This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3648857

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication ..., date of current version ...

Digital Object Identifier ...

Real-World Gait Speed Estimation: an AI-based Approach for Adaptive Wearable Devices Integration

Michele Zanoletti[1], (Student Member, IEEE), Carlo Vallati[2], Nicola Carbonaro[2], (Member, IEEE), Alberto Greco[2], (Senior Member, IEEE), Alessandro Tognetti[2], (Senior Member, IEEE), and Marco Laurino[1]

[1]Michele Zanoletti and Marco Laurino are with the National Research Council, Institute of Clinical Physiology, Pisa, Italy. (e-mails: michelezanoletti@cnr.it, marco.laurino@cnr.it)

[2]Michele Zanoletti, Carlo Vallati, Nicola Carbonaro, Alberto Greco, and Alessandro Tognetti are with the University of Pisa, Dept. Information Engineering, Pisa, Italy. (e-mails: carlo.vallati@unipi.it, nicola.carbonaro@unipi.it, alberto.greco@unipi.it, alessandro.tognetti@unipi.it)

Corresponding author: Michele Zanoletti (e-mail: michelezanoletti@cnr.it).

Funded by the European Union. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency (HADEA). Neither the European Union nor the granting authority can be held responsible for them.

ABSTRACT Real-world gait speed assessment, has recently gained recognition as an important health indicator particularly in patients with chronic conditions such as Chronic Obstructive Pulmonary Disease (COPD). This study proposes an AI-based method for estimating gait speed in the real-world context in COPD patients using different combinations of three wearable devices: a smartphone, a smartwatch, and a pair of sensorized shoes. An analysis pipeline is developed and trained using laboratory datasets, then validated on an independent dataset acquired under free-walking conditions, demonstrating strong performance in both walking detection and gait speed estimation with mean accuracies across different device combinations of 0.81, 0.93, and 0.90 for resting, walking, and stair climbing activities, respectively, as well as a mean RMSE of 0.118 m/s and an ICC of 0.80. The estimated speed, obtained by applying the developed pipeline to daily life data of COPD patients, showed a strong and significant correlation with the clinical standard of functional capacity, the six-minutes walking distance (6MWD) across 34 patients (Spearman coefficient = 0.854, p-value = 1.38e-10). These findings highlight the feasibility of unobtrusive, continuous gait speed monitoring in real-life conditions, offering a promising integrative tool to traditional clinical assessments. The results suggest that wearable-based mobility monitoring could enhance remote patient management and personalized care strategies for COPD and other chronic conditions.

INDEX TERMS Daily Life Monitoring, Gait Speed Estimation, Mobility Analysis, Smart Sensors, Smartwatch, Smartphone, Smartshoes, Telemedicine, Wearable Devices, Six-Minute Walking Test, COPD.

I. INTRODUCTION

MOVING beyond traditional clinical settings, the assessment of real-world mobility, with gait speed as a primary indicator, is gaining prominence as a critical measure of health, particularly for older adults and individuals with chronic conditions. Gait speed represents a powerful predictor of critical health outcomes, from disability to cognitive decline, and ultimately mortality [1] [2]. It serves as a practical and informative tool for assessing health status, predicting prognosis, and monitoring changes over time. It has been referred to as the "sixth vital sign" [3], placing it on par with fundamental health parameters such as blood pressure, heart rate, and body temperature. Despite its apparent simplicity and ease of measurement, gait speed estimation has significant methodological complexities that have persistently challenged researchers, creating heterogeneous approaches that are difficult to compare across different clinical and real-world contexts. Variability in testing protocols - such as differences in starting positions, walking speeds, and distances - can affect the reliability of results. The available literature suggests that real-world activity monitoring might provide more objective, ecological, and reliable observations of functional ability, in comparison with laboratory, offering significant potential to improve clinical trials [4] [5]. In particular, some studies [6] have shown that gait patterns differ significantly between laboratory and real-world settings, with older people walking more slowly and exhibiting longer cycle times in real-world settings. These discrepancies represent fundamental obstacles to valid clinical interpretation, potentially leading to misguided treatment decisions and incomplete understanding of mobility limitations. These findings underscore the importance of ecological gait analysis in accurately assessing walking disorders, particularly in older adults. In recent years, the Mobilise-D consortium has made significant strides in leveraging advanced inertial sensor technology for real-life mobility assessment, particularly through the identification of digital mobility outcomes (DMOs) for various health conditions, including Parkinson's disease, multiple sclerosis, femoral fracture recovery, and COPD [7]. COPD, in particular, has drawn significant attention, with recent advances in artificial intelligence and digital health technologies showing promise for disease monitoring and management [8] [9] [10], and with growing evidence supporting the role of mobility metrics in disease management and prognosis [11] [12] [13]. The relationship between walking patterns and clinical outcomes in COPD patients reveals profound insights into disease progression and treatment efficacy. Moreover, a higher real-world walking cadence correlates with better outcomes in COPD, including a reduced risk of severe exacerbations [14]. This finding underscores the gait analysis potential as a valuable prognostic marker. An open challenge lies in ensuring the unobtrusive collection of real-life patient data and validating the quality of extracted mobility parameters by comparing them against clinically collected data. Despite the growing interest in smartphone apps for health monitoring and activity tracking, several studies have identified significant limitations, particularly concerning reliability and accuracy in real-world conditions [15] [16] [17]. Nevertheless, promising advancements have been made in wearable technologies. For example, accelerometer-based gait monitoring has been shown to enhance fall risk assessment for individuals with osteoarthritis and Parkinson's disease [18]. Additionally, high-precision deep learning models for IMU-based gait analysis have demonstrated robust performance, even in populations with mobility impairments [19] [20]. Moreover, recent deep learning approaches employing Transformers architectures have also been succesfully applied to others gait-related tasks [21], [22]. These and other studies on wrist sensors [23] [24] and lower-back sensors [25] highlight the transformative potential of wearable devices in gait-related healthcare, paving the way for innovative solutions to improve patient outcomes. We believe that the key to reliable gait speed estimation lies in using everyday objects that people already use, turning them into passive medical data generators. This approach embraces an ecological perspective, leveraging naturally worn or carried devices without requiring users to change their habits, perform manual calibrations, or initiate measurements. The system should seamlessly adapt to patient preferences, which may evolve over time or across different contexts. To achieve this, it must dynamically recognize both the user's activity and the available sensing devices, ensuring continuous and unobtrusive gait speed estimation in real-world conditions. Building on this foundation, our previous work [26] demonstrated that a combination of wearable devices — including smartphones, smartwatches, and sensorized shoes — could reliably estimate gait speed in laboratory conditions. In this work, we present a comprehensive approach for ecological monitoring of gait speed that addresses several critical barriers that have previously limited real-world application through three key innovations: (i) an adaptive sensor fusion architecture that dynamically accommodates any combination of wearable devices according to user preferences, operating effectively even when not all sensors are simultaneously available, thereby ensuring flexibility and adaptability under real-world conditions; (ii) a deep learning model that operates passively without requiring user interaction or subject-specific calibration, demonstrating direct applicability to new datasets; and (iii) validation in the daily lives of COPD patients under uncontrolled environmental conditions, demonstrating a strong correlation with the functional capacity clinical standard in naturalistic settings. This approach could help advance ecological monitoring by enabling the system to seamlessly identify walking periods and estimate gait speed while maintaining clinical accuracy in real-world, free-living conditions. To develop and validate our approach, we collected datasets in both laboratory and free-walking conditions, thereby bridging the gap between controlled assessments and ecological validity. Subsequently, we applied the complete pipeline to data from COPD patients recruited within the clinical study of the TOLIFE project [27] [28] [29] over a two-week period during their daily activities, generating new insights into real-world mobility patterns and their clinical implications. The impact of our work extends beyond technical achievement to potential clinical transformation: enabling continuous, non-invasive monitoring of real-world mobility in both healthy subjects and patients with COPD, facilitating earlier therapeutic intervention, providing objective measures for evaluating therapeutic efficacy, and empowering patients through greater insight into their condition. This work represents a significant progress in our ability to unobtrusively monitor and meaningfully interpret human movement in the context of daily life, potentially improving the management of chronic respiratory conditions and other mobility-affected disorders.

