical for TAS and the optimal system configurations practical system implementation are suggested (this is highlighted in Chapter 7).

A number of papers within the research domain have been presented at refereed journal and conferences and a short description for the published papers are summarized below:

- Activity Recognition Using Wearable Computing

  N. Al-Naffakh, N. Clarke, P. Dowland and F. Li, Proceedings of the 11th International Conference for Internet Technology and Secured Transactions (ICITST), Barcelona, 2016, pp. 189-195.

- A Comprehensive Evaluation of Feature Selection for Gait Recognition Using Smartwatches

  N. Al-Naffakh, N. Clarke, P. Dowland and F. Li, International Journal for Information Security Research (IJISR), Volume 6, Issue 3, September 2016.

- Unobtrusive Gait Recognition using Smartwatches

  N. Al-Naffakh, N. Clarke, F. Li, and P. Dowland, Proceedings of the 16 International Conference of the Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany, 2017.

- Continuous User Authentication using Smartwatch Motion Sensor Data

  N. Al-Naffakh, N. Clarke, and F. Li, Proceedings of the 12 International Conference for Trust Management (IFIPTM), Ontario Canada, 2018.

The first two studies (i.e., Al-Naffakh et al., 2016; Al-Naffakh et al., 2017a) explored the feasibility of activity recognition using smartwatches and proposed a feature selection approach that helped to improve the system performance. The latter two studies (i.e., Al-Naffakh et al., 2017b; Al-Naffakh et al., 2018) concentrated on comprehensive analysis that involved collecting the largest dataset in the research area, developing a novel dynamic feature selection approach for each user independently, identifying the optimal source sensor for the authentication task, highlighting the impact of the majority schema on the system accuracy, and the feasibility of using multiple activities Recently, a journal article submitted for publication that included a comprehensive and more realistic investigation by using unlabelled movement data. To this end, it is believed that the research has successfully achieved valid and useful contributions to the biometric-based user authentication field.

7.2 Limitations of the research

Although the overall objectives of the research have been achieved, there is still a number of limitations that are summarised below

- Despite the collected samples of this study are fairly acceptable and, to the best of the author's knowledge, represent the biggest dataset for activity - based user authentication using smartwatch, it would be recommended to capture data from a large number of users (e.g., between 100 to 200) over a prolonged period of time (i.e., months). As a result, the performance of the proposed system was not tested over a long period time to claim an allegation of robustness, despite the literature provided evidence that building the user's reference template over a long period of time could improve the system performance.

- While the proposed activity detection method (Chapter 6) successfully improved the authentication rates, it was limited to divide the real-life data into gait and non-gait samples only. Moreover, the proposed technique does not thoroughly examine or provide a better understanding of the nature of signal, although the findings in Chapter 6 suggesting that identifying the activity type accurately could significantly reduce the EER.

- Given that the aim of this work is verify the user's identity of smart devices and due to the time constraint, designing a framework to evaluate the proposed system could be useful although the development of such a system is considered outside the scope of the research.

7.3 Suggestions and Scope for Future Work

The conducted research by this thesis has successfully presented alternative user authentication solution for smartphones and smartwatches devices. However, a number of ideas has been identified in which a more direct continuation of the research programme could be carried out. The details of future work are listed as follows.

- Developing an application that transparently and continuously collect the user's samples and negligible resources consumption.
- Further investigation is required to explore different feature reduction approaches in order to remove the redundant features that might negatively influence the classification results and consume more computational power.
- More experimental work should be carried out in order to understand how the user's template might be changed over the time and make sure that template will be always appropriate to identify the legitimate user versus other users (i.e., imposters)
- Future work will focus on better optimization such as extracting new features, evaluating different machine learning classifiers (e.g., Random Forest, Naive Bayes, and SVM) and combining the smartphone and the smartwatch movement data.
- Although this study was able to divide the uncontrolled data into gait and non-gait data, a context aware approach could be useful to predict a wider variety of activities hence, improving the recognition rates. For example, using GPS and the calendar, it would be possible to identify not only that an individual is running but that he is running to catch a train to the airport – thus likely to be carrying or pulling luggage. In this scenario, a composite classifier could be

used that not only focusses upon running but running and carrying luggage. This can be achieved by incorporating other sensor-based information (e.g., GPS) to provide some situational awareness of what a user might be doing at a specific point of time.

- Implementing the proposed system in a real-world scenario is required; this can be achieved by developing a framework prototype on the device in order to evaluate the activity-based user authentication technique on live user and analysis real user feedbacks. The storage space of deploying the framework is also important parameter that required further investigation.

7.4 The Future of Activity-Based User Authentication for Smart Devices

Mobile and smartwatch devices have become an irreplaceable part of people's and currently utilized for various purposes (e.g., personal communication, online payment, and office work); also, these devices have increased amount of access to sensitive information such as financial or health records. The use of smartwatch and mobile devices have inherently raised security concerns and there exists a prevalent requirement to secure these devices. Despite several techniques are proposed to recognise the owner's identity, the obtrusive implementations of these methods promote users to take no security precautions against unauthorized access, specifically to the smartwatch subscribers due to the small touch screen of these devices. Therefore, protecting the information and continuously checking the user's identity in a more innovative and convenient fashion is pivotal. To this end, this research has designed a novel activity-based user authentication by utilizing the accelerometer and gyroscope sensors of smartwatch and the findings have positively demonstrated that the proposed system is a feasible approach in achieving reliable transparent authentication.

To conclude, verifying the legitimate user of smartwatch and mobile devices will be crucial in the near future as more applications and services emerge to the smart devices. Therefore, the future will see further growth and expansion to perform user authentication in a continuous and user-friendly fashion.

References

1. Acar, A., Aksu, H., Uluagac, A.S. and Akkaya, K., 2018, May. WACA: Wearable-assisted continuous authentication. In 2018 IEEE Security and Privacy Workshops (SPW) (pp. 264-269).
2. Achlioptas, D. (2003) 'Database-friendly random projections: Johnson-Lindenstrauss with binary coins', Journal of Computer and System Sciences, 66(4), pp. 671–687.
3. Adam, M., Rossant, F. and Mikovicova, B. (2009) 'Iris identification based on a local analysis of the iris texture', in 2009 Proceedings of 6th International Symposium on Image and Signal Processing and Analysis. IEEE, pp. 523–528.
4. Ahmad, M., Alqarni, M.A., Khan, A., Khan, A., Hussain Chauhdary, S., Mazzara, M., Umer, T. and Distefano, S., 2018. Smartwatch-based legitimate user identification for cloud-based secure services. Mobile Information Systems.
5. Al Abdulwahid, A., Clarke, N., Furnell, S. and Stengel, I., 2013. A conceptual model for federated authentication in the cloud. Proceedings of the 11th Australian Information Security Management Conference, Edith Cowan University, pp. 1–11.
6. Fahmi, P.A., Kodirov, E., Choi, D.J., Lee, G.S., Azli, A.M.F. and Sayeed, S., 2012, October. Implicit authentication based on ear shape biometrics using smartphone camera during a call. In 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (pp. 2272-2276).
7. Al-Naffakh, N., Clarke, N., Dowland, P. and Li, F., 2016, December. Activity recognition using wearable computing. In 2016 11th International Conference for Internet Technology and Secured Transactions (ICITST) (pp. 189-195).
8. Al-Naffakh, N., Clarke, N., Li, F. and Haskell-Dowland, P., 2017, September. Unobtrusive gait recognition using smartwatches. In 2017 International Conference of the Biometrics Special Interest Group (BIOSIG) (pp. 1-5). IEEE.
9. Al-Naffakh, N., 2017. A comprehensive evaluation of feature selection for gait recognition using smartwatches. International Journal for Information Security Research, 6(3), pp. 691–700.
10. Al-Naffakh, N., Clarke, N. and Li, F. (2018) 'Continuous user authentication using smartwatch motion sensor data', IFIP Advances in Information and Communication Technology, 528, pp. 15–28.
11. Aloul, F., Zahidi, S. and El-Hajj, W. (2009) 'Two factor authentication using mobile phones', in 2009 IEEE/ACS International Conference on Computer Systems and Applications., pp. 641–644.

12. Alsultan, A. and Warwick, K. (2013) 'Keystroke Dynamics Authentication: A Survey of Free-text Methods', International Journal of Computer Science, 10(4), pp. 1–10.
13. Anil, J. and Sharath, P. (2001) 'Fingerprint Classification and Matching', p. 32.
14. Anthony, S. (2014) In 2015 tablet sales will finally surpass PCs, fulfilling Steve Jobs' post-PC prophecy. Available at: http://www.extremetech.com/computing/18 5937-in-2015-tablet-sales-will-finally-surpass-pcs-fulfilling-steve-jobs-post-pc-prophecy (Accessed: 27 February 2016).
15. Arora, P. (2015) 'Survey on Human Gait Recognition', 3(3), pp. 492–497. Available at: http://pnrsolution.org/Datacenter/Vol3/Issue3/66.pdf.
16. Arora, P. and Gandhi, V. C. (2014) 'Survey on Human Gait Recognition', 5(4), pp. 492–497. Available at: http://www.ijcst.com/vol54/3/73-Avani-P-Patel.pdf.
17. Ashbourn, J. (2000) 'Biomtrics Advanced Identity Verification', in. Springer.
18. Aune, S. P. (2011) Google Introduces Face Unlock for Android Ice Cream Sandwich. Available at: http://www.technobuffalo.com/2011/10/18/google-introduces-face-unlock-for-android-ice-cream-sandwich/ (Accessed: 7 May 2016).
19. Aupy, A. and Clarke, N. (2005) 'User Authentication by Service Utilisation Profiling', Proceedings of the ISOneWorld 2005.
20. Aviv, A.J., Gibson, K.L., Mossop, E., Blaze, M. and Smith, J.M., 2010. Smudge attacks on smartphone touch screens. Woot, 10, pp.1-7.
21. Baca, M., Grd, P. and Fotak, T. (2012) 'Basic Principles and Trends in Hand Geometry and Hand Shape Biometrics', in New Trends and Developments in Biometrics. InTech.
22. Banerjee, S. P. and Woodard, D. (2012) 'Biometric Authentication and Identification Using Keystroke Dynamics: A Survey', Journal of Pattern Recognition Research, 7(1), pp. 116–139.
23. Benzekki, K., El Fergougui, A. and Elalaoui, A. E. B. (2018) 'A context-aware authentication system for mobile cloud computing', Procedia Computer Science. Elsevier B.V., 127, pp. 379–387. doi: 10.1016/j.procs.2018.01.135.
24. Bhagavatula, R., Ur, B., Iacovino, K., Kywe, S.M., Cranor, L.F. and Savvides, M., 2015. Biometric authentication on iphone and android: Usability, perceptions, and influences on adoption.
25. Bhushan, A., Kalyani2, P. and Supriya, J. (2001) 'Handwritten Script Recognition', Journal of Computer Engineering (IOSR-JCE), pp. 30–33.

26. Boodoo, N. B. and Subramanian, R. K. (2009) 'Robust Multi biometric Recognition Using Face and Ear Images', International Journal of Computer Science and Information Security, 6(2), pp. 164–169.

27. Buchoux, A. and Clarke, N. L. (2008) 'Smartphone Deployment of Keystroke Analysis', pp. 190–197. doi: 10.1.1.678.9212.

28. Bursztein, E. (2014) Survey: Most people don't lock their Android phones but should. Availableat:https://www.elie.net/blog/survey-most-people-dont-lock-their-android phones -but-should (Accessed: 5 May 2016).

29. Chang, T. Y., Tsai, C. J. and Lin, J. H. (2012) 'A graphical-based password keystroke dynamic authentication system for touch screen handheld mobile devices', Journal of Systems and Software. Elsevier Inc., 85(5), pp. 1157–1165.

30. Chiasson, S. and Biddle, R. (2007) 'Issues in User Authentication', (April), pp. 1–4.

31. Choraś, M. (2005) 'Ear Biometrics Based on Geometrical Method of Feature Extraction', in Electronic Letters on Computer Vision and Image Analysis, pp. 51–61. doi: 10.1007/ 978-3-540-30074-8_7.

32. Clarke, N. (2011) Transparent User Authentication Biometrics, RFID and Behavioural Profiling, Transparent User Authentication. London: Springer London.

33. Clarke, N. L. and Furnell, S. M. (2007) 'Advanced user authentication for mobile devices', Computers & Security, 26(2), pp. 109–119.

34. Clarke, N., Karatzouni, S. and Furnell, S. (2008) 'Transparent facial recognition for mobile devices', Proceedings of the 7th Security Conference.

35. Cola, G., Avvenuti, M., Musso, F. and Vecchio, A., 2016. Gait-based authentication using a wrist-worn device. In Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (pp. 208-217).

36. Costello, K. (2018) Gartner Says Worldwide Wearable Device Sales to Grow 26 Percent in 2019. Available at: https://www.gartner.com/en/newsroom/press-releases /2018-11-29-gartner-says-worldwide-wearable-device-sales-to-grow- (Accessed: 21 June 2019).

37. Damaševičius, R., Maskeliūnas, R., Venčkauskas, A. and Woźniak, M., 2016. Smartphone user identity verification using gait characteristics. Symmetry, 8(10), p.100.

38. Daugman, J. (2009) 'Iris Recognition at Airports and Border-Crossings', 25(12). Available at:http://www.cl.cam.ac.uk/~jgd1000/Iris_Recognition_at_Airports_and _Border-Crossin gs.pdf.

39. Dave, C. (2019) Mobile marketing statistics compilation. Available at: https://www.smartinsights.com/mobile-marketing/mobile-marketing-analytics/mobile-marketing-statistics/ (Accessed: 11 April 2018).

40. Davidson, S., Smith, D., Yang, C. and Cheah, S., 2016. Smartwatch user identification as a means of authentication. Department of Computer Science and Engineering Std.

41. Derawi, M.O., 2010. Accelerometer-based gait analysis, a survey. Available at: http://derawi.com/cv/publications/derawi_nisnet_nisk_gaitsurvey.pdf.

42. Derawi, M. O. (2012) Smartphones and Biometrics: Gait and Activity Recognition. Available at: http://hdl.handle.net/11250/144369.

43. Derawi, M.O., Nickel, C., Bours, P. and Busch, C., 2010, October. Unobtrusive user-authentication on mobile phones using biometric gait recognition. In 2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (pp. 306-311). IEEE.

44. Ehatisham-ul-Haq, M., Azam, M.A., Loo, J., Shuang, K., Islam, S., Naeem, U. and Amin, Y., 2017. Authentication of smartphone users based on activity recognition and mobile sensing. Sensors, 17(9), p.2043.

45. Ehatisham-ul-Haq, M., Azam, M.A., Naeem, U., ur Rèhman, S. and Khalid, A., 2017. Identifying Smartphone Users based on their Activity Patterns via Mobile Sensing. Procedia computer science, 113, pp.202-209.

46. Enge, E., 2019. Mobile Vs. Desktop Usage In 2019. [online] Available at: <https://www.perficient.com/insights/research-hub/mobile-vs-desktop-usage-study> [Accessed 14 June 2020].

47. Faundez-Zanuy, M. (2007) 'On-line signature recognition based on VQ-DTW', Pattern Recognition, 40(3), pp. 981-992.

48. Feng, T., Yang, J., Yan, Z., Tapia, E.M. and Shi, W., 2014. Tips: Context-aware implicit user identification using touch screen in uncontrolled environments. In Proceedings of the 15th Workshop on Mobile Computing Systems and Applications (pp. 1-6).

