



# User gait biometrics in smart ambient applications through wearable accelerometer signals: an analysis of the influence of training setup on recognition accuracy

Maria De Marsico<sup>1</sup> · Andrea Palermo<sup>1</sup>

Received: 30 June 2021 / Accepted: 13 March 2024  
© The Author(s) 2024

## Abstract

Gait recognition can exploit the signals from wearables, e.g., the accelerometers embedded in smart devices. At present, this kind of recognition mostly underlies subject verification: the incoming probe is compared only with the templates in the system gallery that belong to the claimed identity. For instance, several proposals tackle the continuous recognition of the device owner to detect possible theft or loss. In this case, assuming a short time between the gallery template acquisition and the probe is reasonable. This work rather investigates the viability of a wider range of applications including identification (comparison with a whole system gallery) in the medium-long term. The first contribution is a procedure for extraction and two-phase selection of the most relevant aggregate features from a gait signal. A model is trained for each identity using Logistic Regression. The second contribution is the experiments investigating the effect of the variability of the gait pattern in time. In particular, the recognition performance is influenced by the benchmark partition into training and testing sets when more acquisition sessions are available, like in the exploited ZJU-gaitacc dataset. When close-in-time acquisition data is only available, the results seem to suggest re-identification (short time among captures) as the most promising application for this kind of recognition. The exclusive use of different dataset sessions for training and testing can rather better highlight the dramatic effect of trait variability on the measured performance. This suggests acquiring enrollment data in more sessions when the intended use is in medium-long term applications of smart ambient intelligence.

**Keywords** Gait recognition · Wearable sensors · Relevant feature extraction

## 1 Introduction

The services in a smart city, as well as in a more restricted context of a smart home or office, technically rely on ubiquitous sensors and algorithms that transparently support automation, safety, and structural integration (Hizam et al. 2021). Among the enabling technologies and methods, biometric-based authentication has extended beyond securing highly protected or classified zones or services. It has quickly moved from applications oriented to authentication to ambient intelligence and user-tailored services (Barra et al. 2018). Of course, the basis is always the verification

or identification of a user. In verification, a claim of identity allows one to compare the template only with a reference template for that identity. The claim could be implicit. For instance, the owner of a smartphone generally is the only enrolled user for the device. Therefore, the submitted probe is compared with the only reference identity even without a claim. In identification, it is up to the system to determine the correct identity by comparing the probe with all enrolled users. This is the most common scenario in ambient intelligence, where the presence of a registered user can trigger automatic and transparent personalization and adaptation of the environment. This is much more comfortable and natural than having to issue a password or present a token. However, biometric traits must fulfill some requirements to be effectively used, including universality, uniqueness, permanence, acceptability, and robustness to circumvention (spoofing, i.e., an attempt to steal the identity of another person through a counterfeit biometric trait). Appearance-based traits, like face or popular fingerprints, are stronger in the

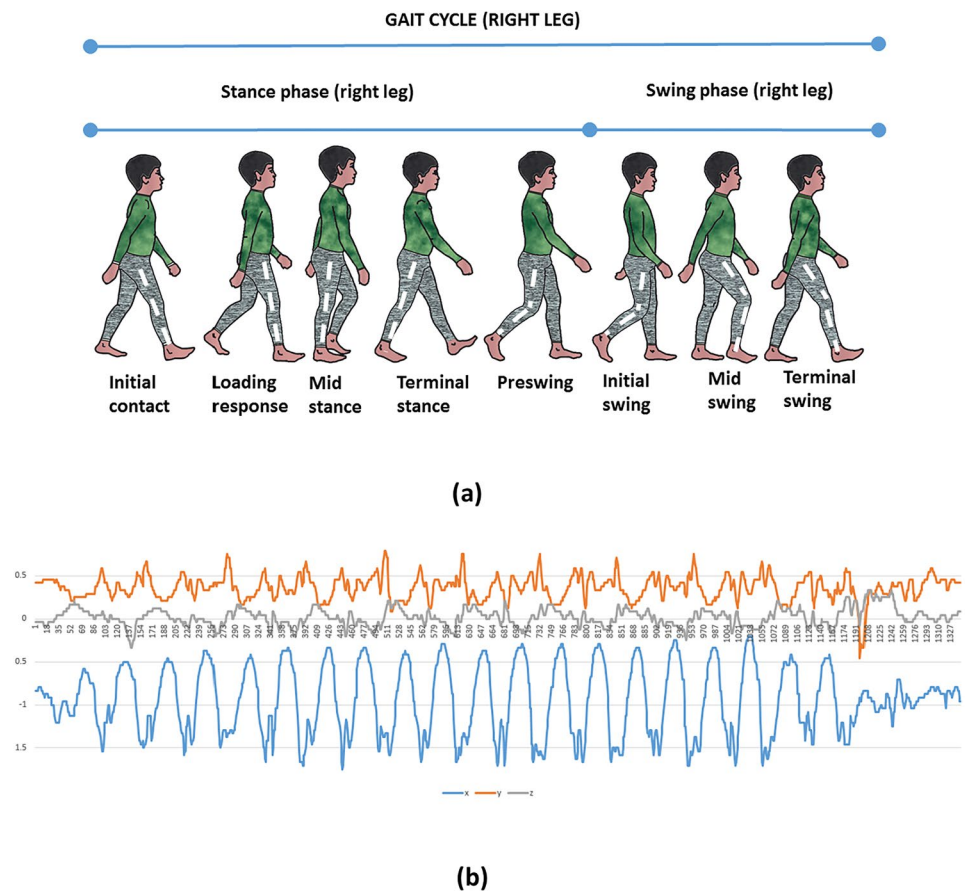
---

✉ Maria De Marsico  
demarsico@di.uniroma1.it

Andrea Palermo  
palermo.1810218@studenti.uniroma1.it

<sup>1</sup> Sapienza University of Rome, Via Salaria 113, Rome, Italy

**Fig. 1** **a** Stereotypical gait cycle and its phases. **b** Example of signal in ZJU-gaitacc dataset



sense of uniqueness and permanence but may be easier to reproduce (Kavita et al. 2021). On the other hand, according to several studies, behavioral traits, though lacking permanence, appear to be more robust to presentation attacks (submitting a false sample to the biometric system) (De Marsico and Mecca 2019).

It is appropriate to wonder whether gait can be considered as a biometric trait to reliably discriminate among different subjects. The basic walking characteristics and the basic kinematic patterns are clearly driven by common stereotypes (Borghese et al. 1996), but the individual energy-saving strategies produce qualitative and quantitative differences that can make individual walking styles well recognizable. The analysis of electromyographic (EMG) signals (Pedotti 1977) shows “that locomotion cannot be considered as a completely stereotyped function” because, “despite the similar kinematics, the torque time courses of different subjects present significant differences in agreement with different temporal sequence of muscle activation.” Despite the common dynamic aspects of the gait pattern, precise kinematic strategies (Bianchi et al. 1998) explain the behavioral variations able to produce individual differences. These possibly support biometric

identification. In more detail, a gait sequence (gait cycle) is conventionally divided into stance and swing phases, each further divided into phase components. A cycle is usually referred to either the right or left leg (in normal conditions the two cycles are symmetrical). Figure 1a shows the reference leg by the light dashed line.

