Smart Health 19 (2021) 100162

Contents lists available at ScienceDirect

Smart Health

journal homepage: www.elsevier.com/locate/smhl

GaitCode: Gait-based continuous authentication using multimodal learning and wearable sensors

Ioannis Papavasileiou [a,c,1], Zhi Qiao [b,1], Chenyu Zhang [a], Wenlong Zhang [b,*], Jinbo Bi [a], Song Han [a,*]

[a] Department of Computer Science & Engrineering, University of Connecticut, USA
[b] The Polytechnic School, Ira A. Fulton Schools of Engineering, Arizona State University, USA
[c] Manufacturing Technology & Engineering, Corning Incorporated, Corning, NY, USA

ARTICLE INFO

Keywords:
Biometric authentication
Gait authentication
Autoencoders
Sensor fusion
Multimodal learning
Wearable sensors

ABSTRACT

The ever-growing threats of security and privacy loss from unauthorized access to mobile devices have led to the development of various biometric authentication methods for easier and safer data access. Gait-based authentication is a popular biometric authentication as it utilizes the unique patterns of human locomotion and it requires little cooperation from the user. Existing gait-based biometric authentication methods however suffer from degraded performance when using mobile devices such as smart phones as the sensing device, due to multiple reasons, such as increased accelerometer noise, sensor orientation and positioning, and noise from body movements not related to gait. To address these drawbacks, some researchers have adopted methods that fuse information from multiple accelerometer sensors mounted on the human body at different locations. In this work we present a novel gait-based continuous authentication method by applying multimodal learning on jointly recorded accelerometer and ground contact force data from smart wearable devices. Gait cycles are extracted as a basic authentication element, that can continuously authenticate a user. We use a network of auto-encoders with early or late sensor fusion for feature extraction and SVM and softmax for classification. The effectiveness of the proposed approach has been demonstrated through extensive experiments on datasets collected from two case studies, one with commercial off-the-shelf smart socks and the other with a medical-grade research prototype of smart shoes. The evaluation shows that the proposed approach can achieve a very low Equal Error Rate of 0.01% and 0.16% for identification with smart socks and smart shoes respectively, and a False Acceptance Rate of 0.54%–1.96% for leave-one-out authentication.

1. Introduction

Passwords and keys allow users to access their personal information, while protecting against unauthorized attempts. However, studies have shown that users often choose weak and easy to remember passwords like “12,345”, “abc 1234” or even “password” to

* Corresponding authors.
E-mail addresses: ioannis.papavasileiou@uconn.edu (I. Papavasileiou), zqiao7@asu.edu (Z. Qiao), chenyu.zhang@uconn.edu (C. Zhang), wenlong.zhang@asu.edu (W. Zhang), jinbo.bi@uconn.edu (J. Bi), song.han@uconn.edu (S. Han).
[1] The first two authors have equal contribution to this work.

https://doi.org/10.1016/j.smhl.2020.100162

Available online 13 December 2020

protect their data, even though those passwords are easy for an unauthorized user to guess Patel et al. (2016). In addition, the rapid advancements in mobile and wearable devices encourage the users to use these devices to monitor and store their sensitive health data, however this brings up challenges such as privacy leakage and financial damage once the devices are lost or compromised Liu and Sun (2016). Strong passwords that combine characters, numbers and symbols are more difficult to hack but can be easily forgotten. In order to bridge the gap between secure authentication and usability, there has been a shift towards biometric authentication methods which take advantage of biological features, such as fingerprints and face characteristics De Luca et al. (2015), or behavioral features like speech, keystroke dynamics Banerjee and Woodard (2012), swipe patterns Patel et al. (2016) and gait patterns Ngo et al. (2015); Juefei-Xu et al. (2012); Sprager & Juric (2015a); Derawi et al. (2010); Kale et al. (2004). These features cannot be forgotten and thus biometric authentication methods significantly improve the usability. Furthermore, continuous biometric authentication systems, especially on mobile devices, are gaining more attention in recent years. Instead of authenticating the user only at the entry point when the device is locked, biometric authentication methods determine whether biometric traits correspond to a respective user in a real-time and continuous manner. In this way, users can be continuously monitored after initial access and thus do not need to constantly worry about security and privacy in case their devices are lost Patel et al. (2016).

Gait-based authentication is among the most popular behavioral biometric authentication methods. Gait refers to locomotion achieved through the movement of limbs and due to the different properties of an individual's muscular-skeletal structure, gait patterns are fairly unique among individuals Zhong & Deng (2014). Gait-based continuous authentication seeks to verify whether the user is genuine in a periodic or constant manner without interrupting the user's normal interaction. It requires little cooperation from the user, and is usually an inexpensive option. Gait is also difficult to mimic Mjaaland et al. (2011, pp. 361–380); Muaaz and Mayrhofer (2017); Gafurov et al. (2007), making spoofing of gait a hard task for an adversary. To perform gait-based authentication, multiple technologies have been used such as cameras for computer vision-based gait recognition Han and Bhanu (2006); Chen et al. (2009), or smart mats Pataky et al. (2012) and plates Derlatka and Bogdan (2015) for floor sensing. With the rapid development of wearable sensors and mobile phones, an increased amount of works that utilize those technologies has been performed recently. However, the majority of those devices are equipped with an inertial measurement unit (IMU) and a common issue of IMU measurements is the lack of accuracy, robustness and reliability Sprager & Juric (2015a).

Motivated by these recent technological advances, in this paper, we present a gait-based continuous authentication framework using multimodal learning. Specifically, our approach aims to support a more user friendly and robust authentication method by combining two sensing modalities, i.e., accelerometer (ACC) data and ground contact force (GCF) data. We employ a multimodal learning approach based on autoencoders to explore the relationships between these two different modalities of the data and thus build more robust learning models leading to more accurate authentication results. Two types of sensor fusion techniques are explored, i.e. early and late sensor fusion. Early fusion is based on the hypothesis that it is possible to develop models that use simple time-domain features for authentication, while the hypothesis for the late sensor fusion is that more complex and abstract features are required for gait-based authentication, and thus extraction of higher-order features based on simple time-domain features is required.

The proposed authentication method can be used in broad application scenarios. Specifically, it may be used as part of a multi-factor authentication framework on mobile phones Muaaz (2013), which is considered stronger than single factor or multi-layer authentication Al Abdulwahid et al. (2016). It may also be used in situations where other strategies such as facial recognition and fingerprints cannot be applied, but continuous authentication is required as part of an ongoing task. For example, continuous biometric authentication could be used to secure sensitive biomedical data recorded from smart wearable health devices. Furthermore, it can help enhance the UI experience, through adaptive interfaces based on different active users, or it can be used along with access control schemes, where access to more sensitive features or data of the device can be done with other more secure biometric methods or strong passwords. In this way, usability is increased, while security is kept high.

The effectiveness of our approach is evaluated through extensive experiments on datasets collected from two case studies, one with commercial off-the-shelf (COTS) running smart socks and the other with a research prototype of smart shoes designed for lower-extremity rehabilitative training, both of which can record GCF and ACC data. With the use of these two different sensing platforms we can evaluate the generalizability of our approach. In the experiments, we first evaluate the robustness of the proposed authentication framework under different attack scenarios, such as passive attack and active attack, i.e. when impostors perform gait mimicking while observing their victim's gait in real time. On top of that, we further evaluate the robustness of the proposed approach while controlling two parameters that effect gait patterns, i.e. walking speed Kirtley et al. (1985) and fatigue Qu & Yeo (2011); Helbostad et al. (2007). Finally, a per modality and per fusion technique evaluation is performed, based on individual sensing platforms. Our extensive results show that between the two modalities used in this approach, GCF is more robust than ACC. In addition, our evaluation shows that an early fusion of the ACC and GCF modalities is the most robust approach, compared to the GCF modality only or any other fusion method. By utilizing information in ACC and GCF data, the models can achieve equal error rates (EER) of as low as 0.01% for the smart socks platform and 0.16% for the smart shoes platform. The leave-one-out approach, which evaluates the generalizability of the proposed method when a never-seen-before impostor tries to be authenticated, achieves a false acceptance rate (FAR) of 0.54% for smart socks and 1.96% for smart shoes. Walking parameters, such as speed and fatigue from everyday activities are shown to have significant impact on the authentication performance as well. This suggests that providing diverse gait samples can further improve the robustness of gait-based authentication models.

The remainder of this paper is organized as follows. Section 2 reviews the related work. Section 3 gives an overview of our approach. The two sensing devices used for data acquisition are discussed in Section 4. Section 5 describes our filtering and data segmentation techniques. Section 6 discusses the feature extraction and classification models of the gait patterns. Section 7 describes the data collection protocol and the evaluation results. Finally, we conclude the paper in Section 8 and discuss the future work.

2. Related work

Gait-based behavioral biometrics can be used in both identification and authentication scenarios Zhong & Deng (2014). For identification, a sample gait is compared to a database of enrolled gait samples with known identities to determine whom the unknown sample belongs to; for authentication, a gait sample is compared to the enrolled sample gait data for a known person to validate his or her identity. Gait biometrics have attracted tremendous research attentions in recent years due to two main reasons: the rapid development of sensing technologies on mobile devices, and the increasing popularity and usability of biometric authentication compared to traditional authentication methods. Apart from that, gait-based authentication may be harder to spoof when compared to other widely used physiological biometric methods. For example, fake fingers have been used to get access in fingerprint based systems; recorded voice has been used in voice recognition systems; and pictures and masks have been used for face recognition based authentication systems Patel et al. (2016); Gafurov et al. (2007); Al Abdulwahid et al. (2016). In addition, gait-based biometric authentication may be preferred from individuals when compared to other biometric methods which may be considered more privacy intrusive, e.g. face recognition that requires the use of cameras Vildjiounaite et al. (2006).

In general, there are three main approaches in terms of hardware used for gait-based authentication, i.e., computer vision techniques, sensing on the floor with smart mats and plates, and wearable devices. Computer vision techniques are based on video recordings of the subject to be authenticated or identified Han and Bhanu (2006); Chen et al. (2009). They have drawn great attention recently due to their promising application to security, monitoring and surveillance systems in public places, such as airports. Floor sensing has been performed with the use of smart mats Pataky et al. (2012), force plates Derlatka and Bogdan (2015) and floor vibration measurement Pan et al. (2017). Its application mostly focuses on the identification of people entering a restricted area.

