Towards Automated Assessment of Frailty Status Using a Wrist-Worn Device

Domenico Minici, Guglielmo Cola, Antonella Giordano, Silvana Antoci, Elena Girardi, Mauro Di Bari, and Marco Avvenuti

***Abstract*—Wearable sensors potentially enable monitoring the user's physical activity in daily life. Therefore, they are particularly appealing for the evaluation of older subjects in their environment, to capture early signs of frailty and mobility-related problems. This study explores the use of body-worn accelerometers for automated assessment of frailty during walking activity. Experiments involved 34 volunteers aged 70+, who were initially screened by geriatricians for the presence of frailty according to Fried's criteria. After screening, the volunteers were asked to walk 60 m at preferred speed, while wearing two accelerometers, one positioned on the lower back and the other on the wrist. Sensor-derived signals were analyzed independently to compare the ability of the two signals (wrist vs. lower back) in frailty status assessment. A gait detection technique was applied to identify segments made of four gait cycles. These segments were then used as input to compute 25 features in time and time-frequency domains, the latter by means of the Wavelet Transform. Finally, five machine learning models were trained and evaluated to classify subjects as robust or non-robust (i.e., pre-frail or frail). Gaussian naive Bayes applied to the features derived from the wrist sensor signal identified non-robust subjects with 91% sensitivity and 82% specificity, compared to 87% sensitivity and 64% specificity achieved with the lower back sensor. Results demonstrate that a wrist-worn accelerometer provides valuable information for the recognition of frailty in older adults, and could represent an effective tool to enable automated and unobtrusive assessment of frailty.**

***Index Terms*—Frailty, gait analysis, machine learning, wearable sensor.**

I. Introduction

IN RECENT years, clinical gerontology has devoted a growing interest to frailty, defined as an age-related condition characterized by an increased vulnerability to external stressors and an excess risk of poor clinical outcomes, such as falls and fractures, disability, hospitalization and ultimately death. Not surprisingly, frail older subjects absorbe a major share of health-care resources [1].

Within the general definition given above, diverse conceptual models and, consequently, different diagnostic tools have been proposed for frailty. One of the most accredited among them is the phenotype model, which was developed by Fried *et al.* [2] and considers decline in physical performance as the cornerstone of frailty. Its operational translation is represented by a simple tool including five items: unintentional weight loss, self-reported exhaustion, low energy expenditure, slow gait speed, and weak grip strength. Scoring positive in three or more, one or two, or none of these items classifies older subjects as frail, pre-frail, or robust, respectively.

Although this tool is simple and of rapid application, it requires some specialized clinical setup and trained personnel. Moreover, it might be hypothesized that exploring physical performance of older subjects in their own environment is more appropriate to capture frailty status, as compared to the somewhat artificial setting needed to apply this and other similar tests. For these reasons, many researchers have focused on the use of wearable, sensor-based technology to gather parameters on motor condition [3]–[6], thus obtaining an objective, ecological assessment of frailty status in older subjects [7], [8]. Furthermore, patient monitoring through wearable technology has the potential for enabling prolonged studies, thus expanding the bulk of data available for evaluation, while reducing healthcare costs and the discomfort for the patient when the assessment is done within a specific clinical setting. Among the activities that can be investigated using wearable sensors, gait plays a very important role in identifying age-related conditions. Gait analysis has been the subject of several studies in the context of personalized healthcare, and gait-related parameters derived from wearable sensors have already been associated with conditions such as frailty [9], [10]. The precise procedures and methods to apply this innovative and promising technology in the assessment of

Manuscript received January 12, 2021; revised May 12, 2021 and July 19, 2021; accepted July 20, 2021. Date of publication July 30, 2021; date of current version March 7, 2022. This work was supported in part by the Italian Ministry of Education and Research (MIUR) in the framework of the CrossLab project (Departments of Excellence) and in part by the PRA 2018_81 Project "Wearable sensor systems: personalized analysis and data security in healthcare" funded by the University of Pisa. *(Corresponding author: Domenico Minici.)*

Domenico Minici and Marco Avvenuti are with the Department of Information Engineering, University of Pisa, 56122 Pisa, Italy (e-mail: domenico.minici@phd.unipi.it).

Guglielmo Cola is with the Institute of Informatics and Telematics, IIT-CNR, 56124 Pisa, Italy.

Antonella Giordano and Silvana Antoci are with the Research Unit of Medicine of Aging, Department of Clinical and Experimental Medicine, University of Florence, 50134 Firenze, Italy.

Mauro Di Bari is with the Research Unit of Medicine of Aging, Department of Clinical and Experimental Medicine, University of Florence, 50134 Firenze, Italy, and also with Unit of Geriatrics - Geriatrics Intensive Care Unit, Department of Medicine and Geriatrics, Careggi Hospital, 50134 Firenze, Italy.

Elena Girardi is with the Unit of Geriatrics - Geriatrics Intensive Care Unit, Department of Medicine and Geriatrics, Careggi Hospital, 50134 Firenze, Italy.

Digital Object Identifier 10.1109/JBHI.2021.3100979

frailty are still under investigation, and uncertainties exist on the best position of the sensors, as well as on the parameters to be extracted and the algorithm to process them.

This study aims to devise an automated process based on machine learning, able to classify subjects according to their frailty status, using a set of gait-related parameters extracted from wearable accelerometers. In this paper, we report on the basic approach and the experimental protocol, and we highlight the importance of the position chosen for placing the sensors, comparing the performance of a wrist- and a lower back-worn sensor. We therefore propose a method for distinguishing robust (R) subjects from pre-frail and frail subjects, joined under the common denomination of non-robust (NR). This binary classification still allows geriatricians to identify the individuals who may need further clinical evaluation, namely pre-frail and frail. Indeed, early identification of the latter categories enables prompt clinical intervention, which in turn may lead to better health outcomes for the subject.

A three-stage approach is used for frailty status assessment, exploiting either the wrist or the lower-back sensor. In a first stage, gait cycle detection is executed to identify segments of four gait cycles. In the next stage, gait signals are analyzed to extract 25 features characterizing the subject's gait. Finally, the last stage uses machine learning to classify participants as R or NR.

We compared two solutions based on a single device as we aim to provide an unobtrusive solution for frailty assessment. However, in future work it will be interesting to evaluate whether the joint use of multiple sensors could further improve the assessment accuracy.

The experiments to evaluate this approach involved 34 volunteers, who were asked to walk at preferred speed under the supervision of experienced medical staff. Clinicians were also responsible for the classification of participants according to Fried's frailty phenotype. The results of clinical assessment were used to label accelerometer signals and built a ground-truth dataset that, in turn, was used to train, validate and test machine learning-based classification of frailty.

