OXFORD

The Computer Journal, 2025, 1–17
https://doi.org/10.1093/comjnl/bxae144
Original Article

Real-world continuous smartwatch-based user authentication

N. Al-Naffakh [1,2,3], N. Clarke [1,4,*], F. Li [5], P. Haskell-Dowland [4]

[1]School of Engineering, Computing & Mathematics, University of Plymouth, Drake Circus, Plymouth PL4 8AA, United Kingdom
[2]College of Education, University of Kufa, P.O. Box 21, Najaf Governorate, Najaf 540011, Iraq
[3]School of Computing, University of York, York YO10 5DD, United Kingdom
[4]School of Science, Edith Cowan University, 270 Joondalup Drive, Joondalup WA 6027, Australia
[5]School of Computing, Bournemouth University, Fern Barrow Poole, Dorset BH12 5BB, United Kingdom
[*]Corresponding author. School of Engineering, Computing & Mathematics, University of Plymouth, Drake Circus, Plymouth PL4 8AA, United Kingdom.
E-mail: N.Clarke@plymouth.ac.uk

Abstract

User authentication is often regarded as the "gatekeeper" of cyber security. It has, however, long suffered from significant usability issues that have resulted in research focussing upon frictionless and transparent biometric approaches. Activity-based user authentication—a technique that authenticates a user by what they are physically doing at a specific point in time has attracted significant attention, particularly due to the increasing popularity of smartwatches. This research aims to overcome limitations in prior work by exploring the viability of the approach in real-world conditions. The study presents two principal experiments, one focused upon a constrained environment to provide a control and a second reflecting real-life. With over 1000 h of sampled data across 60 participants, the study sought to explore sensor, feature composition, and classifier design to explore the practical viability of the approach. Whilst the control experiment achieved best case Equal Error Rate of 0.29%, an improvement upon the prior art using optimisation, the best-case real-world results were not too far behind at 0.7%. This demonstrates that whilst the feature generated in the real-life experiment are subject to increased levels of noise, the performance is viable within the context of a transparent and continuous user authentication approach.

Keywords: biometrics; mobile authentication; gait recognition; user authentication; activity recognition

1. INTRODUCTION

Mobile devices have become an irreplaceable part of people's daily life; however, they are arguably more susceptible to risk (e.g. loss or theft) than other digital devices due to their small size and portability. Traditional user authentication methods (i.e. password and PIN) fall short of addressing these security concerns due to a lack of technical sophistication and their often-intrusive implementation. PINs are considered cumbersome, particularly for smartwatch users due to the small touch screen of those devices. Therefore, protecting the information and continuously checking the user's identity in a more innovative and convenient fashion is essential. To overcome those issues, the use of a Transparent Authentication System (TAS) was proposed that seeks to non-intrusively capture biometric information to verify a user's identity in a continuous fashion [1]. It is also referred to as *Active/Frictionless or Implicit* authentication. Such schemes seek to reduce the authentication burden upon the user whilst maintaining or improving the security being provided. Saevanee et al. [2] and Alotaibi et al. [3] highlight that there was an overall reduction of more than 90% in the explicit authentication requests by employing continuous authentication systems. Critical to such schemes working in practice is the availability of appropriate biometric modalities. Traditional biometrics are often not appropriate because they require external factors to be tightly controlled—such as light, orientation, background noise. For example, using a user's typing rhythm, ear, and face recognition techniques [4–8]. As such, the research community is actively developing novel transparent biometric modalities that are able to operate within a less constrained environments [8,9].

Biometric-based user authentication systems, which utilize commercial smartphone accelerometer and gyroscope sensors, have been proposed as a reliable and convenient approach focused upon the user's gait pattern such as walking and jogging [10–15]. More recently a focus has been given to smartwatches to determine whether the feasibility of using the accelerometer and gyroscope on the wrist offers the opportunity to provide more granular monitoring of physical movement. The fixed position of the smartwatch versus the more dynamic handling of mobile devices, offers up a potential benefit with respect to more consistent sensor information. Interesting, this also offers up the opportunity to go beyond simply gait recognition and perform recognition against a wider set of activities—often referred to as activity recognition. Much of the prior art has focused upon highly controlled experiments, where participants undertake specific activities within (often) single sessions. Whilst these limited activities, limited participants, and the highly controlled environment are useful for exploring the feasibility of an approach, they do not necessarily reflect real-world use,

Received: May 17, 2024. **Revised:** September 25, 2024. **Accepted:** December 18, 2024

Table 1. Critical analysis of gait/activity recognition studies

Study	Approach	Features Type	Classification methods	# of Participants	Duration	Device	Accuracy %
[10]	Cycle	Time	DTW	51	CD	Mobile	20.1 (EER)
[11]	Segment	Time	HMM	51	CD	Mobile	10.3 (EER)
[12]	Cycle	Time	DTW	48	CD	Mobile	21.7 (EER)
[13]	Segment	Frequency	KNN	36	CD	Mobile	8.24 (EER)
[14]	Cycle	Time	DTW	48	CD	Mobile	29.4 (EER)
[15]	Segment	Frequency	HMM	48	CD	Mobile	6.15 (EER)
[16]	Cycle	Time	KNN	51	CD	Mobile	33.3 (EER)
[17]	Cycle	Time & Frequency	SVM	14	SD	Mobile	91 (CCR)
[18]	Segment	Time	NN	8	SD	Mobile	98 (CCR)
[19]	Segment	Time	SVM	315	SD	Mobile	92 (TP) 1 (FP)
[20]	Segment	Time	JD	14	SD	Mobile	5.7 (EER)
[21]	Cycle	Time	GMM-UBM	35	CD	Mobile	14.4 (EER)
[22]	Segment	TD&FD	KNN	28	SD	Mobile	93.3 (CCR)
[23]	Segment	Time	RF	57	CD	Mobile	5.6 (EER)
[24]	Segment	Time & Frequency	SVM	10	SD	Mobile	99 (CCR)
[25]	Segment	Time & Frequency	BN	10	SD	Mobile	94.5 (CCR)
[26]	Segment	Frequency	HMM	102	CD	Mobile	4.93 (EER)
[27]	Segment	Time & Frequency	DTW	24	CD	Mobile	3.8 (EER)
[28]	Segment	Frequency	CNN	387	SD	Mobile	98 (CCR)
[29]	Cycle	Time	LSTM	30	SD	Mobile	90 (CCR)
[30]	Segment	Time	CNN & LSTM	20	CD	Mobile	99.5 (CCR)
[31]	Segment	Time & Frequency	LSTM	50	SD	Mobile	88 (CCR)
[32]	Segment	Frequency	-	29	SD	Mobile & Smartwatch	17.6 (EER)
[33]	Segment	Time	RF	18	CD	Mobile & Smartwatch	4 (EER)
[34]	Segment	Time	SVM	20	SD	Mobile & Smartwatch	0.65 (EER)
[35]	Segment	Time & Frequency	CNN	50	SD	Mobile & Smartwatch	98 (CCR)
[36]	Segment	Time	RF	51	SD	Mobile & Smartwatch	0.8 (EER)
[37]	Segment	Time	RF	15	SD	Mobile & Smartwatch	95.3 (CCR)
[38]	Segment	Time & Frequency	KRR	20	CD	Mobile & Smartwatch	8 (EER)
[39]	Segment	Time & Frequency	KRR	35	CD	Mobile & Smartwatch	1.85 (EER)
[40]	Segment	Time	RF	51	SD	Mobile & Smartwatch	9.3 EER
[41]	Segment	Time	KNN	15	SD	Smartwatch	2.9 (EER)
[42]	Cycle	Time	DTW	5	CD	Smartwatch	22.5 (EER)
[43]	Segment	TD&FD	KNN	10	SD	Smartwatch	88.4 (TP) 1.3 (FP)
[44]	Segment	Time	SVM	20	CD	Smartwatch	6.56 (EER)
[45]	Segment	Time	SVM	20	CD	Smartwatch	4 (EER)
[46]	Segment	Time	Man	10	SD	Smartwatch	4.27 (EER)
[47]	Segment	Time	KNN	20	CD	Smartwatch	5 (EER)
[48]	Segment	Time & Frequency	DT	6	CD	Smartwatch	90.4 (CCR)
[49]	Segment	Time & Frequency	MP	34	SD	Smartwatch	1 (EER)
[50]	Segment	Time & Frequency	MLP	20	SD	Smartwatch	1 EER
[51]	Segment	Time	KNN	10	SD	Smartwatch	98 (CCR)
[52]	Segment	Time	LSTM	36	SD	Smartwatch	99 (CCR)
[53]	Cycle	Time	-	20	CD	Smartwatch	5 (EER)
[54]	Cycle	Time	RF	17	SD	Smartwatch	99 (CCR)
[55]	Segment	Time	GMM-UBM	10	SD	Smartwatch	91.5 (CCR)
[56]	Segment	Time & Frequency	SVM	21	SD	Smartwatch	0.2 (EER)
[57]	Segment	Time	DTW	4	CD	Smartwatch	28 (EER)
[58]	Segment	Time	AdaBoost	30	-	Smartwatch	88.4 (CCR)
[59]	Segment	Time	-	10	SD	Smartwatch	2.25 (EER)
[60]	Segment	Time	DNN	20	SD	Smartwatch	94 (TAR) 0.3 (FAR)
[61]	Segment	Time	RNN	20	SD	Smartwatch	0.78 (EER)
[62]	Segment	Time	MLP	49	SD	Smartwatch	87 (TP) 0.2 (FP)
[63]	Segment	Frequency	RNN	22	SD	Smartwatch	0.83 (EER)
[64]	Segment	Time	RF	32	SD	Smartwatch	99.5 (TP) 3.6 (FP)
[65]	Segment	Time	RNN	11	CD	Smartwatch	10.6 (EER)
[66]	Segment	Time	MLP	64	CD	Smartwatch	2 (EER)
[67]	Cycle	Frequency	RNN	30	CD	Smartwatch	6.6 (EER)
[68]	Segment	Time	KNN	36	CD	Smartwatch	3.5 (EER)
[69]	Segment	Time	RNN	22	SD	Smartwatch	0.83 (EER)

Legend: DTW: Dynamic Time Warping; HMM: Hidden Markov Model; SVM: Support Vector Machine; KNN: k-nearest neighbours; RF: Random Forest; NN: Neural Network; RF: Random Forest; BN: Bayesian Network; KRR: Kernel Ridge Regression; GMM-UBM: Gaussian Mixture Model-Universal Background Model; JD: DT: Decision Tree; MP: Multilayer Perception; Jaccard distance; EUC: Euclidean Distance; Man: Manhattan Distance; RNN: Recurrent Neural Networks; CNN: Convolutional Neural Networks, LSTM: Long Short-Term Memory classifier; EER: Equal Error Rate; CCR: Correct Classification Rate; TP: True Positive; FP: False Positive; SD: Same Day; CD: Cross Day.

where signal data collected from sensors are very likely to be considerably noisier, resulting in a more challenging classification task.

This study investigates the viability of activity recognition with a focus upon understanding the real-world performance that can be achieved. Focused upon collecting a large uncontrolled dataset, the study explores the impact upon sensor data and the resulting recognition performance that can be achieved in practice. To provide a basis for comparison, the study also includes a control experiment—following a similar experimental methodology as the prior art to both provide a baseline for performance and to also provide an opportunity to explore additional gaps in the prior knowledge. These advances in feature selection and classifier design in the control experiment are then fed through into the real-world study.