Using wearables to perform gait-based biometric authentication is another popular approach due to their mobility, small size and low cost. That includes the use of IMUs and devices such as smart phones and smart watches. These wearables can record signals at multiple locations on the human body, including wrist Xu et al. (2017), waist Sprager & Juric (2015a); Ngo et al. (2015), breast pocket, trouser pocket Juefei-Xu et al. (2012); Zhong & Deng (2014); Derawi et al. (2010); Zhong et al. (2015), and hip Nickel et al. (2012); Gafurov et al. (2007). A summary of recent research work on using accelerometer for gait recognition can be found in Zhang et al. (2015) and Sprager & Juric (2015b). A common issue of IMU measurements is the lack of accuracy, robustness and reliability Sprager & Juric (2015a). This can be partially attributed to the high sensitivity to the location and orientation of the sensor and increased level of noise present in the recorded data. However, many of these recent research efforts have achieved remarkable performance. For example, a 0.8% equal error rate (EER) was reported in Sun & Yuao (2012) by applying a curve aligning approach on the dataset collected from 22 subjects. Research also shows that EER keeps increasing when the dataset size grows, e.g. Sprager & Juric (2015a) reports an EER of 6%–12% evaluated on a large open source dataset of 744 subjects. This increase may be related to multiple other parameters on top of the population size, such as types of sensing technologies, sensor placement and noise levels in the data.

Besides wearables, smart phones can also be easily used as sensing devices for gait recognition, as they require no additional hardware support Lu et al. (2014). However, the effectiveness of using smart phones for gait authentication heavily depends on the location and orientation of the phone to be deployed on the human body. The performance will further degrade when the subject performs everyday actions with their phones (e.g., making calls or browsing the web Watanabe & Sara (2016)). Among the research attempts in this direction, Shen et al. (2018) studies gait authentication based on multiple placements of phones on the human body; several work rely on using the phones in a fixed position in their pocket Nickel and Busch (2013); Zhong & Deng (2014) addresses the issue of variable phone orientation by computing invariant gait representations and uses gait dynamic images to extract features, achieving an EER of 3.88–7.22% with 55 human subjects; Juefei-Xu et al. (2012) presents a pace-independent gait identification system with 36 subjects, achieving verification rates (VR) of 61.1–99.4% with a False Acceptance Rate (FAR) of 0.1%.

To further improve the authentication/identification performance and alleviate the performance degradation from noisy measurements, multimodal methods and sensor fusion techniques have been used to combine different types of data from multiple sources Zhang et al. (2015). These approaches are attractive as they can effectively relate the increased information available from multiple sources and modalities and thus result in better models that outperform the traditional methods. For example, Crouse et al. (2015) introduces a continuous authentication method for mobile devices based on fusion of face images and IMU data. Zhang et al. (2015) presents a sparse representation method with the use of four accelerometer data sources, achieving an EER of 2.2% for verification. Such approaches however may come with reduced usability as the potential user may need to wear multiple wearable devices, making the setup not practical for everyday use. In addition, most of the studies fail to report leave-one-out cross-validation, which gives a better estimate of the generalizability of the approach when a new impostor subject is tested whose gait data are not used to train the corresponding model. For the studies that do report leave-one-out cross-validation, the best FAR performance achieved is 3% with 11 subjects Trivino et al. (2010) and 6% with 32 subjects Trung et al. (2011).

In this work, we use two modalities for gait-based authentication, which combine accelerometer (ACC) and ground contact force (GCF) data. GCF measurements can be recorded with the use of smart socks or smart shoes, which has recently seen great advancement in multiple domains, especially for gait rehabilitation Kong and Tomizuka (2009); Papavasileiou et al. (2017b, pp. 195–204, a, pp. 34–49); Zhang et al. (2016). We hypothesize that by building models that combine data from multiple sources, we can achieve robust gait-based authentication. To the best of our knowledge, this is the first attempt in gait-based authentication using a combination of ACC and GCF data. Improved usability can be achieved with this approach, as users do not need to use extra wearables, except from a pair of smart socks or smart shoes that can record both modalities. With the rapid development of new wearable technologies, we envision that most shoes and socks in the near future will be equipped with smart sensors to capture ACC and GCF data in a continuous manner. In the following, we will first give an overview of our proposed methodology and then present the technical details.

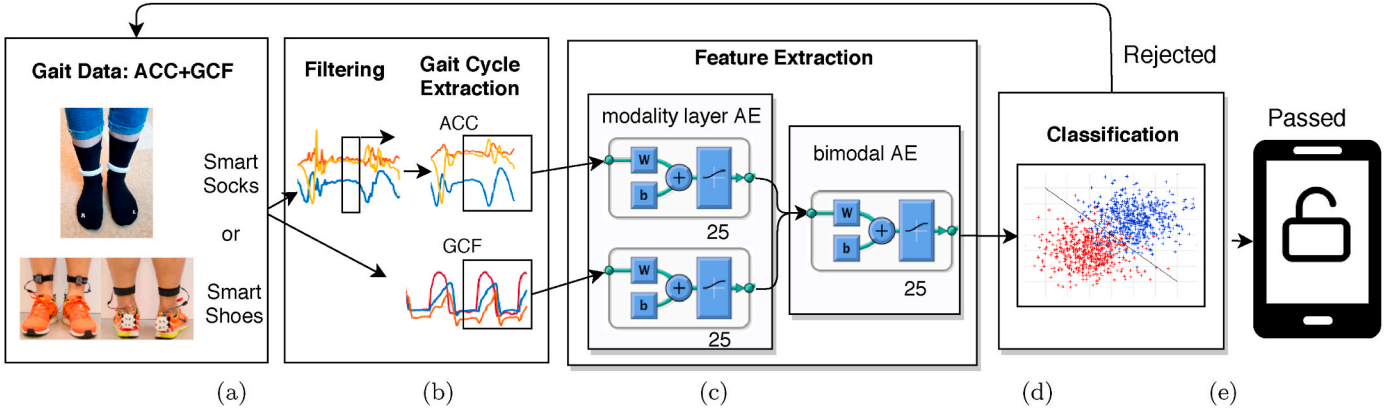
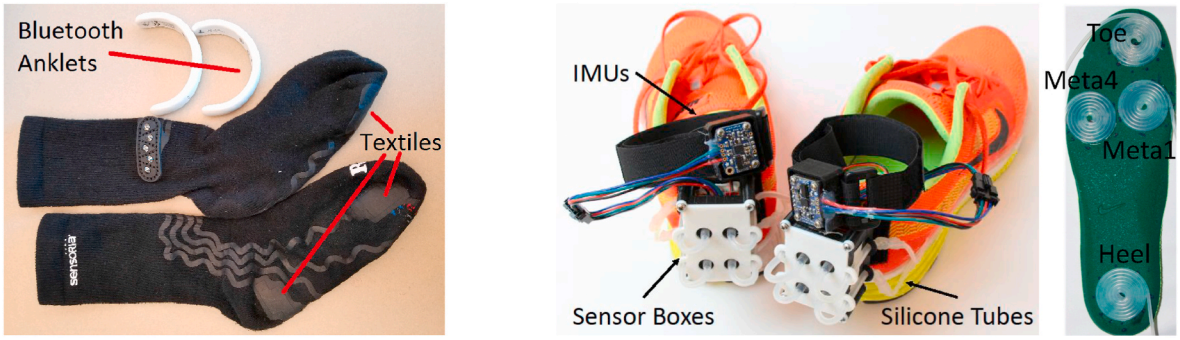


Fig. 1. An overview of the proposed integrative framework for multimodal gait-based continuous authentication: (a) sensing platforms, (b) filtering/cycle extraction, (c) feature extraction, (d) classification, and (e) authentication.



(a) Smart socks from Sensoria (b) Research prototype smart shoes

Fig. 2. The two sensing platforms for data acquisition.

3. Methodology overview

To support multimodal and continuous gait-based authentication we propose an integrative framework that comprises of a data acquisition platform, a gait cycle extraction component, a feature extraction component based on autoencoders and a classification component. An overview of the proposed framework is presented in Fig. 1. The sensing device used in the data acquisition platform (Fig. 1a) is either a pair of smart socks or smart shoes (see Section 4 for the details). Both sensing devices are capable of recording synchronized GCF and ACC motion data, and are equipped with wireless modules to transfer the recorded data to an application on a phone or a laptop. The corresponding wireless connection between the sensing platform and the mobile phone is considered to be secure and encrypted, so that there is no possibility for a replay attack, i.e. a type of attack in a biometric system where old captured data from previous sessions or other users are used for authentication.

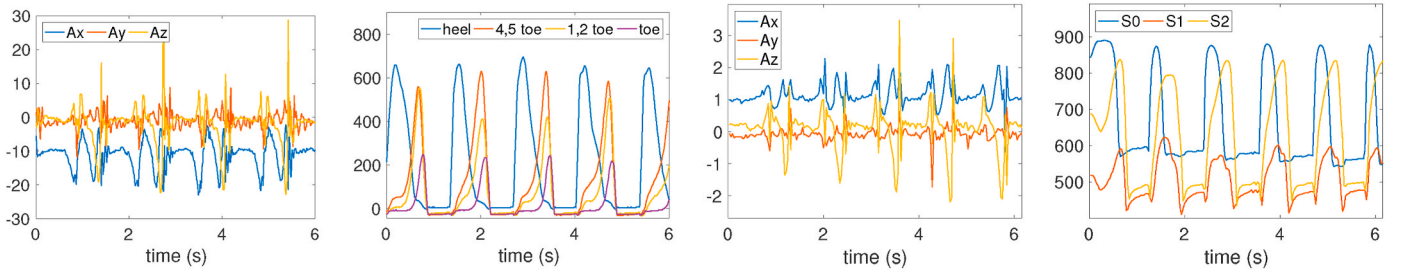
The collected data samples are then sent to the filtering and gait cycle extraction component (Fig. 1b) which is responsible for filtering the ACC data, segmenting the GCF and filtered ACC gait data into individual gait cycles, and storing the data within each gait cycle in a vector and forward them to the feature extraction component (Fig. 1c). For feature extraction, we use autoencoders. Different sensor fusion techniques can be achieved, depending on the way how the selected autoencoders are connected and what sensor types they get their input from. An overview of the five investigated sensor-fusion techniques is summarized in Fig. 6. The first model receives the stacked raw ACC and GCF gait cycle data as input and learns a shared representation. The next two models each use only raw gait cycle data from one of the two modalities. The fourth model stacks the two individual encoder outputs from each modality. Finally, the last model forms a bimodal-deep autoencoder network that learns higher order features Ngiam et al. (2011). The number of nodes in each autoencoder has been empirically selected, aiming to reduce the initial raw data dimensionality, and improve the feature quality and class separation (see Section 5).