The main contributions of this study are to: (i) gain a better understanding of the possibility of using practical wearable devices in combination with machine learning-based classification to discriminate R from NR older subjects, (ii) show that walk-related features in the time-frequency domain, computed by means of Wavelet analysis, can improve performance in frailty status assessment, (iii) compare the performance of sensors applied in two different body parts, the wrist and the lower-back.

The paper is organized as follows. First, a discussion of related work is presented in Section II. Next, Section III describes the method we proposed for automatic frailty assessment based on wearable accelerometers. In Section IV, the setup of the experiment, the design choices and their validation are presented. Finally, in Sections V and VI the results achieved and the conclusions reached by this study are presented and discussed.

II. Related Work

Frailty is common in later life across different countries [11]; the reported prevalence increases progressively with aging and reaches approximately 30% in subjects aged 85+ years [12]. As pointed out in several studies [12], [13], there is a lack of consensus in the definition of frailty and on its clinical identification. A turning point in the conceptualization of frailty was represented by a study by Fried et al. who in 1991 hypothesized the existence of a stage of preclinical disability in which there is a greater risk of functional decline [14]. A few years later, Guralnik *et al.* [15] demonstrated that measures of lower extremity functioning, obtained with a simple battery of three physical performance tests, may be a hallmark of frailty: in non-disabled subjects, poorer scores on this battery predicted the subsequent development of disability in the two domains of mobility and basic activities of daily living (BADL), thus providing evidence that a state of preclinical disability can be clinically identified [16]–[18]. Further advancement was offered by Fried *et al.* in a 2001 landmark study, where the concept of frailty phenotype was introduced, accompanied by the proposal of a simple tool to recognize frailty in the presence of three or more of the following: unintentional weight loss, self-reported exhaustion, weakness, slow walking speed, and low levels of habitual physical activity [2]. A substantial body of literature has shown that older persons identified as frail by these criteria have an increased risk of accelerated functional decline, overt disability, and other poor clinical outcomes, including death. Thus, Fried's frailty phenotype has become the most commonly accepted model to identify frail older subjects [12], [15], [19], [20]. Given the importance of physical functioning and, more specifically, lower extremity mobility, in the detection of frailty, it is reasonable to hypothesize that an extended evaluation of mobility may improve our ability to identify this condition more precisely and at an earlier stage.

Inertial sensors have been proposed as powerful tools to assess functional capacity [21]–[23]. These devices allow collecting a huge amount of data on subjects' mobility in their own environment over prolonged times; data can then be automatically transmitted at a distance and be analyzed with appropriate algorithms, potentially reducing the operator-dependent variability in the assessment of frailty. In particular, accelerometers and gyroscopes have been used to measure parameters related to mobility, or to assess the risk of falls in the elderly [10], [24]–[26]. Several researchers studied the parameters recorded by body-worn inertial sensors to discriminate patients according to their frailty status. The authors in [27] described the differences between frail and non-frail older subjects using parameters extracted from inertial sensors: in particular, they found that frail elderly persons obtained lower maximum and minimum accelerations than fit individuals. The authors in [4] found that unique parameters derived from objective assessment of gait, balance, and physical activities are sensitive for the identification of pre-frailty and the classification of a subject's frailty status. In [28], Greene *et al.* concluded that it is possible to distinguish frail from non-frail subjects by using mobility tests in combination with wearable sensors.

A relevant project is represented by Frailsafe [29], which proposes a framework aimed at highlighting the potential for frailty prediction strategies based on information and communication technology (ICT). Both quantitative and qualitative measures of frailty are used to predict long-term outcomes via advanced data mining approaches applied to multi-parametric data. Frailsafe

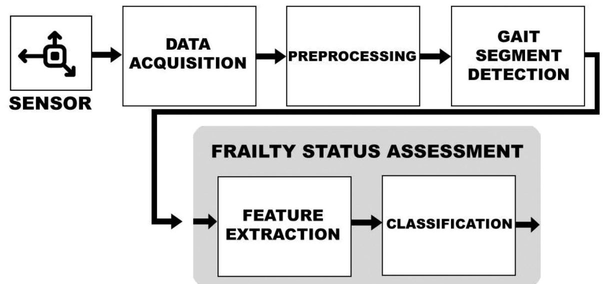


Fig. 1. Flowchart of the proposed method.

uses wearable sensors for monitoring several physical activities and physiological parameters, whereas in our work we focus on the use of gait pattern analysis for frailty assessment, which should be possibly carried out by using only one inertial device.

We argue that the position chosen for wearing the sensor can be of fundamental importance, not only for usability reasons, but also for recognizing frailty. For this reason, our study proposes a performance comparison between two different sensor's positions (lower back and wrist). In the literature, other researchers analyzed the importance of data collected from wrist-worn accelerometers in health, functional, and social assessment of older adults. Huisingh-Scheetz *et al.* used the signal extracted from wrist-worn sensors to determine how frailty and other characteristics relate to activity among older adults [30], [31]. In our case, besides using a machine-learning approach, we focus on the relationship between gait-related movements and the subject's frailty status. Notably, while in many previous works inertial signals are mainly treated as static processes, in our study we devote particular attention to the time-frequency domain by means of Wavelet analysis.

III. The Proposed Method

In this section, we give a system description of the method we propose for automatic frailty assessment. The flowchart of the method is shown in Fig. 1. At a glance, acceleration data is collected through a sensor embedded in a wearable device. After preprocessing, acceleration samples are sent to a gait segment detection module, which works by isolating sequences of consecutive gait cycles from the incoming signal trace. Once the gait segments have been detected, a subset containing the most regular walks is selected and used for the feature extraction phase. Finally, extracted features are used to feed a machine learning classifier which assesses whether the bearer is a robust (R) or non-robust (NR) subject.

A. Data Acquisition and Preprocessing

Acceleration is collected at a sampling rate of 102.4 Hz. All the acceleration components (**x**, **y**, **z** axes and acceleration magnitude **m**) are converted into g units. As body movements typically have frequency components below 20 Hz [32], acceleration components are passed through a second order Butterworth low-pass filter with a cut-off frequency of 20 Hz. Filtered acceleration is then provided to the gait segment detection module.