The remainder of the paper is organized as follows: Section 2 reviews the state of the art in TAS specifically focusing upon approaches that have utilised the accelerometer and gyroscope sensors. Data collection, pre-processing, feature extraction, and a new feature selection approach are outlined in Section 3. Sections 4 and 5 present the experimental results and discussion. Section 6 presents the conclusions and future directions of research.

2. RELATED WORK

Identifying what people are doing based on their daily activities can be useful for several mobile applications (e.g. health-care and fitness tracking) and can offer a robust and continuous biometric-based user authentication system. The first attempt [11] focused upon wearing a dedicated sensor around the human body such as hip, ankle, and arm in order to collect the user's motion data. Although these studies could determine physical actions (e.g. walking and gesture) and reported encouraging results, the need to use costly specialized devices for the data collection and a comprehensive set-up reduced the usefulness and increased the implementation cost in a real-world system. Recently, a growing body of studies [10–27] have utilised smartphone accelerometer and gyroscope sensors for biometric-based authentication systems. However, the majority of these studies relied upon limited activities (i.e. gait or gesture). Despite the increased popularity of smartwatches, little work has been conducted to verify the user's identity based upon their activities. A comprehensive analysis of the prior studies on smartwatch/mobile-based user authentication using focused upon gait an activity recognition is presented in Table 1.

Much of the prior art collected the accelerometer and gyroscope data with a known pocket position. Smartphones, while having the benefit of technological maturity and widespread adoption, arguably suffer from many problems in placement to produce a consistently effective implementation. Orientation and off-body carrying (i.e. when the device is not carried in a pocket or somewhere else close to the body) make obtaining consistent movement data challenging. A study conducted by [70] highlighted that the system's performance was degraded when the position or orientation of the smartphone were changed. In contrast, smartwatches guarantee consistent placement on the body regardless of the clothing choices of an individual user.

The Cross-Day (CD) evaluation, which trains a classifier using one-day and verifies it with data from a different day, is generally accepted as the most robust experimental methodology. However, a considerable amount of gait and gesture-based user authentication studies trained and tested the system based on data that are collected on a Single-Day (SD), which tends to be a less realistic scenario as human behaviour changes over time. Most studies claiming a system resilient to the CD problem either trains on mixed data from both days (thus not making it a true CD test as a user will be required to enrol in the system every day) or has an error rate so high that the system would not be practical. Notably, the lack of realistic data underpins a significant barrier in applying these systems in practice (in both mobile and smartwatch contexts).

Feature extraction is a key component in the development of any classifier. There are two main methods: cycle-based (which segments data into pairs of steps) and segment-based (which focus on fixed-length blocks of data). Previous findings in the literature supported the use of a segment-based approach as it appears to produce a more effective and stable performance, with studies reporting Equal Error Rate (EER) between 0.65% and 28% [34,57]. In contrast, the performance of using a cycle-based approach is typically worse with EERs ranging from 5% to 33.3% [16,53]. This is most likely the result of the complicated and unclear nature of cycle extraction. With respect to features, little attention was given to measuring the impact of time domain and frequency domain features on the system performance.

Previous studies [32,34–37,40,41,43,46,49–52,54,55,59–61,63,64–69] used a smartwatch device in or1-der to collect the accelerometer and gyroscope signals for the same day scenario. However, the authors did not carry out any investigation on the feature selection process in order to identify the most discriminative features. Moreover, a limited range of activities (i.e. gait or gesture) were considered. To improve the classification results and reduce the computation time, it is important to analyse the discriminative ability of the extracted feature set due to their influence on the system performance. As a result, various feature selection approaches have been proposed in the prior art. For example, forward search was based upon testing each feature independently and then selecting a subset of features that reported the best classification performance. However, the implementation of this method is not necessarily accurate as the extracted features can be relatively correlated to each other. In comparison, several biometric-based user authentication systems create the user's reference and test templates based upon selecting the most common features (e.g. features that have the smallest standard deviation [STD] for all the population). Studies such as [33,41,44,45,71] applied different mechanisms when exploring the feature vector. For example, analysis of the user's gait pattern resulted in 95% Correct Classification Rate (CCR) when the SD scenario was applied on a dataset of 40 users [71]. However, the system performance reduced to 86.8% (with a limited dataset of only 13 users) when the CD scenario was used. This can be attributed to the proposed approach not being sophisticated enough to identify a unique feature set for individuals that work overtime.

Other gait-based authentication studies by [33,41] were able to achieve low EERs of 4% and 2.9%, respectively. One of the major drawbacks to adopting the former study is the feature vector size (i.e. 326 features) that was constructed based on data of seven sensors from the smartwatch and phone. Therefore, complex computational processing was required and subsequently a higher demand upon the battery (which is one of the biggest problems of these approaches). While the latter research [41] was able to shrink the user's reference template by identifying the most distinctive features, their finding (i.e. an EER of 2.9%) was based upon a dataset of only 15 users with their movement data being collected on the same day. Therefore, further research is arguably required to devise a novel feature selection strategy that

can offer a delicate balance between usability and security that uses more realistic data.

Motion-based authentication systems hypothesize that each individual has a unique pattern such as handwriting [44,52] and gesture pattern [51,59]. However, such studies focus upon intrusive authentication (i.e. a user is asked to perform a specific activity to be authenticated) rather than seeking to capture data transparently and offer continuous frictionless user authentication. An accelerometer-based study [47] proposed an algorithm that automatically detects the gait pattern while the user walks on different surfaces (e.g. grass and asphalt). Although the proposed system reported encouraging results with an EER of 5%, the users' data was collected within a controlled environment and the authentication process was limited to the availability of the user's gait information only. The most recent reviews of the literature [38,39,58,65] were more realistic in terms of the data collection phase (i.e. users were not asked to perform certain activities, but to wear a smartwatch during their day-to-day activities). Lee et al. [38] show that utilizing the combination of sensor data from a smartwatch and a smartphone could greatly enhance the authentication performance, with an EER of 7.9%. The EER, however, increased to 17.85% and 13.4% by using the mobile motion sensors only (i.e. accelerometer and gyroscope). A major criticism of this study is that the data was not categorized based on the user's daily activities. An extended study [39] examined the possibility of dividing the real-life data into two types (i.e. moving and stationary data) by developing a context detection approach to predict the user's activity. Identifying the activity type before the classification process successfully reduced the error rates (i.e. an EER of 1.85%). Although, the system was evaluated using unlabelled data, the proposed context approach was trained by using constrained data (i.e. all participants were asked to use the smartphone for around 80 min in order to detect different contexts such as using the smartphone while sitting, standing, moving, and sitting on a moving vehicle). While other gait-based studies claimed to use data obtained from an uncontrolled environment, [58] utilised a very limited set of samples (ranging from 10–57) and [65] used a very limited set of participants (only 11).

3. EXPERIMENTAL METHODOLOGY

In order to overcome the shortcomings of prior work, this study identified the following research questions (RQ):

1) How well can an activity-based user authentication system perform in real-life environments?
2) What is the most effective classification strategy—single or an activity-specific multi-classifier?
3) What is the optimum feature vector composition?

RQ1 focusses upon exploring the practical realities of utilising activity recognition. In comparison to the controlled experiments in the prior art, this involves using unlabelled data, in wholly uncontrolled environments. The resulting biometric signals are very likely to much noisier. Building upon that understanding, RQ2 was devised to explore whether a single-classifier or a multi-classifier approach would be beneficial. The working hypothesis being that given more noisier data, a single classifier might struggle in comparison to a multi-classifier approach, which can focus upon specific activities. The challenge is how to determine those activities using an unlabelled dataset. RQ3 looks to explore how to optimise the feature vector—seeking to reduce its size and the resulting classifier complexity—whilst maintaining performance.

To address these questions, the following experiments were conducted:

1) Evaluation of the activity-based user authentication system using a controlled and uncontrolled environment (RQ 1).
2) Exploration of single classifier versus an activity-specific multi-classifier approach (RQ2).
3) Investigation of static feature vectors versus dynamic feature selection (RQ3).
4) Investigation of time versus frequency domain features, acceleration versus gyroscope-based signals, and same day versus cross day evaluation scenario (RQ3).

3.1. Data Collection

This section describes the data collection methodology that was used to collect the two different datasets: controlled and real-life. The former contained labelled data of five different activities that was captured within a laboratory environment, while the latter dataset contains real life data (i.e. captured within an unconstrained environment over a prolonged period and unlabelled).

3.1.1. *Controlled data*

To determine and evaluate the feasibility of the proposed activity-based user authentication, it is important to ensure the population sample being used is as large and significantly reliable as possible. Therefore, this experiment, in comparison to the prior art, aimed to capture a significant number of samples from each individual and to have a substantial population. They study collected over 60 h of movement data from 60 participants. To collect user's movement data, the Microsoft Band 2 was utilized due to its wide range of built-in sensors. A third-party application, which is compatible with all Android smartwatches and smartphones, was used to capture the acceleration and gyroscope signals [72]. Whilst the application offered three different sampling rates (i.e. 16 Hz, 32 Hz, and 128 Hz) and data was captured at 32 Hz as it offered a trade-off between processing and storage limitations of the device versus a desire to capture data with sufficient resolution. As soon as data was acquired by the smartwatch, it was sent to a smartphone residing in the user's pocket via Bluetooth.

As highlighted earlier, the main aim of this study is to remove or minimize the intrusiveness of current user authentication approaches, therefore, certain activities were considered that are non-intrusive, frequently used, and contain unique arm patterns. These included five physical activities: normal walk (NW), fast walk (FW), typing on PC (TPC), playing mobile game (MobG), and typing on mobile phone (TMob). Based on the previous studies [10,12,13,34,46] that captured acceptable numbers of samples (in the range of 36 to 100 samples) and achieved good levels of accuracy, this study aimed to have at least the same amount of data. For each activity, 72 samples were obtained from each participant resulting in a total of 360 samples. This was collected over two days within a time frame of 3 weeks. In total, 60 h of movement data were collected from 60 users (26 males and 34 females) with ages ranging from 18–55. Ethical approval was sought and approved and written informed consent was obtained from each test subject prior to data collection.

Raw gait signals (i.e. NW and FW) were collected by asking users to walk on a predefined route and encouraged to walk normally on flat ground. For consistency, the gait data was collected on the second day in a manner similar to the first, with the user walking over the same route (this is standard practice drawn from the prior work). For a more realistic scenario, the user

had to stop to open a door and take multiple turns. No other variables, such as type of footwear or clothing, were controlled. For the typing activities (not previously researched by the prior art), participants were asked to sit and continuously type a short and predefined text on the touch screen of their smartphone and on a PC keyboard. For the game activity, Candy Crush Saga was selected to capture the user's arm movement. The rationale for selecting this game was based on being the top mobile game by downloads at the time of the experimental design, was a free application to install, simple to play, and contained enough touch gestures to obtain sufficient data for each individual.

3.1.2. *Real-life data*

Whilst the labelled data collected in controlled conditions is useful for initial feasibility studies, it is arguably less realistic than normal use. Therefore, to evaluate the proposed system under an unconstrained real-life environment, users were asked to wear the smartwatch for 10 days. Participants were encouraged to freely undertake their daily routine to ensure the collected data represented the user's normal behaviour. The acceleration and gyroscope data streams were collected from 30 users (17 males and 13 females with an average age of 24); those users were a subset of the 60 users from the original controlled experiment. Each user was asked to wear the watch for at least 4 h per day (or until the smartwatch battery was drained). The total collected data per user was ∼40 h (4 h ∗ 10 days) with around 1200 h in total for all users. Based upon prior art to date, this is the largest smartwatch-based sensor dataset utilised to date.

