2019

Transparent Authentication Utilising Gait Recognition

Al-Obaidi, Hind K

http://hdl.handle.net/10026.1/14898

http://dx.doi.org/10.24382/854
University of Plymouth

UNIVERSITY OF PLYMOUTH

Transparent Authentication Utilising Gait Recognition

By

Hind Khudhair Al-Obaidi

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

DOCTOR OF PHILOSOPHY

School of Engineering, Computing and Mathematics

September 2019

COPYRIGHT STATEMENT

Acknowledgements

First and foremost, all praise and gratitude are due to Allah the All-Compassionate and All-Merciful for everything He has provided me throughout my life, in general, and for giving me the potential and patience to persevere and reach this stage of my PhD research, in particular. Without Him, I would not have achieved anything or even existed.

The work, and indeed this PhD, would not have been possible without the great help of my Director of Studies, Professor Nathan Clarke. Thanks go to him for his endless support and guidance throughout this amazing journey. Without his excellent scientific knowledge, encouragement and enthusiasm, this work would not be completed. Very deep and special thanks also go to my supervisor, Senior Lecturer Dr Fudong Li, who has spent a lot of time and efforts proofreading papers and my thesis, in addition to providing helpful advice throughout my studies. Many thanks to my supervisor, Associate Professor Dr Bogdan Ghita, for his insights and support. Moreover, I would like to thank all staff members in the Centre for Security, Communications and Network Research Group for their support.

My sincere appreciation goes to the soul of my beloved parents and many thanks to my dear brother Muhammad who has always been supportive to me whenever I am in need and without any reservation. May Allah bless him.

My unreserved love, thanks and appreciation must go to my husband (Furat) and children (Danyah and Fahad) who have been very patient, understanding, and inspiring to me throughout this endeavour, spending days, nights, and sometimes even holidays without me. Hope that the potential success of this research will compensate some of what they have missed. May Allah bless them.

I would like to thank all my friends and colleagues, either in the UK or in Iraq, who have contributed positively towards my progress by any means even if it was just a smile. Many of them deserve mentioning, but it is difficult to state all the names. However, it is inevitable not to express my sincere thanks to my younger brother-like colleague Abdulrahman, for all kinds of support he has provided throughout my PhD journey. I wish them all success in all their current and future endeavours.

Finally, I would like to acknowledge with thanks and appreciation the Ministry of Higher for Education and Scientific Research in Iraq and my employer, the University of Baghdad / College of Education for Pure Science – Ibn Al-Haitham / Computer Science Department.

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 the prior agreement of the Graduate Committee.

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

This study was financed with the aid of a scholarship from the government of Iraq. Relevant scientific seminars and conferences attended at which work was often presented, and several papers were published and prepared for publication in the course of this research project. Additional research skills development courses also attended.

1- Al-Obaidi, H. et al., 2018. A Multi-Algorithmic Approach for Gait Recognition. In ECCWS 2018 17th European Conference on Cyber Warfare and Security (p. 20). Academic Conferences and publishing limited. https://researchportal.port.ac.uk/portal/files/11163228/A_multi_algorithmic_approach_for_gait_recognition.pdf.

2- Alruban, A. et al., 2018. Human Activity Recognition for Healthcare using Smartphones. In ICPRAM 2019 8th International Conference on Pattern Recognition Applications and Methods, pp.20-21. DOI: 10.5220/0007271903420351.

Word count of the main body of thesis: 53,467 words

Signed...

Date: 03/09/2019

Abstract

Transparent Authentication Utilising Gait Recognition

Hind Al-Obaidi (MSc)

Securing smartphones has increasingly become inevitable due to their massive popularity and significant storage and access to sensitive information. The gatekeeper of securing the device is authenticating the user. Amongst the many solutions proposed, gait recognition has been suggested to provide a reliable yet non-intrusive authentication approach – enabling both security and usability. While several studies exploring mobile-based gait recognition have taken place, studies have been mainly preliminary, with various methodological restrictions that have limited the number of participants, samples, and type of features; in addition, prior studies have depended on limited datasets, actual controlled experimental environments, and many activities. They suffered from the absence of real-world datasets, which lead to verify individuals incorrectly.

This thesis has sought to overcome these weaknesses and provide, a comprehensive evaluation, including an analysis of smartphone-based motion sensors (accelerometer and gyroscope), understanding the variability of feature vectors during differing activities across a multi-day collection involving 60 participants. This framed into two experiments involving five types of activities: standard, fast, with a bag, downstairs, and upstairs walking. The first experiment explores the classification performance in order to understand whether a single classifier or multi-algorithmic approach would provide a better level of performance. The second experiment investigated the feature vector (comprising of a possible 304 unique features) to understand how its composition affects performance and for a comparison a more particular set of the minimal features are involved. The controlled dataset achieved performance exceeded the prior

work using same and cross day methodologies (e.g., for the regular walk activity, the best results EER of 0.70% and EER of 6.30% for the same and cross day scenarios respectively). Moreover, multi-algorithmic approach achieved significant improvement over the single classifier approach and thus a more practical approach to managing the problem of feature vector variability.

An Activity recognition model was applied to the real-life gait dataset containing a more significant number of gait samples employed from 44 users (7-10 days for each user). A human physical motion activity identification modelling was built to classify a given individual's activity signal into a predefined class belongs to. As such, the thesis implemented a novel real-world gait recognition system that recognises the subject utilising smartphone-based real-world dataset. It also investigates whether these authentication technologies can recognise the genuine user and rejecting an imposter. Real dataset experiment results are offered a promising level of security particularly when the majority voting techniques were applied. As well as, the proposed multi-algorithmic approach seems to be more reliable and tends to perform relatively well in practice on real live user data, an improved model employing multi-activity regarding the security and transparency of the system within a smartphone. Overall, results from the experimentation have shown an EER of 7.45% for a single classifier (All activities dataset). The multi-algorithmic approach achieved EERs of 5.31%, 6.43% and 5.87% for normal, fast and normal and fast walk respectively using both accelerometer and gyroscope-based features – showing a significant improvement over the single classifier approach. Ultimately, the evaluation of the smartphone-based, gait authentication system over a long period of time under realistic scenarios has revealed that it could provide a secured and appropriate activities identification and user authentication system.

Contents

COPYRIGHT STATEMENT iii

Acknowledgements iv

Author's Declaration vi

Abstract vii

1 The Need for Authentication 1

1.1 Introduction and overview 1

1.2 Research Aims and Objectives 8

1.3 Thesis Structure 8

2 Biometrics Authentication 11

2.1 Introduction 11

2.1.1 History of biometrics 12

2.1.2 Biometric system requirement 13

2.2 Classification Approaches 17

2.2.1 Statistical 18

2.2.1.1 Dynamic Time Warning (DTW) Distances 18

2.2.2 Machine learning 18

2.2.2.1 Support-vector machines (SVM) 19

2.2.2.2 A classification tree or a decision tree 20

2.2.2.3 A hidden Markov model (HMM) 20

2.2.2.4 Nearest Neighbour 21

2.2.2.5 Neural Network multilayer perceptron (NN-MLP) 21

2.3 Biometric system performance measurement factors 22

2.4 Biometric Techniques 24

2.4.1 Physiological Biometrics 25

2.4.1.1 Fingerprint Recognition 25

2.4.1.2 Palm print and Hand Geometry 28

2.4.1.3 Facial Recognition 29
2.4.1.4 Iris Recognition and Retina Recognition 30
2.4.1.5 Ear Geometry 32
2.4.2 Behavioural Biometrics 33
2.4.2.1 Voice Recognition 33
2.4.2.2 Signature Recognition 35
2.4.2.3 Keystroke Analysis 37
2.4.2.4 Behavioural Profiling 38
2.4.2.5 Biometric Gait Recognition 39
2.5 Summary of the Biometric Techniques 41
2.6 Conclusions 42
3 A Literature Review of Gait Recognition 44
3.1 Background of Biometric Gait Recognition 44
3.1.1 Gait collection methods 44
3.1.2 Machine Vision (video sensors) 44
3.1.2.1 Wearable sensors 45
3.1.2.2 Mobile-based sensors 46
3.1.3 Gait process approaches 47
3.1.3.1 Cycle-based Method 47
3.1.3.2 Segment-based Method 50
3.2 Review Methodology 52
3.3 Overview of Mobile-based Gait Authentication Related Work 53
3.4 Discussion 88
3.5 Conclusion 98
4 Research Methodologies 100
4.1 Introduction 100
4.2 Research Methodology 102
4.3 Technology Assessment 105
4.3.1 Preliminary Testing 108

4.4 Control Conditions Experimental Methodology 110
4.4.1 Data Collection 112
4.4.1.1 Activity labelling 115
4.5 Uncontrolled Condition Experimental Methodology 116
4.5.1 Real-World Data Collection 118
4.5.2 Activity Identification..119
4.5.3 Activity-Based Recognition...120
4.5.4 Activities Identification Data Modelling 122
4.6 Signal pre-processing and Feature Exploration 127
4.6.1 Segmentation 128
4.6.2 Feature extraction 128
4.6.2.1 *Time-domain features*.. 129
4.6.2.2 *Frequency domain features* 130
4.6.3 Normalisation 131
4.6.4 Feature selection...132
4.6.5 Classification 132
4.7 Conclusion 134
5 Experimental Results of Exploring Classification Strategies 136
5.1 Introduction 136
5.2 Investigation of Feature Vector Composition 137
5.2.1 Investigating the Sensors and Feature Vectors...........................137
5.2.2 Investigation of the Feature Vector across Activities 147
5.2.3 Investigation Using the Feedforward Neural Network Classifier..163
5.2.4 Performance of the Neural Network Feedforward Classifier........165
5.4 Discussion..172
5.5 Conclusion 178
6 Experimental Results: Real-Life Gait Recognition 180

6.1 Introduction ... 180

6.2 Activities Identification Results ... 181

6.3 Exploring More Discriminative Features for Different Real-World Activities 183

6.4 Real-World Gait Authentication and Multi-Algorithm Performance ... 189

6.5 Applying Majority Voting ... 196

6.6 Discussion ... 202

6.7 Conclusion ... 208

7 Discussion ... 210

7.1 Introduction ... 210

7.2 Comparison with the Prior Art ... 211

7.2.1 Real-World Usability Performance ... 218

7.3 Proposed Context-Awareness Model ... 223

7.3.1 Data collection ... 225

7.3.2 Feature extraction ... 225

7.3.3 Classification and Decision Making ... 226

7.4 Gait Recognition Using Context Information ... 227

7.5 Conclusion ... 234

8 Conclusions and future work ... 235

8.1 Contributions and Achievements of the Research ... 235

8.2 Limitations of research ... 237

8.3 Suggestions, Scope for future work ... 238

8.4 The Future of Authentication ... 239

References ... 241

Appendices 258

Appendix A- Publications 258

Appendix B- Consent Form and Information Sheet (Data Collection) 260

Appendix C- Top Ten Discriminative Features for each user in Fast walk 262

Appendix D- Top Ten Discriminative Features for each user in walking with a bag Activity 263

Appendix E- Top Ten Discriminative Features for each user in Down Stairs Activity 264

Appendix F- Top Ten Discriminative Features for each user in Upstairs Activity 265

Appendix G- Top Ten Discriminative Features for Each User in Fast Walking Activity 266

Appendix H- Top Ten Discriminative Features for Each User in Down Stairs Walking Activity..267

Appendix I- Top Ten Discriminative Features for Each User in Walking Upstairs Activity 268

Appendix J- Top Ten Discriminative Features for Each User in All Activities ..269

List of Figures

Figure 2-1: The Components of a Biometrics System (Clarke 2011) 17
Figure 2-2: Classification approaches Taxonomy (Technology at Nineleaps, 2019) 19
Figure 2-3: Biometrics Performance Metrics Factors (Clarke 2011) 24
Figure 2-4: An example of fingerprint recognition (NEURO technology 2015)...25
Figure 2-5: Example Distance Measurements Source: (NSTC, 2006) 28
Figure 2-6: An example of facial recognition (GreekoSystem 2011) 30
Figure 2-7: An example of a human eye (NSTC 2006; Monitgomery, 2014).....31
Figure 2-8: Images of Human Ears (da/sec. 2015) 33
Figure 2-9: Voice Recognition Platform (BiometricSolutions 2015) 35
Figure 2-10: Signature Recognition (Signing Hub 2016) 36
Figure 2-11: Illustrations of periodic motion of the legs (Hoang et al. 2013) 40
Figure 3-1: Background segmentation for extracting the silhouette picture (Bajrami 2011) 45
Figure 3-2: Different locations of the attached wearable sensor (Gafurov et al. 2007b; Gupta & Dallas 2014) 46
Figure 3-3: The three axes in which acceleration is measured and phone position when receiving data (Miguel & Neves 2013) 47
Figure 3-4: A complete gait cycle showing the eight gait phases (Miguel & Neves 2013) 47
Figure 3-5: Phone attached to the subject and the three axes in which acceleration measured (Derawi et al. 2010) 55
Figure 3-6: Sample data collected with the G1 from x, y and z directions 55
Figure 3-7: Gait Cycle Detection 56
Figure 3-8: Quorum voting scheme method (#V total segments of a probe gait signal, #Vg, number of votes for genuine, #GV positive classification results) (Nickel, et al. 2011b) 58
Figure 3-9: Acceleration plots for the (a) walking, (b) ascending, and (c) descending activities (Kwapisz et al. 2011) 62
Figure 3-10: Acceleration waveforms of the smartphone and tri-axial accelerometer (Nishiguchi et al. 2012) 68
Figure 3-11: Flow control of cycle extraction steps, dashed arrow lines indicate optional steps (Muaaz & Nickel 2012).................71

Figure 3-12 the process of extracting BFCC features (Hestbek et al. 2012) 79
Figure 4-1: (a) and (b) show Sensor Tracker and AndroSensor, respectively.. 107
Figure 4-2: Illustrates the accelerometers X, Y and Z-axis signals of three users 108
Figure 4-3: Illustrates the gyroscopes X, Y and Z-axis signals of three users.. 109
Figure 4-4: (a) and (b) the phone is placed in the right or left belt pouch and its orientation 112
Figure 4-5: The Accelerometer and gyroscope raw gait data recorded along X-axis........ 114
Figure 4-6: Authentication Process Systems. Each activity investigated individually then all users' gait activities were treated as a single dataset........ 115
Figure 4-7: Activities Identification Model 123
Figure 4-8: Real-World gait data identification and recognition 125
Figure 5-1: The EER results on Accelerometer data for all activities by using SD scenarios........ 138
Figure 5-2: The EER results on Accelerometer data for all activities by using CD scenarios........ 139
Figure 5-3: The EER results on Gyroscope data for all activities by using SD. 141
Figure 5-4: The EER results on Gyroscope data for all activities by using CD 141
Figure 5-5: The EER results on Accelerometer and Gyroscope data for all activities by using SD and CD scenarios 143
Figure 5-6: The impact of using dynamic feature for all activities (same day scenario) 145
Figure 5-7: The Impact of using dynamic feature for all activities (cross day scenario) 145
Figure 5-8: Impact of the dynamic feature selection technique upon the performance, normal walking (same day scenario) 150
Figure 5-9: Impact of the dynamic feature selection technique upon the performance, normal walking (cross day scenario) 150
Figure 5-10: Impact of the dynamic feature selection technique upon the performance for fast walking (same day scenario) 151
Figure 5-11: Impact of the dynamic feature selection technique upon the performance for fast walking (cross day scenario)........ 151
Figure 5-12: Impact of the dynamic feature selection technique upon the performance for walking with a bag (same day scenario)........ 152

Figure 5-13: Impact of the dynamic feature selection technique upon the performance for walking with a bag (cross day scenario) 152
Figure 5-14: Impact of the dynamic feature selection technique upon the performance for down stairs (same day scenario) 153
Figure 5-15: Impact of the dynamic feature selection technique upon the performance for down stairs (cross day scenario) 153
Figure 5-16: Impact of the dynamic feature selection technique upon the performance for upstairs (same and cross day scenario) 154
Figure 5-17: Impact of the dynamic feature selection technique upon the performance for upstairs (cross day scenario) 154
Figure 5-18: (a) Acc (x-axis) covariance feature values for five users (13F) 162
Figure 5-19 Figure 5-20 178
Figure 6-1: Normalised confusion matrix (%) of the soft voting model 183
Figure 6-2: (a) and (b) show User1's profile signals for feature 243 and 89 from both the controlled and uncontrolled datasets 188
Figure 6-3 (a) and (b) show User1's signals for feature 13 and 56 from both the controlled and uncontrolled datasets 188
Figure 6-4: The distribution of data points using a 2D plot from 243F and 89F 189
Figure 6-5: The EER (%) of individual performance for normal walking activity 194
Figure 6-6: The EER (%) of individual performance for all activities 195
Figure 6-7: Majority voting mean values using different numbers of data samples for multi-algorithm walking activities (10-second sample period) 198
Figure 6-8: Majority voting mean values using different numbers of data samples for single algorithm walking activities (10-second sample period) 198
Figure 6-9: Majority voting median values using different numbers of data samples for (a) normal walking, (b) fast walking. 199
Figure 6-10: Majority Voting Median Values Using Different Numbers of Data Samples for (a) Down Stairs, (b) Upstairs walking 199
Figure 6-11: Majority voting median values using different numbers of data samples for (a) normal and fast walking, (b) all walking activities 200
Figure 7-1: The average of daily gait activity time in minutes for all users 219
Figure 7-2: Context-awareness gait recognition model 225
Figure 7-3: Two days' GPS tracking data and Google Maps direction for User1 229

Figure 7-4: Two days' GPS tracking data and Google Maps direction for User 2 229
Figure 7-5: Four days' GPS tracking data for User 1............230
Figure 7-6: Six days' GPS tracking data for User2 231
Figure 7-7: Four days' GPS tracking data for User3............232
Figure 7-8: Four days' GPS tracking data for User4............233

List of Tables

Table 2-1: shows the aforementioned biometric techniques transparency feature 41
Table 3-1: The number of returned references 53
Table 3-2: Surface and walking distances used in the experimental protocol (Sprager and Zazula 2011). 59
Table 3-3: A comparisons of results with and without PLA-based gait cycle extraction (Muaaz & Mayrhofer 2013) 73
Table 3-4 Evaluation of results with and without wavelet transforms (Hestbek et al. 2012) 79
Table 3-5 shows the evaluation of the essential facts of the SVMs, HMMs, and k-NN in the same database (Nickel et al. 2012) 81
Table 3-6 The stated results FAR (%) and FRR (%) for each holding state by using five algorithms (Watanabe 2014) 83
Table 3-7 Correctly classified rate (%) of states by four algorithms when a subject walked in nine different states (Watanabe 2015) 85
Table 3-8 (a) Improvements of segmentation based on gait cycles compared with a fixed length, (b) The influence of disorientation error to the effectiveness of classification mode 86
Table 3-9: Comprehensive analysis of the prior studies on gait authentication systems using mobile sensors. 90
Table 4-1: The tested software with sensors that could be extracted by the smartphone 106
Table 4-2: Age and gender distribution of participants 113
Table 4-3: Activity states for all users across two days 114
Table 4-4: Comparison of prior studies in activity recognition using smartphone sensors 122
Table 4-5: Activities states for each user for one week 126
Table 4-6: The overview of the unconstrained dataset 127
Table 4-7: Time domain features 130
Table 4-8: Frequency Domain Features 131

Table 5-1: System Performance Utilising Dynamic Feature Selection Technique on Accelerometer Data 139
Table 5-2: System Performance Utilising Dynamic Feature Selection Technique on Gyroscope Data 142
Table 5-3: System Performance Utilising Dynamic Feature Selection Technique on Accelerometer and Gyroscope Data 143
Table 5-4: The Impact of Using Dynamic Feature for All activities (Same and Cross day) 145
Table 5-5: Accelerometer Features list (1-152) and Gyroscope Features list (153-304) 148
Table 5-6: A Comparison between Dynamic and Static features Techniques for Individual and All Activities (SD and CD scenarios) 155
Table 5-7: Top Ten Discriminative Features for each user in Normal walk 158
Table 5-8: A Comparison between Two Users Best Features Patterns 160
Table 5-9: Summary of the Top Repeated Features for Each Activity-Control Dataset 161
Table 5-10: Comparison between FF MLP and SVM performances (for similar feature vector set FN) 164
Table 5-11: The EER (%) of the SD test for the normal walking activity 166
Table 5-12: The EER (%) of the SD test for the fast walking activity 166
Table 5-13: The EER (%) of the SD test for the walking with a bag activity 167
Table 5-14: The EER (%) of the SD test for the walking down stairs 167
Table 5-15: The EER (%) of the SD test for the walking upstairs 167
Table 5-16: The EER (%) of the SD test for all activities 168
Table 5-17: The EER (%) of the CD test for normal walking 169
Table 5-18: The EER (%) of the CD test for fast walking 170
Table 5-19: The EER (%) of the CD test for walking with a bag 170
Table 5-20: The EER (%) of the CD test for walking down stairs 171
Table 5-21: The EER (%) of the CD test for walking upstairs 171
Table 5-22: The EER (%) of the CD test for All Activities 172
Table 5-23: The positive impact of using dynamic features selection techniques on the system performance and the features used ratio 175
Table 6-1: Overall classification accuracy for each model 182
Table 6-2: Top ten discriminative features for each user in the normal and walking with a bag 185

Table 6-3: A comparison between two users' best feature patterns (real dataset) 186
Table 6-4: Summary of the top repeated features for each activity -real dataset 187
Table 6-5 and 6-6: The EER (%) of normal walking activity utilising different feature subsets 192
Table 6-7 and Table 6-8: The EER (%) of fast walking activity utilising different feature subset size 193
Table 6-9: The EER (%) of normal and fast walking activities utilising different feature subset size 193
Table 6-10: The best EER (%) for individual and all activities 194
Table 6-11: Majority voting results for each number of samples across all gait activities 202
Table 6-12: Top six ranked features (time between peaks) 205
Table 6-13: Number of features included in the repeated top-ten 205
Table 6-14: Individual performance for each activity 206
Table 6-15: System performance utilising the majority voting module 207
Table 7-1: A comparison between some common selected smartphone-based databases 212
Table 7-2: Percentage of identified real activities samples 220
Table 7-3: Comparing controlled and realistic system performance with and without using majority voting 221

1 The Need for Authentication

1.1 Introduction and overview

In the last decade, smartphone devices have become a ubiquitous technology, with more than 9.5 billion users globally (GSMA 2018). Currently, smartphones provide a wide range of services and features including (but not limited to): communication (e.g., texting, email, and calling), entertainment (e.g., internet-connected game consoles such as Xbox One and PS4, music, and online streaming), work (e.g., viewing clients' documents), financial services (e.g., transferring money and shopping online), sensors (e.g., accelerometers, gyroscopes, magnetometers, rotation sensors, light sensors, and temperature sensors), accessing multiple networks (e.g., GPS, Wi-Fi, and Bluetooth), and location-based services (LBS) (i.e., identifying the location of a person or object, such as discovering the nearest banking cash machine automated teller machine (ATM)) (WebMapSolutions 2011). Further, with the explosive growth in the number of internet users worldwide reaching around 3.4 billion (40% of the world population) (Internet live stats 2016), mobile traffic is expected at a high acceleration of 150% per year (Meeker 2013), enabling more smartphone users to access internet services. Inevitably, these activities will be associated with personal, financial, medical, and business information that is sensitive and confidential; this means the data stored on smartphones could be more expensive than the cost of the device (Saevanee et al. 2015). As a result, smartphones should be kept secure at all times.

However, smartphones and their services and information are becoming targets of cybercrimes. For example, the UK government found that in 2012, there were, on average, over 260 mobile phones being stolen across England, Scotland, and Wales daily (BBC NEWS 2012). Alarmingly, the number of smartphones being stolen each day increased to two thousand in 2014 within the UK (Mail Online News2014). Furthermore, 35% of (CSID 2017) survey respondents' accounts or personal information were compromised or stolen by imposters. A study by Symantec Corporation (2013) showed that attacks increased by 42% with about 604,826 of accounts hacked, of which 23% was caused by theft or loss of smartphone devices. In addition, a report was published in 2016 depicting that data breach threats increase dramatically with the use of stolen, weak, and default credentials, which represented 63% of data breach corpus (Verizon 2016). Recently, a Home Office report revealed that mobile devices surged last year with over 700,000 handsets stolen. The study rated that the total number of stolen mobiles was more than double the 330,000 figure officially recorded by the police (MailOnline News, 2019). Moreover, (COMPUTERWORLD, 2019) reported that the incidence of smartphone theft has been increasing rapidly through recent years and is fast becoming an epidemic. Serious crimes in San Francisco from November to April recorded 579 thefts of mobile phones or tablets, and this represents 41 percent of all serious crimes. On some days, like Feb. 27, the only serious crimes stated in the daily police log were cell phone thefts.

As a result, it is mission-critical to secure smartphones and their services and information. To secure any system or information, it is essential that confidentiality, integrity, and availability (CIA) is achieved. Without implementing a proper authentication mechanism, it is difficult to achieve these aims. Three major approaches can be used for authentication: something the user knows (e.g., password or PIN), something the user has (e.g., token or smart card), and something the user is (i.e., biometrics). Some systems utilise a single approach to achieve authentication while others combine two or more techniques in order to strengthen the authentication (Karatzouni 2014).

Current protection mechanisms of smartphone devices are usually based on the knowledge-based authentication technique (e.g., PIN, password, and graphical password) (Meng et al. 2015). However, the main disadvantages of this approach can be summarised in the following points:

- Secrecy and public use; users log into various websites using passwords while they are in public places such as libraries or cafes. This leads to many password authentication issues. First, the password could be observed by other users by looking over the shoulder or looking at the keyboard and noting the keystrokes. Second, the password information could be intercepted by someone connected to the network while the user logs in using network programs that monitor the local Wi-Fi hotspot (itstillworks, 2019).
- User Engagement; many people use common password tropes, such as "password", "1234", or "pass" as passwords for sites they use. Also, the same password may be used for multiple sites. That means compromising one site will probably leading to compromising any other site that uses that password (itstillworks, 2019).
- Security could be easily compromised. For instance, currently, individuals have an average of 21 passwords to remember and 81% of them select common words as their passwords (Rana 2015).
- A typical user may use several devices with approximately 13 accounts with different usernames and passwords (Ghazizadeh et al. 2012). Surveys carried out by Cobb (2012) and CSID (2012) found that 46% and 61%, respectively, of their participants, used the same password for multiple

accounts to reduce the burden on memory to save and retrieve various passwords.

- Furthermore, improper use may occur if users do not use the techniques in the right way, such as never changing the PIN code, sharing it with friends, and writing it down. Indeed, 30% of users write their passwords down in an insecure manner (Rana 2015).

The weakness of point-of-entry techniques is documented extensively and is considered as a significant problem of the PIN approach. The user logs in once and gains access to all applications without the need to log in again to each of them or legitimise the user's identity again after obtaining trust and private information (Crouse et al. 2013). Additionally, according to the present NIST guidelines for mobile security, there is no set form to lock the device automatically as long as the device has been used regardless of whether that person is authorised. They suggest locking the device after staying idle for a specific time (NIST 2014). Indeed, a lot of smartphones offer this functionality now; however, it is unknown whether users take the full advantage of it.

From the above limitations, these approaches suffer from the probability of lost, stolen, guessed, shared, forgotten, misplaced, eavesdropping, and repudiation. Also, 67% of mobile users leave their devices without password protection, so their personal information could be accessed by malicious individuals (Crouse et al. 2013).

In order to solve some password problems, token-based authentication could be used. Instead of the human brain, the secret-knowledge is placed in a memory chip. But they are high in cost and the user needs to carry multiple tokens to access many services, which is considered as inconvenient. Additionally, the verification depends on the token itself rather than the individual (Clarke & Furnell 2005). Consequently, in token-based approaches, the user is primarily responsible for maintaining the security of the system (i.e., making sure the token is secure). When the token is lost or stolen, the system is compromised.

Various biometric modalities are currently used generally in order to improve system security. Additionally, biometrics have essential features that can be used to assess a specific link to the identity of the person concerned because biometrics use human physiological and behavioural characteristics to identify individuals. The main advantage of biometric-based approaches is that they cannot be easily stolen or forgotten, unlike passwords, PINs, and tokens (Rana 2015). In smartphone devices, fingerprint and face recognition are already used as an alternative authentication method in addition to the PIN. For example, Samsung Galaxy Note 8 and iPhone 8 and 8s devices provide biometric-based unlock style devices using the face and fingerprint, respectively (SKY BIOMETRY 2018).

However, the face and fingerprint are used to offer a point of entry authentication; hence, they cannot provide continuous protection for smartphones. Also, the high possibility of deceptive actions against biometric security implementations on smartphones could occur. Moreover, several factors may affect facial recognition algorithms, such as the stability of the extracted facial features over time, adjacent lighting, image resolution, face distance and position from the camera, and liveness test provisioning. Also, both fingerprints and facial recognition are intrusive and need user intervention, which could be considered inconvenient with recurrent use. Meeker (2013) pointed out that the smartphone user checks their devices on average 150 times per day. These users could spend over five minutes daily to unlock their devices (based on an average of two seconds for every single unlock process). Besides, mobile devices are still susceptible to data theft when in an unlocked state (Crouse et al. 2013). Therefore, to improve security, more convenient, secure and effective biometric modalities are needed that operate transparently for mobile authentication to minimise user inconvenience and increase user acceptance and security (Rana 2015).

Transparent authentication provides security by “authenticating the user periodically throughout the day/session/use of the device in order to maintain confidence in the identity of the user” (Clarke & Furnell 2005). Transparent and continuous authentication is considered as non-intrusive, more secure, and places less encumbrance on the user. Moreover, the decision based on multiple sources/biometric modalities can provide better confidence in the authenticity of the user (Clarke 2011). Several studies have proposed advanced authentication mechanisms that can provide transparent and continuous authentication to the user by using behavioural biometrics. According to these studies, a number of biometrics could have the probability to be used for transparent authentication on mobile devices, including keystroke dynamics (Crouse et al. 2013; Saevanee et al. 2015), behavioural profiling (Clarke 2011), 3D-facial recognition (Muaaz 2013), Voice recognition (Clarke & Furnell 2005), linguistic profiling (Saevanee 2014), and gait recognition (Mohammad Omar Derawi et al. 2010; Nickel et al. 2011). While much effort has been expanded on conducting and implementing the existing behavioural biometric approaches, less focus has been given by researchers into using a smartphone device to collect realistic data for gait and activity recognition.

Gait recognition recognises a person by how they walk. Many studies in psychology, medicine (kale, 2003), and biometrics indicate that human gait is unique for every person, as well as non-invasive techniques that can be used for identification and verification purposes (Gafurov, 2008; Mäntyjärvi et al., 2005). As a result, this approach has an excellent opportunity to be implemented in a continuous authentication manner (rather than user re-authentication), thus, decreasing the burden on the user and increasing the security. Currently, most smartphones and portable devices have built-in sensors (e.g., accelerometers) that can be used to record the user's gait. Therefore, there is no need to attach further hardware to collect gait features (Rana 2015). By using gait recognition, the user does not need an explicit action for mobile authentication because the related data is continuously recorded while the person is walking (Derawi et al., 2010; Clarke, 2011; Zhong and Deng, 2014). During times when he is not walking, other biometric modalities can be used (Derawi et al.2010).

Moreover, gait recognition can be seen as an advantageous biometric identification technique for the following two reasons: (1) user-friendliness, because the gait of a person can be captured unobtrusively and continuously; and (2) security, because of the fact that the gait of an individual is challenging to mimic (Hoang et al. 2015). However, there are several challenges related to personal identification via gait recognition. Gait will be affected by several situations (1) stimulants, like drugs and alcohol; (2) physical changes, for example pregnancy, an accident or disease affecting a leg or foot, or severe weight gain/loss; (3) psychological changes, where the mood of a person influences his/her gait; (4) clothing, in particular, shoes (Derawi 2012); and the condition of the road surface (e.g., grass or concrete). Also, very few studies have used actual commercial smartphone devices to collect realistic data for gait and activity recognition. In addition, both the number of participants and the amount of data used in existing studies are somewhat limited. The use of real-world data is likely to result in far higher variability in the gait signature because of the variety of situations in which a person might find themselves in (e.g., in a rush to a meeting, carrying luggage to an airport, running because of poor weather, exercising, to name but a few). As such, envisaging the context within which the user finds themselves will be an important factor to take into consideration in order to achieve good recognition performance in practice.

1.2 Research Aims and Objectives

The main aim of this research was to develop a Context Awareness Gait Recognition model that could adapt to different circumstances (e.g., changes in shoe, stress, or carrying a bag). To achieve this, the following research objectives were established:

- To review the current state-of-the-art literature in gait authentication including mobile-based gait authentication.
- To review the biometric authentication techniques including their application in the current research on continuous and transparent authentication systems (TAS).
- The study sought to investigate the performance of gait recognition across a wider range of activities and participants.
- To experimentally investigate the nature of gait features under more realistic real-world scenarios and understand how well existing approaches would work.
- To evaluate the developed system in order to determine the usability, functionality, and appropriateness of the approach.

1.3 Thesis Structure

This thesis is organised into eight chapters. In addition to Chapter One;

Chapter Two reviews the biometric system from many perspectives, including its system components, requirements, techniques, performance measures, and standards for physical and behavioural biometrics with a view of examining its potential to be incorporated in the continuous and transparent authentication proposal.

Chapter Three provides a comprehensive literature review of the existing research on mobile-based gait authentication. The chapter concludes with a discussion section that scientifically identifies the gap that exists in the literature.

Chapter Four represents the data gathering and methodologies that were used to collect and categorise data and the method of preparing the data, the devices, and the software that were employed. The chapter then proceeds to describe the pre-processing, time, and frequency domains feature vector extraction and effective selection feature technique to support the experiments mentioned below.

Chapter Five provides a comprehensive evaluation at gait recognition, including an analysis of motion sensors (i.e., accelerometers and gyroscopes), an investigation and analysis of features, and an understanding of the variability of feature vectors during differing activities across a multi-day collection. Furthermore, it explores the impact of dynamic feature selection for each user to investigate their efficiency to reduce the feature vector size and enhance performance. Moreover, it implements the proposed multi-algorithmic approach and compares its performance with single algorithmic approach (i.e., a dataset treated as one activity).

Chapter Six builds upon the knowledge of Chapter Five to present a novel real-world gait authentication approach that manages the main research gap. This chapter will focus on providing the empirical basis for whether the proposed approach could work – initially through exploring smartphone-based real-world data (rather than highly constrained control data) to understand the variability and difficulty in successfully authenticating individuals.

Chapter Seven discusses the main contributions of this study by comparing the research achievement with relevant studies that employed the mobile-based gait authentication. It also defines the development plan of the proposed context-awareness gait authentication model, including the processes of modelling.

Chapter Eight is the final chapter presents the conclusions accomplished from the research and highlights the key achievements and limitations. It also contains a recommendation on future research and development of this research.

2 Biometrics Authentication

2.1 Introduction

Authentication is a cornerstone of information systems security and authorisation is a process of identifying legitimate users by an effective user authentication technique to prevent unauthorised access to personal or sensitive data. All approaches for human authentication rely on at least one of the following: something the person knows (secret knowledge-based approach, e.g., a password/PIN), something the person has (e.g., a smart card), or something the person is (i.e., biometrics) (Hocking 2014). In the first and second authentication approaches, maintaining the security of the system is dependent on the user. Hence, a lost or stolen token or shared password will compromise the system. Moreover, they have several vulnerabilities (Al Abdulwahid 2015), the (password/PIN and smart card) techniques suffer from many disadvantages, as mentioned previously. Therefore, the operational performance being achieved is highly correlated to the biometric software, which adds another level of security. But, it is not guaranteed that it is impervious to compromise (Clarke 2011).

The chapter presents background information about typical biometric system components and performance metrics used to evaluate such a system, based upon their physiological or behavioural characteristics. It also provides an overview of existing authentication approaches and devices to explore whether they solve some issues related to the research area. Finally, this chapter highlights some of the applicability of the biometric techniques in order to operate transparently.

2.1.1 History of biometrics

The word biometric comes from the Greek word (bio), which means life, combined with metrics, which means measures (McCabe 2005). The first signs of biometrics emerged in 500 B.C. when cavemen used their fingerprints to symbol their drawings (Babich, 2012). Babylonians behaved in the same way to sign business deals, which existed on clay tablets. Ancient Egypt is the birthplace of the first evidence of using biometric authentication ever seen by archaeologists. In that time, in order to summarise the process of providing food, the supervisor of the workers would record information about them, including their name, age, work unit, position, and occupation. Moreover, to avoid cheating, he was enforced to record more individual information/characteristics, such as physical and behavioural ones (Babich, 2012, Page 3).

Early biometric characteristics were simple; one of these is biometrics, inked paper allowed to yield palm prints that can acquiesce from inked paper, while children could be distinguished from each other by their footprints. According to the National Science and Technology Council (NSTC) (2006), in a study on the history of biometrics, the period between the end of the 19th century and the beginning of the 20th saw an acceleration in the use of fingerprint authentication. For instance, in 1892, Galton developed a classification system for fingerprints and, in 1896, Henry developed a fingerprint classification system. Then, in 1903, NY State Prisons began the use of fingerprints. More recently, the first model of acoustic speech production was created in 1960, when researchers drew attention to behavioural biometrics and then the behavioural components of speech were modelled for the first time in 1970. Also, a study on the compatibility of biometrics and a machine-readable travel document was launched in 1999. In the same year, the Integrated Automated Fingerprint Identification System's (FBI's IAFIS) major components became operational (NSTC, 2006).

The International Standard Organisation (ISO) and the International Electro technical Commission (IEC) (ISO/IEC 2012) defined biometrics as encompassing "counting, measuring and statistical analysis of any kind of data in the biological sciences including the relevant a biometric system that provides biometric technology using components from multiple vendors". They also go further to describe biometric characteristics as the distinguishable, repeatable biometric features that can be extracted from an individual for the purpose of identification or verification. This characteristic can be either physiological or behavioural and can be achieved from any part of the individual. ISO-based biometrics should have a high level of performance of data interchange amongst applications and systems, which is an essential characteristic for implementing biometric systems (interoperability). Also, the dependability of utilising biometrics that support frustrate the spoofing and avoidance risks (reliability), alongside with the user-friendliness (usability) and security for future standards-based systems and applications. With better interoperability between biometrics systems, the success of these applications would be much more similar.

2.1.2 Biometric system requirement

Many essential biometric requirements are needed in order to select the best authentication approach to utilise. The suitability of the biometric authentication technique is specified according to the availability of the following requirements on the associated trait, as suggested by Jain et al. (2002):

- Universality: which means that each person utilising the application should have the chosen biometric feature. For example, as all users have fingers, it is possible to use the fingerprint as a biometric identifier.

- Uniqueness: to distinguish people from one another, the specified trait should be befittingly different for persons' relative application environment (e.g., the iris is much more unique than the fingerprint).

- Permanence: shows the constancy of a biometric characteristic over time. For example, while an individual's retina remains stable for the entire life, people's keystroke behaviour varies because of many factors, such as device, mood, and text familiarity.

- Measurability: the ease of collection of a particular biometric trait by employing an appropriate device and how easy it is to extract the feature set from raw traits. For example, the retina needs a specific device and explicit user interactions. In comparison, a person's walk can be collected unobtrusively and easily using standard devices.

- Performance: refers to the recognition accuracy, robustness, and speed, in addition to the appropriateness of the resources used to achieve that accuracy. For instance, individual retina screening is considered constant compared to the keystroke, which can differ because of the device, mode, or text experience.

- Acceptability: this specifies how people are interested in using biometrics as an authentication method in their lives. For example, confidentiality and suitability. Otherwise, they will avoid using it.

- Circumvention: The possibility of imitating a trait and the degree of its vulnerability. For example, the iris scan is almost impossible to imitate and mimic in comparison to behaviour-based biometrics (e.g., keystroke dynamics).

As a result, a perfect biometric authentication system should meet all the requirements mentioned above. However, Jain et al. (2008) claimed there is no biometrics that will fit all the above seven characteristics. In practice, these requirements are varied depending on the specific needs and security of the application.

2.1.3 Verification and Identification

A biometric system could work in two modes, namely verification and identification. Verification is defined briefly by (Clarke 2011): "determining whether a person is who they claim to be". Verification is also referred to as a one-to-one matching. The current captured biometric sample(s) of the claimed person compared with the stored template of the registered person. For example, an individual could access to his/her bank account at an ATM by using an iris scan or scanning a finger to confirm his/her work daily attendance (Jain et al., 2008). Both Jain et al. (2008) and Clarke (2011) pointed out that the biometric verification system is considered more reliable than the traditional systems that use token-based (e.g., ID card) and knowledge-based (e.g., password or PIN).

In contrast, in identification mode, the comparison is one-to-many (i.e., explore whether the identity exists in the database). The current person's biometric sample should be compared with all templates that are stored on the system database to decide if a match exists. Therefore, because of these additional complexities and computation, more time will be needed for the identification mode. It is clear that identification requires a higher level of the system's accuracy and feature uniqueness than verification. As physiological biometrics (e.g., fingerprint, facial recognition, iris and retina scan, and hand geometry) are more unique than behavioural biometrics (e.g., voice, gait recognition and keystroke dynamics), they tend to be more appropriate for identification.

2.1.4 Components of the Biometric System

To complete a biometric process, there are five incorporated components declared in Figure 2-1 (Clarke 2011):

- Sample Capturing: This is the stage of collecting the biometric sample from the genuine user utilising an appropriate capture device or method according to the biometric system (e.g., optical finger scanner for fingerprint recognition, webcam or mobile front camera for facial recognition, and mobile accelerometer sensor for gait recognition).

- Feature Extraction: In this stage, distinctive features of the captured sample(s) are processed to generate a feature extraction template. For instance, after a gait signal is captured, many algorithms are executed to extract many unique features, like average resultant acceleration, binned distribution, and time between peaks for gait recognition.

- Storage: the feature vector (reference template) that resulted from the feature extraction process is stored in the database. This stored template is used as a reference in the matching process.

- Classification (matching): In the comparison phase, the individual's current sample (probe template) is compared with the reference template taken at the enrolment phase. Consequently, a match score is given, indicating the degree of similarity.

- Decision: in this stage, access is permitted or denied according to the comparable score value, which should be equal or more than the previously identified threshold; otherwise, access will be denied.

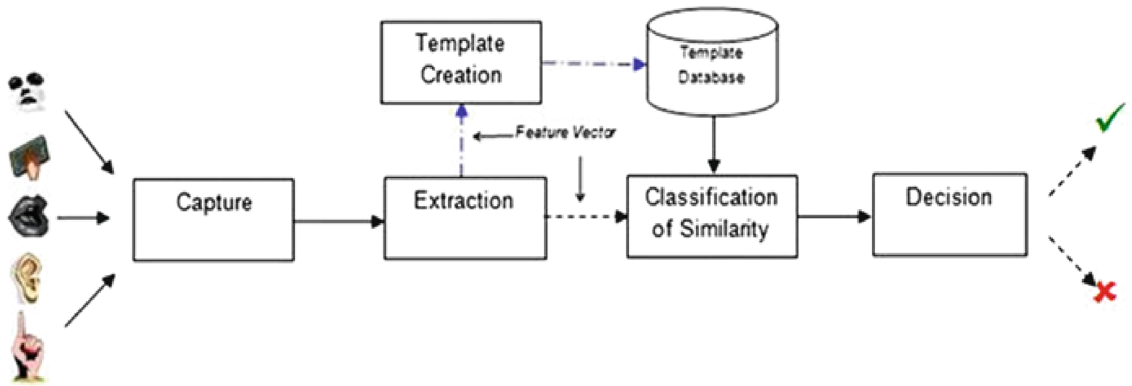


Figure 2-1: The Components of a Biometrics System (Clarke 2011)

2.2 Classification Approaches

This section provides a high-level description of some of the popular classification algorithms. There are two types of classification: Statistical modelling and Machine Learning. Firstly, a statistical model is a family of probability distributions. Basically, Statistical models use mathematical equations (Analytics Vidhya, 2015) by a formalisation of relationships between variables in the form of mathematical equations in order to find the relationship between variables to predict an outcome and applied for smaller data. On the other hand, machine learning is an application of artificial intelligence by learning from data without relying on explicitly programmed instructions (Jordan and Mitchell, 2015; Statistical Models, 2019).

2.2.1 Statistical

2.2.1.1 Dynamic Time Warning (DTW) Distances

Dynamic time warping (DTW) has been widely used for computing similarities between two temporal sequences in time series analysis even with various speeds. For example, similarities in walking could be identified using DTW, even if one person was walking faster than the other was or if there were accelerations and decelerations during the course of observation (Anantasech and Ratanamahatana, 2019).

2.2.2 Machine learning

The main objective of machine learning is to create systems that are able to learn automatically (Henrique, Sobreiro and Kimura, 2019). More specifically, machine learning teaches computers to do what comes naturally to humans by learning from experience and using computational methods to "learn" information directly from data without relying on a predetermined equation as a model. However, the types of machine learning algorithms might differ in their approach, the type of data they input and output, and the type of task or problem that they are intended to solve. Machine learning can be divided into two subdomains: supervised learning and unsupervised learning, as shown in Figure 2-2. Supervised learning requires training with labelled data, which has inputs and desired outputs. There are two types of supervised learning, namely: classification (discrete output variable) and regression (continuous output variable). On the other hand, with unsupervised learning, there is no need for labelled training data and inputs are provided without desired targets, such as the clustering approach, by allocating to groups without class information (Qiu *et al.*, 2016). In the literature, support-vector machines (SVM) and neural networks are considered as the most commonly used models for prediction (Henrique, Sobreiro and Kimura, 2019).

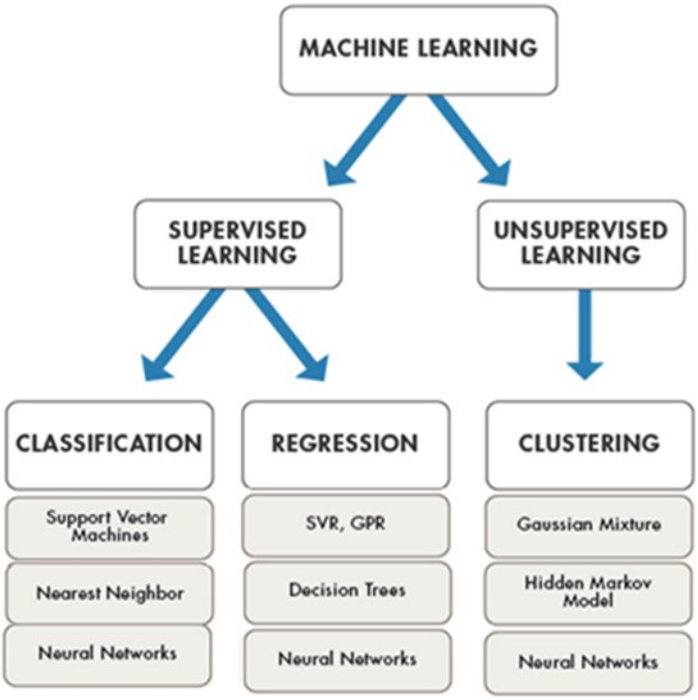


Figure 2-2: Classification approaches Taxonomy (Technology at Nineleaps, 2019)

2.2.2.1 Support-vector machines (SVM)

Support-vector machines are supervised machine-learning algorithms that can be used with learning algorithms by analysing the data used in order to solve classification and regression problems (Tavara, 2019). In SVM, there are two phases: training and testing.

SVM is training by specifying a set of training examples to one or the other of two classes, and an SVM training algorithm builds a model that allocates new samples to one category. These algorithms can efficiently perform a linear and non-linear classification. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces. The kernel methods are a sort of algorithm for pattern analysis, known as the best member, the support vector machine (SVM) (Barber, 2012; Technology at Nineleaps, 2019). After the engine is trained, the SVM model predicts which class label a new unseen test sample should have in the testing phase (Tavara, 2019).

2.2.2.2 A classification tree or a decision tree

For building classification models in the real world, the decision tree is one of the more widely used methods because of its simplicity and ease of interpretation (Kim, 2016). Each interior node corresponds to one of the input variables and is split into child nodes based on the values of the input variable. Each leaf or terminal node represents the particular value of a target variable—for example, the specific class of a categorical variable for the classification problem and the specific real value of a continuous variable for regression problems. During the classification tree learning process, samples at each interior node are split into subsets based on an attribute, and this process is repeated on each derived subset in a recursive manner called "recursive partitioning". The recursion is finished when a subset at a node has the same target value, when splitting does not improve prediction, or when splitting is impossible because of user-defined constraints (Kim, 2016). Generally, decision trees are used in operations research or statistical probability analysis, especially in decision analysis, to help identify a most probable strategy to reach a goal, but are also a popular tool in supervised machine learning (Pao, 2005).

2.2.2.3 A hidden Markov model (HMM)

A Markov chain is a stochastic model explaining a sequence of probable situations in which the possibility of each situation depends only on the state achieved in the prior situation. In a hidden Markov model, there are unobserved or "hidden" states while all states are apparent to the observer in a standard Markov chain. In comparison with the Markov chain, the hidden Markov model aims to predict the future state of the variable utilising probabilities based on the present and previous state. The variability between a Markov chain and the hidden Markov model is that the state in the final is not directly noticeable to an observer, even though the output is. Many machine learning and data mining tasks have been effectively applied to problems including speech, handwriting, optical character and gesture recognition (Franzese and Iuliano, 2019; Techopedia, 2019).

2.2.2.4 Nearest Neighbour

The nearest neighbour (KNN) algorithm is the simplest classification algorithm and one of the popular learning algorithms (Altman, 1992). KNN is a non-parametric, lazy learning algorithm. Non-parametric means it does not make any expectations on the fundamental data distribution and the data usually used to structure the model. The KNN algorithm stores all presented cases and classification procedures based on a similarity measure (e.g., distance functions). The distance function is used by the distance metric, which provides a relationship metric between each element in the dataset. It should be suitable with real-world data when the data mostly does not obey the classic theoretical expectations made (e.g., linear regression models). Therefore, KNN has been used in statistical estimation and pattern recognition and could be more suitable when there is limited or no preceding knowledge about the distribution data.

2.2.2.5 Neural Network multilayer perceptron (NN-MLP)

Artificial neural networks (ANN) are based on a collection of connected units or nodes called artificial neurons. Each connection can transfer a signal from one artificial neuron to another. In addition, there is pattern recognition, feature mapping, clustering, and classification examples of applications of neural networks (Han *et al.*, 2016; Techopedia, 2019). In this approach, the neural network consists of units (neurons) arranged in layers, which convert an input vector into some output. Each unit takes an input, applies an (often nonlinear) function to it, and then authorises the output on the next layer. Mostly, the networks are identified to be feed-forward: a unit supplies its output to all the units on the next layer; however, no feedback will be transmitted to the previous layer. Weightings are implemented to the signals feed-forwarding among layers, and it is the same weightings that are matched in the training stage to adapt a neural network to the particular problem at hand. This is the learning phase. Multilayer perceptron (MLP) is the most used model in neural network applications using the back-propagation training algorithm (Ramchoun *et al.*, 2016). In this approach, the neural network creates a set of outputs from a set of inputs. An MLP is categorised by several layers of input nodes linked as a directed graph between the input and output layers. As there are multiple layers of neurons, MLP is a deep learning technique (Data Science Bootcamp, 2019).

2.3 Biometric system performance measurement factors

Having highlighted that, all biometrics work is based on the result of comparing the individual's current sample (probe template) and the reference template. Two essential error rates reflect the performance of the template matching process: the false acceptance rate or false match rate (FAR or FMR) and the false rejection rate or false non-match rate (FRR or FNMR). Woodward (2003) identifies these error rates as follows:

- FAR: It measures the percentage of biometric technique errors when the imposter is falsely accepted.

- FRR: It measures the rate of biometric technique errors when genuine individuals are incorrectly rejected.

The FAR and FRR are calculated as (Miguel & Neves 2013):

$$\mathrm{FAR} = \frac{\text{accepted imposter attempts}}{\text{total imposter users attempts}} * 100\%$$

$$\mathrm{FAR} = \frac{\text{rejected genuine attempts}}{\text{total genuine users attempts}} * 100\%$$

As highlighted by Clarke and Furnell (2005), in case of legalising a person's identity, it is an unlikely situation to get a perfect 100% match between two samples of an individual's biometric trait because of various issues, such as environmental noise and trait variability. As a result, the security level of a biometric system is based on a pre-set threshold value for the biometric system, which controls the acceptable degree of similarity. The system designer should balance the security of biometric systems and their user's suitability by setting the threshold tightness. As shown in Figure 2-3 , these two-performance metrics (i.e., FAR and FRR) are inversely proportional: as one rate decreases the other increases. A system with tight security can be achieved by increasing the threshold value, which may result in more genuine users being denied access (i.e., high FRR); also increasing the protection will minimise the potential of obtaining access for unauthorised users (i.e., low FAR). In addition to FAR and FRR metrics, a third error rate named the equal error rate (EER) is a measure of where the FAR and FRR curves intersect (i.e., FAR equals FRR) and is frequently used to evaluate and compare the performance of biometric systems (Clarke & Furnell 2005; Jain et al. 2002).

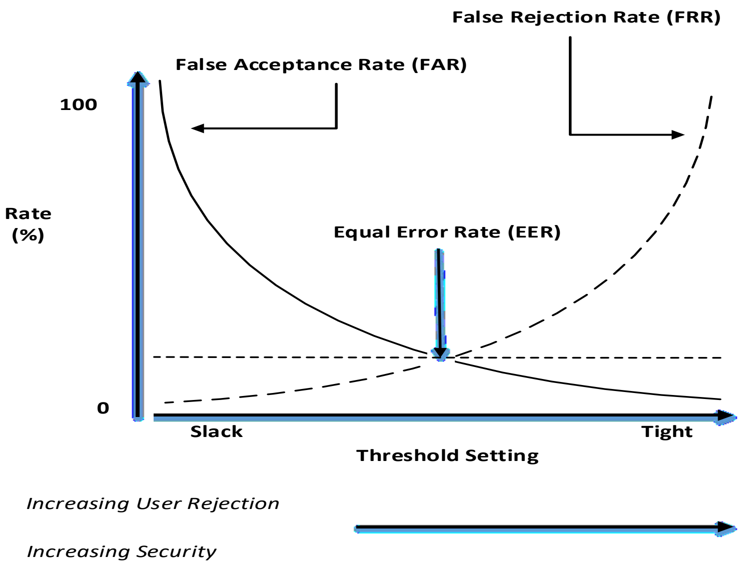


Figure 2-3: Biometrics Performance Metrics Factors (Clarke 2011)

In addition to FAR, FRR, and EER, other metrics are frequently used when testing and evaluating biometric systems. For example, Clarke (2011) defined two rates as follows: The failure to acquire (FTA) represents the rate at which the creation of a valid template is incapable in the capture or extraction stage; on the other hand, the failure to enrol (FTE) effectively means the rate at which the user cannot enter into the system. They measure the error rates that probably happen during the enrolment stage. It usually results when there are inappropriate user features and samples to be used to create a template. For instance, when the system is unable to capture the user's sample(s) affected by an equipment problem.

2.4 Biometric Techniques

Biometric techniques are classified into two main groups based on the environment of the deployed discriminative attribute. The physiological and behavioural details of these two types are described below.

2.4.1 Physiological Biometrics

Physiological biometric approaches aim at distinguishing an individual based on specific physical characteristics, such as the fingerprint and the face, which tend to be invariant and thus applicable to be utilised for both identification and verification (NSTC, 2006).

2.4.1.1 Fingerprint Recognition

Fingerprint identification is the oldest and most widespread, well known, deployed and used a biometric feature for authentication on many systems, such as securing laptops and mobile phones (Clarke 2011). It refers to the automated process of identifying or confirming identity-based on the comparison of two fingerprints (as shown in Figure 2-4). The reasons for it being so popular are the ease of achievement, conventional use and acceptance when compared to other biometrics, and the fact that there are numerous sources (i.e., ten fingers) for biometric data.

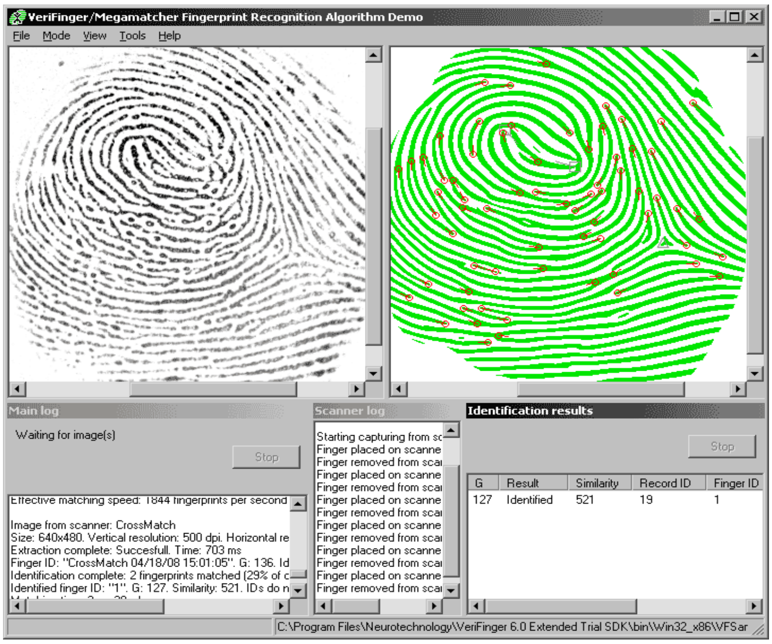


Figure 2-4: An example of fingerprint recognition (NEURO technology 2015)

There are three basic outlines of fingerprint points, which are (BiometricSolutions 2015):

- The arch: a pattern where the ridge reaches one side of the finger, then increases in the centre, creating an arch, and exits on the other side of the finger.
- The loop: Loops are a highly familiar pattern in fingerprints; the ridge arrives on one side of the finger, then creates a curve and exits on the same side of the finger from which it entered.
- The whorl: the pattern when ridges form a circle around a central point.

Minutiae mean specific points in a fingerprint, and it is the slight details in a fingerprint that is of the highest importance for fingerprint recognition (biometric-solutions.com, 2015).

There are four main types of fingerprint reader hardware:

- Optical reader: is a digital camera that obtains a visual image of the fingerprint. They start at low prices, but dirty or marked fingers impact the readings. This type of reader is easier to fool than other types.
- Capacitive reader (CMOS readers): it uses an electrical current to form an image of the fingerprint; they are more expensive than optical readers. A significant advantage, they require a real fingerprint shape rather than only a visual image. This makes CMOS readers harder to trick.
- Ultrasound readers: they use high-frequency sound effects to access the outer layer of the skin. They read on the dermal skin layer, which removes the need for a clean, unharmed surface. All other readers acquire an image

of the outer surface, therefore requiring hands to be cleaned and free of scars before read-out.

- Thermal readers' sensor: on a contacting exterior, there is a variance of temperature between fingerprint ridges and valleys. Thermal fingerprint readers have disadvantages such as higher power consumption and performance that depends on the environmental temperature.

After developing a fingerprint image using the reader hardware, it must be interpreted. Then it is handled in such a way that read-outs can be efficiently related and matched against each other. Generally, one of the three matching classification approaches is utilised: minutiae matching depends on recognition of the minutiae points, this is the most widely used method; ridge-based-matching depends on the number of ridges instead of minutiae points; the correlation-based approach (pattern matching) compares merely two images to see how related they are. It is often used in fingerprint systems to detect duplicates (BiometricSolutions 2015).