49. Finextra (2018) UK contactless mobile payments hit tipping point. Available at: https://www.finextra.com/newsarticle/31753/uk-contactless-mobile-payments-hit-tipping-point (Accessed: 20 June 2019).

50. Firstpost (2018) Over 50 % of smartphone users don't use passwords or anti-theft solutions. Available at: https://www.firstpost.com/tech/news-analysis/over-50-of-smartphone-users-dont-use-passwords-or-anti-theft-solutions-4681761.html (Accessed: 9 March 2019).

51. Gafurov, D. (2007a) ‘A Survey of Biometric Gait Recognition : Approaches , Security and Challenges Davrondzhon Gafurov Gjøvik University College Biometric system’, Gjøvik University College, (NIKConference).

52. Gafurov, D. (2007b) ‘New gait recognition method using Kinect stick figure and CBIR’, in. Available at: http://www.nik.no/2007/14-Gafurov.

53. Gafurov, D. and Snekkenes, E. (2008) ‘Towards understanding the uniqueness of gait biometric’, in 2008 8th IEEE International Conference on Automatic Face & Gesture Recognition. IEEE, pp. 1–8.

54. Gafurov, D. and Snekkenes, E. (2009) ‘Gait Recognition Using Wearable Motion Recording Sensors’, EURASIP Journal on Advances in Signal Processing, 2009(1), p. 415817..

55. Gafurov, D. and Snekkkenes, E. (2008) ‘Arm Swing as a Weak Biometric for Unobtrusive User Authentication’, in 2008 International Conference on Intelligent Information Hiding and Multimedia Signal Processing. IEEE, pp. 1080–1087.

56. Gafurov, D., Helkala, K. and Søndrol, T. (2006) ‘Biometric gait authentication using accelerometer sensor’, Journal of Computers (Finland), 1(7), pp. 51–59.

57. Gafurov, D., Snekkenes, E. and Bours, P. (2007a) ‘Gait Authentication and Identification Using Wearable Accelerometer Sensor’, in 2007 IEEE Workshop on Automatic Identification Advanced Technologies. IEEE, pp. 220–225.

58. Gafurov, D., Snekkenes, E. and Bours, P. (2007b) ‘Spoof Attacks on Gait Authentication System’, IEEE Transactions on Information Forensics and Security, 2(3), pp. 491–502.

59. Gafurov, D., Snekkenes, E. and Bours, P. (2010) ‘Improved Gait Recognition Performance Using Cycle Matching’, in 2010 IEEE 24th International Conference on Advanced Information Networking and Applications Workshops. IEEE, pp. 836–841.

60. Gafurov, D., Snekkenes, E. and Buvarp, T. E. (2006) ‘Robustness of Biometric Gait Authentication Against Impersonation Attack’, in On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops, pp. 479–488. doi: 10.1007/11915034_71.

61. Gamassi, M., Lazzaroni, M., Misino, M., Piuri, V., Sana, D. and Scotti, F., 2004, May. Accuracy and performance of biometric systems. In Proceedings of the 21st IEEE Instrumentation and Measurement Technology Conference (IEEE Cat. No. 04CH37510) (Vol. 1, pp. 510-515).

62. Gascon, H., Uellenbeck, S., Wolf, C. and Rieck, K., 2014. Continuous authentication on mobile devices by analysis of typing motion behavior. Sicherheit 2014–Sicherheit, Schutz und Zuverlässigkeit.

63. Greg, S. (2017) Report: 57% of traffic now from smartphones and tablets - Search Engine Land. Available at: https://searchengineland.com/report-57-percent-traffic-now-smartph ones-tablets-281150 (Accessed: 20 June 2019).

64. Griswold-Steiner, I., Matovu, R. and Serwadda, A. (2017) 'Handwriting watcher: A mechanism for smartwatch-driven handwriting authentication', IEEE International Joint Conference on Biometrics, IJCB 2017, 2018-Janua, pp. 216–224.

65. Habib, K. and Leister, W. (2015) 'Context-Aware Authentication for the Internet of Things', The Eleventh International Conference on Autonomic and Autonomous Systems fined, (c), pp. 134–139.

66. Hayden (2015) Random massive battery drain on Sony SW3 with 5.1.1. Available at:https://productforums.google.com/forum/#!topic/android-wear/mY0678x_a98; context-place=forum/android-wear (Accessed: 10 May 2016).

67. Hestbek, M. R., Nickel, C. and Busch, C. (2012) 'Biometric gait recognition for mobile devices using wavelet transform and support vector machines', pp. 205–210.

68. Ho, C.C., Eswaran, C., Ng, K.W. and Leow, J.Y., 2012, December. An unobtrusive Android person verification using accelerometer based gait. In Proceedings of the 10th International Conference on Advances in Mobile Computing & Multimedia (pp. 271-274).

69. Hoang, T., Nguyen, T.D., Luong, C., Do, S. and Choi, D., 2013. Adaptive Cross-Device Gait Recognition Using a Mobile Accelerometer. JIPS, 9(2), p.333.

70. Hocking, C.G., Furnell, S.M., Clarke, N.L. and Reynolds, P.L., 2013. Co-operative user identity verification using an Authentication Aura. Computers & security, 39, pp.486-502.

71. Hollingsworth, K., Baker, S., Ring, S., Bowyer, K.W. and Flynn, P.J., 2009, May. Recent research results in iris biometrics. In Optics and Photonics in Global Homeland Security V and Biometric Technology for Human Identification VI (Vol. 7306, p. 73061Y). International Society for Optics and Photonics.

72. Hurley, D. J., Arbab-Zavar, B. and Nixon, M. S. (2007) 'The Ear as a Biometric', in Handbook of Biometrics. Boston, MA: Springer US, pp. 131–150.

73. Ichikawa, F., Chipchase, J. and Grignani, R. (2005) 'Where's The Phone? A Study of Mobile Phone Location in Public Spaces', in 2005 2nd Asia Pacific Conference on Mobile Technology, Applications and Systems. IEEE, pp. 1–8.

74. Jadhao, P. and Dole, L. (2013) 'Survey on Authentication Password Techniques', International Journal of Soft Computing and Engineering, (2), pp. 67–68.

75. Jain, A. K., Ross, A. and Prabhakar, S. (2004) 'An Introduction to Biometric Recognition', 14(1), pp. 1–29.

76. Jain, A., Flynn, P. and Ross, A. A. (2008) Handbook of Biometrics. Springer.

77. James, T. (2017) iPhone X's facial recognition is not for children under 13, says Apple. Available at: https://www.telegraph.co.uk/technology/2017/09/27/iphone-xs-facial-recogn ition-not-children-13-says-apple/ (Accessed: 21 June 2018).

78. Jesudoss, a and Subramaniam, N. P. (2014) 'A SURVEY ON AUTHENTICATION ATTACKS AND COUNTERMEASURES IN A DISTRIBUTED ENVIRONMENT', Indian Journal of Computer Science and Engineering, 5(2), pp. 71–77.

79. Johnston, A. H. and Weiss, G. M. (2015) 'Smartwatch-based biometric gait recognition', in 2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS). IEEE, pp. 1–6.

80. Junshuang Yang, Yanyan Li and Mengjun Xie (2015a) 'MotionAuth: Motion-based authentication for wrist worn smart devices', in 2015 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops). IEEE, pp. 550–555.

81. Junshuang Yang, Yanyan Li and Mengjun Xie (2015b) 'MotionAuth: Motion-based authentication for wrist worn smart devices', in 2015 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops). IEEE, pp. 550–555.

82. Junshuang Yang, Yanyan Li and Mengjun Xie (2015c) 'MotionAuth: Motion-based authentication for wrist worn smart devices', in 2015 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops). IEEE, pp. 550–555.

83. K.Rajasri, S.Sathiyadevi and S.Tamilarasi (2013) 'A Survey on Biometric Recognition Techniques and Algorithms', International Journal of Science, Engineering and Technology Research (IJSETR), 2(4), pp. 5708–5711.

84. Karakaya, M., Bostan, A. and Gökçay, E. (2016) 'How Secure is Your Smart Watch?', International Journal of Information Security Science, 5(4), pp. 90–95.

85. Karnan, M., Akila, M. and Krishnaraj, N. (2011) 'Biometric personal authentication using keystroke dynamics: A review', Applied Soft Computing. Elsevier B.V., 11(2), pp. 1565–1573.

86. Karthikeyan, S., Feng, S., Rao, A. and Sadeh, N., 2014. Smartphone fingerprint authentication versus pins: A usability study (cmu-cylab-14-012).

87. Kaspersky (2018) Half of Consumers Don't Password-Protect their Mobile Devices. Available at: https://www.securitymagazine.com/articles/89220-half-of-consumers - dont-password-protect-their-mobile-devices (Accessed: 10 September 2018).

88. Kaur, G., Singh, G. and Kumar, V. (2014) 'A Review on Biometric Recognition', International Journal of Bio-Science and Bio-Technology, 6(4), pp. 69–76.

89. KC, S. and Nattee, C. (2010) 'A Comprehensive Survey on On-line Handwriting Recognition Technology and its Real Application to the Nepalese Natural Handwriting', Kathmandu University Journal of Science, Engineering and Technology, 5(1), pp. 31–55.

90. Khan, W. Z., Aalsalem, M. Y. and Xiang, Y. (2011) 'A Graphical Password Based System for Small Mobile Devices', International Journal of Computer Science Issues, 8(5), pp. 145–154.

91. Khaw, P. (2002) 'Iris Recognition Technology for Improved Authentication', Information Security.

92. Khursheed, F. and Mir, A. H. (2014) 'AR Model Based Human Identification using Ear Biometrics', International Journal of Signal Processing, Image Processing and Pattern Recognition, 7(3), pp. 347–360.

93. Kim, I. (2012) 'Keypad against brute force attacks on smartphones', IET Information Security, 6(2), p. 71.

94. Krupp, A., Rathgeb, C. and Busch, C. (2013) 'Social acceptance of biometric technologies in Germany: A survey', Lecture Notes in Informatics (LNI), Proceedings - Series of the Gesellschaft fur Informatik (GI), P-212, pp. 193–200.

95. Kulkarni, R. and Namboodiri, A. (2014) 'One-Time Biometric Token based Authentication', in Proceedings of the 2014 Indian Conference on Computer Vision Graphics and Image Processing - ICVGIP '14. New York, USA: ACM Press, pp. 1–7.

96. Kumar, R., Kundu, P.P., Shukla, D. and Phoha, V.V., 2017, October. Continuous user authentication via unlabeled phone movement patterns. In 2017 IEEE International Joint Conference on Biometrics (IJCB) (pp. 177-184).

97. Kumar, R., Phoha, V. V and Raina, R. (2016) 'Authenticating users through their arm movement patterns', (July). Available at: http://arxiv.org/abs/1603.02211.

98. Kwapisz, J. R., Weiss, G. M. and Moore, S. a (2010) 'Cell phone-based biometric identification', in 2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS), pp. 1–7.

99. Lashkari, A.H., Farmand, S., Zakaria, D., Bin, O. and Saleh, D., 2009. Shoulder surfing attack in graphical password authentication. arXiv preprint arXiv:0912.0951.

100. Lau, H. and Tong, K. (2008) 'The reliability of using accelerometer and gyroscope for gait event identification on persons with dropped foot', Gait and Posture, 27(2), pp. 248–257.

101. Lee, W.-H. and Lee, R. (2017) 'Implicit Sensor-based Authentication of Smartphone Users with Smartwatch'.

102. Lee, W.H., Liu, X., Shen, Y., Jin, H. and Lee, R.B., 2017, June. Secure pick up: Implicit authentication when you start using the smartphone. In Proceedings of the 22nd ACM on Symposium on Access Control Models and Technologies (pp. 67-78).

103. Lemos, R. (2019) Watch Out: Smartwatches need smarter security. Available at: https://techbeacon.com/app-dev-testing/watch-out-5-reasons-smartwatches-need-smarter-security (Accessed: 21 June 2019).

104. Leswing, K. (2017) Apple says the iPhone X's Face ID is less accurate on kids under 13. Available at: https://www.businessinsider.com/apple-says-the-iphone-xs-face-id-is-less-accurate-on-kids-under-13-2017-9?r=US&IR=T (Accessed: 21 June 2019).

105. Lewis, A., Li, Y. and Xie, M. (2016) 'Real time motion-based authentication for smartwatch', 2016 IEEE Conference on Communications and Network Security, CNS 2016. IEEE, pp. 380–381.

106. Li, F. (2012) 'Behaviour Profiling for Mobile Devices', (February), pp. 1–216. Available at: https://pearl.plymouth.ac.uk//handle/10026.1/1025.

107. Li, F., Clarke, N., Papadaki, M. and Dowland, P., 2011. Misuse detection for mobile devices using behaviour profiling. International Journal of Cyber Warfare and Terrorism (IJCWT), 1(1), pp.41-53.

108. Li, F., Clarke, N., Papadaki, M. and Dowland, P., 2014. Active authentication for mobile devices utilising behaviour profiling. International journal of information security, 13(3), pp.229-244.

109. Liang, G.-C., Xu, X.-Y. and Yu, J.-D. (2017) 'User-Authentication on Wearable Devices Based on Punch Gesture Biometrics', ITM Web of Conferences, 11, p. 01003.

110. Liarna, L. P. (2018) What are app leaks and how do I know if my apps are leaking data? Available at: https://www.wandera.com/mobile-security/app-and-data-leaks/app-leaks/ (Accessed: 20 November 2018).

111. Lifestylegroup (2011) Data stored on a phone more precious than the phone itself. Available at: http://www.lifestylegroup.co.uk/content/Data-stored-on-a-phone-more-prec ious-than-the-phone-itself.html (Accessed: 27 February 2016).

112. Liszewski, A. (2015) Your Smartwatch's Motion Sensors Can Reveal Everything You Type (Including Passwords). Available at: http://gizmodo.com/your-smartwatchs-motion-sensors-can-reveal-everything-y-1750442236 (Accessed: 2 May 2016).

113. López, X. S. P.-S. (2018) Overfitting in Statistics. Available at: https://xaperez sindin.com /2018/03/13/overfitting-in-statistics/ (Accessed: 23 June 2019).

114. Luo, J.-N. and Yang, M.-H. (2015) 'A mobile authentication system resists to shoulder-surfing attacks', Multimedia Tools and Applications, (January), p. 197.

115. Maltoni, D., Maio, D., Jain, A.K. and Prabhakar, S., 2009. Handbook of fingerprint recognition. Springer Science & Business Media.

116. Mantyjarvi, J., Lindholm, M., Vildjiounaite, E., Makela, S.M. and Ailisto, H.A., 2005. Identifying users of portable devices from gait pattern with accelerometers. In Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing. (Vol. 2, pp. ii-973).

117. Mare, S., Markham, A.M., Cornelius, C., Peterson, R. and Kotz, D., 2014, May. Zebra: Zero-effort bilateral recurring authentication. In 2014 IEEE Symposium on Security and Privacy (pp. 705-720).

118. Dong, J. and Cai, Z., 2016. User authentication using motion sensor data from both wearables and smartphones. In Chinese Conference on Biometric Recognition (pp. 756-764). Springer, Cham.

119. Mayhew, S. (2012) Explainer: Verification vs. Identification Systems. Available at: http://www.biometricupdate.com/201206/explainer-verification-vs-identification-systems (Accessed: 10 June 2015).