B. Gait Segment Detection

A gait cycle is the sequence of events that occur during the walking process between two consecutive heel strikes of the same foot. Henceforth, we will use the term *gait segment* to indicate four consecutive gait cycles. In this phase, gait segments are automatically identified by means of the walking detection algorithm described in [33], which is based on the analysis of the acceleration magnitude signal. The number of gait cycles per gait segment (four) was determined empirically. According to our experiments, the selected number of cycles is able to capture the most important information on the subject's gait pattern. At the same time, a gait segment composed of just four gait cycles will enable the detection of such segments in home environments, where spatial characteristics may make long walks unlikely. The action performed by the gait segment detection algorithm is independent of the orientation of the wearable device. However, the signals extracted from sensors positioned in different parts of the body could differ in shape or amplitude. For this reason, it is necessary to adapt the setup operation of the algorithm to the specific body position by tuning the thresholds used for peak-detection.

To prevent anomalous segments from affecting the classifiers' performance, a segment filtering procedure has been implemented based on autocorrelation. More specifically, we compute unbiased autocorrelation coefficients [34] on magnitude m samples as follows:

$$AC_k = \frac{1}{N-k}\sum_{i=1}^{N-k} r_i * r_{i+k},$$

where AC_k represents the k-th unbiased autocorrelation coefficient; k is the considered time lag; r_i represents the i-th magnitude m sample minus the average magnitude of the whole gait segment; N represents the number of samples in the whole gait segment.

Subsequently, a peak detection technique is used to find the first and the second dominant periods of the autocorrelation function (AC_DP1 and AC_DP2), which represent the estimated average duration of a step and average duration of a gait cycle, respectively [34]. The autocorrelation coefficient corresponding to the first dominant period (AC_C1) can be used to evaluate the regularity of consecutive steps, whereas the coefficient at the second dominant period (AC_C2) describes the regularity of consecutive gait cycles. The latter has been exploited to discard highly irregular gait segments. In particular, for each subject:

- the subject's gait segments are ordered in descending AC_C2 order;
- the first M segments, with the highest values of AC_C2, are selected for next phases.

From now on, let us set $M = 9$.

C. Frailty Status Assessment

The M gait segments that are not discarded in the previous phase, are used as input to *Feature Extraction*. Several features are computed in the time and time-frequency domains to capture

Algorithm 1: Frailty Status Assessment Procedure.

Result: Subject's frailty status (R or NR)
Let $G = \{g_1, \ldots, g_M\}$ be the set of subject's gait instances;
foreach $g_i \in G$ **do**
  g_i is classified as R or NR;
end
Let $T = \{t_1, \ldots, t_M\}$, of length $M = 9$, be the set of labels assigned by the classifier;
Let T_{NR} be the total number of gait segments labeled as NR;
Let k_{NR} be the non-robust parameter, and let k_{NR} be equal to $0, 40$;
$threshold \leftarrow \lceil M \cdot k_{NR} \rceil$;
if $T_{NR} \geq threshold$ **then**
  Subject is classified as NR;
else
  Subject is classified as R;
end

the main characteristics of the subject's gait pattern. Features and feature selection are described and discussed in Section IV.

Frailty status is assessed in a two-stage process by means of a machine learning model. Let us define *gait instance* the vector of features extracted from a gait segment. First, each of the subject's gait instances is classified as belonging to the NR or R class. Then, the subject is classified according to the majority voting scheme shown in Algorithm 1.

IV. The Method's Validation Criteria

In this section, we describe the validation procedure for our method. In particular, we illustrate the experimental protocol used for building the ground-truth dataset, the techniques for selecting the set of features to be inputted to the predictive model, and how the performance of frailty status assessment is evaluated.

A. Participants and Experiment Protocol

Adults aged 70+ years, consecutively accessing the Geriatrics outpatients clinic at Careggi academic hospital as patients or patient's caregivers, were assessed by geriatricians. After exclusion of participants who reported physical dependence in at least one of Katz's basic activities of daily living (BADL) [35], as well as of those with conditions causing overt abnormalities of gait (stroke, Parkinson's disease, severe hip or knee osteoarthritis), 34 eligible subjects were enrolled.

The sample included 13 females (80.15 ± 6.80 years old, height 1.58 ± 0.04 m, weight 65.19 ± 11.80 kg) and 21 males (80.05 ± 6.30 years old, height 1.72 ± 0.07 m, weight 75.43 ± 8.91 kg). All the participants were screened for the presence of frailty based on Fried's criteria [2], in order to enroll older subjects in each of the three categories of *robust*, *pre-frail* and *frail*. In particular, the following dimensions were studied on each subject:

1) an unintentional weight loss of 4.5 kg or more in the last year;

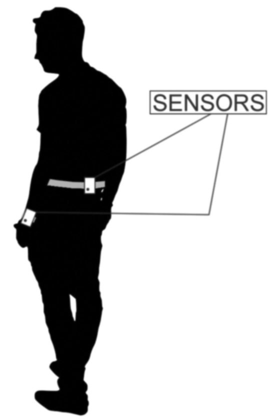


Fig. 2. On-body sensors setup.

2) low energy, identified through the CES-D (Center of Epidemiologic Studies Depression Scale) [36];
3) low physical activity, defined thanks to the Physical Activity Questionnaire for the Elderly (PASE) [37];
4) slowness, defined by the speed measured over a distance of 4.5 m and normalized for height and gender;
5) weakness, meaning reduced muscle strength in the dominant hand.

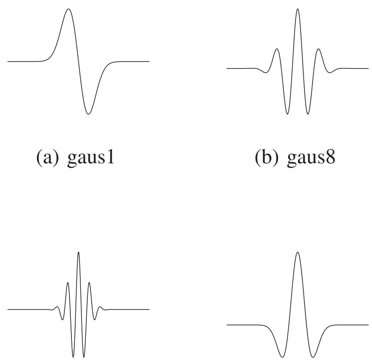
A subject was considered: *frail* if positive for three or more dimensions; *pre-frail* if positive for one or two dimensions; *robust* if negative for all dimensions. As a result, the sample contained 23 NR (non-rubust, including 8 frail and 15 pre-frail) subjects and 11 R (robust) subjects.

After clinical evaluation, participants were asked to walk 60 meters at their preferred pace along a 20 meters long path, under the supervision of geriatricians. In order to study the relevance of the sensor's position, during the test the subjects wore two Shimmer3 devices, one placed on the lumbar back, and the other worn like a watch on the wrist (Fig. 2). Shimmer3 is a wearable sensor embedding a tri-axial accelerometer [38], which was used to collect acceleration samples at 102.4 Hz. Sensors were positioned in the same way for all subjects, so that the directions of the sensor reference system with respect to the subject were consistent throughout experiments. Nevertheless, in the case of use in an uncontrolled environment, simple calibration procedures can be implemented for both considered on-body positions.