All studies asserted that the fingerprint is more secure than typing a PIN (Ferrero et al. 2015). Besides, the fingerprint is robust, unique, and only needs a short time for enrolment with a fingerprint scanning system. Furthermore, it is generally accepted as technology as most people are familiar with the use of the fingerprint for identification purposes. However, fingerprint systems do suffer from many problems, such as injury (whether temporary or permanent), dirtiness, and the poor quality of the finger samples of some people. Moreover, as mentioned previously, there is the possibility of using fake fingers (e.g., silicon or jelly fingers). Fingerprint recognition systems may suffer other problems; for example, fingerprint readers might undergo wear and tear effects over time. As a result, this would weaken the efficiency as error rates go higher and thus increase user inconvenience. In addition, from the user acceptance point, fingerprint scanning mostly resembles the impersonal and non-intrusive nature of passwords and PINs.

2.4.1.2 Palm print and Hand Geometry

Palm prints were used in 1858 physically with ink on employment agreements in India (NSTC, 2006). It was not automated until 1994. However, regardless of the early robotized usage (i.e., personality) in the mid-1970s and other progressive licenses, hand geometry frameworks must be used to check not to distinguish clients because these qualities are not extremely particular. Therefore, hand geometry-based systems usually are used in authentication systems rather than identification.

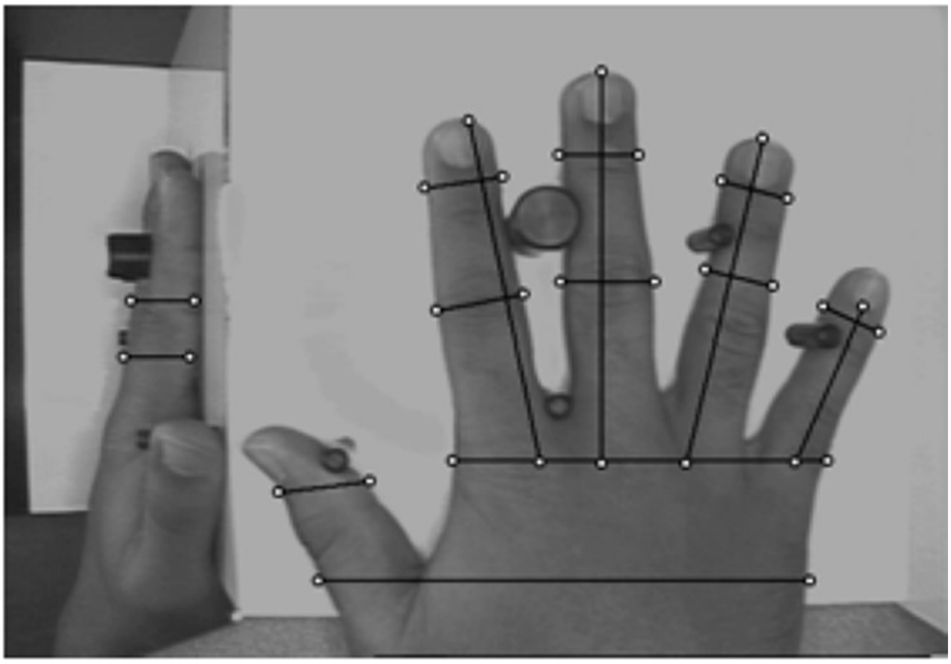


Figure 2-5: Example Distance Measurements Source: (NSTC, 2006)

The palm print recognition method identifies individuals based on the unique features of their palm as shown in Figure 2-5. Hand geometry measures a portion of the hand attributes yet from the external surface, specifically, length, width, thickness, and surface range of the back of the hand and four fingers. It shares the comparison criteria with a fingerprint recognition system, such as size, as well as ridges and the minutiae feature of the palm. In this way, it can be used in verification and identification modes. Be that as it may, it makes them deficient notwithstanding those of the unique mark strategies; for example, the large capturing machine, the generally bigger format size contrasted with finger impression, and the likelihood of palms' geometric elements changing because of maturing or weight (Jain et al. 2005).

2.4.1.3 Facial Recognition

Perceiving and recognising known individuals in light of their faces has been used since the beginning of creating people; however, it was carried out after the first of semi-automated facial recognition systems was evolved in the 1960s. The development of its classification systems has been investigated, developed, and approved gradually to be used in different fields, from visa distinguishing proof and observation applications to physical/virtual access control, to all the more as of late smartphone confirmation. It is viewed as the second biometric after fingerprint concerning users' acceptance and the sale rate (Biometric Institute 2013). This is attributed to its possibility to be used straightforwardly (i.e., without collaboration or association of the client) and using standard cameras (e.g., webcam as the catching sensor). The features depended are general measurements of the eyes, nose, mouth, ears, cheekbones, and separation between most or every one of them takes into account diverse restrictive calculations (as shown in Figure 2-6). The viability of such calculations changes contingent on a few variables: the dependability of the removed face highlights after some time, encompassing enlightenment, picture determination, face separation and position from the camera, and liveness test provisioning.

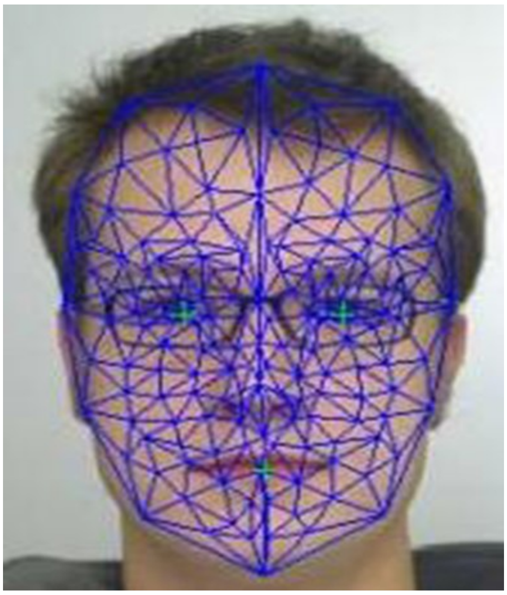


Figure 2-6: An example of facial recognition (GreekoSystem 2011)

Different answers to control some of these components have been proposed. Utilising a three-dimensional picture may help with alleviating the impacts of face introduction and lighting conditions, although the requirement for 3D camera/sensor would obstruct its acknowledgement and proliferation, as they tend to be more costly and slower. Moreover, a more complex composite model in which some of the client's face pictures in various sizes, brightening and dimensions are by and large put away as a format when an example is taken then the latter will be compared with the stored template. However, the balance between ease of use and security is an issue, as there is the probability of declining a genuine user and an increased risk of accepting an imposter (Biometric Institute 2013).

2.4.1.4 Iris Recognition and Retina Recognition

In biometrics, iris and retinal are known as "visual-based" advanced distinctive patterns, which means they depend on unique physiological qualities of the eye to recognise a person (as shown in Figure 2-7). Despite the fact that they both share part of the eye for ID purposes, these biometric modalities are distinctive in the way they work. The concept of using the iris pattern for identification was proposed in 1938 and a patent stating that the iris can be used for identification was awarded in 1978. After several years, many studies on automated iris recognition systems have been developed, but John Daugman applied the most successful patented algorithms that can implement iris recognition automatically in 1994 (Daugman 2003).

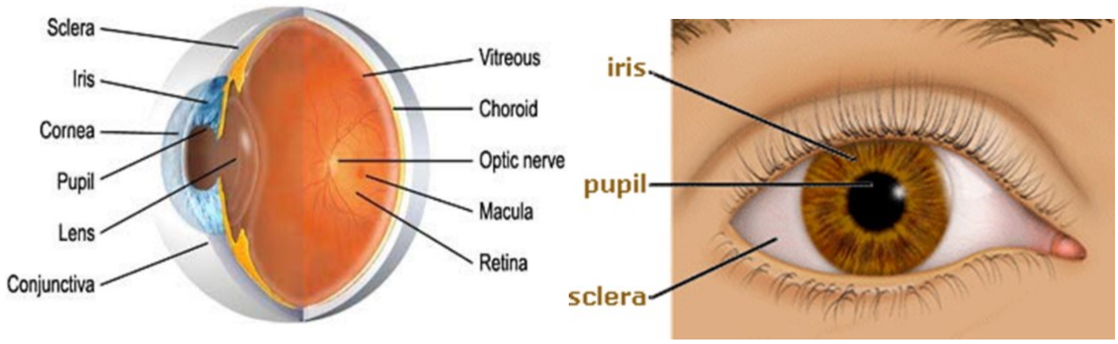


Figure 2-7: An example of a human eye (NSTC 2006; Monitgomery, 2014).

Iris acknowledgement is based on recognising people using their irises, which is the round hued tissue framed with numerous wrinkles and edges and surrounding the pupil of the eye. It trusted as the most accurate biometric strategy because of the exceptionally particular intricate iris designs that are steady for the duration of life and accessible in all healthy individuals. As a result, it can be used to confirm and distinguish people.

The human retina is coloured tissue made out of neural cells that are situated in the back segment of the eye. In view of the perplexing structure of the vessels that supply the retina with blood, every individual's retina is exceptional (Clarke 2011). The system of veins in the retina is complex to the point that even indistinguishable twins do not share a comparative example. Albeit retinal patterns might change in instances of diabetes, glaucoma, or retinal degenerative issue, the retina ordinarily stays unaltered from birth until death (Hocking 2014).

The retina recognition process reads the distinctive pattern of veins in the back of the eye using an infrared camera for a light at a nearby separation. The mastery level of the example is high, as catching it is not visibly accessible without special devices and client collaboration. Thus, it is viewed as exceptionally meddlesome leading to reducing the applicability areas and henceforth low selection. Besides, guiding the infrared wave to such a delicate organ, the eye, may raise some sound issues, which may cause clients' reluctance to acknowledge being presented to it. Despite that, it has been utilised widely in high-security areas (e.g., military buildings) for physical access control. Since 2012, the UK Border Agency has been using the Iris Recognition Immigration System (IRIS) in several airports around the UK for authenticating passengers (UKBA 2011).

However, practically, because of the fact that this particular modality requires the user to directly and intrusively align their eyes with the camera, it is classified as an intrusive approach to be utilised in continuous and transparent systems. Moreover, broadly speaking, the importance of the eye for humans makes the idea of using the iris for authentication purposes not comfortable for some people.

2.4.1.5 Ear Geometry

The ear geometry apperception technique distinguishes individuals predicated on the unique structural pattern of their auditory perceivers, including the concha, helix, antihelix, and other discriminative features (Ross 2011) (as shown in Figure 2-8). It has been evidenced that the auditory perceiver's unique characteristics are relatively stable throughout the life span, unlike those of the face that have salient effects of ageing. This approach offers a high level of flexibility, which

Fahmi et al. (2012) encouraged employing for transparent user authentication with mobile devices. During a telephone call time, they captured a series of ear images to be used for authentication purposes. The more perfect images were captured, the more succeeded authentications achieved.

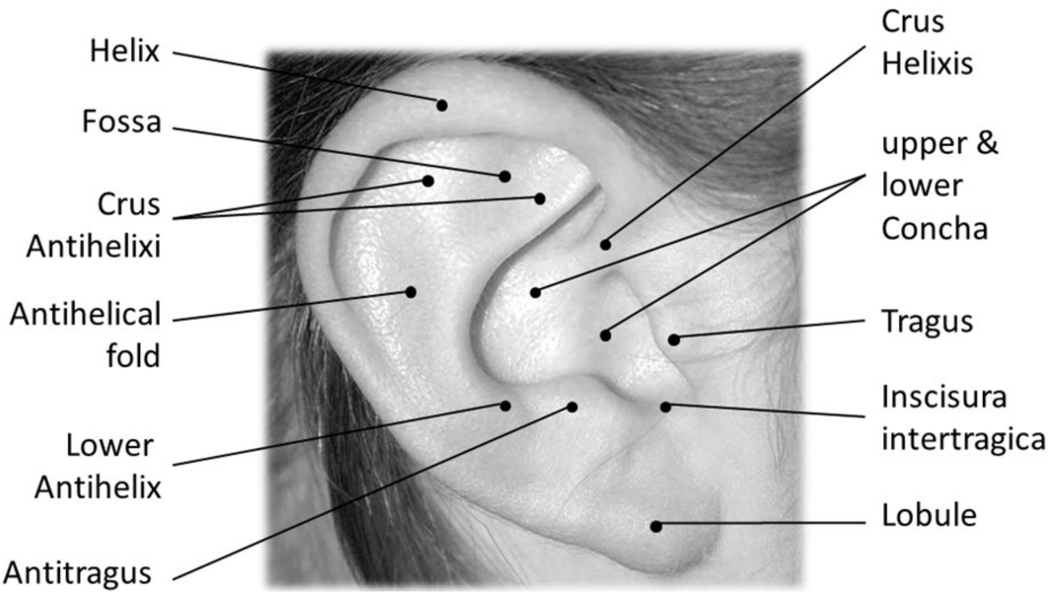


Figure 2-8: Images of Human Ears (da/sec. 2015)

2.4.2 Behavioural Biometrics

Behavioural biometrics modalities are based on people's psychological attributes because of their variability over time. This includes voice, signature, keystroke dynamics, and hand-written signatures (Gamboa & Fred 2004). Regardless of the lower level of uniqueness and perpetual quality created by behavioural characteristics, for example, evolving state of mind, well-being, and environment, they tend to be more general, straightforward, and usable than physiological ones (Clarke 2011).

2.4.2.1 Voice Recognition

The use of voice recognition is increasing with the sensor modules that are used for verification in mobile devices because of their mobility, miniaturisation, decreasing price, and increasing computational power. Voice recognition is divided into two modes: obliged (content ward) and unconstrained (content autonomous). Gunnar Fant, in 1960, spearheaded an x-beams based model for the acoustics of discourse creation. From that point forward, numerous related explorations have been carried out (e.g., the NIST Speech Group has been established, multiple relevant licenses have been issued, numerous exploratory and assessment studies have been conducted to upgrade the voice recognition frameworks, and numerous exploratory and assessment studies have been conducted to improve the voice recognition frameworks).

The voice recognition system digitises the voice and segments them into frames from the vocal signal frequencies. These vocal signals need to be of good quality for successful comparison in the future because vocal sounds can be dynamic (Marzotti and Nardini, 2006). The voice pattern is obtained from the organs that enhance speech. These include the laryngeal pharynx (below the epiglottis), the oral pharynx (behind the tongue, between the epiglottis and vellum), the oral cavity (forward of the velum and bounded by the lips, tongue, and palate), the nasal pharynx (above the velum, at the rear of the nasal cavity), and the nasal cavity (above the palate and extending from the pharynx to the nostrils) (Saquib et al., 2011).

There is an advantage inherent in the use of voice for user authentication; it is non-intrusive and is among one of the most used methods. It can be used in any smart mobile phone, and it can be carried out via the internet too, which gives it an advantage over others for over-the-phone verification from the other end of the communication channel (Kounoudes et al., 2006) as shown in Figure 2-9. Usually, the performance of voiceprint recognition is influenced by the quality of the available hardware on the smartphones the user usually interacts with. Additional to the aforementioned and ease of use, the speech recognition techniques embedded in modern smartphones (e.g., Siri on the Apple iPhone) offer a promising future for voice recognition (Huntington 2012). However, some issues can affect the quality of the sample (i.e., pattern) of people's voice, such as their emotional state, health conditions, and possible background noise. Consequently, it used in verification rather than identification purposes. Nevertheless, it may be an efficient, transparent, authentication technique.

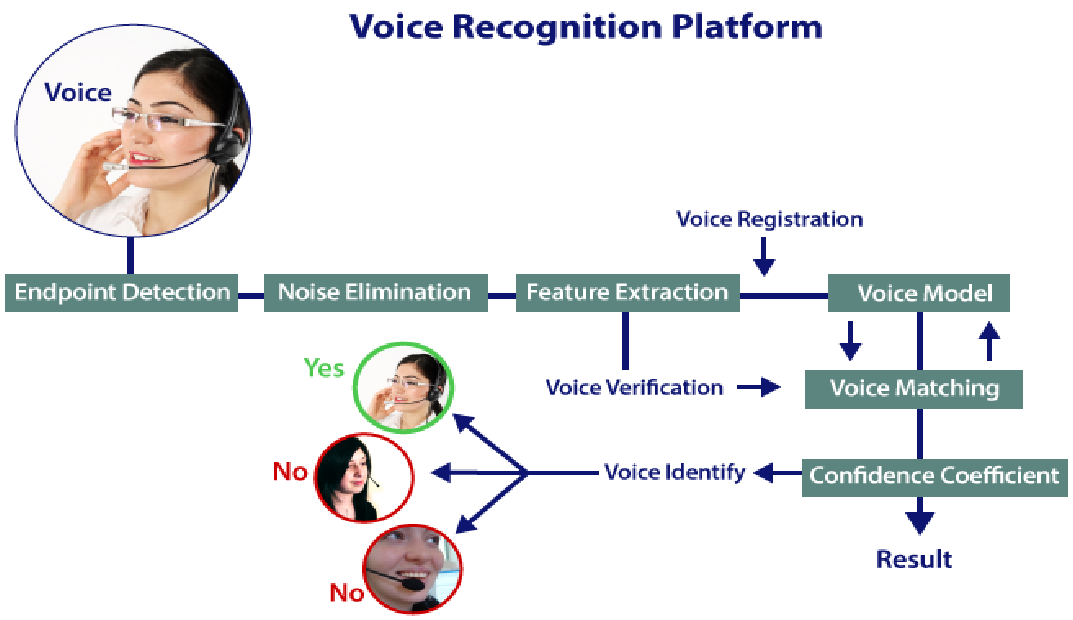


Figure 2-9: Voice Recognition Platform (BiometricSolutions 2015)

2.4.2.2 Signature Recognition

Signatures have been in use for decades now and many approaches have been developed over time on the use of this method of verification (Plamondon, 1994). This method is considered as an attribute of an individual, which has been developed over a long period. Thus, handwritten signatures have been used commonly for many laws, official business, and financial transactions for identity authentication (e.g., signing a contract).

Osborn published the first acknowledgement of its potential in verification in 1929 and then it evolved from being purely manual, using pen and paper, to a digitised recognition system in the 1980s. Subsequently, the proliferation of touchscreen devices has led to applying the similar individuality to handwriting verification as shown in Figure 2-10. Verifying the authenticity of the handwritten signature and handwriting can be conducted using static (off-line) or/and dynamic (on-line) approaches. The former is merely carried out by examining the handwriting appearance (e.g., the curvatures, angles, and patterns of letters or symbols) and comparing it with the genuine image. In comparison, information about how the handwriting was generated is involved with the latter, including pace, movement changes, and pressure.

The advantage of using the signature is that it is non-intrusive and less time consuming while its disadvantage is that sometimes there is inconsistency in the signature. As a result, it becomes challenging to enrol and verify a user (Hirsch and Pearce, 2000).

Figure 2-10: Signature Recognition (Signing Hub 2016)

2.4.2.3 Keystroke Analysis

Keystroke dynamic is the ability to verify a subject based on the discrimination of the typing pattern on a keyboard or keypad. This is done while using a computing device with the typing characteristic is recorded. There have been works done in this domain with a different characteristic of the typing pattern achieving satisfactory performance. This work includes work by (Clarke and Furnell 2006) using inter-keystroke latency (the interval time between two successive keystrokes) and Saevanee and Bhatarakosol (2008) using finger pressure on keys. Other studies have been done on using keystrokes to enhance the security of mobile devices using different keystroke characteristics. The classification of keystroke dynamics uses methods like statistics and neural network as presented in works by (Bergadano et al. 2002; Brown et al. 1993; Joyce et al. 1990; Leggett et al. 1988; Spillane 1975) with good result output. Keystroke analysis is categorised as static (text-dependent) and dynamic (text-independent) (Banerjee et al. 2012). This categorisation is based on a predetermined and non-predetermined text typing either at the point of entering or after entering by comparing those against the reference template as defined as follows:

- Static (text-dependent): The subject's typing behaviour is analysed at the point of entry during authentication using password and user identification (ID) or later after gaining entry into the system through regular interaction
- Dynamic (text-independent): This typing is analysed with no predetermined text used by the subject. A comparison template stored in the computing system is used for analysing the input to authenticate the subject.

The use of keystroke analysis has an advantage of easy deployment because it can be easily integrated using the existing computing system without any additional hardware. Another advantage of the keystroke is uniqueness, low implementation and deployment cost, transparency and non-invasiveness, increase password strength and lifespan, replication prevention and extra security, continuous monitoring, and authentication (Teh et al. 2013). The use of keystroke analysis has some drawbacks, which include low accuracy as a result of some variation. The position of the subject (either sitting or standing) while typing can affect the pattern and typing frequency can increase. Also, using different languages might affect the typing rhythm.

2.4.2.4 Behavioural Profiling

Behavioural profiling classifies users based on the distinct pattern(s) of their usage of devices' applications and/or services, such as specific applications and websites they access, specific time of day, and for how long (Aupy and Clarke 2005). A profile template is created from the user's historical behavioural interactions to be used subsequently for the authentication process while the regular interactions determine whether it is the genuine user identity or vice versa when the usage pattern differs.

Research into behavioural profiling started in the late 1990s. However, the focus has been mainly on utilising the mechanism in intrusion detection systems (IDS) and fraud detection of telephony and credit card systems (Stolfo et al. 2000). The technique takes various aspects into consideration, such as network-based, device/host-based, desktop or mobile environments, and deploying it alone or coupled with other authentication techniques (Aupy & Clarke 2005, Li et al. 2011, Saevanee et al. 2012). The user's location information also can be incorporated based on either the mobile cellular network (i.e., cell ID), the global positioning system (GPS) (i.e., longitude, latitude), or/and the IP address. Nevertheless, it might be considered as a third approach of authentication as proposed by the International Information Systems Security Certification Consortium (ISC2) and can be referred to as “what the user is” (Conrad et al. 2012). Notwithstanding, it can be argued that it is under the behavioural profiling biometrics because location alone would not be sufficient to verify the user; hence, it is a measurable feature rather than a category. There are some advantages and disadvantages of this method. Behavioural profiling biometrics have the potential to monitor behavioural patterns on most types of devices without interrupting the user from their everyday interaction, which makes them a good alternative for transparent and continuous authentication. While a disadvantage is that it suffers from privacy and acceptability issues. The fear of private information leakage during behaviour monitoring tends to affect the level of user acceptance. Furthermore, because of the high comparative probability of changing over time along with the low individuality of user behaviour (as most of the behavioural biometrics), it is probably more feasible to be incorporated with a multi-factor/biometric authentication system.

2.4.2.5 Biometric Gait Recognition

Gait recognition discriminates people based on the patterns associated with their walking stride. Figure 2-11 shows the periodic motion of the legs. The person’s gait data is initially collected and enrolled to generate a template, which is used to compare with other samples; if the samples match it, the user is considered as legitimate; otherwise, some security processes should be completed (Nickel et al. 2011). The first serious discussion and analysis of human gait emerged during 1977 with Cutting & Kozlowski (1977); they experimented and proved the plausibility of identifying individuals based on their gait. Later, many studies have emerged concerning gait recognition from various perspectives. It is primarily used for surveillance purposes, then deployed to authenticating users using wearable sensors (Gafurov et al. 2007a) or on mobile devices (Derawi et al.

2010). Regarding surveillance, a camera is used to capture the gait motion from a distance (without the user's intervention). Regarding user authentication, wearable sensors can be worn in various places such as on the ankles, hip, or arms (Gafurov et al. 2007a). When smartphones are employed, the user's gait information can be captured while they interact with the device or even carry it in their pocket.

It is considered an unobtrusive authentication method with more user-friendliness. Recently, researchers have shown an increased interest in mobile gait authentication (Hoang et al. 2013). Gait recognition can be seen as an advantageous biometric identification technique for many reasons. Firstly, the gait of a person can be captured unobtrusively and continuously via acceleration sensors, which are already contained into most smartphones as long as the user walks; therefore, there is no need for additional hardware costs for using this method (Derawi et al. 2010). Secondly, gait recognition does not require explicit user interaction during verification or identification (Nickel & Busch 2013). The third reason is the security, because of the fact that the gait of an individual is challenging to mimic (Hoang et al. 2015). However, this technique is relatively affected by many factors such as clothing (e.g., footwear), health condition (e.g., pregnancy), and ground condition (e.g., grass or concrete) (Derawi 2012).

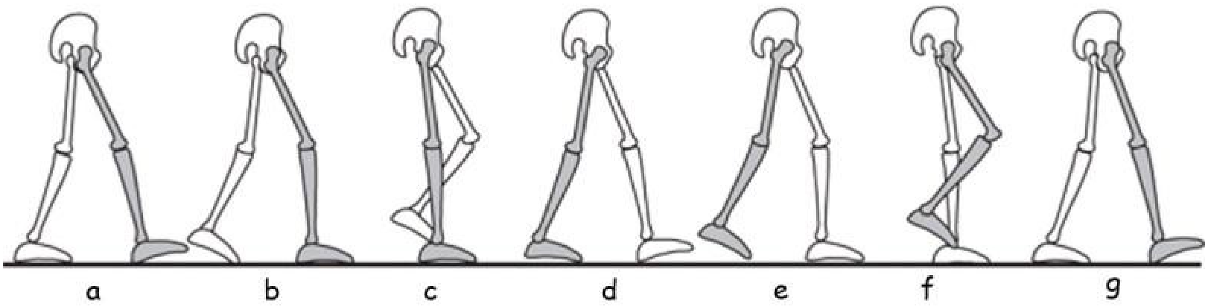


Figure 2-11: Illustrations of periodic motion of the legs (Hoang et al. 2013)

2.5 Summary of the Biometric Techniques

In order to highlight the effectiveness of these biometric techniques to a universal advanced authentication solution, Table 2-1 shows the aforementioned biometric techniques' transparency feature against a continuous smartphone-based authentication approach (where ✓ and X represent yes and no, respectively). Regardless of the lower level of uniqueness and perpetual quality created by behavioural characteristics, for example, evolving state of mind, well-being, and environment, they tend to be more general, straightforward, and consequently usable than the physiological ones (Clarke 2011).

Biometric Techniques		Transparency	Usability (Intrusive/ Non-intrusive)	Anti-spoofing safeguard	Applicability in Smartphones
Physiological	Fingerprint Recognition	X	✓	X	✓
	Palm Print & Hand Geometry	X	X	X	X
	Facial Recognition	X	✓	X	✓
	Iris & Retina Recognition	X	X	✓	✓
	Ear Geometry	✓	✓	X	✓
Behaviour	Voice Recognition	✓	✓	X	✓
	Signature Recognition	X	✓	X	✓
	Keystroke Analysis	✓	✓	X	✓
	Behavioural Profiling	✓	✓	X	✓
	Biometric Gait Recognition	✓	✓	X	✓

Table 2-1: shows the aforementioned biometric techniques transparency feature

It is also apparent that biometric behaviour techniques outperform physiological techniques based on usability and transparency requirements. It is also obvious that none of them is free from scoring X (no). However, dependent on context requirements, behaviour biometric gait recognition would feasibly be suitable to some extent. Moreover, this would be more practical to apply significantly if a real-life dataset is used as a transparent authentication approach.

2.6 Conclusions

The smartphone and its services and information are becoming targets of cybercrimes and have serious security concerns as any other technology. As mentioned before, current approaches of user authentication (e.g., secret knowledge and token-based authentication) suffer from security and usability issues. Many studies confirmed that the password might be easy to guess by attackers, forgotten, written down, shared with friends, discovered by eavesdropping, or even social engineering while the token can be lost or stolen. As a result, system security will be compromised and misused by attackers. They are also used to offer a point of entry authentication; hence, they cannot provide continuous protection for smartphones. Furthermore, they tend to be intrusive and fail to take into account user satisfaction. In consequence, an authentication method needs to improve the level of security being afforded while reducing user inconvenience. Therefore, users do not need to carry or remember anything. Therefore, the operational performance being achieved is highly correlated to the biometric software. Biometrics is a method of recognising and thus authenticating subjects based on their physiological (e.g., face, fingerprint, or hand geometry) or behavioural (e.g., gait, signature, and voice) features.

Modern smartphones contain various mobile sensors, such as accelerometers, gyroscopes, magnetometers, rotation sensors, vision sensors (cameras), audio sensors (microphones), light sensors, temperature sensors, GPS receiver, Wi-Fi and Bluetooth receivers. Hence, a number of biometric techniques can be applied to these devices, such as gait, face, iris, and voice verification. It is hard to deploy all of them on mobile phones because existing mobile resources cannot guarantee the acquisition of specific data, such as iris data, fingerprints, etc., accurately. Moreover, all these seem to be intrusive (e.g., typing passphrases, facing the front camera, etc.). Therefore, we need more convenient, secure, and effective biometric modalities that operate transparently for mobile authentication. Motion sensors, such as accelerometers and gyroscopes, are most commonly used as sensors for data collection and can be used to collect the data transparently. Moreover, the definite advantage is that no more hardware is needed; merely a software needs to be developed. Hence, researchers started using mobile phones to record the accelerometer data, which offers a user-friendly, unobtrusive, and a periodic way of authenticating individuals on personal mobile devices — all they need to carry is their cell phone.

3 A Literature Review of Gait Recognition

This chapter presents the state of the art in the academic literature on transparent and continuous authentication utilising gait recognition where gait data is recorded using accelerometer and gyroscope sensors, which are included in smartphone devices. The chapter begins with a detailed description of the methodology used for this review, followed by the literature review of the mobile gait-based authentication studies. Then the chapter concludes with a discussion that presents and analyses the research gaps.

3.1 Background of Biometric Gait Recognition

3.1.1 Gait collection methods

Biometric gait recognition can be categorised into three main approaches (captured using three different types of equipment): machine vision-based, wearable sensor-based, and mobile sensor-based.

3.1.2 Machine Vision (video sensors)

Machine vision (MV) uses video from one or more cameras to capture gait data (movement of the whole body), as shown in Figure 3-1. The video/image processing methods are applied in order to detect and extract static-like stride length, which is determined by body geometry and dynamic features from body silhouettes (Nickel et al. 2011a). MV systems can be used remotely without any user interaction; however, they are expensive and involve the use of background subtraction. MV-based gait recognition is mainly used in surveillance and forensics applications (Holien 2008).

Figure 3-1: Background segmentation for extracting the silhouette picture (Bajrami 2011)

3.1.2.1 Wearable sensors

The second class uses wearable sensors (WS), where the gait data is collected using a body-worn recording sensor(s). High-quality dedicated devices are used for data collection containing high-grade accelerometers, which can be placed on the hip, waist, pockets, arm, or ankle to record the acceleration while the subjects are walking, as shown in Figure 3-2 (Gafurov et al. 2007a). The accelerometers used for gait recognition usually are tri-axial and the acceleration signals are measured backwards-forward, sideways, and vertical. The collected acceleration signals are the result of the acceleration of the person's body, gravity, external forces like the vibration of the accelerometer device and sensor noise (Nickel, Brandt, et al. 2011c). Then the raw accelerometer data is segmented into cycles or fixed time windows to extract discriminative gait information (e.g., average cycle, standard deviation, energy, frequency-domain entropy, mean, variance, window mean difference, and the Bark frequency cepstral coefficients) (Gupta & Dallas 2014). The wearable dedicated sensors for gait are simple, small, and inexpensive devices and can be employed in the field of transparent and continuous user verification and identification settings, respectively. However, they require a lot of computational power, so they are not well suited for real-time detection of activities on low-powered devices and can be readily integrated into smartphone devices to reduce cost.

Figure 3-2: Different locations of the attached wearable sensor (Gafurov et al. 2007b; Gupta & Dallas 2014)

3.1.2.2 Mobile-based sensors

Modern smartphones contain various built-in mobile sensors (such as gyroscopes and magnetometers) that are most commonly used for gait data collection. The substantial advantage of this method is that no additional hardware is needed, merely software. In addition, they provide user-friendly, unobtrusive, and a periodic way of authenticating individuals on personal mobile devices. All they need is to carry their mobile phones, as shown in Figure 3-3. Also, Figure 3-4 shows the gait recognition is captured when a user carries their mobile device in their trouser pocket. Their gait information can be collected as they walk.

Figure 3-3: The three axes in which acceleration is measured and phone position when receiving data (Miguel & Neves 2013)

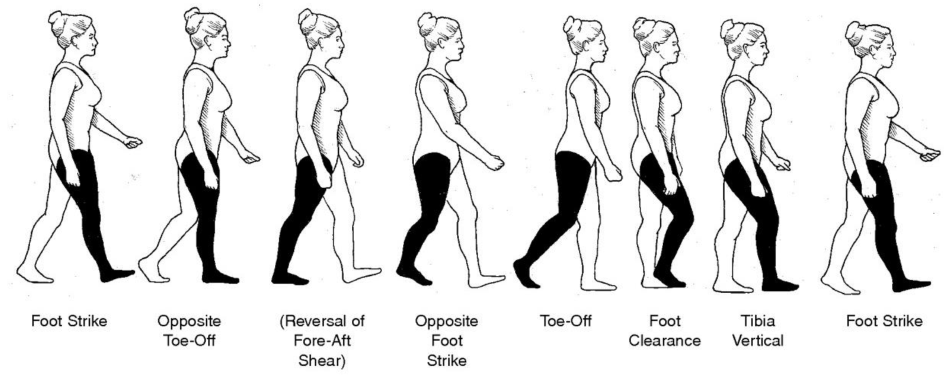


Figure 3-4: A complete gait cycle showing the eight gait phases (Miguel & Neves 2013)

Furthermore, most people have smartphones, but not all of them like to wear additional equipment. Hence, the smartphone will be suitable for designing an efficient, transparent, and continuous user authentication system.

3.1.3 Gait process approaches

There are various methods to process the gait data retrieved from the sensor. Most of the studies apply one of the following two methods to the gait signal: 1) cycle-based and 2) segment based. A brief description of each method is described below:

3.1.3.1 Cycle-based Method

The cycle-based process can be considered as the most common approach used in gait recognition. In cycle-based segmentation, the gait is supposed to be a periodic signal in which each gait cycle begins as soon as the foot touches the ground and finishes when the same foot touches the ground for the second time (i.e., two steps of a human) (Derawi et al. 2010).

Usually, two different methods are used to detect gait cycles from acceleration signals, namely local minima and the salience vector (Ferrero et al. 2015). The first step in cycle-based segmentation requires recognising local minima and maxima over a selected period, thus requiring a peak-detection algorithm. In many studies, only the minima are used to identify a cycle start. In case that the minima are not clear enough, there are often distinct local maxima (Nickel 2012). Then the data points between two sequential minima/maxima salience vectors are considered as one cycle. These determining cycles are considered the actual walking pattern (Nickel et al. 2011). Further analysis may be needed if there are some cycles in the acceleration signal that are different than others. This has been generally accomplished by using a distance function (e.g., dynamic time warping (DTW) or Manhattan) to omit in irregular cycles that are significantly different than other cycles (i.e., unequal length) as indicated in Derawi et al. 2010, Nickel et al. 2011, and Nickel 2012.

Afterwards, the regular cycles are averaged and the gait template concluded from subsequent average cycles for each acceleration direction, which is used for biometric template creation and sample comparison. Gait cycles based on pattern similarity estimation usually rely on simple metrics that measure dissimilarity of compared gait patterns (i.e., reference and probe templates), including standard classification methods (e.g., Manhattan, DTW, Euclidean distance, principal component analysis (PCA), and the cyclic rotation metric (CRM)) (Derawi & Bours 2013; Derawi et al. 2010; Nickel et al. 2011; Marsico & Mecca 2015).

As long as the achieved distance scores for the user's samples is low enough, that means the reference and probe samples are related to the same person. Otherwise, (when the distance score is high) it signifies they are not associated with the same person. Below is an explanation of the most common classification algorithms used in the cycle extraction approach (Derawi 2012) :

- Absolute distance (Manhattan distance metric)

Absolute distance is a straightforward metric that computes the sum of the absolute values of the differences between all the values in the template and the input value. However, it requires that the reference and probe templates have a similar length, as shown in Equation 1 (Derawi 2012).

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

- Euclidean Distance

The Euclidean distance is a modified process of the absolute range. It calculates the square root of the sum of all differences squared between the values of the stored template and the equivalent values in the test template, as given in Equation 2 (Derawi 2012).

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

- **DTW Distance**

DTW is an algorithm for calculating the optimal distance between two feature vectors regardless of their variation in length or speed. The DTW distance is different from absolute or Euclidean metrics and there is less restriction to the differences of features of the matching cycles.

In conclusion, the cycle-based segmentation suffers from some drawbacks, such as finding the best approach to specify the start and endpoint of each cycle. Moreover, irregular cycles and unclear boundaries between two cycles result in the possibility of cycle extraction failure methods and increase the error rates of these methods.

3.1.3.2 Segment-based Method

In this approach, the gait data is divided into a fixed time-length window (e.g., 5 or 10 seconds) from which the gait features are extracted depending on the acceleration values within the window. A time-window approach is considered simple and more accessible to apply than the cycle-based method (Nickel & Busch 2013).

Ordinarily, two types of gait features can be extracted from these windows: statistical and cepstral coefficients. The statistical characteristics, such as standard deviation (Std), minimum (Min), maximum (Max), mean value (Mean), and root mean squared (RMS). Furthermore, these features are created easily and do not need complicated calculations; they can achieve a high level of accuracy. These features are computed for a single axis (e.g., vertical, horizontal, and sideways directions) or with the three-acceleration axis (x, y and z). Likewise, the cepstral coefficient features, which are already used and had great success in speech recognition and speaker identification systems, have shown promising results in gait recognition, such as Mel-frequency cepstral coefficients (MFCCs) and Bark-frequency cepstral coefficients (BFCCs) (Nickel 2012; Nickel & Busch 2013). In order to construct more sophisticated feature vectors and perfect recognition, some studies merge both types of features (i.e., statistical and cepstral coefficients) (Nickel, Brandt, et al. 2011; Nickel et al. 2012; Hestbek et al. 2012). For classification of non-cycle-based feature vectors, the supervised machine learning algorithms are usually used, such as the support vector machine (SVM), hidden Markov model (HMM), and neural network, to classify the segment-based features.

The supervised learning in wearable gait recognition is a type of machine learning method, and it is used to get a general function derived from gait signal training data (i.e., the data obtained from the accelerometer signals). The function output should be a value continuously extracted, which is used to predict a class label for each person, and later utilised for classification, as well as a supervised learning technique that is commonly used for activity recognition in a majority of the researches (Bajrami 2011).

Mobile-based gait authentication utilises different supervised learning techniques, which are perfect performance results. These promising approaches include neural networks (Kwapisz et al. 2010; Watanabe 2014; Watanabe, 2015). Other studies by Nickel, Brandt, et al. (2011) and Nickel and Busch (2013) used the HMM classifier. Also, SVMs perform well in gait recognition, according to Hoang et al. (2013) and Phan and Dam (2015). Furthermore, many classification techniques from the collection of machine learning algorithms in the WEKA (Waikato Environment for Knowledge Analysis) data mining suite, such as decision trees (J48), neural networks, Bayesian networks (BN), random forest (RF), and radial basis function (RBF) were also used by (Kwapisz et al. 2010; Kwapisz et al. 2011; Watanabe 2014; Watanabe 2015).

3.2 Review Methodology

The methodology presents a comprehensive review of relevant studies. In order to ensure relevant literature was identified and analysed, a review protocol was developed to describe how the collected data were selected.

The main research questions highlighted by most researchers were:

- How reliable is the gait-based user authentication? (M. O. Derawi, 2012; Nickel, 2012; Muaaz, 2017)
- What are the best feature extraction and classification algorithms for gait recognition and to what extent can they adapt to recognise a person under different circumstances? (Holien, 2008; Bajrami, 2011; Nickel, 2012)
- How do external factors, such as different walking speeds and surfaces, influence accelerometer and/or gyroscope-based gait recognition? (Holien, 2008; Nickel, 2012; William A. Parker, 2014)

The following databases were considered in this review because of their popularity and relevance to the chosen research domain:

1. IEEE Xplore: http://ieeexplore.ieee.org/Xplore/home.jsp
2. ScienceDirect: http://www.sciencedirect.com/
3. ACM Digital Library: http://dl.acm.org/
4. SpringerLink: http://link.springer.com/
5. Google Scholar: http://scholar.google.co.uk

The following compound search expression was used to find the current state of the art in "Transparent Authentication Utilising Gait Recognition":

"(mobile OR smartphone OR gait) AND (transparent OR continuous OR unobtrusive) AND (authentication OR verification)"

The number of the most related references published to date is shown in Table 3-1. The initial number of references was 98. After applying an additional filter (i.e., transparent authentication using gait recognition on smartphone devices), a final 35 papers were selected.

Database	Number of References	Final Selected
IEEE	13	12
Science Direct	30	2
Springer	3	2
Scholar	48	15
ACM Digital Library	4	4
Total	98	35

Table 3-1: The number of returned references

3.3 Overview of Mobile-based Gait Authentication Related Work

This section will present a comprehensive analysis of the prior studies on gait recognition systems using the acceleration sensors embedded in a smartphone environment. Furthermore, several key areas will be discussed, including devices types and positions, types of sensors, the datasets and numbers of participants employed, pre-processing data approaches, features created, classification, and an evaluation of using test data recorded under different conditions. These studies are fully described and they are listed in chronological order.

Sprager (2009) reported the first successful attempt of gait recognition based on the inertial data acquired by smartphones. In work, the Nokia N95 was attached to the hip to collect the vertical and horizontal acceleration data that were divided into cycles. A cumulant-based method for the identification of accelerometer-based gait data was used. Cumulant coefficients of order 1 to 4 were extracted from each gait cycle, and each cycle was converted into a feature vector. They used a Gaussian radial basis kernel function of the SVM classification as the classifier. Using a cross-day test database of six subjects in two different days for two weeks, they obtained a recognition rate of 92.9%. The tested rate was high, but it was expected that the performance would drop if more subjects were used in the evaluation process (i.e., it is clear that the data set and the participant number are limited). Besides, several walk styles (e.g., normal, fast, and slow) were tested; but that was determined with pre-identified speeds, which are not realistic to apply in practical life. Also, they walked on a surface made of stone plates and ignored the effect of other surfaces.

The research study by Derawi et al. (2010) also utilised a mobile phone to collect gait data. In their experiment, the G1 phone with embedded accelerometers was placed horizontally on the right-hand side of the hip for each of their subjects to collect the data (as shown in Figure 3-5). The software was written for the Android platform in order to transfer the data from the accelerometer to a file. The program for data analysis was based on the work of (Holien 2008), which utilised a dedicated accelerometer. The data was collected from two different days (a cross-day test) from 51 volunteers at their normal walking speed. Each subject was asked to walk two sessions a day (37 meters for each session). The signal was captured through a 3-axis accelerometer (for each of the three directions x, y and z) with 40-50 samples per second (as shown in Figure 3-6). The repeated gait cycles that were extracted from the acceleration in the x-direction showed better results. The data average cycle length was computed, after time interpolation and filtering, by using dynamic time warping (DTW). This was used to identify minima, which equate to cycle starts (as shown in Figure 3-7). For each walk, the most regular cycle was used as a feature vector (in terms of dynamic time warping distance). The average cycle vector length was about 45 samples used as the feature vector for this walk, and again dynamic time warping was used for distance calculation. For data from those four sessions, one session was used as the training dataset and the other three sessions were used for the testing purpose. The result of an EER of 20.1% for the gait recognition indicates promising performance for the mobile accelerometer. Nonetheless, the EER is still high as only the normal walking speed of the participants was tested. Hence, more research on other walking styles should be considered.

Figure 3-5: Phone attached to the subject and the three axes in which acceleration measured (Derawi et al. 2010)

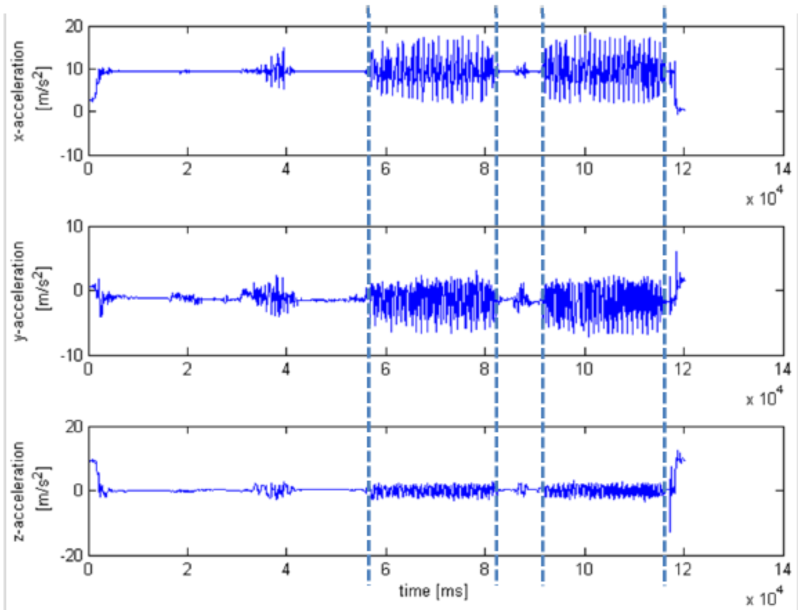


Figure 3-6: Sample data collected with the G1 from x, y and z directions

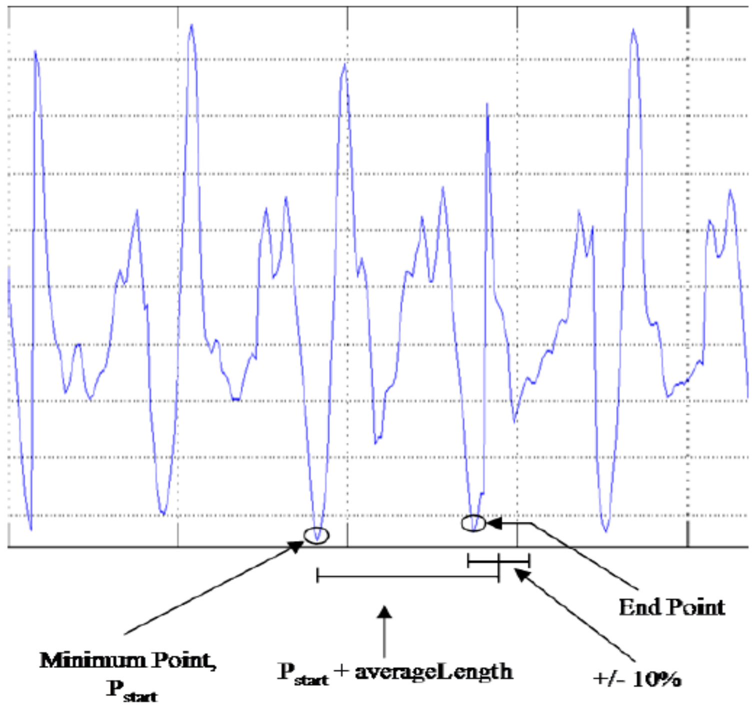


Figure 3-7: Gait Cycle Detection

Nickel et al. (2011) suggested using HMM for classification, which has been applied successfully in speaker recognition systems. This work applied the same data set from Derawi et al. (2010). The advantage of using HMM is to overcome that restriction of irregular and unclear cycle minima and decrease the error rates by using the accelerometer data directly to build up the model and thus help to obtain better recognition. The authors stated that HMM or DTW could be directly applied to raw time series data of acceleration, instead of feature extraction. The data collected using a G1 mobile phone with a 3-axis accelerometer was placed horizontally in a pouch, attached to the belt of 48 subjects on the right-hand side of the hip. Two sessions were captured on two different days and the subjects walked at their normal speed. All data from the first session and parts of the second session were used for training and the remaining parts of the second session were used for testing. The data were interpolated to have a fixed sampling rate of 200 samples per second; then it was divided into fixed-length parts of three seconds, which were used directly for training and testing. The results achieved an FRR of 10.42% and a FAR of 10.29%, almost a 50% performance improvement from EER of 20% (Derawi et al. 2010). The authors claimed that processing steps do not need distinctive minima or more particular properties as in cycle extraction methods. However, the performance could be improved if several pre-processing methods were utilised, e.g. extracting up-to-date feature extraction with advanced HMM patterns.

A further experimental study was carried out by (Nickel et al. 2011b); they employed the same data set used in (Derawi et al. 2010). They proposed a surrogate approach, a non-cycle-based gait representation. There was no need for previous identification for the gait cycles. Otherwise, the features were extracted from the time-series data from a selected time window. For the two data sessions, half of the data from both sessions was used for training; the other half of the data from both sessions was used for testing. In this study, the segment length was set to 5s, 7s, 10s and overlaps of 50% result were evaluated. The partition into training and testing data was not the same in the three evaluations. The finding of the present study highlighted that the segment length of 10s outperformed the other two settings. However, if a more extended time (e.g., 15s) was tested, maybe a more meaningful result could be obtained. Single features and various combinations of the features were also examined. For each segment, one feature vector was created. As a starting point, statistical features were calculated for the acceleration signals (mean, maximum (Max), minimum (Min), binned distribution (BD), root mean squared acceleration (RMS), zero-cross and Std) and extracted in addition to the Mel- and Bark-Frequency Cepstral coefficients (MFCC, BFCC), which are usually used in speech and speaker recognition. Previous studies on gait recognition have not dealt with MFCCs and BFCCs. These authors were the first to use them in the context of gait recognition. They found that the features that were extracted from the accelerometer and magnitude sensors values presented the best performance. SVMs were used as the classification method. Usually, the FAR and FRR are directly calculated from each classification result for each segment. An alternative approach was proposed based on a quorum voting system method presented by the authors who merged many genuine classification results (#gV) into one and accepted a user as legitimate if the obtained classification results are positive #GV; otherwise, the probe signal was rejected as shown in Figure 3-8.

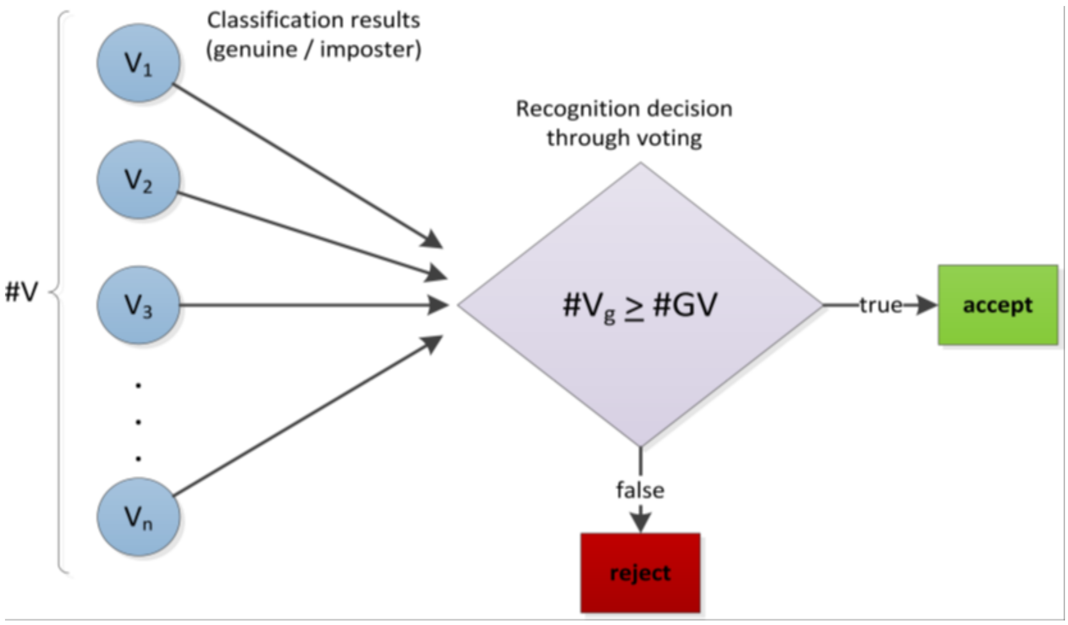


Figure 3-8: Quorum voting scheme method (#V total segments of a probe gait signal, #Vg, number of votes for genuine, #GV positive classification results) (Nickel, et al. 2011b)

The voting results were reported separated by three experimental setups: cross-day (enrolment and probe data were collected in two different days), same-day (enrolment and probe data were collected on the same day), and mixed-day (database consists of data of two separate sessions, but enrolment and probe data were taken at least partly from the same session). The reason for analysing the three different setups was to get an impact on the influence of time on the recognition results. The experiments contradicted these results; the cross-day scenario showed 5.0% FAR and 25.0% FRR, which is considered high. For the mixed-day setup, the results showed an FRR of 6.3% at FAR of 5.9%, which can be regarded as better in comparison to the previous studies by (Derawi et al. 2010) and (Nickel et al. 2011) with an EER of 20% and 10%, respectively. However, the experiments were conducted in a very controlled environment: participants walked at their normal speed, and no mention was made to the surface type that would positively affect the results.

Furthermore, as highlighted by the author, the same-day scenario had more training data with higher intra-class variability results in better trained SVMs and, hence, better recognition rates. Also, they mentioned that the error rates increased significantly when data is collected on continual days. So, the system was more suitable for a single-day scenario.

Sprager & Zazula (2011) extended their previous work to investigate the influence of the different solid surfaces on the human gait pattern efficiency and gait identification based on accelerometer data. Data were collected by using the same method as described by their previous experiments (Sprager 2009). They obtained the gait samples by using accelerometer data from Nokia N95 with a built-in 3-axis accelerometer attached to the right hip of the subjects. Five users were asked to walk across four different surfaces with their usual speed on three different days with various walking distances shown in Table 3-2.

Surface	Length
Ground	25 m
Stone Plates	30 m
Gravel	15 m
Grass	25 m

Table 3-2: Surface and walking distances used in the experimental protocol (Sprager and Zazula 2011)

The gait cycles were extracted based on wavelet transform, and high-order statistics were used to calculate all gait cycle features. Cumulant coefficients of the order of 2, 3 and four were calculated for all time lags. Discrimination of the different subjects was done by principal component analysis (PCA). The study claimed that the identification performance of subjects based on their gait was not affected significantly by different solid surfaces when no evaluation was presented. Even so, the short distances were used in the experiment with the same (standard) pace. Additionally, the mobile phone was attached steadily to the hip, i.e., it did not rotate; however, in real life, it does rotate when in the pocket according to pocket movement. Thus, the accelerometer noise within their experiment was reduced.

Kwapisz et al. (2011) evaluated and described their scheme-based accelerometers to identify users on smartphones based on physical activity performed by the user. They collected data from twenty-nine users as they performed and executed daily activities like jogging, walking, climbing stairs, standing, and sitting. They used Android phones from different brands (Nexus One, HTC Hero, and Motorola Backflip); the data were collected using the Android applications for the accelerometer sensor on the mobile phone. In all cases, the accelerometer data used a default frequency of 50ms (20 sample/second). The data were supervised by one of the research team members to ensure the quality of the data. The classification algorithms that were used in their study could not directly learn from time-series data; to achieve this, they divided the data into 10-second segments and then generated features from the accelerometer values contained in each 10-second interval (since acceleration data is collected for three axes 20 times per second for a 10-second range, there are 600 total values). Then they generated useful features based on the 600 raw accelerometer readings. Then they created forty-three features based on variations of six basic features (i.e., average, standard deviation, average absolute difference, average resultant acceleration, time between peaks, and binned distribution). Once the data set was prepared, they used three classification techniques from the Waikato Environment for Knowledge Analysis (Weka) data mining suite to induce models for predicting user activities: decision trees (J48), logistic regression, and multilayer neural networks. In each case, they used the default setting ten-fold cross-validation for all experiments, and all results are based on these ten runs. They claimed that most cases achieved good accuracy. For the two most common activities, walking and jogging, they generally achieved accuracies above 90%. Jogging appeared to be easier to identify than walking, which seems to make sense, as jogging involves more extreme changes in acceleration. However, there were few examples of sitting and standing; they identified these activities quite well because the two activities cause the device to change orientation and this is easily detected from the accelerometer data.

The authors indicated that it was more challenging to identify the two stair-climbing activities (i.e., ascending the stairs and descending the stairs). This was because the two similar activities are often confused with each other. The confusion matrices specify that many of the prediction errors are because of confusion between these two activities. The experiments showed that when a subject is climbing upstairs, the most mutual improper classification happens when expecting “downstairs,” which occurs 107 times and accounted for a decrease in accuracy of 19.6% (107 errors out of 545). However, when the actual activity was ascending downstairs, walking out-paces slightly “upstairs” in terms of the total number of errors (99 vs 92), but this is because walking occurs more than three times as often as climbing upstairs as in their data set. Figure 3-9 shows that the patterns in the acceleration data between “walking”, “ascending stairs”, and “descending stairs” were somewhat similar. To limit the confusion between the ascending and descending stair activities, another set of experiments was made. They combined ascending stairs and descending stairs into one activity. The resultant confusion matrix for the J48 algorithm was significantly improved. Despite the fact of providing some activities in this experiment, it was limited for a few supervised activities, which were fundamentally far from realistic to be applicable and usable.

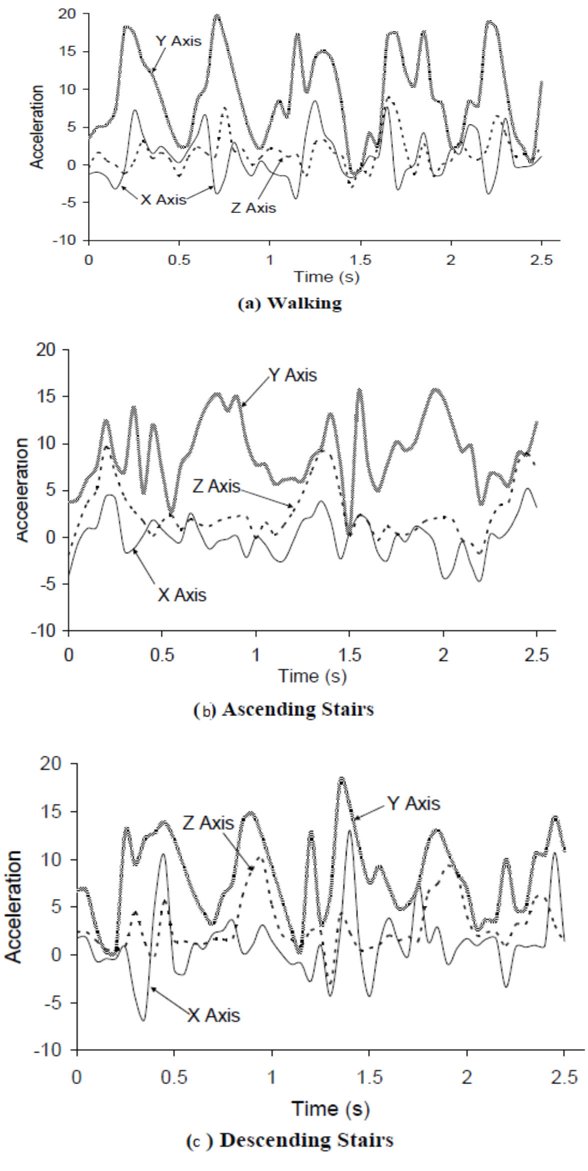


Figure 3-9: Acceleration plots for the (a) walking, (b) ascending, and (c) descending activities (Kwapisz et al. 2011)

Kwapisz et al. (2010) aimed to assess their previous experiments to identify and authenticate users' mobile phones. The data collected by an Android phone device was placed in the front pocket for thirty-six users, monitoring several daily activities (e.g., walking, jogging, and climbing stairs) for a predetermined time in one session only. In total, 10 minutes of activity was captured, and the time-series acceleration data were segmented into 10-second partitions. The data collection and feature extraction were performed as described in their previous work (Kwapisz et al. 2011). In this study, four separate data sets were created, each containing examples from only one activity (i.e., walking, jogging, ascending stairs, and descending stairs) for the authentication purpose. Their results were used to examine the suitability for each activity discerning between users. The authors created a fifth data set, which they refer to as "aggregate (Oracle)", identical to the complete data set but which contained the activity label as a feature. They used two classification techniques from the WEKA data mining suite to induce models for person identification: decision trees (J48) and neural networks. They changed the multiclass identification problem into a binary classification problem. Where the positive class indicated to the user was authenticated and the negative class to the other (thirty-five) users. As the positive class is so infrequent (on average it contains 1/36 of the data), most classification approaches tend to generate classifiers that do not perform well at predicting this (occasional) class. This was not desirable; therefore, they under-sampled the negative type, such that the resultant ratio of positive examples to negative examples was 1:3. The identification results indicated that the models were successful at recognising users' identities based on only 10 seconds of accelerometer data. However, while some of the precisions may not appear to be that good, they were fundamentally quite impressive when considering that for this 36-class classification problem, the straw man strategy of always guessing the most frequent class yields accuracy in the 4-7% range. The results achieved for identification-decision trees (J48) and neural networks were 72% and 69% identification rates, respectively.

