sensors

MDPI

Article

A New Post-Processing Proposal for Improving Biometric Gait Recognition Using Wearable Devices

Irene Salvador-Ortega , Carlos Vivaracho-Pascual * and Arancha Simon-Hurtado

Departamento de Informática, Escuela de Ingeniería Informática de Valladolid, Universidad de Valladolid, Paseo de Belén 15, 47011 Valladolid, Spain
* Correspondence: cevp@infor.uva.es

Abstract: In this work, a novel Window Score Fusion post-processing technique for biometric gait recognition is proposed and successfully tested. We show that the use of this technique allows recognition rates to be greatly improved, independently of the configuration for the previous stages of the system. For this, a strict biometric evaluation protocol has been followed, using a biometric database composed of data acquired from 38 subjects by means of a commercial smartwatch in two different sessions. A cross-session test (where training and testing data were acquired in different days) was performed. Following the state of the art, the proposal was tested with different configurations in the acquisition, pre-processing, feature extraction and classification stages, achieving improvements in all of the scenarios; improvements of 100% (0% error) were even reached in some cases. This shows the advantages of including the proposed technique, whatever the system.

Keywords: gait recognition; smartwatch; accelerometer sensor; window fusion technique; cross-session tests

check for updates

Citation: Salvador-Ortega, I.; Vivaracho-Pascual, C.; Simon-Hurtado, A. A New Post-Processing Proposal for Improving Biometric Gait Recognition Using Wearable Devices. *Sensors* **2023**, *23*, 1054. https://doi.org/10.3390/s23031054

Academic Editor: Miguel Velhote Correia

Received: 24 November 2022
Revised: 30 December 2022
Accepted: 9 January 2023
Published: 17 January 2023

1. Introduction

Biometric recognition is an important field for both economics (with a market in continuous growth) and research, where the search for and study of new alternatives is a question of continuous interest. From these, those related to wearables have aroused great interest in the biometric community in recent years due to the popularization of these devices. It can therefore be considered an emergent biometric technology with a promising future [1].

In this work, user authentication, or verification is approached by means of the behavioral biometric characteristic gait using a smartwatch. Biometric recognition encompasses two different tasks: verification and identification. In verification, the goal is to authenticate the user (Am I the person I claim to be?). In identification, the goal is to find the owner of the characteristic (To which person does it belong?).

Gait, as a behavioral biometric characteristic [2] like speech, signature, or keystroke, is based on *measurements and data derived from an action of the user*, in this case, the user's personal movement while walking. That is, the goal in biometric gait recognition is to characterize individuals by their way of walking.

The use of gait shows interesting advantages with regard to other more mature biometric characteristics (generally, physiological ones, e.g., fingerprint, face or iris): the capture is unobtrusive (the user does not need to carry out a special action, only a common one as is walking) and it is also difficult to steal or falsify. In addition, with the appearance of wearables, its capture can be performed continuously (while the user wears the device), and it is easy to obtain by simply accessing the corresponding sensors.

The use of gait for user recognition has been posed since the beginning of the research in biometric recognition [3]. The main problem, until the appearance of smartphones, was that dedicated devices were used to capture the user's way of walking, such as cameras or floor sensors [3]. In addition, its performance has usually been lower than other behavioral characteristics, for example, speech or signature.

A first step to make gait acquisition easier with the use of smartphones, more specifically, through the use of their inertial sensors (accelerometer and gyroscope). The use of these devices prevents the need to use external or specialized ones to capture user gait, such as the above-mentioned cameras or floor sensors. However, the signal acquired with these devices has the problem of being body-place-dependent [3,4]; i.e., the signal is different when the mobile is located, for example, on the belt, in the trouser pocket, or in the hand.

The appearance of Wearables solved this problem, since they are worn on a specific part of the body (head, neck, wrist, etc.), which is always the same.

With the popularization of these devices, the biometric community has shown considerable interest in them, due to their ability to capture both physiological and behavioral personal data, their use not being restricted to gait recognition. Some examples of these different approaches are heart-based authentication (using PhotoPlethysmoGraphy sensors [5] or Electrocardiogram [6,7]), speech [8,9], keystroke [10], galvanic skin response [11], and non-volitional brain response [12].

Nevertheless, gait has a major advantage with respect to others: it is easy to collect, as all current wearables have inertial sensors (accelerometer and gyroscope), these being the signals usually used for gait recognition [13]. Therefore, it is very suitable for general practical applications, which has increased the interest of the gait research community in these devices in recent years [3,4]. Here, our interest focuses on wrist wearables, since they are very popular, increasing their real usage possibilities.

From the two inertial sensors, accelerometer and gyroscope, the first is the most commonly used and the first referenced [14–16]. However, the gyroscope was rapidly incorporated [17]. From comparative works performed [18–21], the accelerometer has shown the best performance in general, so it is the one used here.

The systems proposed in the literature are based mainly on the standard biometric system [22,23] (Section 2) schematized in Figure 1. To improve recognition, different alternatives in the pre-processing, feature extraction, or classification stages have been proposed, as seen in Section 2. Here, we address the problem with a different approach, proposing the introduction of a post-processing stage between classification and decision making (Figure 2) where the Window Score Fusion Post-Processing (WSFPP) proposal is applied.

This proposal is based on several previous ideas:

- *Fusion of features* that can be implemented at distinct levels and used in biometrics to improve accuracy [24]. Here, we use score-level fusion. Two main approaches can be found at this level: fusing scores (outputs) of different classifiers over the same input (e.g., [25]) and fusing the scores of different characteristics, *multibiometric system* (e.g., [26,27]). We propose a different and simpler approach here, since the scores of a single classifier and characteristic (gait) are combined (Section 4).
- *Exploiting context* to enhance system accuracy [28]. This is habitual in problems where the context is relevant, e.g., those related with video (e.g., [29]), prosody recognition (e.g., [30]), or speech. Since, in gait, a step is related to the ones that go before and after, our proposal aims to exploit this dependence (Section 4).

When analyzing the literature, only a few works were found that used post-processing, but they do not exploit the context and are not performed at the decision level (Section 2).

To test our proposal, a strict biometric testing protocol has been followed, taking reproducibility principles into account in the experimental protocol description. Testing is performed under a cross-session scenario: the data acquired for training and testing are from different days, which this is important in order to include the intra-user variability of the biometric traits [31] and to consider the "template ageing" problem [22], essential in real systems.

Besides the main contribution shown, two new and original proposals for signal cleaning and for gait cycle extraction (gait cycle is defined in Section 3) are set out. Furthermore, from the different comparisons carried out (Section 5.1), the time domain vs. frequency domain in the feature extraction (the two main approaches) has been included, which has not been included in previous work to the best of our knowledge.

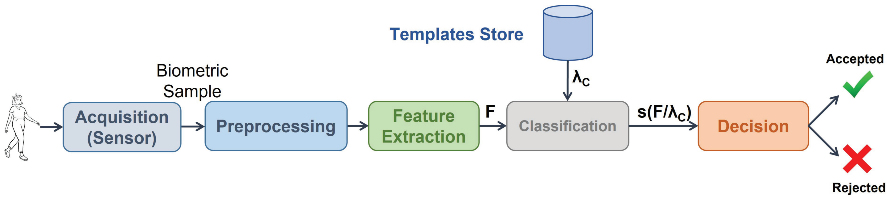


Figure 1. Main modules in a biometric system, where F stands for the feature vector (mathematical representation of the biometric sample), λ_C stands for the template (model) of the user C to be authenticated (Claimant), and $s(F/\lambda_C)$ is the score (classifier output), which is a measure of the degree of belonging of the biometric sample to the user.

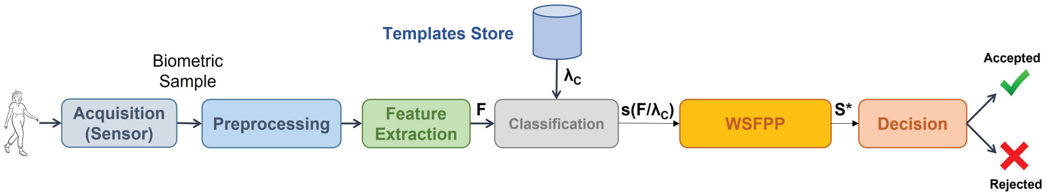


Figure 2. Proposal of system with Window Score Fusion Post-Processing (WSFPP) stage.

The rest of the paper is organized as follows. In Section 2, the state of the art in gait recognition by means of wearables is analyzed, focusing on the configurations of the systems proposed. After describing the reference biometric system (a state of the art system) in Section 3, our novel WSFPP proposal is set out in Section 4. The experimental methodology can be seen in Section 5. The results are shown in Section 6, followed by their discussion in Section 7. To finish, the conclusions are given in Section 8.

The above shows the main narrative of the work, which is complemented with the content included in two Appendices: an original proposal for gait cycle extraction in Appendix A and another for signal cleaning in Appendix B.

2. Related Works

Different approaches to biometric recognition using wearables can be found in the literature [5–12]. However, those based on gait can be considered one of the most important and popular due to its ease of acquisition: all current wearables have inertial sensors, which are the signals usually used for gait recognition [13].

From the many related works that can be found, only a very small number include post-processing stages. Of these, to the best of our knowledge, none have proposed a technique similar to our proposal (Section 4). We have found two main related approaches, both based on *Fusion of features* [24]:

- At the score level, using scores of different classifiers [32,33].
- At decision-level, with the voting scheme being the most frequently used [18,19,32].