Once the smartphone and watch were turned on, the signal was captured in a continuous and transparent manner. For consistency, the sampling rate was fixed for all participants. The dataset consisted of 32 327 gait samples (i.e. for both normal and FW). This compares with the prior accelerometer-based studies that collected limited datasets ranging between 900 and 1000 samples. A further 93 637 non-walking (Non-W) samples were also collected, which is the first acceleration/gyroscope-based smartwatch study that used a stationary signal for the TAS.

3.2. Data pre-processing

Pre-processing provides a mechanism to remove unnecessary or irregular data and transform the raw signal into discriminative information prior to its use in calculating the feature vector. Based upon the techniques used in the prior art, this study took the following steps:

- **Removing unworn signals**: As long as the smartwatch is on, the application would keep running in the background and capture the movement data in a continuous manner. Therefore, data from the galvanic skin sensor was used to identify periods where the band was removed.
- **Time interpolation**: Due to the limited accuracy of the sensors in the Microsoft Band, the smartwatch was not able to record data at a fixed sample rate (i.e. the time intervals between two successive acceleration values were not fixed). Therefore, time interpolation was applied to ensure that the time between two successive data points was always equal.
- **Filtering**: a low pass filter was applied in order to enhance the accuracy of the signal. This was carried out with several settings (i.e. 10, 20, and 30 Hz), and through experimentation a cut-off frequency of 20 Hz achieved the best accuracy (examples of the filtering are shown in Fig. 1).
- **Segmentation:** Classification approaches do not typically operate directly on time-series data and require data to be represented as a set of samples. The prior literature supports the use of segment-based rather than the cycle-based approach as it appears to be more effective and stable. Therefore, the tri-axial raw format for both acceleration and gyroscope signals were segmented into 10-s segments, which ensures that each sample includes several cycles and any brief period of non-movement signal (e.g. a pause) will not dominate the sample. This was achieved by using a sliding window approach with no overlapping.

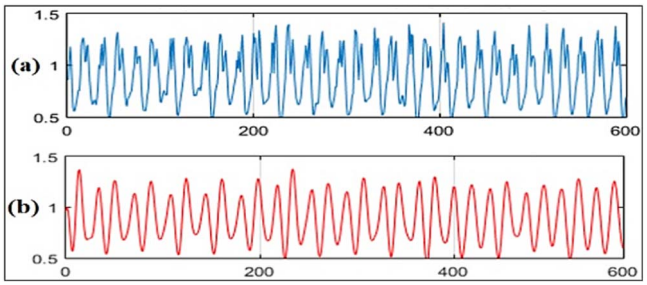

Figure 1. Acceleration signal (a) before filtering and (b) after filtering.

3.3. Feature extraction and feature selection

Feature extraction is a key component of any biometric system and needs to contain the discriminative information necessary for classification. Therefore, a comprehensive feature extraction process was carried out on both the accelerometer and gyroscope sensor data. Based upon prior art, features were extracted in both the time and frequency domains and resulted in 140 features [20–46]. These features are the same regardless of whether the sample is being generated from the acceleration or gyroscope sensor data. Since most features were generated on a per-axis basis and each sensor has three axes, most features were represented by a vector of three values.

The process of extracting frequency domain features is somewhat different from the time domain. Before extracting a frequency domain feature, a Fourier transform is applied to the motion data. A set of frequency domain features are calculated which might be useful to create a discriminative feature vector for each individual. Details of these features (e.g. what they are and how they are calculated) are described in Table 2.

Feature selection is used to select an appropriate feature subset from the entire set of features by identifying the most optimal and user discriminative features for the machine learning algorithms. When the feature set size is relatively large, feeding all features to a classifier without selecting a distinguishing feature subset might negatively affect the system performance and results in classification complexity problems due to the curse of dimensionality [73,74]. Therefore, feature selection has become the focus of many research studies in order to reduce the potentially large dimensionality of input data, with the resultant effect of enhancing performance and reducing the computational complexity of the classifier [41,44,45,71,73–76]. This becomes even more important when considering the limited processing capability and battery limitations of mobile devices.

A variety of feature selection approaches exist in order to explore and show the effectiveness of the extracted feature vector; whilst some of the prior gait/activity-based studies [13,15,18,40,48] proposed strategies for identifying the more discriminative features from the population, the feature selection approaches that have been presented so far do not seem to excel

Table 2. Time and frequency domain features

Feature	Description	NF	TD	FD
Interquartile range	The range in the middle of data. It is the difference between the upper and lower quartiles in the segment.	3	✔	✔
Skewness	A measure of the symmetry of distributions around the mean value of the segment.	3	✔	✔
Kurtosis	A measure of the shape of the curve for the segment data.	3	✔	✔
Percentile 25,50	The percentile rank is measured using the following formula: R = (P/100) ∗ (N + 1). Where R represents the rank order of the values, P: percentile rank, and N is the total number of data points.	6	✔	✔
Correlation Coefficients	The relationship between two axes is calculated. The Correlation Coefficients is measured between X and Y axes, X and Z axes, and Y and Z axes	3	✔	✔
Difference	The difference between the maximum and minimum of the values in the segment.	3	✔	✔
Median	The median values of the data points in the segment.	3	✔	✔
Root Mean square	The square root of the mean squared.	3	✔	✔
Maximum	The largest 4 values are calculated and averaged.	3	✔	✔
Minimum	The smallest 4 values are calculated and averaged.	3	✔	✔
Average	The mean value of the values in the segment for each axis	3	✔	✔
Standard Deviation	The standard deviation is a measure of how spread the data points from the mean. It is calculated for each axis.	3	✔	✔
Average Absolute Difference	The average absolute distance of all values in the segment from the mean value over the number of data point in the segment (for each axis).	3	✔	✔
Time Between Peaks	During the user's walking, repetitive peaks are generated in the gait signal. Thus, the time between consecutive peaks was calculated and averaged (for each axis).	3	✔	-
Peaks Occurrence	Determines how many peaks are in the segment	3	✔	-
Variance	Average of the sum of the squared differences of each value in the segment from the mean over the segment size (for each axis).	3	✔	✔
Cosine Similarity	All pairwise cosine similarity measurements between axes.	3	✔	-
Covariance	All pairwise covariances between axes.	3	✔	-
Entropy	Spectral entropy describes the complexity of the signal based on the Shannon entropy	3	-	✔
Energy	The summation of the mean square of each frequency component multiplied by time interval of the signal	3	-	✔
Binned Distribution	Relative histogram distribution in linear spaced bins between the minimum and the maximum acceleration in the segment. Ten bins were used for each axis	30	✔	-
Average Resultant Acceleration	For each value in the segment of x, y, and z axes, take the square roots of the sum of the values of each axis squared over the segment size (i.e. 10 s).	1	✔	✔

in terms of performance. The wider literature on biometrics has demonstrated that a user-specific feature vector will seek to maximize the discriminative information per user rather than across the whole population—resulting in a stronger and more unique set of features. To this end, this study carried out an exhaustive exploration of data using descriptive statistics for a better understanding of the nature of features and to explore the relationship between inter and intra variance that might exist.

The output of this exploration process was used to develop a dynamic feature vector algorithm to see how that would impact system performance. For each individual, a unique feature subset was generated (i.e. creating a dynamic feature vector that contains distinctive features for each user). This was achieved by calculating the mean and STD for each feature individually for all users. Thereafter, comparison of the authorized user's results against impostors to select the feature set with minimal overlap; for each feature, a score was calculated based upon the following condition:

- If the mean of imposter's feature was not within the range of the mean +/− STD of genuine, add 1 to the total score.
- Dynamically select the features according to their score order from high to low. The highest means less overlap between imposters and the genuine user.

Figure 2 shows an example of applying the proposed feature selection method (using the real-life dataset) on two different features for *User 1*. Based upon the overlap percentage, it is clear from Fig. 2 that the Kurtosis feature has the lowest overlap score compared to the Variance feature. As a result, the Kurtosis feature was selected to form the feature vector, while the second feature (i.e. Variance) was neglected. This procedure is repeated for each individual and each feature resulting in a bespoke and prioritized feature set. Given that the proposed feature selection method is highly dependent on the dataset (especially data for imposters), the proposed dynamic feature algorithm is designed to automatically update the user's feature vector. Once the biometric data is obtained, the proposed algorithm samples the data and acquires the most appropriate feature vector to adapt to the changes of the user's feature vector.

3.4. Experimental procedure

The aim of biometric-based authentication is to determine if a system can classify a user correctly (i.e. a genuine user or as an imposter). This study utilized two approaches namely, single and an activity-specific multi-classifier. The former was created by using unlabelled training data, which means there was no prior knowledge about the activity while a separate reference template was generated for each activity in the latter. Therefore, in total, six unique reference templates were created for each user (i.e. five activity-based templates and one template was built by using the unlabelled samples). Once these models were prepared, the

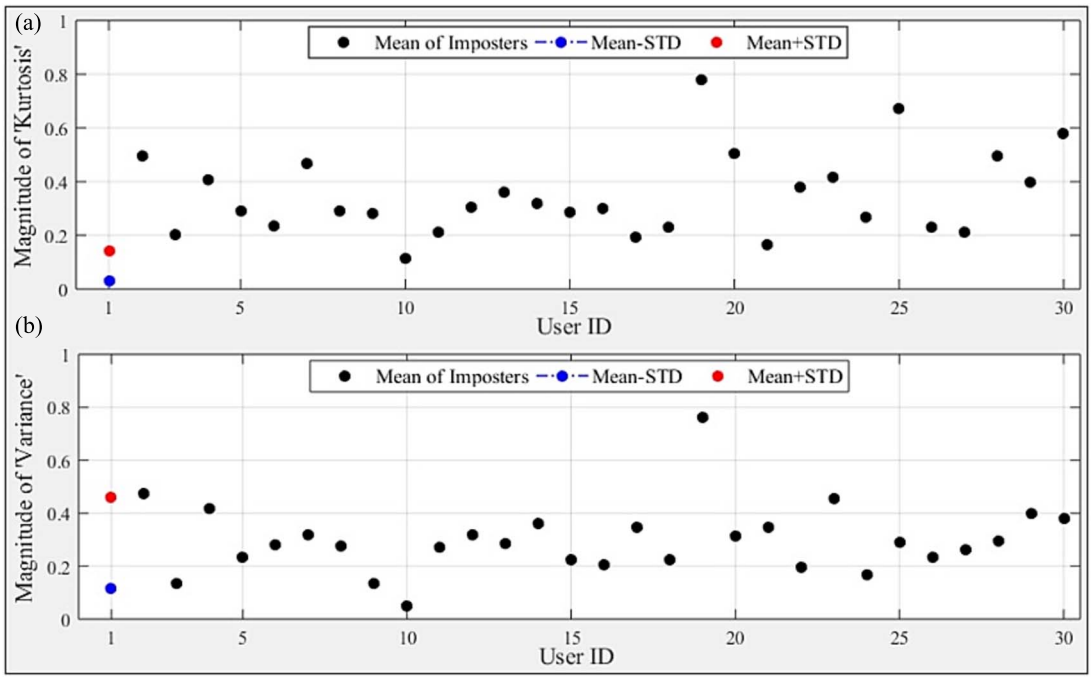

Figure 2. (a) Example of features with minimum overlap (b) example with features with maximum overlap.