II. METHODS

A. WEARABLE DEVICES

In this study, we utilized the wearable device set from the TOLIFE sensor platform [27] [29], specifically designed for home monitoring of COPD patients. Figure 1 illustrates both the device setup and the reference system. The selected devices enable the collection of raw data related to mobility parameters, essential for AI-based algorithms aimed at detecting COPD exacerbations and assessing health-related quality of life. The wearable set includes a smartphone (Samsung Galaxy A14), a smartwatch (Samsung Galaxy Watch 5), and smart-shoes. The smartphone and the smartwatch collect inertial sensor data: accelerometers and orientation for the smartphone, and accelerometers and gyroscopes for the smartwatch. The smart-shoes, a research prototype tailored for the TOLIFE project, integrate a 3-axis accelerometer, gyroscope (LSM6DSL, STMicroelectronics), Bluetooth low-energy module, and three pressure sensors (FSR 402, Interlink) under the insole to monitor foot-ground interaction [30]. Data collection is managed through two custom Android applications, with the smartphone app also retrieving data from the smart-shoes. Both applications timestamp the data using device clocks synchronized to a common server, with periodic resynchronization. All signals are acquired at a 50 Hz sampling frequency, with the accelerometer measuring acceleration in m/s^2, the gyroscope capturing angular velocity in rad/s, the orientation sensor providing angular measurements in degrees, and the pressure sensor recording values in millivolts (mV). To ensure reliability and consistency for subsequent analysis, the raw signals from the devices were subjected to a series of preprocessing steps. First, a 4th-order Butterworth low-pass filter (ft = 20 Hz) was applied to remove high-frequency noise. The filtered signals were then segmented into 5-second time windows. Within each window, the data were resampled by interpolation at a uniform sampling rate of 20 Hz, resulting in a total of 100 data points per window. Overlapping was applied to increase the number of windows, with 80% overlap used in the first two datasets and 90% overlap in the third one.

B. MODEL DEVELOPMENT AND DATA PROCESSING PIPELINE

Two types of machine learning models were developed to process and analyze motion data collected from the wearable devices: (i) features-based models for walking recognition and (ii) a deep learning model for gait speed estimation. The Walking Recognition Models (WRMs) were designed to perform device-specific walking detection with computational efficiency. Each WRM utilizes a single fully connected layer applied to a vector of extracted temporal and frequency domain features. Building upon the temporal segmentation provided by the WRMs, the Gait Speed Estimation Model (GSM) integrates data from up to four devices through a deep learning architecture. As depicted in Figure 2, the GSM employs a complex architecture that combines convolutional neural network (CNN) layers for feature extraction, a self-attention module for adaptive feature weighting, and fully connected layers for the final regression output. The convolutional layers extract temporal patterns from the synchronized sensor streams, while the self-attention mechanism identifies the most relevant features across the different sensor locations during various phases of the gait cycle. A key methodological innovation in the system is the implementation of an adaptive sensor fusion approach. To accommodate real-world scenarios where subjects may not consistently wear all devices, a binary "wearing mask" is derived from the outputs of the WRMs. This mask indicates which devices are actively providing valid walking data at any given time point. The fusion process incorporates this mask to dynamically adjust the architecture's attention mechanisms and weighting schemes, ensuring robust performance across varying device configurations. This approach enables the system to maintain estimation accuracy even when not all devices are worn, addressing a significant challenge in longitudinal real-world monitoring where device adherence may fluctuate.

The analysis pipeline consists of the cascading application of these two types of models to 5-second signal time windows. To support the development and evaluation of the models, four different datasets were collected (Table 1), including both healthy individuals and COPD patients. Three of these datasets were collected from healthy subjects. The first two were acquired under laboratory conditions and utilized to train the walking recognition models and the gait speed estimation model. The third, collected under free-walking conditions, served to validate the performance of the learned algorithms. The fourth dataset was obtained from a clinical trial conducted in the Tolife project involving COPD patients. A total of 37 patients were enrolled and underwent a recruitment visit (RV) during which they performed two six-minute walking tests (6MWT). Then each patient was provided with the Tolife kit [27] [29], which included a smartphone, a smartwatch, and a pair of sensorized shoes [30] [31]. The analysis pipeline developed from the first two datasets was then applied to the third and fourth datasets. For each patient, the mean gait speed was estimated based on the data recorded during the first two weeks after RV. To evaluate the performance of the pipeline, a correlation analysis was performed between the estimated speed during these two weeks and the mean distance walked during the two 6MWTs performed at RV (see figure 3).

C. ACTIVITY RECOGNITION

1) Experimental Protocol

The first dataset is dedicated to train the activity recognition models and involved 15 healthy subjects (M: 7, F: 8, mean age: 27.3±3 years, 176.8±9.4 cm, 72.1±12.7 kg, BMI: 23.0±3.37 kg/m^2). Participants were instructed to perform specific activities. These activities are resting, walking on flat ground (level walking), and climbing stairs (ascending and descending). Resting consisted of three distinct postures: lying on a bed, sitting on a chair, and standing in a room. While standing, participants were permitted to move freely but were restricted from walking for more than 5 consecutive seconds. Level walking included three sessions: two involved walking in a straight line along a path of 180 meters, while in the third session, the path was 450 meters long with directional changes. Stair climbing consisted of continuous ascending and descending of a staircase for a couple of minutes. The data collected following this protocol were processed to extract activity windows. These windows served as input for training activity detection algorithms, with a separate model developed for each device. The objective was

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3648857

FIGURE 1: TOLIFE wearable set: smartphone (left), smartwatch (center), smart-shoe (right). Local reference frames associated with the inertial sensors are reported. The left shoe reference is defined symmetrically to the right shoe.

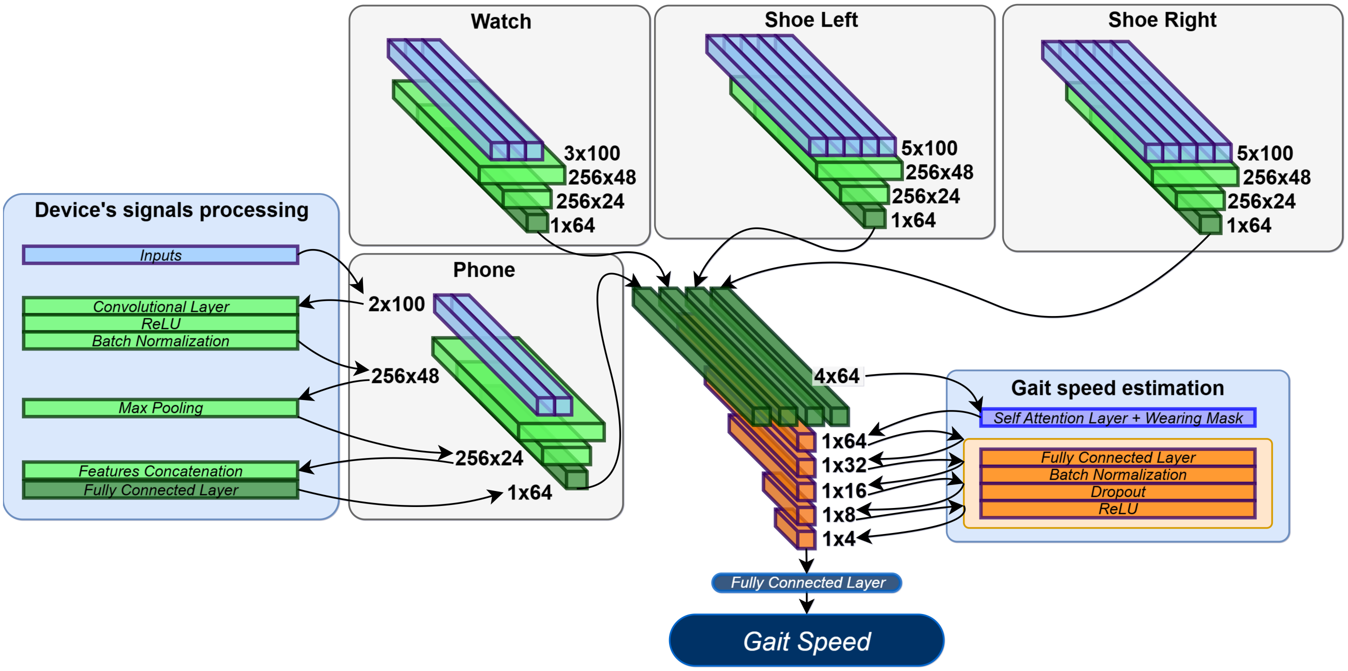


FIGURE 2: The Gait Speed Estimation Model (GSM) architecture consists of four modules, each corresponding to a specific wearable device. Each module processes raw sensor data to generate a feature vector, which is then analyzed by an attention module. This module integrates these feature vectors with device usage information, represented by the WRMs as a binary Wearing Mask, to produce the final gait speed estimate.

