# Towards the suitability of gait wearable signal processing for long term recognition

Anonymous IJCB 2022 submission


## Abstract

*One of the present approaches to gait recognition exploits the signals captured by wearable sensors, especially the accelerometers embedded in modern smartphones. However, the different speed, the ground slope, or simply the time lapse between captures cause variations that negatively affect long term recognition in a dramatic way. The proposed procedure aims at extracting gait characteristics that are as invariant as possible, and therefore useful for accurate long term recognition. The experiments compare the performance of the proposal with others in state-of-the-art that use the same benchmark, namely the ZJU-gaitacc dataset. This dataset includes a high number of samples per subject, captured in two time-separated sessions. This allows to assess the performance of the proposed method also in the long term, i.e., when comparing templates captured in different times. Most works using the same benchmark so far have not exploited both sessions. They use samples captured in the same time, constraining the use of this trait to continuous recognition, e.g., of the smartphone owner. The obtained results testify that, in this condition, the proposed feature-based method outperforms competitors in the current literature. The experiments also compare the results from a session-based partition with those obtained from a training that mixes-up samples from different sessions. As expected, the latter strategy can dramatically improve the measured performance. The significantly different results seem to suggest that the session-based partition, when feasible, can provide more realistic results, closer to the real-world application context when behavioural traits are involved in the long term. The same results seem also to testify that there is still need to improve the accuracy of gait recognition via wearable sensors. This calls for further investigation of the problems related to the variability over time in the pattern of individual gait signals.*


## 1. Introduction

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 [2], 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 [21] 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.*" Precise kinematic strategies [1] explain the behavioral variations that produce the individual differences that can allow biometric identification.

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 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 $x_i$, $y_i$, and $z_i$ 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. However, 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 a novel automatic procedure to create individual gait models using the most relevant information from accelerometer data, in the attempt to distill the invariant dynamic regularities present in the walking pattern. The first step extracts a number of aggregate features from the time series of the accelerometer events; afterwards, a two-phase procedure, including two steps of feature selection, produces feature vectors that better capture the invariant gait kinematics. Logistic Regression produces a trained model for each identity in the gallery after suitable upsampling/downsampling to balance the individual training sets. The contributions are:

- the user modeling procedure that, though exploiting popular algorithms well-known in literature, combines them in an original way to extract and select gait features and to further reduce template dimensionality; an upsampling algorithm is further used to balance the sample classes during the one-vs-rest training procedure; the achieved results are, at the best of our knowledge, the best in literature using the same dataset;
- experiments to assess wearable gait recognition in the long term; the used ZJU-gaitacc dataset, though dated and including signals with poor quality with respect to present accelerometers, differently from other available gait datasets has the advantage to include many samples per identity that are captured in two sessions separated in time; this allows a session-based partition in order to train the identity models on samples acquired in a time different from probe capture time, as it happens in realistic recognition applications in the medium/long term; in this way it is possible to evaluate the recognition performance of this behavioural biometric trait when it is not intended for continuous re-identification only.

The achieved results are compared with other approaches tested over the same dataset, considering verification and identification closed-set. Further study should allow extracting from the stereotypical gait pattern those features that are better related to the personal kinematic strategies and therefore allow for more reliable recognition.

## 2. Some related work

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 [13], can dramatically suffer from time misalignment, especially without a preliminary interpolation and for signals captured at different times. Therefore, they are rather exploited within 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 [7] or cycles [10, 23, 11] (a cycle is composed by a pair of steps) that are compared afterwards, 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 form 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 points in the time series. A Hidden Markov Model (HMM) is computed for each user in [19]. The acceleration data from each fragment is used in [18] 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 [20] 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 [12] 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. A large wide dataset freely available is ZJU-gaitacc [26] 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 is worth considering that the approaches in both works fuse the results of 5 accelerometer 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 [14] 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 can be found in [9].