The second finding was the authentication results, which were reported for only five users, which was a minimal data sample. It was a similar case for person identification. They first presented the effects related to individual examples and then applied the most frequent user strategy to determine the actual authentication presentation statistics. The key statistic for authentication was the positive authentication (a user is correctly granted) and the result was 85.9% positive authentication rate at 95% negative authentication rate (the imposter was correctly recognised as an imposter). They achieved 100% positive and negative authentication degrees for all five users by applying majority voting to all test data. In contrast, the authentication was based on a limited number of users (only five people). Furthermore, they implemented majority voting, which is legitimising the user when half of the test samples or more are positive. This might significantly increase the acceptance of the users' verification claim wrongly when the system is applying on a more considerable amount of data.

Derawi and Bours (2013) were the first to utilise the mobile smartphone in both data collection and real-time analysis. This work extended the experiments of (Derawi et al. 2010) and (Kwapisz et al. 2011). In comparison with the authors' previous work (Derawi et al. 2010), various walking speeds and analysis methods were evaluated. In the work of (Kwapisz et al. 2011), the walking data was analysed into segments of 10-seconds that affected the activity recognition negatively. So, any changes of activity or speed within the 10-second segment may have caused the segment to be unidentified as the features extracted from that segment may have been a mixture of two different activities. In this work, the author focused on single cycles in the walking data, so if the walking speed changed, then only a few walking cycles would be affected, and recognition before and after these few cycles would not be disturbed. In this study, five participants were asked to walk at three different speeds (i.e., three templates for each user). The purpose of the system was to identify the user or the walking activity. They tested 20 new users and with the five enrolled users. The Manhattan distance metric was used for the comparison applied on the phone, and Euclidean distance and DTW were used for the comparison on the PC. The accuracy of the activity recognition was 99%, and the users were identified successfully with 89.3% of the cases with 1.4% false positive probability. While the results look good, the experiment worked on activity recognition rather than verifying the person who was doing the activity.

In comparison with previous studies, Nickel et al. (2011) experienced a more realistic application scenario with gait recognition on mobile devices. The Motorola Milestone sensor was used to collect data from 48 subjects who walked at their normal speed in the same shoes in two different days. The participants needed to walk straight on a flat floor for 10 seconds through the enrolment. Then they were asked to walk on a predefined route in a realistic scenario (i.e., walk on linoleum and tiled floor in a non-straight line, open the glass door, walk upstairs, and stop at some points). During the walk, they were asked to stay in nine authentication points defined previously. They collected 28 data sets in each session for each of the 48 subjects, for a total of 2,688 data sets.

An effective cycle extraction was used, which is capable of handling irregularities occurring in the data. In case the gait data do not have specific minima (maxima), the usual formula for cycle extraction methods will be less effective. Therefore, the authors proposed using the salience vector (which is known as the right salience vector) to determine the cycles starting point in order to get an adapting process that depends on the data. Only a vertical acceleration (x-direction) is considered from the measured acceleration. The acceleration values of the cycles were compared using two different gait recognition methods: DTW and the cyclic rotation metric (CRM). First, DTW was used as a distance function for comparison and for applying majority voting; after computing the distance between the reference cycles with all probe cycles, matching occurred only if a pre-selected threshold was above the distance between two cycles. Otherwise, there was non-matching (i.e., if at least 50% of the results were a match, the whole comparison was accepted, and the subject was authenticated). Secondly, they applied a CRM distance metric which cross-compared two sets of cycles (a reference cycle and an input cycle) with a cyclic-rotation mechanism to find the best matching pair. This comparison was used to find the most optimal and the best distance score when cross-comparing two sets of the cycle.

The subject was authorised if at least half of the results was matched. Their database contains 48 subjects gait data collected from the same day and cross days. Because it corresponds to a realistic scenario, they involved walking for about 15 minutes on a predefined route around corners, opening and closing doors, walking up and down stairs, having to cross doorways, and walking on different surfaces (linoleum and tiles). Data were collected in two sessions and compared with the data from the same session and the other session (on a different day) to see the influence of the period between enrolment and authentication. The best results obtained in terms of same-day were EER of 21.7% for the module using CRM as a distance and EER of 28.0% for the module using majority voting. The distinctiveness of this evaluation is that it is completely performed on the smartphone. The authentication results relied on the subjects as they could change their shoes and trousers during the different sessions. The device position may be influenced by many factors such as the height and angle of the pouch, as well as its stability. That means that each subject could enrol many times with different pants and shoes.

As a consequence, the threshold needed to adapt to the subjects, and that was impractical as a higher limit increases the probability of authentication for the attacker. In addition, the computational time for the CRM module was longer, around 32 seconds, and for the MV-module about 27 seconds. These intervals were far too high for a real authentication application. Another hurdle was that the authentication was started just once, when the user needed to use his phone again and switched off the screen saver. It can be noted that the error rate that was achieved was significantly high for both classification methods. Furthermore, the user needed to wait 30 seconds until the phone unlocked itself because the system would extract cycles from 30 seconds of data and the comparison that was done with the reference template took about 30 seconds, which took more time than entering a PIN, and this as less user-friendly.

Nishiguchi et al. (2012) presented a study to demonstrate the reliability and validity of a smartphone accelerometer for gait recognition. They used two devices: a smartphone and a tri-axial accelerometer and taped them together for the data collection. Data were collected from 30 volunteers in controlled walking conditions, and the trunk accelerations were more secure over the L3 spinous (i.e., body centre mass) at normal speed. After signal processing, the study computed the gait parameters of each measurement: peak frequency (PF), root mean square (RMS), autocorrelation peak (AC), and coefficient of variance (CV) of the acceleration peak intervals. All the results of the gait parameters captured by the smartphones significantly correlated with the same parameter results gained by the tri-axial accelerometer. The authors had demonstrated that the smartphone with a gait analysis application used in their study could quantify gait parameters with a relative degree of accuracy that is equivalent to that of the tri-axial accelerometer key, as shown in Figure 3-10. As a result, this evaluation showed the reliability and validity of gait analysis by Android-based smartphones.

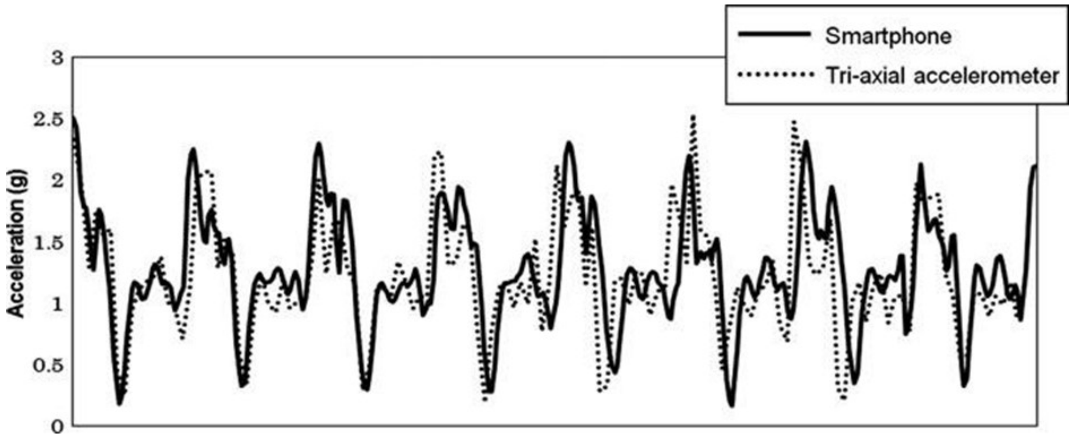


Figure 3-10: Acceleration waveforms of the smartphone and tri-axial accelerometer (Nishiguchi et al. 2012)

Hoang et al. (2012) analysed acceleration signals of gait biometric using the two main methods in gait identification: template matching in the time domain and machine learning in the frequency domain. DTW was used to calculate the similarity score of the extracted gait templates while the SVM was used to classify extracted features in the frequency domain. Eleven participants were employed and they achieved the recognition rate of both methods, respectively 79.1% and 92.7%, which are considered good results. But the data collection session from the volunteers was from the same day, and it did not take into account the biometric gait fluctuation for each person day by day. Also, weaknesses of their work include the static position of the mobile phone and the limited evaluated dataset (only 11 subjects).

The first attempt at providing a more detailed investigation regarding the effects of different walking speeds and surfaces on gait recognition was conducted by Muaaz and Nickel (2012). They utilised Google's G1 smartphone-based accelerometers and cycle extraction approach. The cell phone was placed inside the pouch and attached to the subject's belt or trousers as its screen facing the subject. The accelerated values were accessed by Android API software. The cycles were extracted to create a template for each subject based on the same cycle extraction process of (Nickel et al. 2011) and considered the measured acceleration of the vertical x-axis as it was more distinguishable than y and z-axis. Cycle length was estimated and detected by computing the min-salience and max-salience vectors. Outliers (e.g., unusual cycles) were removed from a set of cycles using DTW distance.

After computing the distances between cycles, a threshold value was specified. The cycle distance had to be less than the threshold value, otherwise at least half of the cycle was cancelled. At least three cycles were needed as remaining cycles (i.e., name of all cycles remaining after deleting unusual cycles) or the threshold value increased, and the deletion of the unusual cycles was repeated. The mean or median cycle of the normalised cycles or the lowest DTW distance value compared with other cycles was considered as typical cycles. After a generation of reference and probe cycles by the cycle extraction process, they were compared against each other to compute intra-class (genuine attempts) and inter-class distances (impostor attempts). The distances calculated for all cycles of one walk by DTW then passed to the majority voting module, which was used as a present threshold to calculate for each cycle if it matched the reference cycle. The result of the walk was accepted in the case that at least 50% of the cycles matched. Gait data were collected from 48 subjects walking at different velocities (slow, normal, and fast) on four variant surfaces (flat carpeted grass, gravel, and inclined). Six different walk settings for each subject were appraised in order to get a realistic experiment. The data capturing was in two sessions on two separate days to show the efficiency of the same-day and cross-day measurements. A walk setting framework measures changes in speed and surfaces when gait data is collected. The subjects regularly walked on different surfaces (carpeted, grass, gravel, and inclined) in four walk settings and changed the speed (slow, fast) on a carpeted surface.

Two experiments were implemented. In their first experiment, the typical cycle was used as a reference cycle and the remaining cycles as probe cycles. For each walk set, 34 different tests were performed, 24 of them used normalised cycles, and the other ten tests used cycles in their original length. The flow control of cycle extraction for each test in each group is shown in Figure 3-11. Cycle length parameters estimated and detected cycles were counted on the interpolation rate. The normalisation lengths of the cycles were 120, 100, 80, and 40 data values in different tests and evaluated threshold values were 80, 50, and 30 for the deletion of unusual cycles. In their second experiment, they extended their first experiment by attempting to improve the results by increasing the number of reference cycles. They used all the remaining cycles as reference and probe cycles. So, their second experiment was implemented only on the lowest EER tests of their first experiment. In both experiments, the resultant EER were high, as presented in Table 3-3. Nevertheless, increasing the number of reference cycles didn't enhance the results. Generally, most cases of their first experiment were better than those of the second experiments.

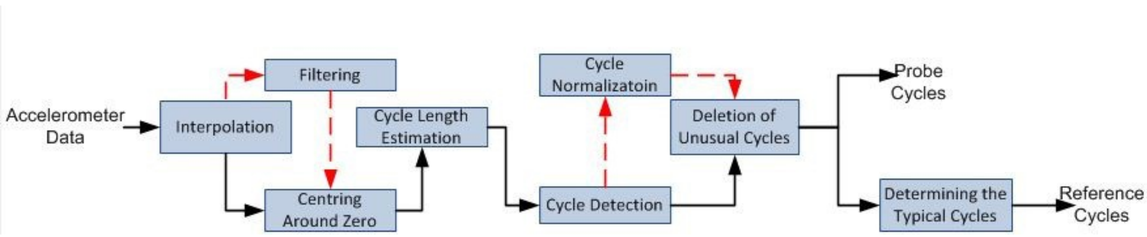


Figure 3-11: Flow control of cycle extraction steps, dashed arrow lines indicate optional steps (Muaaz & Nickel 2012)

The results of six different walk settings indicated that time interpolation with rate 100, cycle normalisation with length 100, a threshold value of 50 to delete unusual cycles and best cycle as the typical cycle produced the best results. The experiments in this study were more realistic, and the findings highlighted the effect of different walking speed and surfaces on gait recognition. In addition, they changed the space on a flat surface and regularly walked on different surfaces. The collected data was for each of the six walk settings: six different reference cycles for the same day and another six reference cycles for the cross day (e.g., normal walk, fast walk, slow walk, gravel walk, grass walk, and inclined walk). That means the importance of creating a reference template of data for each specific walking speed and surface type. Consequently, activity recognition should be applied to recognise the type of the test vector before matching with the correct authentication template. Furthermore, the finding of the same day EER ranged between 16.26% for the normal walk and 37.24% for the inclined walk. The cross day EER results were noticeably higher, ranging between 29.39% for the normal walk and 35.18% for the inclined walk.

Also, this work used a cycle-extraction approach to recognise accelerometer-based gait recognition. In comparison, Nickel et al. (2011b) performed a study on the same database where support vector machines were used for classification of non-cycle-based feature vectors. By applying the more flexible quorum voting, the FRR could be lowered to 16.2% while increasing the FAR to 20.8%. This is in comparison with previous publications that used another approach by creating feature vectors from segments of a fixed time length and using machine learning algorithms like HMMs for classification to give promising results (Nickel & Busch

2013). Because of the smaller size of the database containing different surfaces and paces, it could not be used to train HMMs. Nickel et al. (2011c) tested using the k-nearest neighbour algorithm as a classifier and applying majority voting produced non-acceptable FRR of over 80% (normal walk) while the FAR was 0%, which is considered unacceptable at the yielded 40% EER. Other experiments were done by Muaaz and Mayrhofer (2013) to classify the normal walk style. The Android phone Google G1, which was utilised to collect the data from 51 participants was attached to their hips, with an application that was improved to record three-dimensional (X, Y, and Z) accelerometer data to a text file with time stamps. Gait data was recorded at 40-50 Hz sampling frequency interpolated at 100 Hz. The recorded text files were stored on the SD-card. They used the same data pre-processing and segmentation method for gait cycle extraction as had been done in their previous study (Nickel, Brandt, et al. 2011b). Two types of experiments were conducted used two different classification methods: DTW and SVM.

Experiment one was based on the template-based classification. They investigated the effects of using the piecewise linear approximation (PLA) technique after the data pre-processing steps and just before the cycle length estimation module in cycle extraction steps. After cycle detection, the DTW was used as a distance function. The authors noted that the process of gait cycle extraction and their recognition was faster than the approach presented in (Muaaz & Nickel 2012) with about 2-3 minutes. However, the results obtained on the same-day and cross-day scenarios (with PLA) were EER 22.49% and 33.3%, respectively, which are more than those achieved without using PLA. Table 3-3 gives a comparison of results with and without PLA gait cycles.

Cycle extraction	Same-day	Mix-day	Different-day
	EER (%)		
With PLA	22.49	29.4	33.3
Without-PLA (Muaaz & Nickel	16.26	29.39	28.21

Table 3-3: A comparisons of results with and without PLA-based gait cycle extraction (Muaaz & Mayrhofer 2013)

Experiment two used the gait cycle as a feature and classified them using a machine learning (SVM) approach with the custom kernel (i.e., Gaussian dynamic time warping (GDTW) kernel). Entire gait cycles were used as a single feature. Also, a Gaussian kernel function was used with the Euclidian distance, which requires the same length input feature vector. In this experiment, DTW was used as an elastic similarity with the Gaussian kernel function instead of the Euclidian distance, as it works efficiently with unequal gait cycles to solve the problem of the fixed-length input feature vector. However, the approaches presented in this study are more appropriate for gait cycles of different length. Even though variable-length gait cycles were not used, the GDTW kernel approach that had been presented achieved a total error rate EER of 18.41%. In addition, this approach suffers from the indefinite kernel matrix. Furthermore, the data was collected in two different sessions and in controlled conditions (the subjects were asked to walk two times at their normal pace on a straight carpeted corridor) for a limited period (one minute in two days). Nonetheless, the achieved results were in terms of EER of 22.49% for same-day, 29.4% for different-day, and 33.3% for mix-days which are considered high.

The above finding is consistent with the study by (Ottomoeller 2014). The author examined a fixed-mobile location on the waist with standard walking conditions, referred to as a fixed method. A more natural position like a pocket, referred to as an unfixed method, with various subject tasks such as standing, sitting, walking, biking, running, driving, and random movements. They used PLAs for gait classification for the first experiment; in comparison, the other test used SVMs for classification. Although the processing time was faster with about 2-3 minutes when using PLAs, the EERs increased from 16.26% (without PLAs) to 22.49% (with PLAs), which is considered high especially for a controlled experiment (walking straight on the flat floor). The achieved results were EER of 18.41% when SVMs were used, which is also very high. The two experiments focused on the training data. The author mentioned that using larger data sets, up to 10% of the data, achieved better accuracy. The best EER was 14%, and by using the unfixed approach, the EER was 12%. Whilst they experienced different implementations of the strategies, the achieved results added no more to their previous work.

Nickel and Busch (2013) used the same data collected in (Nickel et al. 2011). The authors took into account data sets that must be capable of enough training data for HMMs and contained more realistic data from two different days to calculate cross-day results. This data was pre-processed as in prior studies, and the features were extracted from each segment for each of the three-acceleration axes (x, y and z) and the magnitude vector. There were five steps to pre-processing the data. Firstly, the walks were extracted from the data (i.e., non-walking parts were removed from the recorded section). Then the data was interpolated to a fixed sampling rate, centred around zero, and divided into segments. Features were extracted from each segment. This study focused on Mel-frequency (MFCCs) and bark-frequency cepstral coefficients (BFCCs), which were used before in prior work by the author (Nickel, Brandt, et al. 2011c). The HMM was used as a classification approach. Different amounts of training data, various feature sets, and segment length were tested in addition to same-day and cross-day results, which were computed in order to analyse the influence of time on gait. Each of those sections was evaluated separately, and the results for the parts with no stairs were even better as follows:

Firstly, the subjects had to walk straight on the flat floor in the enrolment phase of the data collection. Only 10s (one section) of training data per subject when all data from the second day was used as probe data, which was not enough, and resulted in an EER of 31.6%. When they increased the training data sections gradually, it was found that the minimum time required for adequate training of HMMs was 33s of enrolment data, and the more data, the better the performance.

Secondly, the HMM classification results were calculated for each subject using a BFCC2MFCC feature set. The overall FAR for most of the users was 10%, with a high variability detected with FRR. They assumed that the outliers might be caused by changing shoes or the phone position because of the different trousers.

For all tested feature sets and segment lengths (2, 3 and 4 seconds), the range of the EER was from 15.77% to 18.94%, and the best performance was BFCC2MFCC employing a segment length of two seconds. In the same-day experiments, they obtained an EER of 7.88% for a BFCC2MFCC, which was about half of the reported cross-day results. In addition, the HMM was trained with approximately two minutes (10 sections) of walking data (including walking around corners and upstairs), an EER of 7.45% was achieved with mixed test data of all route sections.

All previous FAR and FRR results were calculated directly from classification results for each segment. As in their prior work (Nickel et al. 2011c), they combined many classification results into a single decision by using the voting approach. There were multiple classification decisions (#V) instead of one, by employing quorum voting. This quorum needed to have at least #GV of the #V classifications vote in order to authenticate the users; otherwise, the test signal was excluded. The calculation in this work suggested using #V = 60 and #GV = 1; good results were achieved with a segment size of 4s. The EER could be decreased from 15.77% to 7.45%. Best results also were obtained when using a segment size of 2s; the reported EER was 7.33%. The same-day results could even be decreased to 0.71% EER. After voting, the best results were achieved for the feature MFCC. When using about one minute of walking data for training, they got an EER of 6.15%. However, it was found that increasing the amount of training data (about two minutes in this work) achieved lower EER of 5.81%. Nevertheless, even if the suggested values (#V = 60 and #GV = 1) may be considered odd, the achieved results were good. Furthermore, in comparison with the author's previous work (Nickel et al. 2011), which utilised a cycle extraction approach, their finding for EER was 21.7%. This result reduced to 6.15% by using the segment-based approach. Consequently, it can be noted that the segment-based approach provides better performance compared to a cycle-based method.

In another significant study, Nickel et al. (2011a) presented a comparison between two classification methods (SVM and HMM) on the same database. Data was collected using a typical mobile phone on a cross day from 36 subjects. Each subject walked about 32 minutes on a flat carpet during two sessions in normal and fast speed. More than 19 hours of accelerometer data were collected. The data was divided into fixed-length time segments. Different feature sets were evaluated, and the best for each classifier was assigned. Furthermore, different amounts of training data were tested. As the fixed-size sections are less complicated and less error-prone than discovering the beginning of the cycle. The data was divided into segments of size 3, 5, and 7.5s with an overlap of 50%, then two types of features were extracted: Several statistical features were extracted from each segment for each of the three-acceleration axis (x, y and z) and the magnitude vector. These were Min, Max, Mean, Std, Bin, RMS, zero crosses. They also extracted MFCCs and BFCCs from the segments for each axis. In this study, different feature sets were evaluated and the best for each classifier was assigned:

- **Single Features**

In this experiment, they investigated the discrimination possibility of single features by testing them individually. Cross-day datasets were tested with a different interpolation rate (50, 100 and 200 samples) for each segment length (3, 5 and 7.5 seconds).

Regarding the HMMs, changing the interpolation rate and segment size did not affect the error rates. Whilst, the SVMs presented results when using a segment length of 5s and a low interpolation rate of 50. So, the setting of 50 as the interpolation rate and segment length 5s was applied to all the results of this work. In case of SVMs, they attained for cross-day results for single features the FRR obtained for the SVMs, which were between 99.18%- 47.90% for the Diff and the BFCC2, respectively. For the HHMs, the EER was between 46.23%- 17.06% for the Diff and the MFCC, respectively.

- **Combined Features**

Features were combined with having enough information because single statistical features alone did not have sufficient information. For SVMs, the achieved results were 20.26% of EER using a combination of the BFCC2 of x-axis acceleration and the magnitude vector. For HMMs, the achieved results were

17.30% of EER using MFCC of all axes when a normal walk was used for testing. However, the best error rates were obtained when testing with fast steps for SVMs and HMMs with 15.43% and 14.52% of EER, respectively.

This study highlights that the Cepstral coefficients showed better performance than the statistical features. Also, the finding indicates that without voting, both classifiers results were approximately similar, with the SVMs being slightly better for normal walking. To obtain a more acceptable result, the quorum voting was applied to 70 samples from a user’s test data (equivalent to about three minutes of the walking data), which showed a notable reduction. For the SVMs, the EER was reduced to 10% for normal walking while the EER of HMMs decreased to 12.63%. Most experiments used about four minutes of training data for each subject. The error rates reduced by 25% for SVMs when they doubled this quantity; in contrast, the minor improvement was experienced in the HMMs. While using mixed training data containing fast and normal walks raises the error rates when the subjects were walking fast or normal during the authentication phase, suggesting dedicated training samples for different walk styles are required for obtaining better performance and more realistic situations. By applying the quorum or majority voting methods, the results were enhanced. However, they needed to double the testing data for more accurate results (i.e., decrease the error rates by 25% for SVMs, but insignificant enhancement for HMMs).

Moreover, compared to previous research findings into HMMs (Nickel et al. 2011) and SVMs (Nickel et al. 2011c) had been stated as EER are 20.71% for HMMs and 30.0% for SVMs, which are incredibly high. In this work, the evaluated results were about 10% better for the SVMs. However, the HMMs were evaluated on an experimental database (i.e., containing walking straight on the flat floor).

In 2012, Hestbek et al. (2012) published a paper that is similar to the one in (Nickel, Brandt, et al. 2011a). They used the same database, modifying the feature extraction by applying wavelet transform. Gait templates were created by BFCC and standard deviation (SD) from the wavelet coefficients, as shown in Figure 3-6. The data were divided into fixed-length time segments size of 5 seconds with a 50% overlap then SVM was used for gait template classification. The experiments showed the possibility of using wavelet transform as a feature extractor for gait recognition. The achieved result was a FAR of 9.82% and FRR of 10.45%. The error value was not affected by the inclusion of the wavelet transform in the feature extraction process. The results in Table 3-4 also show the approximate EER with the results calculated when using the wavelet transform as a method to extract the gait feature templates and when no wavelet transform is applied.

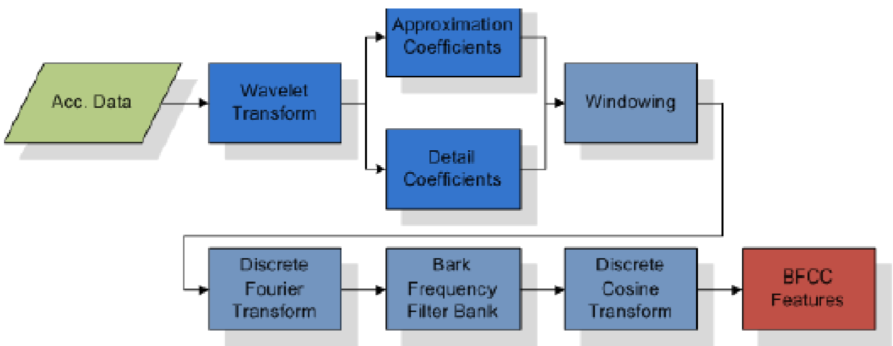


Figure 3-12 the process of extracting BFCC features (Hestbek et al. 2012)

Study	FAR	FRR	EER
Wavelet	9.82%	10.45%	10.14%
Acceleration	10.01%	10.00%	10.01%

Table 3-4 Evaluation of results with and without wavelet transforms (Hestbek et al. 2012)

In comparison with the authors' prior studies (Nickel; Brandt, et al. 2011a) and (Nickel, Brandt, et al. 2011c) in which no wavelet transform is employed, no performance improvement was made, as they reported a FAR of 10.01% at an FRR of 10.00% after applying quorum voting (see Table 3-4).

In another study, Nickel et al. (2012) extended their previous work using the k-nearest neighbour (k-NN) algorithm to evaluate the same database obtained in (Nickel et al. 2011a). Thirty-six users were asked to walk normally on a flat floor attaching a mobile phone at their hip pouch and data were collected on two different days (five minutes of gait data for each day). The segment of raw data was divided into 7.5 seconds; the segments overlap by 50%. In this work, some single features that provided better performance were combined (they selected the features that had low intra-class variability and a high inter-class variability). Otherwise, related features were extracted from gait data based on their discriminating potential score (DPS) and added to feature vectors. The experience shows that the results based on DSP features are sensible (Nickel et al. 2012). The feature vectors were considered only of function BFCC, calculated for all three-acceleration axes and the magnitude vector was preferred for normal walking. Three different algorithms HMM, SVM, and K-NN were evaluated to find out the more convenient method to classify users' gait templates. The k-nearest neighbour algorithm showed better efficiency than the machine learning algorithms, e.g. hidden Markov models and support vector machines (Nickel, Brandt, et al. 2011a), as shown in Table 3-5. The table presents the results after applying a quorum voting method. The walking data used for voting were based on 1.7 to 3.2 minutes, which is the same walking data time that the authentication must be based on.

Algorithm	Length of the best feature vector	Authentication based on x minutes	Lowest EER (%)
HMM	104	2.5	8.75
SVM	52	2.5	8.85
k-NN	52	1.7	8.24

Table 3-5 shows the evaluation of the essential facts of the SVMs, HMMs, and k-NN in the same database (Nickel et al. 2012)

The lowest EER is between 8.24% and 8.85%. Also, there was no distinct difference between the results of the three methods. Albeit, the k-NN was based on a short walk duration and gave good results when classified before and after voting. The enrolment of the steps and authentication were applied to a standard smartphone. However, the proposed approach achieved low processing time and revealed that it was efficient enough to be used in practice. But the K-NN technique needs calculating the distance between vectors of attributes (test case) and stored training case. Also, predominantly, the FRR was still very high while the FAR was low. Although one classification applied to each segment, it is based on less than eight seconds walking period, which is considered to be short. Therefore, to obtain the best results, they can combine several successive classification results and transform them into a single result. In contrast, this work showed that the k-NN implementation in a controlled walking condition when subjects were walking on flat floor achieved good results. Therefore, analysing the influence of covariates like clothes, carrying bag, and surfaces are needed, in addition to different spaces and the position of the phone.

Frank et al. (2010b) presented two papers to investigate the activity and gait recognition. Acceleration data was collected from 25 subjects by placing the smartphone in the trouser pocket while they were walking, lingering, running, upstairs, downstairs, and riding in a vehicle. Also, additional data was collected during the other study by the authors (Frank et al. 2010a) from four persons while they were carrying out a sequence of fitness activities: riding a stationary bike, using an elliptical machine, using a stationary rowing machine, using a stair-climbing machine, and running on a treadmill. In both studies, the author created a time-delay embedding model for each subject. Then the nearest neighbours in the model were calculated and the new data segments of time series were compared using an algorithm called geometric template matching (GTM). The two experiments focused on identifying activities and users, respectively. Even so, the activity classification algorithm and gait identification achieved 85.48% and 100%, respectively. On the contrary, the recorded data was limited (20 seconds for each subject). Also, the participants joined with an observer who recorded the labels as the activities were performed, which is not realistic. Including other types of sensors, e.g., GPS, may help to provide further context for identifying more activities such as doing sport in a gym or driving to work.

Recently, there has been increasing attention in mobile-based activity recognition. An experimental investigation was conducted by Watanabe (2014) to explore the possibility of authenticating a mobile user while the phone is not fixed in the pocket (i.e., walking, making a call, and touching the mobile screen). They achieved initial experiments using enhanced application of four subjects only to collect the client-generated acceleration data on a normal walk. A better accelerometer record function on the iPhone was used to collect the superior three-axis accelerometer data during walking. Each had made one round trip along the corridor with about 50m distance for two minutes, whereas they took the phone in three holding states as follows: in the pocket, squeezing the phone to the ear, and imagining touching on the screen. Additional information has been calculated, such as the phone rotation around the 3-axis utilising the gyroscope and the electromagnetic compass. In their work, they split the data from time series acceleration for each axis into windows of 3-second intervals.

Furthermore, they extracted characters from the data in each window. They used ten-fold cross-validation (i.e., windows divided into ten equal size samples) to train and test the system. This method used 90% of samples for training and 10% for testing and the cross-validation process was repeated ten times. This study employed J48, NN, RBF, BN, and RF algorithms and neural networks (NN) from the Weka. Table 3-6 shows the achieved results for each activity; the best results gained when the phone was in the pocket were 1.30% FAR at 2.34% FRR. Results for touching on the screen were extremely bad (state 3), indicating that the location of the mobile was adversely affecting the outcome significantly in addition to the limited collected data set (four users) within the same day.

Algorithm	State 1		State 2		State 3	
	FAR	FRR	FAR	FRR	FAR	FRR
J58	3.39	15.63	7.03	22.66	6.51	29.69
NN	1.30	2.34	3.65	7.81	9.38	22.66
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
FR	0.26	7.81	1.82	17.19	2.86	32.03

Table 3-6 The stated results FAR (%) and FRR (%) for each holding state by using five algorithms (Watanabe 2014)

Watanabe (2015) extended his previous work by carrying out additional experiments by increasing the number of subjects to eight participants and adding the data recorded on a different day. The same techniques for collecting data, pre-processing, and feature extraction were used as their previous work (Watanabe 2014).

One metric used correctly classified the rate to evaluate the recognition performance. In this work, two experiments were carried out for identification of users and states.

The goals of the first and the second experiments were different; these were the identification of users and states. In all the tests for each subject, the collected dataset was: logged time, gravity, user-generated acceleration, rotation rate, magnetic field, and three angles. Each individual made one round trip along the corridor with about 50m distance for about one minute.

- The first was implemented as increasing the number of subjects and added the data recorded about one month later.

- The second was carried out to examine the influence of different walking states when a subject placed the phone in the right or left pants pocket or shirt pocket, called or touched using his right or left hand, or wore slippers or shoes.

However, they observed that the additional data collected before have a different effect on identification performance according to phone position states. Also, they have noted there are influences of various walking states for one subject. The author discussed in his paper the application of an "immunity-based diagnosis model to gait recognition to integrate the identification results from multiple smartphone sensors".

Their study results show that:

The classification rate according to the phone's position (in the pocket) was not affected by an individual number.

A small number of subjects influenced performance when calling and touching.

The author employed four algorithms (BN, NN, RBF and RF). The results are shown in Table 3-7. For each algorithm, the default settings were used, that is, automatic optimisation methods. In comparison with the author's previous work, this study set that radial basis function (RBF) was the best and Bayesian network (BN) and random forest (RF) also performed better, but decision trees (J48) was the worst in all cases.

	Bayesian Network	Neural Network	Radial Basis Function	Random Forest
To classify into 9 states	94.44	93.52	91.67	93.52
To classify into 3 holding states	96.30	99.07	98.14	98.14

Table 3-7 Correctly classified rate (%) of states by four algorithms when a subject walked in nine different states (Watanabe 2015)

Wolfe (2013) accomplished a study on both authentication and identification modes. Both built-in accelerometers and magnetometers were used to deal with disorientation and misplacement errors in mobile installation problems. Realistic data, including the influences of mobile installation errors and shoes, were collected from 38 subjects asked to walk in their average speed on the ground floor, and the mobile was located in a narrow pocket (e.g., the jean trouser). Then the processing steps to the authentication model were analysed thoroughly.

The signal was segmented into separate gait cycles instead of a fixed time interval and both time and frequency domain features were employed. The support vector machine classifier and the radial basis function kernel were used to classify users from their gait features. The good results are shown in Table 3-8 (a), which were achieved by using segmentation based on the gait cycle. However, the users walked in a regular style (i.e., the cycles were distinguished easily). Table 3-8 (b) proves the performance of their method with/without fixing disorientation error (in transformed Z-signal). It can be noticed that better accuracy was attained when they applied the proposed method with the transformed Z-signal (i.e., the phone position was always fixed in parallel to the ground and its Z-axis direction to the sky and perpendicular to the ground). However, adjusting the phone position always parallel was considered unrealistic because the users needed to put their phone in various places around their body wherever there was a pocket (i.e., back pocket and inside coat pocket). They achieved accuracy of about 94.93% under the identification mode, the FAR, FRR of 0%, 3.89% and processing time of fewer than four seconds under the authentication mode. And the best-achieved classification rate at length 3s was also worse (79.53%).

Segmentation Method		**Accuracy**
Fixed length	3000ms 6000ms 9000ms	**87.88%** **87.78%** **84.73%**
Gait cycle	2 gait cycles 4 gait cycles 8 gait cycles	**92.26%** **94.93%** **90.94%**

(a)

Segmentation Method	**Fixing disorientation**	**Accuracy**
Fixed length	No	79.53%
Our algorithm	No Yes	84.03% 94.93%

(b)

Table 3-8 (a) Improvements of segmentation based on gait cycles compared with a fixed length, (b) The influence of disorientation error to the effectiveness of classification mode

Hoang et al. (2013) examined the influences of the sampling rate on creating an adaptive gait recognition model with two different mobile phones. They discussed the impact of the sampling rate on the pre-processing steps, such as noise elimination, data segmentation, and feature extraction. Gait features were extracted from two different mobiles signals. The feature extraction and classification methods were used as same as in their previous work (Wolfe 2013). In addition, both the average error rate (AER) and intra-class correlation coefficients (ICC) were calculated to evaluate the probability of creating a device-independent mechanism. They claimed that the sampling rate of 32-36 Hz was most appropriate to build an efficient gait recognition system, which is considered low. They achieved the classification accuracy of about 91.33 ± 0.67% for both devices.

More recently, Hoang et al. (2015) proposed a different security and privacy gait authentication system on a smartphone by using the fuzzy commitment scheme. The fuzzy commitment scheme is one of the biometric cryptosystems aimed at securing cryptographic keys using biometrics. They stored a key that was biometrically encrypted by gait templates gathered from a mobile accelerometer in order to authenticate the user as an alternative of storing archetype gait patterns for user authentication as usual methods. The binary BCH code was used in this work as the error-correcting code to discriminate differences between biometric measurements. The system was evaluated on the dataset including gait signals of 34 subjects and achieved the zero- FAR and the FRR of around 16.18%. However, as they used a simple quantisation scheme, the achieved error rate of FRR was still rather high, which could affect the friendliness of the system.

Marsico and Mecca (2015), tested different methods of walk recognition to investigate gait identification by utilising smartphone accelerometers. Twenty-six subjects were asked to fix the phone vertically in the belt, either on the right or the left side of the hip. The participants were then asked to keep the feet together and start walking by the leg opposite to the phone location and walk in controlled or adverse conditions. The system records ten steps along a straight line in the most regular way. The DTW for each axis was used for classification. They achieved a recognition rate above 0.95 and EER 7.69%. However, the system was so controlled and constrained with a fixed number of steps in the walks. In other words, it was not realistic and not suited for individual mobile devices.

The smart kiosk model was used by (Phan & Dam 2015) to research choosing gait item as a biometric factor, then to design a well-matched scheme for their smart kiosk system. There were two mechanisms in their smart kiosk system: continuous authentication based on gait in mobile devices, and interactive kiosk to afford users with facilities corresponding to their identities.

The authors used different procedures to recognise clients from their gait characters and other schemes. They used Android mobile devices for real-time authentication. This article specified that the authentication with biometric structures, a source of high-entropy information, for authentication and identity has the following advantages: cannot be lost or forgotten, difficult to copy or share, hard to forge, and cannot be guessed easily. The smart kiosk system allows clients to access online services related to their individual identities using indirectly continuous gait-based recognition. The authors proposed a user organisation method based on gait using multiple SVM classifiers and a secure scheme with biometric information. Experiments with a dataset of 38 people presented the results accuracy of this method was up to 92.028 %.

3.4 Discussion

As illustrated in the previous section, gait recognition can be captured using different acceleration sensors embedded in devices (e.g., wearable devices and smartphones). Table 3-9 displays a comprehensive analysis of the prior studies on gait recognition systems using the mobile sensors that have been discussed in this literature. A thorough discussion on several key areas based on the information presented in Table 3-9 (including sensors, data pre-processing, features, and classification) of the gait recognition within the smartphone environment is presented as follows:

Table 3-9: Comprehensive analysis of the prior studies on gait authentication systems using mobile sensors.

Legend: T-Pocket: Trouser Pocket; Acc: Accelerometer; ML: Machine Learning, SD: Same day; CD: Cross day; CCR: Correct Classification Rate

No.	Author/year	Device	Position	Sensor/ Sampling rate	Segmentation	Match algorithm	# users	Performance		Data Collection Scenario	Walking Type
								Measure	Value		
1	(Sprager 2009)	Nokia N95	Right hip	Acc. /37 Hz	Cycle-based	ML SVM & PCA	6	CCR	93.3%	Normal walk, fast walk, slow walk	Normal walk, fast walk, slow walk
2	(Derawi et al. 2010)	Google G1	Right hip (horizontally)	Acc. /45Hz	Cycle-based	DTW	51	EER	20%	Normal walk	Normal walk
3	(Frank et al. 2010b)	HTC G1	T pocket	Acc. /25 Hz, Barometric pressure	Fixed-length segment	ML SVM &PCA	25	CCR	100%	Running, walking up or down stairs	Running, walking up or down stairs
4	(Frank et al. 2010a)	Android phone	T pocket	Acc. / 32Hz	Fixed-length segment	ML Nearest-Neighbour	40	CCR	100%	Normal walk	Normal walk
5	(Kwapisz et al. 2011)	Nexus One, HTC Hero, and Motorola Backflip	Front pants leg pocket	Acc.	Fixed-length segment	Decision Trees (J48), Multilayer NN	29	CCR	90%	Normal walk	Normal walk
6	(Kwapisz et al. 2010)	Nexus One, HTC Hero, and Motorola Backflip	Front pants leg pocket	Acc.	Fixed-length segment	J48,NN	36	CCR	82%	Normal walk	Normal walk
7	(Nickel et al. 2011)	Google G1	Right hip-pouch	Acc. /40 sample per second	Fixed-length segment	ML HMM	48	EER	6.15%	Normal walk	Normal walk
8	(Nickel, Brandt, et al. 2011b)	Google G1	Hip- pouch	Acc. / 45Hz	Fixed-length segment	ML SVM	48	EER	6.1%	Normal walk	Normal walk
9	(Nickel, Brandt, et al. 2011a)	Motorola Milestone	Right hip-pouch	Acc.	Cycle-based	ML SVM, HMM	36	EER	10 % 12.36%	Normal walk	Normal walk
10	(Wolfe 2013)	Google Android HTC Nexus one mobile	T pocket	Acc./27 Hz Magnetometer	Cycle-based	ML SVM, RBF	38	CCR	94.93%	Normal walk, Three types of footgear:	Normal walk, Three types of footgear: sleeper, sandal, shoe

No.	Author/year	Device	Position	Sensor/ Sampling rate	Segmentation	Match algorithm	# users	Performance		Data Collection Scenario	Walking Type
								Measure	Value		
										sleeper, sandal, shoe	
11	(Nickel et al. 2011)	Motorola Milestone	Right hip-pouch	Acc.	Cycle-based	DTW, Manhattan	48	EER	21.7%, 28%	Normal walk, climbing of stairs	Normal walk, climbing of stairs
12	(Boyle et al. 2011)	Motorola Droid phone (API) O.S	Motorola Droid phones,	Acc.	Fixed-length segment.	K-NN	2	CCR	Normal walk 90%, Different speed. 85%-98%	Normal walk	Normal walk
13	(Nishiguchi et al. 2012)	Smartphone	Body centre mass.	Acc./ 7.68 sample per second	Cycle-based	Spearman's correlation coefficient	30	-	-	Normal walk	Normal walk
14	(Nickel et al. 2012)	Moto Milestone	Right hip-pouch	Acc. / 127 sample per second	Fixed-length segment.	ML HMM, SVM, k-NN	36	EER	8.24%	Normal walk	Normal walk
15	(Hestbek et al. 2012)	Motorola Milestone	T pocket	Acc.	Fixed-length segment.	ML SVM	36	EER	10.45%	Normal walk	Normal walk
16	(Hoang et al. 2012)	Google Nexus	T pocket	Acc. / 27 Hz	Cycle-based	DTW SVM	11	CCR	79.1%, 92.7%	Normal walk	Normal walk
17	(Muaaz & Nickel 2012)	WS &Google G1 Android by HTC	Right hip-pouch	Acc. / 40-50 samples per second	Cycle-based	DTW	48	EER	Normal 29.39%, Fast 33.81%, Slow 35.31%	Normal walk Different walk speed and surface	Normal walk Different walk speed and surface
18	(Ho et al. 2012)	Android phone	-	Acc.	Cycle-based	SVM	32	CCR	100%	Normal walk	Normal walk
19	(Derawi & Bours 2013)	Samsung Nexus S	T pocket	Acc./ 150 Hz Magnitude	Cycle-based	Manhattan distance , Euclidean distance, DTW	5	CCR	89.3%	Normal walk, fast walk, slow walk	Normal walk, fast walk, slow walk
20	(Nickel & Busch 2013)	Motorola Milestone	Hip-pouch	Acc. Magnitude	Cycle-based, Fixed-length segment.	ML HMM	48	EER	15.8%	Normal walk	Normal walk
21	(Hoang et al. 2013)	Google Android HTC Nexus One, LG Optimus G	T pocket	Acc. / 32-36Hz , 100Hz	Cycle-based,	SVM& RBF	14	CCR	91%	Normal walk	Normal walk

No.	Author/year	Device	Position	Sensor/ Sampling rate	Segmentation	Match algorithm	# users	Performance		Data Collection Scenario	Walking Type
								Measure	Value		
22	(Muaaz & Mayrhofer 2013)	Google G1	Right hip-pouch	Acc. / 40-50 Hz	Cycle-based	DTW	51	EER	33.3%	Normal walk	Normal walk
23	(Watanabe 2014)	IOS app.	Pocket	Acc./20 samples per second	Fixed-length segment.	J48, NN, RBF, BN, and RF.	-	FAR, FRR	1.30%, 2.34%	Normal walk	Normal walk
24	(Ottomoeller 2014)	Android phones	Fixed on the waist	Acc.	Cycle-based	PLA- DTW, SVM, Gaussian kernel	51	EER	14%	Normal walk	Normal walk
25	(Hoang et al. 2015).	HTC Google Nexus one	Pocket (Vertically)	Acc. / 32 Hz	Cycle-based	Hamming distance	34	FAR, FRR	0%, 16.18	Normal walk	Normal walk
26	(Phan & Dam 2015)	Android phones	-	Acc.	Smart kiosk system	SVM	38	CCR	92.028%	Normal walk	Normal walk
27	(Watanabe 2015)	iOS iPhone 5	T pocket, Shirt pocket	Gyroscope, Magnetometer /20 sample per second	Fixed-length segment	NN	8	CCR	94.44%	Normal walk	Normal walk
28	(Marsico & Mecca 2015)	One pulse smart phone	Right pouch, Left pouch (vertically)	Acc.	Fixed-length segment	DTW	26	EER	10.46%	Normal walk Different shoes no high heel	Normal walk Different shoes no high heel

The majority of studies have used a fixed position to collect gait data. The smartphones were attached to the person's hip or trouser pocket. This position turned out to be the most appropriate for the cell phone users and performed better when the orientation of the device remained constant throughout the transition between activities (France 2014). As such, many studies have required the attachment of the device in a known position on the human body. However, this is not the normal behaviour of individuals in the real world, who may place their mobile devices casually and even randomly (e.g., put their mobile phones on the desk).

Sensors that were used in early studies were limited; in comparison, current smartphones contain various sensors, such as accelerometers, gyroscopes, magnetometers, rotation sensors, and GPS receivers. Amongst these sensors, no study has been carried out using GPS information despite the fact that it can reveal critical location information that might aid in the decision-making process. This might be because GPS was not available on those devices. Nonetheless, GPS may be helpful to provide further context for identifying more activities, such as doing sport in a gym and driving to work. The triaxle motion-based signal can be obtained using accelerometers or gyroscopes. Both of them seem to provide the same information. Previous studies have primarily concentrated on using accelerometers alone. By using the data from two sensors or more, the performance and accuracy are expected to be better than using a single sensor. However, it remains a challenge for real-world applications imputable to data reliance on sensor placement (i.e., the device position may be influenced by many factors such as height and angle of the pouch, as well as its stability). Hence, there is little research that has used multi-sensors.

In terms of pre-processing the gait data, most of the studies applied one of the following two methods: cycle extraction and segmentation. In cycle-based, the gait is supposed to be a periodic signal in which each gait cycle begins as soon as the foot touches the ground and finishes when the same foot touches the ground for the second time (i.e., two steps of a human). Many studies depend on the cycle-based approach (Derawi et al. 2010; Nickel et al. 2011; Muaaz & Nickel 2012; Muaaz & Mayrhofer 2013; Hoang et al. 2015), and generally, the accuracy results of using the cycle-based method are relatively low. In the best cases, they achieved a 16.18% EER. These high error rates mostly result from using the cycle-based approach, which suffers from some drawbacks, such as finding the best approach to specify the start and endpoint of each cycle. Moreover, the cycle can be irregular (i.e., vary in length and width following different user speeds). Hence, unclear boundaries between two cycles result in the possibility of cycle extraction failure methods and increases the error rates. In comparison, the segmentation-based approach divides the gait data into fixed time-length windows. A time-window approach is considered uncomplicated and more comfortable to apply than a cycle-based method. In spite of the simplicity of the segmentation approach, it seems to be the most commonly used by studies (Nickel et al. 2011; Nickel, Brandt, et al. 2011b; Nickel & Busch 2013; Watanabe 2014; Watanabe 2015). It can be noted that the segment-based approach provides better performance in comparison to the cycle-based method (e.g., the worst EER achieved was about 10%).

Concerning features, two main approaches can be used to extract information from the acceleration signal; the statistical features and cepstral coefficient features from a fixed-size window could achieve better results. The statistical features, such as Std, Min, Max, Mean, and RMS, were used by (Nickel, Brandt, et al. 2011b; Sprager & Zazula 2011; Nishiguchi et al. 2012). The cepstral coefficient features, which have already been used and have had great success in speech recognition and speaker identification systems, have shown promising results in gait recognition, such as Mel-frequency cepstral coefficients (MFCCs) and Bark-frequency cepstral coefficients (BFCCs)(Nickel, Brandt, et al. 2011a; Nickel & Busch 2013; Hestbek et al. 2012). In order to construct more sophisticated feature vectors and better recognition, some studies merge both types of features (i.e., statistical and cepstral coefficients) (Nickel, Brandt, et al. 2011b; Nickel 2012; Hestbek et al. 2012). More features often provide better performance and accuracy. However, consideration must be given to take into account the length of the feature vector and, subsequently, the processing power and memory that will be needed, especially when the whole biometric process may exist in the smartphone.

For the matching algorithms, they can be classified into the following categories: cycle based (TM) and fixed time windows (ML). The gait cycles correspond to two steps of a human and, based on pattern similarity estimation, usually rely on simple metrics that measure dissimilarity of compared gait patterns, including Manhattan and Euclidean distance (Derawi & Bours 2013). Besides simple metrics, advanced metrics are commonly used such as DTW or DTW-derived metrics (Derawi et al. 2010; Nickel et al. 2011; Marsico & Mecca 2015), principal component analysis (PCA) (Sprager & Zazula 2011), or the cyclic rotation metric (CRM) (Nickel et al. 2011). However, these classification algorithms achieved high EERs ranged between (19%- 33%). The high error rate may be consequent to the complicated nature of cycle extraction. Gait signal is assumed to be periodic, and the mobile base signal is very noisy and commonly influenced by many factors (i.e., device orientation, type of the sensors, and many other environmental factors). Also, the cycle changes according to the person's speed (i.e., cycle length varies according to walking speed). Then each separate gait cycle length will need to be normalised, and this increases the computational effort. This indicates these algorithms do not operate well with different walking templates and behavioural biometric techniques in general because of fluctuating human behaviour. Therefore, it is more suitable to collect multiple templates for different days and apply advanced algorithms as in the recent studies that utilised the prominent approach for comparison of feature vectors, such as machine learning algorithms that are well established in other pattern recognition domains such as speaker recognition (Nickel, Brandt, et al. 2011a). These promising approaches include neural networks, k-NN, HMMs classifier, SVM, and the Gaussian mixture model (GMM) classifier (Ottomoeller 2014) (Kwapisz et al. 2010; Kwapisz et al. 2011; Watanabe 2014; Watanabe 2015)(Boyle et al. 2011; Nickel et al. 2012)(Nickel, Brandt, et al. 2011a; Nickel et al. 2011;(Sprager 2009; Frank et al. 2010b; Nickel, Brandt, et al. 2011c; Nickel, Brandt, et al. 2011a; Hestbek et al. 2012; Hoang et al. 2012; Muaaz & Nickel 2012; Ho et al. 2012; Hoang et al. 2013; Muaaz & Mayrhofer 2013; Phan & Dam 2015). Many classification techniques from the WEKA data mining suite (decision trees (J48), neural networks, Bayesian network (BN), random forest (RF), RBF) were also used by (Kwapisz et al. 2010; Kwapisz et al. 2011; Watanabe 2014; Watanabe 2015). Generally, they achieved better accuracy. So, it can be noticeable from the previously conducted evaluations that the recognition rates obtained from segments based are better than those of cycle-based. As gait is assumed to be periodic, each time segment is reasonably expected to contain similar signal features. This approach requires fewer computational operations than cycle-detection and thus is more suited to use with mobile devices. Moreover, irregular cycles and unclear boundaries between two cycles result in the possibility of cycle extraction failure methods and increase the error rates in these methods.

Regarding the experiment setup, three configurations can be applied: whether enrolment and probe data are collected on the same day, on different days (i.e., cross-day scenario when the acceleration signals obtained on the first day are used for training and the signal obtained from the second day are used for testing), or if the database consists of data of two different sessions but enrolment and probe data are taken at least partly from the same session (mixed-day). Analysing these three setups gives the possibility of evaluating the impact of template ageing on the recognition results. The cross-day performance represents the most realistic results because in real-life training and testing data are from different days. However, the cross-day results are much lower than the same-day results. Furthermore, there is no common standard on how to collect the data sets' subject of experiments regarding the number of walk sessions, distance, speed and the time of the dataset measured in different ways such as seconds, minutes, hours, and days for each subject.

Prior literature has shown the accelerometer-based biometric gait recognition is still a new field of research and the majority of researchers were focused on the evaluation of using test data recorded under laboratory conditions (i.e., "assumes the fact that natural and unaffected gait has been performed during the measurement" (Sprager & Juric 2015)), containing just walking straight on a flat floor ((Mohammad Omar Derawi et al. 2010; Nickel, Brandt, et al. 2011c; Nickel et al. 2011; Boyle et al. 2011; Nishiguchi et al. 2012; Nickel et al. 2012; Hestbek et al. 2012; Hoang et al. 2012; Ho et al. 2012; Nickel & Busch 2013; Hoang et al. 2013; Muaaz & Mayrhofer 2013; Ichino et al. 2013; Hoang et al. 2015; Phan &

Dam 2015). The reported EER ranged between 6% and 20%, and the reported FRR ranged between 6.33% and 10.29%. The recognition rate ranged between 79% and 100%.

In contrast, less focus has been given to several studies proposing partly realistic strategies, such as different walking speeds (normal, fast and slow) (Sprager 2009) and the impact of different surfaces (ground, stones plates, gravel, grass, etc.) and different shoes (Sprager & Zazula 2011; Muaaz & Nickel 2012; Wolfe 2013). Also, researchers have studied the effect of holding the phone in different places (Watanabe 2014; Ottomoeller 2014; Watanabe 2015). Finally, in terms of researcher concentration, the realistic activity can be considered a very new and limited approach such as climbing stairs, jogging, running, sitting, standing, opening the door and walking around corners (Frank et al. 2010b; Frank et al. 2010a; Kwapisz et al. 2010; Kwapisz et al. 2011; Nickel et al. 2011; Derawi & Bours 2013). However, the reported EER fluctuated because of the noisy data resulting from the influence of different conditions. Moreover, the participants were joined with an observer who recorded the labels as the activities were being performed, which was considered non-realistic.

3.5 Conclusion

In recent years, several studies have focused on smartphone-based biometric gait authentication. It should be noted from the above literature review that there has been a dramatic improvement in their level of performance. This improvement accrued as a result of the development of smartphone devices with built-in sensors and constructing more sophisticated feature vectors and better recognition algorithms (e.g., artificial algorithms). Thereby, they prefer better recognition. It appears from the investigations mentioned above that the accelerometer-based biometric gait recognition is still a new field of research and most attention has been paid to the evaluation of using test data recorded under laboratory conditions. Limited studies have used an actual commercial mobile device to collect realistic data for variant gait signal such as climbing stairs, jogging, and running. Also, most of the studies used only the accelerometer sensor and very little works utilised two sensors. However, no research has been found that seeks to employ additional information in the process (such as GPS or weather info) to advance the state of knowledge and enable a better decision-making process. Furthermore, we have seen, in previous work, their experiments were (same day, cross day, and next day) with a limited number of users and restricted datasets. In simple comparison, none of the previous systems had attempted to cover a wide variety of data sets in seven consecutive days a week (i.e., study the potential for the general use in realistic circumstances).

4 Research Methodologies

4.1 Introduction

The use of gyroscope and accelerometer signals for gait authentication was investigated to provide an empirical basis for supporting its use in transparently authenticating users. The feasibility of using a wearable mobile device has increased because of increased demand for smartphone devices; however, the majority of previous studies were applied within a highly controlled environment (i.e., a present set of activities for the participant to undertake, such as walking on a flat floor at their normal pace). While this approach is suitable when first evaluating whether an approach has merit (i.e., discriminate information exists), it does not reflect the type of use one might expect in practice with a large number of variables playing a role that could impact the reliability of the approach. Very few studies have used actual commercial smartphone devices to collect real data for gait recognition (i.e., suffered from the absence of real-world datasets, which lead to verifying individuals incorrectly). In those studies that have, the volume of data and number of participants have been minimal. Therefore, the PhD research was focus on getting data that is richer and more experiential in terms of real-life experiences (i.e., free (uncontrolled) conditions) to assist in improving the validity. However, there is concern that real live data will be very noisy - it was the critical reason previous studies have focussed on particular well-defined activities. To assist, two main experiments were conducted:

- Control conditions experiment to largely duplicate previous studies; this helped to provide a baseline understanding performance and aid in direct comparison to the prior art. Multiple gait-based activities were collected in a controlled and separate manner (such as normal and fast walking speed, climbing up and down stairs, and carrying a bag in different days). These activities have been identified from the analysis of the prior art. The research examined a variety of activities (i.e., five types of walking activities) rather than doing a subset, to offer the opportunity to learn the users' walking behaviour across more realistic scenarios than simply walking under laboratory conditions. Consequently, this will help to determine how gait works in a wider set of activities, as the prior work was limited in that sense. This was framed into two phases. The first explored the classification performance of individual activities to understand whether a single classifier or an activity-based classifier (multi-algorithmic approach) would provide a better level of performance. The second phase explored the features vector (comprising of a possible 304 unique features) to understand the variability of feature vectors during differing tasks (walking with variable speed, stairs) across same and multi-day collections.

- An uncontrolled conditions experiment duplicated the control experiment phases with an entirely different data set. Real-life data was used to evaluate how well the approach works in practice. The first phase of the research work (activities identification), a human physical motion activity identification model, was built to classify a given individual's activity signal into a predefined class. A model was designed for identifying four types of unlabelled activities depending on the controlled experiment samples for each activity. These samples provided the basis for training multiple reference templates for each user, each template containing a specific gait activity.

This chapter represents the following novel and investigated aspects of this study:

- Gathering the largest controlled dataset containing different gait activities of 60 users over multiple days.