to independently determine whether each device was being worn, and to provide this information, along with the raw data, to the gait estimation model that will perform the gait speed estimation. The smartphone could be worn in various orientations. In this protocol, participants were instructed to wear the smartphone in a consistent manner in their front trouser pocket while varying randomly its orientation. This approach aimed to enable the algorithms to recognize a wide range of possible orientations. For the smartwatch, participants were allowed to wear it on their preferred wrist.

2) Walking Recognition Model Architectures and Training

WRMs are architectures consisting of a single fully connected layer tailored to process device-specific features set. They are designed to map the feature vectors to a 3-dimensional output space representing the three possible activities. The extracted features capture statistical, temporal, and frequency-domain characteristics of the input signals. In order to generalize the extracted features with respect to the wrist on which the watch is worn, we used the absolute value of the X axis. Acceleration along this axis (towards the arm, see Figure 1) is inverted with respect to zero when the wrist is changed, while the Y and Z axes remain unchanged. To identify the most relevant features for the classification task, we applied a feature selection technique based on L1 regularisation, also known as lasso. L1 regularisation promotes sparsity in the model by driving the coefficients of less important features to zero. Features were selected based on a coefficient magnitude threshold of 1, retaining only those with absolute values exceeding this limit. By enforcing sparsity, this approach mitigates overfitting, ensuring that the model is both interpretable and generalisable.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3648857

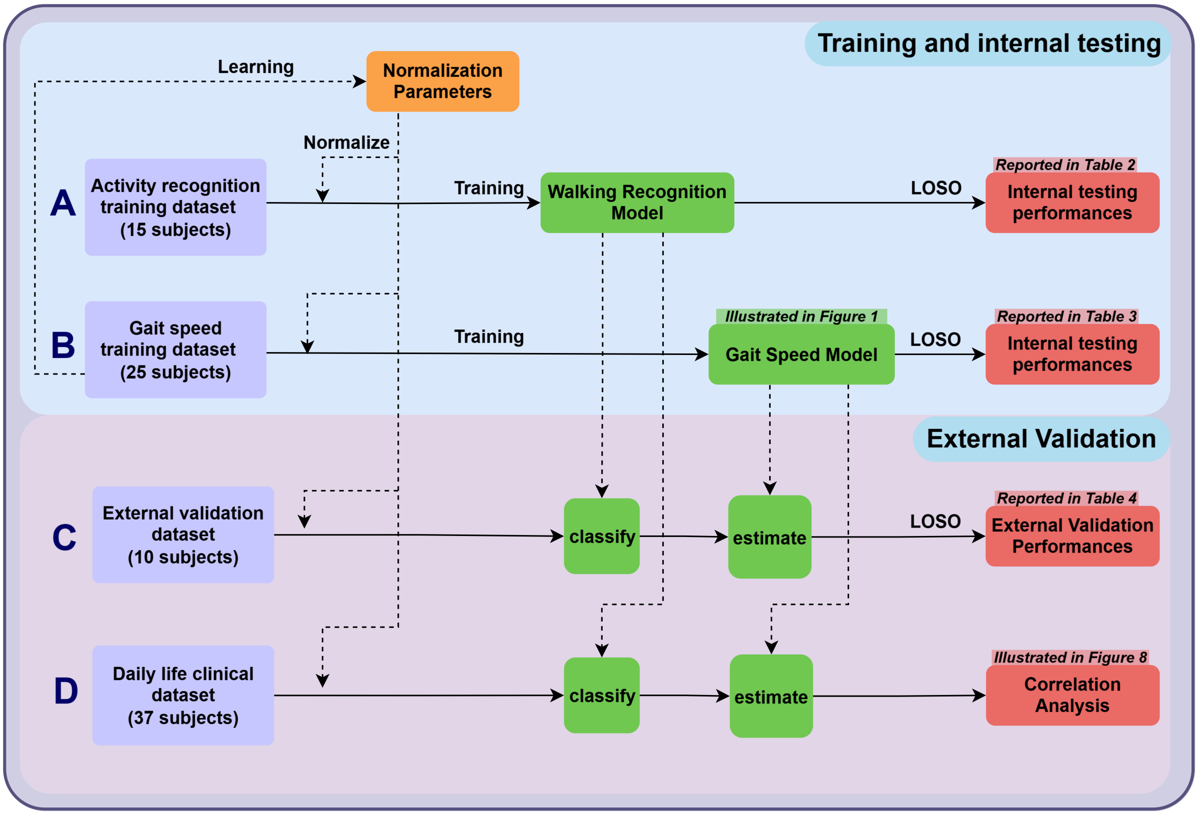


FIGURE 3: Four datasets were involved in the development and validation of the pipeline. Dataset A: 15 healthy subjects performed different activities; from these data the WRMs were trained and the performances are reported in Table 2. Dataset B: 25 healthy subjects performed the 6MWT; from these data the GSM (shown in Figure 2) was trained and the performances are reported in Table 3. Dataset C: 10 healthy subjects were recorded in free-walking conditions; the pipeline was validated on Dataset C and the results are reported in Table 4. Dataset D: 37 COPD patients' daily-life activities over two weeks were monitored; the correlation analysis between the estimated gait speed and the clinical 6MDW is reported in Figure 7.

TABLE 1: Summary of datasets used for activity recognition model training and validation.

Dataset	**N**	**Sex (M/F)**	**Age (years)**	**Height (cm)**	**Weight (kg)**	**BMI (kg/m^2)**	**Acquisition conditions**
Activity Recognition Dataset	15	7 / 8	27.3 ± 3.0	176.8 ± 9.4	72.1 ± 12.7	23.0 ± 3.37	Laboratory setup, $\sim$ 20 minutes per subject
Gait Speed Estimation Dataset	25	15 / 10	28.9 ± 4.8	175.3 ± 8.9	68.1 ± 10.8	22.1 ± 2.4	Laboratory setup, $\sim$ 18 minutes per subject
External Validation Dataset	10	5 / 5	43.8 ± 12.1	169.0 ± 9.6	65.7 ± 4.0	22.9 ± 4.23	Free walking, $\sim$ 10 minutes per subject
Clinical Dataset	37	23 / 14	63.3 ± 5.82	169.7 ± 8.73	74.2 ± 15.62	25.6 ± 4.63	Daily-life conditions, two weeks per subject

This model was trained for a total of 100 epochs with an 80% window overlap. The learning rate was set to 0.01, and the batch size was fixed at 256 samples per iteration. The optimization process was guided by the cross-entropy loss function, using the Adam optimizer. The performance evaluation is carried out taking into account metrics such as precision, recall, precision, and F1 score for every device. The training and the performance evaluation process is carried out with a leave-one-subject-out cross validation.

D. GAIT SPEED ESTIMATION

1) Experimental Protocol

The second dataset is composed of 25 healthy subjects (is an extension of the previous published dataset [32] [26] (M: 15, F: 10, age: 28.9±4.8 years, height: 175.3±8.9 cm, weight: 68.1±10.8 kg, BMI: 22.1±2.4 kg/m^2) was devoted to the training of the gait speed estimation model. This dataset was collected in a laboratory setting using the smartphone, the smartwatch and the sensorized shoes. The reference system is the Awinda inertial motion tracker coupled with the MVN Analyse software, both supplied by Xsens [33] (Enschede, Netherlands). To each participant was asked to perform a modified version of the 6MWT, a widely recognised and standardised assessment tool used to evaluate functional autonomy, particularly in people with impaired lung function [34]. The original clinical version of the test consists of measuring the distance a person walks in six minutes along a 30-meter flat path, walking as fast as possible [35]. Our modified version took place on a 10 meter flat path, with a turning radius of about 50 centimetres to change direction. Each participant was asked to perform the test three times at three different speeds: medium, slow and fast, to cover a wider range of speeds (from 0.3 to 1.85 m/s). During data collection, the phone and the watch were consistently worn in the same positions: the phone was placed in the left front trouser pocket, with its screen facing the thigh and the microphone pointing upwards, while the watch was worn on the left wrist.