## 3. Gait information extraction

During the walking process, the accelerometer embedded in a smart device, typically a smartphone, produces a set of three time series (one for each axis in the 3D space), where each value represents an acceleration measure. Figure 1 shows an example taken from the ZJU-gaitacc. Each

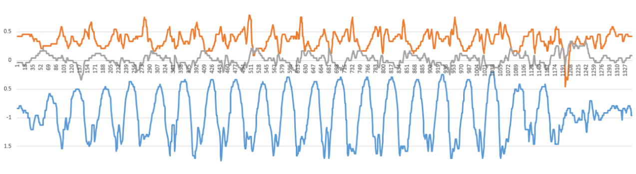

Figure 1. Example of the time series from the accelerometer signal

time series contains values that are influenced by both the stereotypical walking dynamics and by personal kinematic strategies aiming at maintaining the body's equilibrium during the advancement of the walk. The factors affecting the latter, and distinguishing the walking patterns of different individuals, include, e.g., the length of the legs, possible asymmetry of hips or of other parts of the upper body, the unconscious and automatic effort to maintain the head in a balanced position. The proposed approach to gait recognition relies on a novel procedure to extract the most informative features, i.e., the most distinctive ones, from the accelerometer time series. The following paragraphs report both the elements of the theoretical approach, and some details useful for experimental reproducibility.

The procedure of feature extraction from the accelerometer temporal series exploits the *tsfresh* [1] library, while the selection of the most relevant ones, the training procedure and the classifier test exploit the *scikit-learn* [2] library and, for some specific tasks, *imbalanced-learn* [3] (all written in Python). The extracted features consist of aggregate values obtained from the time series relative to each single axis. This process initially returns 763 features per axis for a total of 2289 features. Obviously, it is not possible to list them all, and moreover not all of them, tough being suitable for time series analysis, meaningfully apply to gait acceleration series, as it is highlighted by the following selection procedure. Some interesting features are:
- autocorrelation, i.e., the correlation of a signal with a delayed copy of itself for finding repeating patterns, such as the presence of a periodic signal obscured by noise, that distinguishes regular gait patterns from more rambling ones; the formula of the autocorrelation $r_k$ for lag k is:

$$r_k = \frac{\sum_{t=k+1}^{n}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{n}(y_t - \bar{y})^2} \tag{1}$$

- features computed from the absolute Fourier transform spectrum when treating the time series as a 1-D signal, e.g., the spectral centroid (mean), the variance, the skew, and the kurtosis; this seems reasonable due to nature of normal gait that is almost periodic; for instance, the spectral centroid is computed as the weighted mean of the frequencies present in the signal, determined using a Fourier transform, with their magnitudes as the weights:

$$spcentroid = \frac{\sum_{n=0}^{N-1} f(n)x(n)}{\sum_{n=0}^{N-1} x(n)} \tag{2}$$

where $x(n)$ is the weighted frequency value of the bin number $n$ (possibly overlapping signal segment), and $f(n)$ represents the center frequency of that bin;
- the binned entropy of the power spectral density of the time series obtained using the Welch method, that provides an estimate of the series informative content.[4]

The training and testing of classifiers with these full vectors achieved unsatisfactory results, as reported in the section about experimental results. This is possibly due to information redundancy, to noisy or especially outliers-affected features, or simply by the fact that some of them have a poor meaning in gait pattern modeling and rather excessively sparsify the feature space. Therefore, the next step selects the most relevant information for recognition purposes. A two-phase procedure allows to significantly reduce the dimension of the template vectors. This includes a sequence of two steps of feature selection to reduce the

[1] https://tsfresh.readthedocs.io/en/latest/, https://tsfresh.readthedocs.io/en/latest/api/tsfresh.feature_extraction.html

[2] https://scikit-learn.org/stable/

[3] https://imbalanced-learn.org/stable

[4] A complete list is available in tsfresh documentation at https://tsfresh.readthedocs.io/en/latest/text/list_of_features.html

dimensionality of the obtained vector space .

**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 within a set by a supervised approach taking into account the identity associated to each vector. This strategy selects 1421 features out of the 2289 initial ones.