The extracted features are finally sent to the last component of the processing pipeline for classification (Fig. 1d). The classifier decides whether the given feature vector from the corresponding gait cycle belongs to the legitimate user or not. In order to achieve this, we build classification tasks for each individual user. Training data belonging to the corresponding owner of the model are marked to be in the positive class, while data from all the other subjects in the training set are marked to be in the negative class. In order to achieve continuous gait-based authentication, the pipeline can be repeatedly invoked to decide whether a given gait cycle belongs to the corresponding owner of the mobile device. If the device is unlocked there could be a time interval before the method is invoked again to save resources, while keeping the user authenticated continuously. In addition, to improve the overall user experience and authentication performance, an additional ensemble or voting layer can be used to decide whether to lock or unlock the device based on the classification results from a number of past gait cycles. In this paper, we focus on the first four components of the proposed framework as depicted in Fig. 1 and plan to extend our study on the additional layers in the future work. The following three sections give a more detailed description of the components in the proposed gait authentication framework.

4. Data acquisition platforms

Multimodal learning essentially reveals correlations among different modalities from multiple data sources to build stronger and more robust learning models than those learned from individual modality Ngiam et al. (2011). In this work, we relate features extracted from users' ACC and GCF data to build an improved model for gait-based behavioral biometric authentication. To demonstrate that our algorithms can be applied on different sensing platforms, we use both a commercial off-the-shelf (COTS) smart socks and a research prototype smart shoes for data acquisition. As will be shown in Section 7, despite the numerous differences between the two sensing devices, our algorithms can achieve similar performance and are not affected by the different characteristics of the sensing platforms.

The COTS smart socks are purchased from Sensoria Sensoria (2020) (see Fig. 2a). The socks are designed for runners who need to improve their running skills and get real-time feedback on multiple parameters, such as cadence, foot landing position, pace and speed. The Sensoria Software Development Kit (SDK) includes a license to support raw data collection, a pair of smart socks and a pair of



(a) ACC raw data (left shoe) (b) GCF raw data (left shoe) (c) ACC raw data (left sock) (d) GCF raw data (left sock)

Fig. 3. Raw ACC and GCF data from two platforms.

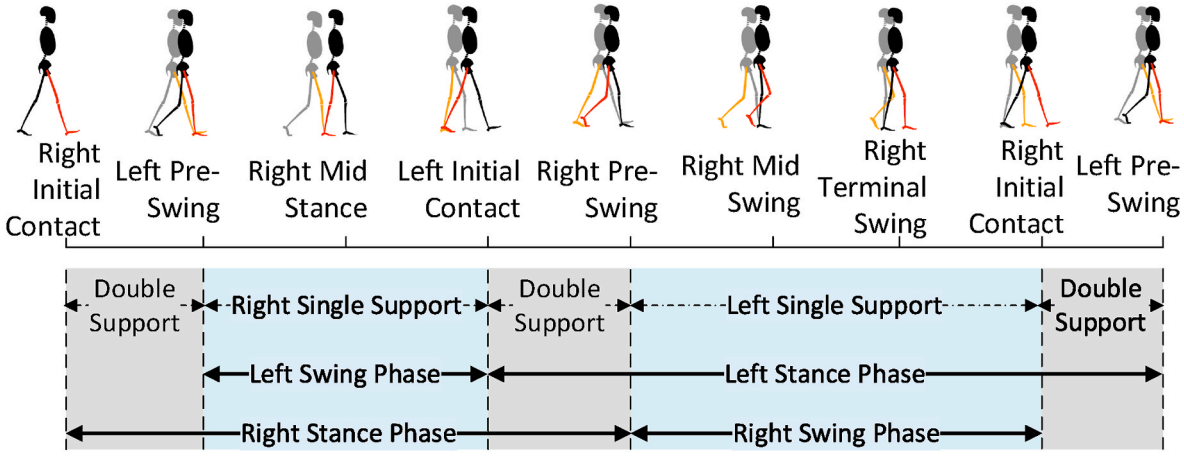


Fig. 4. An overview of gait cycles and gait phases.

Bluetooth anklets for wireless data collection. The socks are embedded with 3 proprietary textile pressure sensors, attached at the bottom of the sock, one in the heel area under the calcareous bone and two in the metatarsal area, at the first and fifth joints, respectively. The collected GCF signals are relayed through conductive fibers to the anklet. The attachable Bluetooth anklet contains a 3-axis accelerometer, making the hardware completely mobile. The GCF signals along with the ACC signals, are sent to the SensoriaLab iOS application, where data can be stored locally or uploaded to the cloud for further processing. The battery of the anklet allows about 6 h of operation and the socks' sampling frequency is set to 32 Hz.

The smart shoe is a novel wireless human motion monitoring system for gait analysis in rehabilitation training Zhang et al. (2016); Deng et al. (2018). It is developed to measure the GCF at four points: toe, first metatarsal joint (Meta 1), fourth metatarsal joint (Meta 4) and heel (see Fig. 2b). The silicone tubes are wound into air bladders and connected to barometric pressure sensors. In addition, an IMU sensor is attached to the distal end of the shank to measure the accelerations and rotations in three dimensions. The sampling rate is set to 30 Hz and the data is sent to a high-performance laptop through WiFi.

Despite the fact that both sensing devices provide similar sensing modalities, there are multiple differences between the two. For example, the smart socks are equipped with three textile pressure sensors at the bottom of the socks, while the smart shoes have four air bladders embedded in the sole and connected to the barometric sensor on the back. In the smart socks, the ACC data are collected by the attachable Bluetooth anklet, while in the smart shoes, the IMU sensor measurement is sent through the WiFi. These characteristics can cause evident differences in the collected raw data. For example, in Fig. 3b and d, we can observe that the data collected from smart shoes change more sharply than the measurements from the socks. This could be due to the participants' walking patterns since the data are collected from two individuals. In addition, a second factor for this difference could be the nature of the textile material which is less sensitive, compared to the barometric sensors that can capture small variations in data.

5. Data filtering and gait cycle detection

After receiving the ACC and GCF measurements from the data acquisition platforms, a pipeline of processing components are employed in order to provide continuous and robust gait-based authentication. In this section, we examine the data filtering and gait cycle detection components, which are used to reduce the noise levels in the ACC signals and segment the data stream to perform further processing.

5.1. Filtering

Both sensing devices provide raw unfiltered ACC and GCF data. As it can be seen in Fig. 3a and c, the ACC data contain higher noise levels compared to the GCF data. In order to reduce the noise levels in the ACC data, we employ a moving average approach. Moving average filters are low-pass filters, which are easy to implement, and provide great smoothing performance. This greatly helps in the following analytic layers to prevent overfiting and improve generalization. In our design, the ACC raw data from every sensor channel (x, y and z directions) from both sensing devices will pass through a 5-point moving average filter. The filter length was chosen to be small as we do not want to miss any important information, such as spikes from the feet movement, that could help distinguish a subject. The filter output is given by the following difference equation:

$$y[i] = \frac{1}{N}\sum_{j=0}^{N-1} x[i-j] \tag{1}$$

where $y[i]$ is the filter output at timepoint i, $x[i]$ represents the input data at timepoint i, and $N = 5$ is the filter length. If $i < j$ we can set

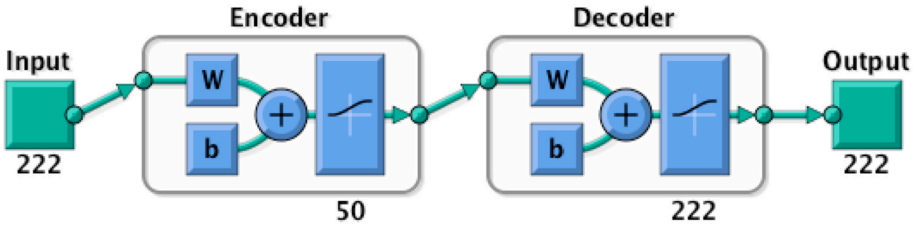


Fig. 5. The conceptual structure of the autoencoder.

$y[i] = x[i]$ and skip filtering the first N observations.

5.2. Gait cycle detection

Gait cycle is the time interval between the same repetitive events of walking. Each gait cycle can be further divided into gait phases. Fig. 4 gives an overview of two gait cycles at the lower two horizontal solid lines and gait phases at the top part. Typically there are eight gait phases for a healthy subject, i.e. initial contact, loading response (or pre-swing), mid-stance, terminal stance (or initial contact), pre-swing, initial swing (not shown in Fig. 4), mid-swing, and terminal swing. Detecting gait cycles can be challenging, especially when only ACC data are used and the IMU sensor recording point is far away from the foot. However, the heel GCF sensor data can help easily detect heel strikes. By using the measured contact force of the heel with the ground, we can accurately detect the repeating heel strike gait phases. Based on this, we can define a gait cycle to be the time interval between two consecutive left heel strikes. Since the gait cycle length is not constant and depends on the walking speed, we use a fixed-size window that starts with a left foot heel strike, and has a length of 37 samples, i.e. 1.156 s at 32 Hz or 1.233 s at 30 Hz (Fig. 1b). This window length is empirically chosen to capture any walking speeds of a subject, except extremely slow walking, such as walking with less than 48 strides per minute. Note that on the two data acquisition platforms, both ACC and GCF data are recorded with the same timestamp. Thus data synchronization between the two modalities is not necessary during the gait cycle extraction.