2) Gait Speed Model Architecture and Training

The proposed GSM is specifically designed to estimate gait speed using data from up to four devices in combination with the output of the activity recognition models, the "Wearing Mask". The model uses a modular architecture that processes data from each device independently, allowing it to effectively handle different combinations of device availability (Figure 2). These combinations include the use of only the phone, only the watch, only the shoes, any pair of devices (phone and watch, phone and shoes, or watch and shoes), or all three devices together. The model combines convolutional layers, a self-attention module and fully connected layers to extract and integrate features across the available devices. To ensure robust performance, the model incorporates regularization techniques such as batch normalization which mitigate overfitting and improve generalization. The model takes as input raw data from the devices and employs device-specific feature extraction modules. The GSM processes 5-second window of sensor data from wearable devices using accelerometer, orientation, and pressure data. Each device module applies 1D convolutions (kernel size 10, output size 256, stride 2, padding 2) followed by ReLU activation, batch normalization, and max-pooling. After flattening, each device produces 6,400 features, which are projected to a 64-dimensional feature space via device-specific fully connected layers. Specifically, all modules utilize accelerometer signals, while the phone module also incorporates orientation data, and the shoe modules include pressure signals. From the phone, we extracted the magnitude signal from the three accelerometer axes and the orientation sensor. This approach was chosen because the phone is a wearable device that can be oriented in various ways, and as a result, the meaning of the axes may change depending on the device's orientation. By computing the magnitude, we obtain a value that is independent of the orientation facilitating generalization. The watch utilizes all three accelerometer axes separately, as do the shoes. Regarding pressure signals, the two signals originating from the forefoot are averaged. Summarizing, the phone has an input dimension of 2, the watch has 3, and the shoes have 5. The four device-specific feature vectors are processed by a multi-head self-attention module with 4 attention heads. The Wearing Mask derived from the WRMs is applied to zero out features from devices identified as not worn during walking and is also integrated as a key padding mask in the attention module, ensuring that such features are excluded during fusion. This combined representation passes through a fully connected network with progressive dimension reduction (64→32→16→8→4), where each layer includes batch normalization, dropout (rate=0.3), and ReLU activation. Finally, a linear layer predicts gait speed. The designed model contains 1.65 million trainable parameters and requires 7.27 million FLOPs per 5-second input sample. This model was trained for a total of 30 epochs with an 80% window overlap. The learning rate was set to 0.001, and the batch size was fixed at 2048 samples per iteration. The optimization process was guided by the mean squared error loss function, using the Adam optimizer. To address variations in orientation and the potential absence of devices, a data augmentation strategy was employed by simulating all possible rotations and missing data scenarios. Specifically, we adopted two key augmentation techniques: simulated device rotations, which account for natural variations in device orientation, and simulated missing devices, which enhance model robustness by randomly omitting devices. To comprehensively capture these variations, we define a mask for each possible device configuration: the phone can appear in one of four positions or be absent (5 states), the watch can be worn on either wrist or be absent (3 states), and each shoe can be present or absent (2 states per shoe). This results in 60 possible configurations, 59 of which include at least one device. By training the model on this diverse set of conditions, we ensure adaptability to real-world scenarios. The Root-Mean-Square-Error (RMSE) and the Intraclass Correlation Coefficient (ICC) [36] are used as performance metrics of the gait speed estimation. These metrics are evaluated for each possible combination of devices.

The Intraclass Correlation Coefficient (ICC) quantifies the reliability or consistency among raters. The ICC of type (2,1) is defined as:

$$\mathrm{ICC}(2,1) = \frac{MS_R - MS_E}{MS_R + (k-1)MS_E + \frac{k}{n}(MS_C - MS_E)} \tag{1}$$

The calculation of ICC(2,1) [36] is performed using the Python package `pingouin`, version 0.5.5 [37]. In this formulation, MS_R represents the mean square between raters (Mean Square for Raters), MS_E the mean square for error (Mean Square for Error), k the number of measurements for each object, n the number of objects measured, and MS_C the mean square for targets (Mean Square for Targets).

E. PIPELINE VALIDATION

1) Experimental Protocol

A free walking data collection session was then organized in an outdoor environment. Ten new healthy subjects were recruited, with a mean age higher than that of the subjects who participated in the acquisition of training datasets (M: 5, F: 5, 43.8±12.1 years, 169±9.6 cm, 65.7±4 kg, BMI: 22.9±4.23 kg/m^2). During the execution of this protocol, participants were asked to keep the phone in their front pocket in the position they preferred, and the watch on their favorite wrist. Each of these participants followed a protocol that involved the continuous execution of resting, level walking, and stair climbing activities. Specifically, for the level walking activity, each participant walked four times along a straight path of known length (100 or 280 meters). The path was predetermined, but participants were free to walk at their own pace, and time stamps were annotated at predetermined checkpoints.

2) Performance Evaluation

The evaluation metrics for both activity recognition and gait speed estimation remained consistent with those previously described. Classification accuracy was calculated using all activity windows. For gait speed estimation, both the root mean square error (RMSE) and the intraclass correlation coefficient (ICC) were assessed for each walked path. These metrics were evaluated in two ways: (i) by averaging the estimated gait speeds across all windows within a given path to assess the algorithm's overall performance, and (ii) by considering only the windows classified as walking, thereby focusing on true positives identified by the WRM algorithm and accounting for the entire processing pipeline. A total of 40 paths were analyzed, with each of the ten participants walking four distinct paths. The reference gait speed was derived from the known distance and time for each path, yielding 40 data points. Since only four data points were available per participant, a single ICC value was computed for each combination. To further examine the robustness of the reported performances against classification errors, we introduced increasing percentages of false positives into the validation dataset. These artificial misclassifications were randomly selected from the windows originally identified as false positives by the gait recognition algorithm. Finally, we analyzed the impact of this manipulation on RMSE and ICC by plotting their trends across different device combinations as the proportion of false positives increased. In the described validation we excluded windows showing more than 80% missing data, which occurred in less than 5% of collected windows, exclusively affecting the shoes due to occasional Bluetooth connection interruptions.

F. VALIDATION ON REAL-WORLD DATA

1) Experimental Protocol

The pipeline was subsequently applied to data collected from 37 COPD patients from two clinical centers (sex: 23 M, 14 F; age: 63.3 ± 5.82 years; height: 169.7 ± 8.73 cm; weight: 74.2 ± 15.62 kg, BMI: 25.6±4.63 kg/m^2) in daily life conditions. Each participant underwent a clinical recruitment visit (RV), during which they performed two 6MWT administered by an expert pulmonologist. The average six-minute walking distance from these two tests served as a reference value to assess the correlation with the estimated gait speed during daily life walks. At RV, each patient received a kit containing a smartphone, a smartwatch, and a pair of sensorized shoes. Participants were instructed to use these devices freely in their daily lives (both indoors and outdoors) under conditions they found comfortable. For each patient, the two weeks following the RV were analyzed, with data evaluated daily between 5:00 AM and 10:00 PM. These 17-hour periods were segmented into 5-second windows, for which an activity mask was generated and subsequently used in the gait estimation model. Daily analysis focused on identifying walking periods. Given the large number of windows analyzed (12,240 in one day), a significant number of false positives —windows incorrectly classified as walking— were likely. Furthermore, very short walks are less representative of aerobic capacity or endurance, whereas longer walks, closer in duration to the six-minute walking test, are more likely to yield meaningful insights. To ensure a reliable daily estimate of walking speed, the analysis prioritized walks exceeding a defined minimum duration. The process began with a threshold of 10 minutes. If not long enough walks were detected, the threshold was iteratively reduced—first to 9 minutes, then to 8 minutes, and so on—until walks as short as 1 minute were accepted. Single-window gaps ("not walking" preceded and followed by walking windows) were permitted to account for potential misclassifications and avoid discarding informative walks. For each identified walk, the average gait speed was calculated by averaging the gait speed of all windows within the walk. If multiple walks were identified in a single day, their gait speeds were averaged to determine the daily gait speed. Finally, for each subject, the daily gait speed estimates over the 14-day period were aggregated by calculating the median of all daily values. The whole approach is illustrated in Figure 4. We considered as valid only windows with at least 80% of expected samples.