Computer vision-based techniques still represent the most active research line to tackle the problem of recognizing an individual walking pattern. However, they suffer from the well-known problems of trajectory crossing, total or partial occlusion and variation of the point of view also affecting other video-based applications. The increased presence of embedded sensors in smart devices has more recently inspired the investigation of alternative methods in which the gait dynamics are captured in the form of time series of, e.g., acceleration values. The accelerometer collects one signal per each 3D space axis: the data recorded at time  $t_i$  is the triplet  $\langle x_i, y_i, z_i \rangle$  of acceleration values measured on the three axes. It seems to provide better performances than other wearable sensors for recognizing the subject carrying the device, by also avoiding the problems related to video processing. For instance,

the problem of trajectory crossing is automatically solved since signals are acquired singularly. Figure 1b shows examples of signals captured by the accelerometers used to collect ZJU-gaitacc dataset<sup>1</sup> (Zhang et al. 2015), that is used for this work and will be described in more detail later.

The reverse of the medal is that these time series suffer from a significant variability, especially related to different capture times and other conditions like ground slope or kind of shoes.

This paper proposes two contributions:

- a novel automatic procedure to extract the most relevant features from accelerometer data, in the attempt to distill the invariant dynamic regularities present in the individual walking pattern;
- the investigation of the critical influence of the training/testing benchmark partition when different capture sessions are available, as for the exploited ZJU-gaitacc dataset; the experimental results are expected to hold in general for behavioral biometric traits when they are not used in the traditional way, i.e., for continuous re-identification of the owner of a smart device, but in a wider smart ambient context.

The achieved results are compared with other approaches tested over the same dataset, considering verification and identification closed-set (all probes are assumed to belong to known/enrolled subjects). The novelties of the proposal can also be summarized as follows:

- Recognition: the approaches to gait recognition via wearables mostly exploit the accelerometer signal (e.g., segmented into cycles, steps, or fragments), or features like the FT coefficients or MEL/BARK cepstral coefficients; fewer approaches use aggregate features extracted from the signal; most of all, the double-step feature selection with Principal Feature Analysis (PFA) instead of much simpler variance analysis improves the quality of the final selection.
- Training: the imbalance in the one-vs-all approach to model learning is solved by resampling.
- Analysis: the training/testing partition is an obviously important step in experimental setup; we further assess the importance of considering the different acquisition sessions, if available, in such partition and analyze its influence; this aspect is mostly neglected in the compared literature and, in general, avoiding a session-based partition improves the results, as the experiments demonstrate; most datasets except ZJU-gaitacc (the one that the work exploits), at the best of our knowledge, include

a single capture session: this compromises the generalization of results to evaluate gait recognition over time, e.g., to provide automatic access to user-granted facilities; a thorough comparison is mostly impossible, due to the use in most literature works of proprietary datasets, either not publicly available or containing a low number of subjects, therefore the analysis deals with works using ZJU-gaitacc.

The paper continues as follows. Section 2 introduces some basic concepts and terms related to biometrics and to gait recognition. Section 3 briefly presents a possible scenario for the use of wearable device-based gait recognition in a smart ambient. Section 4 summarizes some works related to wearable-based gait recognition. Section 5 describes the steps of feature extraction and selection. Section 6 discusses the training procedures and the related issues. Section 7 presents the results of the experiments, and in particular highlights the importance of a fair partition when multiple-session datasets are exploited. Section 8 draws some conclusions.

## 2 Gait as a biometric trait

Before tackling the main paper topic, it is worth reminding some basic concepts related to biometric systems and their performance evaluation. The section continues with an introduction to gait as a biometric trait.

### 2.1 Biometrics at a glance

A biometric sample is a digital information obtained from a biometric capture device, for instance, an image of a face or a fingerprint, or another type of signal or time series representing behavioral traits, for instance, voice or keystroke patterns. Samples are generally not used directly, but a template is built from each sample. This can include a feature vector, collecting the values of hand-crafted (selected by the system developer) characteristics. Alternatively, the embeddings computed by convolutional neural networks (CNNs) can be used as templates. In these cases, the popular distance or similarity measures can be used to estimate whether two templates belong to the same subject. Reference templates of registered users make up the system gallery. The templates extracted from incoming samples to recognize make up the probe set used for performance evaluation. As a further possibility, feature vectors or embeddings extracted as above can be used to train a model for each subject using machine learning methods. This returns a classification response when an unknown probe is submitted to the system, after the same template extraction is performed. Once decided the structure of biometric templates, they can be extracted with the same procedure also from unknown samples.

<sup>1</sup> <http://www.cs.zju.edu.cn/~gpan/database/gaitacc.html>.

Biometric applications can be classified according to the number of comparisons carried out to recognize a subject. The template extracted from the submitted probe sample is compared against the templates in the system gallery to either obtain the confirmation of the claimed identity or the identification of an “anonymous” subject. If the subject claims an identity, the identity verification entails a 1:1 comparison (in terms of identities) between the template(s) of the unknown subject and the templates of the claimed identity. The identity claim can be implicit when a single subject is enrolled. For instance, when the owner of a smart device records his/her biometric template (e.g., fingerprint or face), all the following verification operations take this template as a reference. A distance/similarity threshold determines either an Accept or Reject response for the claim. For a certain acceptance threshold, the False Non Match Rate (FNMR) [often substituted by False Rejection Rate (FRR)] expresses the probability of rejecting a legitimate user by computing the rate of false reject responses to the total number of genuine probes (that should have been accepted). Even though they are often used interchangeably, the second is different since it also counts the number of operations where rejection was not caused by the algorithm but by other causes, for instance, a failure to enroll the probe. For a certain acceptance threshold, False Match Rate (FMR) [(often substituted by False Acceptance Rate (FAR)] expresses the probability of accepting an impostor user by computing the rate of false accept responses to the total number of impostor probes (that should have been rejected). In the following, we will use the more popular FRR and FAR since, in our case, we do not have rejects other than those produced by the comparison, and therefore the metrics are equivalent. Their curves for different thresholds intersect in the point of Equal Error Rate (EER), at the threshold producing  $FAR = FRR$  (the lower, the better). Other performance-related items are the Receiver Operating Characteristic (ROC) curve and its Area Under Curve (AUC) (the larger, the better).

Without any identity claim, the identification entails a 1:N comparison between the template(s) of the unknown subject and all the templates in the gallery. The returned result is the closest identity, and, especially during testing, the entire gallery is ordered by similarity with the probe. Identification can be Closed Set or Open Set. In the first case, the system assumes each probe belongs to an enrolled subject. The main performance measure is the Recognition Rate (RR), i.e., the rate of correct identities returned as the most similar ones. Further information on system performance can be derived by the Cumulative Match Characteristic (CMC) curve, plotting the values of Cumulative Match Score at rank  $k$ —CMS( $k$ )—i.e., the rates of correct identifications within the  $k$ -th list position. The lower the rank where the CMS( $k$ ) reaches the value of 1, the better. Identification can be Open Set, too. A Reject response is returned

when, according to an acceptance threshold, the probe is not recognized as belonging to an enrolled subject. Detection and Identification Rate at rank 1 for threshold  $t$ — $DIR(1, t)$ —expresses the rate of probes belonging to enrolled subjects that have been returned in the first position of the ordered list and whose similarity meets the threshold;  $FRR(t) = 1 - DIR(1, t)$ , while  $FAR(t)$  is the rate of False Accepts (with any identity) for threshold  $t$  to the number of probes belonging to unregistered subjects. FAR and FRR curves determine EER and allow to compute the ROC and its AUC.

It is to point out that performance evaluation is done over a benchmark, including a ground truth for the recognition, i.e., each sample is annotated with its correct identity. In real operations, the ground truth is not available. This is why it is important to thoroughly evaluate the system before deployment in the real world.