120. Mir, A., Rubab, S. and Jhat, Z. (2011) 'Biometrics verification: a literature survey', International Journal of Computing and ICT Research, 5(2), pp. 67–80. Available at: http://docs.mak.ac.ug/sites/default/files/Volume 5 Issue2.pdf#page=67.

121. Monrose, F. and Rubin, A. D. (2000) 'Keystroke dynamics as a biometric for authentication', Future Gener. Comput. Syst., 16(4), pp. 351–359.

122. Moren, D. (2015) Face Recognition Security, Even With A 'Blink Test,' Is Easy To Trick. Available at: http://www.popsci.com/its-not-hard-trick-facial-recognition-security (Access ed: 7 May 2016).

123. Muaaz, M. and Mayrhofer, R. (2013) 'An Analysis of Different Approaches to Gait Recognition Using Cell Phone Based Accelerometers', in Proceedings of International Conference on Advances in Mobile Computing & Multimedia - MoMM '13. New York, USA: ACM Press, pp. 293–300.

124. Muaaz, M. and Mayrhofer, R. (2014) 'Orientation Independent Cell Phone Based Gait Authentication', in Proceedings of the 12th International Conference on Advances in Mobile Computing and Multimedia - MoMM '14. New York, New York, USA: ACM Press, pp. 161–164.

125. Muaaz, M. and Nickel, C. (2012) 'Influence of different walking speeds and surfaces on accelerometer-based biometric gait recognition', in 2012 35th International Conference on Telecommunications and Signal Processing (TSP). IEEE, pp. 508–512.

126. Nanavati, S., Thieme, M. and Nanavati, R. (2002) Biometrics: Identity Verification in a Networked World.

127. Nazarian, R. (2015) Hackers could steal Android users' fingerprints: HTC and Samsung comment. Available at: http://www.digitaltrends.com/mobile/hackers-can-steal-fingerprin ts-android-phones/ (Accessed: 22 January 2016).

128. Ngo, T.T., Makihara, Y., Nagahara, H., Mukaigawa, Y. and Yagi, Y., 2014. The largest inertial sensor-based gait database and performance evaluation of gait-based personal authentication. Pattern Recognition, 47(1), pp.228-237.

129. Nickel, C. and Busch, C. (2011) 'Classifying accelerometer data via Hidden Markov Models to authenticate people by the way they walk', in 2011 Carnahan Conference on Security Technology. IEEE, pp. 1–6.

130. Nickel, C. and Busch, C. (2012) 'Does a cycle-based segmentation improve accelerometer-based biometric gait recognition?', in 2012 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA). IEEE, pp. 746–751.

131. Nickel, C., Brandt, H. and Busch, C. (2011a) 'Benchmarking the performance of SVMs and HMMs for accelerometer-based biometric gait recognition', in 2011 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), pp. 281–286.

132. Nickel, C., Brandt, H. and Busch, C. (2011b) 'Classification of Acceleration Data for Biometric Gait Recognition on Mobile Devices', Special Interest Group on Biometrics and Electronic Sginatures, pp. 57–66.

133. Nickel, C., Busch, C., Rangarajan, S. and Möbius, M., 2011, March. Using hidden markov models for accelerometer-based biometric gait recognition. In 2011 IEEE 7th International Colloquium on Signal Processing and its Applications (pp. 58-63).

134. Nickel, C., Derawi, M.O., Bours, P. and Busch, C., 2011, May. Scenario test of accelerometer-based biometric gait recognition. In 2011 Third International Workshop on Security and Communication Networks (IWSCN) (pp. 15-21). IEEE.

135. Nickel, C., Wirtl, T. and Busch, C. (2012) 'Authentication of Smartphone Users Based on the Way They Walk Using k-NN Algorithm', in 2012 Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing. IEEE, pp. 16–20.

136. O'Gorman, L. (2003) 'Comparing passwords, tokens, and biometrics for user authentication', Proceedings of the IEEE, 91(12), pp. 2021–2040.

137. Okumura, F., Kubota, A., Hatori, Y., Matsuo, K., Hashimoto, M. and Koike, A., 2006. A study on biometric authentication based on arm sweep action with acceleration sensor. In the IEEE International Symposium on Intelligent Signal Processing and Communications (pp. 219-222).

138. Oz, C., Ercal, F. and Demir, Z. (2003) 'Signature recognition and verification with ANN', Proceeding of the Third, 7782, pp. 1–5. Available at: https://mospace. library.umsyste m.edu/xmlui/handle/10355/32109.

139. Palmer, D. (2017) Face, fingerprint, passwords, or PIN: What's the best way to keep your smartphone secure? Available at: https://www.zdnet.com/article/face-fingerprint-passwords-or-pin-whats-the-best-way-to-keep-your-smartphone-secure/(Accessed:21 June 2019).

140. Jonsson, P., Carson, S., Blennerud, G., Shim, J., Arendse, B., Husseini, A., Lindberg, P. and Öhman, K., 2019. Ericsson Mobility Report. [online] Available at: <https://www.ericsson.com/4acd7e/assets/local/mobility-report/documents/2019/emr-november-2019.pdf> [Accessed 14 June 2020].

141. PIERRE, V. (2019) What to do if your smartphone gets stolen. Available at: https://www.androidpit.com/what-to-do-if-your-smartphone-gets-stolen (Accessed : 20 June 2019).

142. Plamondon, R. and Srihari, S. N. (2000) 'Online and off-line handwriting recognition: a comprehensive survey', IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(1), pp. 63–84.

143. Polin, M. Z. H., Kabir, A. N. M. E. and Sadi, M. S. (2012) '2D human-ear recognition using geometric features', in 2012 7th International Conference on Electrical and Computer Engineering. IEEE, pp. 9–12.

144. Phaneuf, A., 2020. Latest Trends In Medical Monitoring Devices And Wearable Health Technology. [online] Available at: <https://www.businessinsider.com/weara ble-technology-healthcare-medical-devices?r=US&IR=T>[Accessed 14 June 2020].

145. Primo, A., Phoha, V.V., Kumar, R. and Serwadda, A., 2014. Context-aware active authentication using smartphone accelerometer measurements. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 98-105).

146. Rachubiński, M. (2009) 'Iris Identification Using Geometrical Wavelets', in 2009 Proceedings of 6th International Symposium on Image and Signal Processing and Analysis. IEEE, pp. 322–332.

147. Ratha, N. K., Connell, J. H. and Bolle, R. M. (2001) 'Enhancing security and privacy in biometrics-based authentication systems', IBM Systems Journal, 40(3), pp. 614–634.

148. Raza, M., Iqbal, M., Sharif, M. and Haider, W., 2012. A survey of password attacks and comparative analysis on methods for secure authentication. World Applied Sciences Journal, 19(4), pp.439-444.

149. Ricci, J., Baggili, I. and Breitinger, F. (2017) 'Watch What You Wear', in, pp. 47–73.

150. Smith, R.E., 2001. Authentication: from passwords to public keys. Addison-Wesley Longman Publishing Co., Inc..

151. Robbins (2019) How Many Emails Are Sent Per Day | Campaign Monitor. Available at:https://www.campaignmonitor.com/blog/email-marketing/2019/05/shocking-tru th-about-how-many-emails-sent/ (Accessed: 20 June 2019).

152. Roy, A., Memon, N. and Ross, A., 2017. Masterprint: Exploring the vulnerability of partial fingerprint-based authentication systems. IEEE Transactions on Information Forensics and Security, 12(9), pp.2013-2025.

153. Ross, A. (2013) 'Activity and User Recognition Using Mobile Phone Accelerometer and Gyroscope'.

154. Saevanee, H., Clarke, N. L. and Furnell, S. M. (2012) 'Multi-modal Behavioural Biometric Authentication for Mobile Devices', in IFIP Advances in Information and Communication Technology, pp. 465–474.

155. Choi, S., Youn, I.H., LeMay, R., Burns, S. and Youn, J.H., 2014, February. Biometric gait recognition based on wireless acceleration sensor using k-nearest neighbor classification. In 2014 International Conference on Computing, Networking and Communications (ICNC) (pp. 1091-1095). IEEE.

156. Schlöglhofer, R. and Sametinger, J. (2012) ‘Secure and usable authentication on mobile devices’, in Proceedings of the 10th International Conference on Advances in Mobile Computing & Multimedia - MoMM ’12. New York, New York, USA: ACM Press, p. 257.

157. Scott, M. (2017) Smartwatch Ownership Expected to Increase Nearly 60 Percent Into 2019. Available at: https://www.npd.com/wps/portal/npd/us/news/press-releases/2017/us-smartwatch-ownership-expected-to-increase-nearly-60-percent -into-2019/ (Accessed: 21 June 2019).

158. Shanmugapriya, D. and Padmavathi, G. (2009) ‘A Survey of Biometric keystroke Dynamics: Approaches, Security and Challenges’, International Journal of Computer Science and Information Security, 5(1), p. 5.

159. Sharif, M., Mohsin, S. and Javed, M. Y. (2012) ‘A Survey: Face Recognition Techniques’, 4(23), pp. 41-50.

160. Shen, C., Li, Y., Chen, Y., Guan, X. and Maxion, R.A., 2017. Performance analysis of multi-motion sensor behavior for active smartphone authentication. IEEE Transactions on Information Forensics and Security, 13(1), pp.48-62.

161. Shrestha, B., Mohamed, M. and Saxena, N. (2016) ‘Walk-Unlock: Zero-Interaction Authentication Protected with Multi-Modal Gait Biometrics’. Available at: http://arxiv.org/ abs/1605.00766.

162. Shrestha, B., Saxena, N. and Harrison, J. (2013) ‘Wave-to-Access: Protecting Sensitive Mobile Device Services via a Hand Waving Gesture’, in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 199-217.

163. Singh, Y. N. and Singh, S. K. (2013) ‘A taxonomy of biometric system vulnerabilities and defences’, International Journal of Biometrics, 5(2), p. 137.

164. Sonkamble, S., Thool, R. and Sonkamble, B. (2010) ‘Survey of Biometric Recognition Systems and Their Applications.’, Journal of Theoretical & Applied Information Technology, 11. Available at: http://www.jatit.org/volumes/research-papers/Vol11No1/6Vol11No1.pdf.

165. Stables, J. (2015) Apple Watch review. Available at: http://www.wareable.com/ apple-watch/apple-watch-review (Accessed: 10 May 2016).

166. Steve, S. (2018) Smart watches and internet security: Are my wearables secure? Available at:https://us.norton.com/internetsecurity-iot-how-to-protect-your-connected-wearable s.html(Accessed: 21 June 2019).

167. Subramanian, H. (2004) 'Audio signal classification', M. Tech Credit Seminar Report, pp. 1–17.

168. Sui, Y., Zou, X. and Du, E. Y. (2011) 'Biometrics-Based Authentication: A New Approach', in 2011 Proceedings of 20th International Conference on Computer Communications and Networks (ICCCN). IEEE, pp. 1–6.

169. Summerson, C. (2015) Massive Battery Drain Still An Issue On The Smartwatch 3, Google And Sony Said A Fix Was In The Works Back In June. Available at: http://www.androidpolice.com/2015/09/11/massive-battery-drain-still-an-issue-on-the-smartwatch-3-google-and-sony-said-a-fix-was-in-the-works-back-in-june/ (Accessed: 10 May 2016).

170. Sunnebo, D. (2017) Nearly 16% of US consumers now own wearables. Available at: https://www.kantarworldpanel.com/global/News/Nearly-16-of-US-Consumers-and-9-in-EU4-Now-Own-Wearables (Accessed: 21 June 2019).

171. Tanviruzzaman, M. and Ahamed, S. I. (2014) 'Your Phone Knows You: Almost Transparent Authentication for Smartphones', in 2014 IEEE 38th Annual Computer Software and Applications Conference, pp. 374–383.

172. Teh, P. S., Teoh, A. B. J. and Yue, S. (2013) 'A Survey of Keystroke Dynamics Biometrics', The Scientific World Journal, 2013, pp. 1–24.

173. Templafy (2017) How many emails are sent every day? And other top email statistics your business needs to know - Templafy. Available at: https://www.templafy.com/blog/how-many-emails-are-sent-every-day-top-email-statistics-your-business-needs-to-know/(Acce ssed: 20 June 2019).

174. Tanvi, P., Sonal, G. & Kumar, S.M., 2011. Token Based Authentication Using Mobile Phone. In 2011 International Conference on Communication Systems and Network Technologies. IEEE, pp. 85–88.

175. Walters, R. (2012) Make your PIN code more secure using three unique numbers. Available at:http://www.geek.com/mobile/make-your-pin-more-secure-using-three -unique-numbers-1454201/ (Accessed: 5 May 2016).

176. Wang, Z., Shen, C. and Chen, Y. (2017) 'Handwaving Authentication: Unlocking Your Smartwatch Through Handwaving Biometrics', in Machine Learning Techniques for Gait Biometric Recognition, pp. 545–553.

177. Watanabe, Y. (2014) 'Influence of Holding Smart Phone for Acceleration-Based Gait Authentication', in 2014 Fifth International Conference on Emerging Security Technologies. IEEE, pp. 30–33.

178. Watanabe, Y. (2015) 'Toward Application of Immunity-based Model to Gait Recognition Using Smart Phone Sensors: A Study of Various Walking States', Procedia Computer Science. Elsevier Masson SAS, 60, pp. 1856–1864.

179. Wayman, J.L., Jain, A.K., Maltoni, D. and Maio, D. eds., 2005. Biometric systems: Technology, design and performance evaluation. Springer Science & Business Media.

180. Wiedenbeck, S., Waters, J., Birget, J.C., Brodskiy, A. and Memon, N., 2005, July. Authentication using graphical passwords: Effects of tolerance and image choice. In Proceedings of the 2005 symposium on Usable privacy and security (pp. 1-12).

181. Winder, D. (2015) Is your smartwatch tracking what you type? Available at: https://itsecuritything.com/mole-a-smartwatch-poses-no-real-world-threat/ (Accessed: 8 November 2016).

182. Witte, H., Rathgeb, C. and Busch, C., 2013, September. Context-aware mobile biometric authentication based on support vector machines. In 2013 Fourth International Conference on Emerging Security Technologies (pp. 29-32). IEEE.

183. Woodward, J. D., Orlans, N. M. and Higgins, P. T. (2003) Biometrics: Identity Assurance in the In-formation Age. New York.

184. Xiaoyuan Suo, Ying Zhu and Owen, G. S. (2005) 'Graphical Passwords: A Survey', in 21st Annual Computer Security Applications Conference (ACSAC'05). IEEE, pp. 463-472.

185. Xu, W., Shen, Y., Zhang, Y., Bergmann, N. and Hu, W., 2017, April. Gait-watch: A context-aware authentication system for smart watch based on gait recognition. In Proceedings of the Second International Conference on Internet-of-Things Design and Implementation (pp. 59-70).

186. YourSecurityResource (2013) My new Android phone has facial recognition. Is that a safe way to secure my device? Available at: http://www.yoursecurity resource.com/tech_tips/expertqa/facial_recognition/index.html#axzz47ytLFkCI (Accessed: 7 May 2016).

187. Zhang, Z., Hu, M. and Wang, Y. (2011) 'A Survey of Advances in Biometric Gait Recognition', in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 150–158.

188. Zhong, Y. and Deng, Y. (2015) 'CHAPTER 1 A Survey on Keystroke Dynamics Biometrics: Approaches, Advances, and Evaluations', Recent Advances In User Authentication Using Keystroke Dynamics Biometrics, 2, pp. 1-22. doi: 10.15579 /gcsr.vol2.ch1.