The dataset is available on Zenodo (https://zenodo.org/records/1694 accessed on 21 September 2025).

III. RESULTS

A. ACTIVITY RECOGNITION

We selected the most relevant features using L1 regularization, highlighting those most frequently used by different devices. Shannon entropy (13), mean (12), and standard deviation (12) emerged as the most common, followed by range (8), number of peaks (4), root mean square (3), mean crossing rate (1), and dominant frequency (1). The phone and the right shoe utilized the highest number of features (16), followed by the left shoe (13) and the watch (8). These results illustrate how feature importance varies across devices, providing valuable insights for the analysis.

To assess the WRM models performances we performed leave-one-subject-out (LOSO) cross-validation on the 15 subjects of the activity recognition dataset (Dataset A). The obtained results are reported in Table 2. In the table's rows, we presented the four different devices, providing a detailed

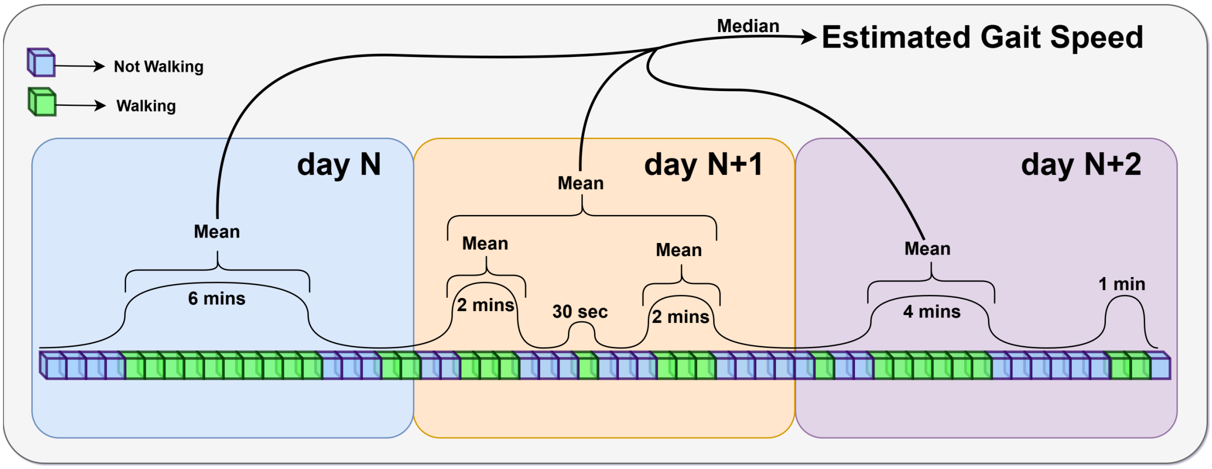


FIGURE 4: The analysis pipeline processed all daily-life recordings for each patient, estimating gait speed over every 5-second window with available data. These estimates were then aggregated by day, prioritizing the longest recorded walks, and the median of the daily estimates was used to obtain the overall gait speed value.

overview of the average accuracies and the corresponding standard deviations calculated for each individual activity.

B. GAIT SPEED ESTIMATION

The assessment of GSM performance is conducted using leave-one-subject-out (LOSO) cross-validation on the 25 subjects from the gait speed training dataset (Dataset B). This process is repeated for each device combination, yielding seven distinct performance evaluations. The first column of Table 3 presents the mean and standard deviation of the RMSE, while the second column reports the mean intraclass correlation coefficient along with its standard deviation.

C. PIPELINE VALIDATION

Table 4 presents the performance results for both activity recognition and gait speed estimation obtained on the validation dataset (Dataset C). For activity recognition, accuracy scores are reported for each activity and every device combination, providing an overview of classification performance. Additionally, precision, recall, and F1 scores, which offer a more comprehensive evaluation of the classifier's effectiveness, are detailed in the appendix (Table 4). For gait speed estimation, RMSE and ICC were computed for each walked path using two distinct approaches. The first approach involved calculating the mean gait speed across all windows within a given path, which provides an overall assessment of the estimation algorithm's performance. The second approach considered only the windows classified as walking, thus isolating true positives identified by the WRM algorithm. This method allows for a more targeted evaluation of the system's accuracy while capturing the effectiveness of the entire processing pipeline, from activity classification to gait speed estimation.

To provide a clearer visualization of the agreement between the estimates and the actual speed values, we have also included correlation and Bland-Altman (BA) plots and we added BA bias and limits to Table 4. We computed both of them for all possible device configurations. The plot for the configuration including all devices is shown in Figure 5. A Spearman correlation coefficient of 0.763 was obtained, with a bias of –0.032 and limits of agreement ranging from –0.252 to 0.189. The remaining plots are presented in Figure 10 in Appendix I, showing correlation coefficients between 0.633 and 0.825. The only configuration exhibiting a positive bias was the “Only shoes” configuration (0.028). The widest range of limits of agreement was observed for the “Only Watch” configuration [–0.349, 0.279], whereas the narrowest range was found for the “Watch and Shoes” combination [–0.201, 0.190]. Notably, the “Watch and Shoes” configuration also demonstrated the best performance in gait speed estimation (RMSE = 0.096).

1) Robustness Analysis

Figures 6a and 6b illustrate the results of the robustness analysis, which investigates the impact of increasing false positive rates on RMSE and ICC. Specifically, the false positive rate was artificially varied from 0% to 100% in 2% increments to systematically assess how classification errors influence gait speed estimation performance. Each curve in the figures corresponds to a different device combination, as specified in the legend. This visualization enables a comprehensive evaluation of how various device pairings respond to increasing false positives, providing insights into the resilience of the estimation pipeline under different error conditions. Comparison of RMSE and ICC versus false positive rate in walking detection across all device combinations. Beyond a 20% false positive rate, the RMSE of all combinations begins to increase, reaching values above 0.15 between 40% and 60% false positives, except for the "Only Watch” configuration, which starts at 0.15 and rises to approximately 0.2. As the false positive rate increases further, RMSE continues to rise, reaching up to 0.3. Similarly, the

TABLE 2: Performance Metrics on Internal Testing for Activity Recognition

	Resting				Level Walking				Stair Climbing			
	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
Smartphone	0.97±0.064	0.98±0.013	0.97±0.064	0.97±0.036	0.88±0.117	0.87±0.092	0.88±0.117	0.87±0.071	0.87±0.056	0.83±0.178	0.87±0.056	0.84±0.113
Smartwatch	0.95±0.062	0.93±0.055	0.95±0.062	0.94±0.043	0.82±0.101	0.82±0.209	0.82±0.101	0.80±0.149	0.84±0.121	0.80±0.138	0.84±0.121	0.80±0.082
Left Shoe	0.99±0.016	0.99±0.012	0.99±0.016	0.99±0.012	0.98±0.037	0.96±0.078	0.98±0.037	0.97±0.043	0.96±0.061	0.97±0.046	0.96±0.061	0.96±0.036
Right Shoe	0.99±0.012	0.99±0.017	0.99±0.012	0.99±0.012	0.92±0.087	0.91±0.194	0.92±0.087	0.90±0.150	0.93±0.112	0.91±0.109	0.93±0.112	0.91±0.087

The table reports values in the format [Mean±std].