Once gait cycles have been identified, normalization across the two different modalities is performed, so that normalized data lie in the $[0, 1]$ space. Normalization is important not only to help utilize the autoencoders in all the models for feature extraction, but also to eliminate differences between body-weight across individuals. In this way adversaries may not benefit when they try to match their body-weight to their victim. The formula used for normalization for both ACC and GCF data is as follows.

$$y_j[i] = \frac{x_j[i] - \min x_j[i]}{\max x_j[i] - \min x_j[i]}, i \subseteq \mathbb{I}, j \in \mathbb{C} \tag{2}$$

$\mathbb{I}$ refers to the set of indices, i, that belong to one gait cycle, while $\mathbb{C}$ refers to all the channels that correspond to this modality. By taking the *min* and *max* across all the channels in a gait cycle, we are able to keep the relative differences across sensor channels, and thus let the models used at the later components benefit from differences in force levels, or timing of gait patterns.

The last step of gait cycle detection is to concatenate the feature vector into a single feature vector. The vector length is $L_v = L_w \times 2 \times N_c$, where $L_w = 37$ $(observations/window)$ is the window length, 2 is the number of feet (left and right), and N_c is the number of channels for each sensor. For example, the ACC sensors have 3 channels (x, y and z directions), the shoe GCF sensors have 4 channels (heel, meta12, meta 45 and toe) and the sock GCF sensors have 3 channels (heel, meta12 and meta). This results in a vector of 222 input features for the ACC modality for both smart socks and shoes, 222 features for the GCF modality of the smart socks and 296 features for the GCF modality of the smart shoes. Before these gait cycle vectors are sent to the next component for feature extraction, in order to train clean models and improve generalizability, outliers removal is performed. If the corresponding gait cycle vector contains more than 10 features that have observations more than 3 standard deviations away from their mean across all the gait cycle vectors belonging to the corresponding user, this feature vector is considered outlier and will be discarded.

6. Feature extraction and classification

To extract features from the detected gait cycles we employ autoencoder as a building block for early and late fusion of ACC and GCF data within a gait cycle. Early fusion technique tries to learn models that extract simple temporal and amplitude features, while late fusion technique uses per modality features to extract higher order and abstract features. In the following, we first describe the detailed characteristics of autoencoders and then discuss how they are used for feature extraction and sensor fusion. Finally, we describe the classification algorithms used on the extracted features to support gait-based authentication.

6.1. Feature extraction with auto-encoders

Autoencoders have recently been applied in a broad range of applications (e.g., image, video and audio processing Ngiam et al. (2011)) to find higher order correlations from different data sources and extract meaningful features that can better represent the data. Compared to linear methods such as the principal component analysis (PCA), autoencoders can achieve better feature extraction and

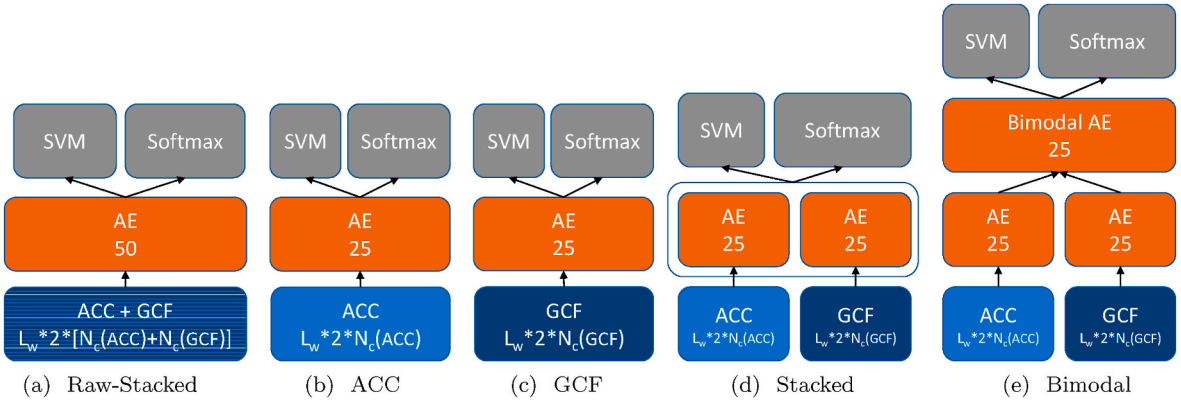


Fig. 6. Models used for sensor fusion and classification.

dimensionality reduction, due to its non-linear transfer function. For these reasons, in this work we use autoencoders to extract features separately for each vector from the two different modalities and then add another layer for bimodal feature extraction.

An auto-encoder is a neural network that is trained to replicate its input at its output (see Fig. 5 for a conceptual structure). Training an autoencoder is unsupervised, in the sense that no labeled data is needed, and is based on the minimization of the error between input $\mathbf{x}$ and its reconstruction at the output $\widehat{\mathbf{x}}$. An autoencoder is composed of an encoder and a decoder. Given an input vector $\mathbf{x} \in \mathbb{R}^D$, the encoder tries to map vector $\mathbf{x}$ to a hidden representation $\mathbf{z} \in \mathbb{R}^{D'}$ by learning a function $h_{W,b}$ as follows: $\mathbf{z} = h_{W,b}(\mathbf{x}) = s(\mathbf{W}\mathbf{x} + b)$. S is a transfer function for the encoder (we use the sigmoid function), $\mathbf{W} \in \mathbb{R}^{D' \times D}$ is a weight matrix, $b \in \mathbb{R}^{D'}$ is a bias vector, and D, D' are the number of nodes at the input and hidden layers, respectively. During training the weight matrix $\mathbf{W}$ and the bias vector b are learned.

The decoder maps the encoded representation $\mathbf{z}$ back to a reconstructed vector $\widehat{\mathbf{x}}$ in input space by learning a function $g_{W',b'}$ as follows: $\widehat{\mathbf{x}} = g_{W',b'}(\mathbf{z}) = s(\mathbf{W}'\mathbf{z} + b')$, where $\mathbf{W}' \in \mathbb{R}^{D \times D'}$ is a weight matrix, and $b' \in \mathbb{R}^D$ is a bias vector. Autoencoders can achieve better feature extraction and dimensionality reduction, compared to linear methods (e.g., PCA) due to the non-linear transfer function s.

The loss function used to train autoencoders is typically the mean squared error loss:

$$E = \frac{1}{N}\sum_{n=1}^{N}\sum_{j=1}^{D}\left(\mathbf{x}_n^j - \widehat{\mathbf{x}}_n^j\right)^2, \tag{3}$$

where N is the number of observations, and D is the number of variables in the training data, $\mathbf{x}_n^j$ is the j-th variable of the n-th training sample, and $\widehat{\mathbf{x}}_n^j$ is the j-th variable of the reconstruction of the n-th training sample from the autoencoder. To avoid over-fitting, a regularization term Ω_w (weight decay) is typically introduced which favors small weights in $\mathbf{W}$. In addition sparsity on the encoded representation can be enforced by adding a regularization term Ω_s that takes a large value when the average activation value, $\widehat{\rho}_i$, of a neuron i and its desired value, ρ, are not close in value Olshausen and Field (1997). By adding these two regularizes to the loss function (3) we get:

$$E = \frac{1}{N}\sum_{n=1}^{N}\sum_{j=1}^{D}\left(\mathbf{x}_n^i - \widehat{\mathbf{x}}_n^j\right)^2 + \lambda * \Omega_w + \beta * \Omega_s \tag{4}$$

where:

$$\Omega_w = \frac{1}{2}\sum_{i=1}^{D'}\sum_{j=1}^{D}\left(w_{ij}\right)^2 \tag{5}$$

is a regularization term that is used to decrease the magnitude of the weights and helps prevent over-fitting (it is also called weight decay), λ is a hyper-parameter, which controls the importance of weights regularization, w_{ij} is the parameter (or weight) associated with the connection between the i-th unit in the hidden layer and j-th unit in the input layer and

$$\Omega_s = \sum_{i=1}^{D'} KL\left(\rho||\widehat{\rho}_i\right) = \sum_{i=1}^{D'}\rho\log\left(\frac{\rho}{\widehat{\rho}_i}\right) + (1-\rho)\log\left(\frac{1-\rho}{1-\widehat{\rho}_i}\right) \tag{6}$$

is the sparsity regularization term, β is a hyper-parameter which controls the importance of sparse representations, $\widehat{\rho}_i$ is the average activation value of a neuron i and ρ is its desired value called sparsity parameter, typically set to $\rho = 0.05$. $KL(\rho||\widehat{\rho}_i)$ is the Kullback-

Leibler (KL) divergence and is measuring how different two distributions are. In this case, it takes the value zero when ρ and $\widehat{\rho}_i$ are equal to each other, and becomes larger as they diverge from each other. The decoder objective function minimization is performed in a similar fashion.

By using the autoencoders as a building block, we can develop different models that perform feature extraction with different characteristics. We discuss the employed approaches in the following subsections.

6.2. Early and late sensor fusion techniques

When dealing with multimodal sensor data in neural networks, it is possible to fuse the data at different stages of the network, achieving different sensor fusion techniques Münzner et al. (2017). Here we investigate early fusion and late fusion approaches, while also evaluating no-fusion between modalities to understand the importance of GCF and ACC modalities. Early fusion has the advantage of a simple model, that lets the algorithm figure out which feature from which modality is important for authentication. On the other hand, late fusion is designed to identify more abstract characteristics, as it is the task of the lower layer encoders to learn specific features from each modality.

An overview of the investigated sensor fusion techniques is presented in Fig. 6. The first model (Fig. 6a) performs early fusion by receiving as input a stacked vector of raw ACC and GCF gait cycle data and treating them equally. Its advantage lies in extracting temporal and amplitude related features that may be easier to detect with a simple model. The next two models (Fig. 6b and c) each uses only raw gait cycle data from one of the two modalities, performing no sensor fusion with the other modality. They have similar characteristics with the first model in extracting simple features for each modality, but do not perform any fusion of features between the two modalities.

Late sensor fusion is performed by the last two models. The fourth model (Fig. 6d) is an extension of the two per-modality feature extraction models. It combines the outputs of these two models and lets the classification algorithm to learn a good combination of features extracted from the two modalities separately. Finally, the last model (Fig. 6e) forms late fusion with a bimodal-deep autoencoder network that fuses the two individual models at the second layer in order to learn higher order, more complex and abstract features given features that were extracted at the first layer, per modality Ngiam et al. (2011). Its advantage lies in extracting features that are not easy to detect with a single layer of encoding and may require more complex, non-linear transformations, such as those required for audio-visual classification Ngiam et al. (2011).