Another important consideration is that most of the works about wearables and gait recognition use non-commercial devices. This can be seen in the public databases available: OU-ISIR [34], ZJU-GaitAcc [35], HuGaDB [36], or that acquired in [37], where devices built by the authors or simulated by means of Wii Remote are used. An exception can be found in the most recent WISDM dataset [18], but with the limitation of having only a single acquisition session.

From our point of view, the use of non-commercial wearables makes it difficult to extrapolate the results to real applications, so we are interested in works that use smartwatches or smartbands to test our proposal, since these are the ones most closely related to our approach. From a review of the literature, the following works using commercial wearables were found: [18–21,38–46].

In [20,39,43,46], the task addressed is identification, not verification. The approach of both problems is different, even the error measure is different. Here, we are interested in verification, and only the works that address this problem are used as reference.

Taking into account the aforementioned, Table 1 gives a brief description of the state of the art in the problem approached. In this table, besides the works that use commercial wearables, some recent and relevant ones that do not use these have been added to show a more complete view of the current state of the topic.

Table 1. Brief description and best results with state-of-the-art proposals.

Work	Device	Classifier	Features	Window Size	Performance
Verma_22 [38]	WISDM-Database	Random Forest	Time domain: statistical, max-min value, time between peaks	10 s (10 cycles **, approximately)	EER *** = 11%
Vecchio_22 [40]	TicWatch E2	K-NN *	Based on [47]	Based on [47]	EER = 5%
Lee_22 [48]	Own-built wrist device	SVM	2D cyclogram features	Tested from 1 to 9 cycles	EER = 5.8%
Cola_21 [47]	Shimmer3	SVM *	Time domain: statistical and autocorrelation-based	Tested from 2 to 6 cycles (called gait segment)	EER = 3.5%
Giorgi_21 [41]	WISDM-Database	RNN *	raw data	2.56 s	EER = 2.4%
Kececi_20 [49]	Own-built	Ripper, MLP *, Random Forest, Decision Tree, k-NN, Bagging, Linear Regression, Random Tree, Naive Bayes, Bayesnet	Not found	Not found	FNMR = 0.3% FMR = 0.01%
Cheung_20 [42]	Smart-Watch	SVM	Time domain: statistical features	10-sample	EER = 6%
Weiss_19 [18]	Smart-watch	k-NN, Decision Tree, Random Forest	Time domain: statistical, max-min value, time between peaks	10 s (10 cycles, approximately)	EER = 6.8%
Musale_19 [44]	Smart-watch	Random Forest, K-NN, MLP	Time domain: statistical, correlation-based, physical (pitch, roll and yaw), force	Tested from 1 to 10 cycles	EER = 8.2%
Al-Naffakh_18 [19]	Smart-band	MLP	Time domain: statistical, correlation-based, max-min value, peaks-based	10 s (10 cycles, approximately)	EER = 0.05%
Wu_18 [50]	Own-built	SVM, ANN *, k-NN	Time domain (statistical, correlation, power, max-min) + Frequency domain (mean frequency, Bandwidth, Entropy) + Wavelet-domain (FFT Coefficient, Wavelet Energy)	Tested from 2 to 11 s (cycles, approximately)	FNMR = 5.0% FMR = 4.7%
Xu_17 [45]	Smart-watch	Sparse Fusion	Sparse Fusion Classification	Tested from 1 to 6 cycles for identification task and fixed to 8 cycles for verification task	EER = 3.1%
Johnston_15 [21]	Smart-watch	MLP, Random Forest, Rotation Forest, Naive Bayes	Time domain: statistical, time between peaks, max-min	10 s (10 cycles, approximately)	EER = 1.4%

* SVM: Support Vector Machine. MLP: Multilayer Perceptron. k-NN: K Nearest Neighbor. ANN: Autoencoder Neural Network. RNN: Recurrent Neural Network. ** Cycle is defined in Section 3 and Window (cycles set) in Section 3.3. *** Performance measures are defined in Section 5.4.

3. State-of-the-Art (Reference) System

In this section, we present a state-of-the-art biometric system (Figure 1) in gait recognition by means of wearables, which will be used as reference to test our proposal. This system is based on the most relevant related literature (review shown in the previous section).

Before continuing, it is important to define two concepts to be used from now on:

- **Biometric sample**, or **sample** for short: the analog or digital representation (Figure 3) of biometric characteristics [51] (Figure 1), gait in our case. This is specified in more detail in Section 5.2.
- **Gait cycle**, or **cycle** for short: time period from when one foot contacts the ground until the same foot again contacts the ground (Figure 4). Therefore, the sample is a periodic signal where each period corresponds to a cycle (Figure 3).

For greater readability, some parts are described in the final appendixes.

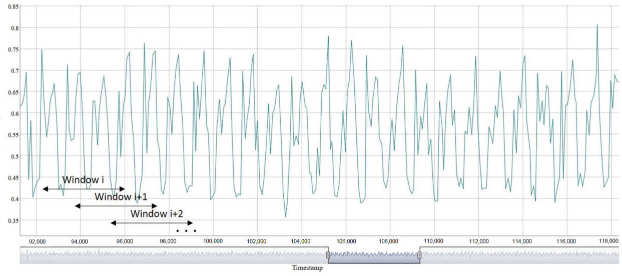


Figure 3. Example of sample windowing.

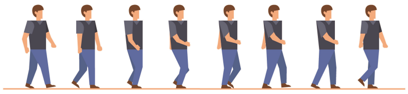


Figure 4. Gait cycle. Free image download from www.vecteezy.com (accessed on 4 November 2022).

3.1. Acquisition

While the user is walking, the raw signal from each coordinate (X, Y and Z) of the accelerometer sensor of the wearable is captured. This signal is divided into cycles, approaching this operation in a different way here from what appears in the literature. Our original proposal is shown in Appendix A.

3.2. Preprocessing

In this stage, the biometric sample is adapted and enhanced to improve its performance in the following stages. This stage encompasses the following tasks:

- **Sample cleaning,** which consists of eliminating the noisy parts of the signal and detecting and correcting acquisition errors. Not much work has been carried out into acquisition problems with real devices, so this has led us to propose our own alternative, which can be seen in Appendix B.
- **Period Normalization.** With real devices, it is not possible to ensure a fixed sampling rate. This can be seen in Figure 5, where the distribution of the time between two consecutive datum of our database is shown. To fix this, the sample must be resampled [13,21]. To perform this operation, the following must be set: (i) the interpolation method and (ii) the sampling rate. For the first, and following the literature [13], linear interpolation has been used. For the second, after analyzing the frequency components of the data, we saw that components bigger than 6 Hz were negligible; so, following the Nyquist–Shannon sampling theorem, a sampling rate of 12 Hz (a period of 83.3 ms) was fixed. This value is in accordance with that shown in [52], where it is demonstrated that the arm moves at a maximum of 8.6 Hz, making the movement as fast as possible. As our data are collected from walking, the sampling rate selected seems reasonable.
- **Amplitude Normalization.** The goal is to change the value of the data to a common scale [13]. The need to perform this operation is machine learning algorithm dependent, so the default option for each classifier in the software used for the experiments (RStudio) is used.

- **Filtering** to soften the signal. Here, one of the most commonly employed algorithms [13,45,53], the Weighted Moving Average (WMA), has been used. This algorithm is defined as: $wma = \frac{d_{t-1}+d_t+d_{t+1}}{3}$, with d_t, d_{t-1}, d_{t+1} being the data at the instant t, the instant before and the instant after, respectively.

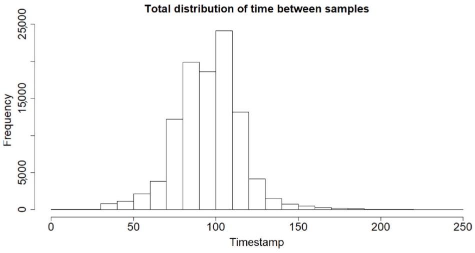


Figure 5. Histogram of the time interval between two consecutive datum. Timestamp is measured in milliseconds.

3.3. Feature Extraction

The output of this stage is a mathematical representation of the biometric sample, suitable to be processed by a learning algorithm. For this, as can be seen in Table 1, most of the works propose to extract features in the time domain. However, since the sensor signal is a time series, feature extraction in the frequency domain has also been suggested [50,54]. Here, we test and compare the results with both.

The feature extraction from a sample is accomplished as follows:

1. The cycles of the sample are grouped into segments called windows, with 20% of overlap [45] (Figure 3). Therefore, a window, w_j, is composed of m consecutive gait cycles. A cycle is a piece of signal that is too short to be representative of the user's gait, so these are grouped into windows, which, from now on, will be the basic unit of information used to model and recognize the user.
2. From each window, w_j, a feature vector, F_j, is extracted as shown in the next two sections.

3.3.1. Time Domain Feature Extraction

Following the literature [2,18,45,47,55], the following features are extracted: mean, median, maximum, minimum, standard deviation, maximum range (maximum–minimum), kurtosis, 25th percentile, 75th percentile, asymmetry coefficient, energy, and maximum value of the autocorrelation.

3.3.2. Frequency Domain Feature Extraction

First, the Fast Fourier Transform (FFT) is applied to each window to convert the signal to a representation in the frequency domain. After eliminating the zero component, the following features are extracted: the same statistics as in the time domain plus the maximum amplitude, the second maximum amplitude, the first dominant frequency, the second dominant frequency, and the under curve area.