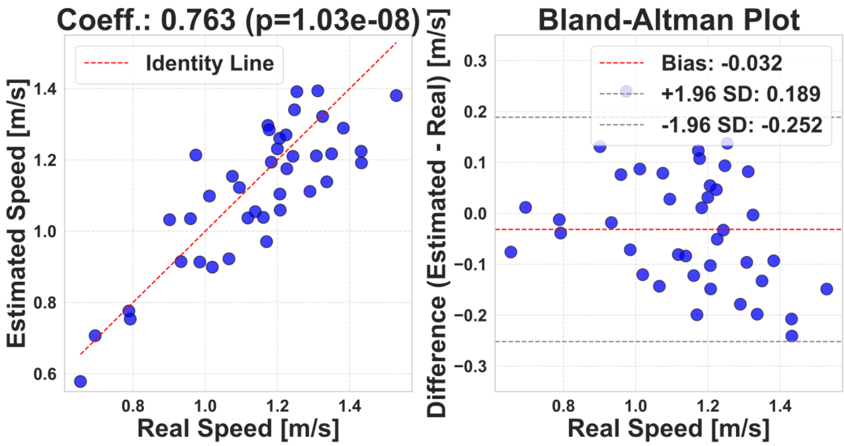


FIGURE 5: Correlation (with Spearman coefficient, left) and Bland-Altman plot (right) of the configuration with all devices in the Dataset D.

TABLE 3: Internal Testing Performances on Gait Speed Estimation [RMSE]

Condition	Gait Speed [m/s]	Intraclass Correlation Coefficient
Smartphone	0.139 ± 0.0642	0.776 ± 0.2090
Smartwatch	0.167 ± 0.0504	0.697 ± 0.2446
Shoes	0.093 ± 0.0368	0.891 ± 0.091
Smartphone and Smartwatch	0.120 ± 0.0462	0.800 ± 0.2169
Smartphone and Shoes	0.090 ± 0.0372	0.893 ± 0.9024
Smartwatch and Shoes	0.095 ± 0.0333	0.874 ± 0.1297
All Devices	0.096 ± 0.0376	0.863 ± 0.1578

The table reports values in the format [Mean±std].

ICC decreases as the false positive rate increases, dropping from good to moderate agreement between 20% and 40% false positives, and from moderate to poor for the "Only Watch" combination.

D. REAL-WORLD VALIDATION ON TOLIFE'S DATA

Of the 37 initial patients, following the procedure illustrated in Figure 4, it was possible to obtain an estimate of the average gait speed for 34 of them. Figure 7 shows the correlation between the daily-life estimated mean gait speed, over the two week monitoring period, and the six-minute walking distance at RV. The colored sections of the circles in the correlation plot represent the proportion of the different configurations in the subject's daily data (Dataset D). The circle size instead indicates the number of days (from 1 to 14) for which a gait speed estimate was available for the subject.

IV. DISCUSSION

This study addresses the critical need for unobtrusive assessment of gait speed, a key biomarker of functional capacity particularly in COPD, while acknowledging challenges from methodological heterogeneity and limited comparability between clinical and real world settings. Our primary contribution focuses on the development and validation of an integrated system capable of accurately estimating gait speed in real-world environments, overcoming several limitations inherent in the laboratory-based assessments. We demonstrate the feasibility of continuous gait speed monitoring by using a combination of wearable sensors (smartphones, smartwatches, and sensorized shoes), by providing a pathway for objective assessments of mobility in real-world conditions. Our approach is based on a machine learning framework, combining walking recognition with gait speed estimation models. We validated our pipeline by testing it on previously unseen individuals. With the external validation dataset (Dataset C), we focused on assessing the accuracy of our gait speed estimates, demonstrating the pipeline's ability to effectively process free-walking data. Our model provides valuable insights into mobility assessment, as demonstrated by its strong correlation with the 6MWD, a gold standard clinical measure of mobility. This validation with the daily-life clinical dataset (Dataset D) underscores the reliability and applicability of our model in real-world scenarios, particularly for COPD patients. Notably, no changes to clinical practice were required, as the devices seamlessly integrated into participants' daily activities. This unobtrusive and natu-

TABLE 4: Activity recognition accuracies and gait speed estimation performances [RMSE] with Bland–Altman analysis on the external validation dataset in all possible configurations

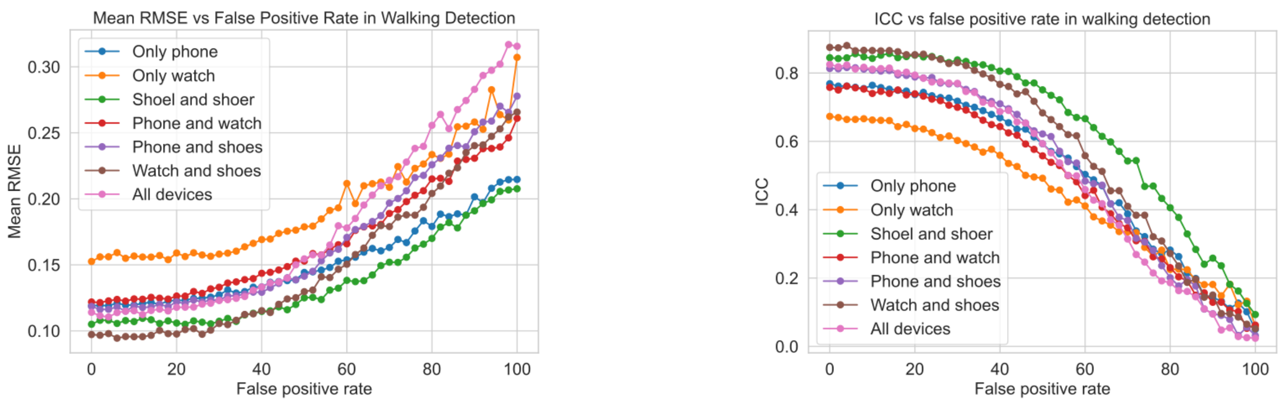
	Activity Recognition			**Gait Speed – Overall**		**Gait Speed – True Positives**				
	Resting	Level Walking	Stair Climbing	RMSE [m/s]	ICC	RMSE [m/s]	ICC	Bias	[LL, UL]	CV
Smartphone	0.85±0.042	0.91±0.097	0.92±0.046	0.119±0.055	0.767	0.119±0.056	0.764	-0.015	[-0.272, 0.241]	11.52
Smartwatch	0.81±0.070	0.68±0.272	0.74±0.185	0.144±0.071	0.708	0.154±0.059	0.677	-0.035	[-0.349, 0.279]	14.20
Shoes	0.84±0.059	0.99±0.015	0.99±0.010	0.115±0.056	0.824	0.107±0.049	0.846	0.028	[-0.198, 0.255]	9.96
Smartphone and Smartwatch	0.79±0.076	0.98±0.021	0.68±0.173	0.122±0.065	0.763	0.122±0.067	0.760	-0.021	[-0.289, 0.247]	12.05
Smartphone and Shoes	0.82±0.068	0.99±0.006	0.99±0.003	0.117±0.033	0.816	0.116±0.031	0.817	-0.035	[-0.262, 0.192]	10.29
Smartwatch and Shoes	0.79±0.086	0.99±0.013	0.99±0.005	0.096±0.027	0.874	0.095±0.030	0.874	-0.005	[-0.201, 0.190]	8.73
All Devices	0.77±0.082	0.95±0.050	0.99±0.013	0.112±0.031	0.825	0.112±0.032	0.821	-0.032	[-0.252, 0.189]	9.95

Values reported as Mean±std for RMSE; ICC = Intraclass Correlation Coefficient; Bland–Altman limits reported as Bias and [Lower Limit, Upper Limit], ±1.96 SD; CV = Coefficient of Variation (%).