## 2.2 Gait recognition approaches

Gait is one of the biometric traits that can be used for recognition. Three main tracks of research relate to the kind of sample that they take into account. The approach that is by far the most popular relies on computer vision. Interesting surveys appear in Lee et al. (2014), Prakash et al. (2018) and Joseph Raj and Balamurugan (2017). Computer-vision-based techniques are often used for re-identification and identification in video surveillance applications. Another approach exploits sensor-equipped floors to capture gait features. Three proposals about floor sensor-based recognition can be found in Orr and Abowd (2000), Suutala and Rönning (2004) and Middleton et al. (2005). This approach, though being quite expensive and requiring a dedicated setup, is generally used only for gait analysis in the biomedical field (Kamruzzaman and Begg 2006; Klucken et al. 2013; Muro-De-La-Herran et al. 2014). The third line exploits wearable sensors. An extensive survey of related proposals can be found in De Marsico and Mecca (2019). The experiments presented in literature mostly exploit data captured in a single session so that the natural variation of the gait pattern is somehow reduced. To this respect, it is worth underlining that behavioral traits, though unique, lack permanence in the medium-long term. Natural variability is caused by different conditions, e.g., voice distortion due to a cold or gait variations due to speed or ground slope. Therefore, they are mostly used for continuous re-identification in the short term, e.g., to assure to be in the presence of the same subject during a communication session, or for (implicit, continuous) verification of a device owner. For instance, methods based on wearable signals are mostly exploited in this way. Evaluation using data from the same capture session effectively addresses this kind of applications. Differently from short-term recognition, this work specifically explores possible issues related to smart city-related case studies, e.g., a

transparent authentication to access controlled city spaces. The definition of “transparent” is because the user does not need to issue any identity claim, and the identification is based on the recognition of the personal gait dynamics among those of the enrolled users. Of course, a reliable evaluation of the system performance requires using a dataset acquired in more sessions at different times and a session-based fair partition of such dataset into training and testing sets. While this issue may appear trivial, the experimental results demonstrate that, for multi-session datasets, it must be taken into careful account. The reverse of the medal is that separating the data from different sessions can reduce the number of training samples when using machine learning training methods. A typical example is when we have only two sessions, like in our case, and either training contains one of the sessions plus part of the other, or both training and testing contain a mixture. This allows to create a larger training set. However, the possibility during testing to compare templates extracted from samples captured at a short time distance from each other or from the training data is non-negligible. This causes better results than comparing samples systematically captured at different times. Lower but more realistic results are obtained by training the system with data separated in time from the testing data, even though this means using less training data.

### 3 A possible biometric identification for smart cities: walk & go

The term *Smart City* defines an (urban) area that, thanks to technology, is much more safe and livable with respect to traditional urban settings. This is achieved through several innovative solutions: intelligent agents support and simplify everyday activities; the use of resources and the transmission of any kind of information is optimized; building design, transportation networks, and all services meet advanced criteria of efficiency and effectiveness; services are customized according to personalized profiling (Barra et al. 2018). This epochal revolution is made possible by Internet of Things (IoT), and more recently Internet of Everything (IoE), by cloud computing, and by smart and mobile devices. Interconnected sensors of several kinds, both wearable and ambient-specific, make up a net that continuously collects context information (Barra et al. 2018). Such information includes data about the physical ambient but also about people, their actions, and context-relevant objects and devices. This huge and varied amount of data requires distributed intelligent systems for both collection and processing in order to trigger appropriate actions. These are executed by suitable actuators, either agents or services. In these technologically advanced scenarios, biometric recognition plays a crucial role in assessing the identity of a person. This can either

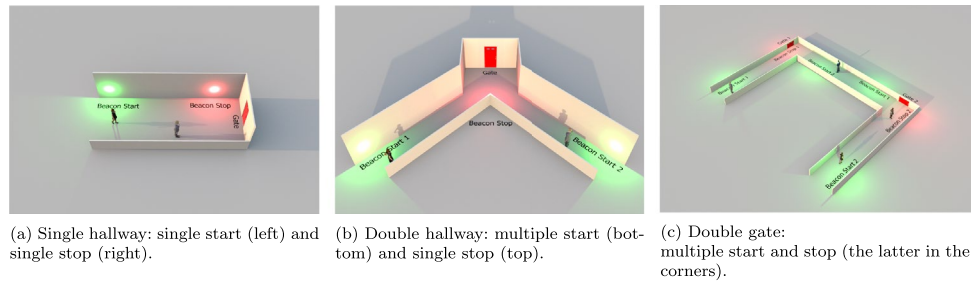
(more or less transparently) grant access to resources/spaces or provide user-customized services (De Marsico et al. 2011; Abate et al. 2011). Authenticated access to restricted areas via biometrics can overcome or complement traditional solutions like passwords or tokens. Biometric systems just require one to show one’s trait, e.g., the face or the fingerprint (probe), that will be compared with the system collection (gallery) of registered (enrolled) users. The reverse of the medal is represented by the so-called “in-the-wild” conditions that often affect biometrics-related operations in smart cities. The acquisition of user samples, e.g., the face, is often unattended, i.e., it is often the case that no operator guides the user for a correct capture. For instance, mobile biometrics implemented through personal devices play an increasing role: users can autonomously exploit their own devices to capture their physical and behavioral features for authentication instead of passwords or cards, and this can be achieved even without an explicit interaction given that the devices run appropriate applications. In addition, sensors should interfere in everyday life the least possible, with citizens possibly unaware of eventual *smart technologies*. As a further issue in uncontrolled and unattended settings, it is possible that the acquisition conditions are not always the best possible, and they can even be adverse, e.g., for uneven illumination or user pose. These challenges call for more reliable and robust biometric algorithms, and for more trustworthy testing providing fairly generalizable results.

The use of personal mobile devices is becoming increasingly popular. The embedded sensors can provide many different user information. Accelerometers and other sensors capture dynamic information about motion and actions. These “sightless” sensors (Kanev et al. 2016) can be used for biometric recognition. While smartphones are the most popular and widespread devices embedding them, the increasing miniaturization allows more options. For instance, wrist-worn devices like smartwatches are a possible alternative. In the case of gait, the natural arm swing can interfere with the walk signal. However, in a normal condition of regular and continuous walking, the swing pattern could also be part of the individual kinematic strategy. Tablets are a further possibility, though being quite impractical for the envisaged type of application. Some recent works suggest embedding sensors in shoe soles (Brombin et al. 2019). Finally, the new frontier is represented by biometric jackets.<sup>2</sup> Other behavioral traces extracted by mobile devices can provide information about users’ habits and social behaviors, still respecting the users’ privacy and the data’s security (Kröger et al. 2019). As anticipated, mobile biometrics raise both new opportunities

<sup>2</sup> Biometric Jacket—Arduino-controlled raincoat with a bunch of sensors and lights embedded and sown into it. <https://www.hackster.io/it-worked-yesterday-i-swear/biometric-jacket-975db8>.



**Fig. 2** Three example scenarios: green and red are devices triggering start and stop of capture



and challenges: the quality of ubiquitous and easy data capture can be hindered by unattended acquisition and by possibly adverse conditions.