A signed, written consent to participate in the study was obtained from all participants. The study was conducted in accordance with the ethical principles of the Declaration of Helsinki. Identifiable information was removed from the collected data to ensure participant anonymity. Ethical approval for this study was obtained from *Comitato Etico Regionale per la Sperimentazione Clinica della Regione Toscana* (approval n. 14 834_oss of May 7, 2019).

B. Feature Extraction

The set includes common statistical parameters used in signal processing (mean, median, standard deviation, minimum and maximum values, interquartile range (IQR), mean absolute deviation (MAD), root mean square (RMS), kurtosis, skewness



(a) gaus1 (b) gaus8 (c) morl (d) mexh

Fig. 3. Wavelet families used in Continuous Wavelet Analysis. In (a) and (b), two different numbers of vanishing moments have been used for the Gaussian Wavelet.

and zero-crossing rate (ZCR)), calculated on acceleration components or Wavelet coefficients.

In addition, we considered some other features previously used in gait analysis and fall detection studies: the *cadence*, defined as the ratio between duration of the gait segment and the number of performed steps, and the average absolute acceleration variation (AAV), which is computed on consecutive acceleration samples [33], [39].

Features bringing information in the time-frequency domain were extracted by applying the Wavelet analysis. In particular, we used the *Continuous Wavelet Transform* (CWT) on the acceleration magnitude signal to study variations of power within gait segments.

For the sake of completeness, we recall here that the Wavelet Transform method is a signal processing technique for analyzing a time series containing non-stationary power at different frequencies [40]. In particular, by decomposing a time series into different frequency components, it is possible to analyze local variations of power for each component. Wavelet theory uses a set of Mother Wavelets (MW), which are scaled and translated to obtain a time-frequency representation of the time series.

In this study, we used the CWT W_n of a discrete sequence r_n with equal time spacing δt, which is defined as the convolution of r_n with a scaled and translated version of the MW ψ_0 [41]:

$$W_n(s) = \sum_{n_i=0}^{N-1} r_n \cdot \overline{\psi}\left[\frac{(n_i - n)\delta t}{s}\right],$$

where $\overline{\psi}$ indicates the complex conjugate of the wavelet function, s is the scale factor, $(n_i - n)$ denotes the translation along the gait segment and δt is the inverse of sampling rate.

Fig. 3 shows the MW families used in our analysis: Gaussian (*gaus*), Morlet (*morl*), and Mexican Hat (*mexh*). Let *gausN* indicate a gaussian wavelet with N vanishing moments: N is related to the approximation order and smoothness of the wavelet, so that a wavelet with N vanishing moments can approximate polynomials of degree $N - 1$. In this study, $N \in [1, 8]$.

Tables I and II list the full set of features considered in the feature extraction phase, for time and time-frequency domains, respectively.

TABLE I
List of Extracted Features in the Time Domain

Feature	Components
Mean	x, y, z, m
Median	x, y, z, m
Standard deviation	x, y, z, m
Minimum value	x, y, z, m
Maximum value	x, y, z, m
Interquartile range	x, y, z, m
Kurtosis	x, y, z, m
Zero Crossing Rate	x, y, z, m
Mean Absolute Deviation	x, y, z, m
Root Mean Square	x, y, z, m
Average Absolute Variation	x, y, z, m
AC_C1	m
AC_C2	m
AC_DP1	m
AC_DP2	m
Cadence	-
Duration	-

TABLE II
List of Extracted Features in the Time-Frequency Domain

Feature	Components
Mean	CWT coefficients
Median	CWT coefficients
Standard deviation	CWT coefficients
Minimum value	CWT coefficients
Maximum value	CWT coefficients
Interquartile range	CWT coefficients
Kurtosis	CWT coefficients
Skewness	CWT coefficients

C. Feature Selection and Frailty Status Assessment

Frailty status is assessed in a two-stage process by means of a machine learning model. In the first stage, gait instances are classified as belonging to a NR or R class. In the second stage, the subject is classified according to a majority voting scheme such as the one described in Algorithm 1. In order to maximise the performance of the classifier, we evaluated five different machine learning models: Random Forest, Gaussian naive Bayes, Logistic Regression, Multilayer Perceptron, and Support Vector Machine. All machine learning models have been implemented by means of the Python module Scikit-learn [42]. Details are shown in Table III. The models' parameters have been chosen empirically. If the parameters are not given in the table, the default values have been used.

Models were tested by means of a Leave-One-Subject-Out cross-validation procedure: at each iteration, the gait instances of one subject were used as testing set, while the gait instances of other subjects were used as training set to build a classification model. A *feature selection* step was performed within the cross-validation procedure. At each Leave-One-Subject-Out iteration, features were selected using only the training set, so that the

TABLE III
PYTHON SCIKIT-LEARN IMPLEMENTATION OF TESTED CLASSIFIERS

Model	Parameters
RandomForestClassifier()	n_estimators=50
GaussianNB()	default parameters
LogisticRegression()	penalty='l1', solver='liblinear'
MLPClassifier()	solver='lbfgs', activation='relu', hidden_layer_sizes=(12)
SVC()	kernel='poly', degree=2

TABLE IV
LIST OF EXTRACTED FEATURES IN THE FREQUENCY DOMAIN

Feature	Components
Mean	FFT coefficients
Median	FFT coefficients
Standard deviation	FFT coefficients
Minimum value	FFT coefficients
Maximum value	FFT coefficients
Interquartile range	FFT coefficients
Kurtosis	FFT coefficients
Skewness	FFT coefficients
First dominant frequency	-

model was built without any test set information. This led to a different feature set at each iteration, according to the subjects belonging to the training set. The trained model was then used to classify the instances of the left-out subject as NR or R. Finally, subject classification as NR or R was based on the majority voting scheme. This procedure was repeated, each time leaving out a different subject as the testing set.

As far as *feature selection* is concerned, we used the One Way ANalysis Of VAriance (ANOVA) test between groups of features and the target. Specifically, for each feature, the values extracted from the R and NR subjects in the training set were compared, and the ANOVA F-statistic was computed. Features were then sorted in descending order by F-statistic, and the best k features were selected. In this case, $k = 25$.

As stated in the *Introduction*, one of the aims of our study is to show that walk related features in the time-frequency domain, computed by means of Wavelet analysis, can improve the accuracy of frailty status assessment. More precisely, we wanted to verify if the variation of power over time, captured by CWT, could bring valuable information for frailty status assessment purposes. To do this, we compared the use of CWT-based features against a system based on frequency domain features. Features in the frequency domain have been extracted by means of Fast Fourier Transform (FFT). Specifically, we computed some statistics to describe the distribution of FFT coefficients. Besides, we added the *1st_dominant_freq* feature, which represents the frequency with the highest corresponding value in FFT power spectrum. The list of FFT-based features is shown in Table IV.