For all the models, the decoders of any autoencoder are discarded after training as the goal is for feature extraction and no reconstruction is required. The number of nodes for each encoder is set to 25, with the exception of the raw-stacked model where 50 nodes are employed. This selection was decided after trying different number of nodes and selecting those leading to increased performance. Higher number of nodes for the raw-stacked model could be explained since this model has to fuse features from both modalities, so there could be more information potential to be learned from the raw data.

6.3. Classification of gait features

Most related work in the literature studies the authentication or identification of gait using distance metrics or pattern similarity measures, such as dynamic time warping (DTW) Sprager & Juric (2015b). One drawback of those methods is that DTW might warp the series too much so that the series lose its discriminative patterns Zhang et al. (2015). In this work, we use autoencoders to perform feature extraction and rely on a classification algorithm to make the final decision of authentication. It is a standard practice for many neural network algorithms to attach an additional final layer (called soft-max layer) of one output node per data class, that will provide a probability that the corresponding input belongs to each class. Here we define a data class to be a human subject, so after a new gait cycle is passed through the model, we have class probabilities, i.e. the probability that it belongs to any class, with the sum of probabilities equal to one.

In addition, we use support vector machine (SVM), which is one of the most popular machine learning algorithms. SVM can achieve improved performance in binary classification, and especially in authentication, since the algorithm is trained to learn a separating hyper-plane that separates the positive class (genuine user) from the negative class (impostor users) in the best possible way. By doing so, we expect that unknown gait samples that belong to an unknown impostor user can be more easily rejected by SVM, which is trained to specifically identify gait patterns of the corresponding genuine user. SVM maps the input points in a high dimensional space using a kernel. A hyper-plane is used to divide the geometric space into two parts for classification. The main advantage of SVM is that it solves a convex problem and is suitable for classification of continuous features. The regularization parameters can be tweaked to control over-fitting. We use SVM by defining a classification task for authentication of each of the subjects, which predicts whether a new gait sample originated from that user or not. To train SVM the user's gait is marked as positive, while all other subjects' gait samples are marked as negative.

7. Performance evaluation

To evaluate the effectiveness of the proposed gait-based authentication framework, we design two case studies. We use smart shoes in one of the studies and smart socks in the other. This helps evaluate the robustness of the approach on different sensing platforms. In addition, different testing scenarios are considered for each of the studies that can affect gait dynamics, such as walking speeds and fatigue levels. In the following, we first summarize the design principles of the case studies and the experiment setup. We then present

Table 1
A summary of the two case studies with different sensing platforms and testing scenarios.

Case Study	Sensing Platform	Testing Scenarios
First	Smart socks	Fast walking
First	Smart socks	Slow walking
Second	Smart shoes	Fast walking in the morning
Second	Smart shoes	Fast walking in the afternoon
Second	Smart shoes	Slow walking in the morning
Second	Smart shoes	Slow walking in the afternoon

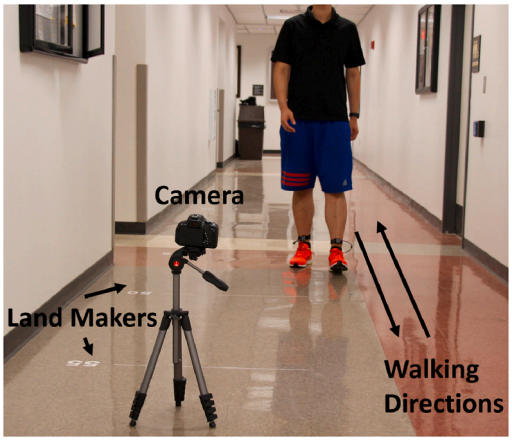


Fig. 7. Test environment for the smart shoes case study.

the details of the experimental results.

7.1. Case studies

In both case studies, we select healthy and young adult subjects to participate in the experiments. The studies are designed to test the efficacy of the proposed algorithmic framework under different scenarios (see Table 1). Our first case study uses Sensoria smart socks (Fig. 2a). Seven female and eight male healthy students participate in this pilot study and their ages range from 20 to 29. For each subject we demonstrate how to use the hardware and give instructions for the data collection sessions. Specifically, the subjects are asked to walk normally on a 55 feet long hallway back and forth for 5 min. To collect the two modalities of the gait data, the subjects wear the smart socks and carry a smart phone in their pocket with the SensoriaLab iPhone application for data storage. Before the 5 min recording session, we make sure that the socks are worn correctly, the Bluetooth anklet is adequately charged, the wireless connection with the phone is in good condition, and the ACC and GCF signal quality from the socks and anklet do not show any abnormalities or excessive noise. During the 5 min recording session, the subjects are asked to walk in two different walking speeds, i.e. slow and fast walking. At the first two and half minutes they walk at slow speed, between 3 and 4 feet per second and at the remaining time they walk between 5 and 6 feet per second. In order to make sure that the subjects follow the walking speed requirements, the student researcher kept walking on the side of the subject for the first two round-trip walks for both walking speed recording sessions in order to keep the speed constant. The researcher had performed the walking session multiple times using a timer to measure the exact time requirements for the given hallway length. The sampling rate of the smart socks is 32 Hz and the total number of gait cycles across all subjects for all sessions recorded is 4004.

In the second study, 10 healthy male subjects aged from 21 to 27 are invited. Limited by the shoe size, only male participants are selected, whose shoe sizes are either 10 or 11. The subjects are asked to walk on a 50 feet long hallway at two speed cases: slow and fast. The speed for slow and fast walking are set as 3–4 feet per second and 5–6 per second correspondingly. To ensure the recorded data are within the speed range, makers are labeled on the hallway floor for every 5 feet and a camera is used to record the whole experiment, shown in Fig. 7. After synchronizing the video with the shoe measurements, the data where the speed is not within the requirement are removed. For each subject, the experiments are finished twice at around 11 a.m. in the morning and 4 p.m. in the afternoon on the same day. The goal of this study is to investigate whether the reduced energy level in the afternoon would affect the performance of gait-based biometric authentication. In total, 9357 gait cycles are recorded from all the subjects. The data are collected by the smart shoes at a sampling rate of 30 Hz to match the experiment with smart socks. Joint acceleration data in three dimensions are collected and pressure data at four different locations (toe, the first and second metatarsophalangeal joint, the fourth and fifth metatarsophalangeal joint, and the heel) at feet are collected.

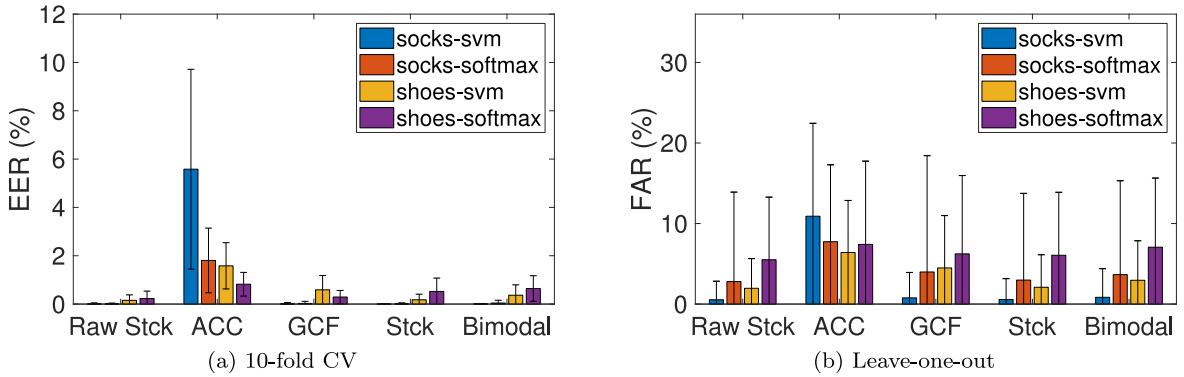


Fig. 8. 10-fold and leave-one-out CV for the two studies.

Table 2
EER and FAR performance per modality and fusion model, with SVM classification.

Experiment Set	Platforms	Metric	Raw stck	ACC	GCF	Stck	Bimodal
1	Shoes	EER	0.16 ± 0.22	1.58 ± 0.96	0.59 ± 0.59	0.18 ± 0.23	0.37 ± 0.43
1	Socks	EER	0.01 ± 0.04	5.58 ± 4.13	0.01 ± 0.05	0.00 ± 0.00	0.01 ± 0.02
1	Shoes	FAR	1.96 ± 3.69	6.40 ± 6.46	4.48 ± 6.50	2.09 ± 4.04	2.96 ± 4.90
1	Socks	FAR	0.54 ± 0.23	10.91 ± 11.53	0.77 ± 3.16	0.56 ± 2.60	0.84 ± 3.56

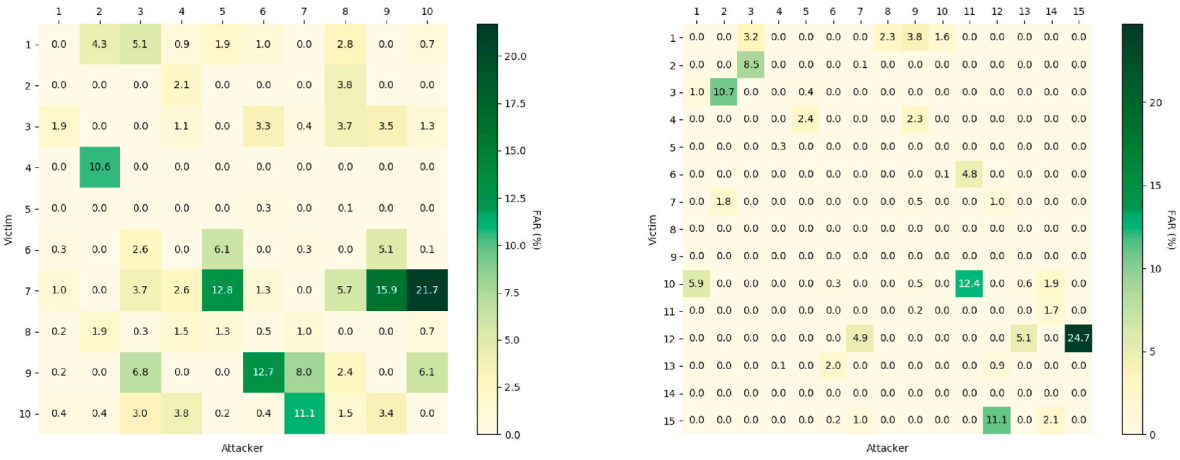
7.2. Experimental setup

To test the effectiveness of the proposed method, we train a binary classifier for each of the subjects in the dataset. Data originating from the corresponding subject are marked to be in the positive class, while all the others are marked negative. To report the generalizability of the models, *k*-fold cross-validation (CV) is adopted. In *k*-fold cross-validation, the dataset is split into k separate equal subsets, one subset is used for testing and the rest $k-1$ subsets are used for training, and this is repeated for all the k subsets. In addition, we perform leave-one-out cross-validation to report how well the models can generalize and predict a never-seen-before impostor. In this case, data belonging to the impostor subject are left out for testing, and the rest data from all the other subjects are used to train the models. For all the experiments, average and standard deviation of the performance metrics are reported. The available training set for each experiment is used to train both the auto-encoders network and classifiers. The test set is tested against the models returned by the training set.