This work focuses on biometric recognition through the individual gait pattern. Video-based approaches traditionally tackle gait, but more recently, the increased embedding of accelerometers and other motion sensors in mobile devices has encouraged using gait acceleration dynamics instead (De Marsico and Mecca 2017). Video-based approaches traditionally present some limitations that can be overcome, e.g., camera perspective and trajectory crossing/occlusion. However, new issues arise, e.g., the ground slope and the walking speed. It is worth underlining that, in wearable-based approaches, the sensor is embedded in user equipment. Therefore, the signal is univocally identifiable, even in a crowd, because it is directly transmitted from the individual device and, differently from video traces, there is no possibility of overlap or crossing. It is also interesting to notice that if no cyber attack is launched, the user is aware of the recognition since authentication requires a suitable app. Therefore, there is little risk of unauthorized data capture. At the same time, the user is not necessarily aware of the exact moment of authentication that can be triggered transparently by unobtrusive ambient elements. A possible architecture adds these few elements to the personal mobile devices, together with a processing facility. For instance, beacons (low energy Bluetooth devices) can send their code to a receiving apparatus (the mobile device). These codes are associated in the app with either start or stop signals that trigger start recording or stop & send the recorded acceleration time series to the authentication server, or to a cloud service. As a response, if the user is identified as an authorized one, suitable doors or services can automatically open or start. The most realistic kind of recognition in this scenario is the identification open set. However, an identification closed set can be used as well in a context where, for example, a two-phase recognition is carried out: a preliminary identification open set or verification, carried out through possibly different biometric traits or through more traditional methods, may allow access to a controlled zone to a limited (closed) set of subjects. Afterwards, further access to a sub-zone or single ambient can be granted by just identifying the approaching subject (identification

closed set) without any further explicit request by the subject itself. In summary, wearable-based gait recognition is flexible and usable enough for different scenarios, e.g., for entire government buildings, banks, etc., and individual restricted areas, e.g., offices (Figure 2).

It is worth underlining that the above approach is definitely different from identifying the subject through his/her device using, e.g., RFID or other proximity technologies. This kind of identification is broken if the device is stolen. On the contrary, studies in literature demonstrate that the individual gait pattern is very difficult to spoof, i.e., it is very difficult to imitate the gait pattern of an authorized user, also for trained impostors (Muaaz and Mayrhofer 2017).

#### 4 Some work related to wearable device-based gait recognition

Thanks to miniaturization, accelerometers can be embedded in wearable and smart devices, e.g., a smartphone. Due to the wide availability they are presently the most used sensors for gait recognition. The three recorded acceleration time series can contain relevant elements to characterize an individual walking style. In some approaches, the whole signal is substituted by relevant features that are extracted from it and used for further processing.

It is possible to identify two main categories of methods that use acceleration time series. The works in the first category exploit signal matching algorithms. Simple distance measures, like Manhattan or Euclidean (Gafurov et al. 2010), can dramatically suffer from time misalignment, especially without a preliminary interpolation and for signals captured at different times. Therefore, they are rather exploited by the variations of Dynamic Time Warping (DTW). In order to tackle the problems of gait signals including a different number of steps, some approaches preliminarily segment the signals to match into steps (De Marsico and Mecca 2017) or cycles (Derawi et al. 2010; Rong et al. 2007; Fernandez-Lopez et al. 2016) (a cycle is composed by a pair of steps) that are compared afterward, while fewer ones compare unsegmented signals.

The second category of methods exploits Machine Learning. These either work on fragments or on features extracted

from them. Differently from the above, in these cases a fragment (or chunk) has not a kinematic basis but is rather a part of the walking signal of fixed time length or number of time series points. A Hidden Markov Model (HMM) is computed for each user in Nickel et al. (2011b). The acceleration data from each fragment is used in Nickel et al. (2011a) to build a feature vector of fixed size. Features are mostly statistical ones, e.g., mean, maximum, minimum, binned distribution, etc., with the addition of the Mel and the Bark frequency cepstral coefficients. Vectors of these features are used to train a Support Vector Machine (SVM) for each user. Similar vectors are used in Nickel et al. (2012) to train a k-NN approach with fragments of different length.

Of course, it is also possible to find in literature approaches based on Neural Networks. IDNet (Gadaleta and Rossi 2016) is an authentication framework relying on Convolutional Neural Networks (CNN). The CNN is used as feature extractor, and then One-Class SVM (OSVM) is used for verification.

Unfortunately, the experiments in most of the first proposals rely on in-house collected datasets, seldom available to the research community. On the contrary, a large wide dataset freely available is ZJU-gaitacc (Zhang et al. 2015) that represents a shared benchmark to compare the results of different approaches, even though the accelerometers used to collect the data are quite dated and the signals are already interpolated. The dataset will be described in the experimental section, and the results reported by the authors will be compared with those achieved in this work, together with other works using the same data. The authors also propose an approach to recognition based on Signature Points. The data are converted into magnitude vectors. The Signature Points are taken as the extremes of the convolution of the gait signal with a Difference of Gaussian pyramid, that is the points that are greater or smaller than all of their eight neighbors. Signature Points are stored as vectors and clustered, linearly combined and saved into a dictionary, therefore obtaining a single element for each cluster. The system considers recognition as a conditional probability problem and uses a sparse-code classifier. The results are very interesting and reach an up to 95.8% of RR (identification), and a down to 2.2% of EER (verification), even if it is worth considering that the approaches in both works fuse the results of 5 accelerometers worn in different body locations. The best results achieved by an individual accelerometer are 73.4% of RR and 8.9% of EER.

The work in Giorgi et al. (2017) exploits Deep Convolutional Neural networks for subject re-identification and exploits the ZJU-gaitacc dataset to test the performance. The processing includes three steps: cycles extraction, filtering, and normalization. The cycle extraction exploits the significant changes of values (peaks) on the z-axis caused by the heel impact with the ground. The deep network architecture

includes two convolutional layers, a max pooling layer, two fully connected layers, and a final softmax layer for template classification. The tests use 5 out of the 6 walk signals of a session per subject as training set, and one for test experiments. Using only a single session in the dataset causes signals from the same subject to be relatively similar, but is consistent with the re-identification application. Overfitting is prevented by creating artificial training samples. The re-identification accuracy of the proposed scheme is quantified as the average number of correctly recognized cycles for each identity, i.e., appearing in the first, second, and third place of the list of results ordered by similarity.

More references and a more detailed description of the mentioned works is in De Marsico and Mecca (2019).

## 5 Feature extraction

The procedure of feature extraction from the accelerometer temporal series exploits the *tsfresh*<sup>3</sup> library. The selection of the most relevant features, the training procedure and the classifier test exploit the *scikit-learn*<sup>4</sup> library and, for some specific tasks, *imbalanced-learn*<sup>5</sup> (all written in Python 3.8). The extracted features consist of aggregate values obtained from the time series; there is an acceleration time series for each axis x, y, and z, so every possible feature is computed relative to a single axis. This process returns 763 features per axis for a total of 2289 features. Obviously, it is not possible to list them all: only some of those selected as the most relevant will be described below; a complete list is available in the *tsfresh* documentation.<sup>6</sup>

Once a single vector of features represents each walk, the proposed method selects the most relevant ones for recognition. In fact, the training and testing of classifiers with vectors consisting of the full set of 2289 characteristics achieved definitely unsatisfactory results (more than 50% less), demonstrating that many of them were either irrelevant to the case at hand or redundant or misleading. The implemented two-phase selection procedure allows for significant reduction of the number of features, and a final scaling allows fair participation of all of them in the classification.

**First phase** Regarding the first phase, two strategies were compared. The first one further exploits the *tsfresh* library, which provides functions to identify the most relevant characteristics in a set using a kind of supervised approach, that takes into account the identity associated with each vector. Actually, it would be possible to unify this phase and the

<sup>3</sup> [https://tsfresh.readthedocs.io/en/latest/api/tsfresh.feature\\_extraction.html](https://tsfresh.readthedocs.io/en/latest/api/tsfresh.feature_extraction.html).