In summary, we performed our evaluation using (i) only time domain features, (ii) a combination of time and time-frequency domains (i.e., CWT-based) features, and (iii) a combination of time and frequency domains (i.e., FFT-based) features. Therefore, the results of (i), (ii) and (iii) were compared to test our hypothesis.

From now on, let us consider NR subjects as positive and R subjects as negative classification results. *Accuracy*, *sensitivity* and *specificity* have been used to evaluate and compare performance of the classification models. Accuracy measures the ratio between the number of correctly classified NR and R subjects and the total number of subjects. Sensitivity, or true positive rate, measures the proportion of NR subjects correctly identified. Specificity, also called true negative rate, measures the proportion of R subjects correctly identified. In addition, *Receiver Operating Characteristic* (ROC) *Area Under the Curve* (AUC) has been calculated for every tested model.

Finally, let us recall that all the operations described so far were performed independently using only information extracted from the lower-back or the wrist sensor, hereafter called *LOWER BACK* and *WRIST* approaches, respectively.

V. RESULTS AND DISCUSSION

In this section we report on the experiments we carried out following the validation procedure described in Section IV.

A. Performance of Frailty Status Assessment

Data collected during the campaign (Section IV-A) was used for the Leave-One-Subject-Out cross-validation procedure of the machine-learning models (Section IV-C). Acceleration signals from both the wrist and the lower-back sensors were processed to detect gait segments (Section III-B), from which a subset of twenty five features from the one listed in Section IV-B were extracted, in order to create gait instances. The frailty status assessment scores obtained by the five chosen classifiers are summarized in Table V, for both the *WRIST* and the *LOWER BACK* approaches.

In the *WRIST* approach, for all three feature sets, Gaussian NB was able to classify subjects with good classification accuracy. The best scores were achieved using the combination of time domain and CWT-based features, with a ROC AUC of 0.87, 91.3% sensitivity (21 correctly classified NR participants out of 23) and 81.8% specificity (9 correctly classified R participants out of 11). Similarly, Random Forest achieved valuable results, with an AUC of 0.80, 95.6% sensitivity and 63.6% specificity.

As far as the *LOWER BACK* approach, Gaussian NB achieved the best performance using features in time and time-frequency domains, with an AUC of 0.75, 87.0% sensitivity (20 correctly classified NR subjects out of 23) and 63.6% specificity (7 correctly classified R subjects out of 11).

From the results, it turns out that the wrist-worn device matched or outperformed the lower back-worn device for almost all the considered machine learning models. Therefore, according to our experiments, we argue that the signal extracted from the wrist-worn device is most effective for discriminating between NR and R subjects. Such a result can be explained by considering the arm swing involved in human walking, to which a wrist-worn device is more exposed. In fact, due to the subject's

TABLE V
AVERAGE RESULTS OF FRAILTY STATUS ASSESSMENT, BOTH FOR WRIST AND LOWER BACK

		WRIST				LOWER BACK			
Features	**Model**	**Acc.**	**Sens.**	**Spec.**	**AUC**	**Acc.**	**Sens.**	**Spec.**	**AUC**
TIME DOMAIN	**Gaussian NB**	0.82	0.91	0.64	0.77	0.76	0.87	0.55	0.71
	Random Forest	0.74	0.78	0.64	0.71	0.68	0.83	0.36	0.59
	Log. Regression	0.76	0.87	0.55	0.71	0.59	0.70	0.36	0.53
	ML Perceptron	0.71	0.78	0.55	0.66	0.65	0.70	0.55	0.62
	SVM	0.74	0.78	0.64	0.71	0.56	0.61	0.45	0.53
TIME DOMAIN + CWT-BASED	**Gaussian NB**	0.88	0.91	0.82	0.87	0.79	0.87	0.64	0.75
	Random Forest	0.85	0.96	0.64	0.80	0.76	0.87	0.55	0.71
	Log. Regression	0.79	0.91	0.55	0.73	0.74	087	0.45	0.66
	ML Perceptron	0.76	0.87	0.55	0.71	0.71	0.83	0.45	0.64
	SVM	0.68	0.74	0.55	0.64	0.74	0.83	0.55	0.69
TIME DOMAIN + FFT-BASED	**Gaussian NB**	0.82	0.91	0.64	0.77	0.76	0.87	0.55	0.71
	Random Forest	0.76	0.91	0.45	0.68	0.71	0.83	0.45	0.64
	Log. Regression	0.71	0.83	0.45	0.64	0.74	0.83	0.55	0.69
	ML Perceptron	0.53	0.48	0.64	0.56	0.50	0.35	0.82	0.58
	SVM	0.74	0.78	0.64	0.71	0.68	0.70	0.64	0.67

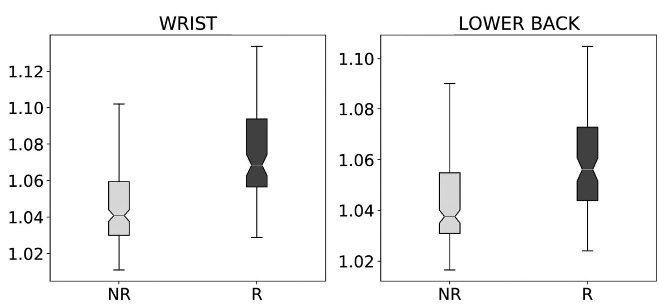


Fig. 4. Box plot representation of the RMS computed on the acceleration magnitude of R vs NR signals, in both WRIST and LOWER BACK.

TABLE VI
RESULTS OBTAINED IN ANOVA STATISTICAL TEST, FOR RMS IN *WRIST* AND *LOWER BACK*

	R	**NR**	$p - value$
WRIST	1.085 ± 0.291	1.047 ± 0.114	< 0.001
LOWER BACK	1.057 ± 0.037	1.043 ± 0.063	< 0.001

better stability, the energy produced by wider arm oscillations and wrist rotations of a R subject is likely to be greater than that produced by a NR subject.