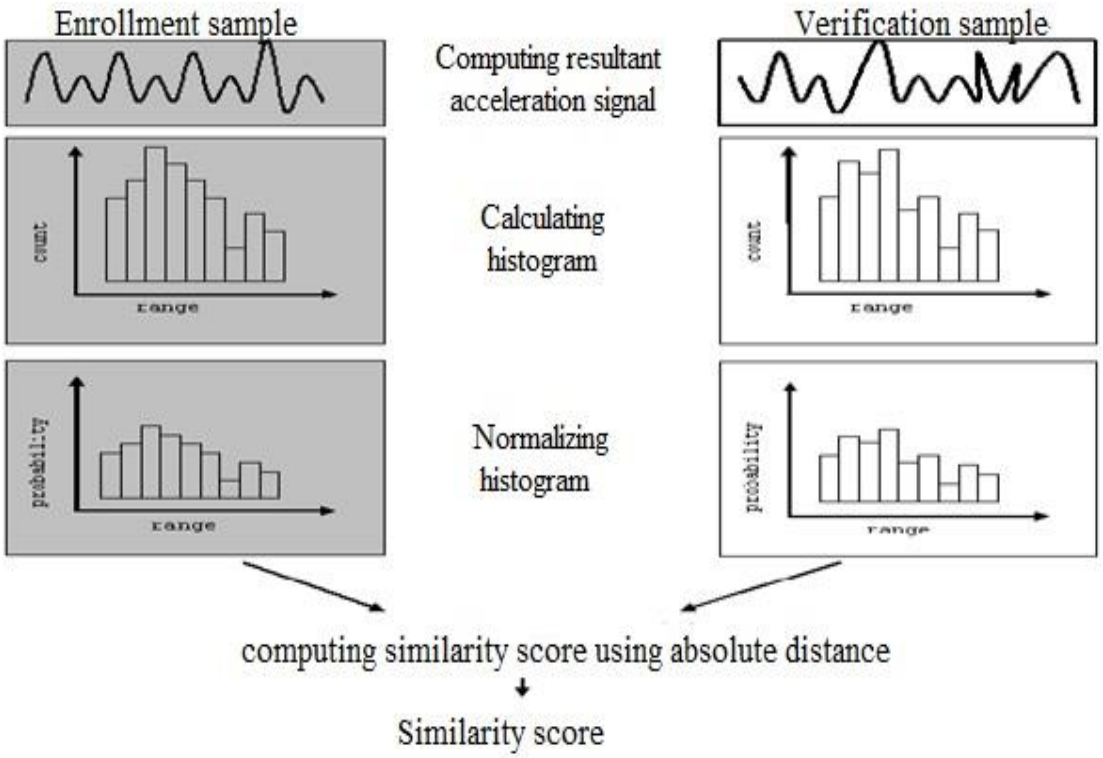
Appendix A: Details analysis of the prior art

A comprehensive analysis was conducted for each individual study in the prior art, details of each study (including, technology used, data collection methodology, feature types and feature selection approaches, classification and decision making) is described below:

Mantyjarvi et al., (2005) placed a recording device on the user's belt; data was collected from 36 participants, who each provided data on two different days within controlled conditions (i.e., laboratory dataset). In each session, the user was asked to walk around 20 meters using the normal, fast and slow paces (the first session was used for training and the second was used for testing). To construct the feature vector, local minima and maxima of each step were detected in the first method while 40 Fast Fourier Transform (FFT) coefficients were computed in the second approach. For the last two methods, data were segmented into histograms, and in case of higher order moments, skewness and kurtosis were extracted to form the reference template. Four different classification methods were utilized: signal correlation, frequency domain, histogram, and higher order moments. The reported EERs were 7%, 10%, 18% and 19% for the aforementioned algorithms respectively. However, the amount of the collected data from each user was small (in total about 30 seconds) for each speed.

Gafurov et al., (2006a) conducted a study by attaching the sensor to the lower leg of 21 participants who walked 1 minute using their normal speed within a constrained environment (i.e., experiment focused on controlled data). Half of the collected samples was used for the training, and the remaining samples was utilized to test the system. The feature extraction method involved the use of histogram similarity and cycle length. The former calculated 10-bin histogram and

applied Absolute distance metric for classifying the user's pattern. This distance was considered as a similarity score (Figure 45 explains the steps used to calculate the reference and probe histograms). The latter (cycle length) was based upon the number of observations inside the cycles to form the feature vector. The findings were EERs of 5% and 9% for the histogram similarity and cycle length respectively. Nevertheless, the obtained results were based upon only one attempt to calculate the FRR, and 20 trails to measure the FAR of each user.



Source (Gafurov et al. 2006a)

Figure 45: Applying the histogram similarity method on the acceleration signal

Gafurov et al., (2006b) carried out an experiment by attaching a motion device to the user's hip. Due to the fact that sideway direction has less movement at this position, the tri-axis signal was combined into a single dimension. The data collection involved the participant of 22 users, each walked approximately 2 minutes within a predefined hall (i.e., controlled dataset) using their normal pace.

Once all cycles were detected, the average cycle was calculated to form the user's templates. Using Euclidean distance, the obtained EER was 16%. Nevertheless, the author did not explain their cycle extraction process hence, it is challenging to reason about the causation of the high EER.

Research by Okumura et al. (2006) studied the ability to discriminate between individuals based upon their arm movement. The cycle-based approach was used in order to divide the raw time series acceleration data of 22 users (signals were collected under a constrained environment). The Dynamic Programming Matching (DPM) algorithm was utilized to identify the user's identity and an EER of 5% was reported. Apart from the limited amount of the collected sample (i.e., 5 samples from each participant), data was captured on the same day which does not show the variability of the human gait behaviour over the time. Moreover, the proposed system does not provide continues and transparent authentication as a user needs to shake the smartphone to gain access.

Gafurov, et al., (2007a) proposed to place a motion-recording device in the user's pocket, which is more realistic in terms of the sensor location for a system that is to be implemented. For the experiment, 50 subjects were involved, and four different methods were used to classify the labelled gait samples (i.e., absolute distance, correlation, histogram, and higher order moments). Cycles were detected by identifying a sequence of local minima in the acceleration signal. The initial minima was found at the following equation:

$$M_{i1} = \min (A_w - d_1, ..., A_w + d_2)$$

Where (d_1 = 50, d_2 = 150, i =100 acceleration values while A_w was the first acceleration value greater than 1.3m/s). The first minima was considered the start

point of the first cycle, and the second local minima was selected as the terminus of the cycle. To compute the second minima, the following equation was used:

$$M_{i2=} \min (M_{i1}+D-d,..., M_{i1}+D+d), \text{ where } D=100 \text{ and } d=20.$$

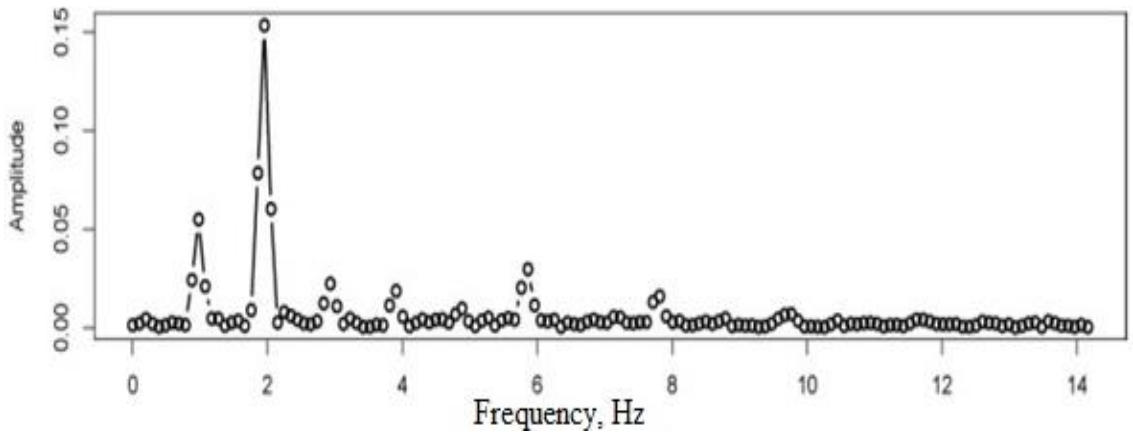
This procedure was repeated until all remaining minima were found in the signal. The end point of one cycle was considered as a start point for the next cycle and so on. To construct the feature vector, the averaged cycle for the first two methods (absolute distance and correlation) was computed. For the last two methods, 10-bin histogram of gait cycles was measured. In case of higher order moments, two additional features (i.e., skewness and kurtosis) were calculated. The result showed an EER of 7.3% using absolute distance while the EERs increased to 9.3%, 14% and 20% when correlation, histogram and higher order moments methods were utilized respectively. A fundamental problem with this approach lies in the cycle detection algorithm. The process of determining where a signal ends will most likely fail for unusual (i.e., both slow and fast) paces. Furthermore, only 24 cycles were used for training and testing purposes, which is limited amount of data.

Gafurov et al., (2007b) involved 100 participants in their controlled experiment with each participant providing one minute of data. Instead of considering the collected acceleration signals of three axes (x, y, and z) separately, the signal was combined into a single axis, denoted as R. Cycles were detected by identifying the all local minima (R_m) in the combined gait signal. The first local minima was found from the first 250 acceleration values (e.g., R_{m1}= min (R_1, R_2, R_3,…, R_{250}) and considered as the start point of the first cycle. The remaining minima were calculated as follows:

$$R_{m2=} \min (R_{m1}+100-20,..., R_{m1}+100+20)$$

After all minima were identified, data points between two consecutive minima were considered as one cycle. To generate the reference and probe templates, the detected cycles were normalized in length, and the median average of normalized cycles was computed. Using Euclidean distance, an EER of 13% was achieved. Nevertheless, data was captured on the same day and within a controlled environment (i.e., walking on flat floor only). Notably, the decision to use fixed window sizes and Euclidean distance were not justified within the work; therefore, it is unknown whether there exists alternative parameters that would improve the accuracy of the system.

A further study by (Gafurov and Snekkkenes, 2008a) had examined the potential of natural arm movement to support a gait recognition system by involving 30 users for the data acquisition. During data collection scenario, a dedicated sensor was attached to the user's wrist. Users were asked to walk at their natural speed in four different sessions on the same day (in total 40 seconds of data was obtained from each participant). Frequency domain was used to analysis the signal rather than time domain. Subsequently, the amplitudes, which are the maximum value in the signal, in a specified frequency range were detected. Varying quantities of amplitudes (2, 4, and 6) were evaluated to build the optimal reference template for each individual (see Figure 46). Employing Euclidean distance, the experimental results showed that using 6 amplitudes yielded better performance with an EER of 10 % compared to EERs 13% and 16% when 4 and 2 amplitudes were utilized respectively. However, the amount of test data to evaluate the system efficiency was limited, where the calculation of FRR was based on only two comparisons.

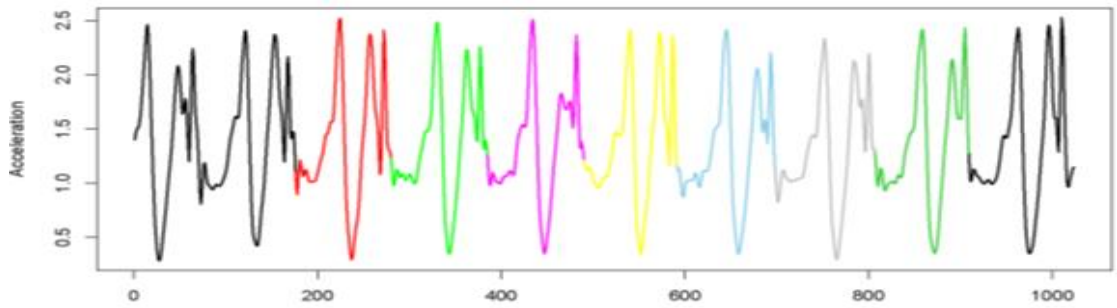


Source (Gafurov and Snekkkenes 2008a)

Figure 46: The amplitudes in the frequency domain signal

An expanded subsequent implementation of gait authentication was investigated by Gafurov and Snekkenes (2008). A motion recording device was attached to the participant's ankle. Data was collected from 30 participants under a constrained environment, with each of them asked to walk 4 sessions on the same day using their natural walking style (each session contained about 15 seconds of motion data). To extract the gait features, cycles were detected by identifying the user's standing phase within the gait signal. This standing phase was detected by filtering out accelerations that were above or below chosen thresholds. This procedure was repeated until all standing phases were detected within the dataset. The distance between two successive phases was marked as one cycle. After all cycles were detected, the median values of the extracted gait cycles were then used to compute the average cycle. Using Euclidean distance, an EER of 5.6% was obtained by employing the sideways-direction data only. However, the reported performance was achieved by requiring all participants to wear the same shoes as each other, which is not a realistic expectation for a practical system. Moreover, Figure 47 shows a signal after cycle detection with this algorithm has been performed; each cycle is a different colour with the black portion being discarded, as it is an irregular length. Different cycle lengths are common as users change their speeds, so it is apparent from the Figure that this

algorithm (and thus this system) would perform poorly in real-world conditions where users do not always maintain a constant speed.



Source (Gafurov and Snekkenes 2008b)

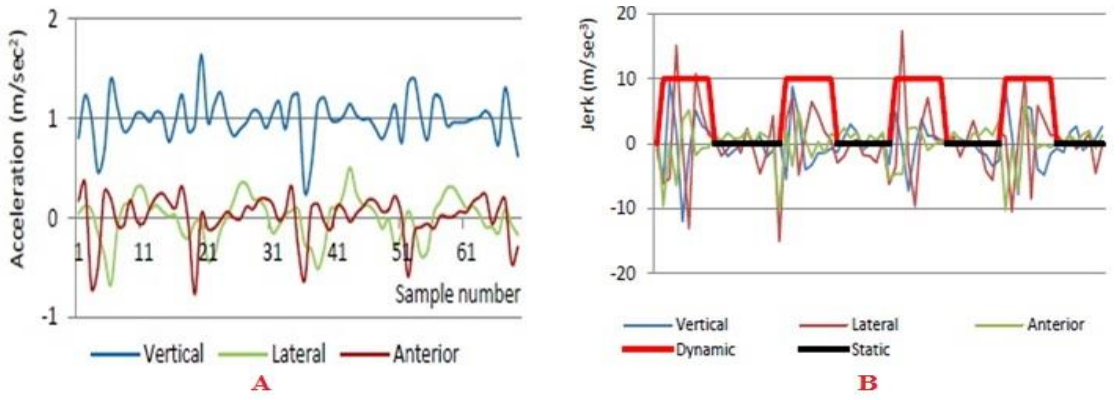
Figure 47: An example of detected cycles (in colour) from the signal.

Gafurov et al., (2010) employed the same dataset and cycle extraction method that was used by Gafurov and Snekkenes (2008). However, their resulting EER improved from 5.6% to 1.6% by using cycle matching instead of computing an average cycle. Euclidean distance was used to calculate the similarity score between two sets of cycles. Subsequently, from multiple comparisons between cycle pairs, the lowest similarity score between two cycles was considered as the best matching. Nevertheless, reducing the threshold used to mitigate false negatives would also influence the accuracy of the system. Moreover, sensor placement on the lower leg is unrealistic for the real life-based authentication applications.