<sup>4</sup> <https://scikit-learn.org/stable/>.

<sup>5</sup> <https://imbalanced-learn.org/stable>.

<sup>6</sup> [https://tsfresh.readthedocs.io/en/latest/text/list\\_of\\_features.html](https://tsfresh.readthedocs.io/en/latest/text/list_of_features.html).

previous extraction to directly obtain only the most relevant features, 1421 out of the 2289 initial ones.

The second strategy exploits a two-step deeper feature analysis and achieves a better selection. The obtained final vectors have lower dimensions and allow better classification results. In this second strategy, the first step discards all 0-variance (constant) features, i.e., those with no useful information for classifier training. The output of this step is a decreased set of 2158 out of the initial 2289. The next step tries again to determine the relevance of the different features with a different approach though still based on variance. Using the *scikit-learn* function *SelectPercentile* only 45% of the analyzed features are maintained, based on a score assigned by the *ANOVA F value* function. This function assigns the F-score to each feature according to the following formula:

$$F = (\text{inter-group variance})/(\text{intra-group variance}) \quad (1)$$

where “inter-group variance” indicates the squared average of the distances of the average values of the individual features from the global average value of all features of all samples; “intra-group variance” is the normal variance of the features seen above. At the end of this computation, the final feature vector includes the highest scored 45% of the input features. This procedure is generally suited to compute the relevance of quantitative features. After this step, the number of features passes from 2158 to 971.

**Second phase** The second phase relies on a further feature selection based on Principal Feature Analysis (PFA) (Lu et al. 2007). It could be worth underlining the difference between PFA and the more familiar Principal Component Analysis (PCA). In both cases, the dataset is interpreted as a matrix  $X$  of dimension  $m \times n$ , consisting of  $n$  (walk) vectors  $x_i$  of size  $m$  (features). Then the matrix is transformed into a new matrix  $Z$ , aimed at bringing to 0 the average of the values of all the columns of  $X$ : this is obtained by subtracting the average of the corresponding values from each value of the whole column. PCA works by transforming the space of characteristics into one of smaller dimension  $q$  whose axes are represented by the  $q$  eigenvectors associated with the highest eigenvalues of the covariance matrix obtained from  $Z$ : they have no correlation with the features initially present in the dataset. The further steps entailed by PFA further determine the most relevant projections of the original features in the new space (represented by the rows of the transformation matrix). At the end of the second phase, the number of surviving features is 400 (out of the initial 2289).

**Feature scaling.** Once the most relevant features have been selected for the classification, it is necessary to normalize their values. This is essential as, even after filtering, there is a large number of features anyway: if they get values in different scales, such values may not be interpreted

correctly for the purpose of training the classification model. The formula used for normalizing a feature value  $x$  is the familiar one:

$$\frac{x - \mu}{\sigma} \quad (2)$$

where  $\mu$  is the average of all values taken by the feature across the vectors and  $\sigma$  is their standard deviation.

## 6 Classifier training

### 6.1 Upsampling of the minority class

The chosen strategy has been to train a classifier for each identity. This allows to easily add subjects to the gallery by simply computing the model associated with their identity. On the other hand, these are typical conditions where, having a subset of samples for each of several subjects, the individual training sets are highly unbalanced due to the huge amount of vectors from the other identities. The classical solutions to this problem are based on resampling (Cervantes et al. 2020) according to two different methods. The first one entails downsampling, i.e., discarding as many samples of the over-represented class(es) up to reach the desired proportion. In Machine Learning applications, due to the value of data, since each sample can take precious information, this is seldom applied. The second method entails upsampling, i.e., increasing the number of samples of the minority class. The strategies normally used for this purpose are varied: the least elaborate and best known is the one that simply duplicates some instances chosen within the minority class; its only advantage, however, is to balance the distribution of the instances, but without bringing any additional information to the training set. This is why it was decided to adopt another strategy for the purposes: it consists of the application of SMOTE (Synthetic Minority Oversampling Technique) (Chawla et al. 2002), which is capable of creating synthetic instances of the minority class starting from those already present. The exploited implementation is the one provided by *scikit-learn* in the library *imbalanced-learn*. In short, the algorithm is based on a very simple idea: it begins by randomly selecting an instance  $I$  in the minority class and looking for the  $k$  closest instances in the same class in the space of characteristics ( $k$  is a parameter of the algorithm, in this case 5 was used). Then it randomly chooses one of these  $k$  neighbors and connects it to  $I$  through a segment; the new synthetic instance is obtained by randomly selecting a point belonging to the segment (obviously different from the extremes). Figure 3 shows an example of the algorithm processing.



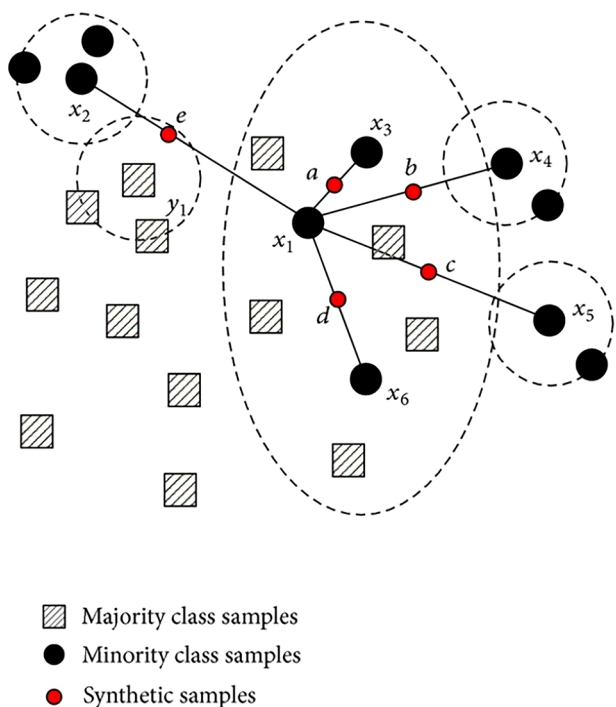


Fig. 3 Visual example of SMOTE processing

### 6.2 Logistic Regression and final calibration

After balancing the positive and negative samples for each model training, the output is predicted by Logistic Regression. Logistic Regression uses an equation as the representation, in a way very similar to linear regression. Input values are combined linearly using weights or coefficient values to predict an output value. A key difference from linear regression is that the output value being modeled is a binary values (0 or 1) rather than a numeric value (see Fig. 4a). This is obtained by combining simple linear regression with a logistic function as follows:

$$P = \frac{1}{1 + e^{(\beta_0 + \sum_{k=1}^n \beta_k x_k)}} \tag{3}$$

where  $P$  is the predicted output,  $n$  is the dimension of the input,  $\beta_0$  is the bias and all the  $\beta_i$  are the coefficients for the input values ( $x_i$ ). During training the model is learning the  $\beta$  coefficients for each column in the input data from the training data. This technique is generalized to the multiclass classification problem by using a one-vs-rest approach.

The model obtained for each user during the training phase is used to make up the system gallery. A final step concerns a further calibration of the decision function; for the purposes of system testing, it is not enough that the models predict the class of a sample, but also a corresponding probability is needed. In this context, every machine learning algorithm produces values that are calibrated in different ways, often not between 0 and 1. The goal of the calibration is to ensure that the predicted probabilities can be directly used as confidence values: this means, for example, that approximately 60% of the instances whose predicted probability is close to 0.6 will actually have to belong to the positive class. The technique used for calibration is isotonic regression. It searches for the non-decreasing broken line as close as possible to the observations (see Fig. 4b) The values obtained from the above computation are used as similarity values. In verification, a single similarity value is obtained per testing operation, which is matched against a threshold to determine the final response. In identification closed-set, the values obtained by comparing the probe with each gallery model, in turn, are ranked in decreasing order to return the recognized identity and also the ordered list of the other gallery models. Re-identification entails determining whether the accelerometer is worn by the same person during a recognition session. It can be considered a special case where the model and the probe are captured close in time.