The second strategy first discards all 0-variance (constant) features, i.e., those that have no useful information for classifier training. The surviving features are 2158 out of the initial 2289. Then, the procedure makes a further selection according to a percentile of the highest scores returned by ANOVA F-value between label/feature for classification tasks. This function assigns the F-score to each feature according to the following formula:

$$F = (\text{inter-group variance})/(\text{intra-group variance}) \quad (3)$$

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 samples of all features; "intra-group variance" is the normal variance of the features. At the end of this computation, the final feature vector includes the highest scored 45% of the input features, so that the number of features decreases from 2158 to 971.

**Second phase**. The second phase relies on a further selection based on Principal Feature Analysis (PFA) [16]. In its first steps PFA uses the same computation of the PCA by transforming the space of features into one of smaller dimension. The further steps entailed by PFA determine the most relevant projections of the original features in the new space. In more detail, the PFA algorithm starts from the matrix having as columns the eigenvectors identified as axes by PCA. The purpose is to analyze the structure of the most informative components just obtained to find the largest informative subset of features of a predefined dimension . 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. The formula used for normalizing a feature value $x$ is the familiar one:

$$\frac{x-\mu}{\sigma} \quad (4)$$

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

# 4. Classifier training

The chosen strategy for the classification is to train a model for each class. The multi-class classification is split into one binary classification problem per class adopting a one-vs-rest strategy. Given the number of the "other" classes there is a clear unbalance between positive and negative samples. From one side, this may decrease the ability to learn the distinctive characteristics of the modeled class. From the other side, it is important to exploit the full variety of inter-class differences. In other words, it is necessary to compensate for prior class probabilities. This holds for both more traditional machine learning approaches [15] and for the more recent deep learning ones [3]. The classical solutions to the class unbalance problem are based on resampling [4] 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, this is seldom applied. The second method entails upsampling, i.e., increasing the number of samples of the minority class.

## 4.1. Upsampling 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 from the minority class; its only advantage, however, is to balance the distribution of the instances in terms of probability of class belonging, but without bringing any additional relevant discriminative information to the training set. The adopted strategy rather applies SMOTE (synthetic minority oversampling technique) [5], which creates synthetic instances of the minority class starting from those already present. In short, the algorithm begins by randomly selecting an instance $x_I$ in the minority class $I$ and looking for the $k$ closest instances in the same class in the space of characteristics ($k$ is a parameter of the algorithm, in these experiments $k = 5$). Then it randomly chooses one of these $k$ neighbors and connects it to $x_I$ through a segment; the new synthetic instance is obtained by randomly selecting a point belonging to the segment (obviously different from the extremes). Figure 2 shows a picture of the algorithm behaviour.

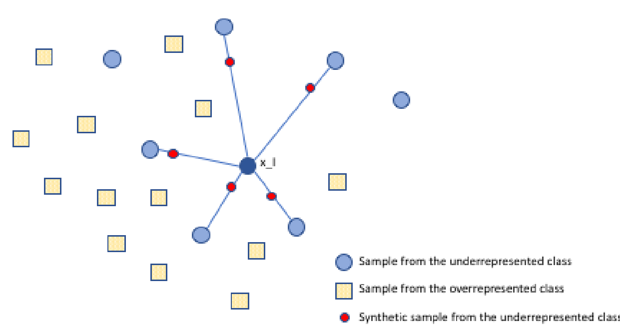


Figure 2. How the upsampling procedure works.

IJCB 2022


### 4.2. Logistic Regression

The popular Logistic Regression uses an equation as the representation, similarly 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 value (0 or 1) rather than a numeric value. This is obtained by combining simple linear regression with a logistic function in the following way:

$$P = \frac{1}{1 + e^{(\beta_0 + \sum_{k=1}^{n} \beta_i x_i)}} \tag{5}$$

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 $(x_i)$. So each individual model is computed by learning the $\beta$ coefficients for each column in the input data from the training data. As already mentioned above, this technique is generalized to a multi-class 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.

### 4.3. Final calibration

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, that can be exploited during the evaluation as a confidence or similarity measure. The goal of the calibration is to ensure that the predicted probabilities can be directly used as confidence values: this means for example that, among the instances whose predicted probability is close to 0.6, approximately 60% of them will actually have to belong to the positive class. The technique used for calibration is isotonic regression (see Figure 3). The obtained values are finally used as similarity values for the evaluation of verification and closed-set identification performance. Re-identification entails determining whether the accelerometer is worn by the same person during a recognition session. It can be considered as a special case of recognition where the model and the probe are captured close in time.