3.4. Classification

Biometric verification is a binary classification problem, where the goal is to classify input data as belonging or not belonging to a certain target class, in our case, the claimant (subject to be authenticated) identity. For this, following what performed in biometrics, a different classifier, λ_C, is trained for each subject or claimant C. Its output (score), $s(F/\lambda_C)$, with F as the output of the previous stage, will be a measure of the degree of belonging of the biometric sample to the user.

As can be seen in Table 1, a wide variety of machine learning algorithms of diverse types have been used. From these, the most used (used in three or more works) are the Support Vector Machine (SVM), Multilayer Perceptron Artificial Neural Network (MLP), K-Nearest Neighbor (K-NN), and Random Forest (ensemble method based on Decision Trees).

Except for K-NN, the rest of the classifiers must be trained with samples of the claimant and examples of the "non-claimant" class, i.e., "other" subjects, which in biometrics is called the impostor class. The data used for this classifier training are described in Section 5.3.

4. Window Score Fusion Post-Processing Proposal

Our system proposal is shown in Figure 2, where the new post-processing stage for the window score fusion is added to the *reference system*.

The classifiers proposed, both here and in the bibliography, have the problem that the temporal relation between windows is lost, i.e., that each window is treated separately. In a temporal series, as the gait is, it is normally important to capture the relations between consecutive events (gait cycles in our case). This is the goal with the technique proposed and described in this section, where this relation is "captured" at score level.

The scores (classifier outputs, Figure 2), $s_j = s(F_j/\lambda_C)$, of several n consecutive windows are fused to obtain the final output of the system (Figure 6) as shown in Equation (1).

$$s_l^* = f(s_k, s_{k+1}, ..., s_{k+n-1}) \tag{1}$$

where $s_j = s(F_j/\lambda_C)$ is the output of the classifier λ_C for the feature vector F_j extracted from the window w_j and $f()$ is the fusion function. Now, s_l^* will be the measure of the degree of belonging of the biometric sample to the user.

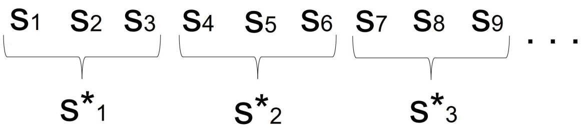


Figure 6. Example of window fusions at score level, for $n = 3$. s_j is the original score (classifier output for each sample window), while s_l^* is the new score as a result of fusing n consecutive original scores.

Focusing on the fusion function, $f()$, we can find many options in the bibliography [32,56]. Here, fusion techniques based on weighting each element to be fused (e.g., weighted sum [56]) are not suitable, since we have no prior knowledge of their values or advantages because the scores to be fused come from the same classifier. Moreover, for practical considerations, simplicity is advisable, so complex fusion techniques, e.g., [57], were discarded. Thus, based on the above and our own experience [56,58], we selected the following simple fusion techniques: mean, max, min, and median.

5. Experimental Methodology

5.1. Experiment Design

The goal with the experiments is to prove that the WSFPP proposal improves the recognition rates independently of the system configuration, i.e., to demonstrate that the WSFPP performance does not depend on the previous stages of the system, so it is suitable for any system.

For this, we start from the review of the state of the art shown in Sections 2 and 3. From this review, we select the following system/experiments configuration to test our proposal:

- With regard to the **sensor coordinate** in the **acquisition stage**, the performance of using each sensor coordinates (X, Y and Z) separately, or fusing all by means of the *module* [27,59] ($Module = \sqrt{X^2 + Y^2 + Z^2}$), is calculated and compared.
- With regard to the **window size** in the **pre-processing stage**, from the values in Table 1, the following number of cycles in the window were selected to be tested: $\{2, 4, 8, 12\}$. This set includes both small and high values, being representative of those used in the state of the art.
- With regard to the **features** extracted in the **feature extraction stage**, as shown in Section 3.3, typical features are extracted in both the time and frequency domains. In addition, this allows the performance of both approaches to be compared.

- With regard to the **classifier** in the **classification stage**, two criteria were fixed:

  1. Variety in the tested algorithms.
  2. Most used in the reference works.

  The most used were shown in Section 3.4: SVM, MLP-ANN, K-NN, and Random Forest. These also fulfill the first criterion, since their theoretical bases are completely different. A deep study of each classifier is beyond the scope of this work, so a brief description of each one is included, focusing on the main differences between them:

  – **K-NN.** Unlike the rest of the classifiers, this does not need to be trained to build a model. The training or enrollment sample(s) (see Section 5.3) is (are) used directly to create the user template, λ_C. More specifically, the user template is made up of the feature vectors extracted from the enrollment sample(s): $\lambda_C = \{F_i^e\}\ 0 \leq i \leq N$, where N is the number of windows of the enrollment sample(s). The classifier output is based on distance; to be precise, given a test trial feature vector F_j^t (see Section 5.3), its score is calculated as shown in Equation (2).

  $$s(F_j/\lambda_C) = Mean(kmin_i(EuclideanDistance(F_j^t, F_i^e))) \tag{2}$$

  Since the K-NN output is based on distance, the $s(F/\lambda_C)$ interpretation is as follows: the lower its value, the higher the degree of belonging of the biometric input to the user.

  – **SVM** [60]. This classifier is based on separating two classes by means of an optimal hyperplane $\vec{w}^T\vec{x} = 0$. The parameters of the hyperplane are fixed using the so called "support vectors" (Figure 7a). To avoid overfitting, a soft margin solution is used in the training phase (calculation of hyperplane parameters), allowing "mistakes" in the training samples (Figure 7a); this is controlled by the regularization parameter C: a small C allows a large margin of mistakes, while a large C makes constraints hard to ignore. With the hyperplane set, the classification is performed as shown in Equation (3).

  $$\begin{aligned} \vec{w}^T\vec{x} \geq 0 \Rightarrow Class\ A \\ \vec{w}^T\vec{x} < 0 \Rightarrow Class\ B \end{aligned} \tag{3}$$

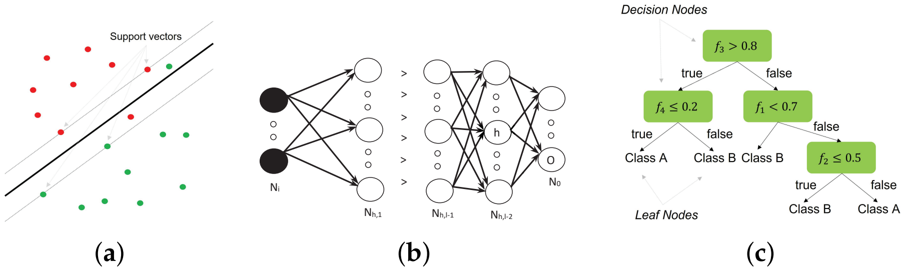


Figure 7. (**a**) SVM classifier. (**b**) MLP with l layers. (**c**) Decision tree example diagram.

  The problem is that real-world data are rarely linearly separable. The solution is to increase the dimensionality of the feature space, aiming to map the input space into a linear separable one, where the linear classifier will be applicable. This is performed by means of the "kernel trick"; i.e., a kernel function (e.g., lineal: $k(x, y) = x \cdot y$, radial: $k(x, y) = e^{\gamma\|x-y\|^2}$) is used, allowing the mapping to be performed without increasing the complexity of the training algorithm.
  As can be seen in Equation (3), the sign of the output is used to classify the input. However, here we need a score, i.e., a level of belonging to each class. The Platt scaling [61] is used to accomplish this. Therefore, the score here is a probability, $s(F_j/\lambda_C) = P(F_j/\lambda_C)$. Therefore, with a different interpretation regarding K-NN,

the higher its value, the higher the degree of belonging of the biometric input to the user.

- **MLP** [62]. This is a net composed of a set of neurons or units organized in layers (Figure 7b). The architecture of the net is defined by the number of layers and neurons in each layer. Each neuron in a layer is connected (its outputs are the inputs) with all the neurons of the following layer, except the last one, whose neurons will be the output(s) of the net. The first layer is the input of the net, which will be the feature vector. The operation performed for each neuron is that shown in Equation (4), where y_h^p is the output of the neuron h for the input p, w_{jh} is the weight (real number) that connects the neuron h with the neuron j of the previous layer, y_j^p is the output of this neuron j for the input p, and θ_h is the *bias* or *offset* of the neuron h. $\mathcal{F}$ is a function that must be derivable; typical functions are the sigmoid or the hyperbolic tangent.

$$y_h^p = \mathcal{F}(\sum_j w_{jh} y_j^p + \theta_h) \tag{4}$$

During the learning or training stage, the weights of all the neurons are set using the backpropagation algorithm, so that the value of an error function, E, will be minimized. The most common error function is the squared error, $E = \frac{1}{2}(d_o^p - y_o^p)^2$, where d_o is the desired output for the output neuron of the net o for the input p and y_o is the output of the neuron.

In our problem, the net has a single neuron in the output, being trained to obtain 0 (the desired output) for training examples of the impostor class and 1 for training examples of the subject (authentic class), using the sigmoid as the activation function, $\mathcal{F}$. Therefore, for the MLP, $s(F_j/\lambda_C) = y_o^{F_j} = \mathcal{F}(\sum_j w_{jo} y_j^{F_j} + \theta_o)$, with j being each neuron of the last hidden layer ($l - 2$ in Figure 7b). Although the output is not really a probability, due to the values of the desired outputs, it can be considered as such, so its interpretation is the same as that seen with SVM.