- Gathering a unique real-life dataset covering unconstrained data over seven days for 44 users.
- Discussing the devices and the software that are employed; methodologies were used to collect the datasets and categorise them.
- Focusing on the method of preparing the data to support the experiments mentioned above.
- Highlighting related work in the area of human activity identification using mainly smartphone sensors.
- Explaining the activities identification model.
- Investigating of the feature vector, time, and frequency domains feature vector extraction and dynamic section feature technique.
- Exploring the multi-algorithmic approach for classification.
- Introducing novel techniques that mainly focus on utilising the use of real-life uncontrolled data.

4.2 Research Methodology

Choosing and deciding on the research methodology is important as this leads to finding the correct answers for research questions accurately and precisely. Conversely, inadequate selection leads to an inaccurate response to research questions and queries. Generally, there are four research methodologies: quantitative, qualitative, pragmatic (mixed approach), and the advocacy/participatory approach (Morgan 2007).

- Qualitative research includes collecting and altering or converting data into numerical values. Therefore, a statistical scheming can be made and conclusions strained. This method has a process in which the researcher has to present one or more hypotheses. Hypotheses are questions researchers want to address; they include guesses around possible relationships between the elements that need to be investigated (variables). Data are posed by various means following a firm procedure and organised for statistical analysis. For instance, there are online surveys, mobile surveys, paper surveys, face-to-face conversations, telephone interviews, longitudinal studies, online elections, and systematic observations (Creswell, 2013). Objectivity is very significant in qualitative research. Accordingly, researchers take reasonable care to avoid their occurrence or presence, behaviour or attitude from influencing the results.

- Quantitative research is typically related to the positivist/post positivist pattern. It is mainly investigative research, and it is used to achieve an understanding of essential reasons, opinions, and motivations (Given, 2008). Quantitative analysis is usually related to the social constructivist model, which highlights the socially constructed nature of reality. This approach is used to measure the problem by way of producing numerical data or data that could be transformed into statistics. It is used to quantify attitudes, opinions, behaviours, and other defined variables and generalise results from a large population sample (DeFranzo, 2011). This method is about recording, investigating, and trying to uncover the hidden connotation and consequence of human behaviour and experience, with conflicting beliefs, behaviours and emotions. Researchers using this method are concerned with acquiring a rich and complex understanding find tolerant of people's knowledge and not in gaining information that could be generalised to other larger groups (e.g., individual interviews and participation/observations).
- The pragmatic research approach accepts ideas to be relevant only if they support action. Pragmatics "recognise that there are many different ways of

interpreting the world and undertaking research, that no single point of view can ever give the entire picture and that there may be multiple realities" (Saunders, Lewis, & Thornhill, 2012). It involves using the method that looks best suitable to the research problem and not getting trapped up in philosophical arguments about which is the best approach.

- The advocacy/participatory approach is sometimes called (emancipatory) researchers adopting "an advocacy/participatory approach feels that the approaches to the research described so far do not respond to the needs or situation of people from marginalised or vulnerable groups. As they aim to bring about positive change in the lives of the research subjects, their approach sometimes described as emancipatory" (alzheimer-europe.org, 2009).

The qualitative approach is utilised as a primary method in this research; furthermore, the participants in this study were comfortable with this approach.

To achieve experiments mentioned above effectively, two types of procedures were conducted: a control experiment that investigates different classifier strategies and real-life user's gait signal; both kinds of datasets were collected locally from the mobile device itself. Users' gait signal was continuously gathered from the accelerometer and gyroscope as long as the user walked or was doing his/her different types of gait activities. This mainly aimed to explore the following related aspects.

- Understanding the performance of gait recognition under both controlled and uncontrolled environments.
- Investigation of the feature vector and its impact on performance.

- The comprehensive set of activities to understand how gait recognition can operate and how better performance could be achieved across a complete set.

4.3 Technology Assessment

As clarified from previous studies presented in the literature review survey, Table 3-9, sensors that were used in early studies were limited. Amongst these sensors, very limited studies were carried out using the gyroscope information, despite it being able to reveal additional details and more features that might aid in the decision-making process. This study used accelerometer and gyroscope readings to explore the efficiency of these two sensors within the Transparent Authentication System (TAS). Accelerometers measure linear acceleration, which is a different physical measurement from the device orientation rate measured by gyroscopes (Heng et al. 2014).

To select an appropriate smartphone and application, this should be installed on the mobile phone for the data collection phase. Two phones were considered: the Samsung Galaxy S6 (32GB) smartphone and the Motorola G5 Inch 13MP (8GB) Android mobile phone, as a result of the wide range of built-in sensors (e.g., accelerometers, gyroscopes, barometers, gesture sensor, GPS, heart rate monitor, and proximity sensor) (Carphone warehouse, 2018). In addition to its lighter weight with more significant storage that is enough to extract the real gait signals for seven to fourteen days, the Samsung Galaxy S6 smartphone was employed to gather individuals' data. Seven third-party applications were evaluated in order to select the most suitable one for the gait signal acquisition. Table 4-1 shows the tested software that was reviewed after theoretical examination of several days experiment with sensors data that could be extracted by a mobile application when installed on the smartphone.

#No.	Sensor	Application Name						
		Sensor Tracker	Galaxy Sensor Explorer	Sensors Multitool	AndroSensor	Android Sensors	Sensorstream MU+GPS	Sensor Monitor
1	Accelerometer	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2	Gyroscope	Yes	Yes	Yes	Yes	Yes	Yes	Yes
3	GPS Position	Yes	No	Yes	Yes	Yes	Yes	No
4	Orientation	No	Yes	No	Yes	No	Yes	Yes
5	Gravity	No	Yes	Yes	Yes	No	Yes	Yes
6	Magnetometer	Yes	Yes	Yes	Yes	Yes	Yes	Yes
7	Pressure	Yes	Yes	No	Yes	Yes	Yes	Yes
8	Light sensor	Yes	Yes	Yes	Yes	Yes	No	No
9	Relative	Yes	Yes	No	Yes	No	No	Yes
10	Temperature	Yes	Yes	No	Yes	No	No	Yes
11	Proximity	Yes	Yes	Yes	Yes	Yes	No	No
12	Elevation	Yes	No	No	No	No	No	No
13	Speed	Yes	Yes	No	No	Yes	No	No
14	Sound	No	No	No	Yes	Yes	No	No

Table 4-1: The tested software with sensors that could be extracted by the smartphone

From the table, the accelerometer, gyroscope, GPS position and orientation sensors were the most important for the research. The accelerometer and gyroscope were the main sensors for extracting the gait signal. The orientation was the physical position of the mobile phone, which was used to determine the position of the phone in the pouch or pocket. GPS can be used for the context-awareness purpose. Figure 4-1 shows that the Sensor Tracker, Galaxy Sensor Explorer and AndroSensor which were used to gather information from most of the sensors. Nonetheless, the Galaxy Sensor Explorer does not have GPS positioning; therefore, the Sensor Tracker, and AndroSensor were considered for use.

Figure 4-1: (a) and (b) show Sensor Tracker and AndroSensor, respectively

The AndroSensor resolution was higher than the Sensor Tracker for the gyroscope and accelerometer. Additionally, this application was developed to record the biometric gait samples from the sensors and store them in comma-separated value (CSV) format on the participants' devices' local storage in order to analyse them later. Therefore, the AndroSensor smartphone application was adopted in order to be installed on smartphone devices for the research data collection process.

The Google Android OS was employed, as it is open-source and easy to use. Both an Android Samsung Galaxy S6 smartphone and the 'AndroSensor' application were able to capture reliably the related signal information required for a real data collection; therefore, it was decided to use the mentioned device and app. There was no need for any modification on the device's OS/applications before, during, or after the collection of data because the software collected the data most satisfactorily. To start recording sensor data, the participant merely needed to click on the application 'AndroSensor', swipe their finger, and tap the record button to started recording then end the recording by the end of the day.

4.3.1 Preliminary Testing

Prior to the engagement in activity application, the application was installed on the smartphone to examine the application functionalities. Three participants were asked to walk normally on a predefined route (along a flat corridor). The accelerometer and gyroscope signals were continuously collected during his/her walking to investigate the signals extracted from the mobile device using the software installed. From the graphical representation of the signals derived from the accelerometers X, Y, and Z-axis and the gyroscope's X, Y, and Z-axis, they show a good level of discrimination between the three users that participated as illustrated by Figure 4-2 and Figure 4-3.

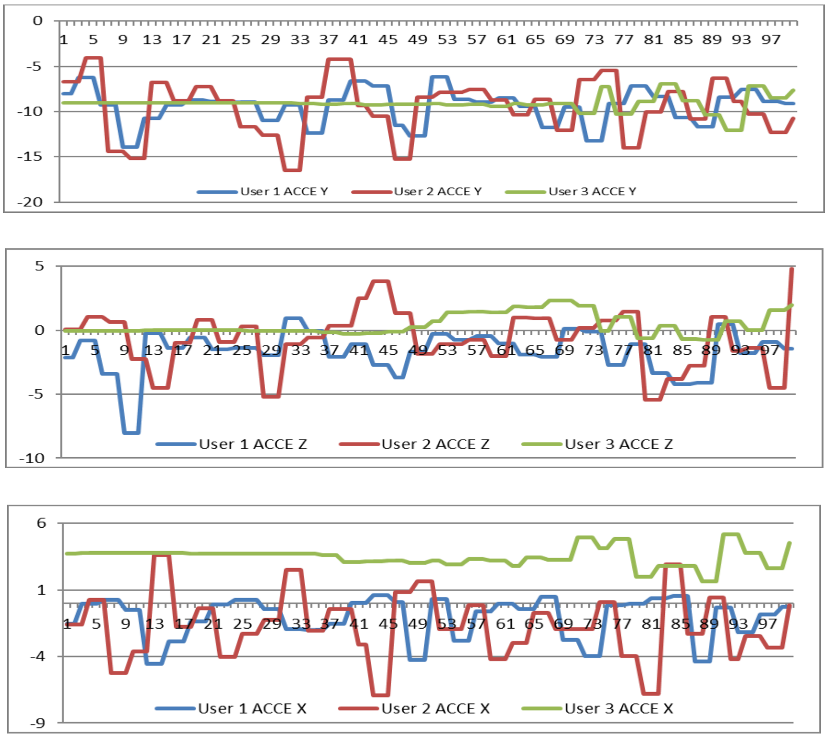


Figure 4-2: Illustrates the accelerometers X, Y and Z-axis signals of three users

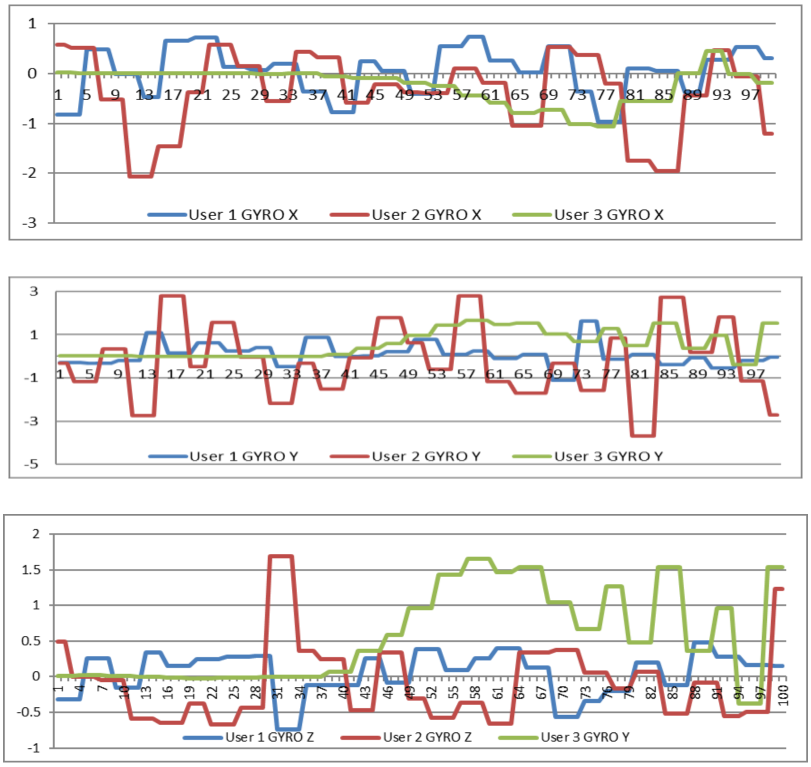


Figure 4-3: Illustrates the gyroscopes X, Y and Z-axis signals of three users

After testing the functionality and reliability of the mobile and the software signal for both the accelerometer and gyroscope sensors and guaranteeing that the application was working correctly and as required, ethical approval was attained from the university's research ethics committee (Appendix A), and participants were needed and invited by sending emails through Plymouth University's internal communication and the Plymouth Conservatoire web page to collaborate in this biometric data collection experiment.

In order to enable analysis, the targeted total number of themes was 60 for the control experiment and 44 for the uncontrolled experiment, as a minimum which was considered a satisfactory baseline grounded on other preceding research using similar sample sizes. The participants were instructed about the environment of this research, which was monitored by giving them the agreement form at the beginning of the examination (Appendix B) should the participant wish to complete the investigation.

4.4 Control Conditions Experimental Methodology

In addition to the primary research objectives, the researcher mainly aimed to explore the following related aspects with the controlled dataset:

1. Evaluate the performance of gait recognition across a wide range of walking activities;
2. Investigate the reliability of both accelerometer and gyroscope sensors;
3. Investigate the effectiveness of time and frequency domains-based features on system performance;
4. Explore the impact of dynamic feature selection techniques and the value of the feature space on the performance for different activities;
5. Investigate whether a multi-algorithmic approach is more viable than a single classifier approach;
6. Investigate the impact of static vs dynamic feature vectors;
7. Investigate the most discriminative features of each activity;

Aiming to contribute to the field of smartphone security, a comprehensive evaluation of users' gait biometric signal across a wider range of user walking activities has been conducted. The research examined a variety of activities to offer the opportunity to learn the users' walking behaviour across more realistic scenarios than simply walking under laboratory conditions. Consequently, this will help to know how gait works in a wider set of activities, as prior work is limited in that sense.

To create and evaluate the effectiveness of a more significant feature vector, the research provided a complete evaluation, including an analysis of motion sensors (accelerometers and gyroscopes). An investigation and analysis of the effectiveness of the time and frequency domain features on the system performance, understanding the variability of feature vectors during differing tasks (walking with variable speed, stairs) across same and multi-day collections were conducted. Furthermore, the impact of the dynamic feature selection technique (i.e., the dynamic feature vector contains the most distinctive features for each user) was explored, which successfully reduced the feature vector size and enhanced the performance for different activities.

This is framed into two experiments involving five types of activities: normal, fast, with a bag, downstairs, and upstairs walking. The five activities were derived from an analysis of prior work. These activities have been identified and rather than do a subset, we did all the five activities. However, no other studies have ever done that. Firstly, the motion sensor analysis experiments focused on the classification performance of individual activities against all through using a multi-algorithmic approach to classify individual activities (separate the classifiers depending on activities), and then a combination of all activities was verified. Secondly, the feature vector experiment was focused on discovered the feature vector (comprising of a possible 304 unique features) to understand how its composition affects performance.

To collect a more distinctive walking style, the phone must be placed close to the body. Otherwise, much noise might be collected randomly (Muaaz and Mayrhofer, 2015). Accordingly, people need to always wear trousers with “not-too-loose” front pockets. On the other hand, the controlled experiment was in two days and a real-life test was carried out for seven days, including the weekend and the participants needed to be free to choose clothes. Consequently, one of the more practical ways was that the smartphone needed to be put in the belt pouch around the waist ‘upside-down facing the body’ while the data was continuously collected during their movements, as shown in Figure 4-4 (a) and (b). That guaranteed the devices were always placed in a fixed place and orientation.

(a) (b)

Figure 4-4: (a) and (b) the phone is placed in the right or left belt pouch and its orientation

4.4.1 Data Collection

During the data collection process, users were asked to walk normally, fast, and normally with a bag on a predefined route (along a flat corridor) for a period of three minutes for each activity; this was followed by walking downstairs for three levels and upstairs for the same three levels, which resulted in a total number of 126 steps (63 for each direction). Between each activity, the participants were asked to stop for 15 to 20 seconds to rest, as well as to later separate the generated signals into their corresponding activities. As illustrated in Figure 4-6, the period of inactivity can be seen visibly between two activities as a more or less flat line. For a more realistic scenario, the participant had to stop in order to open the door and walked along the corridor back and forth many times for three minutes for walking, faster walk, walking with a bag, downstairs and upstairs activities. Ten sessions of activities were collected per user: five sessions were from one day, and the other five sessions were collected a week later. The users were free to change their footwear and clothes for the second day's data collection. In total, 60 users participated in the data collection exercise. Soft biometrics (i.e., age, gender, height, weight) were gathered in addition to gait pattern behavioural characteristics; 35 participants were male, and 25 participants were female, and they were aged between 18 and 56. The participants' weight was between 42-101 Kg and height between (145-187) cm. The age and gender demographics of the users are shown in Table 4-2 below.

Age	<20	21-30	31-40	41-50	51-60	Total
Male	3	6	20	3	2	34
Female	3	11	8	4	0	26

Table 4-2: Age and gender distribution of participants

The accelerometer and gyroscope gait data were recorded along the x, y, and z-axis, respectively; this means that six different signals were captured for the chosen activities. Figure 4-5 shows the accelerometer and gyroscope x-axis signals being generated from the following activities, normal, fast, with a bag, downstairs, and finally upstairs walking.

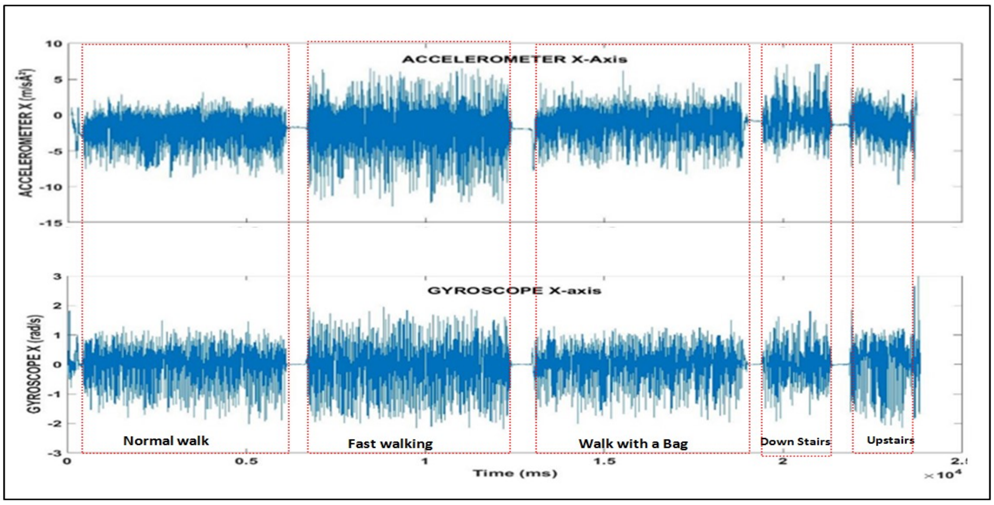


Figure 4-5: The Accelerometer and gyroscope raw gait data recorded along X-axis

Upon completing the data collection process, user's activities were divided into five files for each activity (namely: normal walk, fast walk, walk with a bag, downstairs walking, and upstairs walking) as explained in Section 4.4.1.1 below. Then the tri-axial raw accelerometer and gyroscope signals were segmented into 10-second segments by using a sliding window approach with no overlapping. As a result, 74 samples were collected for each user per day. In total, 8,880 samples were collected for the entire control conditions dataset (60 users, across two days) as shown in Table 4-3.

Activity Type	# Samples
Normal	2,640
Fast	2,640
Carrying a Bag	2,640
Down Stairs	480
Upstairs	480
Total Samples	8,880

Table 4-3: Activity states for all users across two days

To validate the effectiveness of the created features for authentication methods, the datasets were collected in two scenarios; same-day (SD) and cross-day (CD). In the SD scenario, the dataset split in 60-40: 60% of the data was used for the classifier training and the remaining 40% was utilised for testing. In the CD scenario, the first-day data was used for training and the second-day data was used for testing. For each scenario, all users' gait activities were treated as a single dataset; then each activity was studied individually (i.e., a multi-classifier created to every single gait motion type (e.g., walking, running, walking with a bad etc.)), as shown in Figure 4-6.

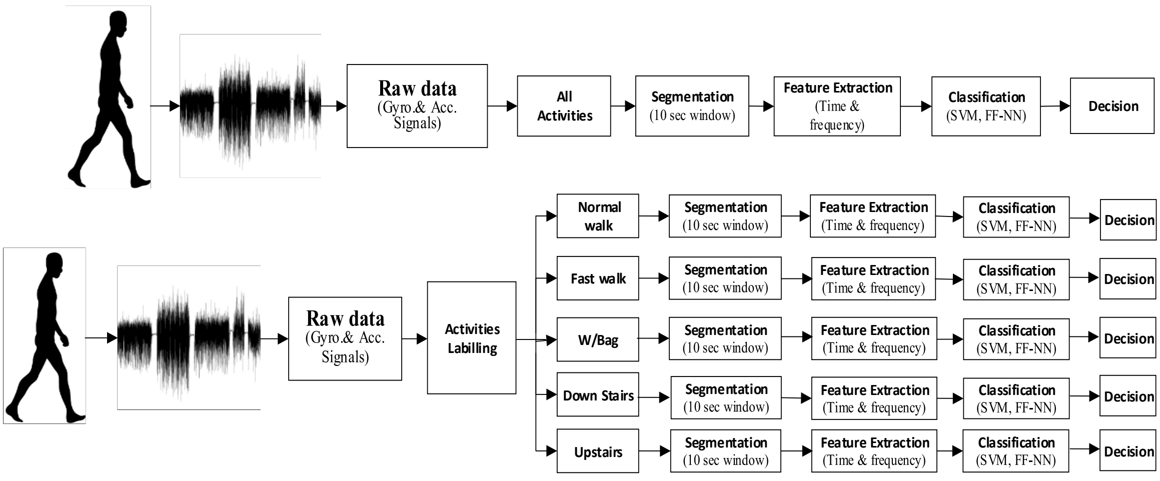


Figure 4-6: Authentication Process Systems. Each activity investigated individually then all users' gait activities were treated as a single dataset

4.4.1.1 Activity labelling

After the data collection phase, the user's gait signal log file was extracted from the mobile phone and then divided into five activity files in the CSV format for further data processing and analysis. As shown in Figure 4-5, the signal pattern of the chosen five activities is clearly different from the pattern of the standing activity. As a result, the setting was three minutes as an initial time for each activity and the space of standing period among them (i.e., 20 seconds). An algorithm for activity extraction was devised. The algorithm read and analysed the (.CSV) file depending on activity time and the break was started with setting up an activity type parameter to 1, then for each file there was a loop to read one record and save the record with the current data file during the time in a range of activity. Otherwise, go to the next file and stop the current step and as in break time will keep reading the main file without saving record until break time finish; followed by saving and close the existing file. The algorithm kept doing the same process for the next activity types until the end of the (.CSV) data.

4.5 Uncontrolled Condition Experimental Methodology

The research also sought to explore how users' gait signal was intelligently utilised for authentication for real and live free use of the mobile device. Accordingly, the work aimed to explore the following related aspects with uncontrolled conditions dataset:

1. Explore the accuracy of the user authentication designed model by analysing real-life mobile sensor signals;
2. Investigate the impact of static vs dynamic feature vectors;
3. Investigate the most discriminative features of each activity;
4. Evaluate the performance of gait recognition across different walking activities identify from a real and live unconstrained use of the smartphone signal;
5. Investigate whether the multi-algorithmic approach is more viable than a single classifier approach through real-life use;
6. Highlight the influence of the majority voting technique on system accuracy;

The eventual aim of uncontrolled (real-life) gait recognition research work is to shape a new approach to transparent authentication. The majority of the previous transparent authentication system (TAS) outlines based on controlled environment data (i.e., not real-world data). Also, it is envisaged that the real practise and system performance might differ from those that are obtained under controlled environments and limited numbers of users, samples, and prearranged tasks. Consequently, conducting an extensive evaluation for the previous work would be ideal for the research to build on. Thus, it would be wise to clarify the control study with live usage and unrestricted data, and a bigger dataset and number of participants to investigate and understand the actual performance in practice with real-world dynamic data. Furthermore, the findings from the unrestricted experiments could provide more accurate and fair insight into the performance evaluation.

A data gathering process was required to create a real dataset with a reasonable number of people during real-life gait activities (i.e., completely free activities) over a significant period. The hardware and software that were used for collecting user's gait data in the real-life environment were the same as the ones utilised for a controlled environment (described in detail in 4.3). The individuals' gait samples were collected continuously and transparently and stored on their devices' local storage. In total, seven days of user gait data were collected as the amount of unconstrained data over seven days was deemed to be sufficient and the relatively shorter timeframe could have attracted more participants to take part in the data collection exercise. Five smartphones were purchased; that means five users participated at a time. The average days of data collected were seven days per user. Accordingly, the approximated time taken for data collection from 44 users was nine weeks.

Participants were not given any specific task to carry out during the seven-day experiment period. Moreover, no particular constraints or conditions were specified, so data was collected in an uncontrolled manner. That was confirmed to be followed to guarantee meeting the stated requirements of the aimed dataset being a real picture of natural practice patterns. Accordingly, once participants accepted the consent, the application was installed on Android smartphones and given to them with the belt pouch. The gait biometric data samples were captured and stored as CSV files on the local storage of participants' smartphones. The users were also asked to synchronise their smartphones with the Dropbox cloud storage service to share the data automatically.

4.5.1 Real-World Data Collection

In the uncontrolled condition experiment, the real-life data was used to evaluate how well the experiment one approach worked in practice. It was anticipated that the use of gyroscope and accelerometer signals without additional information would be sufficient for reliable authentication.

As mentioned in Section 4.4, the control conditions experiment investigated the performance of transparent gait verification. However, this experiment used the identical experimental setting because it was believed that these parameters and classifiers may differ when they applied in such a dataset of real and uncontrolled live usage data with all illuminated varying conditions (changing clothed and shoes, in a rush, carrying luggage, running because of poor weather, exercising, to name but a few).

A user purely needed to put the smartphone in the belt pouch while his/her data was continuously collected at a rate of 30-32 samples per second for the x, y, and z-axis of both the accelerometer and gyroscope sensors. They were asked to start recording by the 'AndroSensor' application every day and to stop recording by the end of the day, for the purpose of automatic data sharing and storing the data daily in cloud storage services in .CSV file format, which contained raw and sensor data. The user's needed to synchronise with a cloud storage application service (e.g., Dropbox). If the transfer failed for any reason, it will be retried the next time.

Then the uncontrolled tri-axial raw accelerometer and gyroscope signals were segmented into 10-second segments by using a sliding window approach with no overlapping as provided in 4.6 (signal pre-processing). Then a significant feature vector of time and frequency domains were extracted and analysed for both motion sensors. Afterwards, a predictable model was designed that was able to classify a given individual's activity signal into a predefined class belonging to, based on the features extracted from the raw sensor data collected from the control environment, as training data (e.g., normal walk, fast walk, walk with bag, downstairs, upstairs, and sitting).

The previous controlled experiment hypotheses were duplicated with real-life data over a significant period. It was almost replication to what was done before but using an utterly different dataset. In other words, exploring to what extent the multi-algorithmic approach and variability of feature vectors (dynamic feature selection technique) during differing tasks (walking with variable speed, stairs) were reliable compared to the single classifier approach.

4.5.2 Activity Identification

The purpose of smartphone-based activity recognition is analysing the continuous inertial sensor data and identifying the actions carried out by a person (Poppe 2007). The activity recognition procedure has become a vital process in determining what activity a user is doing while the people perform a different set of activities in various environments, whether in laboratories and in real life. The raw signal of the inertial smartphone sensors (in the controlled or uncontrolled real-life settings) was employed for recognition of human life activities.

In this research, the activity recognition process was essential to classify different gait activities from the real-life smartphone signal. Therefore, a process was required to be able to identify what activity a user was doing to enable the selection of the correct classifier. This improved the potential to develop more specialised activity-based classifiers (multi-algorithmic approach). Consequently, predictable data modelling was built to classify a given individual's activity signal into the predefined class it belongs to, based on the features extracted from the raw sensor data (in this study, normal walk, fast walk, walking with a bag, downstairs, upstairs, and sitting). That was used in the advanced authentication phase.

The activity identification model was shared with another researcher, including preparing the data, the pre-processing steps, time and frequency domains feature vector extraction, and the activity identification model.

4.5.3 Activity-Based Recognition

It is evident from the literature in Table 4-4 that for activity identification prior research has suggested various approaches to identifying different activities. However, these latest studies amongst the others were evaluated based on using mainly smartphone sensors, so most of the systems experimented on same-day data, a small number of users, short durations, specified tasks, and controlled environments. Moreover, to the best of the author's knowledge, there is an apparent lack of realistic data, which was considered a significant barrier that prevented applying activity recognition in practice. Therefore, this study presents a real-world unconstrained environment over a reasonable period (i.e., real live movement training and testing data over a continuous seven days). Accordingly, there is a need to implement this model with real data to comprehend how they work in practice.

In Kwapisz et al. (2010), the study used a neural network to model human activity and achieved high accuracy (CCR 100%) in identifying the correct class to which the activity signals belonged. However, the limited number of population samples (i.e., 5-30 users) opens the possibility that the learned algorithm is overfitted and has memorised the training samples.

Other studies (Anguita et al. 2012; Ganti et al. 2010; Nakano 2017; Bhanu Jyothi & Hima Bindu 2018; Ogbuabor & La 2018a) have used a sliding window approach with an overlap of 50% in segmenting the raw activity signals. This could, however, lead to an overlap in the subsampling between the training and testing sets, which means that unless the splitting of the two sets occurs before the segmenting of the raw data, the data are only partially seen by the learning algorithm in both the training and testing sets. In terms of the correct classification rate, it can be seen that SVM, neural network, and CNN achieve the highest performance among the techniques shown.

Study	Approach	Performance (CCR %)	Population	Activity Type
(Kwapisz et al. 2010)	NN	100	5	Standing, sitting, walking, jogging, downstairs, upstairs
(Anguita et al. 2012)	SVM	89	30	Standing, sitting, walking, lying down, downstairs, upstairs
[16]	SVM	96		
(Nakano 2017)	CNN	90		
(Bhanu Jyothi & Hima Bindu 2018)	RF PCA	94 89		
(Ogbuabor & La 2018a)	MLP	95		
(Jiang & Yin 2015)	CNN	99	10	Standing, sitting, walking, jogging, running, biking, downstairs, upstairs
(Heng et al. 2016)	SVM	85	5	Standing, walking, running, upstairs, downstairs
(Saha et al. 2018)	Ensemble	94	10	Sitting on a chair, sitting on the floor, lying right, lying left, slow walk, brisk walk

Table 4-4: Comparison of prior studies in activity recognition using smartphone sensors

In this study, a segment-based approach was used to extract features from raw sensor signal data with a sliding window of 10 seconds with no overlap. The extracted features were used to compute various statistical features, such as the mean, median, maximum, and minimum of a given sensor axis within a specific segment window.

4.5.4 Activities Identification Data Modelling

Data modelling aims to build a predictive model that can classify a given individual's activity signal into a predefined class, based on the features extracted from the raw sensor data (e.g., normal walk, fast walk, walking with the bag, downstairs, upstairs, and sitting), as illustrated in Figure 4-7.

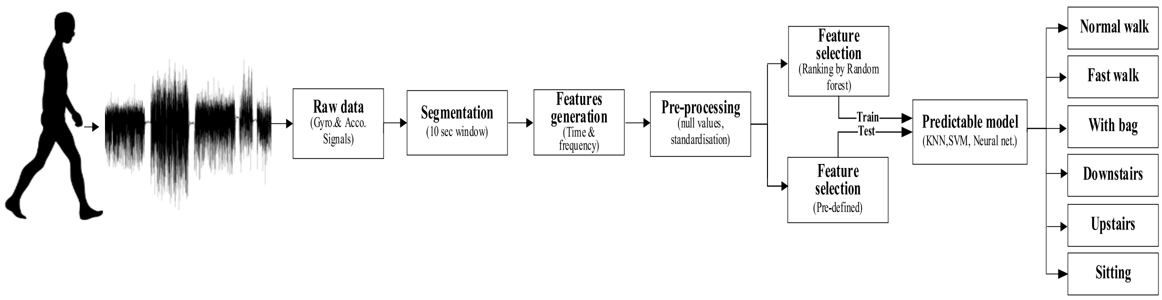


Figure 4-7: Activities Identification Model

The overall process starts by capturing the raw activity signals from smartphone sensors, followed by the subsequent steps undertaken to pre-process the data and form the model:

- **Data pre-processing**

Two approaches (i.e., normalisation and standardisation) were examined for transforming data. The dataset was normalised by scaling the input vectors individually to the unit norm (vector length). The other transformation approach was to standardise the features by removing the mean and scaling to the unit variance. The latter approach (standardisation) emerged as better than the former (normalisation) in discriminating the activity samples for the tested dataset.

- **Feature importance analysis** (Ranked Features)

To reduce the feature vector dimensions, only those ranked as being of higher importance in contributing most effectively to discriminating individuals' activities by the random forest algorithm were included in training the predictable model. The variable importance measure of the random forest calculates how significantly a given feature is biased towards correlated predictor variables (Strobl et al. 2008). Feature importance analysis using random forest reduced the feature vector from 304 to 190 features in the final model. Reducing the feature space dimensionality not only improved the overall model performance but also lowered the probability of the algorithm being over fitted to the training data.

- **Train/Test splitting ratio:** for training the base model, the controlled data were split into 60/40 training and testing sets, respectively. Once the best model was chosen (the one that achieved the highest performance), the model was retrained using all the controlled dataset for training the final model, which was used to predict the uncontrolled activities.
- **Classification Modelling:** several supervised classification algorithms were examined. Finally, three algorithms were the best candidates for the ensemble; these are feedforward neural network (FF-NN), SVM, eXtreme Gradient Boosting (XGB). By using the ensemble model, it improved the overall accuracy compared to a single model-based approach. Two ensemble methods were examined:

   i. ***Hard voting,*** if two algorithms (out of three) agreed for a given activity.

   ii. ***Soft voting,*** the probability of a given sample that belongs to a specific class was averaged using the mean among the three classifiers, and the highest rank is chosen.

The soft voting approach was slightly better than the single algorithm and hard voting approach. Figure 4-8 exhibits the real-world gait data identification and recognition model. Chapter 6 provides more details about the activity identification results.

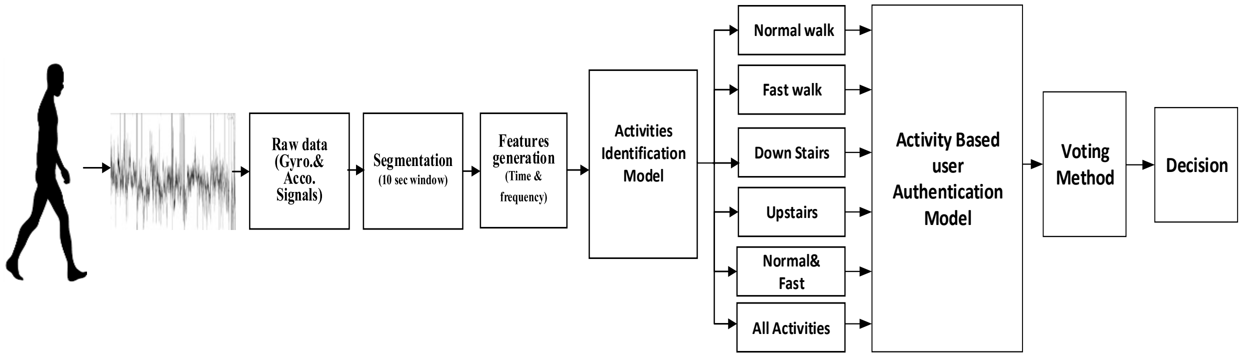


Figure 4-8: Real-World gait data identification and recognition

Table 4-5 shows the activity states for each user for one week. Moreover, Table 4-6 provides the summary statistics for data encapsulated from the 44 contributors, which could be considered deep enough to enable a significant analysis. Even if the engaged subjects were asked to let the software run for at least seven days, some had more the demanded period and a little less. There have not been any researches examining this real dataset (to the best of our knowledge).

It is hard to expect the per cent of the data included in one day; every day has a different number of samples depending on how much each user walked. As shown in Table 4-5, the normal walking (i.e., including carrying a bag) samples reported significantly more than the other types of gait walking activities, about 80% of the total gathered samples. As expected, people mostly walk normally. While fast walking was only 7%, down and upstairs were 3% and 10%, respectively. Concerning other types of samples, which were classified as non-gait activities, has the most proportion of all gathered samples (i.e., 67%) whereas people were doing different scenarios in their life except gait activities. To name but a few, there was sitting, standing, shopping, and using various means of transport (i.e., car, bicycle or motorbike) and fitness.

#User	#Days	Normal	Fast	Downstairs	Upstairs	Other
1	11	2,864	62	1,027	240	18,749
2	9	3,139	66	1,003	195	14,706
3	8	4,450	162	19	713	18,146
4	10	1,418	205	57	202	19,545
5	7	2,087	28	11	96	3,825
6	9	2,678	501	56	521	17,233
7	10	4,979	76	120	844	24,547
8	10	4,224	114	63	386	23,634
9	9	2,192	310	68	258	31,758
10	8	1,504	168	113	304	14,644
11	9	1,999	293	86	216	11,079
12	7	1,607	334	153	238	15,391
13	8	2,940	245	55	357	11,941
14	7	1,830	276	51	75	9,133
15	6	1,726	95	16	10	628
16	6	2,653	41	17	47	6,447
17	6	2,464	42	46	21	12,232
18	6	3,038	402	169	326	8,098
19	10	3,429	95	1,668	930	20,419
20	7	2,634	502	224	102	14,923
21	6	1,627	20	18	125	10,545
22	6	2,398	72	37	309	9,637
23	7	6,267	522	205	1,319	18,517
24	7	4,398	626	266	762	12,073
25	6	1,830	111	39	229	7,925
26	8	6,335	208	179	431	6,756
27	6	4,628	264	118	595	1,783
28	6	4,582	554	265	538	20,571
29	7	3,349	474	99	116	9,207
30	7	1,163	56	49	236	5,651
31	8	4,572	261	44	388	15,472
32	6	2,939	615	454	197	18,439
33	6	1,290	181	8	54	11,155
34	7	4,178	136	75	243	6,389
35	6	2,964	515	105	234	7,574
36	7	2,325	644	181	486	21,055
37	7	4,584	272	185	256	10,210
38	6	2,962	717	38	98	11,189
39	6	2,634	18	11	479	5,351
40	7	2,654	106	7	950	9,985
41	8	4,562	190	33	705	9,481
42	6	3,925	320	34	379	18,799
43	7	3,960	1,067	95	388	11,629
44	9	3,407	256	83	469	19,968
Total	**325**	**137,388**	**12,222**	**7,650**	**16,067**	**576,439**

Table 4-5: Activities states for each user for one week

Total Number of All Users	44
Total Numbers of All Days	325
Average Number of Days per User	7
Total Number of Normal walk samples	137,388
Total Number of Fast walk samples	12,222
Total Number of Downstairs samples	7,650
Total Number of Upstairs samples	16,067
Total Number of Classified Other samples	576,439
Total Number of Gathered samples	749,766
The Number of Hours for the recognised normal, fast, down and upstairs activities	481.46

Table 4-6: The overview of the unconstrained dataset

4.6 Signal pre-processing and Feature Exploration

An examination tool was required to enable the pattern classification procedure of these studies. Consequently, the specialised mathematical modelling package was developed widely for the modelling and validation of the analyses of this research because of its common use and recognition right through engineering and scientific communities in the study of mathematical problems. The mathematical modelling is from MATLAB (R2016b release), developed by Math Works, which was employed on Intel Core i5-4310 CPU, 2.7 GHz and 16 GB RAM hardware and Windows 7 Enterprise 64-bit operating system.

In this study, some scripts were improved in order to perform a variety of functions to implement the experiments. The same as the control condition experiment, the data of each user was split into two subsets: 60% for training the classifiers and generating the user profile and 40% for validation and testing the performance. Given that, we considered one contributor acting as the valid legal user while the remaining other members as imposters. This was repeated to guarantee all users have the opportunity of acting as the authorised user. Results were then averaged across the population sample.

4.6.1 Segmentation

Once the raw gait signals were gathered, pre-processing could be started. According to the analysis of the literature review, the segment-based method outperforms the circle-based technique; as a result, the raw gait signals were divided into a fixed-length window. Obviously, the performance would differ when choosing various segments sizes; hence, an optimum segment length was selected based on the best results of the primary testing. Then the tri-axial raw accelerometer and gyroscope signals were segmented into 10-second segments by using a sliding window approach with no overlapping for both datasets. As a result, for experiment one, 74 samples were collected for each user per day. In total, 8,880 samples were collected for the controlled experiment dataset. Furthermore, approximately 174,713 samples were collected for the un-controlled dataset and the recognised samples for normal and fast activities were 139,907 and 12315, respectively.

4.6.2 Feature extraction

Regarding the feature extraction process, both time domain and frequency domain features were extracted from the users' accelerometer and gyroscope data segments. In total, 304 unique features were generated from both the accelerometer and gyroscope data samples. Details of those features are presented in the following sections.

4.6.2.1 *Time-domain features*

The time-domain features, which refer to variation of the amplitude of the signal with time, were calculated directly from the raw data samples. All the details of those features (including their names and descriptions) are demonstrated in Table 4-7 below:

Features	Description	Studies references
Mean (3)	The mean values in the segment.	(Nakano 2017; Kwapisz et al. 2010; Lu 2014)
Standard Deviation (3)	The standard deviation of the data in the segment.	(Nakano 2017; Kwapisz et al. 2010)
Median (3)	The median values of the data points in the segment.	(Nakano 2017)
Variance (3)	A measure of how far each value in the segment points is from the mean.	(Sprager 2009; Lu 2014)
Covariance (3)	A measure of how much two variables change together.	(Lu 2014; Bashir et al. 2010)
Zero crossing rate (3)	The rate value of sign changes in the segment.	(Derawi 2012)
Interquartile range	The range amidst the data. It is the distinction between the upper and lower quartiles in the segment.	(Nakano 2017)
Average Absolute Difference (3)	The average absolute difference between the value of each of the segment points from the mean value over the segment values (for each axis).	(Kwapisz et al. 2010)
Root mean square (3)	The square root of the mean of the squares of the acceleration values of the segment.	(Nishiguchi et al. 2012; Bajrami 2011)
Skewness (3)	A measure of the symmetry of distributions around the mean value of the segment.	(Nakano 2017; Lu 2014)
Kurtosis (3)	A measure of the shape of the curve for the segment point's values.	(Nakano 2017; Lu 2014)
Percentile 25 (3)	The percentile rank measured by the following formula: R= (P/100) *(N+1). Where R is the rank order of values, P percentile rank, N total number of the data points in the segment.	(Khandnor & Kumar 2017; Schneider et al. 2013)
Percentile 50 (3)	Similar to the Percentile 25feature; but with the setting of P=50.	(Schneider et al. 2013)
Percentile 75 (3)	Similar to the percentile 25 feature but with the setting of P=75.	(Schneider et al. 2013)
Maximum (3)	The largest four values of the segment are calculated and averaged.	(Nakano 2017)
Minimum (3)	The smallest four values of the segment are calculated and averaged.	(Nakano 2017)
Correlation coefficients (3)	The relationship between the two axes is calculated. The correlation coefficient is measured between X and Y axes, X and Z axes and Y and Z axes.	(Nakano 2017)
Average resultant acceleration (1)	Average of the square roots of the sum of the values of each x, y and z-axis in the segment squared.	(Kwapisz et al. 2010)
Difference (3)	The difference between the maximal and minimal value of the segment (each axis).	(Frank et al. 2010b)
Maximum value (4)	The largest four values of the segment are calculated and averaged.	(Hoang et al. 2013)
Minimum value (4)	The smallest four values of the segment are calculated and averaged.	(Hoang et al. 2013)

Binned distribution (30)	Relative histogram distribution in linear spaced bins between the minimum and the maximum acceleration in the segment. Ten bins are used for each segment.	(Kwapisz et al. 2010)
Maximum peaks (3)	The average of the largest four peaks in the segment.	(Nickel 2012)
Minimum peaks (3)	The average of the smallest four peaks in the segment.	(Nickel 2012)
Peak Occurrence (3)	Calculate how many peaks are in the segment.	(Nakano 2017)
The time between peaks (3)	Time in milliseconds between peaks in the sinusoidal waves associated with most activities calculated and averaged (for each axis).	(Kwapisz et al. 2010)
The interquartile range (3)	Calculating the median of the lower and upper half of the data.	(Nakano 2017)
Entropy (3)	The average amount of information produced by a probabilistic stochastic source of data	(Nakano 2017)
Energy (3)	The signal energy is equal to the summation across all frequency components of the signal's spectral energy density.	(Nakano 2017)

Table 4-7: Time domain features

4.6.2.2 *Frequency domain features*

The frequency domain feature refers to the analysis of mathematical functions or signals with respect to frequency, rather than time. There is dissimilarity in the feature extraction process between the time and frequency domains. As in the frequency domain, the data should be processed using a Fourier transform prior to the feature extraction process. Many frequency domain features were calculated in order to produce a unique feature vector; these frequency domain features are presented in Table 4-8, and the feature descriptions are included in Table 4-7 above.

Type and number features	Studies references	Features	Studies references
Entropy (3)	(Nakano 2017; Lu 2014)	Root mean square (3)	(Nishiguchi et al. 2012; Bajrami 2011)
Energy (3)	(Youn et al. 2014; Lu 2014)	Skewness (3)	(Nakano 2017)
Mean (3)	(Lu 2014)	Kurtosis (3)	(Lu 2014; Nakano 2017)
Standard Deviation (3)	(Nakano 2017)	Percentile 25 (3)	(Schneider et al. 2013)
Median (3)	(Nakano 2017)	Percentile 50 (3)	(Schneider et al. 2013)
Variance (3)	(Lu 2014)	Percentile 75 (3)	(Schneider et al. 2013)
Covariance (3)	(Lu 2014; Bashir et al. 2010)	Maximum (3)	(Nakano 2017)
Zero crossing rate Minimum (3)	(Derawi 2012)	Minimum (3)	(Nakano 2017)
The Interquartile range (3)	(Nakano 2017)	Correlation coefficients (3)	(Nakano 2017)
Average Absolute Difference (3)	(Kwapisz et al. 2010)	Average resultant acceleration (1)	(Nakano 2017)

Table 4-8: Frequency Domain Features

4.6.3 Normalisation

In this work, the normalisation approach referee to convert and translate the selected feature values, which were on different or unusual scales in the range of 0-1. This would obtain more effective performances, as well as the mathematical scheming could be faster (Sola & Sevilla 1997). This research involved dividing each feature of a vector by the maximum value of that vector.

4.6.4 Feature selection

To validate the effectiveness of the generated feature vectors (comprising of a possible 304 unique features), the data set was divided to form both reference and testing templates for all users in two scenarios (i.e., same and cross- day).

The impact of the time and frequency domain features on the system performance for accelerometer and gyroscope data was investigated and highlighted. Furthermore, the system performance was evaluated by using the most discriminative feature set (dynamic feature selection technique), and it was also assessed without involving the feature selection technique (i.e., using all feature sets) separately for each activity and all activities.

A large number of features would place a burden on the classification (particularly on processing/battery limited mobile devices). Therefore, a dynamic feature selection approach was devised that can select features based on their uniqueness for individual users. It was envisaged that the effectiveness of each feature towards the classification would vary, with some features having a more significant impact for some users over others. The dynamic feature selection mechanism selected features based on a calculation of the standard deviation of users' features with the smaller standard deviation being selected. Standard deviation was utilised because of the need to reduce the variability of the feature vector and to improve the permanence.

Once the feature vector was formed, it was forwarded to the next phase: either for training or testing purposes.

4.6.5 Classification

In the matching phase, the individual samples compared with the reference template were taken primarily at the setup phase (i.e., the feature vector that resulted from the feature extraction process, which was clarified in Section 4.6.2). Consequently, a match score was given indicating the degree of similarity, which decided acceptance of the users' verification claim based on what the authentication decision was.

As a result of the prior art and preliminary experiments, the support vector machine (SVM) and feedforward neural network classifiers were employed as the default classifier. The system performance was evaluated using the false acceptance rate (FAR), false rejection rate (FRR), and equal error rate (EER) metrics. These metrics were essential to be involved in comparing biometric modalities in the transparent authentication system (TAS).

An examination tool was required to enable the pattern classification procedure of these studies. Consequently, the specialised mathematical modelling package was developed widely for the modelling and validation of the analyses of this research because of its common use and well-recognition right through engineering and scientific communities in the analysis of mathematical problems. The mathematical modelling was from MATLAB (R2016b release), developed by Math Works, and was employed on Intel Core i5-4310 CPU, 2.7 GHz and 16 GB RAM hardware and the Windows 7 Enterprise 64-bit operating system.

In this study, some scripts were improved and in order to perform a variety of functions to implement the experiments. The same as the control condition experiment, the data of each user was split into two subsets: 60% for training the classifiers and generating the user profile and 40% for validating and testing the performance. In view of that, we considered one contributor acting as the valid legal user and the remaining other members as imposters and then repeated to ensure all users had the opportunity to act as the authorised user. Results were then averaged across the population sample.

4.7 Conclusion

This chapter aims to contribute to the field of smartphone authentication. Two experiments were designed, aimed at improving smartphone-based gait authentication, namely control conditions, which explored the classification performance of individual activities (such as normal and fast walking speed, climbing up and down stairs, and carrying a bag on different days) using a multi-algorithmic approach to classify individual activity and the uncontrolled condition experiment, which explored how well the real-life gait recognition approach worked in practice.

In the control conditions experiment, after separating the activities' raw signal, five files were ready: normal, fast, walking with a bag, downstairs, and upstairs. The signal of accelerometer and gyroscope sensors were segmented into 10-second segments by using a sliding window approach with no overlapping. The feature extraction process was carried out on accelerometer and gyroscope data. The number of time-domain features was 97 and the frequency domain was 55 features. In total, 152 features were created for each sensor and 304 features for both sensors. Two scenarios were implemented (same-day and cross-day) and SVM and feedforward neural network classifiers were used to evaluate the extracted features. For each scenario, the features were investigated for the time, and the same process was done for the frequency domain features, then both types of features were investigated.

For the pre-processing phase, 304 features were used as basic samples for training multiple reference templates for each user for different activities in the real-world model. A realistic gait data set was used to evaluate these approaches. The main results in this regard were presented in Chapter 5 and Chapter 6.

5 Experimental Results of Exploring Classification Strategies

5.1 Introduction

This chapter will explore the details of consecutive experiments that were conducted to investigate the uniqueness of users' gait within the controlled environment; the dataset used for the experiment study contains the gait activity of 60 users across multiple days. At the time of writing, the dataset was the most extensive dataset being used within the smartphone gait study domain.

This is through providing a comprehensive evaluation, including:

- An analysis of motion sensors (i.e. accelerometer and gyroscope) signals.
- An investigation and analysis of features understanding the variability of feature vectors during differing activities across a multi-day collection.
- The impact of dynamic feature selection for each user explored, which successfully reduced the feature vector size and enhanced the performance.
- An investigation of a single classifier and a proposal of the multi-algorithmic approach performance.

This is framed into two experiments involving five types of activities: normal, fast, with a bag, downstairs, and upstairs walking. The first experiment explored the features vector (comprising of a possible 304 features) to understand how its composition affects performance and more discriminative features for different activities were identified. The second experiment explored the classification performance of individual activities to understand whether a single classifier or multi-algorithmic approach provides better performance. Both tests were investigated with SD and CD evaluation using accelerometer and gyroscope sensors.

5.2 Investigation of Feature Vector Composition

The following section of the experiments was conducted to address the core research questions which are related to the first dataset (i.e. within the controlled environment). Whereas there were sixty participants employed, each user's data was split into two subsets: 60% for training the classifiers and generating the user profile and 40% for validating and testing the performance.

5.2.1 Investigating the Sensors and Feature Vectors

Several experiments were conducted to evaluate the proposed system by examining the reliability of the accelerometer and gyroscope sensors individually. Then results from both sensors were analysed in the next trial as follows:

- The effectiveness of the time and frequency domain-based features on the system performance was investigated.
- The impact of the dynamic feature selection technique and the value of the feature space on the performance across various activities (i.e. normal, fast, walking with a bag and walking up and downstairs) were explored to investigate the best number of feature subset (NF).
- Two evaluation scenarios (i.e. same day and cross day) were tested; the SVM classification algorithm was used across a combined five gait activities as its execution time is relatively shorter than that of the neural network classifiers.

To achieve this, the following experiments were implemented:

1. **The Accelerometer Data Exploration:** The first experiment was conducted to analyse and highlight the impact of the accelerometer data samples using the

time and frequency domains feature on the system performance by involving the proposed dynamic selection technique (discussed in section 4.6.4). Figure 5-1 and Figure 5-2 presents the results achieved under a complete set of experiments involving various feature vector lengths. Table 5-1 provides the best performance on users' accelerometer data for all activities achieved under a complete set of experiments involving multiple feature vector lengths (under both the same day (SD) and cross day (CD) scenarios).

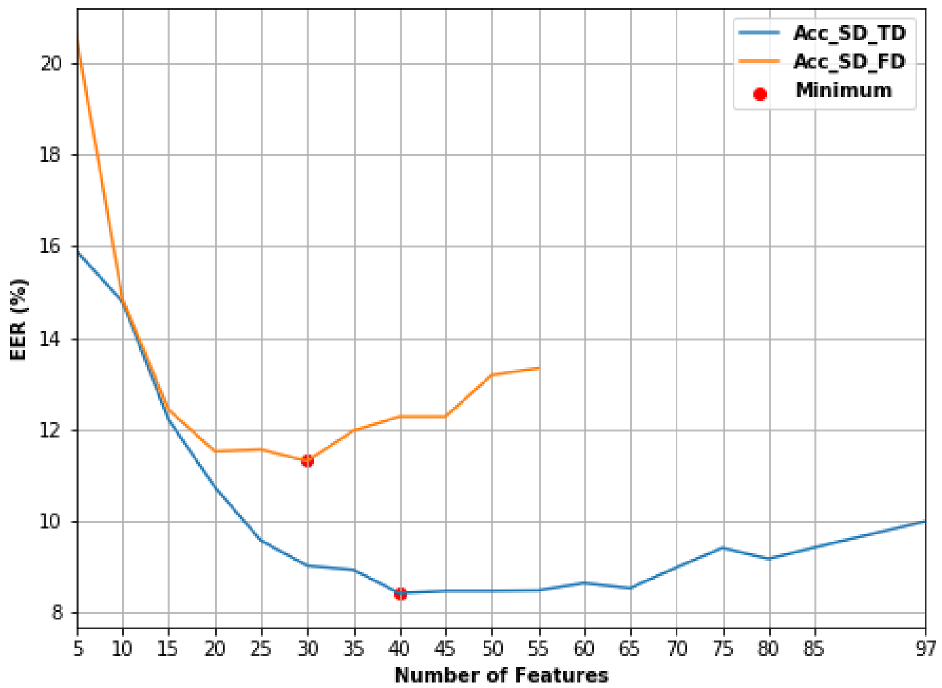


Figure 5-1: The EER results on Accelerometer data for all activities by using SD scenarios

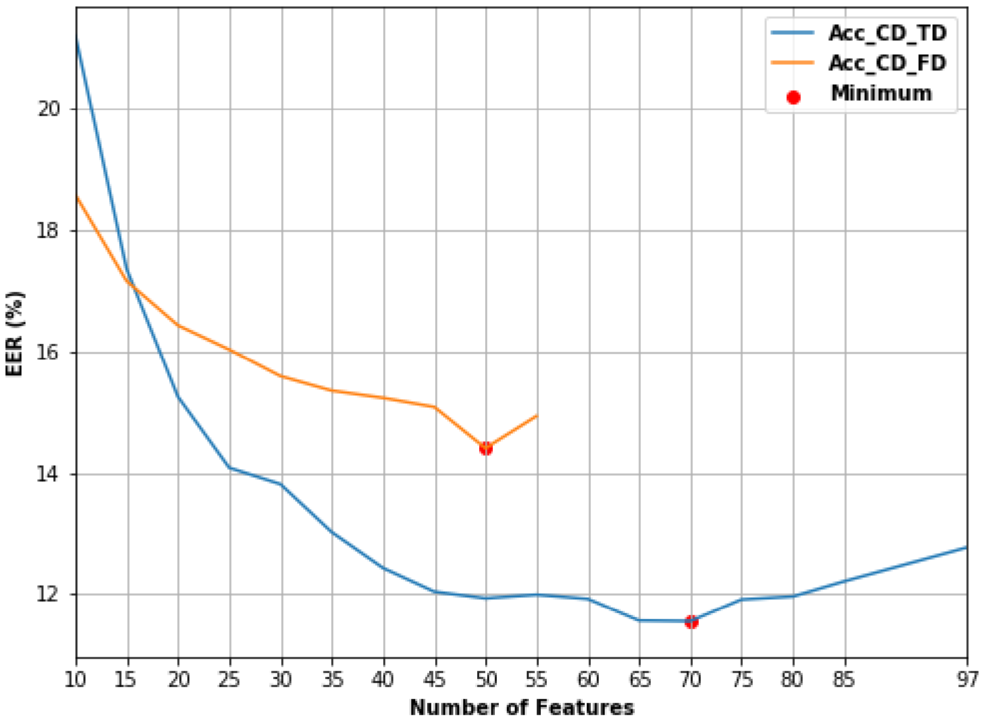


Figure 5-2: The EER results on Accelerometer data for all activities by using CD scenarios

Accelerometer Feature type	Dynamic Feature		Static Feature	
	Number of Features (NF)	EER (%)	All-Time domain Features	EER (%)
SD Time Domain	40	8.4	97	9.9
SD Frequency domain	30	11.3	55	13
CD Time Domain	70	11.5	97	12.77
CD Frequency domain	50	14.4	55	14.93

Table 5-1: System performance utilising dynamic feature selection technique on Accelerometer data

From Table 5-1, it appears that better performance is achieved using the dynamic feature selection techniques across the SD and CD scenarios, by decreasing the number of features used and concentrating on more discriminative information. System performance of the non-realistic SD was significantly better than the results obtained in CD under the time and frequency feature sets. Indeed, the best performance of 8.40% and 11.3% EER was found by using users' accelerometer data with 40 and 30 features under the SD scenario. Meanwhile, 9.9% and 13% EER was achieved by using the complete feature vector data via 97 TD features and 152 FD feature respectively. In addition, when the dynamic feature selection method was used, the biggest performance gap was observed on the users under the more realistic CD scenario: 11.55% and 14.4% EER were obtained using the 70 and 50 dynamic features while 12.77% and 14.93% EER were achieved by utilising all 97 TD features and 152 FD features respectively. It appears that the TD features set achieved better performance than the FD features set in both scenarios. Concerning the feature subset size, it can be seen from Figure 5-2 that the CD test requires more features (i.e. 70 features) than SD (i.e. 40 features) to achieve lower EER. It was apparent that the users' gait manner could differ and fluctuate over time as a result of many reasons (e.g. mood, clothes and shoes).