## 5. Experimental results

### 5.1. ZJU-gaitacc dataset

The dataset ZJU-gaitacc [5] [26] 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 performance analysis that this papers proposes, i.e., the evaluation of recognition in the long term. 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 samples (about 1400) being collected along a hallway 20 meters long. 5 Wii Remote controllers are located on 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 for two main of reasons, though being quite dated and collected using sensors with a lower resolution than present ones. The first reason is that it includes a significant number of subjects, sufficient to reliably test inter-subject gait variability. The second reason is that, differently from other datasets collecting gait signals through wearable devices, the available gait time series are sufficiently long to allow a realistic recognition, assuming that it is quite unrealistic to recognize a person from a few steps unless they are extracted from a longer signal [17]. Last but not least, as mentioned above, the samples belong to sessions sufficiently separated in time to evaluate long term recognition. During three rounds of experiments, the walks of the two sessions were exploited in a different way for training and testing.

In the first round, the walks in the union of the two sessions have been divided in the classical way into 70% training samples and 30% testing samples. This set of experiments aimed at evaluating the influence of using training samples collected in different times on the final testing performance. In order to have a more valid result, of course

[5] http://www.cs.zju.edu.cn/~gpan/database/gaitacc.html

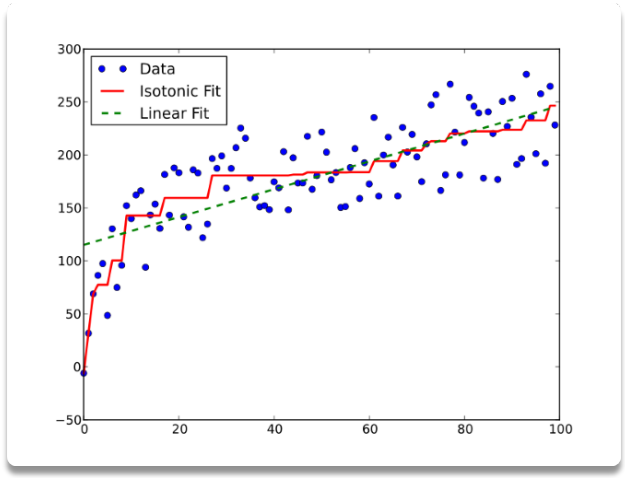

Figure 3. Difference between linear and isotonic regression.

it would have been better having three different sessions at least, and using two of them in turn for training and the remaining one for testing. Unfortunately, no available dataset, to the best of our knowledge, has sufficient data collected in a sufficient number of sessions. Some bias may come from session overlap in training and testing, because some temporary characteristics can find a match. However, since of course there is no sample overlap, this setup should provide useful information on the usefulness of a multiple-session enrolling.

In the second round, the walks of a single session have been divided intro training and testing, in a way similar to short term recognition or re-identification. This setup is the most frequent one in literature.

Finally, in the third round, the walks of each session in turn were used as training set and those in the other session have been used for testing, as it would be in a real-world recognition after a time lapse from enrolling.

It is reasonable to expect that the best performance is achieved in the second setup, due to more homogeneous intra-subject characteristics, and the second best in the first setup due to larger intra-subject variations despite the session overlap (no sample overlap anyway).

The feature extraction and training procedure were always those described above, and experiments were carried out in both verification and identification closed set. It is worth mentioning that identification closed-set can be used in a context where, for example, a two-phases recognition is carried out. For instance, a preliminary identification open set or verification carried out through other biometric traits may allow access to a controlled zone. After this limited and closed set of subjects is defined, further access to subzones can be granted by just identifying the approaching subject without any further explicit request by the subject itself.