- **Random Forest** [63]. This is an ensemble of relatively uncorrelated *decision tree* classifiers. A decision tree is a supervised classifier that has a flowchart-like tree structure (Figure 7c); each *decision node* represents a decision rule, finishing in the *leaf nodes* with the final decisions. This tree is constructed following the algorithm below:

  1. Using Attribute Selection Measures (ASM), select the best feature (attribute) to split the dataset.
  2. Create the decision node with the corresponding decision rule. If the node is the first, it is called the *root node*.
  3. Using the decision rule, divide the corpus into subsets.
  4. Repeat the previous steps recursively for each subset until the nodes cannot be further classified due to all of the subset belonging to the same feature value, due to there being no more features, or due to there being no more data.

  Based on this classifier, Random Forest works as follows:

  * Training stage:
    1. Split the training set randomly into subsets with replacement.
    2. Train a decision tree with each subset.
  * Prediction or test stage:
    1. Each tree predicts a class.
    2. Probabilities are calculated from these classes using the predictions.

  Therefore, for this classifier, $s(F_j/\lambda_C) = P(F_j/\lambda_C)$.

These classifiers were those selected for the tests. In addition to the above, as will be seen in the results, the performance of the selected classifiers is very different, which confirms the variety of the selections.

For their configuration, we tried to use the default options of the software used (RStudio) as much as we could; the reason is to avoid possible bias in the results when optimizing the classifier, since our goal here is not to obtain the best results, but to test our proposal in the most objective way. Under this criterion, only the following particular configurations were posed:

- SVM: radial kernel. From previous experiments, this kernel showed the best performance. R library used: e1071.
- MLP-ANN: *JE_Weights* initialization function was selected. Others were tested, but the system showed inconsistencies. R library used: RSNNS.
- K-NN: $k = 1$ was selected. As with the SVM, from previous experiments, this value showed the best performance, and it is the simplest configuration. R library used: FNN.
- Random Forest: no particular configuration was used in this case. R library used: randomForest.

With regard to the WSFPP technique, the following values of n (number of consecutive windows fused) were tested: $n = \{2, 4, 8\}$. From the different fusion functions proposed, we decided, for clarity in the exposition, to select one: as can be seen in the next section, the results are clear, and adding more comparisons would not, in this case, provide new information and would only complicate the reading of the results. From the preliminary experiment, we saw that *min* and *max* performed worse than *median* and *mean*. The performance of these last two was similar, but *median* was slightly better, so this was chosen as the fusion function to test our proposal.

To achieve the goal posed, the following procedure was used:

- The performance of the reference system (Section 3) was calculated in all of the proposed system configurations.
- The performance, when our proposal was used (Section 4), was also calculated in all of the proposed system configurations, and for the different values of n.
- Both results, under the same system configurations, were compared.
- For objective results, the experimental conditions were the same in all of the experiments performed.

As can be seen (and as can be seen in the results), we compared the performance of using and not using (i.e., the reference system) our proposal in 128 different scenarios (system configurations), and for each one, with three different configurations (values of n) of our proposal.

5.2. Corpus Data Acquisition

The biometric data were captured by means of a Motorola Moto 360 watch (the same smartwatch as in [44]). This device has several sensors, and the data from all of them were collected to be used in future research. In this work, we focus on data from the 3D-accelerometer.

The data were acquired in a normal walking scenario [37,45], the most interesting one in our opinion, since it is a very common daily activity.

To study dependence over time, two different sessions were held, with a minimum separation of two weeks between them; in most cases, the separation was greater than one month, even two to three months for several subjects. As can be seen, the database includes a wide range of time intervals between sessions.

In each session, the subjects walked twice with the device on their wrist, on average 4 min for each walk, with a rest of about three to five minutes between them; therefore, in each session, two *biometric samples* were captured. The walks were outdoors, in different places. The place, walking speed, and type of footwear or clothing between sessions were not controlled; the only variable that was controlled was that the surface was flat (road or sidewalk).

In the end, data from 38 volunteer subjects, 25 men and 13 women, with a wide age range from 16 to 57 years, were captured. As indicated above, each sample consists of the 3D-accelerometer data collected.

5.3. Experimental Sets

Two main scenarios can be considered with the corpus:

- **Short period authentication**: the enrollment and testing samples belong to the same session.
- **Long period authentication**: the enrolment and testing samples belong to different sessions. Testing under this condition is critical as user behavior is different from day to day.

In this work, we only examine the second scenario, as, although less favorable, it is the most realistic. This conditions the division of the database into training and testing, as shown in the following.

We train a different classifier, λ_C, per each subject C in the database using:

- **Enrollment data** (genuine class training set). "Enrollment" is, in biometrics, the step where a subject (claimant) C supplies the biometric data to build their biometric model or template, λ_C. In pattern recognition terminology, they are called *training data*. The samples used to build this model or template are called *biometric enrolment data record* (*enrolment data* in short from now on). The first sample captured is used for enrolment data, i.e., the first sample of the first session of each subject, as is usual in biometrics.
- **Cohort set** (impostor class training set), used to train the classifiers with examples of impostors. This set must be completely different from the impostor class test set in order to obtain objective results. Thus, we randomly split the individuals in the database different from the claimant into two different sets; one for training (cohort set), and the other for testing, as shown below. One sample is randomly selected from each individual in the cohort set. For objective comparisons, the cohort set so formed is always the same throughout all the experiments.

Once the subject model λ_C had been trained, the tests were performed as shown in the following:

1. First, for each subject in the database, we selected the test samples:
    - **Genuine test samples** (for biometric mated comparison trials [51]). For these tests, we used the two samples of the second session of the claimant.
    - **Impostor test samples** (for biometric non-mated comparison trials). For these tests, we used *random forgeries*, i.e., a set of individuals in the database different from the claimant playing the role of impostors (system attackers); this is common in most biometric characteristics for technology evaluation, including gait, e.g., in [42,47], to cite two recent ones. For impostors, as mentioned, we used the subjects of the database different from the claimant not used in the *cohort set*. From each of these individuals, one of their samples was randomly selected to form this set, a set that is the same throughout all the experiments in order to achieve objective comparisons.
2. For both genuine and impostor tests, the corresponding mated and non-mated trials for each subject C are accomplished from each test sample as follows:
    (a) The test sample is windowed, i.e., their cycles are grouped as shown in Section 3.3.
    (b) From each window, w_j, its corresponding feature vector, F_j, is extracted.
    (c) The corresponding score (classifier output), $s_j = s(F_j/\lambda_C)$, is calculated. This output is a *comparison score* [51].

    Therefore, for each test sample, we have as many comparison scores or test trials as windows into which it is divided.

3. With these scores, two sets are created for each claimant C:
    - One Test Set (TS) with genuine comparison scores, TS_g^C, achieved from the *genuine test samples*;
    - Another test set with impostor comparison scores, TS_i^C, achieved from the *impostors test samples*.

    As reference, Table 2 shows the total number of tests performed for each window size, joining the corresponding test sets of all claimants.
4. The system performance is calculated using these two sets, as shown in the next Section (Section 5.4).

Table 2. Test set sizes for the different window sizes tested. Columns #Genuine and #Impostor show the number of genuine and impostor test trials, respectively.

Window Size	#Genuine	#Impostor
2 cycles	11,353	224,650
4 cycles	5486	108,285
8 cycles	2632	51,701
12 cycles	1729	33,824

When the WSFPP proposal is used, the sequence of steps is the same, except that a new one appears before the last. In this, the scores in TS_g^C and TS_i^C are fused in groups of n, as shown in Section 3.2. So, we have the following two sets:

- One test set with genuine comparison scores, but now, these scores will be s_l^*. The set is achieved from the scores in TS_g^C. We call this set TS_g^{*C}.
- Another test set, TS_i^{*C}, with s_l^* fused scores, but now achieved from TS_i^C.