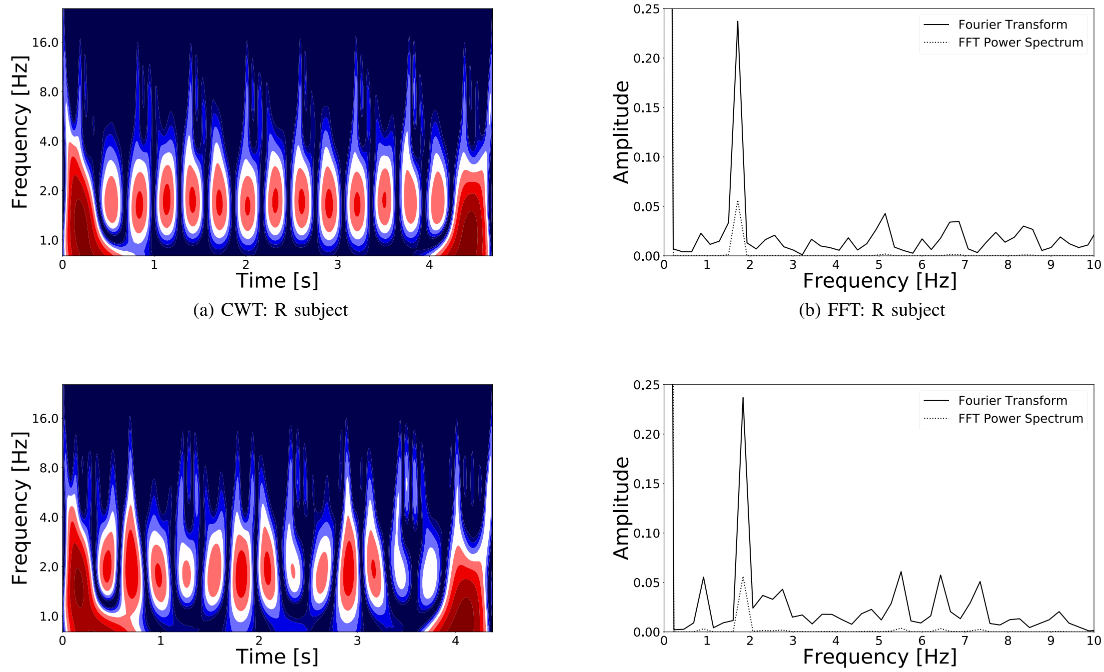
This information pattern can be observed in Fig. 4, which depicts the boxplots generated using RMS values. The RMS, computed as the area under the acceleration amplitude signal, represents the energy of the signal. RMS values of gait segments belonging to R subjects are generally higher than values computed for NR subjects. Even though such a difference is observable in both the *WRIST* and the *LOWER BACK* approaches, in the former the distinction is more evident. To statistically confirm the significance of RMS, we performed a one-way ANOVA test. More precisely, RMS was used as the dependent variable and the frailty category (R or NR) as the independent variable. The results are shown in Table VI. For both *WRIST* and *LOWER BACK*, RMS appears statistically significant in distinguishing R and NR subjects, with a $p - value < 0.001$ (the $p < 0.005$ criterion was used to test for statistical significance).

According to these findings, it appears that arm swing brings valuable information in frailty status assessment. As these gait-related features can be better captured by a wrist-worn device, we can conclude that our method implementation is suitable to be embedded in a common smartwatch, thus enabling continuous assessment of frailty without requiring the adoption of additional devices.

B. Wavelet Analysis

As mentioned in Section IV-C, models were trained and tested using (i) only time domain features, (ii) time domain + CWT-based features, and (iii) time domain + FFT-based features. This was done to investigate whether the use of acceleration features in the time-frequency domain may improve the performance of predictive models. As hypothesized, it appears that the CWT-based approach improves the classification task sensibly, independently of the position chosen for the sensor. Gaussian NB proved to be the best model in frailty status assessment, for all three considered feature sets. In the case of training performed by means of time domain + CWT-based wrist-derived features, however, an important step forward was made, with a 0.1 increase in the AUC score.

It is worth mentioning that all the features described in the previous sections are always extracted from gait segments consisting of four gait cycles. This leads to gait segments composed of a variable number of samples, since the length of a gait segment depends on the subject's cadence. Performing frequency-based features extraction on gait segments of variable length introduces differences into the outputs of the frequency analysis, on which the frequency-based features are computed. Nevertheless, in the current study, these differences concern the frequency band $[0, 0.26] Hz$, which from an in-depth analysis appears to contain nonsignificant information for frailty status assessment. We have chosen not to report here the results of the



(a) CWT: R subject
(b) FFT: R subject
(c) CWT: NR subject
(d) FFT: NR subject

Fig. 5. Power spectrum of the CWT (a: R subject, c: NR subject) and the FFT (b: R subject, d: NR subject).

frequency bands significance analysis, as it is beyond the scope of this study.

Now, let us discuss why Continuous Wavelet Transform enhanced the characterization of gait segments in view of the machine-learning classification of frailty status. In particular, we argue that CWT applied to gait segments can better capture the nature of oscillations produced by different cyclic phases involved in a subject's walk, such as arm motions, torso rotation and heel strike.

Let us first recall that, while the FFT only brings information about the frequency components of a signal, the CWT enables the study of signal power distribution in the time-frequency domain. Variations of power over time can be visualized by means of scalograms (such as the ones shown in the left side of Fig. 5), where colors represent the intensity of power - from blue (low power) to red (high power).

In Fig. 5 we report the outputs of CWT and FFT analysis applied to two sample gait segments, in the case of a R and a NR subject, respectively. Signals were recorded by the wrist sensor (similar considerations also apply to the lower back-worn sensor). Again, the gait segments used in this comparison are composed of 4 gait cycles. Here, red areas of scalograms correspond to higher levels of power released during stronger oscillations in gait, within a given frequency range (y-axis values), over a particular time interval (x-axis values). Notably, areas recurring with similar shape can be attributed to a regular gait. In other words, they characterise a walk-related pattern that produces regular oscillations over time. Large red areas at the boundaries of scalograms are the result of truncation of the gait segments and do not bring any useful information.

As can be seen in Fig. 5(a), most of the signal power released during the gait of a R subject is within the $1.5 - 3\,Hz$ frequency range, in accordance with the corresponding peak of the Fourier transform shown in Fig. 5(b). Notably, the scalogram evidences a certain regularity in the of power released during the gait. The same cannot be said of the gait signal produced by a NR subject, whose scalogram is shown in Fig. 5(c). Here, even though red areas still depict a gait activity with a similar level of associated power, it is clear that power is not released with the same regularity. It should be noted that such a difference is not evident when comparing the FFT pictures (e.g., Figs. 5(b) vs. 5(d)).