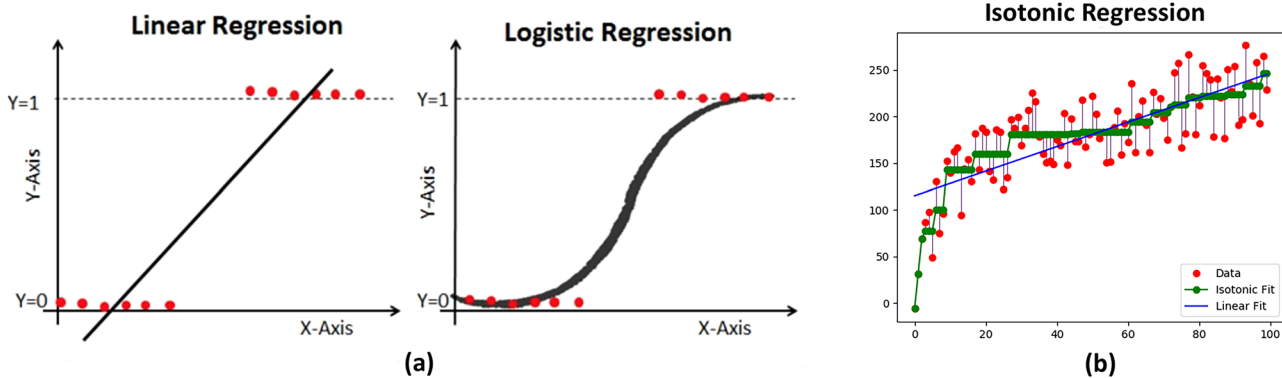


Fig. 4 a Linear vs. logistic regression. b Isotonic Regression

## 7 Experimental results

### 7.1 ZJU-gaitacc dataset

The dataset ZJU-gaitacc<sup>7</sup> (Zhang et al. 2015) collects gait signals from 175 subjects, out of which 153 participated in two capture sessions divided in time with 6 walks each. The delay between two capture sessions for the same subject ranges from one week to six months. The number of walks and the two different sessions clearly separated in time make the results obtained on the dataset more generalizable regarding intra-personal variations and particularly suited for the kind of analysis that this paper proposes. The remaining 22 subjects have 6 walks and can be used as impostors. Unfortunately, no individual demographic information annotates the data. Walk signals have a sufficiently high number of points in the time series (about 1400) being collected along a hallway 20 ms long. 5 Wii Remote controllers are located on the right ankle, right wrist, right hip, left thigh, and left upper arm. The experiments presented here only exploit the signals captured on the hip, being this a reasonable position for a smartphone too. This dataset has been chosen because, besides having a sufficiently high number of users, it has been captured during two different sessions. This is important for testing performance after some time has elapsed between enrollment and recognition.

### 7.2 Experimental results and discussion

For each experiment, the walks of the two sessions were exploited in a different way for training and testing. The discussion following the experiments description analyzes how a different partitioning of the dataset into training and testing subsets, either taking into account the time difference or not, can influence the experimental results. The feature extraction and training procedure were always carried out as described above.

Verification experiments were carried out in two rounds. In the first round, the walks in the union of the two sessions were divided in a classical way into 70% training samples and 30% testing samples. In the second round, the walks of one session, in turn, were used as a training set, and those in the other session were used as testing, and then the results were averaged. It is reasonable to expect that the performance achieved in the first round would be better due to the presence of a subset of the same session walks used for testing in the training set too, i.e., walks captured at the same time of those used for testing. This means that some temporary characteristics can find a match.

The experiments entailing identification closed set were also carried out in two rounds with the same settings as in verification. The expected outcomes are similar for the same reasons. Table 1 reports in the last rows the results achieved by the proposed approach and the comparison with other literature proposals, also taking into account the kind of application and the use of the two sessions of the ZJU-gaitacc dataset. Especially comparing the results obtained in the two rounds of testing, it is possible to observe the dramatic influence of the benchmark partition into training and testing. As expected, when part of the second session is included in the training process, just to reach the “canonical” 70/30 partition, some of the possibly temporary characteristics, peculiar for each user, play a relevant role in testing. It is possible to observe that, in these conditions, the proposed approach outperforms most of the others, which is systematic in a dramatically evident way in verification. It can be hypothesized that when both training and testing are carried out on a single session, like in Zeng et al. (2018), this effect is also stronger. Of course, this is not a limitation if the approach is used for re-identification in the very short term. However, template updating should be executed periodically. On the other hand, the extremely lower results achieved using the sessions in the fairest way demonstrate that even the most significant features do not sufficiently capture individual invariant gait regularities. However, it is further interesting to notice that when averaging the results from the two possible combinations of training and testing (last row in the table), the results are definitely worse than those achieved by De Marsico and Mecca (2017) (the only other work taking into account the session separation) but obtained by a definitely lighter computation (DTW vs. application of the trained function to reduced feature vectors).

Figure 5 visually suggests similar considerations regarding the influence of session partition. While, as expected, there is a negligible difference when using either session 1 or session 2 for training and the other one for testing, the 70/30 partition (one session participates in both training and testing, though obviously with different samples) causes a dramatic boost in performance. However, this latter result is not generalizable because it is obtained in a somehow very specific, if not biased, condition. This suggests that when using behavioral traits that are more variable across time, more data sessions should be captured to try to encompass as much variability as possible and be used appropriately for evaluation according to the expected application. For similar reasons, enrollment in real contexts should entail capturing more samples across different times.

<sup>7</sup> <http://www.cs.zju.edu.cn/~gpan/database/gaitacc.html>.

**Table 1** Experimental results compared with approaches using the same ZJU-gaitacc dataset and same/different use of its sessions