5.4. Performance Measures

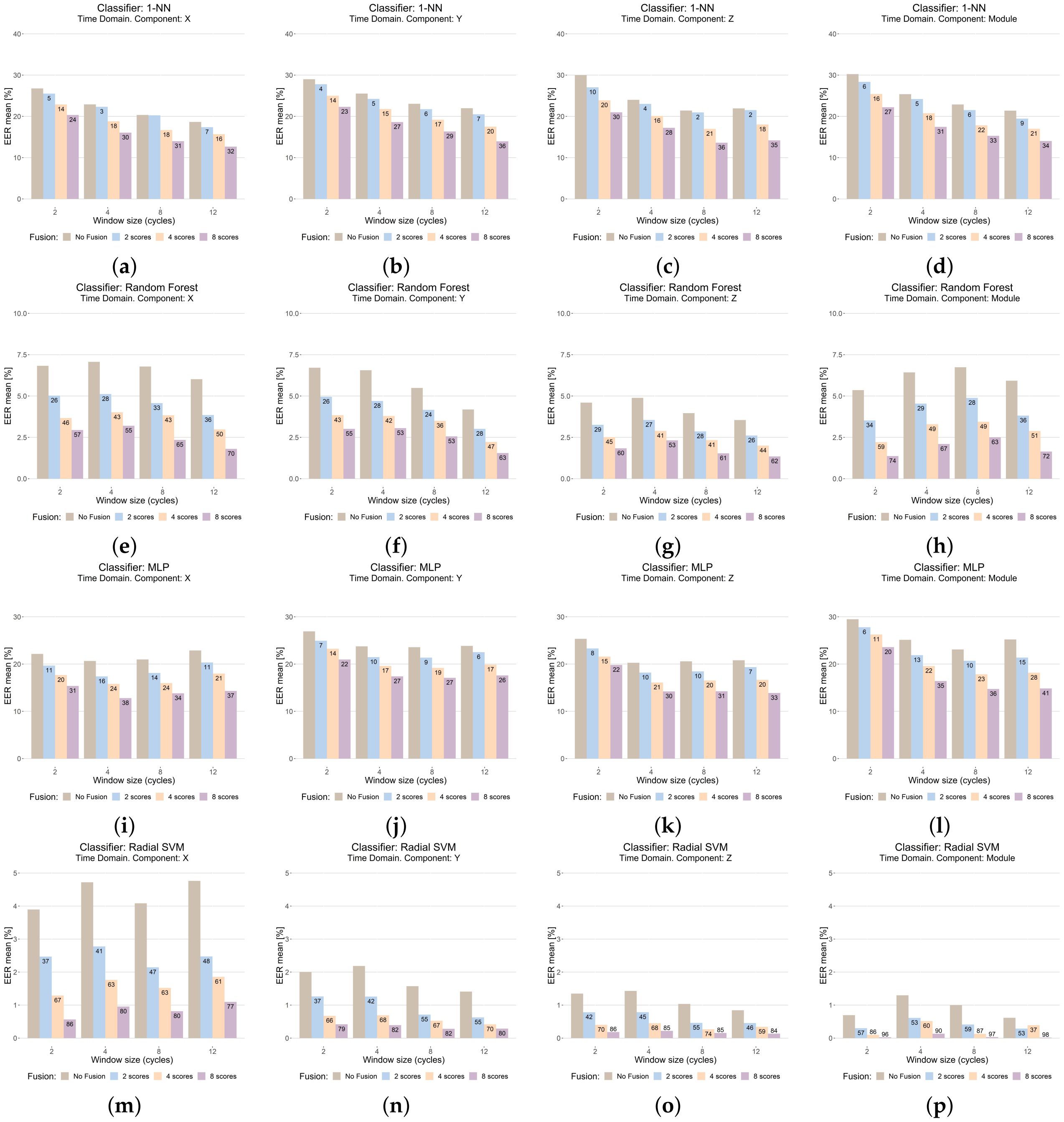
We evaluated Authentication performance using the False Match Rate (FMR, rate of impostor acceptance) and the False Non-Match Rate (FNMR, rate of claimant or genuine rejection). Since these measures are decision-threshold-dependent, graphical representations of the performance, such as the DET (Detection error trade-off) plot or the ROC (Receiver Operator Characteristic) curve, are generally used.

Nonetheless, using a single-number measure is more useful and easier to understand for a high number of comparisons. The Equal Error Rate (EER) is the most widely used in the biometrics literature. EER is the error of the system when the decision threshold is such that the FMR and FNMR are equal (in the ROC curve, the point where the diagonal cuts the curve). We used this measure here.

In order to obtain the final EER of the test, we calculated the EER for each subject C of the database, using TS_g^C (genuine test scores) and TS_i^C (impostor test scores) sets when the reference system was used, or using TS_g^{*C} and T_i^{*C} sets when our WSFPP approach was used. The final EER of the system is the mean of these EERs obtained from each subject.

6. Results

The results in the time domain can be seen in Figure 8, while those in the frequency domain are shown in Figure 9. The figures are organized into a matrix layout, where each row contains the results with the same classifier, and each column shows the results with the same sensor coordinate, including the *module*.



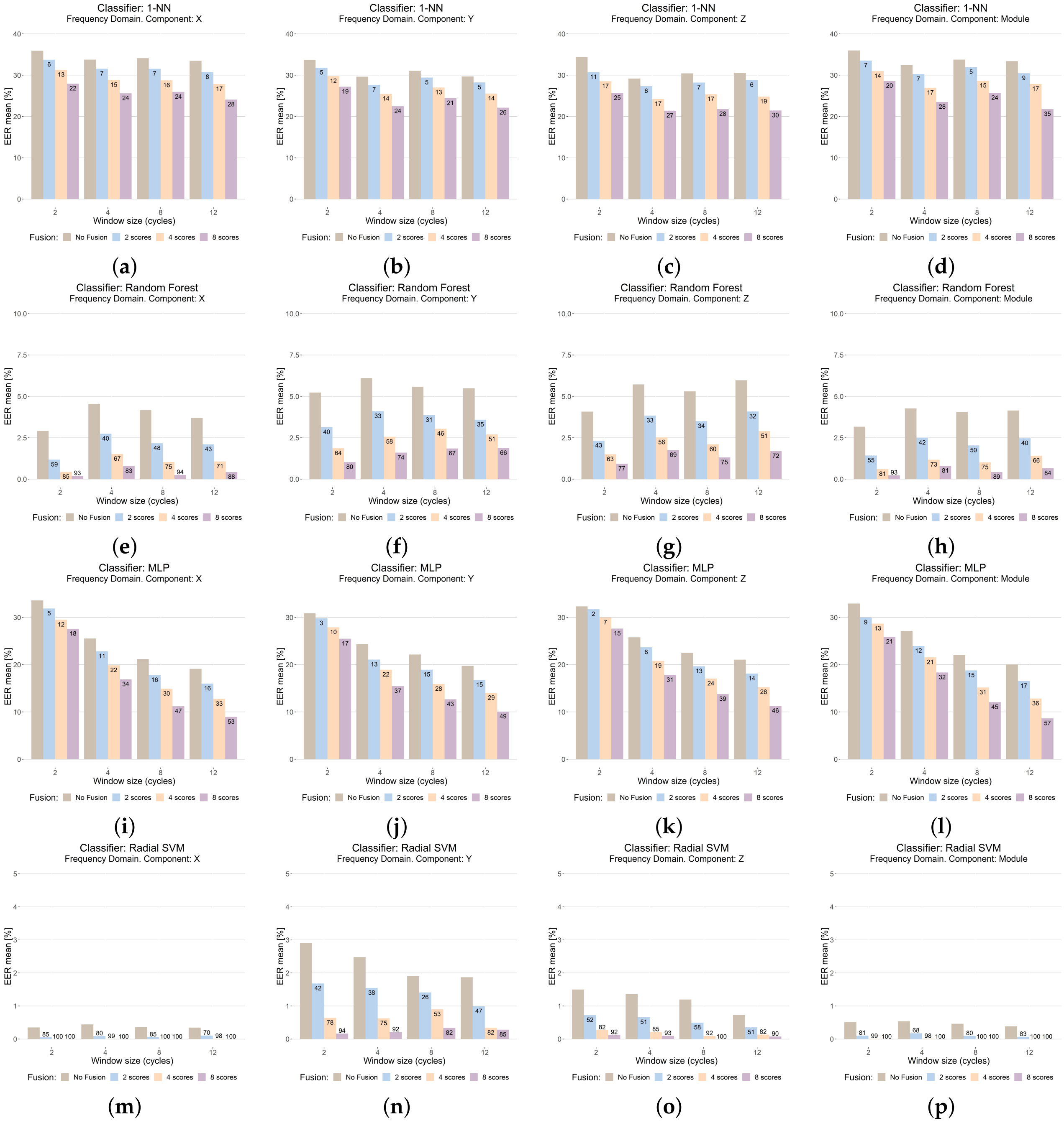
(**a**) (**b**) (**c**) (**d**)

(**e**) (**f**) (**g**) (**h**)

(**i**) (**j**) (**k**) (**l**)

(**m**) (**n**) (**o**) (**p**)

Figure 8. Results in the time domain.



(**a**) (**b**) (**c**) (**d**)

(**e**) (**f**) (**g**) (**h**)

(**i**) (**j**) (**k**) (**l**)

(**m**) (**n**) (**o**) (**p**)

Figure 9. Results in the frequency domain.

To better show and compare the data, a bar plot graph has been selected. Each bar plot contains the following information:

- The title shows the feature extraction domain, the sensor coordinate (or fusing all by means of the module), and the classifier.
- The Y-axis shows the performance (EER in %). This axis has the same scale for each classifier to better compare results.
- The X-axis shows the results for each window size, measured by number of cycles.
- For each window size, four bars are shown. The first (brown) shows the result of the reference system (WSFPP is not used). The rest show the system performance when WSFPP is used, for $n = 2$ (second bar, blue), $n = 4$ (third bar, orange), and $n = 8$ (fourth bar, purple).

For each of these three last bars, the percentage of improvement or worsening with regard to the reference system (first bar) has been added; this calculation has been performed as shown in Equation (5), where EER_{RefSys} is the performance of the reference system and EER_{WSFPP} is the performance when WSFPP is used.

$$((EER_{RefSys} - EER_{WSFPP})/EER_{RefSys}) * 100 \quad (5)$$

The goal of using the proposed data visualization is to make the analysis easier. The performance comparisons of the classifier and sensor coordinates can be carried out without the necessity of "scrolling", since the plots involved are on the same page: those comparisons related to the performance of the classifiers can be accomplished by comparing the results of one row with the others, while those related to the sensor coordinates can be accomplished by comparing the plots of a column with those in the other ones. Due to the great number of experiments, it has not been possible to put the results for time and frequency domains on the same page; however, in this case, these results are on consecutive pages, so the plots in the same position show the results with the same system configuration, except for the feature-extraction domain.

Focusing on each plot, the comparison of our proposal with regard to the reference system can be performed by comparing the first column of each group with the other three columns of the same group; each of these last three columns shows the results for a different value of the n (the number of fused window scores) parameter of our proposal. The analysis with regard to the window size is favored using the bar color; inside a plot, we can compare the results with the different window sizes tested, comparing the results of the bars with the same color, since only this parameter (window size) changes from one to another.