2. **The Gyroscope Data Exploration:** as shown in Table 3-9 (the literature survey), there has been limited used of gyroscope sensors by previous studies. To understand whether the gyroscope could be useful to contribute to differentiating users, a study investigating the impact of the gyroscope data samples using the time and/or frequency domain features on the system performance was conducted. The results achieved under a complete set of experiments involving various feature vector lengths and the achieved performance on users' gyroscope data for all activities (under both the same and cross day scenarios) are presented in detail in Figure 5-3, Figure 5-4, and Table 5-2.

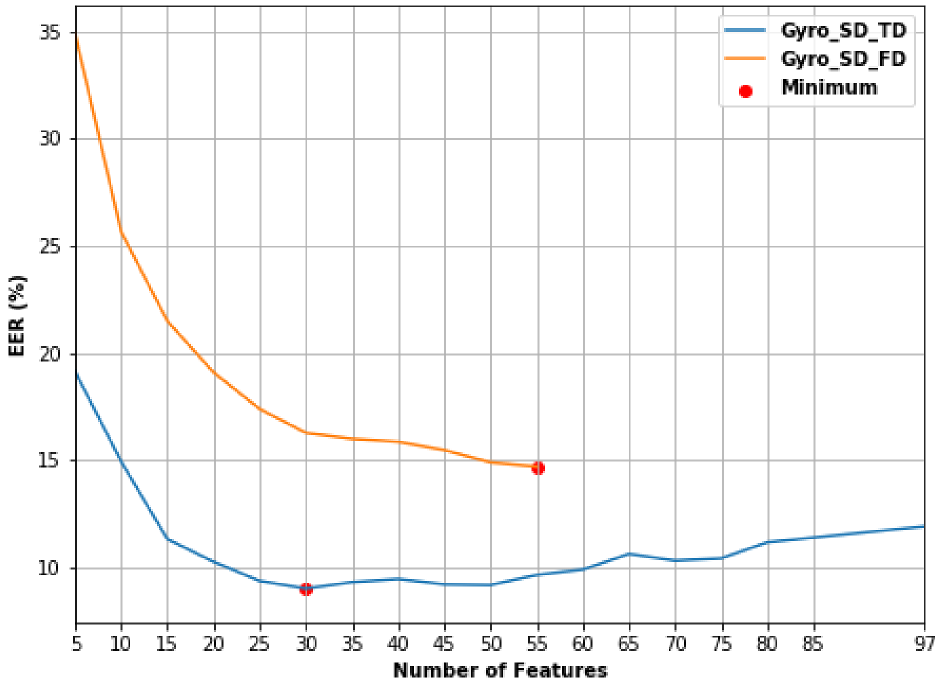


Figure 5-3: The EER results on Gyroscope data for all activities by using SD

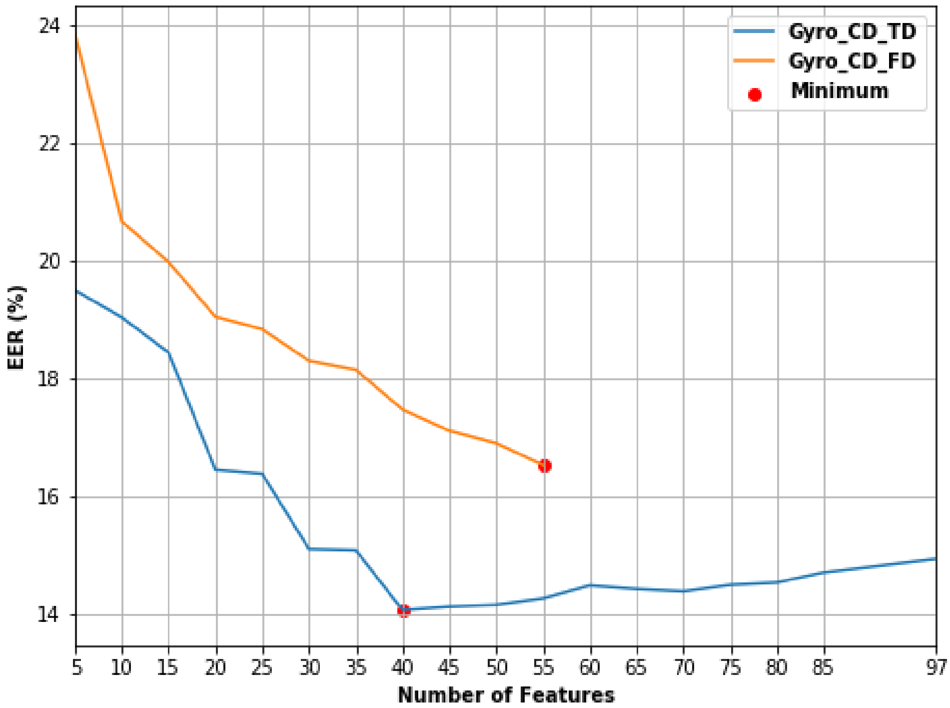


Figure 5-4: The EER results on Gyroscope data for all activities by using CD

Gyroscope Feature type	Dynamic Feature		Static Feature	
	Number of Features	Gyro. EER%	All Features	Gyro. EER%
SD Time Domain	30	9	97	11.9
SD Frequency Domain	55	14.7	55	14.7
CD Time Domain	40	14	97	14.92
CD Frequency domain	55	16.5	55	16.5

Table 5-2: System performance utilising dynamic feature selection technique on Gyroscope data

Interestingly, better system performance was achieved using the dynamic feature selection techniques crossing gait activities. The time-domain feature vectors produced better performance in both the same and the more realistic cross day scenarios (i.e. 9% and 14% EER with SD and CD using 30 and 40 feature sets). Compared to the time domain results, there was 11.9% and 14.92% EER for the SD and CD, respectively, without using dynamic feature selection techniques. In comparison, no difference in the performance was demonstrated in the frequency domain results with and without dynamic feature selection, whereas employing the full feature vector features had the best results in both SD and CD scenarios.

The gyroscope has not been widely studied, which makes the experimental results more interesting to understand the extent to which this signal type affects system achievement. Moreover, the signals extracted from both the accelerometer and gyroscope sensors contribute to creating a more significant feature vector, as well as aim to create a more discriminative feature subset by calculating non-gravitational accelerations using accelerometer data and providing rotational velocities using gyroscope data sensors. All things considered, it seems reasonable to use both sensors together to explore the degree to which they could be utilised for differentiating mobile users.

3. **Both Time and Frequency Domains for Accelerometer and Gyroscope Data**: after checking the relative performance of the two measures, the impact of both TD and FD features for the two different sensors were studied.

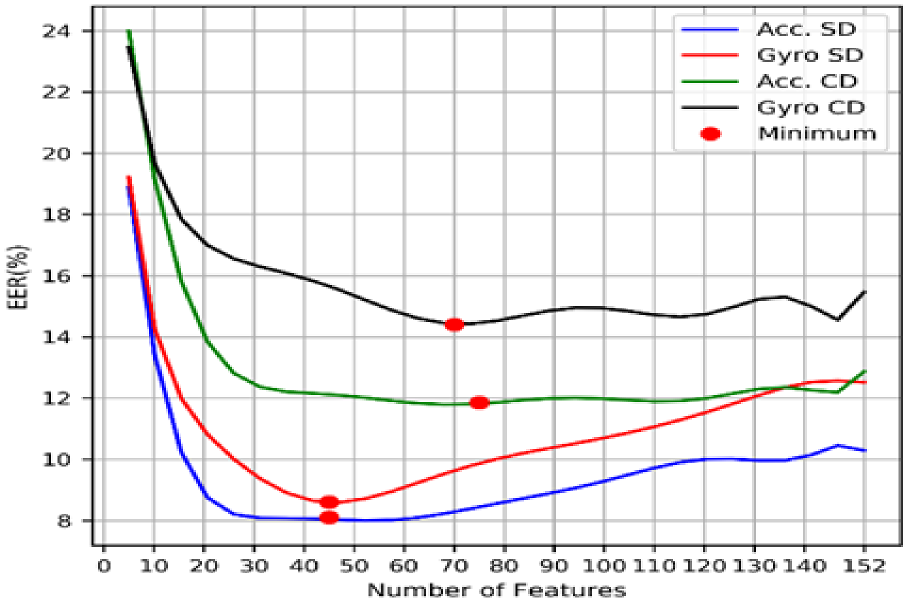


Figure 5-5: The EER results on Accelerometer and Gyroscope data for all activities by using SD and CD scenarios

(TD & FD)/ Features Acc.& Gyro Sensors	Dynamic Features		Static Features	
	Number of Features (NF)	EER (%)	All Features	EER (%)
SD Acc.	45	7.80	152	10.20
SD Gyro.	45	8.39	152	12.50
CD Acc.	75	11.76	152	12.85
CD Gyro.	70	14.25	152	15.45

Table 5-3: System performance utilising dynamic feature selection technique on Accelerometer and Gyroscope data

Figure 5-5 demonstrates that better performance is achieved regarding users' gait activity (both accelerometer and gyroscope data) when the dynamic feature selection technique is applied. Table 5-3 presents the best results achieved under a complete set of experiments involving various feature vector lengths). Indeed, the best performance of 7.80% EER is shown with the user's accelerometer data with 45 features under the SD scenario. Further, when the dynamic feature selection method is used, the most significant performance gap can be observed on the user's gyroscope data under the SD scenario: 8.39% EER is obtained using the 45 selected features while 12.50% EER is achieved using the whole 152-feature set (from both the time and frequency domains).

4. **Accelerometer and Gyroscope Features Dataset**: This concerns the research question of exploring the impact of dynamic feature selection techniques and the value of the feature space on the performance of different activities. Both accelerometers and/or gyroscopes (tri-axial sensors based) are reliable to be employed by sensor-based authentication systems (Lau & Tong 2008). Therefore, more gait features were required as the variability of the signal increased with the real-world dataset. As a result, more experiments were conducted for further verification to conceive that the system performance can be improved if both sensors are used together. To give an illustration of the meaning of a full feature set (i.e. 304 features from both accelerometer and gyroscope signals) both the time and frequency domains were investigated. Figure 5-6 and Figure 5-7 compare the experimental results and the impact of dynamic feature selection techniques on full feature vectors in the same and cross day scenarios.

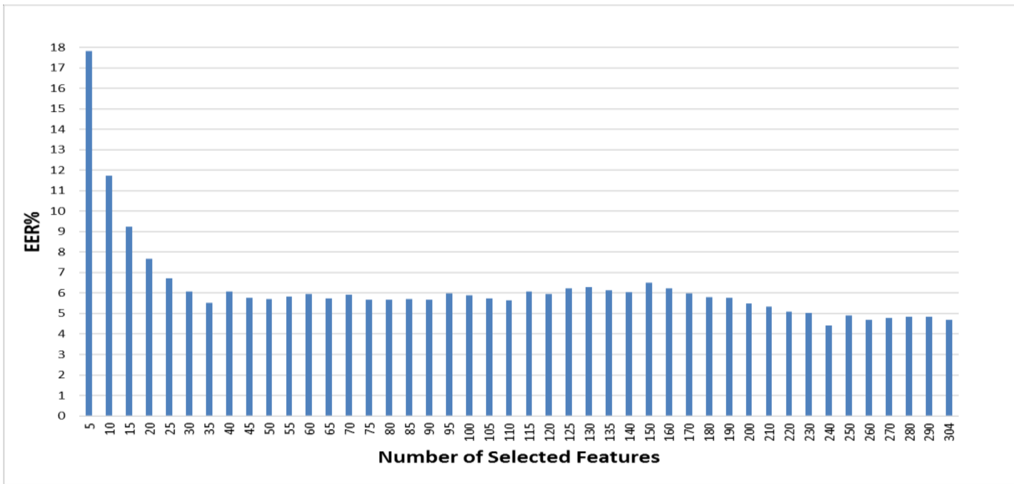


Figure 5-6: The impact of using dynamic feature for all activities (same day scenario)

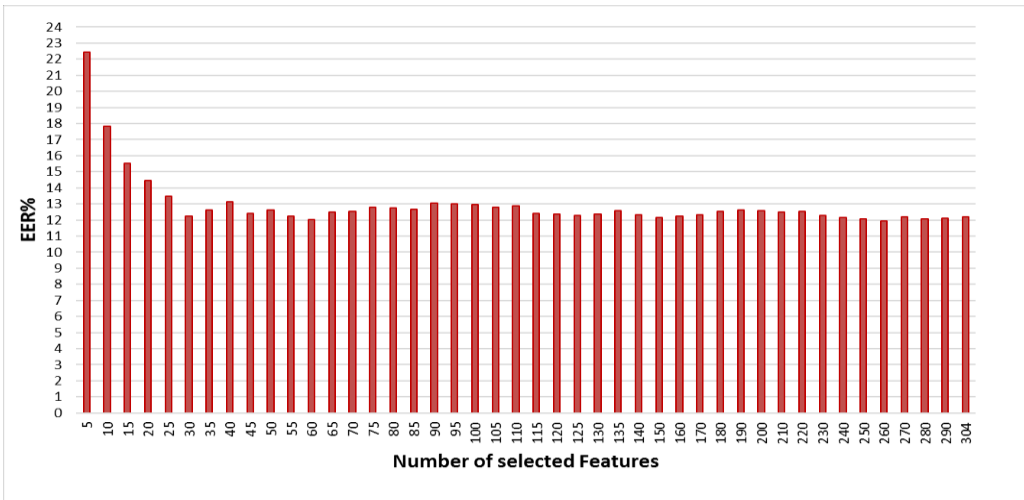


Figure 5-7: The Impact of using dynamic feature for all activities (cross day scenario)

Acc. & Gyro Sensors	Dynamic Feature		Static with All feature (304)
	No. of features	EER (%)	EER (%)
SD	240	4.41	4.70
CD	60	12.00	12.18

Table 5-4: The impact of using dynamic feature for all activities (same and cross day)

Table 5-4 highlights the results with and without using the DF technique. The gait dataset (five gait activities data samples) reported the best EER of 4.41% using 240 feature subsets. It is worth mentioning that a reasonable performance of 5.53% EER can also be achieved using only 35 features where the trend started decreasing under the SD scenario. As shown in Table 5-4, the trend decreases to EER of 12.22% using a 30-feature subset and the best EER of 12.00% via 60 features under the CD scenario. Moreover, using the full feature vector of both situations, the EER was 4.70% and 12.18%. It is worth mentioning that the EER can be reduced by 1.13% by using 240 features although it will require more time to process the data and will increase the computational load on the classifier.

Notably, no significant differences were found between the dynamic and static results with all activities in SD and CD methodologies. However, there is a substantial difference with the number between the two groups of selected features (35 and 60 features versus 304 features). Indeed, the numbers using selected features were around 11.5% and 19.7% of the total number of features for SD and CD. That means the proposed feature selection method effectively discarded a high percentage of inappropriate and/or redundant features and enhanced the system’s accuracy.

Regarding the performance of the SD and CD scenarios, the SD scenario always outperforms its CD counterpart regardless of the dynamic feature selection process used; this is understandable, as human walking behaviour will change over time as a result of various reasons, including changes in shoes, clothes, mood, or health and is in line with what the previous researchers have found. However, notably, better performance is achieved using fewer features in both SD and CD scenarios—although the CD required a more significant number of features than the SD. This phenomenon suggests more gait features are required as the variability of the signal increases and, therefore, additional features are needed when the technique is applied in real life.

5.2.2 Investigation of the Feature Vector across Activities

As demonstrated earlier, the feature sets that were extracted from the accelerometer and gyroscope signals are composed of 304 features. As each sensor has three axes, most features are performed by a vector of three values. Regardless of whether the sample is being created from accelerometer and gyroscope sensor data, the features are the same. From (1-152) are accelerometer (TD and FD) features and from (153-304) are gyroscope (TD and FD) features, as illustrated in Table 5-5.

No.	Time Domain Features	Feature Order for Acc	Feature Order for Gyro	No.	Frequency Domain Features	Feature Order for Acc	Feature Order for Gyro
1	Mean	1-3	153-155	25	Mean	98-100	250-252
2	Standard Deviation	4-6	156-158	26	Standard Deviation	101-103	253-255
3	Average absolute difference	7-9	159-161	27	Average absolute difference	104-106	256-258
4	Variance	10-12	162-164	28	Variance	107-109	259-261
5	Covariance	13-15	165-167	29	Covariance	110-112	262
6	Average resultant acceleration	16	168	30	Average resultant acceleration	113	263-265
7	Binned distribution	17-46	169-198	31	Median	114-116	266-268
8	Difference	47-49	199-201	32	Root mean square	117-119	269-271
9	Median	50-52	202-204	33	Skewness	120-122	272-274
10	Root mean square	53-55	205-207	34	Kurtosis	123-125	275-277
11	Skewness	56-58	208-210	35	Percentile 25	126-128	278-280
12	Kurtosis	59-61	211-213	36	Percentile 50	129-131	281-283
13	Percentile 25	62-64	214-216	37	Zero crossing rate	132-134	284-286
14	Percentile 50	65-67	217-219	38	Maximum value	135-137	287-289
15	Percentile 75	68-70	220-222	39	Minimum value	138-140	290-292
16	Zero crossing rate	71-73	223-225	40	Interquartile range	141-143	293-295
17	Maximum value	74-76	226-228	41	Correlation coefficients	144-146	296-298
18	Minimum value	77-79	229-231	42	Entropy	147-149	299-301
19	Maximum peaks	80-82	232-234	43	Energy	150-152	302-304
20	Minimum peaks	83-85	235-237				
21	Peak Occurrence	86-88	238-240				
22	Time between peaks	89-91	241-243				
23	Interquartile range	92-94	244-246				
24	Correlation coefficients	95-97	247-249				

Table 5-5: Accelerometer features list (1-152) and Gyroscope features list (153-304)

It is worth noting that the effectiveness of each feature towards the classification can vary, with some features having a more significant impact for some users over others. Therefore, a dynamic feature selection approach was devised that can select features based on their uniqueness for individual users. The effect of the dynamic feature selection process (explained in Section 5.2) on performance was successful through consuming fewer resources and time. Therefore, it is essential to know what is composed of mostly repeated features for each user.

In the previous section, all users' gait activities were treated together. In this part of the experiment, each activity was studied individually to address the research question of exploring the impact of dynamic feature selection techniques and the value of the feature space on the performance of different activities, which is related to the first datasets (the controlled environment).

Fundamentally, the research in this part seeks to understand the impact of dynamic feature selection techniques and the value of the feature space on the performance of each gait activity (i.e., normal, fast, walking with a bag, and walking up and downstairs). It also seeks to explore the classification performance of individual activities to understand whether a multi-algorithmic approach would provide a better level of performance. All the following experiments are investigated with SD and CD evaluation and accelerometer and gyroscope feature data for individual activities evaluated by the SVM classifier. Figures 5.8- 5.17 present the system performance depending on various feature subsets. The best EER is selected and summarised in Table 5-6.

Figure 5-8 shows the impact of different feature subsets on the same day scenario system achievement. The performance improved by leveraging more feature subsets. Significant changes started from feature subset 80 with an EER of 1.4%, and the curve seemed to change between 0.70% and 1.8% using different feature subsets. By using 110 features, the best performance of 0.70% was achieved. Figure 5-9 shows the system performance for the cross day scenario. The best EER was 6.30%, and the result of utilising all features was 7.50%. The performance level fluctuated between these values, but beyond feature 100 the resulting effect becomes clear.

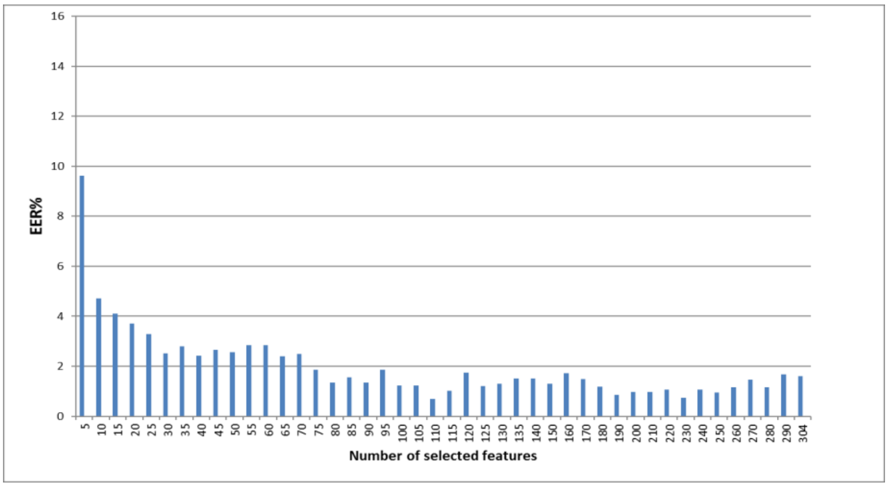


Figure 5-8: Impact of the dynamic feature selection technique upon the performance, normal walking (same day scenario)

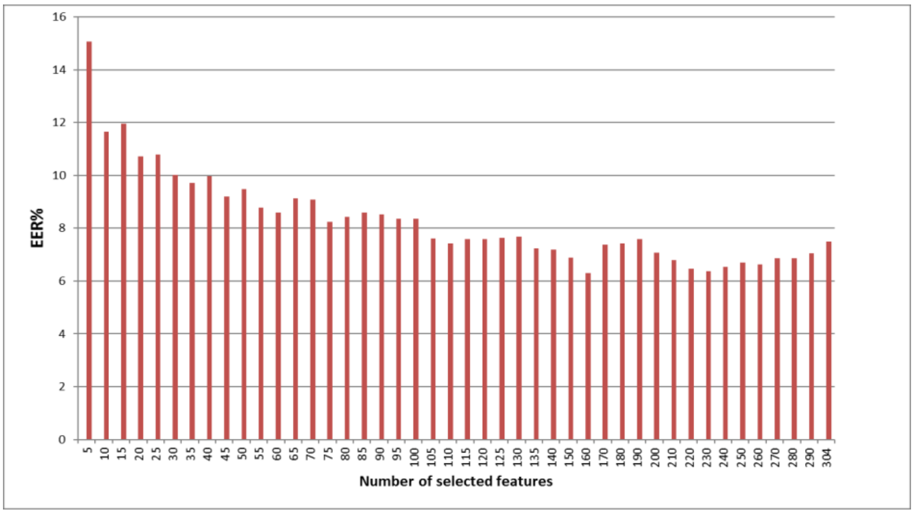


Figure 5-9: Impact of the dynamic feature selection technique upon the performance, normal walking (cross day scenario)

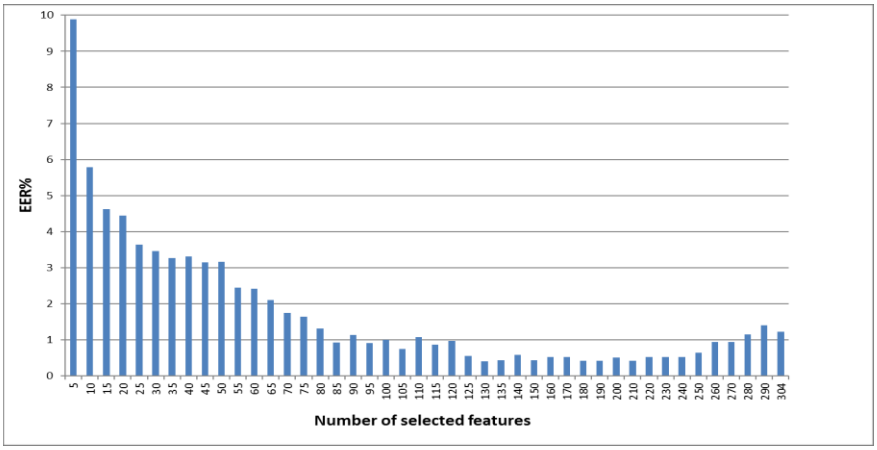


Figure 5-10: Impact of the dynamic feature selection technique upon the performance for fast walking (same day scenario)

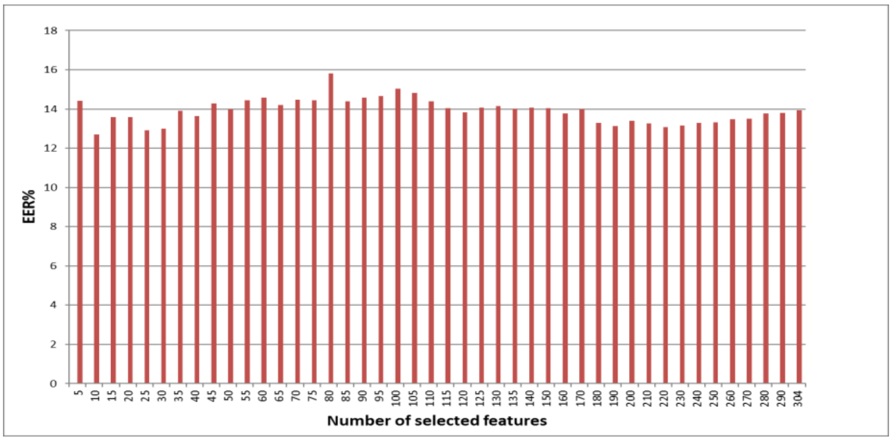


Figure 5-11: Impact of the dynamic feature selection technique upon the performance for fast walking (cross day scenario)

Significant improvement in the result has been shown in Figure 5-10, and the effect started from feature subset 85 downward to feature subset 130, which obtained the best EER of 0.42%. The EER began to increase when more than 130 features were applied gradually. In contrast, there was a small variation, and a small amount was returned in the cross day scenario, which obtained better performance with only a ten feature subset of 12.70% EER. Despite less performance being obtained with the cross day scenario, this finding revealed that fast walking could have more distinctive features with better recognition, using fewer features.

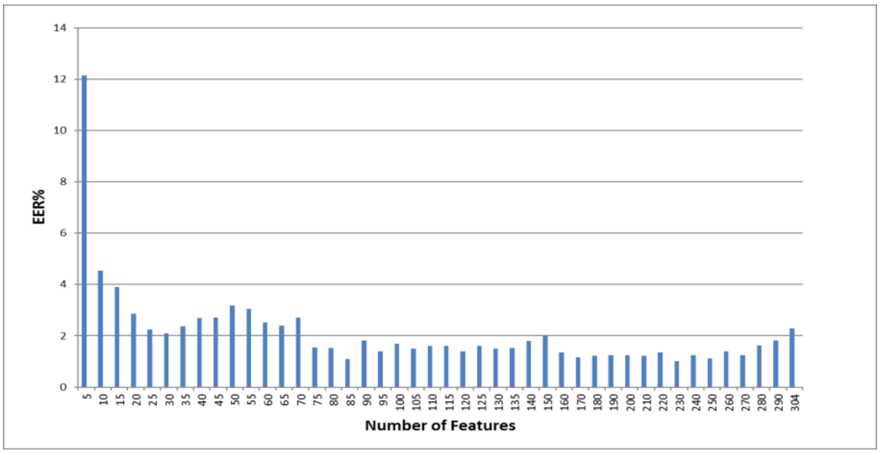


Figure 5-12: Impact of the dynamic feature selection technique upon the performance for walking with a bag (same day scenario)

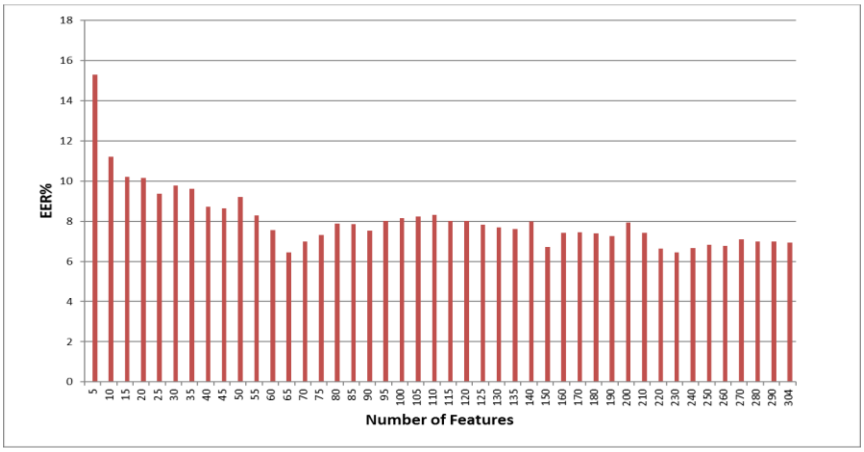


Figure 5-13: Impact of the dynamic feature selection technique upon the performance for walking with a bag (cross day scenario)

The same as previous activities, the same day scenario obtained better performance than the cross day scenario in the walking with a bag activity. Figure 5-12 shows a fluctuating trend in the same day performance; the effect started from feature subset of 30, then begins to reach the plateau and reduces to have the best EER of 1.10% using 85 features. The cross day results presented in Figure 5-13 show the best EER of 6.46% by using 65 features and the result of applying all features is 6.94% EER. The performance level fluctuated between these values.

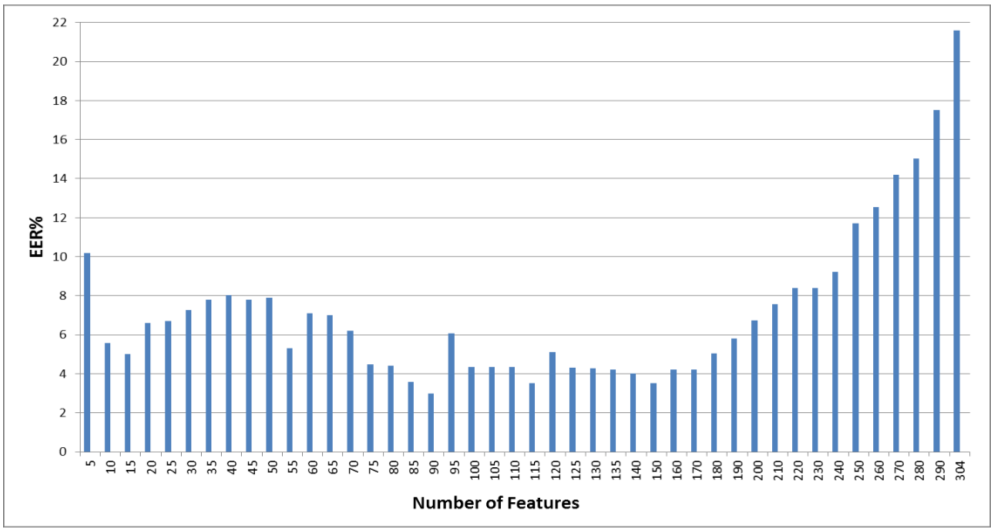


Figure 5-14: Impact of the dynamic feature selection technique upon the performance for down stairs (same day scenario)

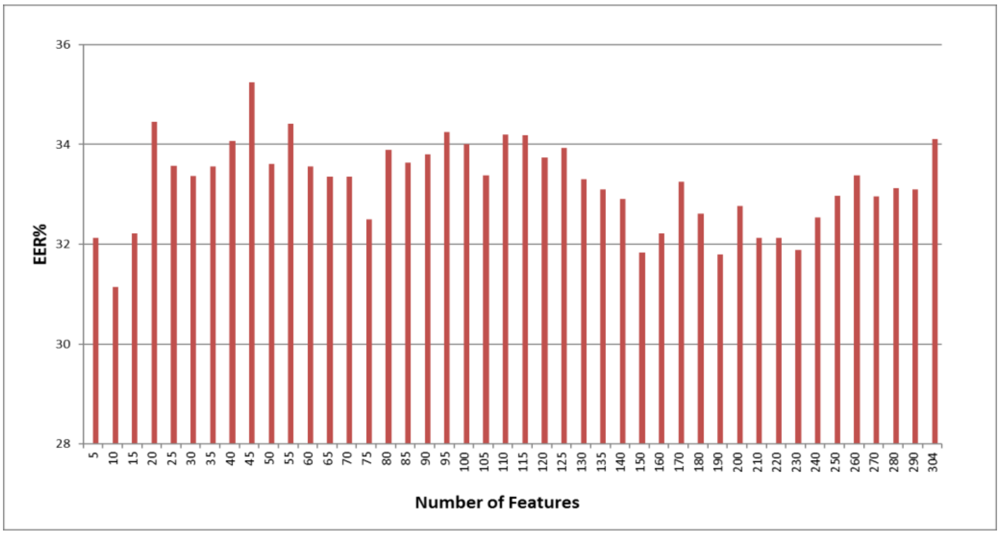


Figure 5-15: Impact of the dynamic feature selection technique upon the performance for down stairs (cross day scenario)

Same day performance for the user's walking down the stairs activity using the dynamic feature selection technique is presented in Figure 5-14. The best EER of 3% was achieved by using 90 features subset and the performance gradually increased. Compared with the cross day scenario, Figure 5-15 shows the range seems to be different with various feature subsets. Moreover, a relatively better result of 31.70% EER was obtained by using only a 10-feature subset.

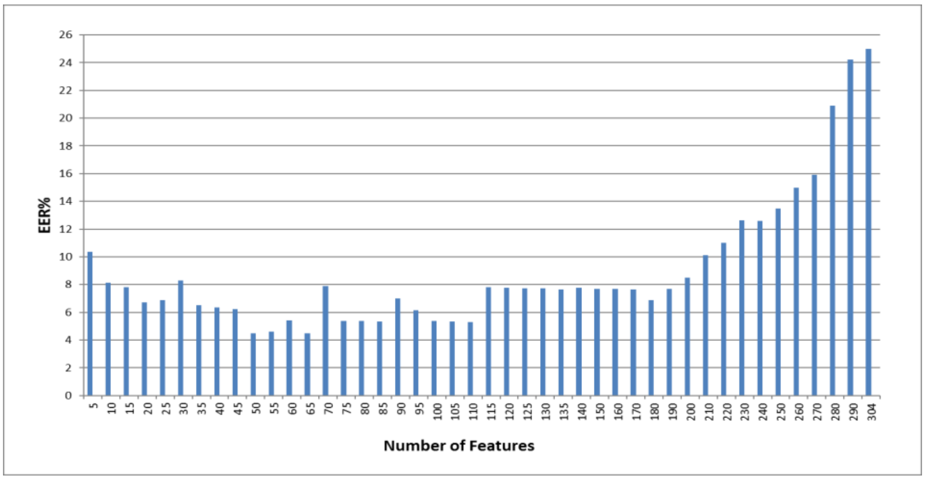


Figure 5-16: Impact of the dynamic feature selection technique upon the performance for upstairs (same and cross day scenario)

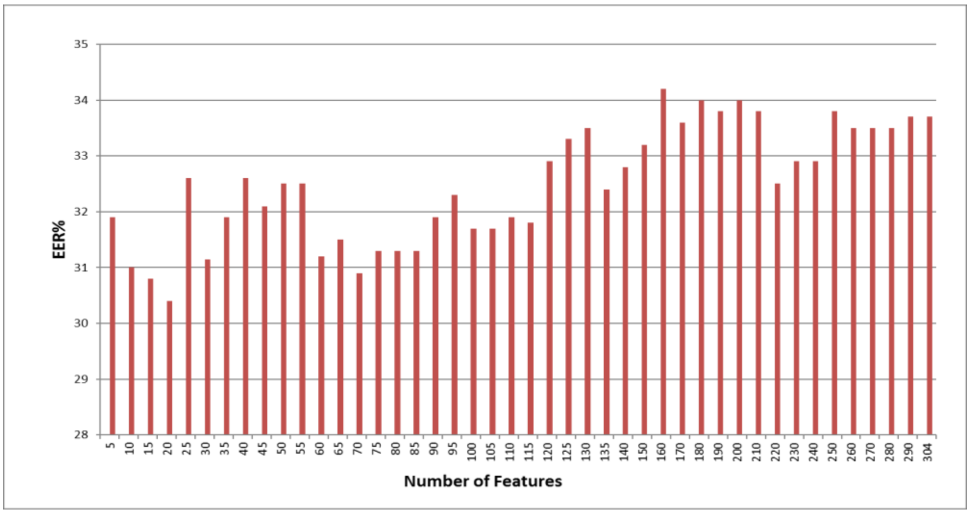


Figure 5-17: Impact of the dynamic feature selection technique upon the performance for upstairs (cross day scenario)

Figure 5-16 and Figure 5-17 provided the same and cross day results for users' upstairs activities. Generally, the cross day scenario achieves less performance than the same day scenario while both scenarios obtained better results employing fewer features (4.50% EER with 50 and 30.40% EER with 20 feature subsets for the same and cross day, respectively).

To investigate the impact of individual gait activities on the classification performance, a multi-algorithmic approach was used to evaluate across the different activities (i.e. normal walking, fast walking, walk with a bag, walk down the stairs, and walking up the stairs). All users' activities from both accelerometer and gyroscope sensors were examined to set a benchmark for comparison purposes.

TD and FD Features/ Acc &Gyro Sensors	Same Day			Cross Day		
	Dynamic		Static with All feature (304)	Dynamic		Static with All feature (304)
	No. of Features	EER (%)	EER (%)	No. of Features	EER (%)	EER (%)
Normal	110	0.70	1.60	160	6.30	7.50
Fast	130	0.42	1.22	10	12.70	13.92
With Bag	85	1.10	2.29	65	6.46	6.94
Down Stairs	90	3.50	21.60	10	31.10	34.10
Upstairs	50	4.50	25.0	20	30.40	33.70
All activities	240	4.40	4.70	60	12.00	12.18

Table 5-6: A Comparison between dynamic and static features techniques for individual and all activities (SD and CD scenarios)

Initially, similar patterns are exhibited by the results regarding the impact of the dynamic feature selection process, with the results using the dynamic feature selection process outperforming those obtained using the full feature set (i.e. 304 features from both accelerometer and gyroscope signals).

As expected, the results from the SD scenario outperform the results obtained from the CD scenario. Usually, the user same-day pattern will not vary as the cross-day activity pattern. The number of features is reduced significantly for the CD scenario, although the performance only increases marginally, which is considered a satisfactory indication regarding using a dynamic feature.

Regarding performance, the best results are 0.70% EER for the normal walking activity with 110 features for the SD scenario and 1.60% EER (using the full 304 features) both of which are better than the performance of existing studies with 1.95% EER (Derawi 2012) and 1.82% EER (Watanabe 2014), using a larger dataset both in terms of participants and samples per user. In addition, regarding the same activity, 6.30% EER is obtained by using 160 features for the CD scenario and 7.50% EER with the full 304 features; these results are in line with prior work including 6.1% EER (Nickel, Brandt, et al. 2011c) and 6.15% EER (Muaaz & Nickel 2012; Watanabe 2015). Nevertheless, those three studies applied decision-level logic (majority or quorum voting), which these results have not applied (at this stage). The decision-level logic techniques may improve their classification results by up to 50%; in addition, they used 20% fewer users for their experiments than this study. Hence, it could be easier to distinguish individual users. Concerning the 'all activities' performance (SD scenario) using the dynamic feature approach, for both sensors (accelerometer and gyroscope), it shows a significant improvement in EER of 4.4% compared to an EER of 7.80% using the accelerometer signal alone. This is consistent with this study's hypothesis, which supports the various features' sources to improve the user recognition process.

As demonstrated in Table 5-6, a significant difference in results can be observed between the multi-algorithmic classifier for individual activity and the single classification approach. With the dynamic feature selection process being used, for the SD scenario, all individual activities (apart from walking down and up the stairs) achieve better performance than the results obtained when all activities are combined. In addition, all individual activities use a smaller number of features in comparison with the number of features used by all activities with a minimum difference of 110 features. In contrast, only two individual activities (i.e., normal walking and walking with a bag) perform better when compared with the result achieved by all users' activities for the CD scenario. Nevertheless, they at least double the number of features used by all activities. The results show gait recognition while walking up and down the stairs was not particularly good, even when applying the dynamic feature selection approach. Further, data analysis showed that this data still suffered from a high degree of variability, which subsequently made classification a challenge (Nickel et al. 2011).

Indeed, it will be more beneficial to define the most discriminative dynamic-based features for most users across each activity type. Therefore, a preliminary analysis will be provided. According to the most repeated features for each activity and all data samples across all user models, the top ten most discriminative features across normal walking activity the complete set presented in appendixes C, D, E, and F. To have more specific features, about 10% of the number of features employed by the classifier are coded in colours where red represents the most repeated features (>40), yellow the second most repeated (<40 and >30), and the green the third most repeated features (>20), all other features coloured in white (<20) repeated. The numbers correspond to the features as listed in Table 5 5. For example, it can be seen from the data in Table 5-7 that the reference pattern of user one could be created using features (10, 13, 159, 220, 217, 56…etc.) while features (67, 70, 43, 13, 10, 114. etc.) may be used to produce user two's reference formula.

#user	Accelerometer and Gyroscope Top Ten Discriminative Features									
1	10	13	159	220	217	56	162	165	266	161
2	67	70	43	13	10	114	34	25	40	162
3	169	105	120	268	266	56	171	170	18	13
4	13	10	41	56	114	17	117	185	34	170
5	266	56	169	34	159	220	217	162	165	105
6	114	10	13	18	170	171	266	57	5	41
7	165	162	220	217	159	43	67	70	164	167
8	56	10	13	117	105	120	32	268	114	41
9	268	162	165	10	13	265	169	43	34	67
10	27	109	56	57	5	123	42	129	117	266
11	34	10	13	56	123	169	41	32	117	171
12	13	10	170	268	171	162	165	41	169	159
13	223	114	10	13	221	218	283	170	254	56
14	56	34	105	120	123	283	268	13	10	169
15	123	170	171	56	135	266	105	120	10	13
16	220	217	159	162	165	287	254	281	266	153
17	162	165	159	217	220	10	13	56	114	41
18	170	171	268	17	169	56	265	18	105	120
19	56	10	13	41	43	57	67	70	5	162
20	114	266	34	169	170	18	268	19	56	111
21	159	162	165	220	217	67	70	10	13	266
22	10	13	32	56	105	120	206	41	57	162
23	43	70	67	164	167	34	12	15	161	219
24	13	10	34	162	165	220	217	159	57	208
25	34	162	165	159	217	220	56	114	287	161
26	19	266	171	114	17	169	13	10	115	18
27	268	56	13	10	170	105	120	171	169	117
28	34	56	13	10	268	41	114	17	171	18
29	56	34	105	120	169	268	170	18	117	171
30	170	171	34	169	111	17	267	114	221	218
31	266	268	171	221	218	169	20	56	123	18
32	10	13	41	114	56	123	165	162	17	4
33	20	22	126	67	70	21	134	221	218	114
34	159	220	217	56	162	165	105	120	268	266
35	114	170	268	171	17	18	13	10	169	56
36	162	165	217	220	159	246	70	67	135	73
37	162	165	217	220	159	246	70	67	135	73
38	10	13	32	170	56	246	105	120	19	18
39	10	13	41	4	70	67	104	56	26	108
40	268	10	13	266	267	56	171	218	221	170
41	169	17	283	56	266	170	34	10	13	171
42	162	165	220	217	159	170	10	13	169	34
43	34	268	18	56	31	169	170	171	114	17
44	34	268	171	169	10	13	56	170	220	217
45	266	169	18	283	105	120	123	10	13	170
46	217	220	159	162	165	123	10	13	18	167
47	13	10	268	41	266	4	267	171	56	17
48	56	114	10	13	221	218	283	170	266	160
49	169	18	19	93	266	287	17	298	223	67
50	169	170	266	268	283	171	221	218	10	13
51	266	18	268	169	282	19	194	171	221	218
52	221	218	266	160	170	255	268	282	163	166
53	169	268	266	19	123	170	18	105	120	283
54	266	13	10	56	41	159	217	220	263	287
55	266	114	171	123	170	13	10	56	41	169
56	268	34	161	223	219	222	167	164	266	215
57	32	34	29	266	105	120	283	17	10	13
58	266	73	171	168	221	218	160	188	194	176
59	266	167	164	169	161	222	219	115	221	218
60	34	56	123	13	10	105	120	169	19	171

(10, 13, 56) Top Repeated	(266, 169) Second Repeated	(170, 171, 268) Third Repeated

Table 5-7: Top ten discriminative features for each user in normal walk

As demonstrated in Table 5-7, the most repeated features used with normal walking activity are as follows:

- Feature numbers 10, 13, 56 are repeated over 40 times. These three features refer to the time domain, accelerometer (x-axis). First is the variance value, which calculates the measure of how far each value in the segment points is from the mean (feature 10). Second is the covariance value, which calculates the measure of how much two variables change together (feature 13). The third is the skewness value, which calculates the measure of the symmetry of distributions around the mean value of the segment (feature 56). Interestingly, the abovementioned three features represent accelerometer (x-axis) measurements. The most discriminative features concentrate on the mathematical value of the distance of each value and the symmetry of distributions around the mean value for each point in the segment (feature 10 and 56). Further, feature 13 calculates the amount of variation between each value in the segment.

- Feature numbers 169, 170, 171, 266, and 268 are repeated between 20 and 40 times and all these features represent gyroscope measurements. Feature numbers 169, 170, and 171 are used to calculate the first three x-axis values located within relative histogram distribution in linear spaced bins between the minimum and the maximum acceleration in the segment. Moreover, features 266 and 268 calculated x and y-axis median values of the data points in the segment.

It is clear from the above table that each user has completely different patterns as compared with other users. In addition, the same user has different pattern according to the activity type. For example,

Table 5-8 compares between user 1 and user 3's top repeated features. It can be seen from the table that each person has completely different patterns for each activity (red related to the first repeated features, yellow is the second repeated features, and the green for the third repeated features).

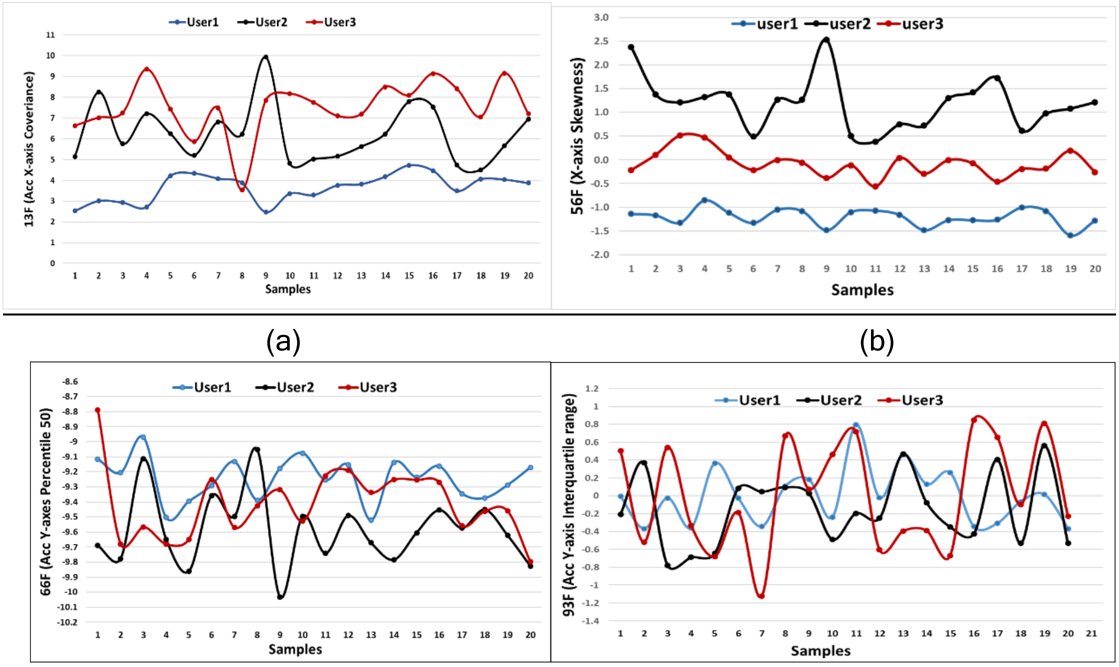
Gait Activity Type	User 1										User 3									
	Top Ten repeated Features										Top Ten repeated Features									
Normal	10	13	159	220	217	56	162	165	266	161	169	105	120	268	266	56	171	170	18	13
Fast	159	217	220	56	281	32	162	165	70	67	56	117	268	105	120	10	13	123	265	166
Carrying Bag	10	13	159	217	220	41	281	162	165	287	105	120	138	56	57	107	122	13	10	117
Down Stairs	32	10	13	268	64	165	162	43	217	220	70	67	61	13	10	3	185	9	268	123
Upstairs	266	27	109	66	69	30	33	42	281	201	16	5	13	10	266	41	123	14	11	27
All Activities	266	193	57	217	220	43	162	165	5	268	127	268	105	120	56	57	138	10	13	106

Table 5-8: A comparison between two users' best features patterns

The most unique dynamic-based features obtained from the preliminary analysis of the above different gait activities tables are summarised in Table 5-9 below. As an example, a part of the full-feature vector (i.e. 304 F) is sorted in ascending order. The total count of feature repetition for each activity is 600. This means 3000 repetition times across five activities (i.e. 5*600). From the data in Table 5-9, it is apparent the accelerometer x-axis covariance feature (13F) was the most repeated feature and the accelerometer x-axis variance feature (10F) was the second repeated across all activities, etc. The three users' patterns plotted in Figure 5-18 (a) and (b) clarify the most discriminated features and Figure 5-18(c) and (d) clearly show user's patterns on less repeated features.

Feature Repetition Order	Feature No.	Normal	Fast	Walk with Bag	Down Stairs	Upstairs	All Activities	Total Repetition
		Feature Repetition						
1	13	37	45	52	44	34	42	254
2	10	36	43	51	45	32	43	250
3	56	34	39	36	25	15	25	174
4	268	22	38	1	28	6	46	141
5	41	14	21	36	14	23	13	121
6	34	20	23	23	13	10	18	107
7	57	5	4	23	10	12	34	88
8	105	15	17	21	13	8	14	88
9	57	5	4	23	10	12	34	88
10	217	15	22	26	2	4	14	83
11	120	14	17	19	12	8	13	83
12	220	15	21	26	2	4	12	80
13	266	28	0	4	5	17	24	78
14	221	11	15	13	14	4	13	70
15	218	11	14	12	14	5	11	67
16	159	15	23	25	0	3	0	66
17	162	19	14	20	1	3	8	65
18	117	6	16	12	15	9	5	63
19	165	16	16	19	3	0	9	63
20	67	11	11	13	12	4	8	59

Table 5-9: Summary of the top repeated features for each activity-control dataset



(a) (b)

(c) (d)

Figure 5-18: (a) Acc (x-axis) covariance feature values for five users (13F)

(b) Acc (x-axis) skewness feature values for five users (56F)

(c) Acc (y-axis) Percentile 50 feature values for five users (66F)

(d) Acc (y-axis) Interquartile range feature values for five users (93F)

Mathematically, covariance (13F) is a measure of how much two variables change together (i.e. when the variables are linearly transformed). From the respect of mean value is the product of the deviations of two variates while the variance (10F) formally measures how far each value in the segment spreads out from the mean. That means that the two functions calculate an approximate value. Therefore, to clarify, the most discriminated features, covariance (13F) and skewness (56F), were selected. As Figure 5-18 (a) and (b) show, there is a significant difference between the three users' patterns; there are almost no common points between them. Consequently, they were vastly differentiated features. Conversely, less repeated features (i.e., percentile 50 (66F) and interquartile range (93F)) were selected to compare with (13F) and

(56F), as there were many of the same or nearly the same values between the users' patterns. Thereby, they appear lower discriminated features.

5.2.3 Investigation Using the Feedforward Neural Network Classifier

As mentioned above, the objective of the experiments is to examine the impact of the dynamic feature selection technique on performance using the SVM classifier owing to its execution time, which is relatively shorter than the neural network classifiers. Depending on those results, a range of the most discriminative features subset for classification will be examined for the different gait activities using the feedforward multilayer perceptron neural network (FF-MLP) to ensure more reliable evaluation, as it exceeded other techniques in previous studies (Karatzouni 2014; Saevanee et al. 2015). For each activity, eleven different FF MLP neural network hiding layer size (i.e.10, 15, 20, 25, 30, 35, 40, 45, 50, 55, and 60) were examined with each being 10 epochs on average to explore the best range (i.e. what is most suitable) of training size for each activity and all activity datasets. The same previous experiment was used. Sixty participants were employed; the data of each user was split into two subsets: 60% for training the classifiers and generating the user profile and 40% for validating and testing the performance. A comparison between the previous section's results using SVM and FF-MLP is presented in Table 5-10 for individual and all activities' datasets.

Type of Activity	FF-MLP				SVM			
	SD		CD		SD		CD	
	Hiding layer size	EER (%)	Hiding layer size	EER (%)	NF	EER (%)	NF	EER (%)
Normal	40	0.08	35	2.09	110	0.70	160	6.30
Fast	40	0.03	40	3.91	130	0.42	10	12.70
With a Bag	50	0.18	45	0.89	85	1.10	65	6.46
Down Stairs	20	1.35	15	23.45	90	3.50	10	31.10
Upstairs	25	4.69	20	23.32	50	4.50	20	30.40
All	20	4.55	50	6.58	35	5.53	60	12.00

Table 5-10: Comparison between FF MLP and SVM performances
(for similar feature vector set FN)

The SVM results were discussed previously in Table 5-6, which is compared between the dynamic and static features techniques for individual and all activities in both the SD and CD scenarios. The same set of best feature subsets are being used in the SVM for each activity, concluded from section (5.2.1), with the above FF-MLP experiments and the best hiding layer size for each activity obtained from the previous series of experiments that were conducted.

In general, as can be seen in Table 5-10, better results were achieved by FF-MLP that outperforms the performance of SVM for both SD and CD scenarios. However, it took a long time in the training and testing phases when the FF-MLP was used. Concerning FF-MLP performance, impressive results were achieved under the SD scenario. The FF-MLP achieved better performance compared with the SVM results that reported an EER in the range of 0.08%-4.69% using FF-MLP against to 0.42%-5.35% using SVM for the individual and all activities dataset

As usual, by applying the more realistic CD scenario, the EERs are increased to a range of 0.89%-6.64% using FF-MLP against to 6.30%-12% using SVM, apart from a walk down and upstairs which achieved poor results especially using the SVM classifier. However, the SVM results still achieved a high level of security.

The dynamic feature technique has shown superior performance over the static feature technique for accelerometer and gyroscope sensors signals. Furthermore, the impact of the proposed multi-algorithmic approach with the FF-MLP classifier is sufficient for the SD and CD scenarios, as most of the individual activities (apart from walking down and up the stairs) achieve better performance than when they are treated as one activity. It is clear that each training size presented different results for each activity. Consequently, each activity is based on different FF-MLP hidden layer size.

The above experiment demonstrated the research questions related to the reliability of both accelerometer and gyroscope sensors, as well as the impact of static versus dynamic feature vectors and the viability of multi-algorithmic versus a single classifier approach. Table 5-10 exhibits the best results obtained from FF-MLP and SVM classifiers that confirmed this research's hypotheses across a wide range of walking activities.

5.2.4 Performance of the Neural Network Feedforward Classifier

This section presents the impact of the dynamic feature selection technique, individual activities, and all activities' dataset with different networks hiding layer sizes on the detailed performance of the SD and CD evaluation scenarios. As aforementioned, the previous comparison between SVM and FF-MLP depended on the best feature subset results obtained from the SVM classifier. It is clear that better performance can be obtained using the FF-MLP classifier with the same feature's subsets. The performance tables are presented in a red-and-white gradient over a range of cells so that white cells represent the smallest value, and the red ones represent the larger values.

As shown in Table 5-11, it is apparent that the best EER is 0.07% using the network hiding layer size with 45 neurons and 130 features. No significant difference with the EER of 0.08% was obtained with the optimum feature subset for the SVM classifier (i.e. 110).

Network hiding layer Size	Number of Features (Normal walk)									
	80	85	90	95	110	115	120	130	140	304
30	0.31	0.20	0.20	0.25	0.15	0.14	0.21	0.28	0.15	0.09
35	0.16	0.19	0.18	0.22	0.16	0.13	0.26	0.12	0.17	0.12
40	0.19	0.16	0.20	0.19	0.13	0.16	0.11	0.08	0.10	0.11
45	0.11	0.14	0.21	0.15	0.08	0.14	0.12	0.07	0.12	0.11
50	0.20	0.13	0.16	0.11	0.13	0.17	0.08	0.09	0.09	0.12

Table 5-11: The EER (%) of the SD test for the normal walking activity

The fast walking data in Table 5-12 shows the best EER of 0.02% could be obtained using various parameters (e.g., 160, 180, and 190 features subset network size with 55, 40, and 35 neurons, accordingly). The results were considered in line with an EER of 0.03% obtained with the best of feature subset for the SVM classifier (i.e.130).

Network hiding layer Size	Number of Features (Fast walking)								
	120	130	140	150	160	170	180	190	304
25	0.08	0.10	0.07	0.09	0.17	0.07	0.06	0.06	0.21
30	0.08	0.09	0.05	0.04	0.04	0.05	0.05	0.05	0.13
35	0.13	0.09	0.09	0.09	0.04	0.17	0.05	0.02	0.16
40	0.04	0.03	0.03	0.05	0.04	0.03	0.03	0.02	0.17
45	0.08	0.09	0.05	0.03	0.05	0.04	0.02	0.04	0.20
50	0.04	0.03	0.04	0.05	0.05	0.07	0.02	0.02	0.19
55	0.03	0.05	0.04	0.04	0.02	0.03	0.03	0.02	0.19

Table 5-12: The EER (%) of the SD test for the fast walking activity

The walking with a bag activity had the best EER result of 0.19% using a relatively small feature subset of only 80 features and a network hiding layer size of 50, as shown in Table 5-13. In addition, this is in line with the best EER of 0.18%, using the ideal feature number (85) to obtain the best results with the SVM classifier.

Network hiding layer Size	Number of Features (Carrying a bag)															
	80	85	90	100	110	120	130	140	150	160	170	180	190	200	230	304
30	0.34	0.38	0.34	0.42	0.47	0.50	0.44	0.49	0.39	0.38	0.41	0.39	0.35	0.31	0.38	0.35
35	0.35	0.33	0.37	0.32	0.42	0.26	0.35	0.31	0.26	0.34	0.36	0.37	0.33	0.23	0.25	0.54
40	0.31	0.35	0.33	0.47	0.39	0.47	0.37	0.31	0.34	0.36	0.37	0.32	0.28	0.24	0.35	0.41
45	0.27	0.40	0.30	0.32	0.37	0.41	0.47	0.22	0.29	0.36	0.30	0.40	0.37	0.28	0.25	0.56
50	0.19	0.28	0.34	0.33	0.29	0.34	0.38	0.27	0.31	0.36	0.36	0.33	0.37	0.24	0.24	0.54

Table 5-13: The EER (%) of the SD test for the walking with a bag activity

Network hiding layer size	Number of Features			
	80	85	90	304
15	2.21	2.58	2.59	6.62
20	2.15	1.35	2.22	6.46
25	3.11	2.72	2.77	8.6

Table 5-14: The EER (%) of the SD test for the walking down stairs

Network hiding layer size	Number of Features				
	45	50	55	60	304
15	2.29	1.7	2.26	2.13	9.38
20	2.13	1.76	2.36	2.44	11.7
25	1.9	1.76	1.64	2.28	9.94
30	2.06	1.98	2.51	2.59	11.32

Table 5-15: The EER (%) of the SD test for the walking upstairs

Table 5-14 and Table 5-15 provide the stairs performance. The walking activities performance exceeded the stairs activities. However, the proposed system can still distinguish the users precisely with an EER of 1.35% and 1.64% with network hiding layer sizes of 20 and 25. This is compared to 1.35% and 4.69% using the FF-MLP classifier for walking down and up the stairs, respectively, for the typical feature subset for each activity used by the SVM classifier. The walking down the stairs performance difference was 3.05%; on the other hand, no difference was shown with walking up the stairs activity results.