To select the hyper-parameters, λ and β, for the auto-encoders, grid search is conducted on the complete training dataset from the first case study and the set of parameters that achieved the highest performance is selected. The kernel scale parameter in SVM, is selected by a heuristic approach in Matlab, which uses sub-sampling. Due to the hardware differences between the two sensing devices, experiments are done independently and only data from one sensing device are used in each experiment.

Biometric authentication methods are typically evaluated using three performance metrics, i.e. false acceptance rate (FAR), false rejection rate (FRR) and equal error rate (EER). Given a new-to-be-tested observation, a classifier returns a score (or probability) that this observation belongs to the positive class, i.e., the class with gait samples of the genuine user. If this score exceeds the acceptance threshold, the observation is accepted, otherwise is rejected. Based on that, FAR is defined as the portion of imposting recognition attempts that are accepted (score above threshold) and FRR is defined as the portion of genuine recognition attempts that are rejected (score below and equal to threshold). A trade-off between these two types of errors is achieved by varying the acceptance threshold, so that as error of one type decreases, error of the other type increases. Thus a common way of evaluating the performance of a biometric system is to estimate the point where FAR and FRR are approximately equal Vildjiounaite et al. (2006), which is called EER. In our experiments, we report EER when we perform *k*-fold CV, since the test set contains observations from both negative and positive classes, so we can report the point where FAR and FRR are equal. If the generated data do not provide scores that set FAR and FRR equal, we report the average of the two metrics at the point where their difference is minimum. Finally, we report FAR for the leave-one-out CV, since the test set of these experiments contains impostor gait samples and thus should be rejected by the algorithms.

We perform extensive experiments to evaluate the performance of the proposed authentication method. First we perform cross-validation and leave-one-out evaluation to assess generalizability and robustness. Next, we conduct experiments to control parameters that affect gait, such as walking speed and fatigue due to daily activities, and report their effects on the proposed gait-based authentication method. In addition, we evaluate how the training set size, in terms of the number of gait cycles, will affect the authentication performance. Finally, we evaluate the robustness of the proposed method when a gait mimicking attack is performed by an adversary.



(a) Results of the shoe study (b) Results of the socks study

Fig. 9. Per-subject leave-one-out performance.

7.3. Cross-validation and leave-one-out evaluation

In the first set of experiments, we first perform 10-fold cross-validation on both case studies. Average EER results across all folds and all subjects are reported in Fig. 8a. These results (SVM only, average and standard deviation) can also be seen in Table 2, since the small differences cannot be visualized. For this experiment, data from both slow and fast walking speeds are used for smart socks and smart shoes, and both morning and afternoon sessions are used for the smart shoes.

To evaluate the generalizability of the models, we further evaluate the effectiveness of the proposed approach when a never-seen-before impostor tries to get authenticated. For this, we perform leave-one-out cross-validation using training data similar to the previous experiment, including both slow and fast walking datasets from the socks and shoes studies and both morning and afternoon data from the shoes study. This experiment can be taken as a passive attack, since the impostors select their victim to attack without performing any active attempt to hack their gait-based authentication model. They passively hope that their gait patterns are close enough to the victim's patterns, so that the models accept them. Both the average and standard deviation of the FAR values are reported in Fig. 8b. A summary of the performance per model is given in the last two rows of Table 2.

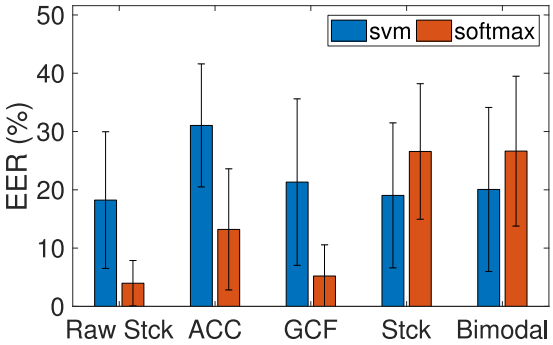
From these results, we first observe that the GCF modality, for both studies, outperforms the ACC modality. This indicates that by adopting advanced sensor technologies that can record GCF data, we are able to significantly improve the authentication performance compared to those methods using ACC data only. In addition, by employing multimodal learning, and fusing the two modalities, we are able to further improve EER and FAR comparing to any single modality. All the fusion models outperform GCF and ACC modalites, except in the socks leave-one-out experiment (last row in Table 2), where GCF performs better than the bimodal fusion model.

Another observation is that authentication in the socks study seems easier (with lower EER and FAR) compared to the shoes study. This could be attributed to the difference in the subjects participated in the study and the increased resolution of the shoe design that may make the authentication task harder, as more variance to the gait patterns is introduced. In addition, differences between shoes worn are eliminated in the shoes study, since all the participated subjects wear the same pair of smart shoes to collect data. While in the socks experiment, subjects are allowed to participate with their own shoes. A detailed discussion on this follows in subsection 7.7.

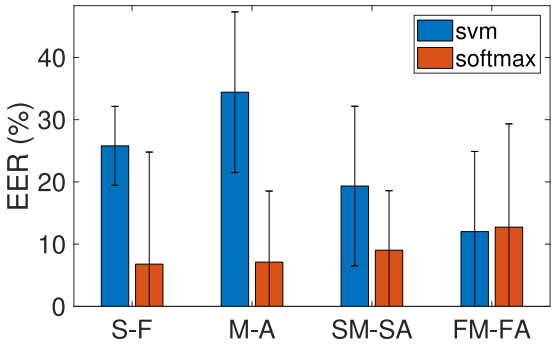
From the results, we also observe that the raw-stacked model, i.e. performing early fusion in the two modalities (Table 2), is better than the other fusion models, as any classifier seems to achieve lower EER with this model. This might indicate that early fusion may be sufficient to learn simple features in the time domain and develop authentication models. In addition, the proposed approach can be successfully applied to gait identification or recognition applications, where the goal is to determine the identity of the subject, given a new observation, based on a database of gait samples from a set of enrolled known subjects.

Finally, we observe increased FAR when compared to EER of the 10-fold CV from the first experiment, in both the shoes and socks studies. This performance degradation is expected, since the test data come from a subject that the auto-encoder and classification models have never seen before. It is important to note that most of the related work on gait-based authentication fail to report the leave-one-out FAR. We believe this experiment should be reported to have a complete and fair evaluation on the effectiveness of the proposed approach as typically in real-life scenario an impostor will not provide his training set to the gait-based authentication system.

A per-subject leave-one-out FAR performance evaluation is given in Fig. 9, for both case studies. Different rows in the tables are the corresponding subjects, for which the model is built. Each column corresponds to an imposing user, who tries to passively attack the corresponding model. From this set of results we observe that for the majority of victim-attacker pairs we get 0% FAR. However, there are specific pairs that generate FAR up to 21.7% and 24.7% in each of the case studies. In addition, there exists specific subjects (like subject 7 in the shoes study) that have worse FAR in their models compared to the others. This suggests that there may be easier and harder to target subjects. Based on this observation, we conduct further experiments in Sec.7.6 that evaluate what consequences there



(a) Socks study with varied (slow vs. fast) walking speeds



(b) Shoes study w/ varied speeds/fatigue (raw-stacked)

Fig. 10. Results from all testing scenarios in Table 1.

may be in the gait-based authentication system when an adversary can identify the best target and attempt mimicking the victim's gait.

7.4. Evaluation of parameters that affect gait

In this set of experiments we evaluate how the walking speed and fatigue level of the subject will affect the authentication performance. The first experiment focuses on the walking speed. For each of the two case studies, we build auto-encoder and classification models based on slow walking data and test the performance of the models against fast walking data. This is repeated with fast walking data being the training set and slow walking data being the test set. Data from both morning and afternoon sessions are used for training on the shoes study in this experiment. The average EER and its standard deviation are reported in Fig. 10a for the socks study and Fig. 10b (first two error-bars, S–F) for the shoes study, respectively. From the results we have the observations that for the raw-stacked model with softmax, the average EER is 4.56% for the socks study with a standard deviation of 6.55%, and the average EER is 6.79% for the shoes study with a standard deviation of 18.01%. SVM seems to perform much worse in this experiment as for the socks study the average EER is 12.22% with a standard deviation of 13.73% and for the shoes study, the average EER is 25.79% with a standard deviation of 6.35%. Compared with the observations from Expt. 1, the results from this experiment indicate that when a gait authentication model is trained only on one pace, the performance drops significantly compared to training the models with data that contain multiple walking speeds. Specifically, the models that are trained on slow gait and tested with fast gait samples, and vise versa, have generated statistically significant differences in the EER scores when compared to the EER scores from models trained on both slow and fast gait samples. The generated T-test p-value for the EER scores of the raw-stacked models with SVM between the two experiments is very small, i.e. 1.0023×10^{-148} for the socks data and 3.8557×10^{-129} for the shoes data. This concludes that the walking pace can greatly affect the authentication performance, and this is verified in both the socks and shoes studies. To reduce this performance degradation, the dataset that is used for training the user's model, needs to include multiple walking speeds. The more significant performance degradation in the shoes study can be attributed to the higher sensitivity of the smart shoes used for this study and the fact that subjects in this study are asked to wear the same pair of shoes. A more detailed explanation of this is given in Section 7.7. Next, we discuss how the fatigue level of the subject will affect the authentication performance.