### 5.2. Comparison of results and discussion

Table 1 reports the results achieved by the proposed approach and the comparison with other literature proposals, also taking into account the kind of application, either verification or identification, and the use of the two sessions of the ZJU-gaitacc dataset. It is worth noticing that, among the compared methods, the proposals with better performance either use Dynamic Time Warping on the whole signal, therefore being computationally demanding, or neural network-based approaches that call for a much higher number of samples for training. A simple ablation study further evaluated the effectiveness of the feature selection strategy, by comparing the performance of both verification and identification with and without the selection steps. For sake of space, the table only reports the results obtained with the full set of features and with the set obtained by the best selection strategy (see Section 3) when mixing samples from both session in the training set (the first two rows for the proposed method). The dramatic improvement is clearly appreciable, with EER passing from 3% to 0.8% and RR passing from 77% to 97%. The next row related to the proposed method shows the best performance for verification among all the compared methods in the same experimental setup (samples from the same session used for training and testing). The improvement with respect to the second best ([8]) is from 4.1% to 0.6%, i.e., a relative 85%. The AUC for the corresponding ROC curve (0.998) further testifies an extremely good verification behaviour. The performance in identification is slightly below the best ones, but the competitor methods achieve significantly worse performance for verification. Therefore, overall, it is possible to state that, to the best of our knowledge, the proposed method achieves the best performance in the state-of-the-art for this dataset, whose characteristics make it more suitable for real-world-like evaluation than others publicly available.

Finally, the last three rows of the table and Figure 4 allow to observe the dramatic influence of the use of both sessions in the benchmark partition into training and testing, that can realistically reproduce the real-world recognition in the long term. The extremely lower results achieved using one session for training and the other for testing demonstrate that even the most significant features do not sufficiently capture individual invariant gait regularities over time. This is further underlined by the different results when alternating the sessions' role. For instance, the value of 1 for the CMS is reached only at rank 66 when using session 1 for training, while it is reached at rank 27 when using session 2. However, the results obtained using samples from two sessions for training encourage further investigations.

## 6. Conclusions

Gait recognition by wearable-sensors has many advantages. It does not suffer from typical computer vision limitations, it is less demanding from a resource point of view, and it can be carried out in specific settings without further equipment. However, the results in literature and also those presented in this paper still highlight the difficulty of gait recognition in the medium/long term and demonstrate the need for searching stronger regularities in individual walking patterns. The presence of more capture sessions separated in time in a suitably large dataset and their partition between training and testing can effectively support further research. To this respect, it is interesting to observe that the influence of different capture sessions is not always sufficiently underlined for behavioural traits, since it is assumed that they cannot be used in the long term. For instance, gait datasets, especially when collected via wearable sensors, generally include data collected in close times (a single session). This constrains the generalizability of results to re-identification applications, or to recognition across a short