Table 5-16 presents the dataset authentication performance for all activities. The best EER was 1.59% using a 230-feature subset and a hiding layer size of 20, compared with the EER of 4.55% conducted using a 35-feature subset (the ideal feature number for the SVM classifier). The FF-MLP, with a variety of activity signals, needs more features to achieve better authentication performance.

Network hiding layer Size	Number of Features (All Activities)															
	30	40	50	60	70	80	110	160	170	180	190	200	210	220	230	304
5	5.82	4.73	4.36	3.65	3.55	3.3	2.92	2.52	2.59	2.3	2.33	2.13	2.07	2.04	1.93	2.11
10	4.92	4.13	3.7	3.46	3.17	3.03	2.59	2.24	2.32	2.17	2.26	2.01	1.89	2.11	1.9	1.84
15	4.57	3.8	3.58	3.15	2.99	2.73	2.43	2.18	2.11	2.18	1.98	1.98	1.85	1.78	1.82	1.77
20	4.55	3.51	3.34	3.01	2.8	2.61	2.42	2.02	2	2.01	2.01	1.89	1.83	1.83	1.59	1.76
25	4.41	3.44	3.27	2.93	2.77	2.61	2.34	2.08	1.91	1.97	1.87	1.82	1.65	1.68	1.74	1.64

Table 5-16: The EER (%) of the SD test for all activities

From the above results (Tables 5.11 - 5.16), it is clear that the selection of the dynamic features algorithm shows better results in this analysis with individual activities. The best performances were achieved with small numbers of feature subsets with the normal, fast, with a bag, and stairs activities. In all activities, the dataset required more features than individual ones to recognise the users. In other words, more feature subsets (i.e. 180, 190, 200, 230, and all features–304) achieved better performance than fewer feature subsets (30, 40, 50, 60, etc.). This is mostly because the variety of walking data signals with a combination of different activity datasets will give more complex gait signals compared to more distinguishable individual activity data signals. Consequently, more feature subsets are needed to perceive the users.

The experimental results for the different walking activities with the more realistic CD scenario are shown through tables 5.17 - 5.22.

Table 5-17 presents the normal activity performance using FF-MLP, best EER of 2.09% obtained using 150 features and a network hiding layer size 35 neurons. That means the same performance conducted using the best feature subset for the SVM classifier used a 160-feature subset.

By analysing individual EERs within the normal activity results, there is an outlier's subject with an EER of 61.99%, 28.85%, 10.33%, and 9.74%. This high error, in comparison to the majority of other subjects, has reduced the averaged EER by 1.84%. Consequently, without an outlier's subject, the calculated average was enhanced with an EER of 0.25% and the computed median shows the EER is 0%.

Network hiding layer size	Number of Features					
	110	150	160	220	230	304
30	2.57	2.59	2.65	2.68	2.41	2.62
35	2.77	2.09	2.69	2.67	2.46	2.49
40	2.66	2.23	2.7	2.65	2.48	2.61
45	2.59	2.3	2.71	2.64	2.5	2.59

Table 5-17: The EER (%) of the CD test for normal walking

The data in Table 5-18 shows the FF-MLP surpasses the results of the SVM classifier with the fast walking activity with the best EER of 3.91%, using the optimum number of features (only ten features) and a network hiding layer size of 40.

The same sufficient feature subset is required for both the FF-MLP and SVM classifiers to obtain the best results.

By analysing individual EERs in the fast activity results, there is an outlier's subject with EER of 58.89%, 36.35%, 30.72%, and 18.77%. This is considered a very high error in comparison to the majority of other subjects. Therefore, the calculated average EER without an outlier subject is 1.49%, and the computed median shows that the EER is 0%. The average EER dropped down to 2.42%.

Network hiding layer Size	Number of Features				
	10	25	30	120	304
35	4.58	5.36	4.8	4.57	4.91
40	3.91	4.61	4.8	4.57	4.74
45	4.14	5.85	4.55	5.25	5.28
50	5.88	6.28	4.8	5.6	5.94
55	5.3	5.45	5.18	5.16	5.15

Table 5-18: The EER (%) of the CD test for fast walking

Table 5-19 shows the performance of walking with a bag activity. The best EER was 0.89% by using a 65-feature subset and network hiding layer size of 45 in comparison with an EER of 6.46% with a fixed selected feature subset using the SVM classifier. That means the percentage level dropped down by 86.23%.

By analysing individual EERs in the walking with a bag activity results, there was an outlier subject with an EER of 17.31%. Thus, the calculated average EER of 0.69% without an outlier subject and the computed median shows that the EER was 0.03%.

Network hiding Layer Size	Number of Features			
	65	150	230	304
35	1.02	0.9	0.92	0.95
45	0.89	1.11	0.92	1.05
50	1.09	0.96	1.02	1.07

Table 5-19: The EER (%) of the CD test for walking with a bag

Network hiding Layer Size	Number of Features				
	10	15	20	150	304
15	23.45	22.84	22.92	22.93	22.56
20	24.11	24.03	24.14	23.21	23.6
25	24.19	23.89	24.44	23.81	23.99
30	23.76	24.09	24.94	24.39	24.22

Table 5-20: The EER (%) of the CD test for walking down stairs

Network hiding Layer Size	Number of Feature						
	10	15	20	60	70	80	304
15	22.89	22.75	22.81	22.46	23.04	23.08	23.09
20	23.06	23.07	23.32	23.2	23.4	23.14	22.94
25	23.61	23.63	23.65	23.27	24.33	23.28	22.85
30	24.13	24.73	24.36	23.33	24.12	24.78	23.18

Table 5-21: The EER (%) of the CD test for walking upstairs

Table 5-20 and Table 5-21 provided the walking down and upstairs walking performance. The walking activities performance exceeded the stairs activities. The walking on the stairs obtained less recognition performance with more realistic CD and EERs of 22.56% and 22.46% with the network hiding layer size of 15 for the down and upstairs walking, respectively. Notably, the results are somehow better than the previous experiments' results using the SVM classifier. However, more features are needed than the SVM classifier. For instance, downstairs utilised the all features vector and upstairs employed the 60-feature subset, compared with 10 and 20 features used with the SVM experiment. Because of the belt phone pouch was mostly wobbling (i.e. unstable) while the participants were walking down or upstairs, the signal was predominantly noisy.

Concerning all activities, it is evident in Table 5-22 that the network hiding layer size of 50 achieves the best EER of 6.58% employing the same feature subset (60) that was most ideal to the SVM classifier and made better results.

By analysing individual EERs in the all activity dataset results, there are ten outlier subjects with EER ranging from 13.10%-35.97%. This is a high error, in comparison to the majority of other subjects. Therefore, the calculated average EER without an outlier subject of 3.35% and the computed median shows that the EER is 2.66%. Accordingly, the average EER decreased by 3.23%, which is considered a superior result.

Network hiding layer size	Number of Features		
	30	60	304
10	7.82	7.86	7.91
20	7.66	7.68	7.2
30	7.2	7.13	6.76
40	6.82	6.76	6.82
50	6.64	6.58	6.64

Table 5-22: The EER (%) of the CD test for All Activities

5.4 Discussion

This research utilised a dataset containing a larger number of gait samples (8,880 samples) across more users (60 users) and covered both same day and cross day scenarios. In addition to regular walking activity, more user gait activities were collected, including fast walking, walking with a bag, and walking up/downstairs, offering the opportunity to learn the user's walking behaviour in a more realistic way rather than under laboratory conditions. Moreover, signals were extracted from both the accelerometer and gyroscope sensors contributing to the creation of a larger feature vector for the time and frequency domains for each sensor. In comparison with existing prior studies; e.g. Osaka University (Ngo *et al.*, 2014) published a dataset employed 744 subjects, accelerometer sensor data, each participant walked normally in a controlled environment for 1 min session, which is not enough for network training. And a study by (Muaaz *et al.*, 2013) employed 51 users walked normally in laboratory conditions.

To improve the system performance, different stratagems were used and their impact investigated on the system's performance, such as a dynamic feature selection technique, the effectiveness of time and frequency domain-based features, in addition to the impact of the proposed multi-algorithmic approach through involving all activities (i.e., five types of activities: normal, fast, with a bag, downstairs, and upstairs walking). In summary, these results show that:

- Accelerometer signal data: the TD features set achieved better performance than the FD features set in both scenarios. As shown in Table 5-1, for the same day scenario, the dynamic feature selection method is applied on all activities with least than 41% and 54% of the number of features being used for TD and FD, respectively. System efficiency using the DF technique outperformed with differences can be obtained of 1.5% and 1.7% in TD and FD, respectively. For the cross-day scenario, the EER difference of 1.22% and the features decreased by 28% with the TD. No significant differences were noticed in EER and the features utilised in FD features.
- Gyroscope signal data: The TD features set achieved better performance than the FD features set in both scenarios. As shown in Table 5-2, for the same day scenario, when the dynamic feature selection method is applied, features decreased by 69% and EER was reduced by 2.9% for the TD features. While, for the CD, time-domain features showed no significant performance. However, the number of features used dropped down by 58.8%. Consequently, the dynamic feature selection succeeded to obtain better results with the TD feature vector while the FD feature vector used all 55 FD features to achieve the best results in both SD and CD scenarios.
- The time and frequency domain features for the accelerometer signal data: Table 5-3 shows the impact of the domains, the TD and FD features,

combined. With respect to the performance, by using dynamic feature selection techniques, the EER differs by 2.40% and 1.09% for the SD and CD accordingly. Furthermore, about 29.60% and 49.34% of features were used with SD and CD, respectively.

Table 5-3 demonstrates that a better performance is achieved by using features from both sensors for both the same day and cross day scenarios.

- Accelerometer and gyroscope features dataset: Notably, as shown in Table 5-4, no significant differences were found between the dynamic and static results with all activities in the SD and CD methodologies. However, the number of feature subsets used was reduced dramatically. This suggests the proposed feature selection technique has an optimistic effect on the system accuracy when applied to all activities (i.e., single algorithmic) with a reduction of 88.5% and 80.3% of the whole features for SD and CD, accordingly. These results indicate that more gait features are required for the more realistic CD scenario as a result of the potential of changing the user behaviour signal over time.
- With respect to the research question (exploring the impact of the dynamic feature selection technique and the value of the feature space on the performance for different activities), which is related to the controlled environment dataset, the impact of the dynamic feature selection technique and the value of the feature space are analysed based on the performance for different activities (i.e. multi-algorithmic). The results using the dynamic feature selection process outperformed those obtained using the full feature set (i.e. 304 features) from both the accelerometer and gyroscope signals, SD and CD scenarios. The ratio of utilised features and the EER differences with and without applying the proposed dynamic feature selection approach are

illustrated in Table 5-23. The 'EER difference' was computed by subtracting the dynamic from the static EER results (i.e. the difference between the performance with and without using dynamic feature selection techniques), displayed in Table 5-6.

Activity Type	Same Day		Cross Day	
	Features Used (%)	EER Difference	Features Used (%)	EER difference
Normal	36	-0.90	52	-1.20
Fast	42	-0.80	3	-1.22
With Bag	27	-1.19	21	-0.48
Down stairs	29	-18.10	3	-3
Upstairs	16	-20.50	6	-3.30
All Activities	78	-0.30	19	-0.18

Table 5-23: The positive impact of using dynamic features selection techniques on the system performance and the features used ratio

This work shows that the proposed feature selection approach has an optimistic effect on system accuracy (under the SD and CD scenarios). With the SD methodology, the utilised features dropped down to a range from 35 features used for the all activities dataset to 160 features used for normal activity (i.e., range ratio from 16%-78% out of the 304 features). Even so, the EER reduced with a range of 0.30%-20.50% with and without using the dynamic features selection technique, but if 35 features were used to classify all activities, the EER differences increased by 0.83% compared with EER using the all features vector. The impact of using the CD scenario on the system performance was also examined. The number of features used to achieve the best performance for individual activities from 10 features was used for the fast and walking down the stairs activities and the maximum features used 160 features for the normal walking activity (i.e., a ratio range of 3%-52% of all 304 features). Nonetheless, only a small improvement on the performance was visible ranging between -0.18% and -0.30%.

Similar patterns were also observed from the impact of dynamic feature selection process on the performance. As shown in Table 5-6, for the same day scenario, at least a 56% decrease in EER can be obtained when the dynamic feature selection method is applied on individual activities with less than 45% of the total number of features being used. In comparison, for the cross-day scenario, the number of features used to achieve the best performance for individual activities (apart from normal walking) decreases dramatically (e.g. with only 65 or 10 features out of the total 304 features); nonetheless, only a small improvement on the performance is visible. It is common that people's walking behaviour can change over time owing to various factors such as weight, mood, and footwear. In addition, there was a 7-day gap between the training and testing data for the cross-day scenario. It is envisaged that the time gap will be reduced for the real-life case, e.g. only the previous two days' data will be used for training and, as a result, better performance will be observed.

- With respect to the research question 5 (To what extent the multi-algorithmic approach is reliable compared to the single classifier approach?) the research examined various activities offering the opportunity to learn the users' walking behaviour across more realistic scenarios than simply walking under laboratory conditions, especially when the prior work is very limited in that regard. As stated earlier, the system evaluated the performance of gait recognition across a wide range of walking activities through involving five types of activities: normal, fast, with a bag, downstairs, and upstairs walking. As demonstrated in Table 5-6, the impact of the proposed multi-algorithmic approach is sufficient for the SD and CD scenario, as most of the individual activities (apart from walking up the stairs for SD) and (fast walking for CD) achieved better performance than when they were treated as one activity. The experiment results have shown that fast walking with the CD scenario is slightly higher than normal and with bag walking. As the methodology suggested, the participants walked in one (the same) corridor, which caused them (in order to add reality to the environment) to turn and walk back when they reached to the end of the corridor. This was repeated about 5-6 times on average with every participant. Consequently, this leads that the speed was not constant for each participant on different days. However, walking on the stairs resulted in poor recognition performance, suggesting that the approach should not be applied to such scenarios.

- The most discriminative features for different activities were investigated. It appears, from Table 5-7, which summarised the top repeated features for normal activity, that the most discriminated features are accelerometer (x-axis) variance (10F), accelerometer (x-axis) covariance (13F), and accelerometer (x-axis) skewness (56F).

As mentioned before, a gyroscope is used to maintain a reference direction in the motion systems by sensing the degree of orientation in the x, y, and z directions of the smartphone. The axis signal is affected by the direction of the device orientation. In addition, the accelerometer sensor measures the acceleration in metres per second squared (m/s2) in the x, y, and z directions of the smartphone.

Figure 5-19 and Figure 5-1920 show the orientation of the positive and negative x, y, and z-axes for a typical smartphone device using the gyroscope and accelerometer sensors, respectively. An Android application called AndroSensor was used to record the sensor data, as it supports most of the sensors an Android device can offer (AndroSensor 2018).

Accordingly, x-axis covariance (13F) measures how much the horizontal movement of the user's leg changes (i.e. when the variables are linearly transformed). This is supposedly distinctive among the users. Regarding the skewness (56F), the measure of the symmetry of distributions around the mean value of the accelerometer x-axis segment, feature 268 calculates the root mean square for the forward movement of the gyroscope z-axis of the leg.

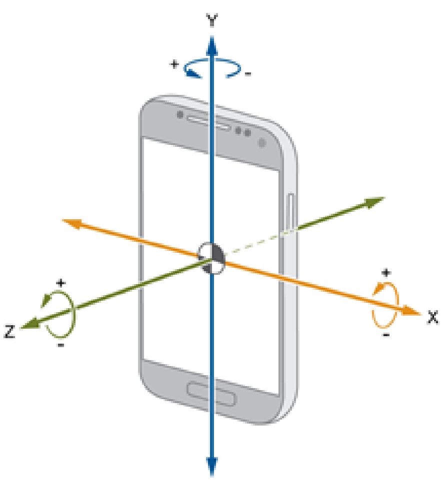


Figure 5-19

The Orientation of the axes relative to a typical smartphone device using a gyroscope sensor.

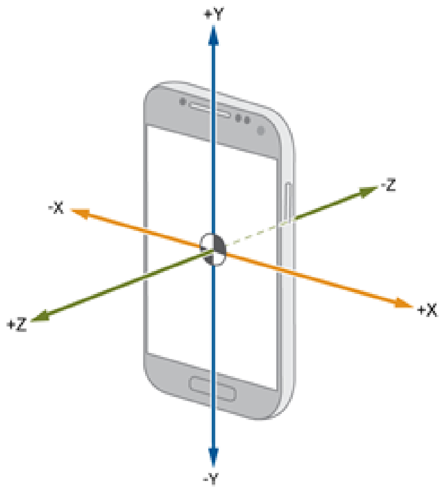


Figure 5-20

The Orientation axes relative to a typical smartphone device using an accelerometer sensor

- Several experiments were conducted to investigate FF-MLP. It can be seen from Table 5-10 that better results were achieved by FF-MLP that outperform the performance of SVM for both the SD and CD scenarios.

5.5 Conclusion

The study sought to investigate the performance of gait recognition across a wide range of activities and participants. Based on 60 participants, the investigation has provided significant evidence to suggest gait-based data can be used as a reliable means of transparently verifying users while moving. However, the performance of the cross-day over the same-day methodology demonstrates feature vector variance that a practical system would need to manage in practice carefully. To aid in this, the study has explored the use of a multi-algorithmic approach (where different classifiers are used based on the nature of the activity) and found that such an approach can achieve a better level of performance over a single classification approach.

The study has also sought to evaluate the feature vector and found that a dynamic approach rather than a static (all feature) approach is beneficial to both the performance that can be achieved but with the added benefit of reducing the computational load on the classifier.

While the study has provided significant evidence to advocate the approach, the experimental methodology still involved relatively controlled data collection. As such, the next chapter will focus on the collection of longitudinal real-life gait-based data to more thoroughly evaluate the recognition performance.

6 Experimental Results: Real-Life Gait Recognition

6.1 Introduction

This chapter explores the details of the consecutive experiments that will be undertaken in developing the proposed real-life gait recognition system. These experiments will focus on providing the empirical basis for whether the proposed approach could work—initially through exploring real-world data (rather than highly constrained control data) to understand the variability and difficulty in successfully authenticating individuals.

This chapter aims to evaluate a real-life gait dataset captured from real and live usage without any restriction conditions. The collected data was employed in a series of investigations to assess the suitability and effectiveness of utilising such kind of gait data for user verification with a view of identifying the attribute types required with a decision being made to verify subject samples, for a successful authentication mechanism. The rest of the chapter is organised as follows The activity identification results are explained, to be able to classify the type of user gait activities. The effectiveness of accelerometer and gyroscope-based features on the system performance was investigated with the impact of static versus dynamic feature vectors. A range of the most discriminative features subset for classification was also examined for the different gait activities using the FF-MLP neural network as the default authentication classifier because of its reliable performance (as demonstrated in Chapter Five). As an additional step, majority voting was applied to the decision to evaluate what impact it would have on performance (inline it being applied by the prior art) to enhance the results of real condition activities. Furthermore, we explored the viability of a multi-algorithmic approach compared with a single classifier approach through actual practice.

6.2 Activities Identification Results

Three different experimental settings were undertaken to study how various activity types affected the identification rate. First, as normal walking and walking with a bag were the most similar activity types, they were merged to form a single activity. The second test combined normal, fast, and walking with the bag into a single activity. As long as the person walked normally, faster or slower in a short period of time during their daily movement (various times of the day, moods, and places), a gradation in the walking speed was expected. The final test examined the correct classification rate for all the activities. As mentioned in Chapter 4 (activity identification model), to train the base model, the controlled data were split into 60/40 training and testing sets, respectively. Once the best model was chosen (the one that achieved the highest performance), the model was retrained using all the controlled dataset for training the final model, which was used to predict the free activities. Three algorithms were the best candidates for the ensemble; these are the feedforward neural network (FF-NN), SVM, and eXtreme gradient boosting (XGB). The highest accuracy was achieved with the FF-NN algorithm (87.67%) and the lowest prediction accuracy was with SVM (84.88%) for all activities (i.e., normal, fast walk, stairs and sitting). The results are illustrated in detail in Table 6-1.

Two types of voting were used: hard and soft majority voting. Hard voting uses predicted class labels for majority rule voting while soft voting predicts the class label based on the argmax of the sums of the predicted probabilities, which is an approach recommended for an ensemble of well-calibrated classifiers (Caruana et al. 2006; Whalen & Pandey 2013). It can be seen in Table 6-1 that the soft voting approach outperformed the other models in all tests.

Activity type	Activity merge	XGB (%)	SVM (%)	NN (%)	Soft voting (%)	Hard voting (%)
Normal, Fast, Downstairs, Upstairs, Sitting	W/bag merged with Normal	93.54	93.35	93.88	94.60	94.27
Walk, Downstairs, Upstairs, Sitting	Fast and W/bag merged with Normal	97.06	97.30	97.65	97.79	97.54
All	None	86.18	84.88	87.67	87.79	87.24

Table 6-1: Overall classification accuracy for each model

The confusion matrix summarises the performance of the classification model for the multi-class classification task in this study (in particular, the soft voting model). It also shows how the predictable model performs on a class level, in which both true-positive and false-negative values can be measured. Figure 6-1 presents the normalised confusion matrix for the percentages for all six activities (normal, fast, W/bag, downstairs, upstairs, and sitting). It is not surprising that sitting had the highest prediction rate of the activities. This is because the uniqueness of its generated sensor signals of the sitting activity. Concerning the downstairs activity, the false-positive samples are misclassified as walking types (either normal, fast, or with a bag), and this could be interpreted as some of the downstairs samples actually containing normal and fast walking types. For example, once a subject reaches the bottom of the stairs, the individual walks a few more steps to complete the activity, which might become a noisy/outlier sample in the downstairs activity dataset.

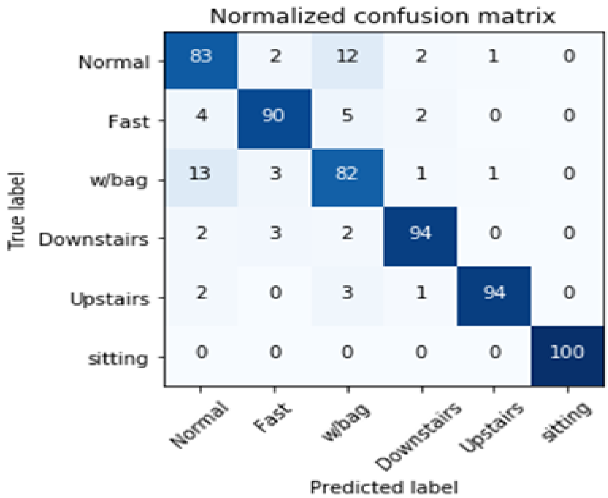


Figure 6-1: Normalised confusion matrix (%) of the soft voting model

6.3 Exploring More Discriminative Features for Different Real-World Activities

In comparison between the controlled and real-world dataset, the best-selected features that significantly contributed to the decision being made in verifying subjects' samples were almost identical among users with the real data signal. This study indicates that the most repeated features for each user were changed with real-world signal dataset along with each single activity data set. Table 6-2 explores the top ten most discriminative features across regular activity model, and the complete set is presented in appendixes G, H, I, J. They are coded in colours where red represents the most repeated features, yellow the second most repeated, and the green the third most repeated features. The numbers correspond to the features as listed in Table 5-5.

Data from Table 6-2 can be compared with the data in Table 5-7, the top ten most discriminative features across the normal activity model for controlled and uncontrolled dataset, respectively. It has been found from the comparison that the most repeated features for an examined subject differed between these two sets (controlled and real). The justification behind such patterns is that the captured signal values increased the variance between the two sets. That is because the real gait patterns were more complex and highly inconsistent than the controlled dataset.

According to the more repeated features regarding every single activity and all activity data samples across all users' models were explored. For example, Table 6-2 explores the top ten most discriminative features across the normal real data activity model. To have more precise features form, about 10% of the number of features employed by the classifier were coded in colours where red represented the most repeated features (>40), yellow the second most repeated (<40 and >30), and the green the third most repeated features (>20). All other white coloured features were repeated less than 20 times.

As demonstrated in Table 6-2, the most repeated features used with normal real data activity as follows:

- Feature numbers 212 and 213 referred to the time domain, gyroscope (y and z-axis), and kurtosis value, which calculated the measurement of the shape of the curve for the segment point's values.
- Feature numbers 89, 90, and 91 referred to the time domain, accelerometer (x, y and z-axis), and the time between peaks, which calculated the time in milliseconds between peaks in the sinusoidal waves associated with most activities calculated and averaged.
- Feature numbers 241, 242, and 243 referred to the time domain, gyroscope (x, y and z-axis), and the time between peaks value, which calculated the time in milliseconds between peaks in the sinusoidal waves associated with most activities calculated and averaged.

  Consequently, the top 6 ranked features (i.e., the most repeated features) are based on the time between peaks feature type for both accelerometer and gyroscope sensors. This could be interpreted that the cycle of walking, as measured by the peak to peak, is distinctively higher for each person.

#user	Accelerometer and Gyroscope Top Ten Discriminative Features_ Real Data									
1	89	90	91	243	72	212	213	241	242	155
2	293	90	89	91	72	241	243	213	242	212
3	90	89	72	91	241	243	213	242	212	2
4	89	91	90	243	241	213	242	212	72	155
5	141	143	212	213	90	89	72	241	16	243
6	89	90	91	243	212	241	242	213	204	219
7	293	72	212	90	213	243	89	241	91	155
8	89	72	90	91	241	243	242	212	213	51
9	293	89	91	90	243	242	213	212	241	2
10	89	90	91	72	241	243	242	213	212	16
11	89	90	91	241	243	242	212	72	213	27
12	293	294	295	213	212	51	66	155	72	2
13	293	294	295	213	212	51	66	155	72	2
14	90	241	89	91	243	242	212	213	51	66
15	89	213	91	212	90	243	242	241	155	153
16	72	212	90	213	241	89	91	162	165	290
17	294	90	89	213	91	72	212	241	243	242
18	293	90	89	91	241	242	243	213	212	155
19	141	294	89	91	90	72	241	109	112	213
20	235	237	236	213	72	212	241	91	204	219
21	212	90	213	89	16	241	91	2	59	243
22	293	90	89	91	241	212	243	213	242	2
23	89	91	241	90	213	243	72	242	212	155
24	89	91	90	72	241	212	213	243	2	153
25	72	295	294	293	204	219	141	143	142	16
26	141	212	90	213	241	243	89	51	66	91
27	89	91	90	213	243	241	212	242	2	51
28	293	89	90	91	241	242	243	213	155	2
29	233	82	80	85	235	81	236	237	83	293
30	89	90	91	212	213	243	59	241	2	242
31	89	91	90	212	241	213	243	72	2	242
32	293	89	90	91	241	242	243	213	212	204
33	212	89	90	243	213	241	91	242	59	16
34	293	89	212	90	213	72	241	243	242	155
35	293	89	90	91	243	241	242	213	212	155
36	294	295	89	90	91	241	213	212	243	242
37	295	89	91	90	241	243	212	242	213	72
38	294	89	90	91	241	243	212	213	72	2
39	212	213	90	89	16	241	59	2	91	155
40	89	72	212	243	213	90	91	241	59	155
41	90	89	72	91	212	213	241	243	155	109
42	89	91	72	90	213	212	243	241	242	155
43	89	91	90	243	212	241	242	213	2	51
44	293	89	90	91	241	242	243	212	213	27

(213, 212, 241) Top Repeated	(89, 90, 91) Second Repeated	(243, 242, 72) Third Repeated

Table 6-2: Top ten discriminative features for each user in the normal and walking with a bag

The above top-ten features tables exhibited the variation of the patterns of the dynamic-based feature selection decision between the users and activities. Table 6-3 illustrates a comparison between user 1 and user 3's top repeated features. It appears from the table that each person has a relative difference pattern for each activity (red colour related to the first repeated, yellow is the second repeated, and the green for the third repeated features with all users).

Activity Type	User 1										User 3									
	Top Ten repeated Features										Top Ten repeated Features									
Normal & W/ Bag	89	90	91	243	72	212	213	241	242	155	90	89	72	91	241	243	213	242	212	2
Fast	89	91	90	243	242	260	263	290	241	162	89	90	91	242	243	164	167	292	241	261
Down Stairs	293	90	91	89	243	242	11	14	241	72	90	89	91	243	151	241	242	169	179	202
Upstairs	270	213	243	170	151	195	109	112	211	187	241	294	89	27	243	2	90	51	66	72
All Activities	293	90	91	89	243	164	167	292	242	212	294	90	91	89	292	164	167	241	212	243

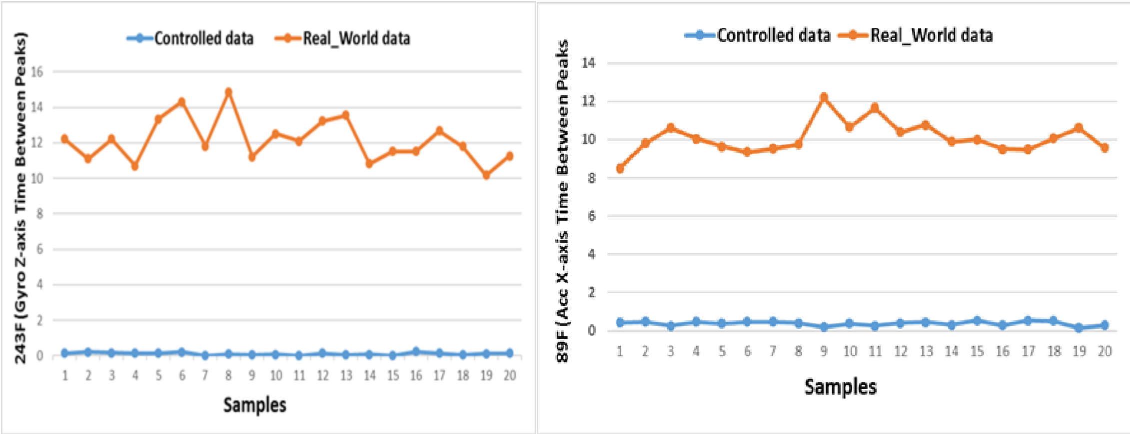
Table 6-3: A comparison between two users' best feature patterns (real dataset)

From the preliminary analysis of the above different gait activities' top repeated tables, the most discriminative distribution-based features that contributed to decision selection are summarised in Table 6-4 below. As an example, a part of the full feature vector (i.e., 304 F) is sorted in ascending order. The total count of feature repetition for each activity is 440, which means 2200 repetition times across four activities (i.e., 4*440). From the data in Table 6.9, it is apparent that the top 6 of the best-selected features significantly contributed to the discussion being made in, verifying subjects' samples are almost identical among users with the real data signal. Our findings revealed that the top 6 are the time between peaks feature type for both sensors (i.e., Gyro, x, y, z-axis and Acc x, y, z-axis). The order is 243F, 89F, 241f, 90F, 91F, and 242F, respectively. To give an illustration of their percentage of accruing are 6.18%, 6.04%, 6%, 5.7%, 5.2% and 4.22% accordingly, across all activities.

Feature Repetition Order	Feature No.	Normal& W/ Bag	Fast	Down Stairs	Upstairs	All Activities	Total Repetition
		Feature Repetition					
1	**243**	36	31	39	30	21	136
2	**89**	39	38	38	18	33	133
3	**241**	40	22	37	33	18	132
4	**90**	39	35	41	11	37	126
5	**91**	38	39	38	0	34	115
6	**242**	28	28	33	4	12	93
7	**27**	2	2	20	25	0	49
8	**212**	40	1	0	3	26	44
9	**213**	42	1	0	0	0	43
10	**72**	25	0	11	6	0	42
11	**164**	0	39	0	2	40	41
12	**167**	0	38	0	0	38	38
13	**292**	0	35	0	0	37	35
14	**162**	1	9	8	16	4	34
15	**165**	1	8	7	15	4	31
16	**165**	1	8	7	15	4	31
17	**2**	14	0	0	15	0	29
18	**290**	1	10	5	13	9	28
19	**151**	0	0	22	4	0	26
20	**261**	0	24	0	0	24	24

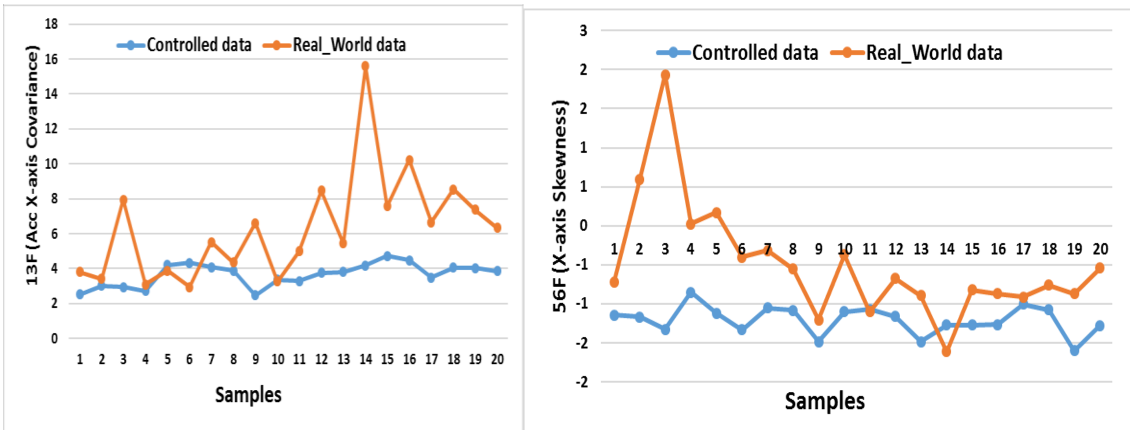
Table 6-4: Summary of the top repeated features for each activity -real dataset

In terms of the user profile, Figure 6-2 (a) and (b) depict examples of the most repeated features, 243F and 89F (i.e., the time between peaks), for the real dataset and, similarly, 13F (accelerometer X-axis covariance) and 56F (accelerometer X-axis skewness) for the controlled dataset. In comparison between the most discriminative feature measurements for the controlled and real-world signal. Figure 6-3 (a) and (b) show a considerable difference of (243F) and (89F) feature measurements between two datasets. Likewise, (a) and (b) show the difference between (13F) and (56F). This finding confirmed that the real-world is not reflected in the control data because it is more variable; it also then highlights the validity of the prior art.



(a) (b)

Figure 6-2: (a) and (b) show User1's profile signals for feature 243 and 89 from both the controlled and uncontrolled datasets



(a) (b)

Figure 6-3 (a) and (b) show User1's signals for feature 13 and 56 from both the controlled and uncontrolled datasets

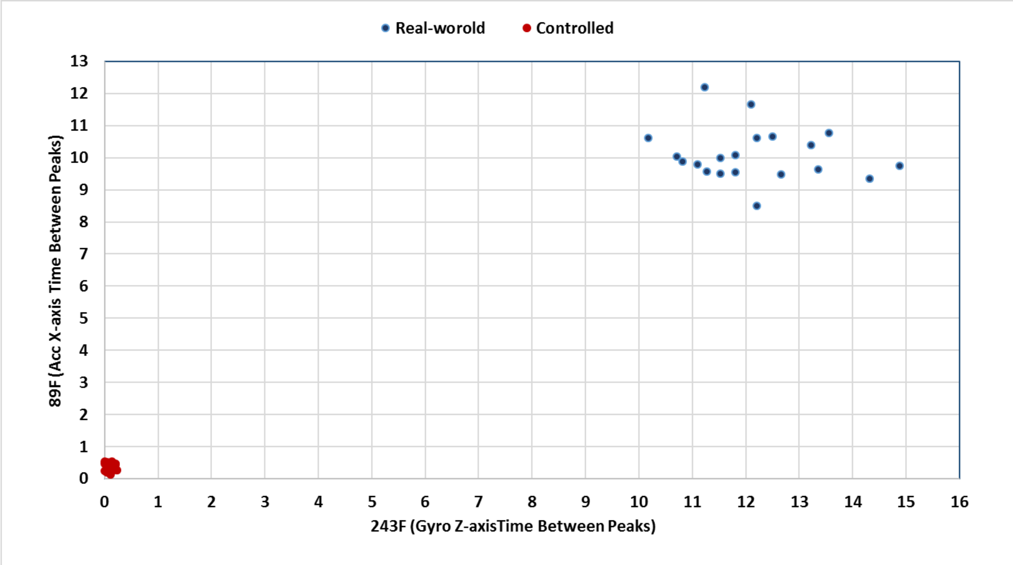


Figure 6-4: The distribution of data points using a 2D plot from 243F and 89F

The most repeated features in the real dataset are 243F and 89F. Figure 6-4 shows the data points of these two features in a 2D plot. Where the red points correspond to the control data and the blue points are the real data. It can be seen that the distributions of these two covariates (features) had shifted in a real dataset in comparison to the control. This means that the STD and mean of these features also changed and this was considered further confirmation that the patterns of the captured signal values were entirely different between the two sets. The dynamic-based feature selection technique output varied between controlled and real datasets. Therefore, all the features were fed into the machine learning algorithm as they were considered more relying than utilising part of them. The following subsection will demonstrate results in the real-world gait authentication system.

6.4 Real-World Gait Authentication and Multi-Algorithm Performance

One of the main contributions of this thesis is to evaluate the performance of gait recognition across different walking activities identified from a real and live unconstrained use of the smartphone sensors' signal (i.e., acceleration and gyroscope). Moreover, it is aimed at investigating the viability of a multi-algorithmic approach through involving all activities vs. a single classifier approach across the real-world dataset. Therefore, the best performance of EERs resulted from the controlled data experiment. The selected FF-NN network sizes classifier (reviewed in Chapter 5) were adopted to be utilised in this and forthcoming real-world experimental studies (as presented in this subsection).

An activity recognition model was applied on the uncontrolled gait dataset, which was already collected from 44 users (7-10 days for each user). Four types of activities were identified (normal, fast, walking with a bag, and down and upstairs) for each user.

As conducted in an uncontrolled experiment to evaluate gait activity verification, the dataset of each subject was divided into four days of the data for training the classifier and creating the user template and the remaining three days for testing and validation (i.e., 100% of genuine data; four days for training/three days for testing). In other words, the training data needs a longer time (around 10 to 15 days) to process, depending on the activity data volume, especially when different activity datasets are merged together. In this setting, mostly the EER is high because of the imbalanced data set problem.

Consequently, and in order to improve the system efficiency, using the undersampling technique as another experiment setting is suggested (e.g., reducing imposters' training set) to perhaps improve the overall EER and to speed up the runtime. Accordingly, the training and testing splitting ratio for the imposter data sample method was determined to 10% randomly selected samples of the first four days were used for training instead of 100% of four days' data.

The previous experiment (demonstrated in Chapter 5) was evaluated by two algorithms, SVM, which is based on a statistical learning technique and FF MLP neural network classifiers. The SVM algorithm was used to train the real dataset. However, because of the nature of it, the process could not accommodate the sizeable real dataset because it can work with a limited volume of data. As a result, an alternative decision tree algorithm, random forest classifier, different tree numbers, and feature subsets were used, but the EER results were high. Therefore, the FF-NN was used instead, as it proves the best results.

The above experiment setting was applied on individual activities; normal and carrying a bag, fast, and down and upstairs, and all activities together, coupled with evaluating a multi-classifier algorithm. The results were somewhat good and better than a single classifier algorithm apart from the walking stairs activities.

Ogbuabor and La (2018b) illustrate the 'Kurtosis' feature, which is a measure of the shape for the values in a particular segment. It is apparent from the 'Kurtosis' descriptive statistic that there is clear variability across the activities examined for this feature. Although normal walk and walk with the bag are two different activities, they are, by their nature, very similar in terms of pace and type of body movement. As well as, the median and first and third quartiles were almost equal for this feature as computed by the random forest algorithm and most of the false positive examining samples for the confusion matrix for the predictable model were also between these two activities, which supports the point being made here. In other words, although normal walk and walking normally with carrying a bag are two different activities, they are, by their nature, very similar in terms of pace and type of body movement. Therefore, they were merged to form a single activity. The activity authentication results are first presented for the "single-sample mode", and then the majority voting scheme was used.

Table 6-5 and Table 6-56 show the accelerometer and gyroscope sensors of the normal walking activity result and the FF-MLP classifier using different feature subsets. The results indicate that utilising a 100% training dataset is not as effective as 10% for classification. For example, the 110-feature subset reported EERs of 15.94% in comparison with 14.53% utilising only 10% training dataset for the same feature subset. Albeit there is no big difference in the performance, there was a significant reduction in the processing time (one to two days), depending on the activity data volume. Depending on the best performance obtained from the previous set of experiments, a network size with 40 neurons was considered, and the full feature vector achieved the best performance of 11.38% utilising 10% of the training dataset.

Network size	Number of Features	10% Training Dataset	100% Training Dataset
40	100	16.39	17.46
40	110	14.53	15.94
40	160	15.90	16.55

Network size	Number of Features	10% Training Dataset
40	10	28.69
40	50	17.50
40	100	16.39
40	110	14.53
40	160	15.90
40	200	14.50
40	250	14.04
40	304	11.38

Table 6-5 and 6-6: The EER (%) of normal walking activity utilising different feature subsets

Concerning the fast walk activity, the reported results can be directly compared; 10% of training dataset results was better. The network size with 40 neurons and the full feature vector achieved the best performance of 11.32% as illustrated in

Table 6-7 and

Table 6-78.

Network size	Number of Features	10% Training Dataset	100% Training Dataset
40	10	26.8	29.8
40	40	19.47	21.3
40	50	17.56	20.10
40	100	15.20	18.25
40	150	14.60	16.10

Network Size	Number of features	10% Training Dataset
40	10	26.8
40	40	19.47
40	50	17.56
40	100	15.20
40	150	14.60
40	200	13.84
40	250	13.38
40	304	**11.32**

Table 6-7 and Table 6-8: The EER (%) of fast walking activity utilising different feature subset size

Network Size	Number of Features	Normal and Fast 10% Training Dataset
40	10	27.74
40	50	17.56
40	100	15.20
40	160	14.56
40	200	14.17
40	250	13.69
40	304	**12.49**

Table 6-9: The EER (%) of normal and fast walking activities utilising different feature subset size

The same network size with 40 neurons and the full feature vector achieved the best performance of 12.49% when normal and fast activities were combined, as shown in Table 6-9.

Activity Type	EER (%)
Normal	11.38
Fast	11.32
Down Stairs	24.52
Upstairs	27.33
Normal & fast	12.49
All Activity	15.08

Table 6-10: The best EER (%) for individual and all activities

Table 6-10 shows the classification system performs with the multi-classifier in comparison with single classifier. As expected, the system performance dropped with the walking up and down stairs activities. Consequently, the proposed multi-algorithmic approach tended to perform better than the individual activities (apart from walking down and upstairs) across the real-world dataset, as most of them achieved better performance than when they were treated as one activity.

Figure 6-5 illustrates the users' profiles for the normal activity as an example of the multi-algorithmic and Figure 6-6 illustrates the users' profiles for all activities as a single classifier.

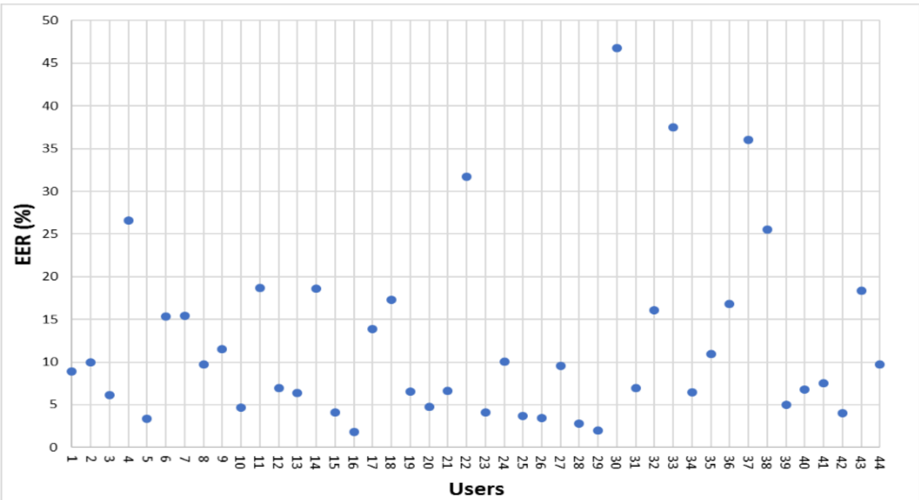


Figure 6-5: The EER (%) of individual performance for normal walking activity

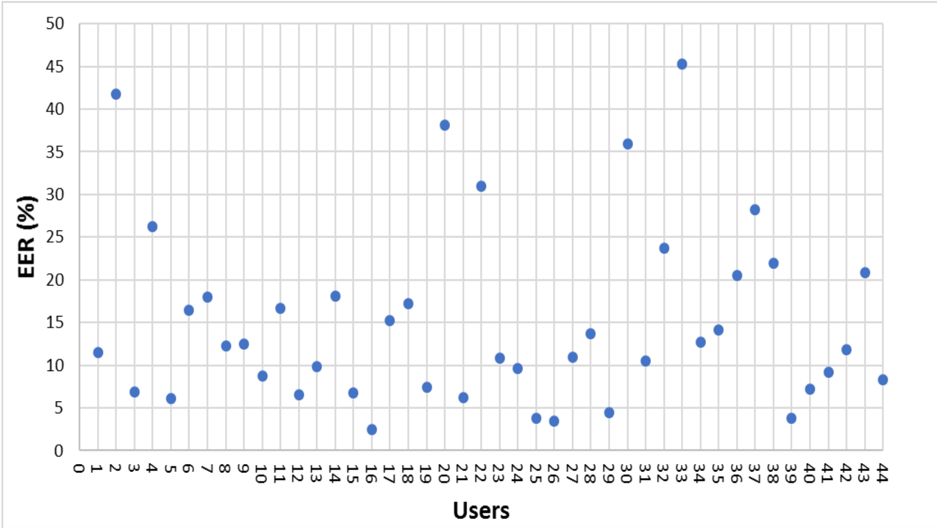


Figure 6-6: The EER (%) of individual performance for all activities

Figure 6-5 illustrates the individual performance from the best EER of 11.38%. The figure exhibits that a significantly wide range of users' performance was under 10%. With 1.94% of User 29 being the best and 46.80% of User 30 being the worst. It was found that 12 participants (Users 5, 10, 15, 16, 20, 23, 25, 26, 28, 29, 39 & 42) achieved an EER of less than 5% each whilst another 6 (Users 26, 22, 30, 33, 37 & 38) accomplished an EER of more than 20% each. From the analysis, it is clear that the majority of the participants scored less than 10% in the multi-algorithmic approach. In contrast, with a single classifier, the majority of participants scored greater than 10%, as shown in Figure 6-6.

The real-world results were not as good as the controlled experiment as human's behaviour does change over time, in addition to the influence of many environmental factors (e.g., human emotion, time effect, and ground substance, changing clothes and shoes. etc.). In spite of this, the results above are presented for "single-sample mode" and none of the voting techniques were employed. However, the presented results are still promising. Therefore, the majority of voting was exploited, and the results are presented in the next subsection.

6.5 Applying Majority Voting

The decision to accept or reject the output done by the system depending on the rating results. Previous studies (Nickel, Brandt, et al. 2011a; Nickel et al. 2011) have primarily concentrated on two standard programs: majority or quorum voting. Improved performance is typically achieved using quorum voting technology. However, the system is more resilient to errors when applying majority voting. With quorum voting, a small number of valid rating results are required for user acceptance. While this improves the users' consolation (a user will maybe get to deploy such a system), this will lead to a high false acceptance rate, and that is, the spoofer is likely to misuse the system. From another view, user behaviour is more distinct when using majority voting; then, the system will produce a high false rate of rejection. The system would provide greater security when using majority voting; at the same time, the system is more invasive (not user-friendly). Consequently, it is necessary to have suitable decision logic to stabilise the system security and user for the authentication procedure. Ultimately, this study applied majority voting rather than the quorum voting schema.

So far, all the results submitted were founded on a single sample classification for the EER calculation; the achievement in Table 6-10 gives good results. It is motivating to regulate the possibility of reducing the number of trials rejected by an original user.

The first is a structure that accepts the user as original if at least half of the user's test samples are positive (i.e., at least 50% of the results are a match); then the biometric resolution merges several classification outputs into one. The latter is a method that authenticates a native user if the required number of user samples is positive.

Three different investigational tests were conducted to explore how various activity types affected the authentication rate.

- First, each activity was tested as a single formula (i.e., normal, fast, down and upstairs walking), which represented a multi-algorithm approach.
- The second test merged normal and fast into a single activity formula.
- The third test studied the correct classification rate for all the activities, which represented a single-algorithm approach.

As expected, the performance of the real-world dataset was poorer than the controlled circumstance. Mean and median measures were considered in this experiment, as the mean was more sensitive to outliers while the median was not, in order to explore to what extent the system performance was affected by the outliers. Figure 6-7 investigated the first experimental setting, comparing different real-world activities and authentication efficiency. Figure 6-8 exhibits the mean of the second and third tests. The majority voting results were obtained when involving a different number of sections (i.e., 15 sections ranging from three to thirty-one, 10 seconds each). Furthermore, Figures 6.9, 6.10 (a) and (b), and 6.10 represented the median for different real-world gait activities' authentication efficiency. In addition, Figure 6-11 (a) and (b) exhibit the normal and fast activities and four activities types, respectively.

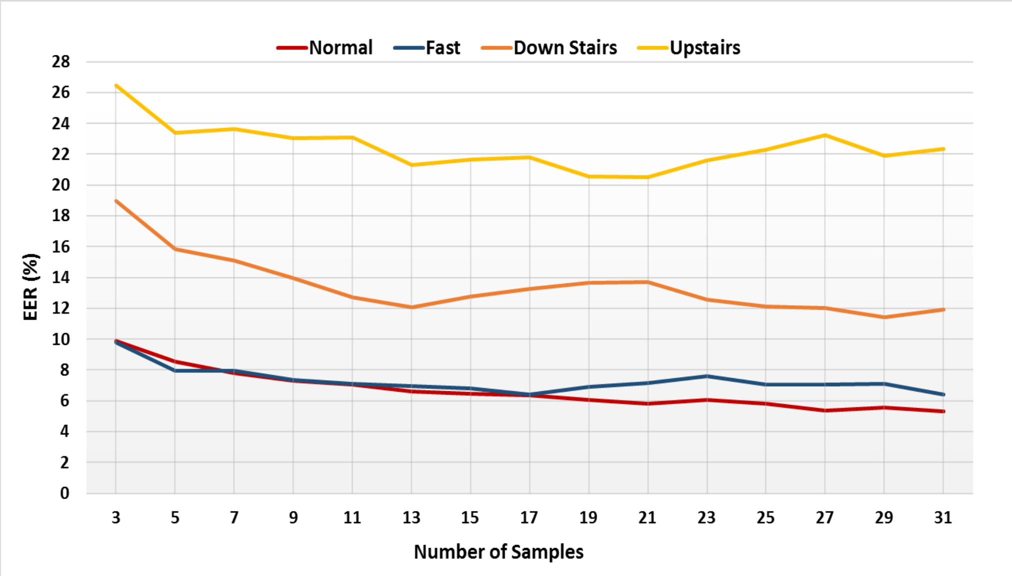


Figure 6-7: Majority voting mean values using different numbers of data samples for multi-algorithm walking activities (10-second sample period)

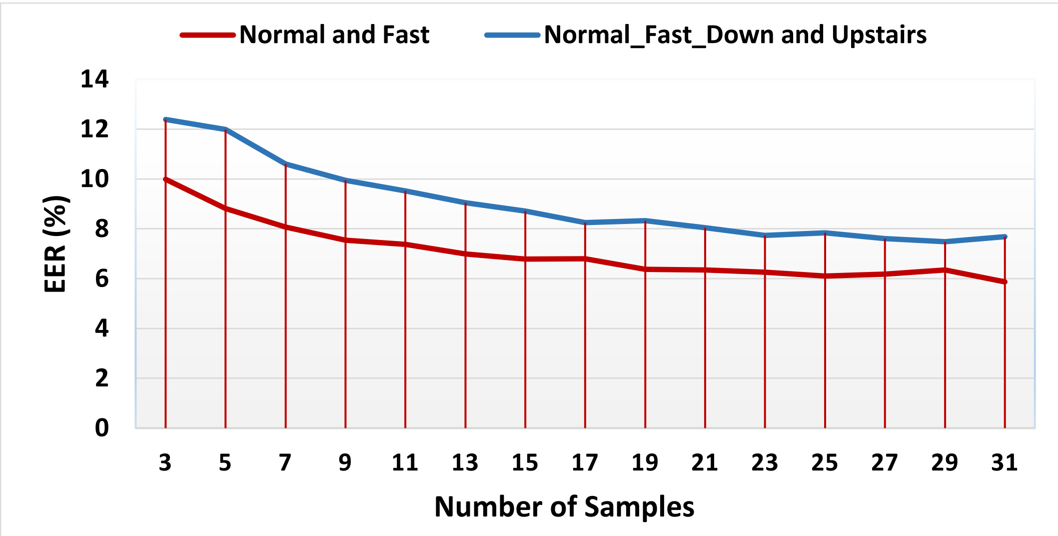
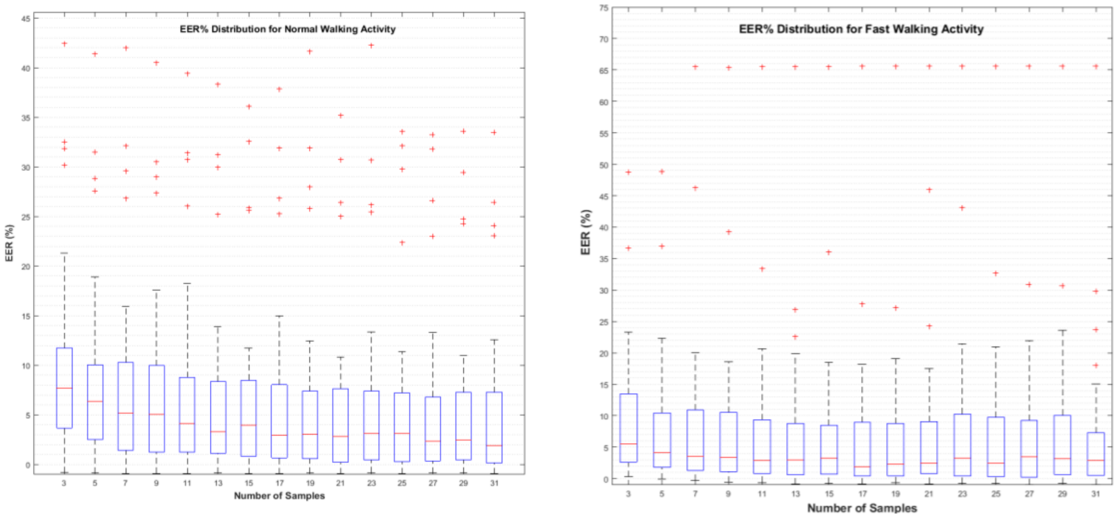
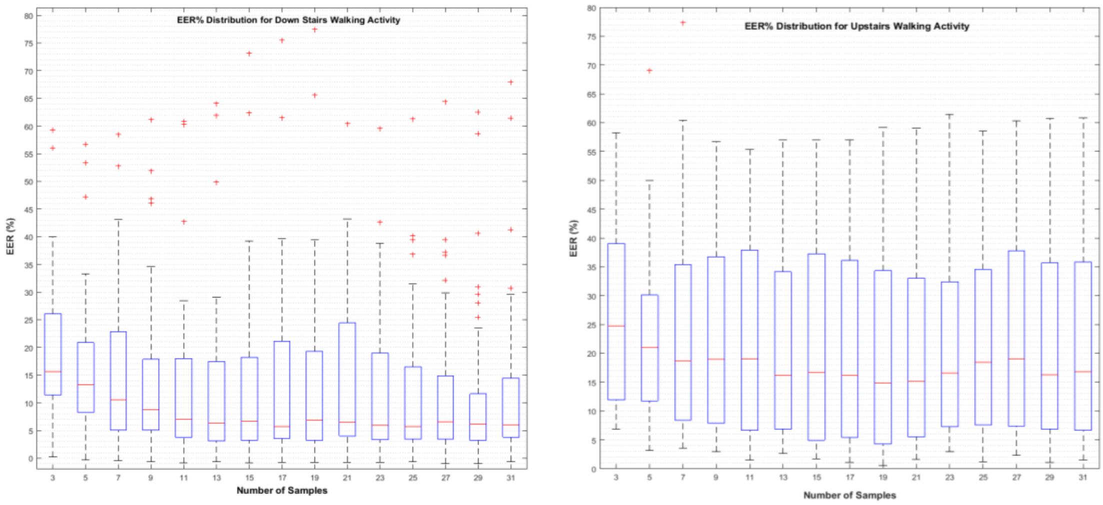


Figure 6-8: Majority voting mean values using different numbers of data samples for single algorithm walking activities (10-second sample period)



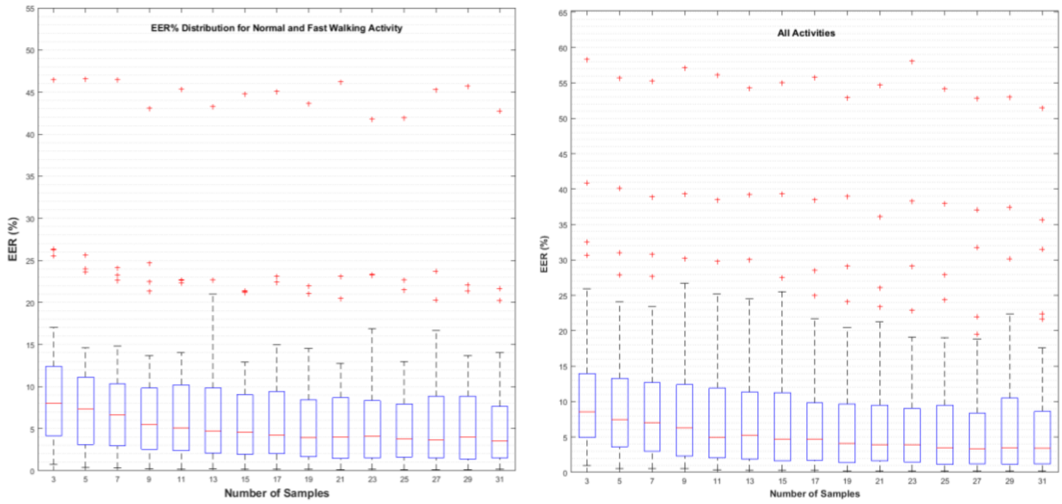
(a) (b)

Figure 6-9: Majority voting median values using different numbers of data samples for (a) normal walking, (b) fast walking.



(a) (b)

Figure 6-10: Majority Voting Median Values Using Different Numbers of Data Samples for (a) Down Stairs, (b) Upstairs walking



(a) (b)

Figure 6-11: Majority voting median values using different numbers of data samples for (a) normal and fast walking, (b) all walking activities

Generally, the normal walking activity achieved better median values with more numbers of samples. It is clear from Figure 6-9 (a) that the lowest median was 2.14% employing 31 samples (5:10 minutes). Also, it can be seen that Figure 6-9 (a) has several outliers values ranging from 22% to 24%, which could affect the performance negatively, while the resulted median reduced by 50% in comparison with the mean considering the same number of samples.