7. Discussion

From the results, the first important general conclusion is that our WSFPP proposal has improved the system performance in all of the scenarios, despite the differences in the system configurations tested. Therefore, the goal posed with the experimental study (Section 5.1) has been achieved, showing that the WSFPP technique is a successful proposal that can be widely used.

Details are provided as follows.

- **With regard to the classifier.** Although the use of WSFPP has improved the results with all, this improvement is higher the better the performance of the classifier. The classifier with the best performance with the reference system is SVM, achieving improvements with WSFPP up to 90% in a lot of cases, even reaching 100%, which allowed 0% of EER to be achieved, a result not shown in any previous work. The second best classifier is Random Forest, which also achieves important improvements (higher than 90% in some cases) when WSFPP is used. The other two classifiers show a worse performance, and although the improvements are lower, these have reached 36% with 1-NN and 57% with MLP.
- **With regard to the feature extraction domain.** The state of the art shows mainly feature extraction in the time domain (Table 1). However, the results show that the features in the frequency domain are an interesting alternative, since similar, and sometimes even better, results have been achieved with these features. Focusing on the case when WSFPP is used, the frequency domain shows higher improvements in general, which has allowed 0% of EER to be reached; except for 1-NN, the best results were achieved in the frequency domain: 0.2% with Random Forest, 9% with MLP and 0% with SVM.
- **With regard to the window size.** There is no clear tendency. Both with the reference system and with WSFPP, the performance is dependent on the rest of the system parameters (sensor coordinates, feature extraction domain, and classifier). An interesting result is that, although not always, very good performances have been achieved with

a size of two cycles, which is very small. Even more, with SVM and frequency domain features, 0% error has been achieved with this size, $n = 4$ and X coordinates; this implies that, with a signal of only about 8 s, it has been possible to recognize a person by means of their way of walking using WSFPP.

- **With regard to the sensor coordinate.** The improvements with WSFPP are similar in all of the sensor coordinates, including the module: once the rest of the parameters of system have been fixed (each row in Figures 8 and 9), the figures of the improvement are, in general, similar for the same values of n. This implies that the performance of WSFPP is independent of this parameter. Focusing on an analysis of the performance, as with the window size, it is dependent on the rest of the parameters of the system. However, if one must be selected, the best alternative is the module; the performance with this is, in the worst case, similar or slightly worst than the best with the other options (X, Y, or Z sensor coordinates), which are almost always better.
- **With regard to the value of n in the WSFPP proposal.** The first important aspect to note is that, with all of the values, the system performance has improved. This improvement is higher the higher the number of fused scores, n, is. Nevertheless, when the reference system has a good performance, very good results have been achieved with low values of n, e.g., with module, a window size of 2 cycles, frequency domain features and Random Forest (EER = 0.2% for $n = 4$), or with the X coordinate, a window size of 2 cycles, frequency domain features, and SVM (EER = 0.05% for $n = 2$).

As a summary, results from the conclusions that can be extracted from the above can be seen in Table 3: the best option is to use WSFPP with $n = 8$, in the acquisition, to fuse the sensor coordinates by means of the module and to use features in the frequency domain except for 1-NN. In each case, the best window size is selected. Although the results in the table are not the best, they are representative of the improvements achieved by applying our WSFPP proposal to a state of the art system.

Table 3. Results summary. Column *WS* shows the window size, *RF* the result with the Reference System, and *WSFPP* when our proposal is used with $n = 12$. *Module* is used in the acquisition stage and frequency domain in the feature extraction stage, except for 1-NN, which performs better in the time domain.

Classifier	WS	RF	WSFPP
1-NN	12 cycles	21.4%	14%
Random Forest	2 cycles	3.2%	0.2%
MLP	12 cycles	20.1%	8.6%
SVM	12 cycles	0.4%	0%

8. Conclusions

In this paper, a new proposal in biometric gait recognition, the Window Score Fusion post-processing technique, has been shown and successfully tested.

Following the state of the art, the proposal has been widely tested with different system configurations in all of the stages of the biometric system, with the goal of proving that the WSFPP proposal improves the recognition rates independently of the system configuration.

Improvements higher than 90% were achieved, e.g., 94% (from 3.2% to 0.2%, Figure 9e), with Random Forest or 100% (from 0.4% to 0%, Figure 9m) with SVM, have been achieved. In the worst cases (using K-NN and MLP as classifiers), the improvements achieved were not less than 30%, e.g., 36% (from 21.4% to 13.6%, Figure 8b) with K-NN, or 57% (from 20.1% to 8.6%, Figure 9l) with MLP.

From the results, it can be concluded that the proposed goal has been achieved, since our proposal has improved the recognition rates in all of the scenarios tested, showing that WSFPP is an interesting proposal that can be widely used.

The very good results achieved in user authentication, by means of biometric gait, allow us to predict a good performance of WSFPP in similar tasks. Among these, we propose, as interesting lines of future works, to approach identification with the same characteristic (gait) or recognition, in general, with other related biometric characteristics, for example, user recognition by means of Electrocardiogram (ECG), a characteristic that can also be captured by means of wearables.

Author Contributions: Conceptualization, I.S.-O., C.V.-P. and A.S.-H.; methodology, I.S.-O. and C.V.-P.; software, I.S.-O.; validation, I.S.-O., C.V.-P. and A.S.-H.; investigation, I.S.-O.; resources, I.S.-O.; data curation, I.S.-O.; writing—original draft preparation, I.S.-O. and C.V.-P.; writing—review and editing, I.S.-O., C.V.-P. and A.S.-H.; visualization, I.S.-O.; supervision, C.V.-P. and A.S.-H.; project administration, C.V.-P. and A.S.-H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study did not require institutional ethical approval.

Informed Consent Statement: All procedures performed in this study were in accordance with ethical and local legal regulations. Before the study, all participants were informed of its purpose, and it was guaranteed that their participation or withdrawal would incur neither reward nor punishment.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Gait Cycle Calculation

The sample is divided into gait cycles (concept defined in Section 3) and for this, we propose a novel and simple technique.

Figure 3 shows a sequence of gait cycles. Most of the systems in the literature are based on exact cycle detection [13]. Here, we follow a different approach, avoiding the problems and errors involved in detecting an exact cycle [35].

Our proposal is based on the autocorrelation of the signal. The coefficient, $R(k)$, measures the periodicity of a time signal or a portion of it, being able to detect if the signal follows a pattern or not. Equation (A1) shows the autocorrelation coefficient calculated at lag k, where d_t is the data acquired at instant t, m is the size of the signal or portion that is analyzed, and μ and σ^2 are the mean and variance of the data. The value of k, where $R(k)$ is the maximum, k_c, gives the duration of the repeated pattern, in our case, the duration of the gait cycle. This is the value that we use here to divide the signal into gait cycles, i.e., given a sample $D = \{d_1, d_2, d_3, ..., d_m\}$, a cycle is a set of c consecutive data, $\{d_j, ..., d_c\}$, so that $\sum_{i=j}^{c-1}(time\ between\ (d_{i+1}, d_i)) \leq k_c < \sum_{i=j}^{c}(time\ between\ (d_{i+1}, d_i))$.

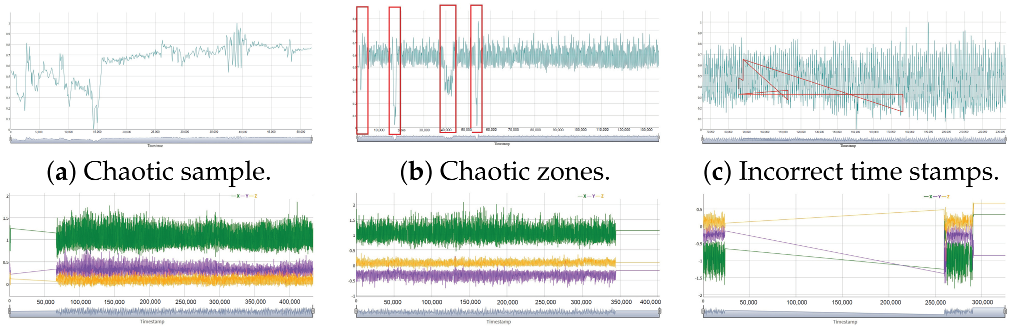
$$R(k) = \frac{1}{m \cdot \sigma^2} \cdot \sum_{t=1}^{m-k} (d_t - \mu)(d_{t+k} - \mu) \tag{A1}$$

In order to optimize the autocorrelation calculation, only values of k around a typical gait cycle duration (about one second) were used in Equation (A1): $830ms \leq k \leq 1245ms$.

Appendix B. Signal Cleaning

Due to the lack of works on signal cleaning with real devices, we had to visually analyze the raw signal for acquisition issues to begin with. Figure A1 shows the problems detected:

1. Loss of connection between device and mobile. This can happen at the beginning (Figure A1d), in the middle (Figure A1f), or at the end (Figure A1e). Its duration is highly variable.
2. Signal without periodicity (chaotic signal). This was observed both in the sample as a whole (Figure A1a) and in parts of it (Figure A1b).
3. Incorrect time stamps. The red line in Figure A1c shows the jumps in time. This happens in isolated data that are moved in time and was observed in a single sample of two users. We have no explanation, leaving their study for future acquisitions.



(**a**) Chaotic sample. (**b**) Chaotic zones. (**c**) Incorrect time stamps.

(**d**) Loss at the beginning. (**e**) Loss at the end. (**f**) Loss in the middle.

Figure A1. Problems detected in the acquired signals.

It is of the utmost importance to detect and eliminate these problems.

Problem 3 can be solved by detecting negative accumulated timestamp values and correcting them.