Table 3. Dataset characteristics

Dataset	Activity	# of samples	# of participants	# of features	% for training	% for testing
Controlled	All	21 600	60	Up to 140	60	40
	Normal Walk (NW)	4320	60	Up to 140	60	40
	Fast Walk (FW)	4320	60	Up to 140	60	40
	Typing on a PC (TPC)	4320	60	Up to 140	60	40
	Playing mobile game (MobG)	4320	60	Up to 140	60	40
	Typing on a mobile (TMob)	4320	60	Up to 140	60	40
Real-Life	Normal Walking (NW)	20,218	30	Up to 140	60	40
	Fast Walking (FW)	10,309	30	Up to 140	60	40
	Non-Walking (Non-W)	94,444	30	Up to 140	60	40

reference and testing templates were created under two different scenarios (i.e. SD and CD). In the SD scenario, data set was divided into two parts: 60% was utilised to train the classifier while the remaining 40% was used to evaluate the performance. The reason for selecting this ratio (i.e. 60% versus 40% for the training and testing respectively) is to ensure that the classifier is trained with sufficient representative samples and to evaluate the robustness of the proposed system. Table 3 provides a breakdown of the datasets, the volume of samples and participants across each of the activities.

To evaluate the controlled experiment under the CD scenario, the first day data was used for training, while the evaluation was carried out by employing the second day data. Once the user's templates were created, a Feedforward Multi-Layer Perceptron (FF MLP) neural network was used as the default classifier as the neural network requires less training data compared to SVM, HMM, and KNN [13,77] and shows reliable performance for the proposed system [74,78,79]. For each experiment, four different FF MLP neural network training sizes were evaluated (i.e. 10, 15, 20, and 25) with each being repeated 10 times in order to account for errors that occur due to the random setting of the neural network weights. Several extensive experiments were conducted; for the controlled data, a total of 432 000 tests (i.e. 7200 tests for each user including the variation of the network, feature subset, the SD and CD evaluation scenario, activity type, and the sensor type). The experimental setup for the real-life experiment included a total of 64 800 tests (2160 tests per user). The step-by-step process for the activity-based authentication system is illustrated in Fig. 3. It highlights the key phases from data acquisition through to authentication decision.

4. RESULTS

The results presented in this study are the core findings from all of the tests undertaken across the controlled and real-life experiments. Based upon the conducted experiments, an FF MLP neural network of size 10 is presented as it resulted in the lowest EER. The EERs were calculated by finding the crossover point between the FRR and FAR as it is the de-facto performance metric used in prior studies. The findings are presented in two sections, reflecting the core experiments identified: controlled activity-based and real-life. In the controlled activity-based experiment, four core aspects were addressed:

- an investigation in time and frequency domain features,
- the impact upon performance given a variety of feature vector compositions

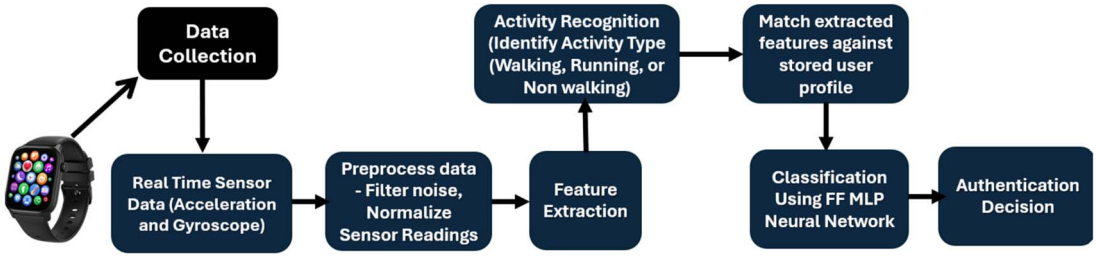

Figure 3. The proposed architecture for user authentication based on activity data from smartwatch.

Table 4. EERs of using the fusion features, time, and frequency domain

Feature type	# of Features	EER (%)	
		Accelerometer	Gyroscope
Fusion features	132	0.13	3.37
Time domain	88	0.15	3.73
Frequency domain	52	3.09	12.69

- highlighting the effect of the SD and CD scenario tests on the classification accuracy.
- investigated a single versus an activity-specific multi-classifier.

The second real-life experiment looks to build upon this knowledge and apply it to the real-life data.

4.1. Activity-specific experiment

The purpose of collecting data under a constrained environment is to investigate and look at an experiment that provides the empirical basis to the feasibility of using smartwatch movement sensors (i.e. accelerometer and gyroscope) and capturing a wide range of activities (not merely gait or gesture). Given the labelled and controlled nature of the activity, it also provides for a stable basis for comparison and analysing feature composition and selection.

4.1.1. *Time vs frequency domain features and sensor selection*

Although the extracted gait/activity feature set of the prior literature consisted of features from both a time and frequency domain, many studies have not considered their impact on system accuracy. Therefore, the aim of the first experiment was to highlight the usability and impact of the time and frequency domain features on the recognition rate. Table 4 displays the EERs of using the fusion of both features against time and frequency domain by utilizing the SD scenario of the NW activity.

Prior studies [33,41,44,45,73] already demonstrated that the incorporation of more features usually degraded the classifier's accuracy (often due to some of the features being irrelevant and/or redundant). Table 4 shows that a good performance was achieved (i.e. 0.15% and 0.13% EERs for acceleration) by using the TD and all feature sets; with little difference in results being observed between the two sets. By using the FD features alone, reasonable performance is obtained (i.e. 3.09% EER); but its performance is far less promising in comparison with the results of using TD features alone and all feature sets. This suggests that FD features add little contribution towards the classification process. Moreover, it is computationally expensive for the system to compute these features in real time on smartphones and/or smartwatches [10,33,41,44,71].

Although sensor based-authentication systems could be implemented using accelerometer and/or gyroscope as the source triaxial sensor, in practice few studies have tested systems using both sensors independently. Intuitively, both should offer similar information and thus similar levels of predictive power; however, the results in Table 4 overwhelmingly supported the use of the accelerometer sensor alone for activity-based user authentication systems. An EER of 0.15% was obtained by using the acceleration data compared to 3.73% when the gyroscope data was tested. A further analysis was conducted to reflect the EER spread within the population and the results are presented in Fig. 4. This shows that the EERs using the gyroscope signal were significantly increased for the majority of users in comparison to the acceleration data. As a result, all the subsequent results for the controlled experiment are based on the use of the acceleration data only.

4.1.2. *Single classifier versus an activity-specific multi-classifier*

It is argued, given the variability of signal data, creating specialized classifiers (i.e. activity-specific multi-classifier) could exhibit better recognition rates than a single classifier. Therefore, it was necessary to design and develop two experiments that investigated this. The generic experiment used a single classifier (which is the most robust case) that utilized the dataset that contained samples from all five activities but with the activity label removed. The multi-algorithmic/classifier experiment evaluated by using the labelled activities. As illustrated in Fig. 5, the activity based multi-classifier outperforms the generic model (single classifier). The NW and FW activities achieved EERs of 0.93% and 3.90% (compared to 7.6% when the generic-based model was used). Similarly, the performance of the remaining activities (i.e. TPC, TMob, and MobG) outperformed the generic-based user authentication model with an average of EER of 5%.

Given the fact that a user can be doing multiple activities at the same time (e.g. sending a text or reading emails while walking), the single classifier approach could be used to classify all other activities that were not recognized by the proposed system. The experimental set up of the single classifier approach bears a close

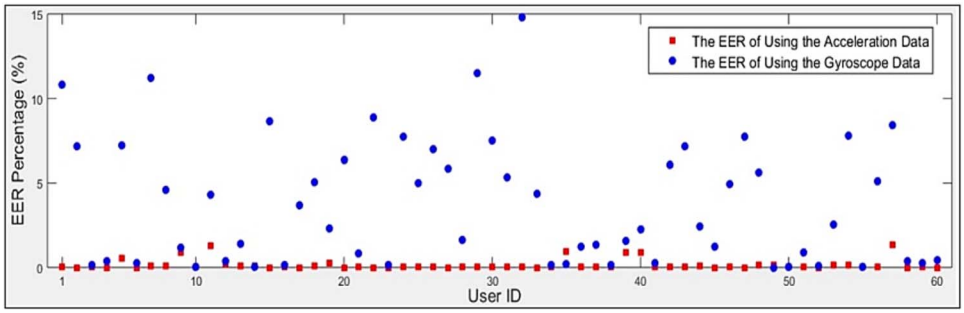

Figure 4. EER performance using accelerometer versus gyroscope sensors.

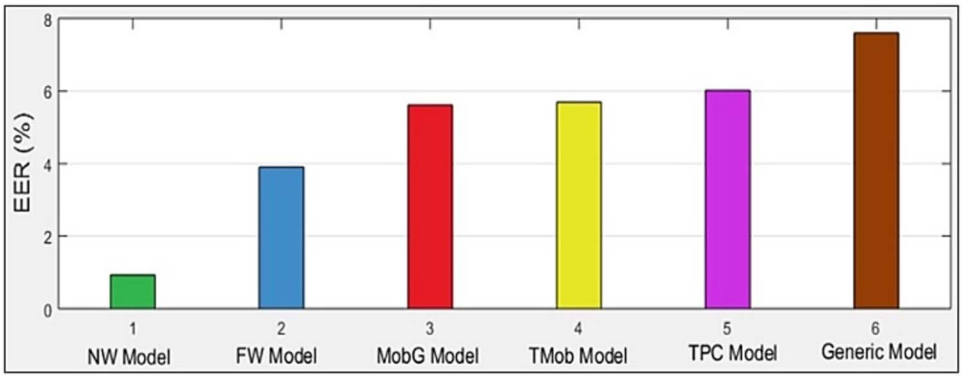

Figure 5. Performance comparison of generic classifier versus an activity-specific multi-classifier.

resemblance to the prior work by [80]. Nevertheless, they have reported poor accuracy (i.e. an EER of 19% compared to 7.1% in this study). The significant improvement could be the result of creating a complex and discriminative feature vector for each individual independently.

4.1.3. *The impact of feature selection on same/cross day methodologies*

So far, the analysis was based upon using all the extracted time domain features (i.e. 88 features). However, as mentioned earlier, it is important to optimize the user's authentication model by selecting the most discriminative feature subset to reduce the cost associated with classification and improve system accuracy/error rates. Another factor that greatly affects the authentication performance is the evaluation scenarios (i.e. SD and CD). It is well recognised that the human behaviour could be influenced by several factors; therefore, a more realistic test for sensor-based biometric system comes by applying the CD scenario to show the variation of the human behaviour overtime.

The CD test does not require the user to re-enrol in the system on a daily basis. Table 5 illustrates the benefit of utilizing the proposed feature selection approach and the impact of the evaluation scenarios (i.e. SD and CD) on the authentication performance.

The results in Table 5 show if the acceleration data is evaluated properly, the proposed authentication system is able to achieve a very high accuracy for both scenarios in comparison to the prior art that reported at best EERs of 0.2% [56] and 1.85% [39] for the SD and CD scenarios. This compares to EERs of 0.13% and 0.69% using the SD and CD tests respectively for the NW. The permanence of the system was evaluated by using the CD scenario with relatively low EERs for all activities (i.e. EERs ranging from 0.69% [for NW] to 5.81% [for TPC]). These results were obtained by using different feature set sizes ranging from 50 features (TPC activity) to 70 (FW activity).

As expected, the system performance decreases under the CD methodology; this is because the behavioural biometric can be affected by several factors such as mood, clothes, and tiredness. Nonetheless, the presented CD results are still promising and offer vital evidence that the collected activities of this study have the potential to accurately recognize the legitimate user in a transparent and continuous fashion. The optimal results of the CD scenario were achieved for the gait–based activities, at best EERs of 0.69% and 3.16% for the NW and FW accordingly. For the non-gait activities, the performances of TMob and MobG were slightly superior to TPC. This could be due to the position of the user's hand being not fixed during the TMob and MobG activities compared to TPC where the hand position was fairly static. Thus, more differential movement data can be observed from the typing or interacting on a smartphone touch screen.