These results suggest that a higher level of power distributed regularly along time is associated with better stability during gait. In contrast, an irregular power distribution over time reflects a high gait variability, whose correlation with frailty has already been explored in previous studies. Montero-Odasso *et al.* demonstrated that a high gait variability is a marker of the loss of complexity in the dynamics of the gait pattern, and it is associated with frailty status [43]. In our experiments, we found that a higher level of power correlates with a more emphasized arm swing, which is also known to be positively related to "global gait stability" [44]. These findings are also in line with those of Mirelman et al. who reported that aging is associated with decreased arm swing amplitude [45].

VI. Conclusion

In this study we have investigated the use of wearable sensors to analyze gait and identify pre-frail and frail subjects. To this purpose, gait segments from 34 older adults were detected by using two wearable devices, placed at the subject's wrist and lower back. These traces were then independently analyzed, in time and time-frequency domains. Five different machine learning models were tested for the classification task: Random

Forest, Gaussian naive Bayes, Logistic Regression, Multilayer Perceptron and Support Vector Machine. The best performance was achieved by Gaussian naive Bayes (91.3% sensitivity and 81.8% specificity) using a combination of time and time-frequency domains features extracted from a wrist sensor-derived signal. Interestingly, the wrist-worn sensor achieved substantially better classification accuracy than the lower-back-worn sensor, indicating arm motion as a distinctive element between robust and frail subjects. Continuous Wavelet Analysis was used to include information about the distribution of power in the time-frequency domain, that allowed us to capture gait regularity and achieve better classification performance.

Our results demonstrate that unobtrusive wearable devices may enable an effective approach in continuous monitoring of human walking, and represent a significant step towards the feasibility of automated frailty status assessment based on machine-learning. Also, from the application point of view, a wrist-worn based implementation of the proposed method may foster user adoption of wearable devices for early detection of the frailty syndrome, as it may be embedded into a smartwatch.

A limitation of this study is represented by the number of subjects involved in the experiments. Future work will need to extend the dataset in order to test the generalization capabilities of the trained machine learning models. For this reason, in future work we aim to validate these results on a larger sample, which will better represent the population of community-dwelling older adults. Moreover, we plan to evaluate our method in a longitudinal study, in order to assess the feasibility of the proposed approach in uncontrolled environments.

References

[1] N. Fairhall *et al.*, "Economic evaluation of a multifactorial, interdisciplinary intervention versus usual care to reduce frailty in frail older people," *J. Amer. Med. Directors Assoc.*, vol. 16, no. 1, pp. 41–48, 2015.

[2] L. P. Fried *et al.*, "Frailty in older adults: Evidence for a phenotype," *J. Gerontol. Ser. A: Biol. Sci. Med. Sci.*, vol. 56, no. 3, pp. M146–M157, 2001.

[3] L. Dasenbrock, A. Heinks, M. Schwenk, and J. Bauer, "Technology-based measurements for screening, monitoring and preventing frailty," *Zeitschrift Für Gerontologie und Geriatrie*, vol. 49, no. 7, pp. 581–595, 2016.

[4] M. Schwenk *et al.*, "Wearable sensor-based in-home assessment of gait, balance, and physical activity for discrimination of frailty status: Baseline results of the Arizona frailty cohort study," *Gerontology*, vol. 61, no. 3, pp. 258–267, 2015.

[5] M. J. Mohler, M. J. Fain, A. M. Wertheimer, B. Najafi, and J. Nikolich-Žugich, "The frailty syndrome: Clinical measurements and basic underpinnings in humans and animals," *Exp. Gerontol.*, vol. 54, pp. 6–13, 2014.

[6] S. Abbate, M. Avvenuti, and J. Light, "MIMS: A minimally invasive monitoring sensor platform," *IEEE Sensors J.*, vol. 12, no. 3, pp. 677–684, Mar. 2012.

[7] M. Schwenk *et al.*, "Frailty and technology: A systematic review of gait analysis in those with frailty," *Gerontology*, vol. 60, no. 1, pp. 79–89, 2014.

[8] R. Thiede, N. Toosizadeh, J. L. Mills, M. Zaky, J. Mohler, and B. Najafi, "Gait and balance assessments as early indicators of frailty in patients with known peripheral artery disease," *Clin. Biomech.*, vol. 32, pp. 1–7, 2016.

[9] N. M. Kosse, N. Vuillerme, T. Hortobágyi, and C. J. Lamoth, "Multiple gait parameters derived from iPod accelerometry predict age-related gait changes," *Gait Posture*, vol. 46, pp. 112–117, 2016.

[10] D. Pradeep Kumar, N. Toosizadeh, J. Mohler, H. Ehsani, C. Mannier, and K. Laksari, "Sensor-based characterization of daily walking: A new paradigm in pre-frailty/frailty assessment," *BMC Geriatrics*, vol. 20, pp. 1–11, 2020.

[11] G. Manfredi, L. Midão, C. Paúl, C. Cena, M. Duarte, and E. Costa, "Prevalence of frailty status among the European elderly population: Findings from the survey of health, aging and retirement in Europe," *Geriatrics Gerontol. Int.*, vol. 19, no. 8, pp. 723–729, 2019.

[12] R. M. Collard, H. Boter, R. A. Schoevers, and R. C. Oude Voshaar, "Prevalence of frailty in community-dwelling older persons: A systematic review," *J. Amer. Geriatrics Soc.*, vol. 60, no. 8, pp. 1487–1492, 2012.

[13] G. A. van Kan, Y. M. Rolland, J. E. Morley, and B. Vellas, "Frailty: Toward a clinical definition," *J. Amer. Med. Directors Assoc.*, vol. 9, no. 2, pp. 71–72, 2008.

[14] L. P. Fried, S. J. Herdman, K. E. Kuhn, G. Rubin, and K. Turano, "Preclinical disability: Hypotheses about the bottom of the iceberg," *J. Aging Health*, vol. 3, no. 2, pp. 285–300, 1991.

[15] J. M. Guralnik, L. Ferrucci, E. M. Simonsick, M. E. Salive, and R. B. Wallace, "Lower-extremity function in persons over the age of 70 years as a predictor of subsequent disability," *New England J. Med.*, vol. 332, no. 9, pp. 556–562, 1995.

[16] J. M. Guralnik *et al.*, "A short physical performance battery assessing lower extremity function: Association with self-reported disability and prediction of mortality and nursing home admission," *J. Gerontol.*, vol. 49, no. 2, pp. M85–M94, 1994.

[17] J. Pritchard *et al.*, "Measuring frailty in clinical practice: A comparison of physical frailty assessment methods in a geriatric out-patient clinic," *BMC Geriatrics*, vol. 17, no. 1, pp. 1–8, 2017.