Sangil Choi *et al.*, (2014) claimed that dividing the extracted cycle into dynamic and static parts can improve the performance of gait recognition. They alleged that the dynamic parts contained more distinctive features due to the significant changes in acceleration values. To separate the dynamic and static parts, firstly the rate of change between two successive acceleration values (Jerk) was calculated in the given equation:

$$\text{Jerk} = \frac{ACC_{t_1} - ACC_{t_2}}{Sampling\ rate\ (30\ ms)}$$

The dynamic part of the gait cycle was characterised with high jerk values. Secondly, a threshold value was computed based upon 100 acceleration values, in the x, y, and z directions. This threshold was used to identify the start and end point of the dynamic section. Figure 48 (A, B) shows the signal before and after separating the dynamic and static parts. Once the two parts have been distinguished, the standard deviation (Std) of each axis (x, y, and z) was calculated and used to construct the feature vector. Two experiments were implemented; the first experiment (called similarity) was to investigate whether two samples from the genuine user were similar to each other. This was achieved by calculating Std of two random gait cycles and comparing them to each other. The second experiment (called individuality) was to determine if the calculated features of each user were distinctive enough to differentiate them from other users. Euclidian distance was applied to calculate the distance between reference and probe templates for each user. Subsequently, the k-Nearest-Neighbours (k-NN) was applied and reported 100% correct classification rate. However, the dataset is considered limited with 10 users only, which was collected in a controlled environment. Moreover, having only 30 seconds of normal walking data, per user, is limited when making an allegation of robustness.



Source (Sangil Choi et al., 2014)

Figure 48: (A) original signal, (B) signal after isolating dynamic and static parts

Recently, Cola *et al.*, (2016) conducted a study with a set of 15 users, each provided fast and normal walking samples (that were collected within a controlled environment) by attaching Shimmer sensor into their trouser pocket and wrist (in such way the sensor looks similar to a watch). Cycle based approach was used to segment the collected data that was obtained within single day and resulted on an average of 70 samples for each user. Correlation-based feature selection method was applied in order to choose the best time domain feature subset (e.g., root mean square, average absolute variation, and median). Several supervised algorithms (i.e., k-NN, Multi-layer Perceptron Neural Network, Random Forest, Rotation Forest, and Multinomial Logistic) were utilized in order to identify the optimal classifier that can provide the best accuracy for gait-based continuous authentication system. The findings showed that there was not a major difference between the evaluated classifiers (in terms of the performance) within an average of 2.9% and 2.5% of EERs for the wrist and leg movement respectively.

Recent technological advances in communication technology and mobile computing have provided new ways to for developing biometric based user authentication systems. Aiming to study the practicality of such a system, Derawi *et al.*, (2010a) used a Google G1 phone to collect the gait signal from 51 volunteers in two sessions, each provided on different days (the gait samples were obtained within a constrained environment). The mobile was placed in a pocket attached to the users' belt, and the user was then asked to walk using their natural speed (this dataset was used in multiple studies in this chapter). In total, data collected from each user amounted to only two minutes; one third of the data was used for training and the remaining was used for testing the system. To extract the gait cycles, the average cycle length was estimated. Subsequently, the minimum peak in the gait signal was considered the start point of the first

cycle (i.e., P_{start} =P_{min}) whereas the terminus of the first cycle was calculated as follows:

$$P_{end} = P_{start} + averageLength.$$

This procedure was repeated until all cycles were detected in the data set. Before calculating the average cycle, Dynamic Time Warping (DTW) was used to omit the cycles that were significantly different than other cycles (i.e., high distance to other cycles). Using DTW, the obtained result was high with an EER of 20.1%. This high error rate could be attributed to the minimal amount of data used to train and test the system.

An alternative solution to segment the raw acceleration signal was proposed by Kwapisz et al., (2010). They raised several concerns about cycle extraction method (specifically, the unclear boundaries between the cycles and the complex computational effort required to detect those cycles). Therefore, the raw accelerometer data (which was collected within a controlled environment) was divided into segments (using sliding window approach) instead of cycle detection method. The data collection process was more thorough, where each user was asked to provide multiple activities (i.e., walk, run, climb up and down stairs) for specific time in one session only. A mobile was placed inside the front pocket of 36 users. Data of each activity was collected separately with the goal of using the dataset for identification and authentication tasks. In total, 10 minutes of data was captured per user for all activities. The raw time-series accelerometer data was then partitioned into 10 second segments. After the collected data was segmented, the statistical features of each axis were calculated. These included average (Avg), standard deviation (Std), average absolute difference (AAD), time between peaks (TBP), binned distribution (BD), and the average resultant change in the acceleration (ARCA). The user's reference template was created

regardless which activity the user was performing (in other words, the user's activity label was removed). Using only a single 10 second segment of walking data, the authors achieved 72% and 69% identification rates using decision trees (J48) and Neural network respectively. In comparison, the authentication result of five participants was 85.9% positive authentication rate at 95% negative authentication rate using J48. By using the majority voting to all test data (5 to 10 minutes), the identification and authentication accuracy were further improved to 100%. However, it has to be noted that the authentication results were based on limited number of participants (specifically five participants). Moreover, using the majority voting scheme, a scheme which accepts a user as genuine if a half or more of the user's test samples are positive, might increase the false acceptance rate. Therefore, applying this schema require a large amount of test data to claim the system is robust to impersonation attacks.

A further study has been conducted by Nickel *et al.*, (2011a). The authors employed the same database that was used by Derawi *et al.*, (2010a). The segment-based approach was used to divide the time-series accelerations data into 3 seconds segments. A total of 28 samples were obtained from each subject's data. In order to create the user's template, 20 samples from the user's data and all samples of 30 imposters were used. The remaining samples from the genuine user and all samples of 17 imposters were used in the testing phase. The proposed system achieved a FRR of 10.42% with a FAR of 6.62% using HMM and the majority voting method. To have a balance system (security and usability), only 8 samples from each imposter was utilized to calculate the FAR. The study reported 10.42% of FRR and 10.29% of FAR. Although the system was able to identify imposters with data not used in the training phase, the false

positives of each genuine user was evaluated based on limited samples of only 17 imposters.

The research study by Nickel et al., (2011b) also employed the same dataset used by Derawi *et al.*, (2010a). Authors demonstrated that increasing the segment size into 10 seconds could enhance the system accuracy. After the raw acceleration data was divided into segments, a total of 12 samples were extracted from each user. To train and validate the system, 50% of the first and second day samples were used for training, and the remaining samples were used for validation (this is known as mixed-day scenario). New features, which have been successfully implemented in speaker recognition, were extracted. Specifically, BFCC and MFCC were calculated. Figure 49 explains the steps that were used to extract the cepstral coefficient features. The extracted features were based upon the acceleration values of each axis and the resultant acceleration (magnitude) as well. Two feature subsets were calculated to build the reference template, the statistical and cepstral coefficient features. The computed statistical features were BD, maximum (Max), minimum (Min), mean, Std, root mean squared acceleration (RMS), and zero cross; on the other hand, only BFCC was measured from the cepstral coefficient features. SVM and quorum voting method (a method that accepts a user as genuine if a requisite number of the user's samples are positive) were applied to classify the user's gait pattern. The proposed system revealed 6.3% of FRR and 5.9% of FAR (compared to roughly EERs of 20% and 10% of the previous studies by (Derawi *et al.*, (2010a) and Nickel *et al.*, (2011a), respectively). However, the findings of this study were based on using a mixed-day scenario. This scenario requires a user to re-enrol in the system every day; effectively; the system is equivalent to a single-day scenario.

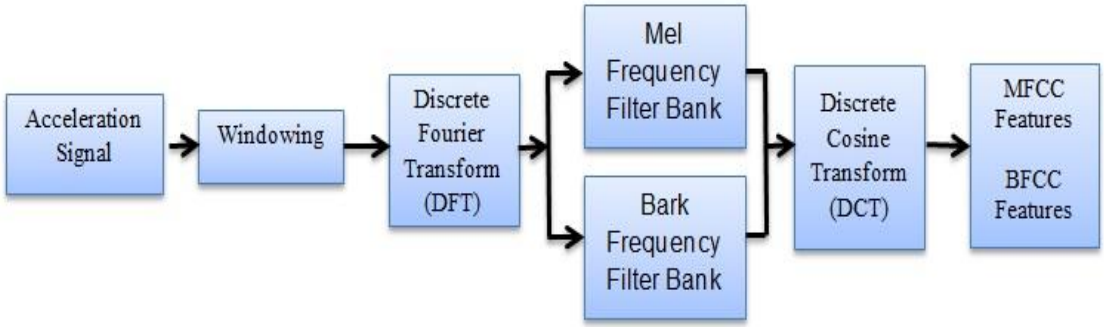
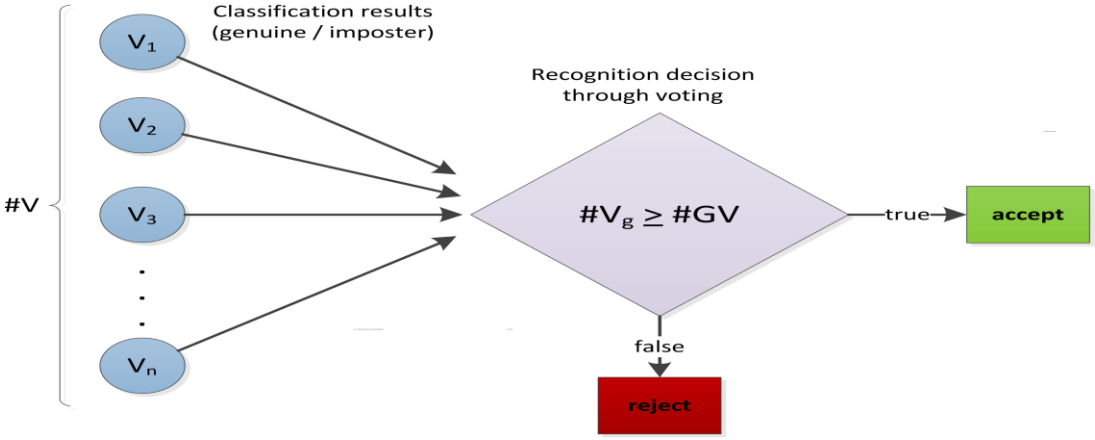


Figure 49: The process of extracting BFCC and MFCC



Source (Nickel et al. 2011b)

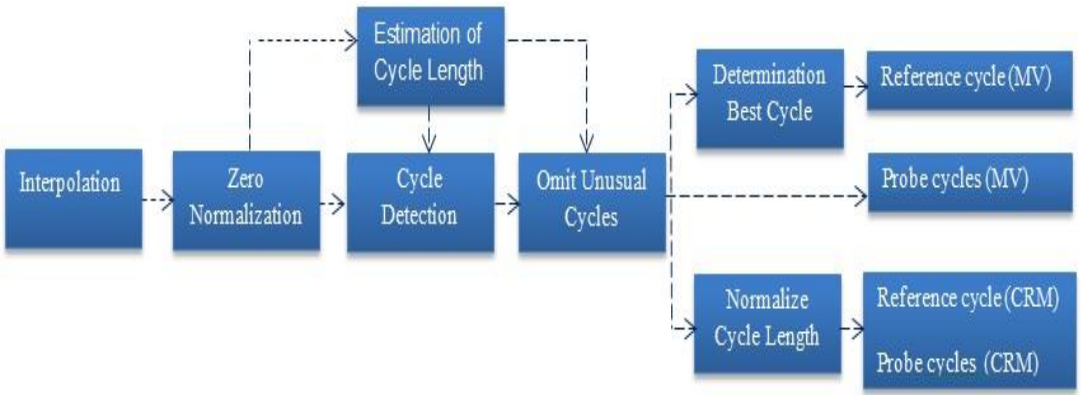
Figure 50: Quorum voting scheme (#V total test segments, #Vg, number of votes for genuine, *#GV* positive classification results)

Nickel et al., (2011c) attempted to compare the performance of two classification algorithms, HMM and SVM. The study argued that SVM slightly showed better accuracy. For the experiment, the normal walking pace of 36 subjects was recorded by placing a mobile phone inside a pouch attached to the user's belt. The subject participated in two sessions, each in a separate day. Once the raw movement data was obtained, the segment- based approach was used to chunk the time-series accelerations data into five seconds windows with an overlap of 50% (this means with five seconds, every two seconds and half, a new segment is generated). The amount of the extracted samples from each subject was fairly acceptable, around 200 samples per session (this dataset is used in multiple

researches in this literature). Statistical (Min, Max, mean, Std, Bin, RMS, and zero cross) and cepstral coefficient features (MFCC and BFCC) were generated for x, y, z axes and magnitude vector (m). Each feature was evaluated, and the user's reference template was constructed by using a feature subset that produced the lowest error rate. Once the feature selection was completed, it was found that only cepstral coefficient features were more consistent and robust enough to report the best performance. When the first day data was used to train and test the system (called the same-day scenario), the reported performance was 16.60% total error rate (TER), TER is the summation of FAR and FRR and 5.86% of EER using SVM and HMM respectively. However, the system performance significantly dropped down to 40.52% of TER and an EER of 17.06% for the aforementioned classifiers by using the first day samples for training and the second day data for testing (called cross-day scenario). Although using this scenario displayed a higher error rate, it avoids training the user's model every day, which is more realistic for real world applications. To improve the system performance, the quorum voting schema was applied to 70 samples from a user's test data (corresponding to about 3 minutes of the walking data). If 3 out of the 70 samples were correctly classified, then the user was considered to be the genuine user. EERs of 10% and 12.63% were achieved by using SVM and HMM respectively. Nevertheless, including more participants might increase the chance of accepting an imposter. Moreover, the decision was based upon continuous 3 minutes of the walking data that would require more processing time and increase the intrusiveness of implementing such system.

Research by Nickel et al., (2011d) proposed a system with real-time testing that took place on the mobile device. During the data collection phase, 48 subjects were participated and each of them provided two sessions on two different days.

In each session, the subject was asked to walk in a natural speed on flat ground for 10 seconds that were later used for training phase. The subject was then asked to walk about 15 minutes in a predefined route during the authentication phase, the route included walking on flat ground, up/down stairs, and opening/closing doors. The authentication process was activated every 30 seconds when the subject stopped at one of the predefined points. Cycles were detected using minima and maxima salience vectors. The extracted cycles were then filtered by calculating the DTW distance between all cycle pairs, and irregular cycles, which had a high distance from other cycles, were removed. The resulting dataset hence is referred to as the remaining cycles. Two different methods were used to pre-process the extracted cycles, majority voting and cyclic rotation metric (CRM), a metric that compares each probe cycle to every reference cycle of a genuine user and stores the largest distance as a similarity score. In the first method (Majority voting), the remaining cycles were further analysed to find the smallest distance between each pair of remaining cycles, this was hence used to create the reference template. In contrast, the second method (CRM) used the remaining cycles for both reference and probe cycles (Figure 51 provides the cycle extraction process used for both methods). The cross-day scenario was applied for both of the enrolment and authentication phases. Using DTW with the majority voting approach, an EER of 28% was achieved. In comparison, applying the Manhattan and DTW distance functions with the CRM method, the reported EER was 21.7%. In addition to the high error rate that were resulted from both approaches, the user needed to wait ~30 seconds to unlock their phone, which is more than the required time to enter the PIN itself. Moreover, some subjects were always rejected by the system (i.e., their FRR was 100%).