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

| **REFERENCE** | **EER (%)** | **RR (%)** | **SESSION PARTITION** | **METHOD** |
|---|---|---|---|---|
| [26] | 8.09 | 73.04 | Identification YES<br>Verification NO | Signature Points |
| [25] | - | 86.9 - 96.2 | NO - Uses only session 0<br>and session 1 separately -<br>2-fold each with leave-one-out | Deterministic Learning |
| [24] | 91.75% User<br>Authentication Rate | 96.9 | NO | Pearson correlation coefficient (PCC)<br>among gait cycles |
| [22] | 6 | **98.4** | NO | Fuzzy based on Neural Networks<br>and Extreme Value Statistics |
| [7] | 9.26 | 92.8 | YES | Dynamic Time Warping (DTW) on<br>unsegmented walk signals |
| [6] | 18.07 | - | Half of each session used<br>for training and the rest for<br>testing | Principal Feature Analysis |
| [17] | 8.24 | 96.49 | NO | DTW on unsegmented signal after<br>removing the starting steps |
| [8] | 4.1 | 97.28 | NO | DTW on unsegmented signal with score<br>level fusion after applying different<br>Gaussian filters |
| [14] | - | 94.00% Re-Id<br>Accuracy | NO - Only 1<br>session used | Deep Neural Network |
| Proposed<br>method | 3.0<br>AUC of ROC = 0.97 | 77<br>CMS(5) = 92.8<br>CMS(10) = 96.7<br>CMS =1 at rank 128 | NO<br>70% of all session walks training<br>30% for testing | Feature extraction<br>No feature selection/dimensionality reduction<br>Logistic Regression |
| Proposed<br>method | 0.8<br>AUC of ROC = 0.99 | 97<br>CMS(5) = 99.8<br>CMS(10) = 99.8<br>CMS =1 at rank 11 | NO<br>70% of all session walks training<br>30% for testing | Feature extraction<br>Feature selection/dimensionality reduction<br>Logistic Regression |
| Proposed<br>method | **0.6**<br>AUC of ROC = 0.99 | 96<br>CMS(5) = 97.5<br>CMS(10) = 99.5<br>CMS =1 at rank 11 | NO - Only 1<br>session used | Feature extraction<br>Feature selection/dimensionality reduction<br>Logistic Regression |
| Proposed<br>method | 15<br>AUC of ROC = 0.85 | 61.2<br>CMS(5) = 88.4<br>CMS(10) = 95.8<br>CMS =1 at rank 66 | YES<br>Session 1 training<br>Session 2 testing | Feature extraction<br>Feature selection/dimensionality reduction<br>Logistic Regression |
| Proposed<br>method | 14<br>AUC of ROC = 0.86 | 59.4<br>CMS(5) = 88.8<br>CMS(10) = 96.4<br>CMS =1 at rank 27 | YES<br>Session 2 training<br>Session 1 testing | Feature extraction<br>Feature selection/dimensionality reduction<br>Logistic Regression |
| Proposed<br>method | 14.5<br>AUC of ROC = 0.86 | 60.2<br>CMS(5) = 88.6<br>CMS(10) = 96.3<br>CMS =1 at a rank<br>between 27 and 61 | YES<br>Average one Session training<br>Other Session testing | Feature extraction<br>Feature selection/dimensionality reduction<br>Logistic Regression |

time lapse. However, a deeper understanding of the regularities that allow to recognize a known person from gait could allow a wider use of this trait. As a final note, it could be interesting to fuse the results of multimodal gait recognition by exploiting the complementary information provided by video and wearable sensors. However, at present no large dataset allowing this exists, and, being the information from the two modalities strictly correlated though different, it is not even possible to use a chimeric dataset.

## References

[1] L. Bianchi, D. Angelini, and F. Lacquaniti. Individual characteristics of human walking mechanics. *Pflügers Archiv*, 436(3):343–356, 1998.

[2] N. A. Borghese, L. Bianchi, and F. Lacquaniti. Kinematic determinants of human locomotion. *The Journal of physiology*, 494(3):863–879, 1996.

[3] M. Buda, A. Maki, and M. A. Mazurowski. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, 106:249–259, 2018.

[4] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez. A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, 408:189–215, 2020.

[5] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. Smote: synthetic minority over-sampling tech-

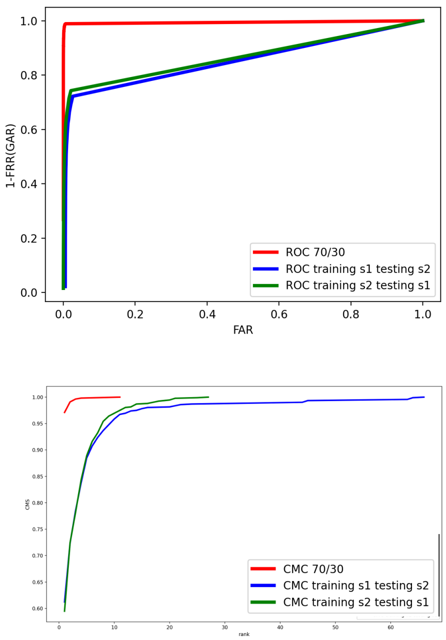


Figure 4. Curves reporting some of the results achieved by the proposed approach according to the kind of application and session partition.

nique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[6] M. De Marsico, E. G. Fartade, and A. Mecca. Feature-based analysis of gait signals for biometric recognition. In *ICPRAM 2018-7th International Conference on Pattern Recognition Applications and Methods*, pages 630–637, 2018.

[7] M. De Marsico and A. Mecca. Biometric walk recognizer - gait recognition by a single smartphone accelerometer. *Multimedia Tools and Applications*, 2016.