In contrast, the fast walking median range decreased gradually to obtain the best median utilising 17 samples (2:50 minute) as presented in Figure 6-9 (b). Then the median values increased with using more samples. Furthermore, fewer outliers appeared. This means fast walking could have more distinctive features helped with better recognition leveraging fewer samples.

Figure 6-10 (a) and (b) show the median values for down and upstairs walking activities in which the downstairs activity experiment achieved better performance than upstairs, by reducing the error rate by 70%. This can be explained as the high values of outliers affected the overall error.

That resulted in the median values of 5.65% utilising 25 samples comparing with 14.81% achieved by upstairs activity using 19 samples.

There was no significant difference between the median values for normal and fast and all activities. It is clear from Figure 6-11 (a) that the lowest median for normal and fast activities was 3.50% employing 31 samples (5:10 minutes) and for all activities, it was 3.63% considering 37 samples (4:30 minutes), as shown in Figure 6-11 (b). The summary of majority voting results for all the experimental settings is presented in Table 6-11.

Table 6-11 shows the mean and median dataset observations for the six activities utilising the majority voting scheme. As shown in Table 6-11, they produces significant enhancement on the system performance. In comparison with a single-sample evaluation, normal, fast, down and upstairs walking activities were improved by an average rate of 53.34%, 45.52%, 53.39%, and 24.81%, respectively. Moreover, analysing the performances for the merged normal and fast and the four combined activities demonstrates quite better improvement with an average rate of 46.99% and 49.40% accordingly.

If the median is considered as a scale of a system performance, which is less sensitive to the outliers and could affect the achievement by shifting the average EER, it can be seen that EER median-based were quite better than the EER mean values. The EER dropped down to 2.14%, 1.89%, 5.65% and 14.81% for individual activities (i.e., normal, fast, down and upstairs). These results demonstrated the adverse effect of the outlier's values, which are shown in the figures above. Whereas, the approximated range of outlier values between is (22%- 80%).

Activity Type / # Samples / Time (second)		Normal	Fast	Down Stairs	Upstairs	Normal And Fast	All Activities
3 (30)	Median	7.93	5.50	15.65	24.72	8.02	8.53
	Mean	9.90	9.77	18.97	26.48	9.99	11.99
5 (50)	Median	6.41	4.19	13.24	21.08	7.39	7.50
	Mean	8.52	7.95	15.85	23.41	8.82	10.60
7 (1:10)	Median	5.38	3.38	10.57	18.69	6.64	7.08
	Mean	7.78	7.96	15.08	23.62	8.07	9.95
9 (1:30)	Median	5.29	3.38	8.74	18.98	5.48	6.27
	Mean	7.30	7.35	13.95	23.05	7.54	9.52
11 (1:50)	Median	4.52	2.91	7.02	19.05	5.10	4.93
	Mean	7.06	7.11	12.71	23.10	7.37	9.04
13 (2:10)	Median	3.56	2.96	6.32	16.13	4.70	5.24
	Mean	6.62	6.96	12.05	21.28	6.99	8.72
15 (2:30)	Median	4.65	3.30	6.65	16.64	4.57	4.70
	Mean	6.44	6.79	12.75	21.65	6.79	8.24
17 (2:50)	Median	3.46	**1.89**	5.71	16.18	4.26	4.71
	Mean	6.34	6.40	13.24	21.79	6.80	8.33
19 (3:10)	Median	3.13	2.24	6.88	**14.81**	3.92	4.10
	Mean	6.05	6.89	13.67	**20.55**	6.37	8.04
21 (3:30)	Median	3.06	2.48	6.48	15.12	4.05	3.96
	Mean	5.79	7.16	13.69	20.49	6.35	7.74
23 (3:50)	Median	3.21	3.25	5.94	16.58	4.06	3.95
	Mean	6.08	7.59	12.54	21.59	6.26	7.84
25 (4:10)	Median	3.32	2.48	**5.65**	18.42	3.82	3.52
	Mean	5.80	7.04	12.12	22.27	6.11	7.61
27 (4:30)	Median	2.48	3.48	6.59	19.02	3.67	**3.36**
	Mean	5.37	7.05	12.03	23.22	6.19	7.48
29 (4:50)	Median	2.90	3.15	6.15	16.25	3.99	3.49
	Mean	5.55	7.09	**11.43**	21.89	6.34	7.69
31 (5:10)	Median	**2.14**	2.89	6.06	16.76	**3.50**	3.38
	Mean	**5.31**	**6.43**	11.91	22.33	**5.87**	**7.45**

Table 6-11: Majority voting results for each number of samples across all gait activities

This is an interesting result outcome with the real-world dataset, that multi-algorithm authentication seems to be more reliable than a single algorithm, apart from the stairs activities, whether before and after applying the majority voting scheme and with using mean or median performance scales.

6.6 Discussion

In this study, a novel multi-algorithm approach was evaluated using real-world data gait recognition. Although the quality of real-life data was considered noisier and less reliable than the controlled data, the accomplished results are promising.

It reflects high possibilities to deploy the proposed mechanism to support existing active mobile authentications such as PIN or password in reality. As there is currently no real-life dataset in the mobile gait authentication field, the comparison with related works will be no relative. In comparison with existing prior studies, this research utilised a real-world dataset containing a more significant number of gait samples employing 44 participants during (7-11) days.

As the performed real-life activity dataset in various environments, the activity recognition is considered a crucial process to split the data into multiple activities (i.e., normal, walking with a bag, fast, down and upstairs). As presented before in Section 4.5.3, predictable data modelling has been built that can classify a given individual's activity signal into a predefined class, based on the features extracted from the raw sensor data. Three different investigational settings were conducted to study how various activity types affected the identification rate. First, as normal walking and walking with a bag were the most similar activity types, they were joined as a single activity formula. The second test merged normal, fast, and walking with a bag into a single activity. The final analysis examined the correct classification rate for all the activities. Two types of voting were employed: hard and soft majority voting. It can be seen that the soft voting approach outperformed the other models in all tests.

The findings of this study provide evidence that it is possible to recognise a person's physical activity with a high degree of accuracy, reaching nearly 98%, based on smartphone-embedded gyroscope and accelerometer sensor signals gathered over two days. This was achieved by leveraging the capabilities of machine learning algorithms in two stages: feature ranking, in which the feature space was ranked based on the multiclass classification approach, followed by activity identification, in which only top-ranked features were included within the classification phase. The soft majority voting approach provides the highest accuracy in comparison with other models, such as single classifier or hard majority voting.

With respect to feature types, all the features for accelerometers and gyroscopes, time, and frequency domains to were utilised, as a result of the potential of changing the user behaviour signal over time with the real-life scenario. Another key thing to remember is the most repeated features, in the comparison between the controlled and real-world dataset, it has been found that the most frequently repeated features for each user changed with the real-world signal dataset along with each single activity data set. The nature of feature measurement was different, which obviously worked better for more variable signals. Whereas, in the controlled data, the feature worked with more limited numbers. Hence, in reality, the feature vectors need to be quite different. This different scenario feature could be explained because of the type of characteristic, realistically and more variable based inputs. Furthermore, in light of Figure 6-4, the data points in the 2D plot of time between peaks for accelerometer and gyroscope sensors (i.e., the most repeated features in the real dataset), demonstrates the obvious shifting between the two covariates (features) with real and controlled datasets.

Above all, these features might be quite different from previous studies because most of the previous studies focus on the control environment and this probably advocates very similar feature sets.

Moreover, one of the most interesting findings of this study is that the top six ranked features (i.e., the most repeated features), as illustrated in Table 6-12, are based on the time between peaks feature type for both accelerometer and gyroscope sensors. This could be interpreted that the cycle of walking, as measured by the peak to peak, is higher for each person.

Feature Order	#of Repetition	# Of Accruing (%)	Sensor/ Axis	Feature Name
243	136	6.18	Gyro/z-axis	Time between Peaks
89	133	6.04	Acc/x-axis	
241	132	6	Gyro/x-axis	
90	126	5.7	Acc/y-axis	
91	115	5.2	Acc/z-axis	
242	93	4.22	Gyro/y-axis	

Table 6-12: Top six ranked features (time between peaks)

Importantly, the number of features utilised substantially throughout the range of top-ten repeated features for all users with real-life dataset effectuated a sharp decrease in a number, about 50% of the same top-ten features calculated with controlled dataset experiments, as provided in Table 6-13 bellows:

Activity Type	# Features Used	
	Controlled Dataset	Real-life Dataset
Normal	84	39
Walking with a Bag	81	
Fast	70	48
Down Stairs	139	58
Upstairs	146	70
All Activities	66	33

Table 6-13: Number of features included in the repeated top-ten

With real-world data, the features are more variable, and in terms of the half, the number of features in the previous is consistent and appears in the top ten.

Overall, these indicate that the data is more variable, and the nature of features is varying. So, the nature of the classification problem is more challenging.

With respect to the performance of gait recognition across different walking activities, Table 6-10 shows that the individual activities normal and fast, apart from the stairs walking activities, succeeded in accomplishing better performance, hence surpassing the results of combing normal, fast and all activities. Therefore, in the comparison between a single classifier and multi-algorithmic approaches across the real-world dataset, the normal and fast activities performance were 11.38% and 11.32% accordingly. While the EER obtained when normal and fast activities were merged was 12.49% and the EER of 15.08% when all activities were combined.

The down and upstairs walking results are considered high during both datasets. That could be attributed to the belt phone pouch mostly wobbling more with walking up and down the stairs and stairs style (e.g., once a subject reaches the bottom of the stairs, the individual walks a few more steps to complete the activity, which might become a noisy/outlier sample in the stairs activity dataset).

Activity Type	# Users				
	ERR (%) <=5	ERR (%) >5-10	ERR (%) >10-15	ERR (%) >15-20	ERR (%) >20
Normal	12	16	4	6	6
Fast	13	13	7	5	6
Normal and Fast	8	18	7	5	6
All Activates	5	12	10	7	10

Table 6-14: Individual performance for each activity

Table 6-14 illustrates the best EER (%) of FF-NN individual performance range for single and multi-algorithmic approaches, with regard to the individual normal and fast activities average EER of 11.38% and 11.32%, respectively. The analysis of individual error rate shows that the majority of the subjects performed better than the average performance for both individual activities. About 63% of the normal users' results and 59% of the fast user's results obtained an EER lower than 10%. Also, the merged normal and fast user's performance found that about 59% of the users obtained lower than the average EER (i.e., 12.49%). Conversely, with all activities and the average EER of 15.08%, only 38% of users had EERs less than 10% compared with individual activities. Also, more than two-thirds of the users encountered less than 10% and under 5% of users' results represented the lowest ratio of users' results.

All things considered, it seems reasonable to assume that the multi-algorithmic approach results are better than the single classifier approach.

As mentioned before, the real-life performance was lower than the controlled dataset. Therefore, implementing the majority voting on the calculation of error rates was deemed essential to enhance the results of real condition activities. Furthermore, the median measurement was considered on result calculations, in order to explore the outliers' effect on system performance.

Activity Type	Best Voting EER (%)		Time
Normal	Median	2.14	5:10s
	Mean	5.31	5:10s
Fast	Median	1.89	2:50s
	Mean	6.43	5:10s
Down Stairs	Median	5.65	4:10s
	Mean	11.43	4:50s
Upstairs	Median	14.81	3:10s
	Mean	20.55	3:10s
Normal &Fast	Median	3.50	5:10s
	Mean	5.87	5:10s
All Activities	Median	3.50	4:30s
	Mean	7.45	5:10s

Table 6-15: System performance utilising the majority voting module

When comparing the majority voting module performance in Table 6-15, one can see that the results decreased when they were based on a longer time (i.e., larger walking samples). If the mean measure considers the overall average population of EER, then it can be seen from Table 6-11 that 31 samples (i.e., 5 minutes and 10 seconds) give the best results, for normal and fast individual activities of EER 5.31% and 6.43%, respectively. Moreover, the same period (5 minutes and 10 seconds) achieved better when combining normal and fast activities and all activities together. The down and upstairs best performance was achieved with 29 (i.e., 5 minutes and 50 seconds) and 19 (i.e., 3 minutes and 10 seconds) data samples accordingly.

On the other hand, if the median is selected to be a metric, mostly a smaller time was needed to have best EER median-based results (i.e., 2:50 s, 3:10s, 3:20s and 4:30s) for fast, upstairs, downstairs and all activities, respectively. However, normal and merged normal and fast activities employed longer time (i.e., 5:10s) to obtain the best results. It is apparent that some outliers sit far from their group. However, outliers were included in the classification tests and were not excluded from any process within this experiment, as they were real-world samples.

6.7 Conclusion

The evaluation of the smartphone-based, gait authentication system over a long period of time under realistic scenarios has revealed that it could provide a secure and appropriate activity identification and user authentication system.

As predicted, the real-life results were higher than the control dataset for all activities. This is because walking behaviour is changed from day to day as the participants mostly were wearing different shoes and clothes. Coupled with the participants' mode in this different day was perhaps different. However, the presented results are still promising with respect to rejection of impostors and accepting genuine subjects, notably when the majority voting techniques were applied, which improved the classification up to 50%, and proved before with the controlled experiment results that multi-algorithm authentication seems to be more reliable. With the results above, especially when a median measurement was employed, normal and fast walking had better performance apart from stairs walking activities. This may give evidence to exploit a multi-algorithmic approach with context awareness data to enhance the performance.

As well, the nature of features measurement was different between the control and real-life data types, which obviously worked better for more variable signals. Whereas, with the control data, the feature is there to work with more confined numbers. The results are shown using the entire feature vector performed better performance (i.e., the longer feature vector is provided to have more reliable achievement). Furthermore, the attribute types required with a decision being made to verify subject samples for a successful authentication mechanism were identified.

7 Discussion

7.1 Introduction

As the smartphone and its services and information are becoming targets of cybercrimes, it is mission-critical to secure smartphones and their services and information. Gait authentication has gained significant attention for use in authentication on mobile devices and this is because of its usability and convenience. The user does not need to provide an explicit action for mobile authentication because the related data is continuously recorded while the person is walking.

A set of experiments were conducted in this work for transparent user recognition utilising gait patterns, evaluated in Chapters 5 and 6. The results of these experiments, nature, and the amount of the collected dataset and classification techniques employed will be compared with previous work related to smartphone-based gait signals to evaluate the viability of the proposed system.

However, the complexity and the high inconsistency of gait patterns limit the capability of gait recognition systems and adversely affect their validation, especially on real environment systems. It is envisaged that other sources of information (including surface material/condition, walking speed, carrying an object, moods, and weather) could be used to understand the context in which the gait information is collected, and more informed and accurate authentication could subsequently be made. Also, there is additional information that can be collected via various sources within the mobile device itself, such as GPS, weather forecast, calendar, and emails.

Therefore, this chapter will present a discussion surrounding factors that can improve performance through integration research by context-awareness gait on real environmental systems information to provide continuous and transparent security for mobile devices using the gait information collected via accelerometers, gyroscopes, and GPS sensors, while the context-awareness data can be gathered from various sources, including, Wi-Fi information and installed mobile applications. Additionally, this could offer the ability to explore the efficiency of these two techniques within the transparent authentication system (TAS).

7.2 Comparison with the Prior Art

As reviewed in Chapter 3, Table 3-9, a comprehensive analysis of prior studies on gait authentication systems using mobile sensors, previous literature on gait recognition has potential and a lot of work in gait recognition has been undertaken. However, the studies were somewhat limited in scope (e.g., limited dataset and very controlled experimental environments). Although their results were desirable, the situation could be very different if the technique was applied to live data, as the information can be very noisy. Also, most of the studies used only the accelerometer sensor and minimal works utilised two sensors. However, no research has been found that seeks to employ additional information on the process (such as GPS or weather info) to advance the state of knowledge and enable a better decision-making process. It is difficult to compare with these studies as a result of the different datasets (e.g., a number of subjects, walking time, and activity type) and data collection settings (e.g., smartphone type or device location). It is clear from

Table 7-1, a comparison between some common selected smartphone-based prior datasets, that none of the previous systems had attempted to cover a wide variety of real-world datasets in seven consecutive days a week (i.e., study the potential for the general use in realistic circumstances).

However, to the best of the author's knowledge, no research has been found that has so far explored mobile-based real-world signals. Accordingly, there is a need to propose a real-life system with more user-friendly, real scenarios (i.e., unconstrained conditions).

No.	Author/ Year	Sensors	#Users	Data Description
1	(Frank, Mannor and Precup, 2010)	Acc	20	A controlled environment, two sessions of 15 minute walks on two different days
2	(Derawi *et al.*, 2010)	Acc	51	A controlled environment, normal walk, two sessions of two minutes each with CD
3	(Nickel *et al.*, 2011)	Acc	48	A controlled environment, normal walk and climbing stairs, two sessions of 15 minute walks on two different days
4	(Ngo *et al.*, 2014)	Acc.	744	A controlled environment, only two data sequences for each participant (session of about 1 min)
5	(Gadaleta and Rossi, 2018)	Acc, Gyro, and manometer	50	A controlled environment, several acquisition sessions, five minutes for each participant
6	Our dataset (1)	Acc, Gyro, GPS	60	A controlled environment, walking normally, fast, and normally with a bag on a predefined route, six minutes each activity; walking downstairs and upstairs for three levels on two different days. In variable conditions, e.g., with different shoes and clothes
7	Our dataset (2)	Acc, Gyro, GPS	44	An uncontrolled environment, longitudinal live usage data (real-life), 7-10 days for each user

Table 7-1: A comparison between some common selected smartphone-based databases

- The dataset is an essential part of the identification and authentication process; an algorithm could give different results depends on the set of data (Gadaleta & Rossi 2018). However, some datasets are publicly available such as the largest set available at the Osaka University (Ngo et al. 2014). This dataset is based on three internal sensors placed on the subject's belt, with a triaxle accelerometer and a gyroscope. However, a smartphone was worn in the centre back waist and only measured the triaxle accelerometer data. This data set contains data collected from 744 subjects. With this high number of contributors, this data set has a significant problem, which is based on a controlled environment. Also, for each participant, there were only two data sequences available (session of about one min), which was not enough for network training. Moreover, the gyroscope (from smartphones) data was not provided. Some other datasets are accessible, but for a much smaller number of participants.

  Consequently, two datasets were constructed: a controlled dataset (as explained in sections 4.4.1 and 4.5.1) and a realistic dataset (unrestricted to influence of many environmental issues, such as changing clothed and shoes, in a rush, carrying luggage, running as a result of poor weather, exercising to human mood, time effect, and ground substances, to name but a few), in order to have a fair and comprehensive evaluation mechanism. There has not been any research examining this real dataset (to the best of our knowledge). Soft biometrics such as (i.e., age, gender, height, weight) were gathered in addition to gait pattern behavioural characteristics, which are easy to collect but not distinct as the physical and behavioural biometric data (Karabatis 2017). However, there is a minor unrealistic restriction in this study; the device is supposed to be fixed in the belt pouch during the data collection phase.

- Research on gait authentication with smartphone-based signal data and dynamic features is relatively very low. Nakano, (2017) studied the impact of the dynamic features on the activity recognition system performance. Their analysis revealed that the performance of the efficiency of dynamic features was better than static features in the classification of different activities, especially with the CNN classifier, which is better than static features with SVM. However, with the cycle- and segment-based approaches, some researchers have utilised deep learning to meet the challenges of the feature extraction process. With recent advances in deep learning algorithms, the use of a convolutional neural network (CNN) learning algorithms to extract a latent pattern from raw data has become common practice (Jiang & Yin 2015; Ronao & Cho 2016). Typically, deep learning approaches require less effort in feature extraction and engineering in comparison to cycle and segment-based approaches. However, a challenging aspect in deep learning-based models is that it is hard to explain and interpret how decisions are made (Weld, D. S., & Bansal 2018). Knowing what drives decisions in models (i.e., the features on which the model relies) is an essential element in some activity recognition applications, such as healthcare-related research.

  As a result of the lack of applying the feature selection process in the literature (reducing the number of features used and attaining more discerning information) and even using this large number of various features for mobile-based two sensors. As extensive feature vectors increase the complexity of classification algorithms and negatively affect the decision speed. Consequently, in a controlled experiment (Chapter 5), in each of accelerometer and gyroscope sensor data signals investigated, nearly half of the feature vectors were used to get the best results with the time domain feature values.

Even so, when time and frequency domains feature vectors were combined for each of accelerometer and gyroscope sensors, the dynamic feature performance was better than static features in classification of various activity datasets (i.e., normal walk, fast, carrying a bag, down and upstairs) while there were no significant differences found between the dynamic and static results with all activities dataset when the signals of two sensors were merged. Nevertheless, in practice, this dramatic dropping down of features utilised (only 11% and 19% of the features were used with the SD and CD respectively) will reduce the system complexity and the burden on the classifier.

Consequently, our finding revealed that the controlled dataset experiment and the dynamic feature selection process outperformed those obtained by using the full feature set (i.e., 304 features) from both accelerometer and gyroscope signals, SD, and CD scenarios.

- It is apparent from the prior studies discussed in Table 3-9 that most of the classifiers utilised are a neural network algorithm, k-NN, HMMs, SVMs, GMM, and random forest (RF). This work employed the SVM, the feedforward neural network, and RF classifiers. Regarding SVM, the results were satisfactory with the controlled dataset, but the performance conducted by the feedforward neural network outperformed the SVM classifier, with different feature vectors subsets and various activities considered in the first experiment. With the real-life seven-day dataset, the SVM did not work correctly because it cannot work with a large data volume, as this type of classifier can be implemented only with a limited capacity of data. Hence, it might not be easy to manage with SVM in practice. Subsequently, an alternative decision tree algorithm, random forest classifier, was selected, but the EER results were high. Accordingly, the FF-NN was used instead, as it gave the best results.

- For smartphone-based gait authentication, single and multi-algorithmic approaches were developed in this thesis. The generative model is a novel multi-algorithmic approach (i.e., where different classifiers were used based on the nature of the activity). Various activity datasets (i.e., normal, fast, carrying a bag, down and upstairs) were employed to evaluate these approaches. The main findings in this regard are presented in Chapter 5. Furthermore, to the best of the author's knowledge, there is no prior work that extensively examined a universal algorithm that can authenticate a smartphone subject with multi-algorithmic gait activity signals. In conclusion, the findings of this study explored that a multi-algorithmic approach can achieve a better level of performance over a single classification approach.

A comparison between the controlled experiment results and the prior studies on gait authentication systems using the mobile sensors is discussed in Table 3-9. In terms of performance, the best results were 0.70% EER for the normal walk activity, which was better than the performance of existing studies' 1.95% EER of (M. O. Derawi 2012) and 1.82% EER of (Watanabe 2014). Under the cross day, 6.30% EER for the same activity was in line with prior work, including 6.1% EER (Nickel, Brandt, et al. 2011c) and 6.15% EER (Muaaz & Nickel 2012; Watanabe 2015). Those three prior studies employed the majority and quorum voting technique, which may improve the classification by up to 50%. In addition, they utilised 20% fewer users for their experiments than this study. Hence, it could be more accessible to distinguish individual users.

The best results were obtained for fast walking of 0.42% EER and 12.70% EER under SD and CD scenarios accordingly, which was better than the performance of the presented study's 14.39% EER and 15.43% EER using HMMs and SVM, respectively (Nickel, Brandt, et al. 2011a), which compared the efficiency of HMMs and SVMs for accelerometer-based biometric gait authentication under the CD scenario.

Concerning the stairs results, these percentages seem large in comparison with other activities. On the other hand, there were considerable reductions of the supposed features used (i.e., 3% and 6% features used for down and upstairs, respectively, under the cross-day scenario).

As mentioned above, an identity recognition algorithm could give different results depending on the set of data. Furthermore, several studies have revealed that each different classifier performance may differ. For instance, down and upstairs classifiers have less discriminative attributes than the walking classifier (e.g., normal or fast walking) (Kwapisz et al. 2010; Nickel et al. 2011; Watanabe 2014; Watanabe 2015). Therefore, the normal, fast walking, and carrying a bag classifiers performed better than the down and upstairs classifiers.

Regarding the 'all activities' dataset EER of 4.4%, this was compared with EER of 18.38% obtained by mixing two speed data (i.e., normal and fast) (Nickel, Brandt, et al. 2011a). As a result, it is apparent from the above analysis that Experiment 1 collecting significant multi gait activities dataset from 60 users over two days obtained better performance than prior work.

The control dataset (Experiment 1) aimed to understand more activities, using more data and more people, to provide intelligence on features and classifiers that could be more feasible with real-life based activity signals. Consequently, the second experiment applied the best parameters learned from experiment one in terms of features, classifiers, the multi algorithmic approach on the real-life dataset, and a sufficient number of people and data to investigate how performance will be in practice.

7.2.1 Real-World Usability Performance

This section will discuss the real-world dataset verification to investigate the capability of performing the proposed smartphone-based gait authentication system in practice.

As the performed real-life activity dataset in various environments, the activity recognition is considered a crucial process to split the data into multiple activities. The proposed approach is evaluated by building a predictive model that can categorise a given individual's activity signals into predefined classes, based on the features extracted from the raw sensor data of the controlled dataset (e.g., normal walk, fast walk, walk with the bag, downstairs, upstairs, and sitting). In comparison with existing studies in which the data were gathered from smartphones (Chapter six, Table 6-1), most of these studies have fewer participants (i.e., 30 or fewer) and the data were all captured on the same day. In this study, most of the data were collected between two days for everyone within the sample set because the probability that users' activity patterns change is higher for data collected across days than it is for data gathered on the same day. Notably, the developed approach reached a high level of accuracy in identifying human physical activity based on raw smartphone motion sensor signals. The findings of this study provide evidence that it is possible to identify an individual's physical activity with a high degree of accuracy, reaching nearly 98%, based on smartphone-embedded gyroscope and accelerometer sensor signals gathered over two days.

The person must be moving to enable the system recognising all of the time. The real dataset for seven days across 44 people was analysed to have a clear conception about the approximate time that the humans are doing the actual walking activities throughout the whole day. Base on real data analysis, it can be seen in Figure 7-1, the average of daily gait activity time for all users is 80 minutes per day. It was also found that for typical users probably walking about 35 minutes a day, another user may walk about 3:30 or more especially during the weekend.

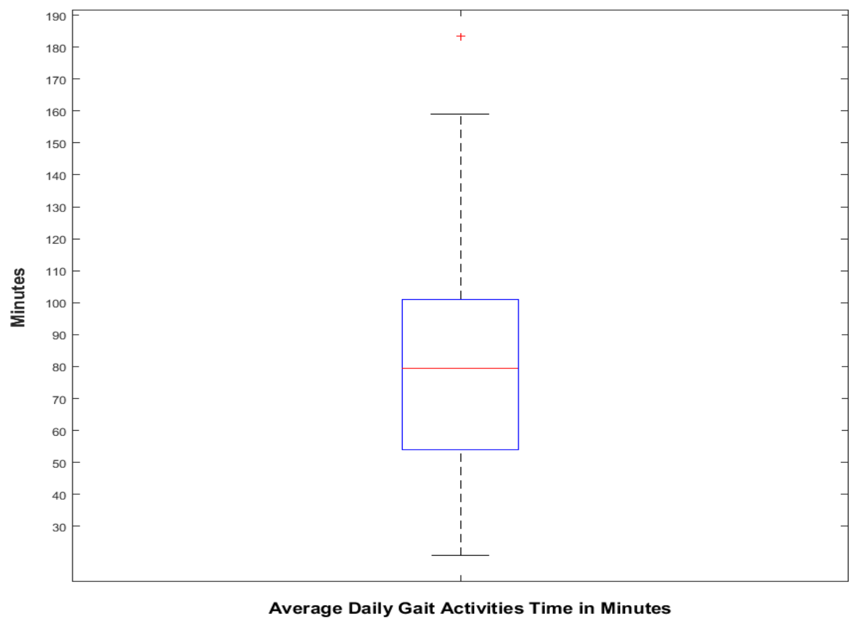


Figure 7-1: The average of daily gait activity time in minutes for all users

In order to have a more accurate insight into the dataset, the percentage of normal, fast, down and upstairs walking activity samples (for a week) are highlighted in Figure 7-1. It is apparent from this table that there is a significant difference between the four groups. What is interesting in this data is that normal walking (including walking with a bag activity) represented the largest percentage of activity samples of 80%, and the lowest ratio of samples was the downstairs activity. Indeed, people are generally walking normally unless they need to walk fast or use stairs.

Activity Type	#Samples	%Samples
Normal	139,907	80%
Fast	12,315	7%
Down Stairs	5,175	3%
Up Stairs	16,999	10%

Table 7-2: Percentage of identified real activities samples

Concerning smartphone resources (e.g., the CPU, battery, and memory), machine learning algorithms require much less computational and memory resources during inference mode. The proposed approach only authenticates smartphone users when a gait activity is detected. In this way, the device processing units (i.e., CPU) and the proposed system uses memory. It can be seen from Figure 7-1 that most of the population performs gait activities from about 35 minutes-3 hours each day. For most of the remaining time, the user either does not use the device or does not perform a gait type activity, in which the proposed system only becomes active during this small period of the day. Therefore, the system consumes a small portion of the device resources during a day of usage.

The performance of the proposed approach with a realistic scenario was the primary concern of Chapter 6. As mentioned before, there has not been any research examining this real dataset. Hence it is challenging to have a fair and comprehensive evaluation mechanism. Therefore, a comparison between the obtained real data error rates with controlled data to perform that the proposed realistic system is not confined to experiment with control conditions. Initially, the results were promising; however; the real-life performance was lower than the controlled dataset. Therefore, in order to have more reliable results, further processing was conducted by combining several consecutive classification results and converting into a single result by employing a majority voting technique. In other words, instead of having one classification result obtained per segment (i.e., the concise walking period of 10-second segments), we designed a more practical classifier arranged around different times.

The majority voting results were obtained when involving a different number of sections (i.e., 15 sections ranging from three to thirty-one, 10 seconds each). The best median and mean performance were calculated to overcome the outlier effect on system performance. Table 7-3 compares the controlled experiment with realistic system performance with and without using the majority voting module for various gait activities.

Activity Type	Controlled Dataset (Cross Day)	Realistic System Without Voting	Realistic System Best Voting		Decision Time
	EER (%)				
Normal	2.09	11.38	Median	2.14	5:10s
			Mean	**5.31**	5:10s
Fast	3.91	11.32	Median	**1.89**	2:50s
			Mean	6.43	5:10s
Down Stairs	23.45	24.52	Median	5.65	4:10s
			Mean	11.43	4:50s
Upstairs	23.32	27.33	Median	14.81	3:10s
			Mean	20.55	3:10s
Normal &Fast	-	12.49	Median	3.50	5:10s
			Mean	5.87	5:10s
All Activities	6.58	15.08	Median	3.50	4:30s
			Mean	7.45	5:10s

Table 7-3: Comparing controlled and realistic system performance with and without using majority voting

It is clear that the majority voting module enhances the results of real condition activities. For example, normal walking results improved by 54% and 88% considering the mean and median measurements, respectively. In general, one can see that the results dropped down when they are based on a more extended period (i.e., larger walking samples). As shown in Table 7-3, thirty-one samples (i.e., 5 minutes and 10 seconds) mostly gave the best results performance for normal and fast individual activities. Moreover, the same period achieved better when combining normal and fast activities and all activities together.

To put it another way, the decision gets better as long as the person walks (better results over a long time). Hence, initially, the classifier could work properly on five minutes because more than this will be quite a long period moreover the high volume of realistic data that supports it.

In contrast, because it is uncontrolled data, there is a need to design a more practical classification system (i.e., majority-based classification system) with the ability to arrange around different times. The time needed for the system to know the user, in case using majority voting, will be determined. For instance, 30 seconds of decision time means three samples (10-seconds-based segment) needed to have a decision in the case that the user walks continuously. Also, 1one minute collected (using six samples), two minutes collected (using 12 samples), five minutes collected (use 31 samples) and so on, and the long period of walking means a better classification rate and tends to produce a better decision. As soon as the sensor information is a flat line (the user stopped walking), the decision will be made. However, in practice, the user maybe walks, for example, 30 seconds and stops. That means it will be challenging to decide using another 30 seconds possibly hours between them. In other words, sample

1 was at 9 o'clock, sample 2 at 10 o'clock, sample 3 at 11:15, the waiting time must be 2:15 hours (from 9 to 11:15). Or short gait disruption such as curbs, sidewalk or slipping because it is uncontrolled data. Thus, there is a need to know how long it takes to this decision to happen or in practice doing classifier arrange around these different times. However, without using majority voting, no need to wait for more samples, the decision could be made in every sample.

With respect to the impact of the proposed multi-algorithmic approach is sufficient for the controlled and real-life experiments as most of the individual activities (apart from walk upstairs for SD) and (fast walking for CD) achieve better performance than when they are treated as one activity.

7.3 Proposed Context-Awareness Model

Providing context is the core of the proposed system. Whilst evaluating the approach, given a collected dataset is possible, the key is to enable an understanding of the context in real-time automatically (not through a manual inspection by a researcher). This process focused on developing automated context-awareness. Whilst information from a variety of mobile sensors and applications can provide underlying information (such as GPS), context needs to provide an understanding of what that information will mean in practice—or at least a probabilistic measure of what it thinks the user is doing. It is envisaged that this will include an investigation of decision support systems and inference engines. For example, the inference engine uses logical instruction or rules to the knowledge base and determines new knowledge. This procedure would emphasise as each new actuality in the information base could trigger additional rules in the inference engine. Inference engines operate mainly in one of two styles (special rule or facts) either: forward chaining and backward chaining.

Forward chaining begins with known facts and broadcasts new facts. Backward chaining starts with goals and performs backwards to verify what points must be maintained so that the goals can be achieved (Jang & Yang 2015).

This experiment suggests involving the use of additional context-based information to enable the biometric system to make a more reliable decision. For example, if a user's gait appears to be faster than normal, an analysis of the calendar might reveal they are running late for a meeting. Therefore, the system could either adopt the classifier (using the fast algorithm) or threshold accordingly because a high degree of availability expected. Likewise, realising a user is heading towards the airport might provide additional information required to understand they are likely to be carrying or pulling a bag and again, the system can adapt appropriately to compensate. This experiment focuses on extracting samples based on context and seeks to develop an algorithm to assist in the decision-making process. This method will lead to an adaptive use that will implement the use of multiple reference templates for users. The proposed experiment aims to provide an empirical evaluation of a realistic gait authentication system. The proposed gait and context model aim to get more reliable authentication decisions acquired from the biometric systems, nevertheless the availability of the signal that causes missing or distorting features of the behavioural biometrics is expected. Details of the key components of the system are described in Figure 7-2, which illustrates the context-awareness gait recognition suggested model.

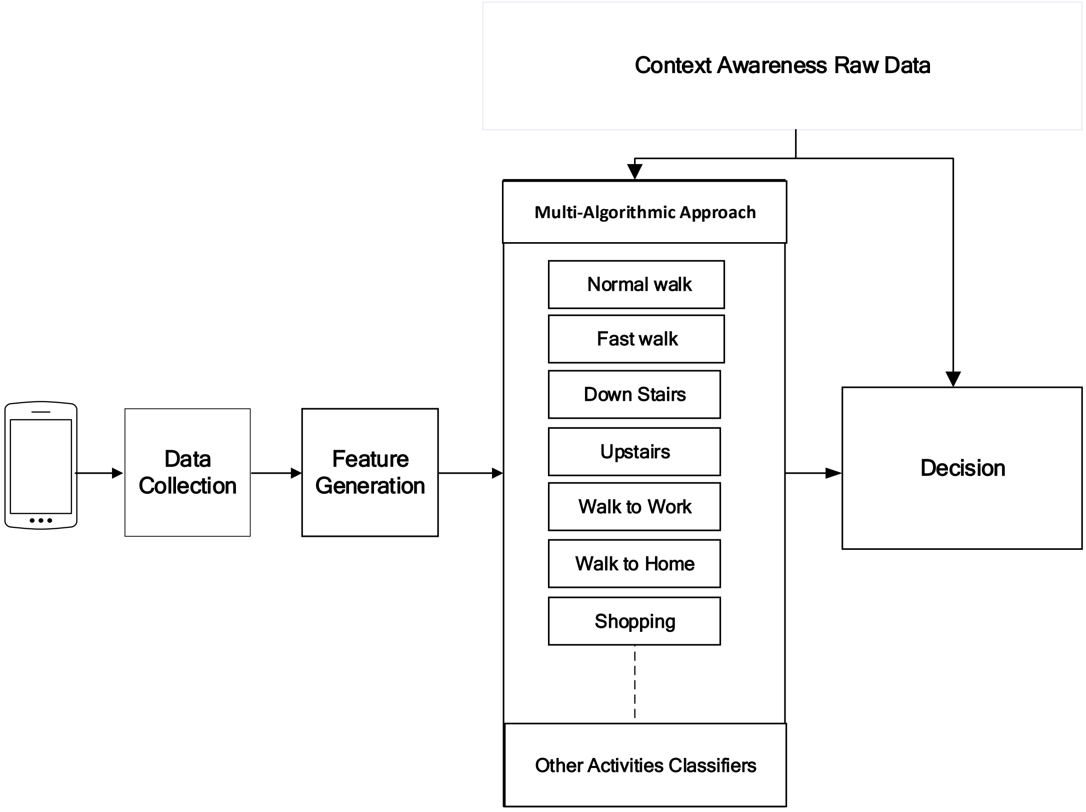


Figure 7-2: Context-awareness gait recognition model

7.3.1 Data collection

In order to allow the proposed model to work effectively, two types of data are required: user's gait signal and context awareness information. Two kinds of data will be collected locally from the mobile device itself. User's gait signal will be continuously gathered from the accelerometer and gyroscope as long as the user walks while the context-awareness information will be fetched under several conditions, including when the gait signal initially occurs, and a drastic change in gait signal happens. Once this information is collected, it will be temporarily saved for further processing.

7.3.2 Feature extraction

Once the raw gait signals are gathered, pre-processing can be started. The raw gait signals will be divided into a fixed-length window. Obviously, the performance will differ when choosing various segments sizes. Then the feature extraction phase, and once the features vector is formed. They will then be forwarded to the next step: either for training or testing purposes.

Regarding the context-awareness raw information, it will be processed into a unified format; the combined information will be used for creating various contexts and contributing the formation of the inference engine, to assist the classification and decision-making processes of the proposed model.

7.3.3 Classification and Decision Making

In the matching phase, the individual samples must compare with the reference template taken primarily at the setup phase (i.e., the feature vector that resulted from the feature extraction process). Consequently, a match score is given indicating the degree of similarity, which decides acceptance of the user's verification claim based on what the authentication decision is. Generally, as noted in chapters five and six, the artificial algorithms (i.e., the feedforward neural network) achieved better performance than statistical methods.

In the proposed system a multi classifier will be created to every single gait motion type (e.g., walking, running, walking under the influence etc.) In each case, an attribute will be added (tag added) to the classification. These attributes will be the output of the context-awareness process (knowledge base and an inference engine results); by adding a tag, it will provide clear identification and indication of making a decision for individual authentication, taking into account the accumulative data will help to create a pattern for an individual. It is envisaged that these will assist in decision making and a more accurate authentication outcome can be obtained.

The proposed system will investigate several techniques to develop a decision support system. It is envisaged that this might involve machine learning, an inference engine, or an expert system.

7.4 Gait Recognition Using Context Information

In a realistic scenario, the classification methods might not be enough to differentiate persons. Therefore, the information that will be provided to the classification methods and decision-making phases (i.e., context awareness, the perception of environmental elements, and the knowledge base) as shown in Figure 7-2, will allow the system to select a proper classifier. Moreover, this will give the decision-making phase a more accurate and precise decision based on the inference engine using forward and backward chaining. For example, some different situations presented as activity detection and gait would make phones applicable as security mechanisms.

- Shopping case: a person shopping would perform large amounts of "**walking and standing"**; in this circumstance, a user performs different activities by walking from one section to another, or from one shop to another etc. In this case data protection is needed to ensure the security of the phone.
- Going to work case: quite often people go to work by different transportation; some people use a personal car, public transportation, bicycle or motorbike. In case a person is sitting in the car and the phone is standing still, the phone will also recognise that a "standing **still**" activity is continuing, and the phone should not be used at all for authentication. For this scenario, a backup solution should be applied, such as using the PIN-code.
- Fitness case: people might lose their phone while they are doing various exercises such as running, playing football, walking outside their home or

going to the gym. “Fitness exercises” are activities and can also be used as a security mechanism towards authentication of the phone for usage.

- Going to the pub case: the person is in a pub. The classifier is still walking, but the additional knowledge (e.g., GPS information) is expected to be slightly intoxicated; therefore, they might be variability in his signal.

These circumstances are a small sample to illustrate which activities can be recognised from gait signal data. Accordingly, it is possible to develop different classifier ideas to have several classifiers per activity. In other words, each person probably could have many repeated journeys. Each journey could be improved by different classifier because that journey was repeated many times. The GPS information and time of the day will help to know what classifier to apply. Furthermore, the pattern of life could mean the classifier ultimately becomes more refined over time.

The interesting and attractive point of these cases is that using the smartphone for “activity recognition” for identifying activities and gait recognition for identifying the individuality of a person, which can establish an access control as a security mechanism for mobility devices. To the best of our knowledge, there is no research using the gait data signal and the context information in one full system.

Above all, to acquire a better understanding of the availability of using context data in order to apply better classifiers, the commonality of existing participants and GPS patterns during the weekdays were examined. The time windows for the same time during the working days were checked (e.g., the pattern for window 8:30-9:30 on Monday the same pattern for window for Tuesday 8:30-9:30). Most of them had a particular common pattern throughout the weekdays (e.g., walking to work, walking home from work, walking to the existing building, going for lunch, shopping, etc.). The time might vary sometimes, but there was a typical pattern. Furthermore, there were some common parts, and there were some errors (differences) between patterns. For example, the morning samples were regularly repeated with mostly walking from home to work as shown in Figure 7-3 and Figure 7-4, two days' GPS tracking data, and Google Maps direction for User1 and User2, respectively.

Figure 7-3: Two days' GPS tracking data and Google Maps direction for User1

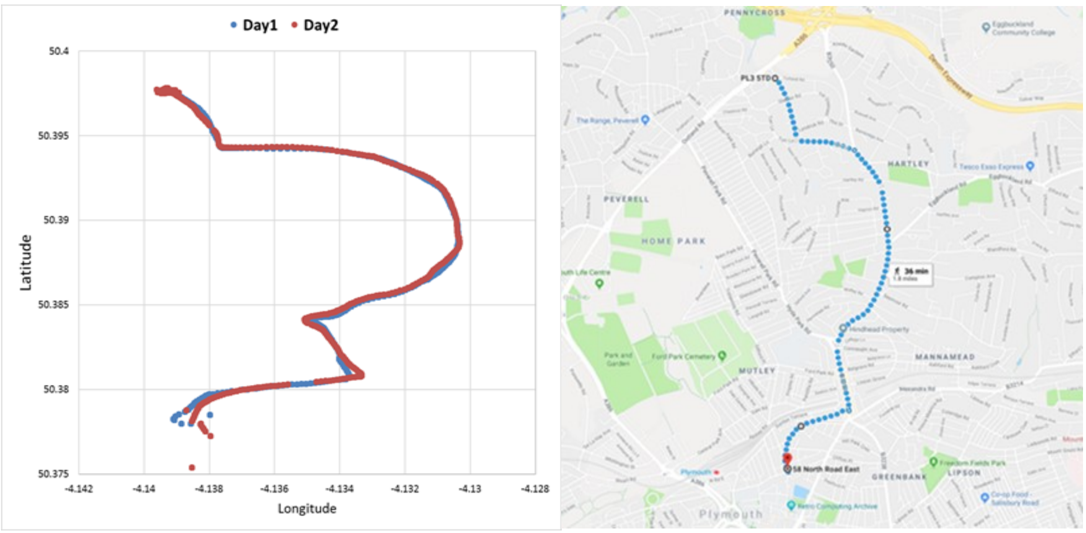


Figure 7-4: Two days' GPS tracking data and Google Maps direction for User 2

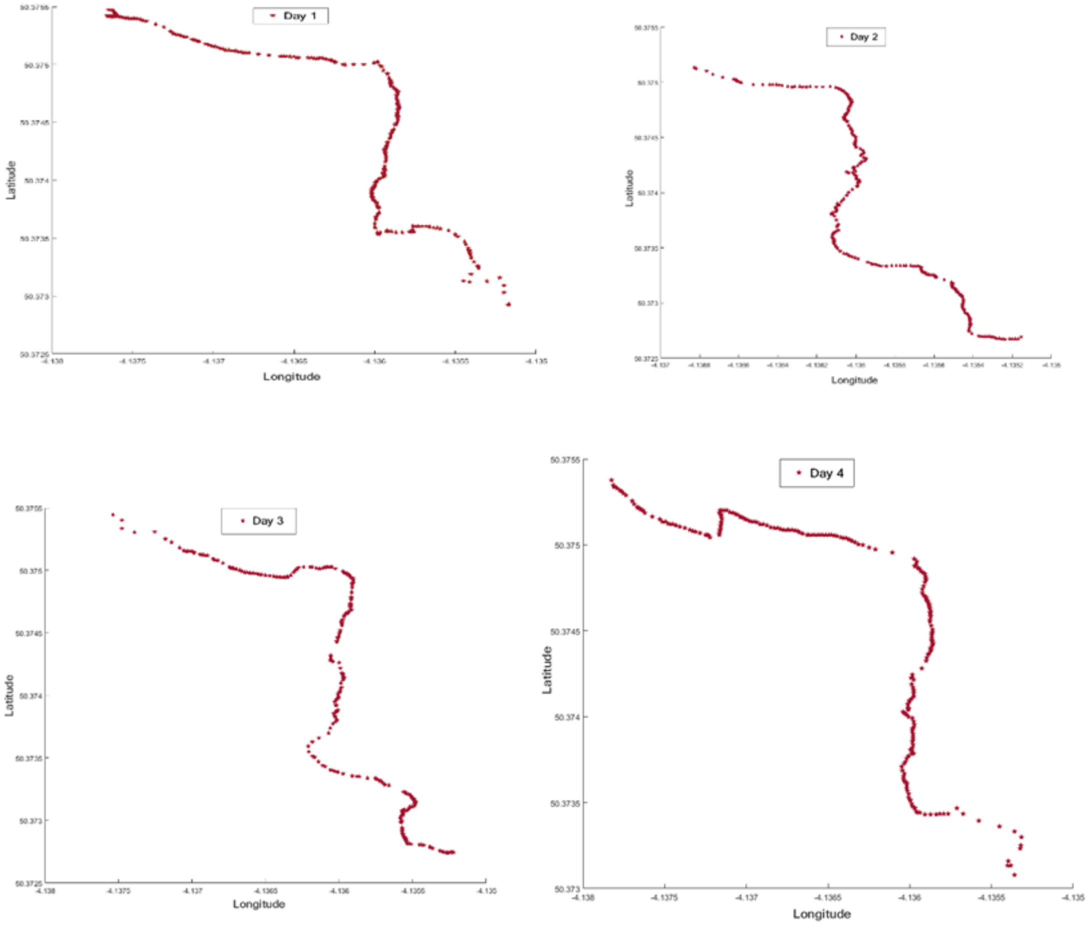


Figure 7-5: Four days' GPS tracking data for User 1

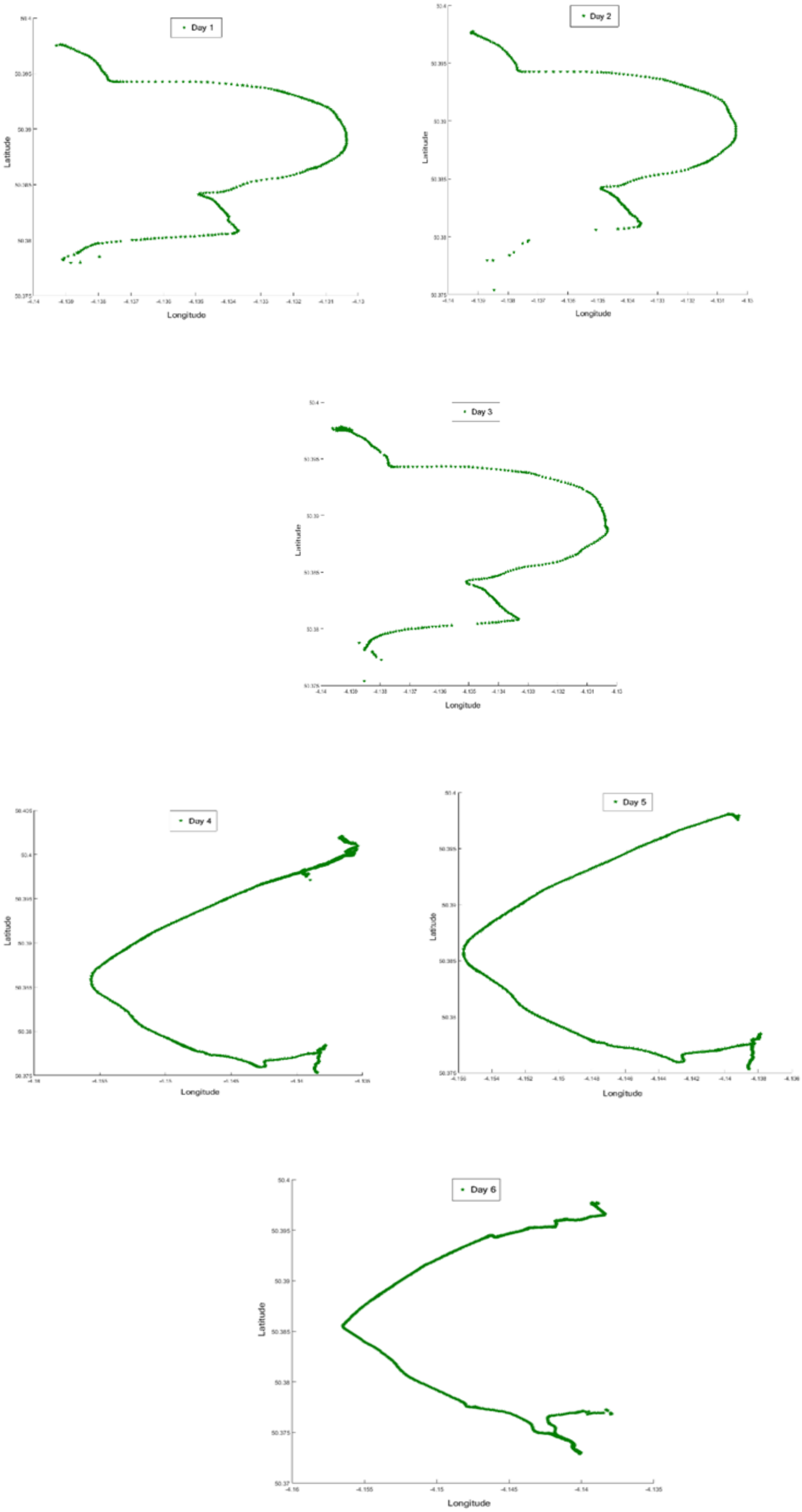


Figure 7-6: Six days' GPS tracking data for User2

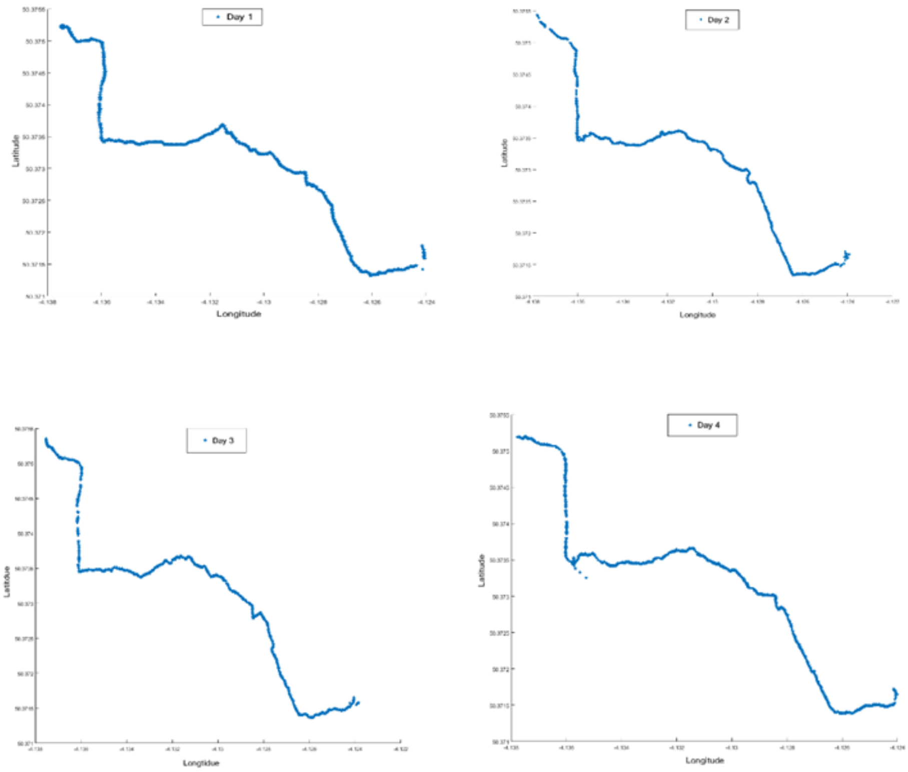


Figure 7-7: Four days' GPS tracking data for User3

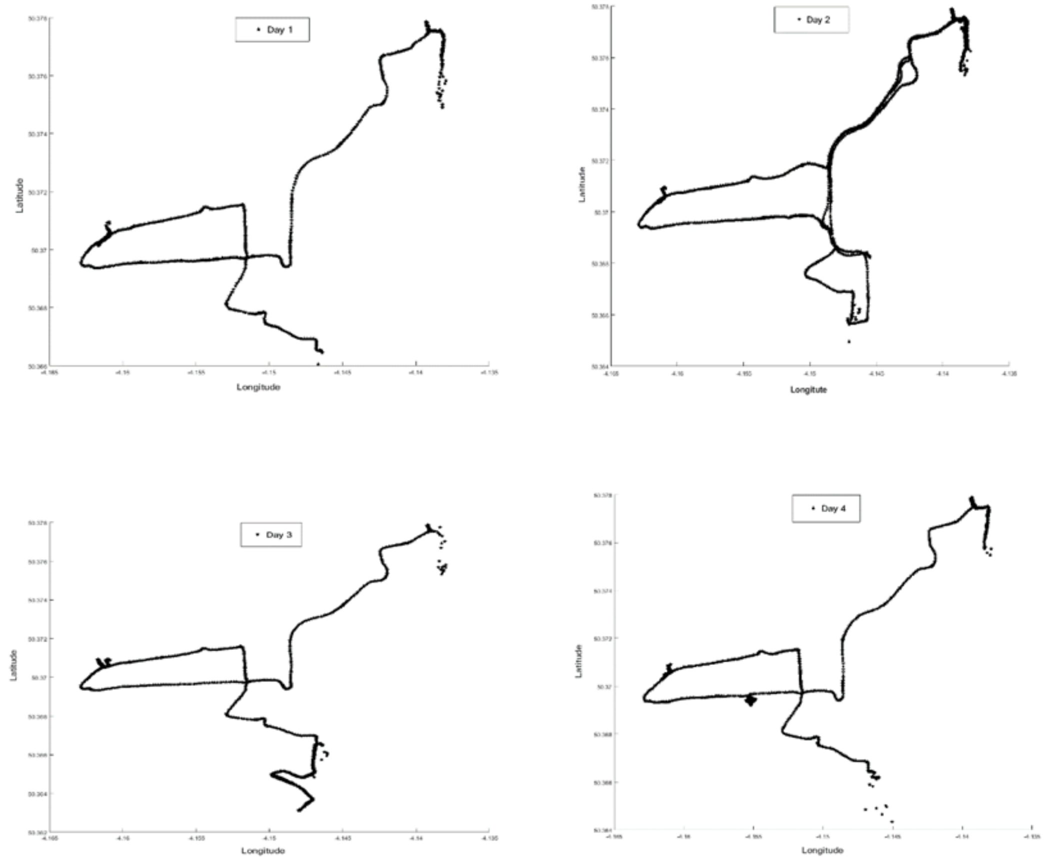


Figure 7-8: Four days' GPS tracking data for User4

As shown above, are examples of participants' GPS patterns during the working days. Users 1, 3, and 4 have the same patterns during the weekdays while user 2 has two patterns throughout the weekdays, as depicted in Figure 7-5. It appears from Figure 7-6 the user root is repeated through the first part of the week (i.e., Monday, Tuesday, and Wednesday) and the rest of the weekdays have a different root. As there are common GPS patterns for most of the users, the GPS picks up patterns and when there are sufficient numbers of these patterns quite for training (e.g., thirty samples) after thirty days walking to work will have additional classifier, rather than just have normal, fast, carrying a bag and stairs walking classifiers. It can be designed for different walking activity classifiers, for instance, the walking to work classifier and the walking home classifier. To put it another way, most people have some actions repeated periodically; consequently, the classifier could be broken still further. According to the GPS pattern, the context-awareness will choose the most proper classifier and decide which algorithm to use. In the decision box, the weather could be an information source that is used to vary the decision, and the context-awareness helps the decision, in raining weather, the person is walking to work in the rain likely bit faster than usual.

As shown in Figure 7-2 is how context awareness might work in practice. The first part of the analysis breaks down the classifier and the second step of context-awareness can help to make more informed decisions in two different stages: the algorithmic approach and decision box.

7.5 Conclusion

With the aims to contribute to the field of smartphone authentication systems without complex algorithms or adding additional cost, a novel multi-algorithmic approach gait recognition system that identifies and recognises the subject utilising a real-life mobile-based signal was introduced. The comparison between this study and previous studies' performance revealed that this research achieved better results than the related works.

Context data with gait recognition may show an improvement over gait-only biometric recognition, where user contextual and behavioural patterns are modelled based on the daily user routine. The data do suggest there is enough pass of life to explore that and context could also be used with a variety of other information resources. Therefore, a context awareness system was proposed. This could offer the ability to get a more reliable authentication decision acquired from these two techniques within the transparent authentication system (TAS).

8 Conclusions and future work

This chapter concludes the key contributions and achievements of the research. This is followed by highlights the research limitations and potential areas for further studies within the continuous authentication field utilising smartphone gait recognition.

8.1 Contributions and Achievements of the Research

The research has fulfilled all the aims mainly set out in chapter 1, with a sequence of experimental studies leading to the enhancement of the transparent mobile user authentication using gait recognition employing a real-life dataset.

The key contributions and achievements of this research are:

- Provided a comprehensive analysis of the prior studies related to transparent and continuous authentication utilising gait recognition where gait data is recorded using smartphone devices sensors. It also identifies the gap that exists in the literature and the need for more transparent and realistic user identification and verification mechanisms and should hopefully suitable for users.
- A mobile software application was installed to extract a real gait activity signal and contextual data. The controlled and real-life collected datasets considered the enormous volume of real and live unconstrained use of the smartphone devices aiming at utilising them in the research experiments. Sixty subjects and forty-four subjects for a controlled and real-world dataset, accordingly, were employed and their walking activities data collected in a period around 7-11days.