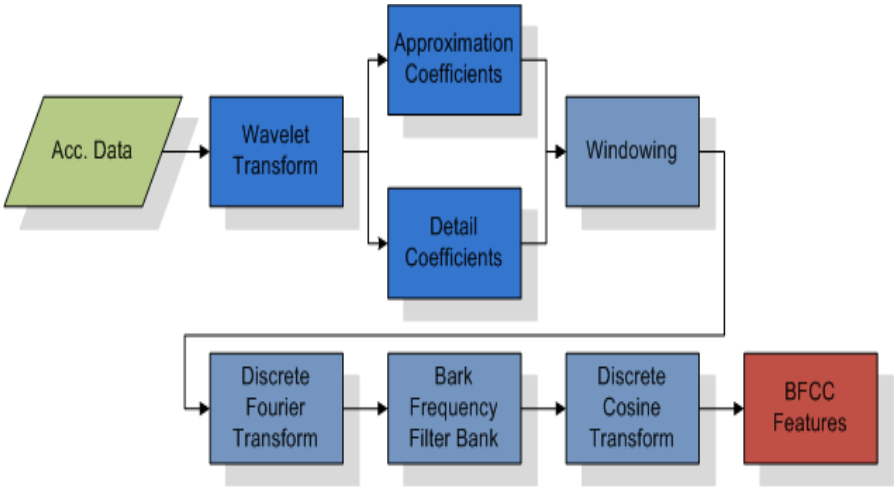
Source (Nickel et al. 2011d)

Figure 51: Cycle extraction steps during the enrolment and verification phase.

Nickel and Busch (2011) investigated the influence of the sample size on the gait-based biometric performance. For the experiment, the authors employed the same dataset of Nickel et al., (2011d). Instead of extracting cycles from the gait signal, segment- based approach was used. Various segment sizes (2, 3, and 4 seconds) were evaluated to select the optimal segment length, which produced a lower error rate. Based on the acceleration values in the segment, the cepstral coefficients feature (MFCC) for a separating axis (x, y, and z) and magnitude were calculated. The user's template was created by generating the MFCC feature from a segment size of 2 seconds as it provided better results. The cross-day scenario was applied to train and test the system; the first 10 seconds of walking data (on flat floor) was employed to train HMM and 5 minutes of mixed gait data (included walking on flat floor, up/down stairs, and opening/closing the doors) were used for testing. When the user's reference template trained with only 10 seconds (i.e., five samples), an EER of 31.6% was achieved. However, increasing the amount of training data into 114 seconds of mixed data greatly reduced the EER into 18.11%. These results were based upon a single 2 seconds instance of the acceleration data. Applying the quorum voting approach to a group of 60 samples of the user's test data (which is equivalent to 2 minutes) resulted in 6.15% of EER. In comparison with the previous work by Nickel et al., (2011d)

that reported an EER 21.7%, this study achieved an improvement of more than 70% (i.e., a low EER of 6.15%). This could be explained that segment-based approach provides a significant performance boost when compared to the cycle method. However, some users in this study were rarely recognized by the system with a high FRR of about 60% or higher.

Hestbek et al., (2012) proposed to use the Discrete Wavelet Transform (DWT) in order to convert the raw acceleration data into signal information, approximation and details coefficients. Again, this study employed the same dataset used by Nickel et al., (2011c). The sliding window approach divided the acceleration signal into 5 seconds with 50% overlap and then the BFCC and Std features were extracted. More details of the feature extraction process are shown in Figure 52. SVM and the quorum voting method were used to classify the user's gait pattern. Fifty segments from the authentication dataset of each user, which typically corresponds to 2 minutes of the walking data, were passed to the quorum voting method. The obtained results were 9.82% of FAR and 10.45% of FRR. Apart from the computational overhead of applying the DWT, the findings of this study did not improve the performance of the prior art by Nickel and Busch (2011), which achieved an EER of 6.15%.



Source (Hestbek et al. 2012)

Figure 52: The steps of extracting BFCC features

To investigate if the cycle extraction method can enhance the gait recognition performance, Nickel and Busch (2012) suggested a new solution. They proposed that each segment contains a number of cycles (called cycle-based segment) instead of windowing the raw data into fixed segment size. The same dataset used in Nickel et al., (2011c) was employed in this study, and the cross-day scenario was used to train and test the system. Cycles were detected in the gait signal based upon the presented cycle extraction method in Nickel et al., (2011d). Around 264 cycles were extracted, half of these cycles were used for training (which corresponds to about 4 minutes of the walking data). From the acceleration values of each cycle, the combination of cepstral coefficient features (BFCC and MFCC) was generated for each axis (x, y, z, and m). A comprehensive evaluation has been conducted to construct a robust reference template. This has been done by testing the impact of the sampling rates, and the number of cycles per segment. Including one cycle per segment and 100 samples per second, an EER of 22.7% was achieved. However, using 50 and 200 sampling rates decreased the performance to 27% and 28.8% respectively. Further experiments were conducted by including four cycles in each segment (corresponding to four seconds); the EER dropped significantly from 22.7% to 17.96%. The aforementioned recognition rates were based upon utilizing only 4 seconds of the accelerometer data (single sample without voting). Considering a group of 14 samples of the user's data (corresponding to about 2 minutes of the walking data) and using the quorum voting method resulted in an EER of 15.46%. To benchmark between cycle- based segment method and segment-based approach, the authors carried an additional experiment. The same features (BFCC and MFCC) were extracted but from a fixed segment size of 5 seconds (without prior identifying the contained gait cycles). The results revealed EERs of 13.89% and 17.28% with and without voting. In addition to the complexity and

computational overhead of identifying the cycles in the gait signal, it can be noted that the proposed method (i.e., cycle-based segment approach) did not improve the system performance before and after voting.

Nickel et al. (2012b) investigated the efficiency of applying three machine learning algorithms (i.e., k-NN, HMM, and SVM) on the gait recognition rates. The normal walking pace data was collected from 36 users by attaching a mobile phone at their hip pouch. Each user took a part twice on two different days; each session contained approximately 10 minutes of the controlled gait data. The segment-based approach was used to divide the raw acceleration data into 7.5 seconds with a 50% segment overlap. On average, about 132 samples were extracted per session from each user. The first day data was used to create the reference template and the remaining samples were used for testing. Once the reference and test templates were generated, the cepstral coefficients and statistical features were extracted. The selected feature subset was based on two main criteria: the performance of individual feature and the combination with other features and the feature's discriminative potential score (the features that had low intra-class variability and high inter-class variability were selected). The evaluation outcomes of the cepstral coefficient and statistical features showed that BFCC was sufficient to create the user's reference template. Euclidean distance was used to compute the distance between reference and test templates and k-NN was applied to select the closest match of the calculated distances. The proposed system achieved unbalanced performance (22.22% of FRR and 3.97% of FAR) which is equal to 13.09% of HTER $\left(\mathrm{HTER} = \frac{\mathrm{FAR+FRR}}{2}\right)$. However, these findings were based upon using a single sample, which was constructed from only 7.5 seconds of the motion data. Therefore, to reduce the FRR (and make the system more tolerance to accept a genuine user), a group of 25 segments of the

user's samples were passed to the quorum voting schema. If two segments were correctly classified, the user was considered as genuine. The reported result was 8.24% HTER. Although HTER gives insight into the average error rate, the metric fails to communicate whether the system strikes a balance between usability and security (FRR and FAR, respectively). To find out the more suitable algorithm to classify the user's gait pattern, HMM and SVM were also evaluated; five minutes of the walking data were used to train each of these classifiers and the verification data were between 1.7 and 3.2 minutes. Table 22 displays the performance of each algorithm after applying the quorum voting method. Although there was no noticeable change in terms of the error rates between the three algorithms, the SVM is more sensitive with the variation of the human gait. Hence, it produced a high FRR (Nickel et al., 2011c). In comparison, k-NN and HMM were less sensitive and performed slightly better. Nevertheless, while the authors claimed the efficiency of the system using a real-world implementation, they fail to state how accurate the system was in practice.

Classification	Verification data	Performance
K-NN	1.7	HTER= 8.24
HMM	2.5	EER= 8.75
SVM	2.5	EER= 8.85

Table 22: The performance of three different classifiers

Muaaz and Nickel (2012) studied the effect of different walking speeds and surfaces on the gait recognition performance. Controlled data was collected from 48 subjects and a Google G1 smartphone was placed inside a pouch attached to the user's hip. Subsequently, the subject was asked to walk at their normal, fast, and slow paces on flat, grass, gravel, and sloping ground. Each subject participated in two sessions on two different days. Every session consisted of six different walks trails from each user. In the first four trials, the subject walked using their natural speed on flat, grass, gravel, and sloping ground. The last two

trails included the fast and slow gait speeds on extrovert ground (each trail contained about 1 minute of walking data). After the raw acceleration data were obtained, cycle extraction method was used to create the reference and probe templates. First, the cycle length was estimated by calculating the minimum and maximum salience vectors then the same method was used to detect the cycles. To remove the irregular cycles from the dataset, DTW was applied to calculate the distances between all cycle pairs. Thereafter, a threshold value has been set to decide which cycles must be deleted. At least six cycles were obtained from each trial (a "remaining cycle"). The remaining cycles were further analysed to select the optimal cycle (a "typical cycle"), which has the minimal distance to other cycles. Two different experiments were conducted; in the first experiment, the typical cycle was used for training and the remained cycles were used for testing. The second experiment employed the remaining cycles for training and testing purposes. In both experiments, the reported EERs were very high, with the first method performing slightly better. Applying DTW and the majority voting method to the normal gait data on flat, grass, gravel, and sloping ground, the best EERs were 29.39%, 32.05%, 36.10%, and 35.18% respectively. For the fast and slow gait signals, EERs was 33.81%, 35.31% were achieved consecutively. The findings of this study highlighted how different walking speeds and surfaces could influence the gait recognition. Therefore, it is important to train multiple reference templates, each contains data of specific walking speed. Thereafter, during verification phase, an activity recognition should be applied to distinguish the speed of probe vector and select the correct authentication template.

Ho *et al.*, (2012) proposed to classify the user's gait pattern by sending the collected acceleration data to an application server for processing. The normal walking pattern of 32 subjects was monitored (which was obtained under a

constrained environment). Only two minutes of the motion data were collected from each user in one session and the tri-axis signals were combined. The periodic motion of one step was considered as one cycle. To detect the cycles in the fused signal, autocorrelation was used to estimate the cycle length and the start point of the first cycle was detected manually. A cycle is defined by two consecutive “zero crossings” or the points where the value of the signal changes sign. In total, 400 cycles were extracted from each subject; the statistical features (mean, variance, Std, Min, Max, and RMS) for each cycle were computed. To train and test the system, the authors divided the dataset into two parts, 70% of the user’s data was used for training and the remaining data was employed for testing. Using the SVM algorithm reported 100% correct classification; this performance was significantly dropped to 69.67% when the ratio was changed to 50% as authorized user and 50% as unknown user. In addition to the higher error rate, the manual cycle detection is not practical for two reasons: first) it does not scale (i.e., it will not work if there are a large number of users); second) it is inaccurate (humans can introduce errors). Moreover, the experiment required a device with network connectivity which increases implementation cost.

A study by Shrestha et al., (2013) proposed a gesture-based authentication system by utilizing the accelerometer and ambient light smartphone sensors. The experimental study included the participants of 20 users, each was asked to perform hand waving gesture 10 times on a single day. In order to classify the user’s identity, the authors developed a wave detection algorithm that are demonstrated in Figure 53. The FAR and FRR were used to evaluate the system accuracy with a FRR of 10% and a FAR of 1% being reported. A major defect of this system is the possibility of high false rejection rate when a user waves the

hand far away from the smartphone (i.e., distance between the hand gesture and a smartphone should be close).

Algorithm 1 Wave Detection using Light Sensor (and Accelerometer)

1: IF sensors are locked THEN wait for $MOVEMENT\_LOCK\_TIME$
ELSE get accelerometer sensor readings x, y and z.
2: IF
$$\sqrt{x^2 * y^2 * z^2} > ACC\_THRESHOLD$$
THEN lock the sensors for $MOVEMENT\_LOCK\_TIME$ and RETURN to step 1.
3: IF sensors are not locked THEN get light sensors reading to check if wave gesture is detected.

1. Analyze $WINDOW\_SIZE\_FOR\_LIGHT$ data to find out how many extremas (maximas and minimas) were there using $LIGHT\_THRESHOLD$.
2. IF $extremaCount > CHANGE\_COUNT\_FOR\_LIGHT$ AND All the light data are recorded within $WAVE\_TIME\_LIMIT\_FOR\_LIGHT$ THEN
SET $unlockAttempted = true$,
RECORD first unlock attempted time
DISPLAY Message "Stop Waving" for $WAVE\_TIME\_LIMIT\_FOR\_LIGHT$.
3. IF $unlockAttempted$ THEN

   (a) IF another unlockAttempt is obtained within less than $WAVE\_TIME\_LIMIT\_FOR\_LIGHT$ THEN Do not unlock, reset everything and start over, i.e., return to Step 2.
   (b) IF another $unlockattempt$ is not obtained within $WAVE\_TIME\_LIMIT\_FOR\_LIGHT$ THEN $Unlock$ the phone for $UNLOCK\_TIME\_FRAME$.

Source (Shrestha, et al., 2013)

Figure 53: The proposed wave recognition algorithm

Another study to classify the normal walk style was by Muaaz and Mayrhofer (2013). The acceleration data was collected from 51 participants, each was asked to place the mobile inside a pocket attached to their hip and walked about 30 seconds down the hall in one session. In total, two sessions were captured per participant on two different days. Cycles were extracted from the gait signal based upon the presented approach by Muaaz and Nickel (2012). Two different experiments have been carried out using two different classification methods, DTW and SVM. The first experiment applied Piecewise Linear Approximation (PLA) to the acceleration data before estimating the cycle length. After the cycles were detected, comparison between the reference and probe cycles was carried out by using DTW as distance function. Figure 54 illustrates the steps used to compare two gait cycles. When the same day scenario and cross day scenario were used to train and test the system, the authors were able to achieve EERs of 22.49% and 33.3% respectively. To train the SVM in the second experiment,

DTW was used to calculate the distance between the reference cycles and probe cycles, and the output distances from the DTW function was used as input to SVM. The achieved result was a FRR of 35.7% against a FAR of 1.1%. Obviously, both experiments reported high error rate.