[8] M. De Marsico and A. Mecca. Benefits of gaussian convolution in gait recognition. In *2018 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–5. IEEE, 2018.

[9] M. De Marsico and A. Mecca. A survey on gait recognition via wearable sensors. *ACM Computing Surveys (CSUR)*, 52(4):1–39, 2019.

[10] M. O. Derawi, P. Bours, and K. Holien. Improved cycle detection for accelerometer based gait authentication. In *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2010 Sixth International Conference on*, pages 312–317. IEEE, 2010.

[11] P. Fernandez-Lopez, J. Liu-Jimenez, C. Sanchez-Redondo, and R. Sanchez-Reillo. Gait recognition using smartphone. In *Security Technology (ICCST), 2016 IEEE International Carnahan Conference on*, pages 1–7. IEEE, 2016.

[12] M. Gadaleta and M. Rossi. Idnet: Smartphone-based gait recognition with convolutional neural networks. *arXiv preprint arXiv:1606.03238*, 2016.

[13] D. Gafurov, E. Snekkenes, and P. Bours. Improved gait recognition performance using cycle matching. In *Advanced Information Networking and Applications Workshops (WAINA), 2010 IEEE 24th International Conference on*, pages 836–841. IEEE, 2010.

[14] G. Giorgi, F. Martinelli, A. Saracino, and M. Sheikhalishahi. Try walking in my shoes, if you can: Accurate gait recognition through deep learning. In *International Conference on Computer Safety, Reliability, and Security*, pages 384–395. Springer, 2017.

[15] N. Japkowicz and S. Stephen. The class imbalance problem: A systematic study. *Intelligent data analysis*, 6(5):429–449, 2002.

[16] Y. Lu, I. Cohen, X. S. Zhou, and Q. Tian. Feature selection using principal feature analysis. In *Proceedings of the 15th ACM international conference on Multimedia*, pages 301–304, 2007.

[17] A. Mecca. 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)*, pages 553–560. IEEE, 2018.

[18] C. Nickel, H. Brandt, and C. Busch. Classification of acceleration data for biometric gait recognition on mobile devices. *BIOSIG*, 11:57–66, 2011.

[19] C. Nickel, C. Busch, S. Rangarajan, and M. Möbius. Using Hidden Markov Models for accelerometer-based biometric gait recognition. In *Signal Processing and its Applications (CSPA), 2011 IEEE 7th International Colloquium on*, pages 58–63. IEEE, 2011.

[20] C. Nickel, T. Wirtl, and C. Busch. Authentication of smartphone users based on the way they walk using k-NN algorithm. In *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2012 Eighth International Conference on*, pages 16–20. IEEE, 2012.

[21] A. Pedotti. A study of motor coordination and neuromuscular activities in human locomotion. *Biological cybernetics*, 26(1):53–62, 1977.

[22] Z. Qin, G. Huang, H. Xiong, Z. Qin, and K.-K. R. Choo. A fuzzy authentication system based on neural network learning and extreme value statistics. *IEEE Transactions on Fuzzy Systems*, 2019.

[23] L. Rong, D. Zhiguo, Z. Jianzhong, and L. Ming. Identification of individual walking patterns using gait acceleration. In *2007 1st International Conference on Bioinformatics and Biomedical Engineering*, pages 543–546. IEEE, 2007.

[24] F. Sun, C. Mao, X. Fan, and Y. Li. Accelerometer-based speed-adaptive gait authentication method for wearable iot devices. *IEEE Internet of Things Journal*, 6(1):820–830, 2018.

[25] W. Zeng, J. Chen, C. Yuan, F. Liu, Q. Wang, and Y. Wang. Accelerometer-based gait recognition via deterministic learning. In *2018 Chinese Control And Decision Conference (CCDC)*, pages 6280–6285. IEEE, 2018.

[26] Y. Zhang, G. Pan, K. Jia, M. Lu, Y. Wang, and Z. Wu. Accelerometer-based gait recognition by sparse representation of signature points with clusters. *IEEE Transactions on Cybernetics*, 45(9):1864–1875, 2015.