With respect to the proposed feature reduction method, the findings in Table 5 have further strengthened the argument that an optimized set of features will provide the best performance. The lowest EERs were obtained by utilizing feature subset sizes between 50 to 70 features. The EERs were further analysed for each participant separately and results on each user's acceleration for both SD and CD scenario are presented in Fig. 6 (for the NW activity). As shown in Fig. 6, high levels of performance (i.e. in the range of 0%–2% EER) were obtained for 90% users, with the exception of users 31, 37, 38, 42, 48, and 51. More than 25% of the participants reported 0% of EERs such as users 2, 4, 13, 15, 17, 21, and 27. This suggests that users have consistent and distinctive feature patterns and optimising the feature vector length for each user independently is beneficial. For example, the reference template of *User 1* might contain 10 features while

Table 5. Results of the SD & CD performance with varying feature vectors

Evaluation Scenario	Activity	10 Features (%)	20 Features	30 Features (%)	40 Features (%)	50 Features (%)	60 Features (%)	70 Features (%)	80 Features (%)	88 Features (%)
SD	NW	1.13	0.78	0.24	026	0.27	0.13	0.20	0.16	0.15
SD	FW	1.55	0.80	0.62	0.36	0.35	0.32	0.28	0.32	0.31
SD	TMob	2.40	1.76	1.38	1.18	0.99	1.20	1.24	1.39	1.43
SD	TPC	2.28	1.36	1.38	1.15	1.15	1.30	1.39	1.33	1.52
SD	MobG	2.40	1.76	1.38	1.18	0.89	1.20	1.14	1.20	1.33
CD	NW	4.68	2.39	1.43	0.9	0.84	0.83	0.69	0.77	0.93
CD	FW	5.42	3.92	3.63	4.17	3.56	3.32	3.16	3.40	3.90
CD	TMob	5.97	5.92	5.93	5.69	5.04	4.94	5.57	5.60	5.69
CD	TPC	8.12	7.21	6.98	6.45	5.81	5.85	5.92	5.91	6.02
CD	MobG	4.97	4.82	4.83	4.79	4.62	4.54	5.17	5.80	5.61

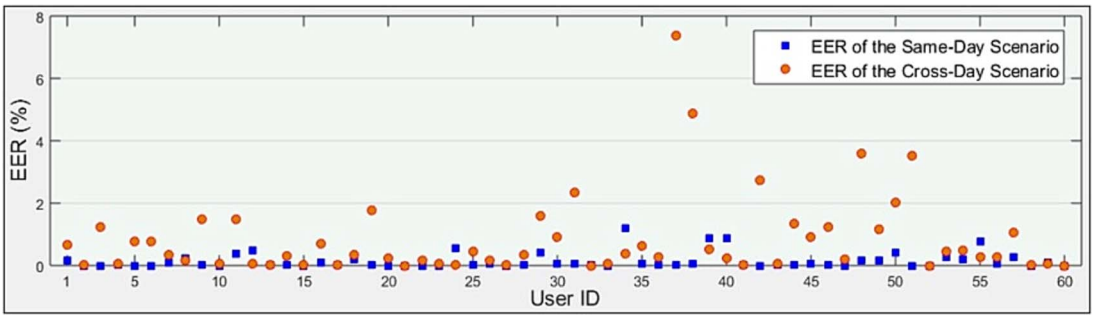

Figure 6. EER of normal walking using same-day/cross-day strategies.

Table 6. System performance using the static feature vector (SFV) and optimized feature vector (OFV)

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

User 2 optimally requires 50. The aim of the investigation was to determine whether the optimized feature vector length for each user independently can further improve the system accuracy. Table 6 shows a comparison of using static and optimized feature vector approaches and the impact upon feature length for each participant.

The findings in Table 6 confirm the hypothesis that creating optimized user vectors for each user independently could further reduce the EER (i.e. the proposed optimized feature vector clearly has an advantage over the static user template). An explanation for the significant improvement on the system accuracy is that the movement pattern of some users requires fewer features to produce the lowest EER and vice versa (i.e. the user's arm movement for users is inconsistent hence, more features are required to obtain the optimal or lowest EER). Fixing the size of the reference template for all users (e.g. 60 features) could negatively affect the overall system accuracy. To support the above assumption, further tests were undertaken for each activity and the evidence presented in Figs. 7 and 8. Apart from the improvement in the system performance of using the optimization technique, Fig. 7 shows the gait templates size was reduced for more than half of the users. For example, the FW-based model of users 3, 5, 8, 10, 24, 32, 37, 46, and 58 were created by utilizing only 20 prioritized features and even less features were used for users 1, 28, 31, 34, and 39 (10 features). In contrast, other users such as 4, 6, 9, 13, 16, and 22 required more features (i.e. 80 to 88 features) to produce the lowest EER. A possible explanation is that the walking pattern of these participants was varied or inconsistent over the time. Therefore, more features are required to generate a reference template that is robust to impersonation attacks and effectively identify the user's identity. Meanwhile, Fig. 8 shows a clear trend that the proposed optimized feature vector technique of the TMob activity successfully reduced the vector size for more than two thirds of users.

4.2. Real-life experiment

Although the findings of the controlled experiment were very compelling, they were, like the prior literature, obtained within a constrained environment (i.e. all users were asked to do the same type of activity). This approach, while standard in assessing the feasibility of a biometric in the early stages of research, is arguably not reflective of real-world use. Therefore, a more realistic investigation was conducted by collecting real life data to make sure the captured signals can be used for practical authentication system. To permit a comparison of the two models (i.e. generic and activity-based authentication models), two experiments were conducted; the first experiment utilized the generic model and

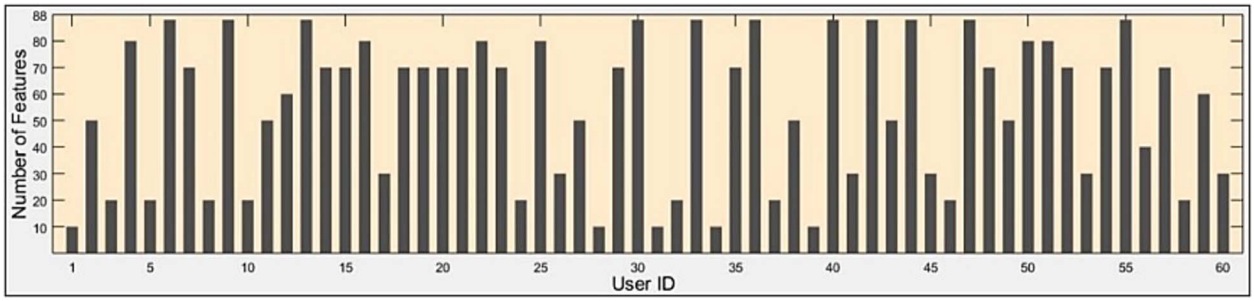

Figure 7. Optimal feature vector size per participant (FW activity).

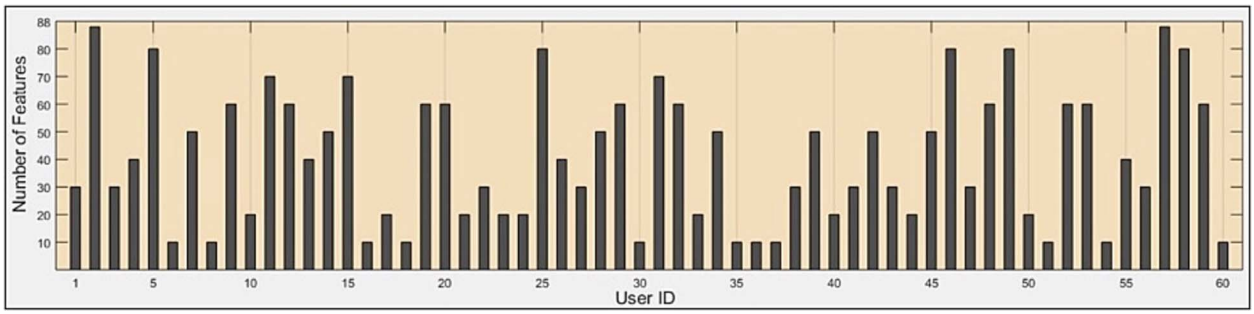

Figure 8. Optimal feature vector size per participant (TMob activity).

reported EERs of 24.54% and 26.11% for the accelerometer and gyroscope respectively. This preliminary result highlights the high degree of variability that exists within the real-life data versus the controlled SD and CD methodologies previously employed. Moreover, it shows the importance of developing an activity recognition approach that automatically detects the user's activity at a particular time in order to improve the recognition rate. The problem with real-life data is how to label the activities.

This study proposed a gait detection method that is based upon detecting the repetitive cycles of the user's walking pattern from the original signal and then classifying the user's identity. This was achieved by analysing the horizontal (x) acceleration signal of different users due to the higher variability compared to the vertical (y) and sideways (z) motions. Based on the observation, it was hypothesized that the detected cycles represent the user's gait pattern while the remaining data is considered as Non-W samples. Primarily, it was important to determine the start point of the actual walking pattern that was identified at ∼1.3 g-force [8] and thereafter detecting the repetitive peaks based upon the initial gait sample. The first minima was considered the start point of the first cycle, and the second local minima was treated as the end of the cycle. This procedure was repeated until all remaining minima were detected in the signal. The end point of one cycle was considered as a start point for the next cycle and so on. Once all gait cycles were extracted, samples were created by using the sliding window approach and the extracted gait samples were then divided into two activities (i.e. NW and FW). This was achieved by detecting the local maxima peaks for each segment and average the values. Segments that have high peak values were considered as FW samples while segments with low peak values reflect the NW data (see Fig. 9). For example, the magnitude range of the FW peaks were between 0 and 2, while it ranged from 0.4 to 1.2 for the NW peaks.

Having devised and applied the proposed detection method to the real-life signal, data was divided into NW, FW, and

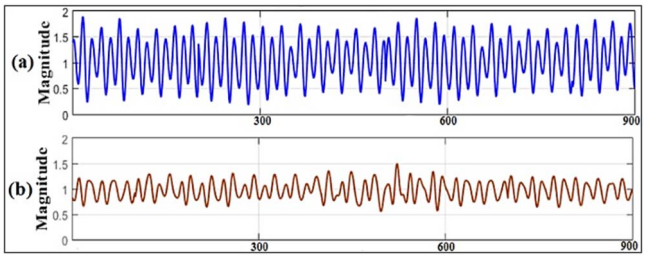

Figure 9. An example of filtering out the real life data for (a) fast walking and (b) normal walking.

Non-W data, and the total extracted samples per participant (over 10 days) for each activity are presented in Table 7.

To evaluate the proposed activity detection approach, real life data that was captured across several days and a fixed feature vector size for all users was applied. The findings are presented in Table 8. The results show that the proposed activity detection method has a significant positive impact on the authentication accuracy with the best EERs of 4.35%, 1.24%, and 7.04% for the NW, FW, and Non-W activities, respectively (compared to an EER of 24.54% by utilizing a generic-based authentication model for the accelerometer data). Although the gyroscope sensor was less effective than the accelerometer with EERs of 8.96%, 5.66%, and 11.25% for the aforementioned activities, these results are considered quite impressive in comparison with the accuracy of a generic-based user authentication model which achieved 26.11% of EER. Table 8 clearly shows that the proposed feature selection approach has an advantage over the verification performance versus all features given the reduction in computation that results in smaller feature vectors. For example, using the whole extracted feature set (i.e. 88 accelerometer features) reported 4.9%, 1.74%, and 8.74% for the NW, FW, and Non-W activities respectively, while the feature reduction method successfully reduced the EERs into 4.35%, 1.24% and 7.04%.