- A series of experiments were conducted to comprehend the effectiveness, viability, suitably and security of the smartphone-based user authentication utilising gait signal to determine to what degree the collected gait signal could be contributing to the system performance. Largest feature vector investigated and evaluated by applying dynamic features selection and using two classification algorithms (FF-MLP, SVM). Many factors were tested including; the impact of accelerometer and gyroscope sensors data (i.e. time and frequency domains), different feature subsets were selected and the neural network sizes of a classifier on the system accuracy.
- One of the more significant findings to emerge from this study is that a dynamic feature methodology rather than a static (all feature) approach achieved better performance and subsequently reducing the computational load upon the classifier especially with controlled experiments.
- The second significant finding was providing a novel comprehensive assessment of the multi-algorithm approach classification design (where different classifiers are used based upon the nature of the activity) support the use of smartphone gait signals. These experiments confirmed that such an approach could achieve a better level of performance over a single classification approach.
- The proposed system employed multi activities extracted from real-life gait-based signals to more thoroughly evaluate the recognition performance under non-lab-based conditions and to add further comprehensiveness feasibility and acceptability of such a proposal system. Moreover, the non-intrusive data collection supported the user-friendliness and transparency of the system.

- Developing an activity identification model which identifying the gait and non-gait activities samples using multi-activity classification algorithm. This resulted in 576,439 samples classified as a non-gait activity sample.

A number of papers related to the research published, and this provided in Appendix A. Overall, the contribution of this study has been to confirmed positive contributions to transparent user authentication for smartphone devices in the application of gait recognition.

8.2 Limitations of research

While the aims of this research have been achieved, some restrictions associated with the research have arisen, which had some had some effect on the work and findings. The fundamental limitations of the study included:

- The collected dataset was acquired using a single type of mobile device (Samsung Galaxy S6). Investigating other widely used devices willing to contribute to the data collection experiments probably conceive of with a larger and better dataset. This can be analysed to show the effect of different devices gait signal.
- Whilst the use of a multi-algorithmic classification scheme would provide better recognition performance, and the problem has now transitioned into how the system will know which classifier to utilise. Therefore, further research will focus on how to determine the nature of the activity the user is undertaking through devising context-awareness.

- The evaluation of this study was conducted offline using a desktop computer. It has not been thoroughly tested in a live environment (smartphone) to measure other operational metrics, such as computational overheads, memory consumption and the time required for the whole pipeline to be completed, starting from acquiring motion signals, to feature extraction, segmentation, pre-processing, and finally inferencing, where the examined data are classified.

8.3 Suggestions, Scope for future work

Although the developed approach reached a high level of accuracy in gait-based activity identification and user authentication based on raw smartphone motion sensor signals, other aspects could be examined and investigated in future research to generate more findings, including the following:

In practice, a two-stage model can be developed, one for detecting state type (non-gait/ gait) activity, followed by an activity identification model to identify the activity type. Once the activity type is identified, the authentication model is legitimate the subject.

As the evaluation of this study was conducted offline using a desktop computer. Further studies need to be carried out in order to validate it in a live environment (smartphone) to fully understand the efficiency of all operational metrics. It is envisaged, much like popular mobile apps, the use of cloud resources will provide a mechanism for off-loading any computationally challenging aspects to relieve local demands upon computation and memory.

Investigating other widely used devices, such as an Apple iPhone, could reveal how similar/different the generated motion signals might be for different devices and to what extent feature space distribution varies.

Future work could also investigate other factors, such as testing various segment numbers of seconds and samples required per individual in order to train a user-dependent predictable model successfully so that it can accurately match a given signal with the similar physical activity.

Further research could investigate context awareness information to enable an intelligent decision process. Through introducing, additional information that can be collected via smartphone itself, leveraging various sources, including, Wi-Fi information, and installed mobile applications, motion sensors, calendar, email, natural languages processing of text messages and weather forecast. In which, this enables biometric systems to make a more reliable decision to leverage a wider range of information. For example, if a user's gait appears to be faster than normal, an analysis of the calendar might reveal they are running late for a meeting– and therefore the system could adapted either the classifier (using fast algorithm) or threshold accordingly because a high degree of availability expected. Likewise, realising a user is heading towards the airport might provide additional information required to understand they are likely to be carrying or pulling a bag and again, the system can adapt appropriately to compensate. Therefore, this could offer a more reliable and robust gait-based Transparent Authentication System (TAS).

8.4 The Future of Authentication

During the last decade, smartphones have become a ubiquitous technology providing a wide range of services and features (e.g. personal communications, entertainment, and business) that are used to access/store sensitive and confidential information. This trend is only set to continue as technology becomes increasingly pervasive and the desire to access information and consume services becomes the norm. Authentication of the user will remain an essential technology to determine who the user is and subsequently what they can and cannot access.

What is clear from current literature is that the authentication burden placed upon the user has increased substantially with an impact on the user experience. It is essential that technologies are continued to be developed that provide a frictionless authentication experience. Gait recognition, as presented in this thesis, provides one such approach that can be used in specific scenarios to aid the authentication decision process, but true frictionless authentication can only be achieved through careful and usable design and multi-modal/factor techniques that are able to adapt to the varying situations, environments, people and technologies.

References

Al Abdulwahid, A., 2015. Federated Authentication using the Cloud,

Al-obaidi, H. et al., 2018. A Multi-Algorithmic Approach for Gait Recognition. In In ECCWS 2018 17th European Conference on Cyber Warfare and Security (p. 20). Academic Conferences and publishing limited.

Altman, N. S. (1992) 'An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression', The American Statistician, 46(3), pp. 175-185. doi: 10.1080/00031305.1992.10475879.

Analytics Vidhya (2015) Analytics Vidhya. Available at: https://www.analyticsvidhya.com/blog/2015/07/difference-machine-learning-statistical-modeling/ (Accessed: 20 July 2019).

Anantasech, P. and Ratanamahatana, C. A. (2019) 'Enhanced Weighted Dynamic Time Warping for Time Series Classification', in, pp. 655-664. doi: 10.1007/978-981-13-1165-9_60.

AndroSensor, AndroSensor. Available at: https://play.google.com/store/apps/details?id=com.fivasim.androsensor&hl=en_GB [Accessed August 14, 2018].

Anguita, D. et al., 2012. Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7657 LNCS, pp.216-223.

Anguita, D. et al., 2013. A Public Domain Dataset for Human Activity Recognition Using Smartphones. European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, (April), pp.24-26. Available at: http://www.i6doc.com/en/livre/?GCOI=28001100131010.

Bajrami, G., 2011. Activity Identification for Gait Recognition Using Mobile Devices. Available at: http://brage.bibsys.no/hig/handle/URN:NBN:no-bibsys_brage_21093.

Barber, D. (2012) 'Bayesian Reasoning and Machine Learning', Bayesian Reasoning and Machine Learning. doi: 10.1017/cbo9780511804779.

Bashir, K., Xiang, T. & Gong, S., 2010. Gait recognition without subject cooperation. Pattern Recognition Letters, 31(13), pp.2052–2060. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0167865510001844 [Accessed December 12, 2014].

BBC NEWS, 2012. BBC NEWS. Available at: http://news.bbc.co.uk/1/hi/technology/8509299.stm [Accessed March 1, 2015].

Bhanu Jyothi, K. & Hima Bindu, K., 2018. A Case Study in R to Recognize Human Activity Using Smartphones. , pp.191–200. Available at: http://link.springer.com/10.1007/978-981-10-6319-0_17.

Biometric Institute, 2013. Biometric Institute. Available at: http://www.biometricsinstitute.org/ [Accessed June 20, 2005].

BiometricSolutions, 2015. Biometric Solutions. Available at: http://www.biometric-solutions.com/solutions/index.php?story=fingerprint_recognition [Accessed May 16, 2016].

Boyle, M. et al., 2011. Poster: gait-based smartphone user identification. Proceddings of the 9th international conference on Mobile systems, applications, and services (MOBISYD '11), pp.395–396. Available at: http://ipsit.bu.edu/documents/mobisys11.pdf.

Breitinger, F. & Nickel, C., 2010. User Survey on Phone Security and Usage.

Biosig, (May 2010). Available at: https://www.fbi.h-da.de/fileadmin/gruppen/FG-IT-Sicherheit/Publikationen/2010/2010_09_Breitinger_Biosig2010.pdf.

carphone warehous, The Lodown. Available at: http://lowdown.carphonewarehouse.com/feature/meet-the-samsung-galaxy-s6-and-its-sensors/30129/ [Accessed April 10, 2018].

Caruana, R., Munson, A. & Niculescu-Mizil, A., 2006. Getting the Most Out of Ensemble Selection. In Sixth International Conference on Data Mining (ICDM'06). IEEE, pp. 828–833.

Clarke N and Furnell S, 2006. A composite user authentication architecture for mobile de- vices.

Clarke, N., 2011. Transparent user authentication: biometrics, RFID and behavioural profiling, London: Springer. Available at: http://books.google.com/books?hl=en&lr=&id=41h4Rk6HANkC&oi=fnd&pg=PR5&dq=Transparent+User+Authentication+Biometrics,+RFID+and+Behavioural+Profilin&ots=0X0uJj8ClA&sig=qxnEs1nsB0OTUEszlQJ4pLF-4Eg [Accessed December 16, 2014].

Clarke, N.L. & Furnell, S.M., 2005. Authentication of users on mobile telephones - A survey of attitudes and practices. Computers and Security, 24(7), pp.519-527.

Cobb, 2012. ESET & Harris Interactive Password Poll. ESET, We Live Security. Available at: http://www.welivesecurity.com/2012/12/04/password-handling-challenges-costs-current-behavior-infographic/ [Accessed February 3, 2015].

Crendant, 2009. "Phone Data makes 4.2 Million* Brits Vulnerable to ID Theft." Available at: http://www.globalsecuritymag.fr/Phone-Data-makes-4-2-Million-Brits,20090318,8005.html [Accessed February 3, 2015].

Creswell, J.W., 2013. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, Thousand Oaks: SAGE Publications.

Crouse, D., Chandra, D. & Barbello, B., 2013. Continuous Authentication of Mobile User : Fusion of Face Image and Inertial Measurement Unit Data.

CSID, 2017. CSID. Available at: file:///C:/Users/hkal-obaidi/Downloads/201617csewtechnicalreportvolumeone.pdf [Accessed February 1, 2015].

da/sec., 2015. da/sec scientific talk on Biometrics. Available at: https://www.dasec.h-da.de/teaching/dasec-scientific-talk/2015-05-07-on-biometrics/ [Accessed May 1, 2016].

Data Science Bootcamp (2019) Data Science Bootcamp. Available at: https://medium.com/data-science-bootcamp/multilayer-perceptron-mlp-vs-convolutional-neural-network-in-deep-learning-c890f487a8f1 (Accessed: 15 July 2019).

Daugman, J., 2003. The importance of being random: Statistical principles of iris recognition. Pattern Recognition, 36(2), pp.279–291.

Derawi, 2012. Smartphones and Biometrics: Gait and Activity Recognition. Gjøvik University College.

Derawi, M. & Bours, P., 2013. Gait and activity recognition using commercial phones. Computers & Security, 39(July 2015), pp.137–144. Available at: http://www.sciencedirect.com/science/article/pii/S0167404813000953.

Derawi, M.O. et al., 2010. Unobtrusive User-Authentication on Mobile Phones Using Biometric Gait Recognition. In 2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing. Ieee, pp. 306–311. Available at:

http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5638036 [Accessed December 3, 2014].

Derawi, M.O., 2012. Smartphones and Biometrics,

Derawi, M.O., Bours, P. & Holien, K., 2010. Improved Cycle Detection for Accelerometer Based Gait Authentication. 2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp.312–317. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5635767 [Accessed December 12, 2014].

E.Cutting, J. & T.Kozlowski, L., 1977. Recognizing friends by their walk: Gait perception without familiarity cues. Bulletin of the Psychonomic Society.

Ferrero, R. et al., 2015. On gait recognition with smartphone accelerometer. In 2015 4th Mediterranean Conference on Embedded Computing (MECO). IEEE, pp. 368–373. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7181946.

Frank, J., Mannor, S. & Precup, D., 2010a. A novel similarity measure for time series data with applications to gait and activity recognition. UbiComp'10 - Proceedings of the 2010 ACM Conference on Ubiquitous Computing, pp.407–408. Available at: http://dx.doi.org/10.1145/1864431.1864460.

Frank, J., Mannor, S. & Precup, D., 2010b. Activity and Gait Recognition with Time-Delay Embeddings Time-Delay Embeddings. In AAAI Conference on Artificial Intelligence. pp. 407–408. Available at: http://www.aaai.org/ocs/index.php/AAAI/AAAI10/paper/viewPDFInterstitial/1666/2260%5Cnhttp://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.224.7630.

Franzese, M. and Iuliano, A. (2019) 'Learn more about Hidden Markov Model

Hidden Markov Models Computational Methods in Cell Biolo- gy'.

Gadaleta, M. & Rossi, M., 2018. IDNet: Smartphone-based gait recognition with convolutional neural networks. Pattern Recognition, 74, pp.25–37. Available at: http://dx.doi.org/10.1016/j.patcog.2017.09.005.

Gafurov, D., Snekkenes, E. & Bours, P., 2007a. Gait Authentication and Identification Using Wearable Accelerometer Sensor. 2007 IEEE Workshop on Automatic Identification Advanced Technologies, pp.220–225. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4263244.

Gafurov, D., Snekkenes, E. & Bours, P., 2007b. Gait Authentication and Identification Using Wearable Accelerometer Sensor. 2007 IEEE Workshop on Automatic Identification Advanced Technologies, (September 2015), pp.220–225. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4263244.

Ganti, R.K., Srinivasan, S. & Gacic, A., 2010. Multisensor Fusion in Smartphones for Lifestyle Monitoring. In 2010 International Conference on Body Sensor Networks. IEEE, pp. 36–43.

Ghazizadeh, E. et al., 2012. A survey on security issues of federated identity in the cloud computing. CloudCom 2012 - Proceedings: 2012 4th IEEE International Conference on Cloud Computing Technology and Science, pp.562–565.

GreekoSystem, 2011. "Facebook Will Go Ahead and Scan Your Face Now." Available at: http://www.geekosystem.com/facebook-face-recognition [Accessed June 15, 2016].

GSMA, 2018. Understanding 7 Billion: Counting Connections and People. Available at: http://www.gsma.com/newsroom/blog/understanding-7-billion-counting-connections-and-people/ [Accessed December 19, 2018].

Gupta, P. & Dallas, T., 2014. Feature selection and activity recognition system using a single triaxial accelerometer. IEEE Transactions on Biomedical Engineering, 61(6), pp.1780–1786.

Han, S. et al. (2016) 'EIE: Efficient Inference Engine on Compressed Deep Neural Network', in 2016 ACM/IEEE 43rd Annual International Symposium on Computer Architecture (ISCA). IEEE, pp. 243–254. doi: 10.1109/ISCA.2016.30.

Heng, X., Wang, Z. & Wang, J., 2014. Human activity recognition based on transformed accelerometer data from a mobile phone. International Journal of Communication Systems, 23(5), pp.633–652.

Heng, X., Wang, Z. & Wang, J., 2016. Human activity recognition based on transformed accelerometer data from a mobile phone. International Journal of Communication Systems, 29(13), pp.1981–1991.

Henrique, B. M., Sobreiro, V. A. and Kimura, H. (2019) 'Literature review: Machine learning techniques applied to financial market prediction', Expert Systems with Applications. Elsevier Ltd, 124, pp. 226–251. doi: 10.1016/j.eswa.2019.01.012.

Hestbek, M., Nickel, C. & Busch, C., 2012. Biometric Gait Recognition for Mobile Devices Using Wavelet Transform and Support Vector Machines. Systems, Signals and Image …, (April), pp.11–13. Available at: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6208109.

Ho, C.C. et al., 2012. An unobtrusive Android person verification using accelerometer based gait. Proceedings of the 10th International Conference on Advances in Mobile Computing & Multimedia - MoMM '12, p.271. Available at: http://dl.acm.org/citation.cfm?doid=2428955.2429007.

Hoang, T. et al., 2013. Adaptive cross-device gait recognition using a mobile accelerometer. Journal of Information Processing Systems, 9(2), pp.333–348.

Hoang, T., Choi, D. & Nguyen, T., 2015. Gait authentication on mobile phone using biometric cryptosystem and fuzzy commitment scheme. International Journal of Information Security. Available at: http://link.springer.com/10.1007/s10207-015-0273-1.

Hoang, T.M. et al., 2012. Gait identification using accelerometer on mobile phone. 2012 International Conference on Control, Automation and Information Sciences, ICCAIS 2012, pp.344–348.

Hocking, C.G., 2014. Authentication Aura Authentication Aura A cooperative and distributed approach to user authentication on mobile devices. , (December).

Holien, K., 2008. Gait recognition under non-standard circumstances. Master's thesis, Gjøvik University College-Department of Computer Science and Media Technology. Available at: http://brage.bibsys.no/hig/bitstream/URN:NBN:no-bibsys_brage_7523/1/Holien.pdf.

Huntington, 2012. Huntington. Available at: http://www.authenticationworld.com/Papers/Creating a User Friendly Authentication and Authorization Architecture.pdf.

Ichino, M. et al., 2013. A S tudy on G ait R ecognition using LPC C epstrum for M obile T erminal.

Internet live stats, 2016. Internet Users. Available at: http://www.internetlivestats.com/internet-users/ [Accessed July 28, 2016].

ISO / IEC, 2012. INTERNATIONAL STANDARD ISO / IEC. , 2012.

Jain, A., Nandakumar, K. & Ross, A., 2005. Score normalization in multimodal biometric systems. Pattern Recognition, 38(12), pp.2270-2285.

Jang, J.H. & Yang, S.H., 2015. Development of the rule-based inference engine for the advanced context-awareness. International Journal of Smart Home, 9(4), pp.195-202. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-84930169216&partnerID=40&md5=8958a2208e7aa5941cacc6a0b62b6e94.

Jiang, W. & Yin, Z., 2015. Human Activity Recognition Using Wearable Sensors by Deep Convolutional Neural Networks. In Proceedings of the 23rd ACM international conference on Multimedia - MM '15. New York, New York, USA: ACM Press, pp. 1307-1310.

John D. Woodward, J. et al., 2003. Biometrics: Identity Assurance in the Information Age. , 13, pp.1-9. Available at: http://web.law.suffolk.edu/highlights/stuorgs/jhtl/book_reviews/2003_2004/burnham.pdf.

Jordan, M. I. and Mitchell, T. M. (2015) 'Machine learning: Trends, perspectives, and prospects', 349(6245).

Karabatis, G., 2017. Information Fusion for Cyber-Security Analytics I. M. Alsmadi, G. Karabatis, & A. Aleroud, eds., Cham: Springer International Publishing. Available at: http://link.springer.com/10.1007/978-3-319-44257-0.

Karatzouni, S., 2014. Non-Intrusive Continuous User Authentication for Mobile Devices.

Khandnor, P. & Kumar, N., 2017. A Survey of Activity Recognition Process Using. , pp.607-612.

Kim, K. (2016) 'A hybrid classification algorithm by subspace partitioning through semi-supervised decision tree', Pattern Recognition, 60, pp. 157-163. doi: 10.1016/j.patcog.2016.04.016.

Kwapisz, J., Weiss, G. & Moore, S., 2011. Activity recognition using cell phone accelerometers. ACM SIGKDD Explorations …, 12(2), pp.74-82. Available at: http://dl.acm.org/citation.cfm?id=1964918.

Kwapisz, J.R., Weiss, G.M. & Moore, S. a, 2010. Cell Phone-Based Biometric Identification. Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on, pp.1-7.

Lau, H. & 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.

Lu, H., 2014. Unobtrusive Gait Verification for Mobile Phones. In ISWC'14. pp. 91-98.

Mail Online News, 2014. mail online. Available at: http://www.dailymail.co.uk/news/article-2747294/Call-alarm-2-000-mobiles-stolen-day-iPhone-target.html [Accessed July 29, 2016].

Marsico, M. De & Mecca, A., 2015. New Trends in Image Analysis and Processing -- ICIAP 2015 Workshops. , 9281, pp.19-26. Available at: http://link.springer.com/10.1007/978-3-319-23222-5.

McCabe, D.L., 2005. Cheating among college and university students: A North American perspective. International Journal for Educational Integrity, 1(1), pp.10-11. Available at: http://ojs.ml.unisa.edu.au/index.php/IJEI/article/view/14.

Meeker Mary, 2013. 2013 Internet Trends. Available at:

http://www.kpcb.com/blog/2013-internet-trends [Accessed July 30, 2016].

Meng, W. et al., 2015. Surveying the Development of Biometric User Authentication on Mobile Phones. IEEE Communications Surveys & Tutorials, PP(99), pp.1-1. Available at: http://ieeexplore.ieee.org/ielx7/9739/5451756/07000543.pdf?tp=&arnumber=7000543&isnumber=5451756%5Cnhttp://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7000543&queryText=authentication+survey&refinements[]=&ranges[]=2012_2015_Year&matchBoolean%5Cnhttp:

Miguel, G. & Neves, G., 2013. Android Gait Recognition System Electrical and Computer Engineering Examination Committee September 2013. , (September).

Monitgomery, T.M., 2014. Anatomy, Physiology and Pathology of the Human Eye.

Muaaz, M. & Mayrhofer, R., 2013. An Analysis of Different Approaches to Gait Recognition Using Cell Phone Based Accelerometers. Proceedings of International Conference on Advances in Mobile Computing & Multimedia - MoMM '13, (August), pp.293–300. Available at: http://dl.acm.org/citation.cfm?doid=2536853.2536895.

Muaaz, M. & Nickel, C., 2012. Influence of different walking speeds and surfaces on accelerometer-based biometric gait recognition. 2012 35th International Conference on Telecommunications and Signal Processing (TSP), pp.508-512. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6256346.

Muaaz, M., 2013. A Transparent and Continuous Biometric Authentication Framework for User-Friendly Secure Mobile Environments. , pp.4–7.

Muaaz, M., 2017. Implicit Biometric Authentication for Smartphones.

Nakano, K., 2017. Effect of Dynamic Feature for Human Activity Recognition using Smartphone Sensors. , (iCAST), pp.539–543.

NEURO technology, 2015. NEURO technology.

Ngo, T.T. et al., 2014. The largest inertial sensor-based gait database and performance evaluation of gait-based personal authentication. Pattern Recognition, 47(1), pp.228–237. Available at: http://linkinghub.elsevier.com/retrieve/pii/S003132031300280X [Accessed November 27, 2014].

Nickel, C. & Busch, C., 2013. Classifying accelerometer data via hidden Markov models to authenticate people by the way they walk. IEEE Aerospace and Electronic Systems Magazine, 28(10), pp.29–35. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6642829.

Nickel, C. et al., 2011. Scenario test of accelerometer-based biometric gait recognition. Proceedings of the 3rd International Workshop on Security and Communication Networks, IWSCN 2011, (July 2015), pp.15–21.

Nickel, C. et al., 2011. Using Hidden Markov Models for accelerometer-based biometric gait recognition. In 2011 IEEE 7th International Colloquium on Signal Processing and its Applications. IEEE, pp. 58–63. Available at: http://www.fbi.h-da.de/fileadmin/gruppen/FG-IT-Sicherheit/Publikationen/2011/2011_03_Nickel_CSPA2011.pdf.

Nickel, C., 2012. Accelerometer-based Biometric Gait Recognition for Authentication on Smartphones. TU Darmstadt.

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

Nickel, C., Brandt, H. & Busch, C., 2011b. Classification of Acceleration Data for Biometric Gait Recognition on Mobile Devices. In Proceedings of the Special Interest Group on Biometrics and Electronic Signatures. pp. 57-66. Available at: http://www.researchgate.net/publication/221234829_Classification_of_Acceleration_Data_for_Biometric_Gait_Recognition_on_Mobile_Devices.

Nickel, C., Brandt, H. & Busch, C., 2011c. GI-Edition Classification of Acceleration Data forBiometric Gait Recognition on Mobile Devices. In Classification of Acceleration Data forBiometric Gait Recognition on Mobile Devices.

Nickel, C., Wirtl, T. & 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. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6274118.

Nishiguchi, S. et al., 2012. Reliability and Validity of Gait Analysis by Android-Based Smartphone. Telemedicine and e-Health, 18(4), pp.292-296.

NIST, 2014. Security and Privacy Controls for Federal Information Systems and Organizations Security and Privacy Controls for Federal Information Systems and Organizations. Sp-800-53Ar4, p.400+.

Ogbuabor, G. & La, R., 2018a. Human Activity Recognition for Healthcare using Smartphones. In Proceedings of the 2018 10th International Conference on Machine Learning and Computing - ICMLC 2018. New York, New York, USA: ACM Press, pp. 41-46.

Ogbuabor, G. & La, R., 2018b. Human Activity Recognition for Healthcare using Smartphones. , pp.20-21.

Ottomoeller, C.R., 2014. Gait Recognition in Mobile Security.

Pao, Y.-H. (2005) Statistical Pattern Recognition. Second edition (Keinosuke Fukunaga), SIAM Review. doi: 10.1137/1035031.

Phan, D. & Dam, N.N., 2015. Distributed, Ambient, and Pervasive Interactions. , 9189, pp.188-200. Available at: http://link.springer.com/10.1007/978-3-319-20804-6.

Poppe, R., 2007. Vision-based human motion analysis: An overview. Computer Vision and Image Understanding, 108(1-2), pp.4-18.

Qiu, J. et al. (2016) 'A survey of machine learning for big data processing', EURASIP Journal on Advances in Signal Processing, 2016(1), p. 67. doi: 10.1186/s13634-016-0355-x.

Ramchoun, H. et al. (2016) 'Multilayer Perceptron: Architecture Optimization and Training', International Journal of Interactive Multimedia and Artificial Intelligence, 4(1), p. 26. doi: 10.9781/ijimai.2016.415.

Rana, J.R., 2015. Obtrusive and Unobtrusive Authentication of Portable Devices for Security. , 2(4), pp.766-771.

Ronao, C.A. & Cho, S.-B., 2016. Human activity recognition with smartphone sensors using deep learning neural networks. Expert Systems with Applications, 59, pp.235-244. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0957417416302056.

Saevanee, H. & Bhatarakosol, P., 2008. User authentication using combination of behavioral biometrics over the touchpad acting like touch screen of mobile device. Proceedings of the 2008 International Conference on Computer and Electrical Engineering, ICCEE 2008, pp.82-86.

Saevanee, H. et al., 2015. Continuous user authentication using multi-modal biometrics. Computers & Security, 53(JUNE), pp.234-246. Available at:

http://linkinghub.elsevier.com/retrieve/pii/S0167404815000875.

Saevanee, H., 2014. Continuous User Authentication using Multi-Modal Biometrics. Continuous User Authentication using Multi-Modal Biometrics, (May).

Saha, J. et al., 2018. An ensemble of condition based classifiers for device independent detailed human activity recognition using smartphones. Information (Switzerland), 9(4).

Schneider, O.S. et al., 2013. Real-time gait classification for persuasive smartphone apps: structuring the literature and pushing the limits. Proceedings of the 2013 international conference on Intelligent user interfaces, pp.161-172. Available at: http://doi.acm.org/10.1145/2449396.2449418.

Signing Hub, 2016. Signing Hub. Available at: http://blog.signinghub.com/ [Accessed August 1, 2016].

SKY BIOMETRY, 2018. SKY BIOMETRY. Available at: https://skybiometry.com/facial-recognition-mobile-phones/.

Sprager, S. & Juric, M.B., 2015. Inertial Sensor-Based Gait Recognition: A Review,

Sprager, S. & Zazula, D., 2011. Impact of different walking surfaces on gait identification based on higher-order statistics of accelerometer data. 2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), pp.360-365. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6144100.

Sprager, S., 2009. A cumulant-based method for gait identification using accelerometer data with principal component analysis and support vector machine. WSEAS Transactions on Signal Processing, 5(11), pp.369-378. Available at: http://dl.acm.org/citation.cfm?id=1853846.

Statistical Models (2019) Statistical Models. Available at: https://www.ibm.com/support/knowledgecenter/en/SS3RA7_17.0.0/clementine/nodes_statisticalmodels.html (Accessed: 20 July 2019).

Strobl, C. et al., 2008. Conditional Variable Importance for Random Forests. BMC Bioinformatics, 9(1), p.307.

Symantec Corporation, 2013. internet security THREAT REPORT.

Tavara, S. (2019) 'Parallel Computing of Support Vector Machines', ACM Computing Surveys, 51(6), pp. 1–38. doi: 10.1145/3280989.

Technology at Nineleaps (2019) Technology at Nineleaps. Available at: https://medium.com/technology-nineleaps/popular-machine-learning-algorithms-a574e3835ebb (Accessed: 5 July 2019).

Techopedia (2019) Techopedia. Available at: https://www.techopedia.com/definition/20879/multilayer-perceptron-mlp (Accessed: 15 July 2019).

UKBA, 2011. UKBA. Available at: http://webarchive.nationalarchives.gov.uk/20140110181512/http://www.ukba.homeoffic e.gov.uk/customs-travel/Enteringtheuk/usingiris/howenterwithiris [Accessed December 8, 2018].

Verizon, 2016. Verizon's 2016 Data Breach Investigations Report finds cybercriminals are exploiting human nature. Verizon Business Journal, 2016(1), pp.1-65. Available at: https://www3.plymouth.ac.uk/CookieAuth.dll?GetLogon?curl=Z2FPGRELO

GZ2FStudentView.aspx&reason=0&formdir=5.

Watanabe, Y., 2014. Influence of Holding Smart Phone for Acceleration-Based Gait Authentication. 2014 Fifth International Conference on Emerging Security Technologies, pp.30–33. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6982770.

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, 60, pp.1856–1864. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1877050915024230.

WebMapSolutions, 2011. Mobile Location Apps Review. Available at: http://www.webmapsolutions.com/mobile-location-apps [Accessed July 28, 2016].

Weld, D. S., & Bansal, G., 2018. The Challenge of Crafting Intelligible Intelligence. Retrieved. Available at: http://arxiv.org/abs/1603.08507.

Whalen, S. & Pandey, G., 2013. A comparative analysis of ensemble classifiers: Case studies in genomics. Proceedings - IEEE International Conference on Data Mining, ICDM, pp.807–816.

Wolfe, H.B., 2013. Security and Privacy Protection in Information "A Lightweight Gait Authentication on Mobile Phone Regardless of Installation Error,"

Youn, I. et al., 2014. New Gait Metrics for Biometric Authentication Using a 3-axis Acceleration. , (Ccnc), pp.596–601.

Appendices

Appendix A- Publications

1- Al-Obaidi, H. et al., 2018. A Multi-Algorithmic Approach for Gait Recognition. In ECCWS 2018 17th European Conference on Cyber Warfare and Security (p. 20). Academic Conferences and publishing limited.

Abstract: Securing smartphones has increasingly become inevitable due to their massive popularity and significant storage and access to sensitive information. The gatekeeper of securing the device is authenticating the user. Amongst the many solutions proposed, gait recognition has been suggested to provide a reliable yet non- intrusive authentication approach – enabling both security and usability. Whilst several studies exploring mobile- based gait recognition have taken place, studies have largely been preliminary, with various methodological restrictions that have limited the number of participants, samples and type of features. Furthermore, prior studies have relied upon evaluating the approach on a limited number of activities - namely walking and running, and there is some concern over the capacity of the approach to correctly verify individuals when the nature of the signals across a wider range of activities is likely to be more variable. This paper has sought to overcome these weaknesses and provide a comprehensive evaluation, including an analysis of motion sensors (accelerometer and gyroscope), an investigation and analysis of features, understanding the variability of feature vectors during differing activities across a multi-day collection involving 60 participants. This is framed into two experiments involving five types of activities: normal, fast, with a bag, downstairs, and upstairs walking. The first experiment explores the classification performance of individual activities in order to understand whether a single classifier or multi-algorithmic approach would provide a better level of performance. The second experiment explored the features vector (comprising of a possible 304 unique features) to understand how its composition affects performance and for a comparison a more selective set of the minimal features are involved. Overall, results from the experimentation have shown an EER of 4.40-12.2% for a single classifier (using

same/cross day methodologies). The multi-algorithmic approach achieved EERs of 0.70%/6.3%, 0.42%/12.68% and 1.10%/6.46% for normal, fast and with a bag walk respectively (using the Same/ Cross Day methodology) using both accelerometer and gyroscope-based features – showing a significant improvement over the single classifier approach and thus a more effective approach to managing the problem of feature vector variability.

2- Alruban, A. et al., 2018. Human Activity Recognition for Healthcare using Smartphones. In ICPRAM 2019 8th International Conference on Pattern Recognition Applications and Methods, pp.20–21.

   Abstract: Human physical motion activity identification has many potential applications in various fields, such as medical diagnosis, military sensing, sports analysis, human-computer interaction and security. With the recent advances in smartphones and wearable technologies, it has become common for such devices to have embedded motion sensors that are able to sense even small body movements. This study collected human activity data from 60 participants across two different days for a total of six activities recorded by gyroscope and accelerometer sensors in a modern smartphone. The paper investigates to what extent different activities can be identified by utilising machine learning algorithms using approaches such as majority algorithmic voting. More analyses are also provided that reveal which time and frequency domain-based features were best able to identify individuals' motion activity types. Overall, the proposed approach achieved a classification accuracy of 98% in identifying four different activities: walking, walking upstairs, walking downstairs, and sitting (on a chair) while the subject is calm and doing a typical desk-based activity.

Appendix B- Consent Form and Information Sheet (Data Collection)

PLYMOUTH UNIVERSITY

FACULTY OF SCIENCE AND ENVIRONMENT

Human Ethics Committee Consent Form

CONSENT TO PARICIPATE IN RESEARCH PROJECT / PRACTICAL STUDY

Name of Principal Investigator

Hind Al-Obaidi

Title of Research

Mobile Authentication

Brief statement of purpose of work

The usability of a system is noticed from the first point of contact of that system more especially if the system is intrusive in perform a task. The usability of a user authentication system should address some key issues which include intrusiveness and user's ability to easily remember user login details. If these issues are met, it will greatly improve the authentication usage

This research seeks to meet these issues by using gait signals from smartphone sensors to overcome intrusiveness and avoid user's ability to know when authentication is done. To use gait signals for user authentication, it has to meet the basic requirement and characteristics needed to create a pattern for user authentication

This study will install software in the smartphone for data collection. As a participant, no modification will be made upon the device before, during and after the collection of data. Please merely put the smartphone in the belt pouch while the data will be continuously extracted during one week duration. Also, a specified exercise of not more than 15 minutes with is done at the beginning and at the end of the data collection. Based upon Plymouth University guidelines, collected data should be stored for ten years. Upon the completion of the ten-year period, the collected data will be securely destroyed.

At all stages of the study, confidentiality of the collected data and subsequent analysis will be maintained. At no time, will any identifying information about the participants be used in any publication or research output.

You have the right to withdraw at any stage upon until the completion of the data collection process. Should you wish to withdraw from the study, please contact Hind Al-Obaidi. Moreover, declining participation and/or asking to withdraw from this study will not affect your study or your relationship with your supervisors or tutors.
For information regarding the study, please contact:

Hind Al-obaidi   hind.al-obaidi@plymouth.ac.uk

For any questions concerning the ethical status of this study, please contact the secretary of the Human Ethics Committee – paula.simson@plymouth.ac.uk

---

The objectives of this research have been explained to me.

I understand that I am free to withdraw from the research at any stage, and ask for my data to be destroyed if I wish.

I understand that my anonymity is guaranteed, unless I expressly state otherwise.

I understand that the Principal Investigator of this work will have attempted, as far as possible, to avoid any risks, and that safety and health risks will have been separately assessed by appropriate authorities (e.g. under COSHH regulations)

Under these circumstances, I agree to participate in the research.

Name: …………………………………….

Signature: ....……………..   Date: ....…………..

Appendix C- Top Ten Discriminative Features for each user in Fast walk

#user	Accelerometer and Gyroscope Top Ten Discriminative Features									
1	159	217	220	56	281	32	162	165	70	67
2	43	70	67	34	15	12	57	117	10	13
3	56	117	268	105	120	10	13	123	265	166
4	10	13	56	41	117	123	105	120	4	34
5	34	135	32	159	217	220	268	281	283	254
6	10	13	268	159	220	217	41	56	93	165
7	268	34	10	13	56	57	80	163	166	265
8	10	13	41	268	117	32	265	56	4	283
9	268	117	105	120	10	13	159	43	217	220
10	56	105	120	10	13	20	41	109	27	5
11	34	10	13	56	105	120	268	117	185	41
12	165	162	217	220	159	43	67	70	15	12
13	268	10	13	221	218	56	160	41	123	166
14	34	123	56	268	105	120	166	163	135	10
15	268	56	34	10	13	123	105	120	159	217
16	159	217	220	165	162	218	221	10	13	160
17	159	217	220	10	13	165	162	268	56	41
18	34	123	56	32	166	163	10	13	221	218
19	67	70	43	10	13	56	12	15	123	41
20	34	268	111	144	31	116	57	115	113	106
21	10	13	67	70	41	43	117	56	159	217
22	56	32	123	268	159	217	220	165	162	117
23	167	164	268	159	217	220	219	222	10	13
24	34	159	217	220	268	165	162	10	13	254
25	159	217	220	268	34	32	56	10	13	222
26	43	12	15	70	67	5	57	123	56	10
27	268	56	10	13	185	105	120	265	138	41
28	268	34	56	105	120	10	13	32	123	185
29	34	117	105	120	185	56	268	138	32	166
30	268	218	221	160	163	166	265	283	10	13
31	20	268	218	221	286	160	166	163	256	265
32	10	13	56	41	217	220	123	159	218	221
33	70	67	20	43	56	126	217	220	55	159
34	165	162	217	220	159	268	70	67	10	13
35	56	34	286	10	13	185	117	123	267	32
36	268	10	13	41	283	265	159	165	162	220
37	268	32	10	13	56	105	120	34	41	116
38	268	32	10	13	286	256	166	163	115	283
39	10	13	254	221	218	67	70	41	56	160
40	268	56	117	32	10	13	163	166	223	221
41	34	117	67	70	43	185	165	162	159	15
42	10	13	34	220	217	159	165	162	56	41
43	268	34	56	267	32	144	150	123	135	265
44	34	268	265	31	56	10	13	32	256	123
45	105	120	56	166	163	268	10	13	283	123
46	12	15	43	159	217	220	67	70	165	162
47	268	10	13	56	34	132	302	256	41	266
48	10	13	56	41	166	163	218	221	105	120
49	20	221	218	160	268	163	166	93	217	220
50	10	13	268	41	56	221	218	117	160	166
51	34	268	221	218	10	13	160	282	223	194
52	10	13	41	56	268	105	120	166	163	218
53	10	13	117	56	281	73	287	105	120	41
54	159	217	220	281	287	165	162	10	13	221
55	56	217	220	159	10	13	254	34	268	165

(10, 13) Top Repeated	(56, 268) Second Repeated	(34, 159,217) Third Repeated

Appendix D- Top Ten Discriminative Features for each user in walking with a bag Activity

#user	Accelerometer and Gyroscope Top Ten Discriminative Features									
1	10	13	159	217	220	41	281	162	165	287
2	10	13	41	12	15	56	43	70	67	34
3	105	120	138	56	57	107	122	13	10	117
4	13	10	34	41	31	185	138	57	4	106
5	34	56	32	57	159	217	220	109	27	13
6	13	10	41	159	220	217	162	165	117	56
7	13	10	41	5	57	43	27	109	70	67
8	13	10	56	41	32	117	105	120	4	57
9	13	10	162	165	117	105	120	159	217	220
10	41	10	13	56	57	218	221	20	27	109
11	13	10	34	56	41	105	120	117	123	100
12	162	165	217	220	159	57	70	67	13	10
13	13	10	41	56	57	5	218	221	281	27
14	13	10	159	217	220	162	165	41	281	287
15	162	165	159	217	220	13	10	41	21	287
16	34	127	13	10	31	32	56	57	123	41
17	70	67	159	217	220	162	165	43	13	10
18	56	13	10	105	120	92	34	58	93	41
19	13	10	41	70	67	162	165	159	217	220
20	13	10	41	34	56	123	218	221	32	57
21	13	10	159	41	220	217	165	162	167	164
22	159	220	217	162	165	13	10	105	120	41
23	162	165	34	217	220	159	10	13	56	167
24	20	115	123	67	70	56	57	138	107	122
25	162	165	34	217	220	159	10	13	56	167
26	20	115	123	67	70	56	57	138	107	122
27	34	13	10	159	217	220	162	165	41	57
28	34	56	13	10	268	117	41	123	146	267
29	34	117	56	13	10	105	120	57	31	5
30	34	160	221	218	163	166	105	120	13	10
31	20	221	218	160	163	166	56	34	285	43
32	162	165	217	220	159	13	10	41	70	67
33	20	70	67	22	43	134	132	56	66	69
34	13	10	70	67	217	220	159	56	138	162
35	56	13	10	105	120	41	34	129	116	135
36	13	10	56	41	217	220	159	105	120	32
37	13	10	56	105	120	32	138	41	117	123
38	13	10	56	41	167	164	57	66	69	109
39	13	10	56	41	67	70	218	221	57	104
40	13	10	41	56	159	217	220	162	165	106
41	290	34	217	220	159	56	107	122	125	105
42	10	13	162	165	217	220	159	41	57	5
43	34	127	56	13	10	116	105	120	57	41
44	13	10	56	117	41	159	217	220	34	185
45	13	10	127	41	223	117	105	120	218	221
46	57	13	10	5	27	109	162	165	217	220
47	34	266	56	13	10	105	120	117	57	31
48	34	13	10	218	221	168	56	67	70	160
49	109	27	5	57	159	217	220	42	56	105
50	10	13	56	41	105	120	123	4	135	176
51	34	13	10	218	221	188	176	191	41	194
52	159	217	220	13	10	162	165	41	221	218
53	10	13	105	120	56	41	57	266	117	74
54	159	217	220	13	10	162	165	281	287	266
55	13	10	56	266	105	120	57	41	123	5
56	159	164	167	217	220	13	10	161	32	219
57	34	105	120	13	10	106	121	123	185	138
58	218	221	160	163	166	215	212	154	261	179
59	20	167	164	161	219	222	43	67	70	221
60	105	120	13	10	56	41	34	221	218	123

(13, 10) Top Repeated	(41, 56) Second Repeated	(217,220, 159, 34, 57) Third Repeated

Appendix E- Top Ten Discriminative Features for each user in Down Stairs Activity

#user	Accelerometer and Gyroscope Top Ten Discriminative Features									
1	32	10	13	268	64	165	162	43	217	220
2	287	195	12	15	10	13	43	93	34	41
3	70	67	61	13	10	3	185	9	268	123
4	115	32	34	10	13	195	161	262	180	219
5	10	13	117	236	267	82	234	41	47	15
6	100	205	34	281	10	13	138	56	117	125
7	10	13	32	41	67	70	138	12	15	43
8	268	56	10	13	41	115	100	265	32	102
9	41	15	12	43	10	13	67	70	265	108
10	48	20	105	120	244	164	167	10	13	56
11	93	32	56	34	268	170	117	265	10	13
12	113	268	116	93	123	12	15	56	73	117
13	215	221	218	124	10	13	106	121	56	67
14	41	108	26	116	114	10	13	82	43	37
15	105	120	10	13	203	56	77	34	268	99
16	57	282	67	70	288	10	13	168	43	194
17	268	10	13	135	36	221	218	70	67	18
18	32	34	10	13	29	31	92	47	138	117
19	29	136	10	13	268	116	67	70	43	41
20	115	74	36	106	121	10	13	100	268	17
21	13	10	41	67	70	47	106	121	114	31
22	32	10	13	160	218	221	171	105	120	117
23	34	268	107	122	265	32	10	13	92	31
24	275	283	268	160	163	166	221	218	12	15
25	263	10	13	227	266	269	57	41	205	47
26	115	57	123	5	90	56	11	14	138	102
27	268	43	129	67	70	106	121	15	12	73
28	200	34	109	27	80	89	141	144	113	42
29	56	113	124	32	116	93	267	268	92	246
30	218	221	160	35	57	13	10	268	166	163
31	139	20	100	12	15	218	221	43	206	287
32	117	13	10	64	15	12	160	56	32	49
33	67	70	43	22	136	107	122	115	132	3
34	206	32	100	13	10	117	145	268	222	219
35	105	120	268	10	13	117	4	41	116	115
36	32	34	136	268	117	29	56	105	120	97
37	32	56	10	13	105	120	27	109	29	57
38	34	207	100	267	5	56	92	10	13	31
39	159	13	10	154	137	168	220	217	67	70
40	268	10	13	194	12	15	168	265	160	117
41	268	221	218	117	267	56	105	120	212	160
42	144	43	12	15	117	201	110	28	138	49
43	57	105	120	56	5	27	109	11	14	42
44	268	34	166	163	157	265	10	13	107	122
45	268	116	57	265	13	10	266	106	121	32
46	56	254	266	57	268	115	10	13	105	120
47	134	268	265	57	116	221	218	144	168	10
48	275	10	13	221	218	56	30	116	33	117
49	10	13	207	56	100	4	106	121	199	138
50	10	13	135	41	47	266	62	56	163	166
51	246	10	13	266	254	168	221	218	12	15
52	247	160	134	154	268	221	218	43	67	70
53	56	245	94	106	121	116	244	100	105	120
54	287	10	13	268	221	218	160	41	56	281
55	34	105	120	290	56	13	10	31	268	231
56	168	221	218	154	160	10	13	170	215	105
57	13	10	34	32	265	41	232	116	56	268
58	218	221	160	163	166	268	215	155	171	212
59	20	10	13	56	57	216	160	116	155	4
60	10	13	117	4	105	120	41	56	160	12

(10, 13) Top Repeated	(268, 56) Second Repeated	(32, 117, 41, 218, 221) Third Repeated

Appendix F- Top Ten Discriminative Features for each user in Upstairs Activity

#user	Accelerometer and Gyroscope Top Ten Discriminative Features									
1	266	27	109	66	69	30	33	42	281	201
2	169	69	66	60	164	167	16	158	171	109
3	16	5	13	10	266	41	123	14	11	27
4	13	10	123	32	17	41	56	117	195	171
5	10	13	106	121	47	266	57	183	5	186
6	99	184	5	144	304	109	27	57	206	42
7	56	133	241	114	37	104	70	67	146	8
8	32	10	13	29	5	57	109	27	41	11
9	13	10	41	266	117	58	32	70	67	170
10	283	116	69	66	104	239	223	60	30	13
11	201	32	10	13	265	34	56	57	283	117
12	66	69	171	23	38	18	12	15	106	121
13	13	10	41	105	120	48	56	4	35	3
14	105	120	123	267	34	124	118	20	162	165
15	159	56	10	13	260	178	220	217	43	135
16	131	30	137	66	69	33	5	27	109	132
17	165	162	159	217	220	69	66	156	132	16
18	57	5	13	10	11	14	123	41	109	27
19	27	109	42	52	5	140	14	11	57	68
20	5	57	11	14	109	27	117	103	265	42
21	109	27	263	42	48	5	106	121	266	57
22	13	10	30	266	33	66	69	41	48	47
23	138	266	12	15	123	43	70	67	13	10
24	10	13	41	56	169	138	4	254	117	34
25	34	30	33	66	69	63	288	60	31	5
26	255	119	169	116	266	18	17	170	171	27
27	118	34	10	13	131	47	268	56	41	124
28	13	10	34	41	47	4	304	125	268	264
29	20	297	266	32	30	146	33	29	171	79
30	136	41	17	13	10	30	266	130	33	66
31	20	29	188	133	66	69	223	42	32	27
32	10	13	56	12	15	4	41	105	120	218
33	20	117	171	17	52	35	114	256	177	192
34	32	266	106	121	303	105	120	29	263	57
35	70	67	34	30	43	99	33	61	169	19
36	32	10	13	134	104	41	129	119	79	29
37	170	41	10	13	104	109	27	68	65	17
38	10	13	138	5	268	11	14	57	30	109
39	10	13	41	56	109	27	105	120	117	42
40	266	13	10	171	162	165	5	156	57	109
41	10	13	56	4	117	41	116	20	27	109
42	177	192	171	13	10	56	105	120	32	251
43	34	31	268	138	214	105	120	156	82	169
44	34	5	31	27	109	57	14	11	42	13
45	13	10	34	56	144	247	47	4	105	120
46	266	153	69	66	133	217	220	30	33	211
47	124	13	10	283	266	123	221	218	117	223
48	217	220	211	159	260	178	13	10	41	56
49	285	125	136	130	48	5	287	123	266	109
50	30	33	48	69	66	285	221	218	263	160
51	10	13	41	171	222	219	155	161	134	164
52	10	13	104	56	124	4	47	41	304	268
53	123	169	266	124	10	13	292	269	151	135
54	10	13	56	41	199	266	116	161	222	219
55	171	13	10	155	37	222	219	161	47	195
56	65	68	59	161	219	222	180	262	41	164
57	10	13	170	125	66	69	30	123	138	41
58	66	69	18	29	42	171	140	2	32	160
59	41	26	108	133	166	163	218	221	160	223
60	112	154	17	221	218	170	268	194	212	160

(13, 10) Top Repeated	41 Second Repeated	(266, 109, 27, 56, 5, 66) Third Repeated

Appendix G- Top Ten Discriminative Features for Each User in Fast Walking Activity

#user	Accelerometer and Gyroscope Top Ten Discriminative Features_ Real Data									
1	89	91	90	243	242	260	263	290	241	162
2	91	89	243	90	242	241	164	167	292	60
3	89	90	91	242	243	164	167	292	241	261
4	91	243	89	90	241	242	290	162	165	164
5	164	167	60	292	290	162	165	211	261	264
6	89	91	164	167	243	292	90	261	264	241
7	243	164	167	90	292	89	91	162	165	242
8	164	167	292	89	91	243	261	264	90	242
9	164	167	292	89	91	90	243	261	264	213
10	241	90	91	243	89	242	151	179	150	189
11	90	91	89	241	164	167	242	292	243	261
12	164	167	292	290	162	165	261	264	235	259
13	164	167	292	290	162	165	261	264	235	259
14	292	164	167	91	241	90	242	261	264	89
15	91	89	243	90	242	11	14	164	167	292
16	164	167	292	241	242	89	91	290	162	165
17	91	164	167	292	163	166	291	260	263	290
18	91	89	90	243	241	242	114	179	211	268
19	91	90	292	164	167	89	261	264	241	11
20	89	243	90	242	91	241	236	235	164	167
21	91	90	164	167	89	292	242	59	282	261
22	89	91	292	164	167	290	243	162	165	261
23	91	89	241	164	167	292	243	261	264	90
24	89	91	294	164	167	292	90	243	242	290
25	164	167	292	261	264	290	162	165	61	11
26	90	243	91	164	167	292	89	59	241	242
27	91	89	90	243	241	242	281	282	164	167
28	91	89	243	90	241	242	164	167	292	211
29	81	82	80	84	236	85	235	83	237	293
30	164	167	292	89	243	90	91	261	264	282
31	164	167	292	91	261	264	241	89	242	90
32	89	91	242	90	243	164	167	292	241	261
33	60	89	90	91	164	167	292	211	11	14
34	91	164	167	292	90	89	261	264	242	212
35	164	167	292	91	89	261	264	243	242	90
36	91	90	89	242	241	243	27	164	167	282
37	91	164	167	292	243	89	261	264	242	90
38	89	91	90	242	243	211	292	164	167	241
39	164	167	292	261	264	60	243	91	89	211
40	164	167	292	91	243	89	90	261	264	242
41	91	89	90	243	163	166	241	291	260	263
42	164	167	292	261	264	91	243	242	90	89
43	91	89	164	167	292	90	243	261	264	211
44	89	91	241	242	243	292	164	167	90	27

(91, 164, 89, 167) Top Repeated	(90, 292, 243) Second Repeated	(242, 261) Third Repeated

Appendix H- Top Ten Discriminative Features for Each User in Down Stairs Walking Activity

#user	Accelerometer and Gyroscope Top Ten Discriminative Features_ Real Data									
1	293	90	91	89	243	242	11	14	241	72
2	91	90	89	243	242	241	72	189	11	14
3	90	89	91	243	151	241	242	169	179	202
4	89	91	90	27	243	241	242	26	190	169
5	163	166	291	90	16	243	162	165	290	260
6	90	89	243	91	242	241	151	27	179	189
7	90	91	16	89	151	243	241	242	303	281
8	90	89	243	91	242	151	72	241	204	219
9	90	89	91	243	151	241	242	16	162	165
10	151	91	90	243	242	89	241	16	203	218
11	91	90	242	243	241	89	151	11	14	189
12	237	90	91	89	243	72	241	80	151	235
13	90	91	89	243	151	189	242	72	241	303
14	90	27	91	89	243	72	242	241	84	151
15	90	89	91	243	27	242	241	301	94	26
16	90	243	89	91	241	151	162	165	242	26
17	90	91	89	242	241	179	243	189	16	151
18	90	89	91	242	243	241	27	179	283	26
19	143	91	90	89	72	243	242	241	11	14
20	295	90	235	91	89	243	236	241	242	80
21	90	243	151	91	241	89	72	26	148	16
22	90	89	27	91	243	72	204	219	241	162
23	90	89	91	243	27	242	241	93	259	262
24	90	91	89	243	242	241	162	165	290	259
25	151	11	14	259	262	178	93	162	165	290
26	143	90	243	151	27	12	15	16	242	241
27	90	91	89	27	243	283	241	242	93	151
28	90	91	89	242	243	241	281	179	189	93
29	143	235	141	142	237	295	293	294	236	83
30	90	89	91	243	27	242	241	151	283	189
31	90	91	89	242	243	241	27	189	281	169
32	27	89	90	151	281	198	241	290	259	262
33	90	89	91	27	243	241	242	28	84	93
34	90	242	243	91	89	241	27	283	51	66
35	90	91	89	243	27	242	151	241	179	80
36	90	89	91	243	242	241	169	189	27	283
37	90	91	89	243	242	27	241	151	189	204
38	72	11	14	163	166	291	90	16	91	151
39	90	243	89	91	259	262	162	165	290	204
40	90	91	27	89	243	72	242	301	283	241
41	90	27	89	91	243	242	241	259	262	28
42	90	91	89	243	242	241	169	204	219	151
43	90	91	89	27	243	259	262	2	162	165
44	27	89	90	151	281	198	241	290	259	262

(90, 243) Top Repeated	(89, 91, 241) Second Repeated	242 Third Repeated

Appendix I- Top Ten Discriminative Features for Each User in Walking Upstairs Activity

#user	Accelerometer and Gyroscope Top Ten Discriminative Features_ Real Data									
1	270	213	243	170	151	195	109	112	211	187
2	241	243	89	16	204	219	90	162	165	290
3	241	294	89	27	243	2	90	51	66	72
4	162	165	290	259	262	241	109	112	302	93
5	27	63	84	241	212	151	16	243	145	149
6	241	243	27	89	189	212	90	17	188	242
7	127	11	14	130	164	167	261	264	292	162
8	127	11	14	130	63	84	241	243	162	165
9	72	241	243	84	235	27	89	150	93	188
10	241	243	89	281	63	235	90	188	189	242
11	63	84	241	127	27	130	243	11	14	2
12	235	237	236	241	27	243	188	82	151	89
13	11	14	164	167	93	292	243	162	165	241
14	243	241	89	281	2	84	51	66	27	169
15	90	243	241	17	27	89	152	179	283	188
16	162	165	290	27	12	15	127	130	11	14
17	188	302	151	16	241	290	162	165	148	152
18	241	243	89	93	188	90	242	27	204	219
19	109	112	241	27	235	243	89	90	140	155
20	243	235	179	302	189	150	198	188	236	72
21	11	14	127	130	162	165	12	15	241	290
22	162	165	241	290	84	243	72	27	63	259
23	2	84	63	51	66	27	241	102	109	112
24	2	84	51	66	109	112	162	165	290	93
25	241	27	243	169	89	188	304	204	219	72
26	243	241	51	66	109	112	84	2	283	3
27	241	243	27	283	84	89	90	17	235	2
28	241	27	2	84	243	51	66	63	109	112
29	127	130	237	11	14	259	262	162	165	290
30	27	241	243	89	51	66	189	2	281	109
31	2	51	66	162	165	290	259	262	93	109
32	241	243	27	235	162	165	89	302	290	259
33	212	2	17	51	66	179	54	124	241	27
34	127	130	11	14	63	84	162	165	290	51
35	243	84	51	66	241	2	27	189	169	93
36	27	241	89	243	188	90	83	189	242	143
37	241	243	27	89	84	63	72	51	66	90
38	243	2	241	27	84	235	89	51	66	102
39	127	130	11	14	63	162	165	290	84	259
40	93	109	112	204	219	11	14	63	140	155
41	162	165	290	259	262	63	12	15	84	109
42	84	2	51	66	27	93	241	109	112	243
43	243	241	27	89	90	294	304	204	219	189
44	241	63	243	109	112	2	51	66	302	17

(241, 243)	27	(84, 89, 162)
Top Repeated	Second Repeated	Third Repeated

Appendix J- Top Ten Discriminative Features for Each User in All Activities

#user	Accelerometer and Gyroscope Top Ten Discriminative Features_ Real Data									
1	293	90	91	89	243	164	167	292	242	212
2	90	89	91	164	167	292	242	241	243	261
3	294	90	91	89	292	164	167	241	212	243
4	91	89	90	243	241	212	242	164	167	292
5	141	143	164	167	292	212	261	264	163	166
6	90	89	91	164	167	292	243	212	241	261
7	295	293	164	167	292	261	264	90	61	212
8	167	164	292	90	91	261	264	89	241	243
9	293	294	91	89	164	167	292	90	212	243
10	241	91	90	61	242	212	89	211	243	164
11	90	91	89	241	164	167	292	242	243	261
12	141	142	143	293	294	295	164	167	292	290
13	164	167	292	90	91	261	264	89	212	163
14	90	241	91	164	167	292	243	89	242	212
15	164	167	91	292	89	90	212	61	261	264
16	164	167	91	292	89	90	212	61	261	264
17	164	167	292	290	162	165	261	264	259	262
18	293	90	91	89	241	242	243	212	211	164
19	143	293	141	294	91	90	89	292	164	167
20	141	142	143	235	237	236	164	167	292	261
21	164	167	292	212	90	261	264	61	290	11
22	293	90	292	164	167	89	91	290	162	165
23	295	164	167	292	261	264	89	91	90	290
24	294	295	91	212	164	167	292	90	89	290
25	164	167	292	261	264	293	294	295	143	141
26	141	143	167	164	292	61	261	264	90	212
27	293	91	90	89	164	167	292	243	241	212
28	294	293	91	90	241	89	242	243	164	167
29	164	167	292	293	294	295	261	264	163	166
30	90	91	89	167	164	292	261	264	243	212
31	293	164	167	292	290	162	165	261	264	91
32	293	90	164	167	292	91	89	241	242	243
33	212	61	90	211	89	91	242	60	241	243
34	293	164	167	292	89	261	264	90	212	91
35	293	164	167	292	90	91	242	243	89	261
36	143	295	294	90	91	89	241	212	242	243
37	295	164	167	292	91	261	264	89	90	241
38	91	90	89	212	241	243	292	164	167	211
39	143	164	167	292	261	264	212	61	90	12
40	164	167	292	212	261	264	89	90	243	91
41	141	142	294	90	91	212	61	89	163	166
42	164	167	292	261	264	91	90	212	89	290
43	294	295	91	89	90	164	167	292	243	241
44	293	91	90	89	241	290	162	165	259	262

164 Top Repeated	(89, 90, 91) Second Repeated	(243, 242, 72) Third Repeated