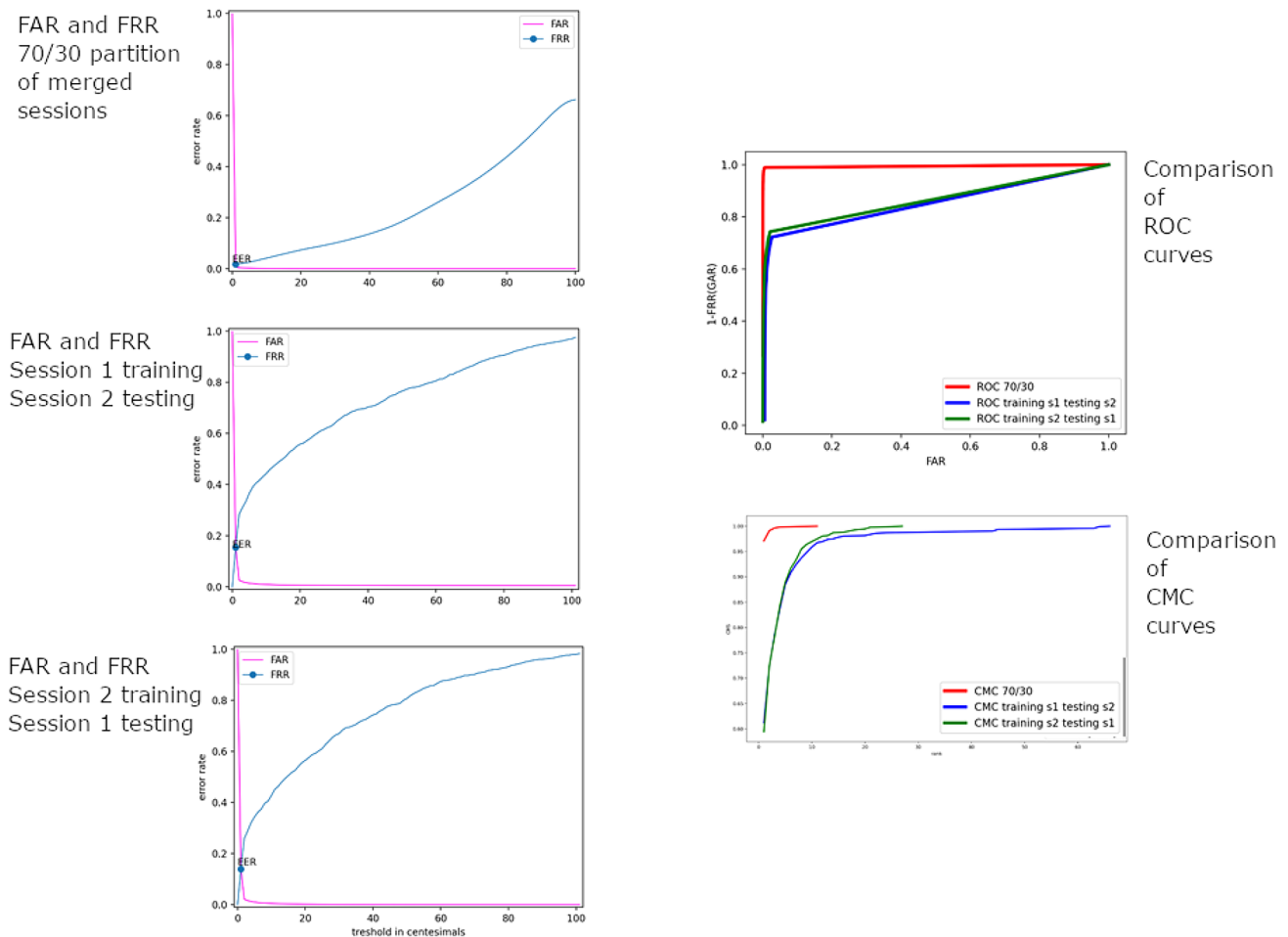
References	EER (%)	RR (%)	Session-based partition	Method
Zhang et al. (2015)	8.09	73.04	Identification YES Verification NO	Signature Points
Zeng et al. (2018)	–	86.9–96.2	NO Uses either session 0 or session 1 separately 2-fold each with leave-one-out	Deterministic Learning
Sun et al. (2018)	91.75% User Authentication Rate	96.9	NO	Pearson correlation coefficient (PCC) among gait cycles
Qin et al. (2019)	6	<b>98.4</b>	NO	Fuzzy based on Neural Networks and Extreme Value Statistics
De Marsico and Mecca (2017)	9.26	92.8	YES	Dynamic Time Warping (DTW) on unsegmented walk signals
De Marsico et al. (2018)	18.07	–	Half of each session for training, the rest for testing	Principal Feature Analysis (PFA)
Mecca (2018)	8.24	96.49	NO	DTW on unsegmented signal after removing the starting steps
De Marsico and Mecca (2018)	4.1	97.28	NO	DTW on unsegmented signal with score level fusion after applying different Gaussian filters
Giorgi et al. (2017)	–	94.00% Re-Id Accuracy	NO—only 1 session used	Deep Neural Network
Proposed	<b>0.8 ROC AUC = 0.99</b>	97 CMS (5) = 99.8 CMS (10) = 99.8 CMS = 1 at rank 11	NO 70% of all session walks for training 30% for testing	Feature extraction and Logistic Regression
Proposed	15 ROC AUC = 0.85	61.2 CMS (5) = 88.4 CMS (10) = 95.8 CMS = 1 at rank 66	YES Session 1 training Session 2 testing	Feature extraction and Logistic Regression
Proposed	14 ROC AUC=0.86	59.4 CMS (5) = 88.8 CMS (10) = 96.4 CMS = 1 at rank 27	YES Session 2 training Session 1 testing	Feature extraction and Logistic Regression
Proposed	14.5 ROC AUC = 0.86	60.2 CMS (5) = 88.6 CMS (10) = 96.3 CMS = 1 at a rank between 27 and 61	YES Average with Session-based partition	Feature extraction and Logistic Regression

The best results are in bold

## 8 Conclusions

Gait recognition by wearable sensors has many advantages. It does not suffer from typical computer vision limitations, is less demanding from a resource point of view, and can be carried out in specific settings without further equipment. However, the results in the literature still demonstrate the need for searching for stronger regularities in individual walking patterns to tackle recognition in the medium/long term. The results presented here further underline that more capture sessions separated in time and their partition between training and testing can dramatically affect the reported recognition performance. Unfortunately, datasets

collecting behavioral data presently present limitations either in the number of sessions or in the number of samples per session. It is interesting to relate the results presented here to the intended use of biometric recognition. When requiring identification of the enrolled subjects at time distance, much better generalizable results can be obtained using multiple-session datasets that better reflect the variability of traits. However, the trait variability can represent an issue also in the case of continuous re-identification of the owner of a smart device (short distance among samples) unless a periodical template updating is planned, i.e., the periodical substitution of the user template used as reference. These results and considerations are expected to extend to other behavioral traits. In the follow-up of this research, we will



**Fig. 5** Left column: FAR and FRR curves reporting the results achieved by the proposed approach according to the kind of session partition; from top to bottom 70/30, s1 vs. s2, s2 vs. s1. Right col-

umn: ROC and CMC curves reporting the results achieved by the proposed approach according to the kind of session partition

continue investigating the possible strategies to extract the truly invariant characteristics related to individual kinematic strategies. We are encouraged by a simple observation: when we know a person very well, we can often recognize it, even from the back, just by looking at the gait pattern. This means that the human cognitive system can extract the essentials from such patterns, and, as for other biometric traits, the challenge is to reproduce this ability for an automatic system.

**Funding** Open access funding provided by Università degli Studi di Roma La Sapienza within the CRUI-CARE Agreement.

**Data availability** The used dataset is public and not owned by the authors. The link to the dataset is reported in footnote no. 7.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated

otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abate AF, De Marsico M, Riccio D, Tortora G (2011) Mubai: multiagent biometrics for ambient intelligence. *J Ambient Intell Humaniz Comput* 2(2):81–89
- Barra S, Castiglione A, De Marsico M, Nappi M, Raymond Choo K-K (2018) Cloud-based biometrics (biometrics as a service) for smart cities, nations, and beyond. *IEEE Cloud Comput* 5(5):92–100
- Bianchi L, Angelini D, Lacquaniti F (1998) Individual characteristics of human walking mechanics. *Pflügers Archiv* 436(3):343–356
- Borghese NA, Bianchi L, Lacquaniti F (1996) Kinematic determinants of human locomotion. *J Physiol* 494(3):863–879
- Brombin L, Gambini M, Gronchi P, Magherini R, Nannini L, Pochiero A, Sieni A, Vecchio A (2019) User's authentication using information collected by smart-shoes. In: *EAI international conference on body area networks*, pp 266–277. Springer, Berlin