The elimination of the zones with the problems 1 and 2, *walking detection*, has been addressed in previous works (e.g., in [45] with real devices or in [47,53] with simulated ones), but with solutions based on a prior exact gait cycle detection. As our approach is different, we propose our own alternative, based on using the following measures:

- **Autocorrelation, R.** The autocorrelation of the signal or a portion of it is defined as $R = \max_k(|R(k)|)$ (Equation (A1)). This measure is between 0 and 1, $0 \leq R \leq 1$, so that a value near 0 means that the signal does not follow a pattern and a value near 1 means the opposite ($R = 1$ implies a perfect periodic signal).
  This coefficient allows chaotic zones to be detected, but not the connection loss problem, since when the signal is plain, the values of R are also high. To detect these zones, the measures in the following items were used.
- **Energy, E.** Defined as $E = \frac{1}{n} \cdot \sum_{t=1}^{n}(X_t^2 + Y_t^2 + Z_t^2)$, where X, Y and Z are the data of each accelerometer coordinate. This measure has also been used in other biometric characteristics to clean the signal and to isolate noise (low values) from the signal (high values).
- **Zero-crossing, Zc.** In a periodic signal, this is the point where the signal crosses the X axis. The number of these points has been used in other biometric characteristics (e.g., voice) under the same interpretation as E.
- **Time between consecutive datum, T.** Figure 5 shows a histogram of the time between two consecutive datum. A high value of T means problems in the acquisition.

The objective here is to fix the optimum value of each measure, so that the noise will be removed, but without any loss of signal. To achieve this, we proceeded as follows.

First, as a reference to compare the cleaning performance, three subjects were randomly selected from the database, and the noisy parts in their samples were manually detected.

The cleaning process started with the sample windowing. From each window, the values of the previous measures were calculated and compared with their mean values across the sample: a "noise" threshold was calculated for each measure as a percent of the corresponding mean value. The exception is the *Time between two consecutive data, T*, where the threshold was based on the histogram shown in Figure 5. Different thresholds were tested, and those that appear in the following list were selected, since the best ratio was achieved between false not-noisy windows removed (4.8%) and true noisy windows not removed (7.1%). The window was removed if any of the following conditions were true:

- $R < Threshold_R = 0.25 \cdot \mu_R$, with $\mu_R = \frac{\sum_{j=i}^{N} R_j}{N}$, where R_j is the autocorrelation of the window j and N is the number of windows of the sample. In the following items, a similar notation is used.
- $E < Threshold_E = 0.1 \cdot \mu_E$
- $Zc < Threshold_{Zc} = 0.25 \cdot \mu_{Zc}$

- $T > Threshold_T = 550$ ms

References

1. Hill, C. Wearables—The future of biometric technology? *Biom. Technol. Today* **2015**, *2015*, 5–9. [CrossRef]
2. Vhaduri, S.; Poellabauer, C. Multi-Modal Biometric-Based Implicit Authentication of Wearable Device Users. *IEEE Trans. Inf. Forensics Secur.* **2019**, *14*, 3116–3125. [CrossRef]
3. Wan, C.; Wang, L.; Phoha, V.V. A Survey on Gait Recognition. *ACM Comput. Surv.* **2018**, *51*, 1–35. [CrossRef]
4. Findling, R.D.; Hölzl, M.; Mayrhofer, R. Mobile Match-on-Card Authentication Using Offline-Simplified Models with Gait and Face Biometrics. *IEEE Trans. Mob. Comput.* **2018**, *17*, 2578–2590. [CrossRef]
5. Cao, Y.; Zhang, Q.; Li, F.; Yang, S.; Wang, Y. PPGPass: Nonintrusive and Secure Mobile Two-Factor Authentication via Wearables. In Proceedings of the 2020 IEEE International Conference on Computer Communications, Toronto, ON, Canada, 6–9 July 2020.
6. Diab, M.; Seif, A.; Sabbah, M.; El-Abed, M.; Aloulou, N. *A Review on ECG-Based Biometric Authentication Systems*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 17–44. [CrossRef]
7. Singh, Y.N. Human recognition using Fisher's discriminant analysis of heartbeat interval features and ECG morphology. *Neurocomputing* **2015**, *167*, 322–335. [CrossRef]
8. Liu, R.; Cornelius, C.; Rawassizadeh, R.; Peterson, R.; Kotz, D. Vocal Resonance: Using Internal Body Voice for Wearable Authentication. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2018**, *2*, 1–23. [CrossRef]
9. Peng, G.; Zhou, G.; Nguyen, D.T.; Qi, X.; Yang, Q.; Wang, S. Continuous Authentication With Touch Behavioral Biometrics and Voice on Wearable Glasses. *IEEE Trans.-Hum.-Mach. Syst.* **2017**, *47*, 404–416. [CrossRef]
10. Acar, A.; Aksu, H.; Uluagac, A.S.; Akkaya, K. WACA: Wearable-Assisted Continuous Authentication. In Proceedings of the 2018 IEEE Security and Privacy Workshops (SPW), San Francisco, CA, USA, 24 May 2018; pp. 264–269. [CrossRef]
11. Blasco, J.; Peris-Lopez, P. On the Feasibility of Low-Cost Wearable Sensors for Multi-Modal Biometric Verification. *Sensors* **2018**, *18*, 2782. [CrossRef]
12. Lin, F.; Cho, K.W.; Song, C.; Jin, Z.; Xu, W. Exploring a Brain-Based Cancelable Biometrics for Smart Headwear: Concept, Implementation, and Evaluation. *IEEE Trans. Mob. Comput.* **2020**, *19*, 2774–2792. [CrossRef]
13. Marsico, M.D.; Mecca, A. A Survey on Gait Recognition via Wearable Sensors. *ACM Comput. Surv.* **2019**, *52*. [CrossRef]
14. Gafurov, D.; Helkala, K.; Soendrol, T. Gait recognition using acceleration from MEMS. In Proceedings of the First International Conference on Availability, Reliability and Security (ARES'06), Vienna, Austria, 20–22 April 2006; pp. 6–439. [CrossRef]
15. Gafurov, D.; Snekkenes, E.; Bours, P. Gait Authentication and Identification Using Wearable Accelerometer Sensor. In Proceedings of the 2007 IEEE Workshop on Automatic Identification Advanced Technologies, Alghero, Italy, 7–8 June 2007; pp. 220–225. [CrossRef]
16. Rong, L.; Jianzhong, Z.; Ming, L.; Xiangfeng, H. A Wearable Acceleration Sensor System for Gait Recognition. In Proceedings of the 2007 2nd IEEE Conference on Industrial Electronics and Applications, Harbin, China, 23–25 May 2007; pp. 2654–2659. [CrossRef]
17. Nowlan, M.F. Human Identification via Gait Recognition Using Accelerometer Gyro Forces. Available online: https://pdfs.semanticscholar.org/a63e/04fefd2be621488646ae11bfe66c98d9649e.pdf (accessed on 23 May 2020).
18. Weiss, G.M.; Yoneda, K.; Hayajneh, T. Smartphone and Smartwatch-Based Biometrics Using Activities of Daily Living. *IEEE Access* **2019**, *7*, 133190–133202. [CrossRef]
19. Al-Naffakh, N.; Clarke, N.; Li, F. Continuous user authentication using smartwatch motion sensor data. In Proceedings of the IFIP International Conference on Trust Management, Toronto, ON, Canada, 29 June–1 July 2018; pp. 15–28.
20. Ahmad, M.; Alqarni, M.A.; Khan, A.; Khan, A.; Chauhdary, S.H.; Mazzara, M.; Umer, T.; Distefano, S. Smartwatch-Based Legitimate User Identification for Cloud-Based Secure Services. *Mobile Inf. Syst.* **2018**. [CrossRef]
21. Johnston, A.H.; Weiss, G.M. Smartwatch-based biometric gait recognition. In Proceedings of the IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS), Arlington, VA, USA, 8–11 September 2015; pp. 1–6. [CrossRef]
22. Mansfield, A.J.; Wayman, J.L. *Best Practices in Testing and Reporting Performance of Biometric Devices*; Technical Report; Centre for Mathematics and Scientific Computing, National Physical Laboratory: Middlesex, UK, 2002.
23. Mandalapu, H.; N, A.R.P.; Ramachandra, R.; Rao, K.S.; Mitra, P.; Prasanna, S.R.M.; Busch, C. Audio-Visual Biometric Recognition and Presentation Attack Detection: A Comprehensive Survey. *IEEE Access* **2021**, *9*, 37431–37455. [CrossRef]
24. Jindal, G.; Kaur, G. A Comprehensive Overview of Quality Enhancement Approach-Based Biometric Fusion System Using Artificial Intelligence Techniques. In *Communication and Intelligent Systems*; Sharma, H., Gupta, M.K., Tomar, G.S., Lipo, W., Eds.; Springer: Singapore, 2021; pp. 81–98.
25. Villalba, J.; Chen, N.; Snyder, D.; Garcia-Romero, D.; McCree, A.; Sell, G.; Borgstrom, J.; Richardson, F.; Shon, S.; Grondin, F.; et al. State-of-the-art Speaker Recognition for Telephone and Video Speech: The JHU-MIT Submission for NIST SRE18. In Proceedings of the Interspeech 2019, ISCA, Graz, Austria, 15–19 September 2019; pp. 1488–1492.
26. Stragapede, G.; Vera-Rodriguez, R.; Tolosana, R.; Morales, A.; Acien, A.; Le Lan, G. Mobile behavioral biometrics for passive authentication. *Pattern Recognit. Lett.* **2022**, *157*, 35–41. [CrossRef]
27. Vhaduri, S.; Dibbo, S.V.; Cheung, W. HIAuth: A Hierarchical Implicit Authentication System for IoT Wearables Using Multiple Biometrics. *IEEE Access* **2021**, *9*, 116395–116406. [CrossRef]
28. Duda, R.O.; Hart, P.E.; Stork, D.G. *Pattern Classification*; John Wiley & Sons: Hoboken, NJ, USA, 2001.