Fatigue from daily activities is considered to be a factor that can affect gait Qu & Yeo (2011); Helbostad et al. (2007). To study the effect of reduced energy levels in the afternoon of the day, we perform a similar experiment to the previous one, with the difference being that the training data are from the morning session and the testing data are from the afternoon session in the shoes study. This is then repeated by taking afternoon session data as the training set and the morning session data as the testing set. The results are summarized in Fig. 10b. From the figure, we have the observation that from the raw-stacked with softmax model, the Average EER and standard deviations are 7.11% and 11.42%, respectively. Comparing to the first set of experiments focusing on the walking speed only, the EER increases by 6%. This indicates that fatigue does play an important role in gait-based authentication. In addition, slightly different sensor placement when the subjects wear the shoes for the afternoon experiment may be another factor for this performance change. To reduce performance degradation from such cases, the dataset used for training the user's model needs to include recordings from a set of subjects that have recorded their gaits in multiple different time-points of the day. By doing so, we can capture the day-to-day variance and variance from different sensor placement in shoe wearing.

To better estimate the effect of fatigue, we perform two more experiments by restricting the training and testing datasets to one walking pace, and the results are summarized in Fig. 10b. More precisely, the first experiment uses training and testing datasets from the slow pace morning (SM) and slow pace afternoon (SA) sessions, and the second experiment uses training and testing datasets from the fast morning (FM) and fast afternoon (FA) sessions. A comparison of the EER statistics can be found in Table 3. Although the performance is similar, it can be observed that the average EER is slightly better when only the fast walking pace data is used. This may be explained by the fact that fast walking can generate gait patterns that are more consistent since there is limited time in each gait cycle for deviations in the movement.

Table 3
The experimental results on the performance of the raw-stacked model. M and A refer to Morning and Afternoon recording sessions, respectively. μ and σ are the mean and standard deviation of the reported metric, respectively.

Experiment Set	Test scenario	Sensing	Metric	Classifier	μ (%)	σ (%)
2	Slow vs Fast (S–F)	Shoes	EER	softmax	6.79	18.01
2	Slow vs Fast (S–F)	Socks	EER	softmax	4.56	6.55
2	Morning vs Afternoon (M-A)	Shoes	EER	softmax	7.11	11.42
2	Slow M vs Slow A (SM-SA)	Shoes	EER	softmax	9.03	9.56
2	Fast M vs Fast A (FM-FA)	Shoes	EER	softmax	12.73	16.60

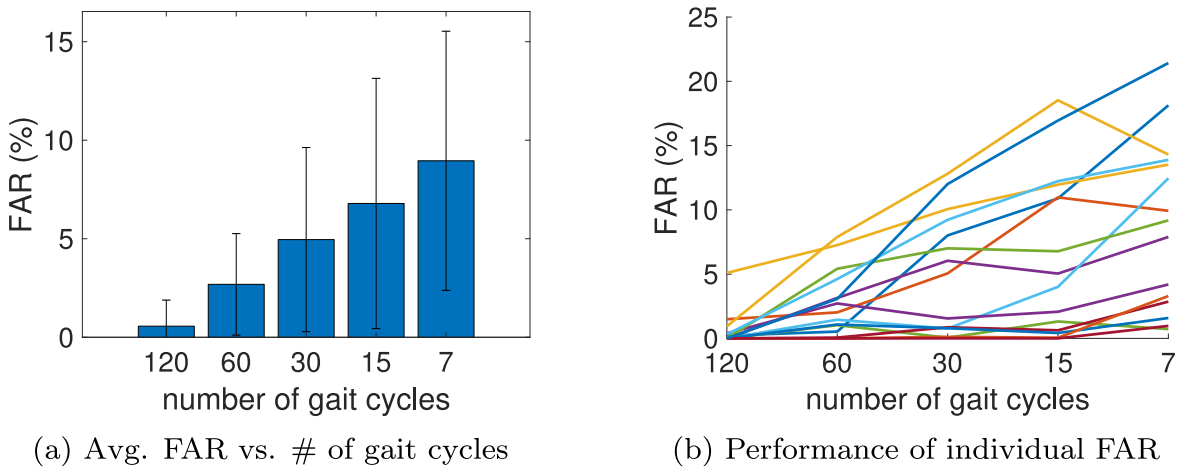


Fig. 11. Impact of the amount of training time on the authentication performance.

7.5. Evaluation on the impact of training time

The amount of data used for training a model can have a big impact on the performance. It is thus important to quantify how much training data a corresponding user needs to provide. In order to quantify that, we perform a similar set of experiments to those in Expt. 1 (Sec. 7.3), but in each iteration a different number of gait cycles is used to train the corresponding model. For a selected number of gait cycles, c, we pick c consecutive gait cycles for training and the rest are used for testing. For each value in c, the experiment is repeated at most 10 times, if there are enough different sets of c consecutive gait cycles. Average FAR results are summarized in Fig. 11. Fig. 11a the average results across all subjects, while Fig. 11a the results from individual subjects. Overall, we observe that FAR increases when the number of gait cycles used for training is reduced. In addition, from Fig. 11b we can observe that the performance depends on the subject as well. Some subjects seem to have their FAR fairly unchanged even when the smallest, i.e. 7, number of gait cycles is used for training. On the other hand, there are subjects that would benefit a lot from increasing their provided number of gait cycles for training their models. By also taking into account the performance changes in gait from parameters such as walking speed and time of the day, it is advised to new enrolled subjects to provide not only increased number of gait cycles for training, but also diverse samples in terms of walking speed and collected time.

7.6. Active attacks through gait mimicking

In this experiment, a human subject is asked to perform gait mimicking to evaluate the robustness of our method. Under this scenario, impostors try to mimic the gait patterns of their victims by generating similar mechanical body movement as their victim do. This scenario assumes the attacker has compromised the system's database and can use their gait patterns against the models belonging to other subjects in the database. By doing that, they are able to identify which victim's model gives the highest FAR. Once a victim is identified, the attackers are able to observe their victim's walking patterns visually, and then mimic the victim's gait patterns in order to increase their FAR even further.

In order to collect data for this gait mimicking scenario, we first identify candidate pairs of attacker-victim whose FAR is the highest compared to others as presented in Fig. 9. Since some subjects are not available to join the gait mimicking experiment, this pair (subject 6 and subject 9) shows the highest FAR among the available pairs. For this experiment, subject 9 was selected to be the victim, and subject 6 to be the attacker (impostor). The victim walks in a hallway similarly to what they are asked to do in the previous experiments. The attacker follows the victim by walking behind him and tries to match his gait patterns and mimic the victim's behavior by visual observation. Specifically, the attacker is required to mimic the victim's walking speed, step length and time duration of stance phase and swing phase. The walking speed for both subjects are the same as the fast walking scenario and the total experiment time is 5

Table 4
FAR scores with gait mimicking under different models. Subject 6 is the attacker and subject 9 is the victim.

Sensing Platforms	Metric	Raw stck	ACC	GCF	Stck	Bimodal
Smart Shoes	FAR	15.18	13.25	15.66	16.75	12.89

min. A camera is used to record the whole experiment and the shoe data is synchronized with the recorded video so that only the data which is within the walking speed range will be kept. In addition, the data are collected from both victim and attacker at same time.

Table 4 summarizes the gait mimicking results under different models. The FAR performance on the raw-stacked model with passive gait leave-one-out (Expt. 1) is 12.7%. After performing gait mimicking, FAR on the raw-stacked model is increased to 15.18%. Although the impostor is able to increase his chance to be accepted, the differences in the FAR scores between the passive and mimicking attacks are not statistically significant (p-value is 0.268).

7.7. Summary and discussion of the experiments

We now summarize our experimental results and findings. First, we observe that the proposed methods can be successfully applied for gait-based authentication. Specifically, it is possible to achieve very low EER and FAR, based on early fusion of data from the two modalities. This indicates that gait data do not require complex and higher order features to improve the authentication performance. Based on the results from the first experiment (Fig. 8) we observe that the average CV EER reaches as low as 0.01% for the smart socks platform and 0.16% for the smart shoes platform. To the best of our knowledge, this result is the best among all the related work, e.g., a 0.8% EER was reported in Sun & Yuao (2012). In addition, the leave-one-out FAR ranges from 0.54%, when using the smart socks to 1.96%, when using the smart shoes. Related studies have reported FAR performance of 3% with 11 subjects Trivino et al. (2010) and 6% with 32 subjects Trung et al. (2011). This indicates that our approach outperforms the literature with similar subject populations.

Another observation is the different authentication performance between the two case studies. In general the socks study yields lower EER and FAR compared to that of the shoes study. This result may be attributed to multiple reasons. First, in the socks study, participating subjects are allowed to wear their own shoes. This indicates that wearing different shoes will contribute in differentiating the gait patterns, and thus make authentication easier in the socks study. In addition, differences in the subjects population and the sensing technology itself may also contribute to the different authentication performance. Nevertheless, this difference may give a hint to understand the extend to which shoe types affect gait-based authentication. Finally, differences in the GCF sensing technology between the two sensing platforms may also play a role in the performance differences.

From both Expt. 2 and Expt. 3, we observe that there are multiple parameters that can affect the gait-based authentication performance. Walking parameters such as speed and time of the day may have a significant impact on the gait patterns. In addition, the amount of gait cycles used for training the corresponding models may greatly affect the performance as well. Based on all these observations, it is recommended that when users provide their training data at the enrollment phase, they should provide longer and more diverse gait examples in terms of both walking speed and collection time of the day.

8. Conclusion and future directions

Gait-based authentication has recently gained great attention in the research community since it has shown promising results towards bridging the gap between usability and effectiveness of authentication methods. In this work, we present our approach to improving the robustness of gait-based biometric authentication with the introduction of multimodal learning. With the use of commercially available smart socks and medical-grade research prototype of smart shoes as our sensing platforms, we jointly collect GCF and ACC data that are then passed through a pipeline of analytic methods for segmentation, feature extraction and classification. The use of early fusion on the sensing data with autoencoders for feature extraction, and SVM for classification is shown to be a very promising design that can capture the specific characteristics of each modality, correlate them and achieve superior performance compared to the methods only using a single modality.