[18] P. Abizanda, L. Romero, P. Sanchez-Jurado, P. Atienzar-Nunez, J. Esquinas-Requena, and I. García-Nogueras, "Association between functional assessment instruments and frailty in older adults: The Fradea study." *J. Frailty Aging*, vol. 1, no. 4, pp. 162–168, 2012.

[19] H. Bergman *et al.*, "The Canadian initiative on frailty and aging," *Aging Clin. Exp. Res.*, vol. 15, no. 3 Suppl, pp. 1–2, 2003.

[20] M. Visser, D. J. Deeg, and P. Lips, "Low vitamin D and high parathyroid hormone levels as determinants of loss of muscle strength and muscle mass (sarcopenia): The longitudinal aging study amsterdam," *J. Clin. Endocrinol. Metab.*, vol. 88, no. 12, pp. 5766–5772, 2003.

[21] N. Millor *et al.*, "Automatic evaluation of the 30-s chair stand test using inertial/magnetic-based technology in an older prefrail population," *IEEE J. Biomed. Health Informat.*, vol. 17, no. 4, pp. 820–827, Jul. 2013.

[22] W. Zhang, G. R. H. Regterschot, H. Geraedts, H. Baldus, and W. Zijlstra, "Chair rise peak power in daily life measured with a pendant sensor associates with mobility, limitation in activities, and frailty in old people," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 1, pp. 211–217, Jan. 2017.

[23] M. Avvenuti, N. Carbonaro, M. G. Cimino, G. Cola, A. Tognetti, and G. Vaglini, "Smart shoe-assisted evaluation of using a single trunk/pocket-worn accelerometer to detect gait phases," *Sensors*, vol. 18, no. 11, 2018, Art. no. 3811.

[24] M. Zivanovic, N. Millor, and M. Gómez, "Modeling of noisy acceleration signals from quasi-periodic movements for drift-free position estimation," *IEEE J. Biomed. health Informat.*, vol. 23, no. 4, pp. 1558–1565, Jul. 2019.

[25] S. Díaz, J. B. Stephenson, and M. A. Labrador, "Use of wearable sensor technology in gait, balance, and range of motion analysis," *Appl. Sci.*, vol. 10, no. 1, p. 234, Dec. 2019.

[26] G. Cola, M. Avvenuti, P. Piazza, and A. Vecchio, "Fall detection using a head-worn barometer," in *Proc. Int. Conf. Wireless Mobile Commun. Healthcare*. Cham, Switzerland: Springer, 2016, pp. 217–224.

[27] A. Galán-Mercant and A. I. Cuesta-Vargas, "Differences in trunk accelerometry between frail and nonfrail elderly persons in sit-to-stand and stand-to-sit transitions based on a mobile inertial sensor," *JMIR mHealth uHealth*, vol. 1, no. 2, 2013, Art. no. e2710.

[28] B. R. Greene, E. P. Doheny, A. O'Halloran, and R. Anne Kenny, "Frailty status can be accurately assessed using inertial sensors and the tug test," *Age Ageing*, vol. 43, no. 3, pp. 406–411, 2014.

[29] E. I. Zacharaki *et al.*, "FrailSafe: An ICT platform for unobtrusive sensing of multi-domain frailty for personalized interventions," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 6, pp. 1557–1568, Jun. 2020.

[30] M. Huisingh-Scheetz *et al.*, "Wrist accelerometry in the health, functional, and social assessment of older adults," *J. Amer. Geriatrics Soc.*, vol. 64, no. 4, pp. 889–891, 2016.

[31] M. Huisingh-Scheetz *et al.*, "The relationship between physical activity and frailty among us older adults based on hourly accelerometry data," *J. Gerontol.: Ser. A*, vol. 73, no. 5, pp. 622–629, 2018.

[32] E. K. Antonsson and R. W. Mann, "The frequency content of gait," *J. Biomech.*, vol. 18, no. 1, pp. 39–47, 1985.

[33] G. Cola, A. Vecchio, and M. Avvenuti, "Improving the performance of fall detection systems through walk recognition," *J. Ambient Intell. Humanized Comput.*, vol. 5, no. 6, pp. 843–855, 2014.

[34] R. Moe-Nilssen and J. L. Helbostad, “Estimation of gait cycle characteristics by trunk accelerometry,” *J. Biomech.*, vol. 37, no. 1, pp. 121–126, 2004.
[35] S. Katz, A. B. Ford, R. W. Moskowitz, B. A. Jackson, and M. W. Jaffe, “Studies of illness in the aged: The index of ADL: A standardized measure of biological and psychosocial function,” *JAMA*, vol. 185, no. 12, pp. 914–919, 1963.
[36] J. G. Orme, J. Reis, and E. J. Herz, “Factorial and discriminant validity of the center for epidemiological studies depression (CES-D) scale,” *J. Clin. Psychol.*, vol. 42, no. 1, pp. 28–33, 1986.
[37] A. J. Schuit, E. G. Schouten, K. R. Westerterp, and W. H. Saris, “Validity of the physical activity scale for the elderly (PASE): According to energy expenditure assessed by the doubly labeled water method,” *J. Clin. Epidemiol.*, vol. 50, no. 5, pp. 541–546, 1997.
[38] Shimmer. [Online]. Available: “http://www.shimmersensing.com,” 2021.
[39] G. Cola, M. Avvenuti, F. Musso, and A. Vecchio, “Personalized gait detection using a wrist-worn accelerometer,” in *Proc. IEEE 14th Int. Conf. Wearable Implantable Body Sensor Netw.*, 2017, pp. 173–177.
[40] I. Daubechies, “The wavelet transform, time-frequency localization and signal analysis,” *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 961–1005, Sep. 1990.
[41] C. Torrence and G. P. Compo, “A practical guide to wavelet analysis,” *Bull. Amer. Meteorological Soc.*, vol. 79, no. 1, pp. 61–78, Jan. 1998.
[42] F. Pedregosa *et al.*, “Scikit-learn: Machine learning in python,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
[43] M. Montero-Odasso *et al.*, “Gait variability is associated with frailty in community-dwelling older adults,” *J. Gerontol. Ser. A: Biomed. Sci. Med. Sci.*, vol. 66, no. 5, pp. 568–576, 2011.
[44] S. M. Bruijn, O. G. Meijer, P. J. Beek, and J. H. van Dieën, “The effects of arm swing on human gait stability,” *J. Exp. Biol.*, vol. 213, no. 23, pp. 3945–3952, 2010.
[45] A. Mirelman *et al.*, “Effects of aging on arm swing during gait: The role of gait speed and dual tasking,” *PLoS One*, vol. 10, no. 8, 2015, Art. no. e0136043.