Table 7. Total samples of the real-life data separated by user

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

Table 8. EERs using accelerometer and gyroscope data

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

It is notable that the proposed feature selection approach does not contribute as much to the system accuracy as the controlled experiment exhibited, the best EERs were achieved by curtailing the number of features from 88 (i.e. all features) to 60, and 30 features (which is a reduction of nearly 32% of the gait features and ∼66% of the Non-W features). Although the findings in Table 8 suggest that the accelerometer sensor resulted in lower EERs (i.e. better performance) than the gyroscope for all activities, prior studies [70,81] have highlighted that sensor-based authentication systems might be susceptible to attacks if a single sensor is used. Moreover, other authors [27,82–84] suggested that the system accuracy might be improved by using fusion features of both sensors.

The fusion schema in biometric-based systems can be implemented at three different levels: sensor level, feature level, and score level. In the sensor level fusion, data of a single modality or multiple biometric traits are combined; e.g. capturing face samples from different cameras and different angles (in the case of a uni-biometric system) or collecting multi-biometric modalities such as face and voice. When it comes to the feature level fusion, features that are extracted from different sensors are fused to generate a resultant reference template. Finally, the score level is an approach that measures the similarity scores between the reference and test templates and combines the resultant scores of each modality together. This study investigated whether the fusion sensor approach can offer better recognition rates. As mentioned earlier, 88-time domain features were extracted for each sensor, so the fusion approach resulted in 176 features for accelerometer and gyroscope sensors (88 features $*$ 2 sensors). Table 9 displays the EERs from using the fusion characteristics of the acceleration and gyroscope signals for the NW, FW, and Non-W activities.

In comparison with the findings in Table 8 of using single sensor data (i.e. accelerometer or gyroscope), Table 9 shows that the authentication performance is improved for all activities (i.e. NW, FW, and Non-W activities). This positive effect from using the fusion approach is more noticeable if it is compared to the EERs obtained by using the gyroscope data (i.e. 8.96%, 5.66%, and 11.25%). The presented results in Tables 8 and 9 were based upon creating a dynamic reference template for each user with the feature vector size fixed for all users; e.g. the best EER for the NW activity was 3.55% by using 110 features for each individual. However, optimizing the feature vector length per user could be useful to maximize the system performance. For instance, some users might require few features to accurately recognize their pattern, while increasing the feature size may offer better accuracy/error rates for other users. Therefore, further analysis was undertaken to explore whether the creation of an optimized feature template for each user independently can further reduce the EER. Table 10 displays the best EERs for activities using static and dynamic feature vector size.

In comparison with the static feature vector, it can be seen in Table 10 that the optimized reference template outperformed the performance of the gait activities (NW and FW) for all the source information (i.e. single sensor and fusion data). For the Non-W activity, there was little difference in the findings between the two approaches (i.e. static and optimized feature vector). The possible explanation for this outcome could be the reference template for the majority of users was nearly optimized (i.e. the best EERs for the Non-W data were obtained by using a small feature subset

Table 9. EER using fusion sensor and CD scenario

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

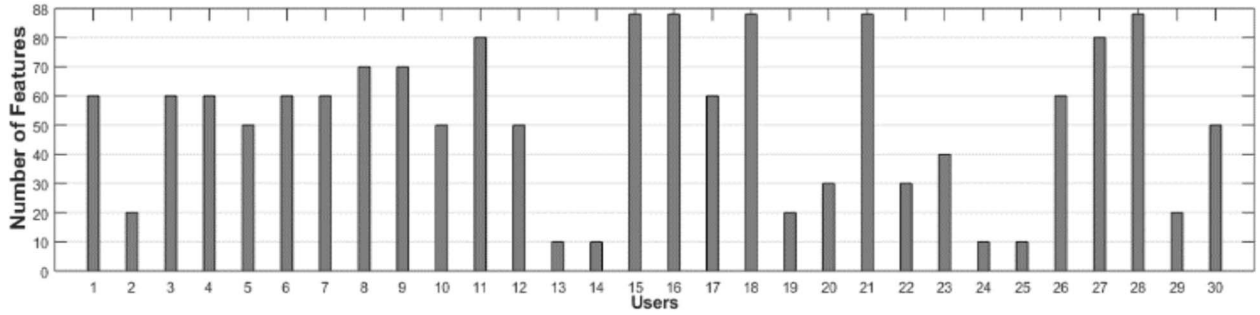

Figure 10. Optimised feature vector for each participant (FW activity).

Table 10. System performance using the SFV and OFV size

Activity	Sensor	EER (%) of SFV	EER (%) of OFV
NW	Accelerometer	4.35	3.83
	Gyroscope	8.96	8.10
	Fusion	3.55	2.89
FW	Accelerometer	1.24	0.73
	Gyroscope	5.66	5.15
	Fusion	0.92	0.70
Non-W	Accelerometer	7.04	6.51
	Gyroscope	11.25	10.47
	Fusion	5.31	4.77

such as 30 and 20 for the acceleration and gyroscope signals respectively as shown in Table 8). Apart from the improvement in system performance using the optimized feature vector, Fig. 10 shows that the user's reference template size was decreased for half of the users. For example, the feature vector of users 2, 19, and 29 was created by utilizing only 20 prioritized features with fewer features required for users 13, 14, 24, 25 (i.e. 10 features). In contrast, other users such as 8, 9, 11, 8, 15, 16, and 18 required more features (i.e. 70 to 88 features) to produce a low EER.

5. DISCUSSION

This study aimed to address the following questions:

- Can smartwatches provide a more reliable and consistent motion signal than smartphones?
- How well can smartwatch-based user authentication perform in the CD scenario testing?
- Does the creation of an activity-based authentication model have a positive impact on the classification performance?
- Can smartwatch-based user authentication be achieved with real-life environment?
- What is the effect of the proposed feature selection method on the biometric performance?

The results suggest that smartwatches have the ability to capture more accurate personal data than smartphones. This was reflected on the system performance where the best achieved EERs in this study were 0.29% and 0.70% (for the controlled and uncontrolled experiments, respectively), compared to the prior art that used smartphone accelerometer sensor and reported EERs between 4.93% and 33.3% [16,27]. The experimental analysis of this study reveals that smartwatch-based user authentication is highly efficient and can be recommended for use in verifying users in a transparent and continuous manner. Moreover, it highlights the effectiveness of wearables in recognizing a wide variety of activities.

To show the influence of the evaluation scenario on the system performance, comprehensive tests were carried out (i.e. SD and CD scenarios). The SD evaluation scenario reported low EERs for all activities (i.e. 0.05%, 0.14%, 0.5%, 0.3%, and 0.25% for the NW, FW, TMob, TPC, MobG, respectively). Improvements on performance have been presented when comparing the results to the prior art on mobile/smartwatch-based acceleration studies under the SD scenario that achieved EERs of 17.6% in the worst case [32] and as low as 0.65% [34]. A common criticism of these studies is the methodology they have adopted with training and test data being captured during a single session, which does not show the variability of the user's movement pattern over time, and limited or unrealistic activities being collected, such as punch gesture or drawing a circle. In addition to the SD, the more realistic CD test was also applied; as expected, the results demonstrated that biometric performance was degraded. Nevertheless, the proposed system is still able to achieve recognition performances typically better than most behavioural-based biometrics. Using the CD scenario in the controlled environment resulted in EERs of 0.29%, 1.31%, 2.66%, 3.85%, and 2.3% for the NW, FW, TMob, TPC, and MobG, respectively (compared to EERs in the range of 4%–33.3% for gait activities [16,33] and 3.3%–6.56% for gesture pattern [44,76].

The effectiveness of creating an activity-based authentication model was also investigated. Although the reported error rate was acceptable for a TAS, it is argued that the system performance

would degrade when using real life data due to the variability of the collected signal. As hypothesized, creating an activity-based authentication model, each trained and tested using CD against a specific activity, greatly enhanced the system performance. For example, the EERs were 0.29%, 1.31% and 2.3% for the NW, FW, and MobG, respectively, compared to 7.03% of EER when the generic authentication model was used. The proposed detection method is capable of identifying incoming activity, extracting features, and training the classifier with the corresponding template. When a particular activity is not recognized or detected, the generic classifier model can be used to accumulate and classify all other activities that were not recognized by the multi-classifier approach. All samples can be stored and at regular intervals an activity detection algorithm can be utilised to identify and group signals into new activities to create the activity-specific classifiers. This process will permit the addition of new activities to be incorporated. Further research is required to understand and optimise the parameters and to ensure signals that are captured and subsequently used to train the classifier are indeed from the legitimate user.

To make sure that the proposed smartwatch-based user authentication system can be used in practice, real life data was collected, and an activity recognition method was developed to divide the user's movement data into gait and non-gait activities. To show how the proposed activity detection affects the authentication accuracy, two experiments were carried out. The first experiment used the whole data without identifying the activity type (i.e. a generic-based model), while the proposed activity method was utilized to divide the real-life data automatically in the second experiment. As expected, high EERs were reported (i.e. 24.54% and 26.11% for accelerometer and gyroscope, respectively) by using the generic-based model due to the nature of the real-life data. Arguably a level of performance that would question the viability of the approach in practice. However, significant improvements were achieved (i.e. at best EERs of 1.24% for accelerometer and 5.66% for gyroscope) when data was split into groups based upon the performed activity. Developing a context aware approach might give better understanding to the user's daily activities (rather than just dividing the data into walking and Non-W activities). This can be achieved by obtaining information from other smartwatch sensors (e.g. GPS, and Ambient temperature) that could be used as a basis for making a more intelligent decision and improving the system accuracy. Further improvements on the recognition rates were obtained by combining the accelerometer and gyroscope data. Apart from the improvement to system accuracy, the fusion-level feature of both sensors is useful to detect and prevent imitation or mimicry attacks.

Further influencing factors on the biometric performance include the selected feature subset. Selecting unique features for each user would improve the results and reduce the complex computations on devices which have limited processing resources when compared to computers. This study presented and evaluated a feature selection method for activity-based authentication, which was based on creating a dynamic feature vector for each user and successfully reducing the dimensionality of the feature space. However, further investigation is required to find out if the user's reference template needs to be updated over time and how best to achieve this. Again, it is likely that the time between template renewals is likely to be dependent upon the user, leading to a dynamic rather than static template renewal process needing to be devised.

Transparent and continuous authentication offers the middle-ground where both security and usability can be accomplished. Nevertheless, it is essential to investigate the implementation cost of applying such a system (i.e. consumption of storage, CPU, and battery life). During the real-life data collection, there was a noticeable impact on battery life due to the continuous collection from all the sensors (e.g. accelerometer, gyroscope, hear rate, and temperature). A commercial solution would need to carefully consider which sensors to utilize and when to utilise them. For example, this study for investigative reasons captured all sensor data, but in reality only used the accelerometer and gyroscope. Careful scheduling of these sensors to coincide with authentication requirements could significantly reduce the demand—reducing battery, memory and CPU load. The study also focused upon a single Smartwatch technology and it would be expected that different Smartwatches would have different capabilities and sensor granularities which might vary the level of recognition performance that can be achieved. However, it is also worth highlighting, the purpose of this approach is to contribute towards a TAS as one of several biometric modalities, with the wider TAS providing the management and intelligence to know what samples to use to ensure a sufficient overall confidence in the user's identity [1]. As such, issues with regards to signal latency are less relevant as the wider TAS management system will manage this.