There are still many open questions that need to be addressed before gait-based authentication systems can be adopted for everyday use on our personal devices. First, more data from more subjects need to be collected to further backup our findings that GaitCode is a feasible and secure biometric authentication method. In addition, the trade-off between the extra security offered by additional wearable sensors versus the additional cost needs to be further quantified. Finaly, biometric systems are vulnerable to different types of attacks, e.g. impersonation, replay, and spoofing. Muaaz and Mayrhofer (2017) gives an overview of all the possible vulnerable points in a generic biometric authentication system, including sensing devices, feature extraction modules, matchers, databases, and all communication channels connecting them. Impersonation attacks on gait-based biometric authentication systems can be a real threat to the security of the system and are very hard to model, since the sophistication level of an attacker and the resources available to them can vary. It will be our future work to overcome these vulnerable points to further improve the robustness of the proposed gait-based authentication system.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to

influence the work reported in this paper.

Acknowledgement

This work is supported in part by the National Science Foundation (NSF) under Grant No. CNS-1718738.

References

Al Abdulwahid, A., Clarke, N., Stengel, I., Furnell, S., & Reich, C. (2016). Continuous and transparent multimodal authentication: Reviewing the state of the art. *Cluster Computing, 19*, 455–474.
Banerjee, S. P., & Woodard, D. L. (2012). Biometric authentication and identification using keystroke dynamics: A survey. *J. Pattern Recognit. Res., 7*, 116–139.
Chen, C., Liang, J., Zhao, H., Hu, H., & Tian, J. (2009). Frame difference energy image for gait recognition with incomplete silhouettes. *Pattern Recognition Letters, 30*, 977–984.
Crouse, D., Han, H., Chandra, D., Barbello, B., & Jain, A. K. (2015). Continuous authentication of mobile user: Fusion of face image and inertial measurement unit data. *Proc. IEEE ICB Conf.*, 135–142.
De Luca, A., Hang, A., von Zezschwitz, E., & Hussmann, H. (2015). I feel like i'm taking selfies all day!: Towards understanding biometric authentication on smartphones. *Proc. ACM CHI Conf.*, 1411–1414.
Deng, W., Papavasileiou, I., Qiao, Z., Zhang, W., Lam, K.-Y., & Han, S. (2018). Advances in automation technologies for lower extremity neurorehabilitation: A review and future challenges. *IEEE Rev. Biomed. Eng., 11*, 289–305.
Derawi, M. O., Nickel, C., Bours, P., & Busch, C. (2010). Unobtrusive User-Authentication on mobile phones using biometric gait recognition. In *Proc. IEEE IIH-MSP conf.* (pp. 306–311).
Derlatka, M., & Bogdan, M. (2015). Ensemble kNN classifiers for human gait recognition based on ground reaction forces. In *Proc. IEEE HSI conf.* (pp. 88–93).
Gafurov, D., Snekkenes, E., & Bours, P. (2007). Spoof attacks on gait authentication system. *IEEE Transactions on Information Forensics and Security, 2*, 491–502.
Han, J., & Bhanu, B. (2006). Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 28*, 316–322.
Helbostad, J. L., Leirfall, S., Moe-Nilssen, R., & Sletvold, O. (2007). Physical fatigue affects gait characteristics in older persons. *J. Gerontol. A Biol. Sci. Med. Sci., 62*, 1010–1015.
Juefei-Xu, F., Bhagavatula, C., Jaech, A., Prasad, U., & Savvides, M. (2012). Gait-ID on the move: Pace independent human identification using cell phone accelerometer dynamics. *Proc. IEEE BTAS Conf*, 8–15.
Kale, A., Sundaresan, A., Rajagopalan, A. N., Cuntoor, N. P., Roy-Chowdhury, A. K., Krüger, V., & Chellappa, R. (2004). Identification of humans using gait. *IEEE Transactions on Image Processing, 13*, 1163–1173.
Kirtley, C., Whittle, M. W., & Jefferson, R. J. (1985). Influence of walking speed on gait parameters. *Journal of Biomedical Engineering, 7*, 282–288.
Kong, K., & Tomizuka, M. (2009). A gait monitoring system based on air pressure sensors embedded in a shoe. *IEEE, 14*, 358–370.
Liu, J., & Sun, W. (2016). Smart attacks against intelligent wearables in people-centric internet of things. *IEEE Communications Magazine, 54*, 44–49.
Lu, H., Huang, J., Saha, T., & Nachman, L. (2014). Unobtrusive gait verification for mobile phones. *Proc. ACM ISWC Conf.*, 91–98.
Mjaaland, B. B., Bours, P., & Gligoroski, D. (2011). Walk the walk: Attacking gait biometrics by imitation. In *Proc. SCCS ISC conf.*
Muaaz, M. (2013). A transparent and continuous biometric authentication framework for User-Friendly secure mobile environments. In *Proc. ACM UbiComp Conf.* (pp. 4–7).
Muaaz, M., & Mayrhofer, R. (2017). Smartphone-Based gait recognition: From authentication to imitation. *IEEE Transactions on Mobile Computing, 16*, 3209–3221.
Münzner, S., Schmidt, P., Reiss, A., Hanselmann, M., Stiefelhagen, R., & Dürichen, R. (2017). CNN-based sensor fusion techniques for multimodal human activity recognition. *Proc. ACM ISWC Conf.*, 158–165.
Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., & Ng, A. Y. (2011). Multimodal deep learning. *Proc. ICML Conf.*, 689–696.
Ngo, T. T., Makihara, Y., Nagahara, H., Mukaigawa, Y., & Yagi, Y. (2015). Similar gait action recognition using an inertial sensor. *Pattern Recognition, 48*, 1289–1301.
Nickel, C., & Busch, C. (2013). Classifying accelerometer data via hidden markov models to authenticate people by the way they walk. *IEEE Aerospace and Electronic Systems Magazine, 28*, 29–35.
Nickel, C., Wirtl, T., & Busch, C. (2012). Authentication of smartphone users based on the way they walk using k-NN algorithm. In *Proc. IEEE IIH-MSP conf* (pp. 16–20).
Olshausen, B. A., & Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision Research, 37*, 3311–3325.
Pan, S., Yu, T., Mirshekari, M., Fagert, J., Bonde, A., Mengshoel, O. J., Noh, H. Y., & Zhang, P. (2017). FootprintID: Indoor pedestrian identification through ambient structural vibration sensing. *Proc. ACM IMWUT Conf., 1*, 1–31.
Papavasileiou, I., Zhang, W., & Han, S. (2017a). *Real-time data-driven gait phase detection using ground contact force measurements: Algorithms, platform design and performance* (pp. 34–49). Smart Health.
Papavasileiou, I., Zhang, W., Wang, X., Bi, J., Zhang, L., & Han, S. (2017b). Classification of neurological gait disorders using multi-task feature learning. In *Proc. IEEE/ACM CHASE conf* (pp. 195–204).
Pataky, T. C., Mu, T., Bosch, K., Rosenbaum, D., & Goulermas, J. Y. (2012). Gait recognition: Highly unique dynamic plantar pressure patterns among 104 individuals. *Journal of The Royal Society Interface, 9*, 790–800.
Patel, V. M., Chellappa, R., Chandra, D., & Barbello, B. (2016). Continuous user authentication on mobile devices: Recent progress and remaining challenges. *IEEE Signal Processing Magazine, 33*, 49–61.
Qu, X., & Yeo, J. C. (2011). Effects of load carriage and fatigue on gait characteristics. *Journal of Biomechanics, 44*, 1259–1263.
Sensoria. (2020). http://www.sensoriafitness.com/.
Shen, C., Chen, Y., & Guan, X. (2018). Performance evaluation of implicit smartphones authentication via sensor-behavior analysis. *Informing Science, 430*, 538–553.
Sprager, S., & Juric, M. B. (2015a). An efficient HOS-Based gait authentication of accelerometer data. *IEEE Transactions on Information Forensics and Security, 10*, 1486–1498.
Sprager, S., & Juric, M. B. (2015b). Inertial sensor-based gait recognition: A review. *Sensors, 15*, 22089–22127.
Sun, H., & Yuao, T. (2012). Curve aligning approach for gait authentication based on a wearable accelerometer. *Physiological Measurement, 33*, 1111.
Trivino, G., Alvarez-Alvarez, A., & Bailador, G. (2010). Application of the computational theory of perceptions to human gait pattern recognition. *Pattern Recognition, 43*, 2572–2581.
Trung, N. T., Makihara, Y., Nagahara, H., Sagawa, R., Mukaigawa, Y., & Yagi, Y. (2011). Phase registration in a gallery improving gait authentication. In *Proc. IEEE IJCB conf.* (pp. 1–7).
Vildjiounaite, E., Mäkelä, S.-M., Lindholm, M., Riihimäki, R., Kyllönen, V., Mäntyjärvi, J., & Ailisto, H. (2006). Unobtrusive multimodal biometrics for ensuring privacy and information security with personal devices. *Proc. IEEE PerCom Conf.*, 187–201.
Watanabe, Y., & Sara, S. (2016). Toward an immunity-based gait recognition on smart phone: A study of feature selection and walking state classification. *Procedia Comput. Sci., 96*, 1790–1800.
Xu, W., Shen, Y., Zhang, Y., Bergmann, N., & Hu, W. (2017). Gait-watch: A context-aware authentication system for smart watch based on gait recognition. In *Proc. IEEE/ACM IoTDI conf.* (pp. 59–70).
Zhang, Y., Pan, G., Jia, K., Lu, M., Wang, Y., & Wu, Z. (2015). Accelerometer-Based gait recognition by sparse representation of signature points with clusters. *IEEE Trans. Cybern., 45*, 1864–1875.

Zhang, W., Tomizuka, M., & Byl, N. (2016). A wireless human motion monitoring system for smart rehabilitation. *Journal of Dynamic Systems, Measurement, and Control, 138*, 111004.
Zhong, Y., & Deng, Y. (2014). Sensor orientation invariant mobile gait biometrics. In *Proc. IEEE IJCB conf* (pp. 1–8).
Zhong, Y., Deng, Y., & Meltzner, G. (2015). Pace independent mobile gait biometrics. In *Proc. IEEE BTAS conf* (pp. 1–8).