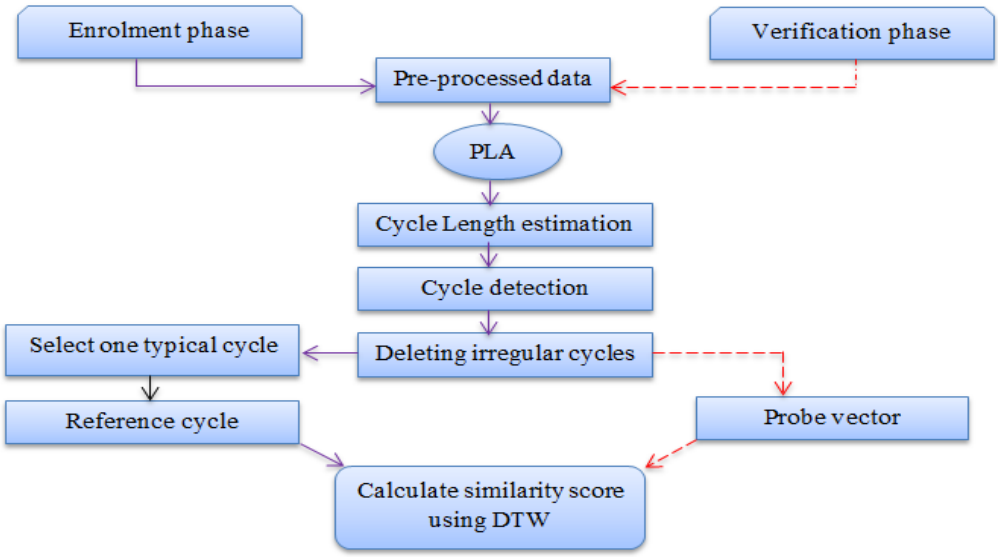


Figure 54: The process of applying DTW

Ross (2013) attempted to identify the user's identity based upon collecting several activities (i.e., walking, running, typing, sitting, standing, and walking up/down stairs). The acceleration and gyroscope motion data were collected from 9 users on the same day and under a controlled environment. The collected data from each user was at least 5 minutes for each activity except the stairs was either one or five minutes. The segment-based approach was used to divide the raw data of both sensors into 10 seconds windows with an overlap of 50% (An average of 260 samples was obtained per user for all activities). A number of statistical features (i.e., Avg, Std, AAD, ARCA, ARCV, TBP, and BD) were generated for the accelerometer and gyroscope data separately, and the findings are presented in Table 23. The results show that gyroscope sensor was more effective than accelerometer. Using J48, the reported results were 86.4%, 95.4%, and 91.5% of correct classification rate for walking, running, and typing activities respectively.

Classification Method	Accelerometer features	Gyroscope features	Features of both sensors
J48	85.3	89.4	90.3
Multilayer neural network	62.4	64.9	70.5

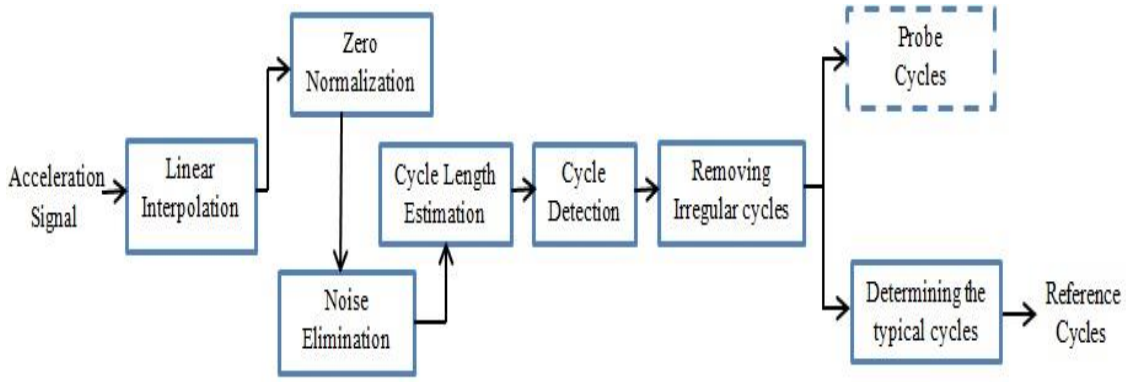
Table 23: Correct classification rate (%) of the proposed system

The previous studies have primarily concentrated on collecting the acceleration data from a single mobile within a specific sampling rate. Therefore, Hoang *et al.*, (2013) conducted a study to identify the user's gait pattern regardless of the sampling rate and utilizing various phones. Two different smartphones (i.e., HTC Nexus and LG Optimus G) were fastened together and placed inside the user's pocket. Two different sampling rates were obtained from both phones (27Hz for HTC Nexus and 100Hz for LG Optimus G). Data was collected from 14 participants, each walked 12 rounds in one day (each round contained about 36 seconds of the controlled gait data). Gait cycles were extracted by identifying the minimum peaks, and the distance between two consecutive peaks was considered as one cycle. Since all cycles were extracted, the gait signal was divided into segments where each segment contained 4 sequential cycles with a 50% overlap windows Based upon the speed of a participant's walk, the extracted segments per subject were between 110 and 167 (half of these segments used for training and the other half for verification). Subsequently, the time and frequency domain features were generated for separating axis (x, y, and z) and magnitude as well. The extracted time domain features were average maximum, average minimum, AAD, RMS, Std, 10-bin histogram, and waveform length. In the case of the frequency domain features, the first 40 Fast Fourier Transform coefficients, and the first 40 DCT coefficients were measured. By applying SVM, the resulted correct classification rates of using HTC Nexus phone and LG Optimus G gait signals were 99.81% and 97.53% respectively. Another experiment has been conducted to create the cross-device gait recognition model

(this model used data from one phone for training and testing the system by employing data from the second smartphone). To build this model, the aforementioned features were further analysed in order to select the features that were more resistant to changes in sampling rate. This was done by calculating the average error rate (AER) and the intra-class correlation coefficients (ICC) for each feature and then selecting the feature subset that showed higher ICC and lower AER. Based upon the conducted analysis, only the time domain features were used to create the feature vector. The result revealed a correct classification rate of about 91%. However, the number of participants was considered limited with 14 users only and data was obtained on the same day.

Usually, gait data contains some noise and errors hence, it requires pre-processing before extracting the features. These errors typically result from the phone movement in the subject's pocket, whilst the noise is produced by impact forces and oscillations caused by the subject walking. From this prospective, Muaaz and Mayrhofer (2014) suggested a solution to handle these issues. The magnitude vector of the three axes was calculated to minimize the errors, and the multi-level daubechies orthogonal wavelet was used to reduce the noise (Percival and Walden, 2000). The normal walking signal (controlled data) was gathered from 35 volunteers, each was asked to place the mobile in their trouser pocket. Each volunteer participated one session per day for two days (the first and second day data were used to train and test the system respectively). Once the acceleration readings were obtained, cycles were detected by identifying the local minima in the gait signal. After all the local minima had been identified, the distance between two consecutive minima was considered to be a one cycle.; at least 12 cycles were detected from each user per session. Figure 55 illustrates the steps used to create the reference and test cycles. An EER of around 19%

was achieved using DTW. However, during the data collection stage the authors asked the subjects to wear a pair of trousers that have a tight front pocket to limit the mobile movement. Therefore, the system performance might decrease if the constraints imposed upon users were more realistic.



Source (Muaaz and Mayrhofer, 2014)

Figure 55: The steps used to create the reference and test templates

In recent years, there have been relatively few studies on activity recognition that are based upon mobile accelerometers. Watanabe, Y. (2014) conducted a study to authenticate the phone's user based upon three different activities (i.e., walking, touching the mobile's screen, and making a phone call). Data was collected from 4 participants only (within a constrained environment), each walked three laps using their normal pace (in total, the participant walked approximately 2 minutes on same day). During each lap, the mobile's position was different. In the first lap, the mobile was placed in the participant's trouser pocket, whereas the second and third laps the participant pretended to pick up a phone call and touch on the mobile's screen respectively. After the accelerometer data was obtained, it was segmented into 3 seconds intervals. Several statistical features (average, Std, AAD, BD, and TBP) for each axis were computed. A ten-fold cross validation technique was used to train and test the system (ten- fold cross means that 90% of the dataset used for training and 10% for testing and the procedure repeats 10 times). Five different classification methods were used,

Feedforward Multi-Layer Perceptron (FF MLP), J48, Radial Basis Function (RBF), Bayesian Network (BN), and Random Forest (RF). Table 24 displays the classification performances for each activity. The results showed that the mobile location influences the system performance significantly. When the mobile was placed in the pocket, only FFMLP showed balance result between the FAR and FRR rates. However, apart from the relatively small number of participants, the samples collected from each user was limited and gathered at same day.

Algorithm	**In trouser pocket**		**Touch on screen**		**Hold calling**	
	FAR	FRR	FAR	FRR	FAR	FRR
FFMLP	1.30	2.34	3.65	7.81	9.38	22.66
J48	3.39	15.63	7.03	22.66	6.51	29.69
RBF	0.52	8.59	4.17	13.28	2.86	22.66
BN	0.26	7.81	8.85	14.06	5.99	21.09
RF	0.26	7.81	1.82	17.19	2.86	32.03

Table 24: The reported FAR (%) and FRR (%) of each activity

A further study was conducted by Watanabe, Y. (2015), where the data collection methodology was similar to his previous study; however, the number of participants was increased to 8 users, and data was collected from each user on two different days. The obtained accelerometer data was segmented into three-second intervals. The user's feature vector was created by using the same feature subset generated by Watanabe, Y. (2014). Two experiments have been implemented: in the first experiment, the same day scenario was used to train and test the system, whilst the second experiment used the cross-day scenario for training and authentication phases. Four different classification methods were utilized for both experiments, BN, RF, RBF, and FFMLP. Table 25 displays the achieved results of both experiments.

Mobile location	Used data	BN	RBF	FFMLP	RF
In trouser pocket	S	98.96	98.96	97.92	96.88
In trouser pocket	C	97.22	97.92	97.92	93.75
Hold calling	S	90.63	90.63	90.63	91.67
Hold calling	C	86.11	86.11	88.89	81.25
Touch on screen	S	87.50	86.46	87.50	83.33
Touch on screen	C	88.89	91.67	86.81	87.50

Table 25: Correct classification rate metrics (%) for each activity, S and C denotes to use the same and cross day data, respectively

While gait recognition offers competitive authentication accuracies, the conducted study by Gascon et al. (2014) attempted to explore the possibility of protecting the sensitive smartphone information based upon the motion data of the typing activity. The data collection was performed in a single day and involved a considerable number of volunteers (i.e., 315 users). The users were asked to type a predefined text message on the touchscreen of their smartphones (controlled experiment). Thereafter, several time domain features were extracted to create the user's reference template such as RMS, mean, and Std. At best, the experimental analysis showed a true positive rate of 92%, which is the rate of classifying the authorized user correctly and 1% of FAR by utilizing SVM. Although the motion signal was captured from a large dataset, at least in the term of behavioural-based biometric, it was collected in single session (which is not a realistic scenario as a more diversity typing profile of the users could be captured during multiple sessions). Moreover, the true positive rate was calculated for limited users (i.e., only 12 genuine users) and data of 302 users was utilized to measure the FAR hence, the recognition rate of identifying the legitimate user might decrease by using the whole dataset.

Another study by Damaševičius et al. (2016) utilized the sliding window approach to divide the walking signals of 14 users (the gait samples were obtained within a constrained environment). Each participant was asked to perform several

activities namely, walking, upstairs, downstairs, running, jumping, sitting, standing, and elevator up and down. Five trails were recorded per activity in a single day. The majority of the extracted features were derived from the prior art (e.g., mean, covariance, and difference). The fusion of the acceleration and gyroscope data was used and applied Random Projections (RP) method (more details about RP method (Achlioptas, 2003) in order to reduce potentially large dimensionality of input data. As a result, the system performance could be enhanced by selecting the most optimal unique features for individual. Once the reference and test templates were created, the Jaccard distance was used for activity classification and reported an EER of 5.7%. Nevertheless, similar to other prior art that collected the movement data on the same day, this study requires a user to re-enrol in the proposed system every day.

A comprehensive gait analysis was conducted by Ehatisham-ul-Haq *et al.*, (2017a) to explore the impact of different smartphone positions on the system performance. This was carried out by placing Samsung Galaxy into five different body positions (i.e., pockets on both sides, right and left wrist, and right upper arm). The acceleration and gyroscope signals of various activities (i.e., walking, upstairs, downstairs, sitting, standing, and running) were collected from 10 participants; each was asked to provide three minutes of the motion data for each activity that was captured in a constrained environment. The segment-based method was utilized to divide the raw data into 5 second segment, which is resulted in a total of 36 samples for each activity. Subsequently, features were extracted in both, the time and frequency domains (e.g., max amplitude, min amplitude, energy, and entropy). SVM, Bayesian network, Decision trees and k-NN were utilized to classify the user's identity, and correct classification rates of

99%, 97.4%, 96.8%, and 93.3% were achieved for the aforementioned algorithms respectively.

So far, all the presented studies were based upon utilizing labelled data (i.e., data was collected within a controlled environment). Therefore, Kumar *et al.*, (2017) proposed a more realistic data collection scenario by capturing unlabelled smartphone movement signals and developed a context authentication model based upon the phone usage. Data was collected from 57 participants over multiple days; the sliding window approach was utilized in order to segment the raw time series data into 10 seconds windows with an overlap of 50%. Time and frequency domain features were extracted in order to form the user's reference template such as mean, Std, median frequency, and spectral entropy. To build a context authentication model for each individual, k-means clustering and Random Forest were used to create several distinctive smartphone usage patterns. Subsequently, multiple authentication models were generated, each trained and evaluated based upon the predicted smartphone usage. Four different classification methods were applied: (i.e., Logistic Regression, FFMLP, K-NN, SVM, and Random Forest) that reported EERs of 13.7%, 13,5%, 12.1%, 10.7%, and 5.6% respectively. However, the proposed system fails to create a distinctive pattern for some users through reporting a high EER of about 40% or higher.

Ehatisham-Ul-Haq *et al.*, (2017b) conducted an activity recognition study by collecting the acceleration data within a single day from 10 users; each was asked to perform three minutes data for each of the predefined activities (i.e., walking, sitting, standing, walking upstairs and downstairs). The raw movement data was portioned into 5 seconds samples with a 50% overlap through using the segment-based approach. Subsequently, a set of frequency and time domain features was used to form the user's reference template such as energy, entropy, mean, and

variance. Thirty presents of the collected data was used to train the machine learning algorithms (i.e., K-NN, BN, and SVM), and the system performance evaluated by utilizing the remaining data (i.e., 70%). The average correct classification rates reported by K-NN, BN, and SVM classifiers were 89.65%, 94.57%, and 94.24% respectively. However, a non-realistic scenario was used to test the efficiency of the proposed system (i.e., the enrolment and authentication phases were based upon data collected in a single session only).

Lee et al. (2017) proposed to use the fusion of accelerometer and gyroscope sensors in order to offer implicit and continuous smartphone-based user authentication. The movement signals were collected from 24 users over a week, each was asked to install an application that continuously record the mobile movement data when users interact with their smartphone. Time and frequency domain features were extracted to train the authentication model such as Std, min, and max. To classify the arm movement pattern, DTW was utilized and reported 0% of FAR and 7.6% of FRR. These errors were increased to 6.4% and 13.7% of FAR against 11.8% and 15.2% of FRR by using the accelerometer and gyroscope sensors respectively. Therefore, the presented results showed the necessity to use the combined signals of both sensors (i.e., acceleration and gyroscope) to improve the system performance. Although each participant provided 7 days data, the proposed algorithm to extract the user's motion samples was able to capture only 18 samples per day from each user, which is limited to train and test a behavioural-based biometric system.

Continuous user identification via touch and movement behavioural biometrics was suggested by Shen *et al.*, (2018); multi-motion sensors were utilized to capture the acceleration, gyroscope, orientation, and magnetometer data from 102 users over multiple days. All participants were asked to provide three touch-

based movement scenarios: touching on a smartphone while a user sitting and/or standing, putting the device on a table and continuously interacting with the touch screen of the smartphone, and interacting with the smartphone on the go. In total, more than 520,000 samples were extracted, each sample contained 0.73 second of the movement data. For each sample, more than 190 frequency and wavelet domain features were extracted such as energy, entropy, mean and cross mean rate. Applying the HMM to a short segment size (i.e., 0.73 second) reported a high EER of about 27%, and the authentication performance was greatly improved (i.e., 4.93% of EER) by increasing the segment size into 8 seconds. However, this would increase the required time for the authentication decision as the authors used a large feature subset size (i.e., 192 features) and did not carry out any feature selection approach to remove the redundant or irrelevant features.