TABLE 5: Comparison of Gait Speed Estimation Performances Reported in the Literature and in the Present Study (RMSE as Performance Metric)

Reference	**Sensor**	**RMSE [m/s]**
Mannini and Sabatini [38]	Accelerometer (thigh), subject-calibrated model	0.083
Mannini and Sabatini [38]	Accelerometer (thigh), non-subject-calibrated model	0.194
Soltani and Aminian [39]	Single inertial sensor (lower back)	0.10
Shresta and Won [40]	Smartphone	0.16
McGinnis et al. [41]	3 accelerometers (sacrum, thigh, shank)	0.136
Soltani et al. [42]	Wrist-worn sensor	0.10
Soltani et al. [42]	Wrist-worn sensor (personalized model)	0.05
Zanoletti et al. [26]	Smartphone, Smartwatch and Shoes	0.111
This study	Smartphone	0.119
	Smartwatch	0.144
	Shoes	0.115
	Smartphone and Smartwatch	0.122
	Smartphone and Shoes	0.117
	Smartwatch and Shoes	0.096
	Smartphone, Smartwatch and Shoes	0.112

Values for this study are reported as Mean ± Standard Deviation. Literature results are taken as reported in the cited works. Best performances in each group are highlighted in bold.



(a) Mean RMSE vs. false positive rate in walking detection across all device combinations.

(b) Intraclass correlation coefficient vs. false positive rate in walking detection across all device combinations.

FIGURE 6: Comparison of RMSE and ICC vs. false positive rate in walking detection across all device combinations. RMSE increases and ICC decreases as false positives rise.

ralistic assessment not only minimizes the burden on patients but also allows for more frequent and continuous monitoring,

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3648857

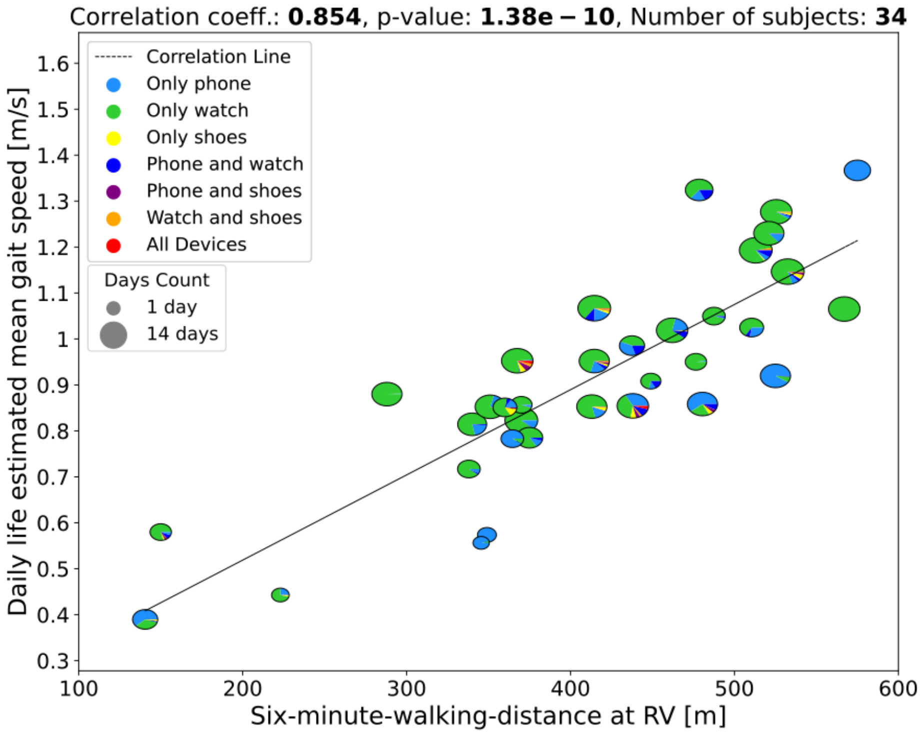


FIGURE 7: Spearman correlation coefficient between the mean daily-life estimated gait speed and the clinical 6MWD at the recruitment visit. The different colors inside the circles represent the proportion of the different configurations in the subject's daily data. The circle size indicates the number of days (from 1 to 14) for which a gait speed estimate was available for the subject.

enhancing the overall efficiency and effectiveness of clinical practice.

Observing the performance metrics reported in table 2 along with the confusion matrices in figures 8 it can be observed that the model achieves high accuracy in distinguishing between activity time windows (level walking and stair climbing) and non-activity time windows (resting). However, when examining the algorithm's ability to discriminate between level walking and stair climbing, some devices —particularly phone and smartwatch— show slightly worse performance. Nevertheless, both the phone and watch maintain consistent accuracy levels. This behavior can be reasonably attributed to the sensor placement: the smartwatch positioned on the wrist and the phone located in the front pants pocket may frequently interpret the limb and wrist movements during stair climbing as similar to those of level walking on flat terrain. In contrast, shoes, acquiring signals directly on the feet where the actual gait pattern occours, have a clear advantage in performing this classification and show better performance. The performance metrics shown in tables 4, 6, along with the cumulative confusion matrices for all subjects in figures 9 reveal a slight decrease in performance when coming to the validation dataset. This reduction could be attributed, particularly for the smartwatch, to the increased complexity of the task. During this acquisition, participants received less guidance on the walking activity than the 6MWT performed in the training dataset acquisition, where subjects were monitored in a controlled laboratory environment and worn with Xsens instrumentation, likely making them more conscious of their movements. In several configurations, the accuracy in identifying resting dropped below 80%. This misclassification may again be due to the greater freedom allowed during this trial. In an outdoor setting, as opposed to a controlled indoor lab, participants were more likely to engage in brief movements, such as taking a few steps during the resting position or keeping their hands closer to the body without using the handrail, making activity discrimination between level walking and stair climbing particularly challenging for the smartwatch. The smartwatch and smartphone configuration (Figure 10 D) shows the poorest performance, with nearly half of the windows labeled as level walking actually from other activities.

In the Results, we identified the most informative features across the four devices. Variability-based descriptors, including Shannon entropy and standard deviation, were consistently favored, with mean and range also frequently selected. These findings indicate that entropy and statistical dispersion capture the complexity and amplitude patterns most relevant to activity classification.

As shown in Tables 3, 4 and 5, the gait speed estimation demonstrates strong performance, achieving a mean RMSE of 0.096 m/s for the best configuration (smartwatch and shoes). Our results are consistent with state-of-the-art methods reported in the literature, with RMSE values comparable to or better than those achieved by single-sensor approaches [39], [40], [41]. Notably, our method achieves these performances without requiring subject-specific calibration. Some approaches show improved accuracy when personalized [38], [42]. The flexibility to use different device combinations while maintaining competitive accuracy represents a key advantage for real-world applications where device adherence may vary. In terms of reliability, all sensor combinations show good ICC values in both internal testing (Table 3) and external validation (Table 4). In the external validation under free-walking conditions, ICC values range from 0.677 to 0.874 across configurations, with the smartwatch and shoes combination achieving the highest value. The smartwatch alone exhibits only moderate reliability in our study, likely due to increased susceptibility to arm movements during non-walking activities. Kirk et al. [25] reported ICC values of 0.91 (laboratory) and 0.88 (real-world) for correctly detected walking periods, and 0.82 (laboratory) and 0.45 (real-world) when considering all detected periods.

Given the significant variation observed among devices, we considered it essential to assess the robustness of our estimates in the presence of false positives (FP). It is important to note that these percentages of false positives refer to the total number of windows within the specific class: for example, a false positive rate of 50% means that half of the non-walking windows were incorrectly identified as walking. The total number of false positives will therefore depend on the distribution of different activities within the whole recording. This variation highlights the importance of assessing model performance under different conditions, as the presence of false positives can significantly affect the reliability and accuracy of the results. Regarding the robustness analysis performed for both RMSE and ICC, it is observed that the values remain quite stable up to about 20-30% of FP rate. This means that a reasonable number of false positives would not have a significant impact on the estimation of average walking speed. This is probably due to two factors: firstly, the neural network was trained on a large number of walking windows with walking speeds ranging from 0.3 to 1.85 m/s; secondly, we use the trimmed mean (threshold of 20%) to calculate the average speed of a walking segment. However, if within a day a subject were to climb stairs for extended periods of time, an activity that is most easily confused with walking, this could affect the estimation.

It is essential to notice that the training of the walking detection, gait speed estimation and validation models was carried out on three completely independent datasets. Not only were these datasets collected from different subjects, but the data were also collected in dissimilar environments, further increasing the robustness of the analysis. The diversity in both participants and environmental conditions contributes to the generalisability of the models.