- Cervantes J, Garcia-Lamont F, Rodríguez-Mazahua L, Lopez A (2020) A comprehensive survey on support vector machine classification: applications, challenges and trends. *Neurocomputing* 408:189–215
- Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res* 16:321–57
- De Marsico M, Mecca A (2017) Biometric walk recognizer—gait recognition by a single smartphone accelerometer. *Multimed Tools Appl* 76(4):4713–4745 (ISSN 13807501)
- Derawi MO, Bours P, Holien K (2010) Improved cycle detection for accelerometer based gait authentication. In: 2010 sixth international conference on intelligent information hiding and multimedia signal processing (IIH-MSP), pp 312–317. IEEE
- De Marsico M, Mecca A (2019) A survey on gait recognition via wearable sensors. *ACM Comput Surv (CSUR)* 52(4):1–39
- De Marsico M, Nappi M, Riccio D, Wechsler H (2011) Iris segmentation using pupil location, linearization, and limbus boundary reconstruction in ambient intelligent environments. *J Ambient Intell Humaniz Comput* 2(2):153–162
- De Marsico M, Fartade EG, Mecca A (2018) Feature-based analysis of gait signals for biometric recognition. In: ICPRAM 2018-7th international conference on pattern recognition applications and methods, pp 630–637
- De Marsico M, Mecca A (2018) Benefits of gaussian convolution in gait recognition. In: 2018 international conference of the biometrics special interest group (BIOSIG), pp 1–5. IEEE
- Fernandez-Lopez P, Liu-Jimenez J, Sanchez-Redondo C, Sanchez-Reillo R (2016) Gait recognition using smartphone. In: 2016 IEEE international carnanah conference on security technology (ICCST), pp 1–7. IEEE
- Gadaleta M, Rossi M (2016) Idnet: smartphone-based gait recognition with convolutional neural networks. arXiv preprint [arXiv:1606.03238](https://arxiv.org/abs/1606.03238)
- Gafurov D, Snekenes E, Bours P (2010) Improved gait recognition performance using cycle matching. In: 2010 IEEE 24th international conference on advanced information networking and applications workshops (WAINA), pp 836–841. IEEE
- Giorgi G, Martinelli F, Saracino A, Sheikhalishahi M (2017) Try walking in my shoes, if you can: accurate gait recognition through deep learning. In: International conference on computer safety, reliability, and security, pp 384–395. Springer, Berlin
- Hizam SM, Ahmed W, Fahad M, Akter H, Sentosa I, Ali J (2021) User behavior assessment towards biometric facial recognition system: a SEM-neural network approach. In: Advances in information and communication: proceedings of the 2021 future of information and communication conference (FICC), vol 2, pp 1037–1050
- Joseph Raj V, Balamurugan S (2017) Survey of current trends in human gait recognition approaches. *Int J Adv Res* 5(10):1851–1864
- Kamruzzaman J, Begg RK (2006) Support vector machines and other pattern recognition approaches to the diagnosis of cerebral palsy gait. *IEEE Trans Biomed Eng* 53(12):2479–2490
- Kanev K, De Marsico M, Bottoni P, Mecca A (2016) Mobiles and wearables: owner biometrics and authentication. vol 07-10-June-2016, pp 318–319. Association for Computing Machinery. ISBN 9781450341318
- Kavita, Walia GS, Rohilla R (2021) A contemporary survey of multimodal presentation attack detection techniques: challenges and opportunities. *SN Comput Sci* 2:1–7
- Klucken J, Barth J, Kugler P, Schlachetzki J, Henze T, Marxreiter F, Kohl Z, Steidl R, Hornegger J, Eskofier B et al (2013) Unbiased and mobile gait analysis detects motor impairment in Parkinson's disease. *PLoS ONE* 8(2):e56956
- Kröger JL, Raschke P, Bhuiyan TR (2019) Privacy implications of accelerometer data: a review of possible inferences. In: Proceedings of the 3rd international conference on cryptography, security and privacy, pp 81–87
- Lee TKM, Belkhatir M, Sanei S (2014) A comprehensive review of past and present vision-based techniques for gait recognition. *Multimed Tools Appl* 72(3):2833–2869
- Lu Y, Cohen I, Zhou XS, Tian Q (2007) Feature selection using principal feature analysis. In: Proceedings of the 15th ACM international conference on multimedia, pp 301–304
- Mecca A (2018) Impact of gait stabilization: a study on how to exploit it for user recognition. In: 2018 14th international conference on signal-image technology & internet-based systems (SITIS), pp 553–560. IEEE
- Middleton L, Buss AA, Bazin A, Nixon MS et al (2005) A floor sensor system for gait recognition. In: Fourth IEEE workshop on automatic identification advanced technologies, 2005, pp 171–176. IEEE
- Muaaz M, Mayrhofer R (2017) Smartphone-based gait recognition: from authentication to imitation. *IEEE Trans Mob Comput* 16(11):3209–3221
- Muro-De-La-Herran A, Garcia-Zapirain B, Mendez-Zorrilla A (2014) Gait analysis methods: an overview of wearable and non-wearable systems, highlighting clinical applications. *Sensors* 14(2):3362–3394
- Nickel C, Brandt H, Busch C (2011a) Classification of acceleration data for biometric gait recognition on mobile devices. *BIOSIG* 11:57–66
- Nickel C, Busch C, Rangarajan S, Möbius M (2011b) Using hidden Markov models for accelerometer-based biometric gait recognition. In: 2011 IEEE 7th international colloquium on signal processing and its applications (CSPA), pp 58–63. IEEE
- Nickel C, Wirtl T, Busch C (2012) Authentication of smartphone users based on the way they walk using k-NN algorithm. In: 2012 eighth international conference on intelligent information hiding and multimedia signal processing (IIH-MSP), pp 16–20. IEEE
- Orr RJ, Abowd GD (2000) The smart floor: a mechanism for natural user identification and tracking. In: CHI'00 extended abstracts on Human factors in computing systems, pp 275–276. ACM
- Pedotti A (1977) A study of motor coordination and neuromuscular activities in human locomotion. *Biol Cybern* 26(1):53–62
- Prakash C, Kumar R, Mittal N (2018) Recent developments in human gait research: parameters, approaches, applications, machine learning techniques, datasets and challenges. *Artif Intell Rev* 49(1):1–40
- Qin Z, Huang G, Xiong H, Qin Z, Choo KK (2019) A fuzzy authentication system based on neural network learning and extreme value statistics. *IEEE Trans Fuzzy Syst* 29:549–559
- Rong L, Zhiguo D, Jianzhong Z, Ming L (2007) Identification of individual walking patterns using gait acceleration. In: 2007 1st international conference on bioinformatics and biomedical engineering, pp 543–546. IEEE
- Sun F, Mao C, Fan X, Li Y (2018) Accelerometer-based speed-adaptive gait authentication method for wearable IoT devices. *IEEE Internet Things J* 6(1):820–830
- Suutala J, Röning J (2004) Towards the adaptive identification of walkers: automated feature selection of footsteps using distinction sensitive LVQ. In: Int. workshop on processing sensory information for proactive systems (PSIPS 2004), pp 14–15
- Zeng W, Chen J, Yuan C, Liu F, Wang Q, Wang Y (2018) Accelerometer-based gait recognition via deterministic learning. In: 2018 Chinese control and decision conference (CCDC), pp 6280–6285. IEEE
- Zhang Y, Pan G, Jia K, Minlong L, Wang Y, Zhaohui W (2015) Accelerometer-based gait recognition by sparse representation of signature points with clusters. *IEEE Trans Cybern* 45(9):1864–1875