Although mobile-based gait authentication provides an unobtrusive and user-friendly method for authentication, the majority of previous studies collected the motion data by placing a mobile phone in a fixed position (i.e., in the trouser pocket or on the hip). However, users can put their phone in numerous locations around their body wherever there is a pocket (i.e., inside coat pocket and back pocket). Moreover, the collected signals by smartphones are too noisy that require extensive pre-processing, which add extra cost in terms of the required resources.

More recently with the introduction of smartwatches, the feasibility of using the accelerometer and gyroscope on the wrist offer the opportunity to provide more granular monitoring of physical movement. Mare et al., (2014) proposed a study to recognize users that were interacting with computer (specifically, interacting with a keyboard and/or mouse). The acceleration and gyroscope data were collected from 20 participants on a single day (data obtained within a constrained environment). An average of 485 samples were obtained from each user (each sample included one second of the user's movement data). Usually, one second of the user's motion data is not sufficient for identifying the user's identity, therefore, the authors included 21 seconds in each single sample (which resulted in a total of 23 samples for each user). To avoid the effect of the orientation, the collected signals across all three axes (x, y, and z) were combined into a single dimension (i.e., magnitude). Thereafter, several statistical features were extracted (e.g., power energy, peak to peak amplitude, mean, and median) and classified by utilizing the Random Forest algorithm. The best obtained results were 90% correct classification rate of the legitimate user against 100% for identifying imposters. However, more investigation for the sample size (i.e., segment size) is required as the authentication accuracies of this study obtained by asking users to type on the keyboard or move the mouse continuously for 21 seconds, which is a bit inconvenient for users.

Johnston and Weiss, (2015) conducted a study to collect the movement data from LG G Watch sensors, accelerometer and gyroscope. For the experiment, 59 subjects were involved; each subject was requested to wear the watch on their dominant hand and walk using their natural gait speed (the gait samples were obtained within a constrained environment. At least 5 minutes of activities were captured from each user in one session, except few users that contributed of only

2 minutes. Subsequently, the raw data of each sensor was divided into 10 seconds segments at the sampling rate of 20Hz. The statistical features (average, Std, AAD, TBP, BD, ARCV, and ARCA) were computed for each sensor separately. Since all features were extracted, a single predictive model for each genuine user was created. To train each model, 80% of the genuine user's data and the data of four imposters were selected (for a 1:4 genuine to imposter ratio of data). Testing the model of each genuine user was performed by selecting four random users (which their data was not in the training dataset) and the remaining 20% of the genuine user's data. Utilizing the acceleration feature vectors only, the authors were able to achieve EERs of 1.4%, 2%, 2.5%, and 4.5% using Random Forest, Multilayer Perceptron, Rotation Forest, and Naive Bayes respectively. However, when the gyroscope features were used, the resulted EERs were significantly increased to 9.6%, 6.3%, 7.0%, and 9.6% (these results were generated via applying the same aforementioned algorithms). With a majority voting scheme, the proposed system managed to attain 0% EER (i.e., 100% accuracy) with 50 seconds of data. Although the reported results were strong and based upon 50 seconds of authentication data, it has to be noted that the data was collected on the same day. Moreover, the testing for authentication is rather strange; besides the genuine user is data of 4 impostor users used in the training, the testing was done with data of 4 impostor users only (as well as the genuine user of course), which is not a comprehensive test to claim the system is robust to impersonation attacks.

A gesture-based user authentication system was suggested by Junshuang Yang et al., (2015) using Samsung smartwatch sensors (accelerometer and gyroscope). Four different gestures were evaluated in this study, forearm rotation about 90 degree clockwise (rotation), drawing a circle (circle), arm down (down),

and arm up (up). In the data collection phase, 26 subjects were involved; each was asked to wear the smartwatch and perform 40 gestures per session (10 times for each of the above gestures). In total, 160 gestures were collected per user over multiple days. The accelerometer and gyroscope readings were then converted into 30 features; these included the magnitude of the acceleration M, the corresponding first and second derivatives of each measurement, and the three angles between M and x/y/z. The histogram and DTW were used as feature extraction methods. The former was used to compute normalized n-bin histogram from the 30 features, and then Manhattan distance function was applied to calculate the distance between two histograms (reference and probe histograms). This distance value represented the similarity score between two gesture samples. The latter (DTW method) was utilised to compute the distance between every two gestures in the training set. Thereafter, the gesture that had a lowest DTW distance to other gestures was selected as a reference gesture. Applying histogram method for the circle, down, up and rotation gestures, EERs of 2.6%, 3.1%, 2.9 and 4.7% were obtained respectively. These results were slightly decreased to 3.8%, 4%, 4.7% and 7.7% when DTW method was applied. Nerveless, the EERs of multiple users were more than 20%.

Another smartwatch- based authentication study was conducted by Kumar, et al.,(2016) by analysing the user's walking pattern. Different segment sizes (i.e., 2, 4, 6, 8, and 10 seconds) were tested and the segment size of 10 seconds achieved the best authentication performance (i.e., correct classification rate of 95%). Two different scenarios were applied, single and cross day; the former scenario included the participation of 40 users while only 13 users participated in the latter scenario (i.e., cross day). The authors utilized two feature selection algorithms, namely Information Gain Based Feature Ranking and Correlation

Feature Selection that successfully reduced 25% of the total time and frequency features. To evaluate the system accuracy, k-NN was utilized and reported 95% correct classification rate by using the single day scenario. However, the system performance was reduced to 86.8% when the cross day scenario was applied. This can be an indication that the feature selection approach that was used is not sophisticated enough to identify an optimal and unique feature set for individuals that work over time. Another reason could be that user's behaviour changes over time; hence a template renewal mechanism is required.

Davidson *et al.*, (2016) utilized the smartwatch acceleration data to capture five simple activities (i.e., walking, typing, open a door, lifting a cup, and interacting with the smartwatch) to offer active and transparent authentication. Their experiment involved the participation of 10 users, and the segment-based approach was used to divide the raw signal into 5 seconds. In total, users provided only 5 samples for each activity, which is limited amount of data to train and test any biometric system. The reference and test templates were generated by extracting time and frequency domain features, which were selected based upon prior work identified in gait recognition studies. The true positive (TP) and false positive (FP) rates were used to evaluate the system efficiency. High true positive rate (i.e., above 90%) represents that the proposed system more frequently accepts a legitimate user while low false positive rate indicates to a high probability of preventing unauthorized access. To distinguish between a legitimate and an imposter, K-NN was used and reported 88.4% TP and 1.3% FP.

The fusion of four smartwatch and smartphone sensors (i.e., accelerometer, gyroscope, magnetometer, and orientation) was investigated by (Shrestha et al., 2016) in order to improve the gait recognition accuracy. The user's walking pattern was captured from a set of 18 users, each user walked naturally about 35

meters per day and repeated the same experiment over multiple days. In total, 50 gait samples were obtained from each participant. More than 300 features (e.g., range, mean, and root mean square) were used as input for training and testing the Random Forest classifier. The findings showed EERs of 8.75% (smartwatch sensors only), 4.5% (data of the smartphone) 2.6% (fusion signals of the smartwatch and smartphone). As expected capturing data from both devices improved the system accuracy. Nevertheless, this would require complex computational processing and hence high demand upon the battery (which is one of the biggest qualms of these devices). Moreover, it was unclear which scenario (i.e., single, mixed or cross day scenario) was applied to train and test the classifier.

A preliminary study by Lewis et al., (2016) considered 3D arm gesture of 5 users over multiple days; cycle based method was used to segment the raw acceleration data. The study reported high error rate of about 30% FRR against 15% FAR by utilizing DTW for classifying the movement pattern of individuals. This high error rates might be the lack of using the appropriate method of dividing the motion signal as well as using small dataset size.

Dong and Cai (2016) conducted a test with 20 users providing single day of the acceleration and gyroscope gait data. The motion signal was collected by utilizing sensors on commercial devices (i.e., Samsung Galaxy Gear 1 and Samsung Galaxy S4). The walking signal was segmented by applying sliding window approach and then transferred into time domain features to represent the reference and probe templates. The division of training and testing data was rather strange where 90% of the collected samples (i.e., 40 samples) was used to train SVM and the remaining 10% samples (i.e., 10 gait samples only) for testing. Using the smartwatch sensors data reported an EER of 4.36% against

2.40% of EER when the smartphone motion signal was utilized. The system performance was significantly improved by using the fusion data of both devices and showed 0.65% EER. Nevertheless, all users were asked to wear flat shoes, which is unexpected condition in the real-life usage.

An empirical authentication system called 'iAuth' was proposed by Lee and Lee (2017); the acceleration and gyroscope data of a smartphone and smartwatch were collected. Considerable number of samples (i.e., 1,200 samples) were acquired over multiple days to generate the feature vector of each individual. Using 6 seconds of data (i.e., the segment size of each sample), the time domain features (e.g., mean and variance) and frequency features (e.g., amplitude of first highest peak) were extracted. Two thirds (i.e., 800 samples) were used to train the Kernel Ridge Regression and the rest of samples for testing the proposed system. An experimental evaluation was carried out that included a set of 20 users; the first experiment that utilized the smartphone acceleration signal reported a high FRR of 22.3% and 13.4% of FAR. Using the combination of smartwatch and smartphone sensors significantly decreased the error rates into 8.3% of FRR against 7.5% of FAR. Nevertheless, the proposed system was carried out on a specific cloud server which is additional cost to consider when implementing iAuth tool. Moreover, the authors did not mention the strategy of selecting the training and test samples (i.e., is their system equivalent to single, mixed, or cross day scenario?).

The feasibility of handwriting based-biometric authentication using smartwatches was investigated by Griswold-Steiner et al. (2017). The acceleration and gyroscope of the writing activity was captured from a group of 20 users over multiple days. Each user was asked to participate in three different experiments (i.e., EXP1, EXP2, and EXP3); EXP1 involved writing pre-defined text, copy a

random text was the task of EXP2, while data of EXP3 was captured by asking users to answer a questionnaire. Different segment sizes were tested (10, 20, 30, ….., and 70 seconds) and the best results obtained by increasing the segment length into 70 seconds. More than 360 time and frequency domain features (e.g., zero cross, absolute difference, mean, and root mean square) were used to train and evaluate the effectiveness of SVM classifier. The lowest EERs were nearly 7%, 10%, and 15% for EXP1, EXP2, and EXP3 respectively. However, in the real scenario, it is not expected that all users can type 70 seconds continuously without pause and this can be cumbersome. When the more realistic test (i.e., using segment size of 10 seconds) took a place, the system performance significantly decreased into 12.5%, 16%, and 20% of EERs. Moreover, the proposed system was limited to a specific surface (i.e., writing on a piece of paper); hence more than one surface should be considered (e.g., typing on touch screen and typing on a PC keyboard), along with their impact upon the performance.

Liang et al., (2017) proposed a system to authenticate individuals based upon their punch gesture; Samsung Gear Fit 2 was used to collect the acceleration signal at a sampling rate of 100HZ from 20 subjects over multiple days. Time domain features were extracted to generate the user's reference and test templates. The reported results of training and testing the SVM algorithm was an EER of 4%. Apart from that the punch activity is not a realistic gesture for transparent user authentication system and does not provide continuous verification, the data collection methodology was quite intrusive (i.e., data was not captured transparently, as each user was asked to hold his/her hand and press a button in order to start and finish the data collection).

Another gesture based- authentication study through handwaving biometrics was suggested by Wang et al., (2017). Segment based-approach was used to divide the raw acceleration data of a single day, which resulted in a total of 15 samples for each volunteer. To classify the characteristic of the user's handwaving pattern, Manhattan distance was utilized and reported 4.3% of EER for a dataset of 10 users only. Nevertheless, the selected gesture seems not robust against imitation-attack scenario as the EER was significantly increased to 14.5% when attackers masquerade as legitimate users.

Xu *et al.*, (2017) carried out a study in order to recognize the user's walking pattern using a smartwatch. Their dataset consisted of 20 users, each provided 20 minutes of the walking signal on two different days. The raw acceleration signal was segmented into cycles, and each segment contained 5 cycles. The detected cycles represented the unique characteristics of the user's gait templates. To validate the effectiveness of the proposed system, K-NN was utilized and achieved more than 96% correct classification rate. Nevertheless, the proposed method to detect the gait cycles is cost to implement on digital devices (i.e., identifying the walking cycles and then normalized the length of each cycle).

Activity based-user authentication for cloud-based services was developed by Ahmad *et al.*, (2018); data of three smartwatch sensors (i.e., accelerometer, gyroscope, and magnetometer) was captured from 6 users only. Each user performed 5 gait activities (i.e., normal walking, walking up and downstairs, running, and jogging) for about one month. The captured signal was then divided into 30 seconds using the segment-based approach. The time and frequency features were fed into four machine learning algorithms (i.e., decision tree, K-NN, SVM, and Naïve Bayes). The findings showed that decision tree overcome other classifiers with an average of 90.4%, 90.2%, and 77% correct classification rates

for the acceleration, gyroscope, and magnetometer data respectively. The results also highlighted that the normal walking was more distinctive than other activities reporting more than 98% of correct classification rate, which confirms the findings of the prior acceleration-based activity recognition studies.

Most recent smartwatch based-user authentication study by Acar *et al.*, (2018) proposed the feasibility of utilizing the user's typing rhythm on a PC keyboard. The collected acceleration and gyroscope signals were segmented into 20 and 30 seconds (using sliding window approach) and then transferred into time and frequency domain features (e.g., covariance, correlation, and entropy). Each user was asked to participate in 3 different sessions on a single day (i.e., the first and second sessions consisted of typing predefined text while imitating a legitimate user in the third session, each trail included 4 minutes of the motion data). A Feedforward Multi-Layer Perceptron was able to predict 34 users with EERs of 1% and 2% using a segment length of 30 and 20 seconds respectively. These findings were not affected when an unauthorized user attempted to imitate someone else with 99% correct classification rate. However, the robustness of such a system requires to collect data on multiple days in order to show the variability of the human's typing rhythm.

Appendix B: Publications

N. Al-Naffakh, N. Clarke, P. Haskell-Dowland, and F. Li, “A Comprehensive Evaluation of Feature Selection for Gait Recognition Using Smartwatches,” International Journal for Information Security Research, vol. 6, no. 3. pp. 691-700, Sep. 2016.

N. Al-Naffakh, N. Clarke, P. Dowland and F. Li, "Activity Recognition using Wearable Computing", in Proceedings of the 11th International Conference for Internet Technology and Secured Transactions (ICITST- 2016), Barcelona, 2016, pp. 189-195.

N. Al-Naffakh, N. Clarke, F. Li and P. Haskell-Dowland, "Unobtrusive Gait Recognition Using Smartwatches", in the 12 International Conference of the Biometrics Special Interest Group (BIOSIG), pp. 1-8, 2017.

N. Al-Naffakh, N. Clarke, and F. Li, " Continuous User Authentication using Smartwatch Motion Sensor Data", to appear in IFIPTM International Conference on Trust Management. Part of the IFIP Advances in Information and Communication Technology book series (IFIPAICT, volume 528, Springer.). Toronto, Canada, Jul 9, 2018

N. Al-Naffakh, Clarke, N. and Li, F. (2019). Implicit Authentication using Activity Recognition on a Smartwatch. Manuscript submitted for publication in MOBILE COMPUTING.