6. CONCLUSIONS & FUTURE WORK

The investigation presented in this study has positively demonstrated that activity-based user authentication using smartwatches is a feasible approach in achieving reliable and transparent authentication. It also examined the effect of using the CD scenario on the system performance and presents a novel feature selection approach that effectively reduced the feature vector size without overtly affecting performance. Experimental results demonstrate the advantage of creating activity-based authentication models in order to enhance system performance. Indeed, results showed that real-life activity-based user authentication is only viable using an activity-based multi-classifier.

Future work will focus on better optimization such as extracting new features, evaluating different machine learning classifiers (e.g. Random Forest, Naive Bayes, and SVM) and combining the smartphone and smartwatch movement data. More experimental research should also be carried out to understand how the user templates might be changed over time and to make sure that the template will always be appropriate to identify the legitimate user versus imposters. Implementing the proposed system in a real-world scenario and capturing real-life data over a longitudinal basis is required to find out what change might exist in the user's behaviour. Additionally, whilst this study was able to identify the real-life data in gait and non-gait data, a context aware approach could be useful to predict a wider variety of activities, with an aim of improving the recognition rates. For example, using GPS and the calendar application, it would be possible to identify not only that an individual is running but that they are running to catch a train to the airport—thus likely to be carrying or pulling luggage. In this scenario, a composite classifier could be used that not only focusses upon running but running and carrying luggage. This can be achieved by incorporating other sensor-based information (e.g. GPS) to provide some situational awareness of what a user might be doing at a specific point in time.

Future research will also consider the use of a wider range of smartwatch technologies and the location for processing the authentication decisions. Currently this is focused upon the Smartphone, but with advances in Smartwatch technology, it would be relevant to investigate the degree to which these devices could manage the process—with the advantage of aligning authentication more closely to the individual and opening further opportunities for developing frictionless authentication across a wider set of technologies.

SUPPLEMENTARY DATA

Supplementary data are available at *The Computer Journal* online.

FUNDING

This research was financially supported by the University of Kufa and the Government of Iraq.

DATA AVAILABILITY

The data underlying this article cannot be shared publicly due to constraints of the ethical approval given to the study by the University of Plymouth. *The data will be shared on reasonable request to the corresponding author.*

References

1. Clarke, N. (2011) *Transparent User Authentication: Biometrics, RFID, and Behavioural Profiling*. Springer, London.
2. Saevanee, H., Clarke, N., Furnell, S. *et al.* (2015) Continuous user authentication using multi-modal biometrics. *Comput Secur*, **53**, pp. 234–246. https://doi.org/10.1016/j.cose.2015.06.001.
3. Alotaibi SN, Furnell S, Clarke N. A novel transparent user authentication approach for mobile applications. *Information Security Journal: A Global Perspective* 2018;**27**(5–6):292–305. https://doi.org/10.1080/19393555.2019.1609628.
4. Krishnamoorthy S, Rueda L, Saad S. *et al.* Identification of user Behavioral biometrics for authentication using keystroke dynamics and machine learning. In: *Proceedings of the 2018, 2nd International Conference on Biometric Engineering and Applications, Amsterdam Netherlands, May 16–18*, 2018, 50–7.
5. Ryu G, Park S, Choi D. *et al.* Active authentication experiments using actual application usage log. In: *Proceedings of the First Workshop on Radical and Experiential Security, Incheon Republic of Korea, 4 June*, 2018, 9–16.
6. Kang, S., Lee, J., Bong, K. *et al.* (2018) Low-power scalable 3-D face Frontalization processor for CNN-based face recognition in mobile devices. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, **8**, 873–883. https://doi.org/10.1109/JETCAS.2018.2845663.
7. Chen Q. Methods of automatically answering a phone call with a mobile terminal and associated mobile terminals. 2017; US 9, 894, 522 B2.
8. Mahbub U, Sarkar S, Patel VM. *et al.* Active user authentication for smartphones: A challenge data set and benchmark results. In: *Proceedings of IEEE 8th international conference on biometrics theory, applications, and systems (BTAS), Niagara Falls, NY, USA ,06–09 September*, 2016, pp. 1–8. IEEE.
9. AI-Naffakh N, Clarke N, Dowland P. *et al.* Activity recognition using wearable computing. In: *Proceedings of the 11th International Conference for Internet Technology and Secured Transactions (ICITST), Barcelona, 05–07 December*, 2016, pp. 189–95. IEEE.
10. Derawi MO, Nickel C, Bours P. *et al.* Unobtrusive user-authentication on mobile phones using biometric gait recognition. In: *Proceedings in Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Darmstadt, Germany,15–17 October*, 2010, pp. 306–11. IEEE.
11. Gafurov D, Snekkenes E, Bours P. Gait authentication and identification using wearable accelerometer sensor. In: *Proc. IEEE Workshop on Automatic Identification Advanced Technologies, Alghero, Italy, 07–08 June*, 2007, pp. 220–5. IEEE.
12. Nickel, C. Derawi, M.O. Bours, P. *et al.* (2011) Scenario test of accelerometer-based biometric gait recognition. In: *Proceedings in the Third International Workshop on Security and Communication Networks (IWSCN), Gjovik, Norway, 18–20 May*, pp. 15–21. IEEE.
13. Nickel C, Wirtl T, Busch C. Authentication of smartphone users based on the way they walk using k-NN algorithm. In: *Proceedings in IIH-MSP Conference, Piraeus-Athens, Greece, 18–20 July*, 2012, 16–20. IEEE.
14. Muaaz M, Nickel C. Influence of different walking speeds and surfaces on accelerometer-based biometric gait recognition. In: *Proceeding in 35th International Conference on Telecommunications and Signal Processing (TSP), Prague, Czech Republic, 03–04 July*, 2012, 508–12. IEEE.
15. Nickel C, Busch C. Classifying accelerometer data via hidden Markov models to authenticate people by the way they walk. In: *Proceeding in Carnahan Conference on Security Technology, Barcelona, Spain*, 2011, 29–35. IEEE.
16. Muaaz M, Mayrhofer R. An analysis of different approaches to gait recognition using cell phone based accelerometers. In: *Proceedings of International Conference on Advances in Mobile Computing & Multimedia, Vienna Austria, 2–4 December*, 2013, 293–300. ACM.
17. Shrestha, B., Saxena, N. and Harrison, J. (2013) Wave-to-access: Protecting sensitive mobile device services via a hand waving gesture. In Abdalla, M., Nita-Rotaru, C., Dahab, R. (eds) *Cryptology and Network Security. CANS 2013. Lecture Notes in Computer Science* (Vol. **8257**), pp. 199–217. Springer International Publishing.
18. Hoang, T., Nguyen, T., Luong, C., Do, S. and Choi, D. (2013) Adaptive cross-device gait recognition using a mobile accelerometer. *Journal of Information Processing Systems*, **9**, 333–348. https://doi.org/10.3745/JIPS.2013.9.2.333.
19. Gascon H, Uellenbeck S, Wolf C. *et al.* Continuous authentication on mobile devices by analysis of typing motion behavior. In: *presented at Proc. GI Conf. Sicherheit (Sicherheit, Schutz und Verlsslichkeit), Wien, Österreich*, 2014, 1–12. Gesellschaft für Informatik e.V.
20. Damaševičius, R., Maskeliūnas, R., Venčkauskas, A. *et al.* (2016) Smartphone user identity verification using gait characteristics. *Symmetry*, **8**, 100. https://doi.org/10.3390/sym8100100.
21. Muaaz M, Mayrhofer R. Accelerometer based gait recognition using adapted Gaussian mixture models. In: *Proceedings of the 14th International Conference on Advances in Mobile Computing and Multi Media, November 28–30, Singapore Singapore*, 2016, 288–91. ACM.
22. Kumar R, Phoha VV, Serwadda A. Continuous authentication of smartphone users by fusing typing, swiping, and phone movement patterns. In: *Proceeding in IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS), Niagara Falls, NY, USA, 06–09 September*, 2016. IEEE.
23. Kumar R, Kundu P, Shukla D. *et al.* Continuous user authentication via Unlabeled phone movement patterns. In: *Proceeding in IEEE International Joint Conference on Biometrics (IJCB), October 01–04, Denver, CO, USA*, 2017. IEEE.
24. Ehatisham-ul-Haq, M., Ehatisham-ul-Haq, M., Azam, M. *et al.* (2017) Identifying smartphone users based on their activity