29. Saitoh, T.; Zhou, Z.; Zhao, G.; Pietikäinen, M. Concatenated Frame Image Based CNN for Visual Speech Recognition. In *Computer Vision—ACCV 2016 Workshops*; Chen, C.S., Lu, J., Ma, K.K., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 277–289.
30. Gonzalez-Ferreras, C.; Escudero-Mancebo, D.; Vivaracho-Pascual, C.; Cardenoso-Payo, V. Improving Automatic Classification of Prosodic Events by Pairwise Coupling. *IEEE Trans. Audio Speech Lang. Process.* **2012**, *20*, 2045–2058. [CrossRef]
31. Tekampe, N.; Merle, A.; Bringer, J.; Gomez-Barrero, M.; Fierrez, J.; Galbally, J. *D6.5: Towards the Common Criteria Evaluations of Biometric Systems*; Technical Report, Biometrics Evaluation and Testing (BEAT) Project; 2016. Available online: https://konfidas.de/downloads/D6.5.pdf (accessed on 23 November 2022).
32. Sprager, S.; Juric, M.B. Inertial Sensor-Based Gait Recognition: A Review. *Sensors* **2015**, *15*, 22089–22127. [CrossRef]
33. Sun, H.; Yuao, T. Curve aligning approach for gait authentication based on a wearable accelerometer. *Physiol. Meas.* **2012**, *33*, 1111–1120. [CrossRef]
34. Ngo, T.T.; Makihara, Y.; Nagahara, H.; Mukaigawa, Y.; Yagi, Y. The largest inertial sensor-based gait database and performance evaluation of gait-based personal authentication. *Pattern Recognit.* **2014**, *47*, 228–237. [CrossRef]
35. Zhang, Y.; Pan, G.; Jia, K.; Lu, M.; Wang, Y.; Wu, Z. Accelerometer-Based Gait Recognition by Sparse Representation of Signature Points With Clusters. *IEEE Trans. Cybern.* **2015**, *45*, 1864–1875. [CrossRef]
36. Chereshnev, R.; Kertész-Farkas, A. HuGaDB: Human Gait Database for Activity Recognition from Wearable Inertial Sensor Networks. In *Analysis of Images, Social Networks and Texts*; Springer International Publishing: Cham, Switzerland, 2017; pp. 131–141.
37. Kork, S.K.A.; Gowthami, I.; Savatier, X.; Beyrouthy, T.; Korbane, J.A.; Roshdi, S. Biometric database for human gait recognition using wearable sensors and a smartphone. In Proceedings of the 2017 2nd International Conference on Bio-Engineering for Smart Technologies (BioSMART), Paris, France, 30 August–1 September 2017; pp. 1–4. . [CrossRef]
38. Verma, A.; Moghaddam, V.; Anwar, A. Data-Driven Behavioural Biometrics for Continuous and Adaptive User Verification Using Smartphone and Smartwatch. *Sustainability* **2022**, *14*, 7362. [CrossRef]
39. Luo, F.; Khan, S.; Huang, Y.; Wu, K. Activity-based person identification using multimodal wearable sensor data. *IEEE Internet Things J.* **2022**, *10*, 1711–1723. [CrossRef]
40. Vecchio, A.; Nocerino, R.; Cola, G. Gait-based Authentication: Evaluation of Energy Consumption on Commercial Devices. In Proceedings of the 2022 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops), Atlanta, GA, USA, 21–25 March 2022; pp. 793–798. [CrossRef]
41. Giorgi, G.; Saracino, A.; Martinelli, F. Using recurrent neural networks for continuous authentication through gait analysis. *Pattern Recognit. Lett.* **2021**, *147*, 157–163. [CrossRef]
42. Cheung, W.; Vhaduri, S. Context-Dependent Implicit Authentication for Wearable Device User. *arXiv* **2020**. [CrossRef]
43. Watanabe, K.; Nagatomo, M.; Aburada, K.; Okazaki, N.; Park, M. Gait-Based Authentication for Smart Locks Using Accelerometers in Two Devices. In *Advances in Networked-Based Information Systems*; Barolli, L., Nishino, H., Enokido, T., Takizawa, M., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 281–291.
44. Musale, P.; Baek, D.; Werellagama, N.; Woo, S.S.; Choi, B.J. You Walk, We Authenticate: Lightweight Seamless Authentication Based on Gait in Wearable IoT Systems. *IEEE Access* **2019**, *7*, 37883–37895. [CrossRef]
45. Xu, W.; Shen, Y.; Zhang, Y.; Bergmann, N.; Hu, W. Gait-Watch: A Context-Aware Authentication System for Smart Watch Based on Gait Recognition. In Proceedings of the 2017 IEEE/ACM Second International Conference on Internet-of-Things Design and Implementation (IoTDI), Pittsburgh, PA, USA, 18–21 April 2017; pp. 59–70.
46. Davidson, S.; Smith, D.; Yang, C.; Cheah, S.C. *Smartwatch User Identification as a Means of Authentication*; University of California San Diego: San Diego, CA, USA, 2016.
47. Cola, G.; Vecchio, A.; Avvenuti, M. Continuous authentication through gait analysis on a wrist-worn device. *Pervasive Mob. Comput.* **2021**, *78*, 101483. [CrossRef]
48. Lee, S.; Lee, S.; Park, E.; Lee, J.; Kim, I.Y. Gait-Based Continuous Authentication Using a Novel Sensor Compensation Algorithm and Geometric Features Extracted From Wearable Sensors. *IEEE Access* **2022**, *10*, 120122–120135. [CrossRef]
49. Kececi, A.; Yildirak, A.; Ozyazici, K.; Ayluctarhan, G.; Agbulut, O.; Zincir, I. Implementation of machine learning algorithms for gait recognition. *Eng. Sci. Technol. Int. J.* **2020**. [CrossRef]
50. Wu, G.; Wang, J.; Zhang, Y.; Jiang, S. Authentication Scheme Based on Physiological and Behavioral Characteristics. *Sensors* **2018**, *18*, 179. [CrossRef]
51. ISO/IEC. ISO/IEC 2382-37:2022(en) Information Technology—Vocabulary—Part 37: Biometrics. Available online: https://www.iso.org/standard/73514.html (accessed on 23 November 2022).
52. Tveit, B. Analyzing Behavioral Biometrics of Handwriting Using Myo Gesture Control Armband. Master's Thesis, Fakultet for Naturvitenskap og Teknologi, The Artic University of Norway, Tromsø, Norway, 2018.
53. Said, S.; Al-kork, S.; Nair, V.; Gowthami, I.; Beyrouthy, T.; Savatier, X.; Abdrabbo, M.F. Experimental Investigation of Human Gait Recognition Database using Wearable Sensors. *Adv. Sci. Technol. Eng. Syst. J.* **2018**, *3*, 201–210. [CrossRef]
54. Sun, F.; Mao, C.; Fan, X.; Li, Y. Accelerometer-Based Speed-Adaptive Gait Authentication Method for Wearable IoT Devices. *IEEE Internet Things J.* **2019**, *6*, 820–830. [CrossRef]
55. Lu, H.; Huang, J.; Saha, T.; Nachman, L. Unobtrusive Gait Verification for Mobile Phones. In Proceedings of the 2014 ACM International Symposium on Wearable Computers, Seattle, WA, USA, 13–17 September 2014; Association for Computing Machinery: New York, NY, USA, 2014; pp. 91–98. [CrossRef]

56. Pascual-Gaspar, J.M.; Faundez-Zanuy, M.; Vivaracho, C. Fast on-line signature recognition based on VQ with time modeling. *Eng. Appl. Artif. Intell.* **2011**, *24*, 368–377. [CrossRef]
57. Navarro, C.F.; Perez, C.A. Color-Texture Pattern Classification Using Global-Local Feature Extraction, an SVM Classifier, with Bagging Ensemble Post-Processing. *Appl. Sci.* **2019**, *9*, 3130. [CrossRef]
58. Vivaracho-Pascual, C.; Faundez-Zanuy, M.; Pascual, J.M. An efficient low cost approach for on-line signature recognition based on length normalization and fractional distances. *Pattern Recognit.* **2009**, *42*, 183–193. [CrossRef]
59. Cola, G.; Avvenuti, M.; Musso, F.; Vecchio, A. Gait-Based Authentication Using a Wrist-Worn Device. In Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, Hiroshima, Japan, 28 November–1 December 2016; MOBIQUITOUS 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 208–217. [CrossRef]
60. Kowalczyk, A. *Support Vector Machines Succinctly*; Syncfusion: Morrisville, NC, USA, 2017.
61. Platt, J.C. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. In *Advances in Large Margin Classifiers*; MIT Press: Cambridge, MA, USA, 1999; pp. 61–74.
62. Kröse, B.; Van der Smagt, P. *An Introduction to Neural Networks*, 8th ed.; University of Amsterdam: Amsterdam, The Netherlands, 1996. Available online: https://www.researchgate.net/publication/272832321_An_introduction_to_neural_networks (accessed on 23 November 2022).
63. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *1*, 5–32. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.