patterns via mobile sensing. *Procedia Computer Science*, **113**, 202–209. https://doi.org/10.1016/j.procs.2017.08.349.
25. Ehatisham-ul-Haq, M., Azam, M., Loo, J. *et al.* (2017) Authentication of smartphone users based on activity recognition and mobile sensing. *Sensors (Basel)*, **17**, 2043. https://doi.org/10.3390/s17092043.
26. Shen, C., Li, Y., Chen, Y. *et al.* (2017) Performance analysis of multi-motion sensor behavior for active smartphone authentication. *IEEE Trans Inf Forensics Secur*, **13**, 48–62.
27. Lee W-H, Liu X, Shen Y. *et al.* Secure pick up: Implicit authentication when you start using the smartphone. In: *Proceeding in the 22nd ACM Symposium on Access Control Models and Technologies, Indianapolis Indiana USA, June 21–23*, 2017, 67–78. ACM.
28. Middya, A.I., Roy, S. and Mandal, S. (2022) User recognition in participatory sensing systems using deep learning based on spectro-temporal representation of accelerometer signals. *Knowledge-Based Systems*, **258**, 110046. https://doi.org/10.1016/j.knosys.2022.110046.
29. Alawneh, L., Al-Zinati, M. and Al-Ayyoub, M. (2023) User identification using deep learning and human activity mobile sensor data. *International Journal of Information Security*, **22**, 289–301. https://doi.org/10.1007/s10207-022-00640-4.
30. Wang Y, Zhang X, Hu H. (2023) Continuous user authentication on multiple smart devices. *Information*, **14**(5), p.274, https://doi.org/10.3390/info14050274.
31. Mahadeen, E., Alghamdi, M., Tarawneh, A.S. *et al.* (2023) Smartphone user identification using accelerometer data. *MDPI*, **15**, 10456. https://doi.org/10.3390/su151310456.
32. Findling R, Muaaz M, Hintze D. *et al.* ShakeUnlock: Securely unlock mobile devices by shaking them together. In: *Proceedings of MoMM, Osaka Japan, 7-11September*, 2014, 165–74. ACM.
33. Babins S, Manar M, Nitesh S. Walk-unlock: zero-interaction authentication protected with multi-modal gait biometrics. 2016; arXiv preprint arXiv:1605.00766.
34. Dong J, Cai Z. User authentication using motion sensor data from both wearables and smartphone. In: *the Chinese Conference on Biometric Recognition*, 2016, 756–64. Springer International Publishing.
35. Luo, F., Khan, S., Huang, Y. *et al.* (2022) Activity-based person identification using multimodal wearable sensor data. *IEEE Internet Things J*, **10**, 1711–1723.
36. Verma A, Moghaddam V, Anwar A. (2022) Data-driven behavioural biometrics for continuous and adaptive user verification using smartphone and smartwatch. *Sustainability*, **14**(12), 7362, https://doi.org/10.3390/su14127362
37. Watanabe K, Nagatomo M, Aburada K. *et al.* Gait-based authentication for smart locks using accelerometers in two devices. In: *Advances in Networked-based Information Systems: The 22nd International Conference on Network-Based Information Systems*. Springer International Publishing, 2020, 281–91.
38. Lee W, Lee R. Implicit sensor-based authentication of smartphone users with smartwatch. In: *Proceedings of the Hardware and Architectural Support for Security and Privacy on – HASP, Seoul Republic of Korea, 18 June*, 2016, 1–8. ACM.
39. Lee WH, Lee R. Implicit smartphone user authentication with sensors and contextual machine learning. In: *Proceedings of the 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), Denver, CO, USA, 26–29 Jun*, 2017, 297–308. IEEE.
40. Weiss, G.M., Yoneda, K. and Hayajneh, T. (2019) Smartphone and smartwatch-based biometrics using activities of daily living. *IEEE Access*, **7**, 133190–133202. https://doi.org/10.1109/ACCESS.2019.2940729.
41. Cola G, Avvenuti M, Musso F. Gait-based authentication using a wrist-worn device. In: *Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services. Hiroshima, Japan, 28Nov-1Dec*, 2016, 208–17. ACM.
42. Lewis A, Li Y, Xie M. Real time motion-based authentication for smartwatch. In: *the IEEE International Conference on Communications and Network Security (CNS), Philadelphia, PA, USA, 17–19 October*, 2016, 380–1. IEEE.
43. Davidson S, Smith D, Yang C. *et al.* (2016) Smartwatch user identification as a means of authentication, *Department of Computer Science and Engineering Std*, Vol **21**(12), pp. 1196.
44. Griswold-Steiner I, Matovu R, Serwadda A. HandwritingWatcher: A mechanism for smartwatch-driven handwriting authentication. In: *Proceeding in the IEEE International Joint Conference on Biometrics (IJCB), USA, 1–4 October*, 2017, 216–24. IEEE.
45. Liang G, Xu X, YU J. User-authentication on wearable devices based on punch gesture biometrics. In: *the International Conference on Information Science and Technology, Shanghai, China 23 May*, 2017, EDP Sciecnes.
46. Wang Z, Shen C. Handwaving authentication: Unlocking your smartwatch through Handwaving biometrics. In: *Proceeding in the Chinese Conference on Biometric Recognition, China, 28–29 October*, 2017, 545–53. Springer International Publishing.
47. Xu W, Shen Y, Zhang Y. *et al.* Gait-watch: A context-aware authentication system for smart watch based on gait recognition. In: *Proceeding in 2nd Int. Conf. Internet Things Design Implement, Pittsburgh, PA, USA, 18 – 21 April*, 2017, 59–70. ACM.
48. Ahmad, M., Alqarni, M., Khan, A. *et al.* (2018) Smartwatch-based legitimate user identification for cloud-based secure services. *Mobile Information Systems*, **2018**, 1–14. https://doi.org/10.1155/2018/5107024.
49. Acar AH, Aksu H, Uluagac AS. *et al.* Waca: Wearable-assisted continuous authentication. In: *Proceeding in Security and Privacy Workshops (SPW), San Francisco, CA, USA, 24–24 May*, 2018, 264–9. IEEE.
50. Acar, A., Aksu, H., Uluagac, A.S. *et al.* (2020) A usable and robust continuous authentication framework using wearables. *IEEE Trans Mob Comput*, **20**, **20**, 2140–2153. https://doi.org/10.1109/TMC.2020.2974941.
51. Zhao, Y., Zhao, Y., Tu, H. *et al.* (2022) Motion gesture delimiters for smartwatch interaction. *Wireless Communications and Mobile Computing*, **2022**, 1–11. https://doi.org/10.1155/2022/6879206.
52. Lopez-Rodriguez, P., Avina-Cervantes, J.G., Contreras-Hernandez, J.L. *et al.* (2022) Handwriting recognition based on 3d accelerometer data by deep learning. *Applied Sciences*, **12**, 6707. https://doi.org/10.3390/app12136707.
53. Vecchio A, Nocerino R, Cola G. Gait-based authentication: Evaluation of energy consumption on commercial devices. In: *Proceeding in 2022 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops), Pisa, Italy, 21–25 March*, 2022, 793–8. IEEE.
54. Hernández Acosta L, Rahe S, Reinhardt D. Does cycling reveal insights about You? Investigation of user and environmental characteristics during cycling. In: *Proceeding In International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services, Switzerland, 14–17 Nov.*, 2022, 172–90. Springer International Publishing.
55. Lee, J., Park, S., Kim, Y.G. *et al.* (2021) Advanced authentication method by geometric data analysis based on user behavior and biometrics for iot device with touchscreen. *Electronics*, **10**, 2583. https://doi.org/10.3390/electronics10212583.
56. Wijewickrama R, Maiti A, Jadliwala M. Write to know: On the feasibility of wrist motion-based user-authentication from

handwriting. In: *Proceedings of the 14th ACM Conference on Security and Privacy in Wireless and Mobile Networks, Abu Dhabi, United Arab Emirates, June 28–July 2, 2021*, 2021, 335–46. ACM.
57. Maes JG, Rahman KA, Mukherjee A. Hybrid smartwatch multi-factor authentication. In: *2023 IEEE 17th International Symposium on Medical Information and Communication Technology (ISMICT), Lincoln, NE, USA, 10–12 May*, 2023, 1–6. IEEE.
58. Khan S, Shah SMA, Arsalan A. *et al.* User recognition based on gait pattern via smartwatch accelerometer in unrestricted environment. In: *Proceeding in 2023 International Conference on Communication, Computing and Digital Systems (C-CODE), Islamabad, Pakistan, 17–18 May*, 2023, 1–6. IEEE.
59. Guerar M, Migliardi M, Palmieri F. *et al.* (2020) Securing PIN-based authentication in smartwatches with just two gestures. *Concurrency and Computation: Practice and Experience*, Vol **32**(18), p 5549, https://doi.org/10.1002/cpe.5549.
60. Buriro A, Fred K. DeepClap: A gesture-based user authentication scheme with deep neural networks. In: *Proc. Twenty Eighth ACM Symposium on Access Control Models and Technologies (SACMAT)*, 2023, 1–4. ACM.
61. Li G, Sato H. Handwritten signature authentication using smartwatch motion sensors. In: *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC), Madrid, Spain, 13–17 July*, 2020, 1589–96. IEEE.
62. Rahman KA, Alam N, Musarrat J. *et al.* Smartwatch dynamics: A novel modality and solution to attacks on cyber-behavioral biometrics for continuous verification? In: *Proceeding In 2020 International Symposium on Networks, Computers and Communications (ISNCC) Montreal, QC, Canada, 20–22 October*, 2020, 1–5. IEEE.
63. Li, G. and Sato, H. (2022) Sensing in-air signature motions using smartwatch: a high-precision approach of behavioral authentication. *IEEE Access*, **10**, 57865–57879. https://doi.org/10.1109/ACCESS.2022.3177905.
64. Zhao, Y., Gao, R. and Tu, H. (2020) Smartwatch user authentication based on the arm-raising gesture. *Interacting with Computers*, **32**, 569–580. https://doi.org/10.1093/iwcomp/iwab013.
65. Mekruksavanich, S. and Jitpattanakul, A. (2022) Deep residual network for smartwatch-based user identification through complex hand movements. *Sensors*, **22**, 3094. https://doi.org/10.3390/s22083094.
66. Yu X, Zhou Z, Xu M. *et al.* Thumbup: Identification and authentication by smartwatch using simple hand gestures. In: *Proceeding in 2020 IEEE International Conference on Pervasive Computing and Communications (PerCom), Austin, TX, USA, 23–27 March*, 2020, 1–10. IEEE.
67. Ramachandra R, Venkatesh S, Raja K, *et al.* (2021) Handwritten signature and text-based user verification using smartwatch. *Proceeding in 2020 25th International Conference on Pattern Recognition (ICPR) Milan, Italy, Milan, Italy*, pp. 5099–5106. IEEE.
68. Xu, W., Shen, Y., Luo, C. *et al.* (2020) Gait-watch: a gait-based context-aware authentication system for smart watch via sparse coding. *Ad Hoc Networks*, **107**, 102218. https://doi.org/10.1016/j.adhoc.2020.102218.
69. Li G, Zhang L, Sato H. In-air signature authentication using smartwatch motion sensors. In: *Proceeding in 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC), Madrid, Spain, 12–16 July*, 2021, 386–95. IEEE.
70. Coskun D, Incel O, Ozgovde A. Phone position/placement detection using accelerometer: Impact on activity recognition. In: *Proc. IEEE 10th Int. Conf. Intell. Sensors, Sensor Netw. Inf. Process. (ISSNIP), Singapore, 07–09 April*, 2015, 1–6. IEEE.
71. Kumar R, Phoha V, Raina R. Authenticating users through their arm movement patterns. 2016; arXiv preprint arXiv:1603.02211.
72. PIMP (2017) *Companion for Microsoft Band*Version 2.9.4. Google Commerce Ltd Publisher Google Play Store.
73. Al-Naffakh, N., Clarke, N., Haskell-Dowland, P. *et al.* (2016) A comprehensive evaluation of feature selection for gait recognition using smartwatches. *International Journal for Information Security Research (IJISR)*, **6**, 1–10. https://doi.org/10.20533/ijisr.2042.4639.2016.0080.
74. Al-Naffakh N, Clarke N, Li F. *et al.* Unobtrusive gait recognition using smartwatches. In: *Proceeding in the International Conference of the Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany, 20–22 September*, 2017, 1–8. IEEE.
75. Al-Naffakh N, Clarke N, Li F. Continuous user authentication using smartwatch motion sensor data. In: Gal-Oz N, Lewis P (eds). *Trust Management XII. IFIPTM. IFIP Advances in Information and Communication Technology, Canada,10-13July*. Cham: Springer, 2018, 15–28.
76. Yang J, Li Y, Xie M. MotionAuth: Motion-based authentication for wrist worn smart devices. In: *the IEEE International Conference on Pervasive Computing and Communication Workshops*, 2015, 550–5. IEEE.
77. Nickel C, Brandt H, Busch C. Benchmarking the performance of SVMs and HMMs for accelerometer-based biometric gait recognition. In: *Proceeding in the IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Bilbao, Spain, 14–17 December*, 2011, 281–6. IEEE.
78. Weiss G, Timko J, Gallagher C. *et al.* Smartwatch-based activity recognition: A machine learning approach. In: *Proceeding in IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), Las Vegas, NV, USA, 24–27 February*, 2016, 426–9. IEEE.
79. Johnston A, Weiss G. Smartwatch-based biometric gait recognition. In: *Biometrics Theory, Applications and Systems (BTAS), Arlington, VA, USA*, 2015. IEEE.
80. Kwapisz JR, Weiss GM, Moore SA. (2011) Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter*, Vol **12**(2), pp. 74–82. https://doi.org/10.1145/1964897.1964918.
81. Mare S, Molina-Markham A, Cornelius C. *et al.* ZEBRA: Zero-effort bilateral recurring authentication. In: *Proceeding in Security and Privacy (SP), Berkeley, CA, USA, 18–21 May*, 2014, 705–20. IEEE.
82. Nickel C, Brandt H, Busch C. Classification of acceleration data for biometric gait recognition on mobile devices. In: *Proceedings of the Biometrics Special Interest Group, Darmstadt, Germany, 08–09 September*, 2011, 57–66. Gesellschaft für Informatik, Bonn, Germany.
83. Hestbek, M.R., Nickel, C. and Busch, C. (2012) Biometric gait recognition for mobile devices using wavelet transform and support vector machines. *Comput Secur*, **113**, 205–210.
84. Kumar R, Phoha VV, Jain A. Treadmill attack on gait-based authentication systems. In: *Proceeding in 2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS), Arlington, VA, USA, 08–11 September*, 2015, 1–7. IEEE.

The Computer Journal, 2025, 1–17
https://doi.org/10.1093/comjnl/bxae144
Original Article