The proposed models are designed to be applicable to a wide range of wearable devices that have characteristics similar to those used in this study. In particular, they are suitable for devices that provide data types such as accelerometry, gyroscope and pressure measurements. This ensures that the methods outlined here can be adapted and extended to other wearable devices that collect comparable sensor data. The models can also be applied using a combination of any of these devices, depending on the data available. This flexibility is particularly important in real-world monitoring contexts where the sparsity of available data is a common challenge. In such settings, the ability to work with incomplete data or data from different sources significantly enhances the practical applicability of these models, making them more versatile and robust, and adaptable to user preferences.

A highly significant result emerging from these data is the ability to reliably estimate the gait speed without requiring active participation from the patients. Traditionally, such measures would necessitate the patient's presence in a clinical setting, such as during the 6MWT. This finding is particularly important for several reasons. Firstly, the ability to obtain these measurements without requiring patients to actively engage, such as traveling to a hospital or participating in a specific clinical assessment, highlights the potential for continuous, unobtrusive monitoring. This method allows for data collection in real-life settings without disrupting patients' daily routines. Moreover, this type of monitoring holds great promise for personalized medicine and remote patient monitoring, enabling healthcare providers to gather clinical data much more frequently than would be possible during periodic in-person visits.

Although the data were collected over a relatively short period (two weeks) the strong correlation with the six-minute walking distance shown in figure 7 (due to the non-parametric distribution of the data we used the Spearman correlation coefficient = 0.854, p-value = 1.38e-10) suggests that data obtained via the wearable device could serve as valid integrative information for assessing patients' motor function, potentially reducing the need for in-clinic measurements. Despite its lower individual performance, the smartwatch demonstrated the highest level of patient adherence, highlighting that ease of use and patient acceptance are critical factors for ensuring long-term compliance in monitoring. This is particularly important, as patient adherence remains a key challenge in long-term monitoring. If we compare the mean gait speed during the 6MWT at RV with the estimated mean gait speed in the subjects' daily life we would observe a mean bias of 0.23 m/s. This means that the speed observed in the subjects' daily life differs from the speed they maintained during the 6MWT in the RV. The fact that patients walk faster in the context of a clinical examination, where they are encouraged to perform at their best, is reasonable and consistent with what has been observed in literature [43] [44] [45].

Furthermore, the observed bias between gait speed during the 6MWT and daily life highlights the critical need to assess patients in their natural environments, rather than relying solely on performance in artificial clinical settings. These

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3648857

findings, consistent with prior research, emphasize the ecological validity of our approach. Continuous gait speed monitoring could complement traditional 6MWT assessments (typically performed at 3–6 month intervals) by enabling real-time detection of functional decline, potentially allowing timely identification of deterioration between clinical visits, in line with the Digital Mobility Outcomes framework from the Mobilise-D consortium [7], [25], [23], [14], [18]. Mobility is already a clinically established outcome, with the 6MWT serving as the in-clinic standard. In the patient population we are considering in the TOLIFE project [27], [28], [29] (i.e., COPD), we aim to leverage wearable-derived gait speed estimates as one of the driving factors for the early identification of exacerbations. Ultimately, achieving this goal requires continuous assessment seamlessly integrated into patients' daily lives. This work represents a significant step forward in real-world gait speed monitoring, offering promising implications for chronic disease management, rehabilitation, and personalized healthcare. By combining wearable sensors with machine learning, we demonstrated the feasibility of unobtrusive and continuous mobility assessment in daily-life settings. Future research should aim to validate these findings in larger and more diverse populations, explore integration with other wearable technologies, and generate clinically meaningful insights to support both patients and healthcare professionals.

V. CONCLUSION

In this study, we developed and validated an analysis pipeline to estimate gait speed from one to three wearable devices, namely smartphone, smartwatch, and sensorized shoes, that is robustly designed without subject-specific calibration and with improved handling of new datasets and variable data inputs. In real-world monitoring over two weeks in 34 COPD patients, estimated gait speed showed a strong correlation with clinical 6MWD (Spearman coefficient = 0.854, p-value = 1.38e-10), offering a valuable complement to traditional in-clinic assessments and potentially reducing the burden of resource-intensive clinical evaluations. Limitations included a high rate of false positives in walking-period detection in some scenarios, particularly with the smartwatch, which often misclassified stair climbing as level walking; nevertheless, its non-invasiveness and ease of use yielded the largest amount of collected data, highlighting its potential for long-term monitoring. Overall, these findings demonstrate the potential of wearable-based gait monitoring for remote patient management, enabling passive, continuous estimation of gait speed without patient intervention. Future work will focus on improving activity-recognition algorithms, by acquiring new datasets or integrating existing public datasets to improve robustness across manufacturers, sensor placements, and environments, expanding validation by extending the monitoring period and including larger, more diverse populations, and integrating these solutions into clinical workflows. Notably, while our current device selection is a first example, the framework is designed to accommodate additional devices over time, expanding the scope of real-world mobility monitoring and clinical applicability.

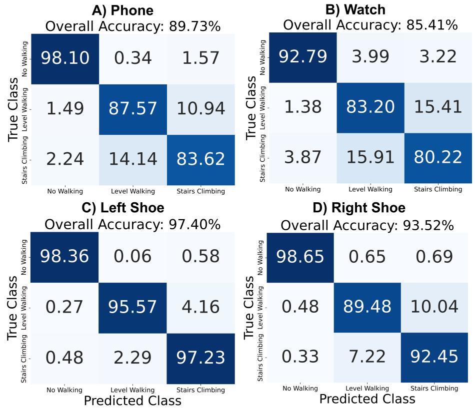


FIGURE 8: Confusion matrix over CV folds for phone (A), watch (B), left shoe (C), and right shoe (D).

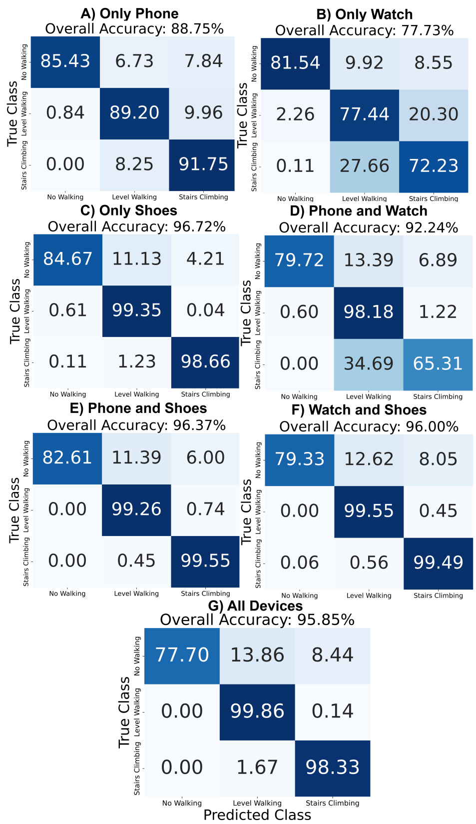


FIGURE 9: Confusion matrix over CV folds for only phone (A), only watch (B), only shoes (C), phone and watch (D), phone and shoes (E), watch and shoes (F), all devices (E).

APPENDIX I: DETAILED VALIDATION PERFORMANCES

The confusion matrices for Dataset A are presented in Figure 8, while those for Dataset C are shown in Figure 9. Additionally, the precision, recall, and F1-score metrics for Dataset C are reported in Table 6.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3648857

TABLE 6: Performance Metrics on External Validation for Activity Recognition

Device	Resting			Level Walking			Stair Climbing		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Smartphone	0.97±0.062	0.85±0.042	0.90±0.029	0.97±0.032	0.91±0.097	0.93±0.051	0.59±0.223	0.92±0.046	0.69±0.173
Smartwatch	0.93±0.120	0.81±0.070	0.86±0.064	0.90±0.068	0.68±0.272	0.75±0.227	0.29±0.114	0.74±0.185	0.40±0.097
Shoes	0.98±0.051	0.84±0.059	0.90±0.033	0.97±0.018	0.99±0.015	0.98±0.010	0.92±0.071	0.99±0.010	0.95±0.039
Smartphone and Smartwatch	0.98±0.052	0.79±0.076	0.87±0.044	0.92±0.029	0.98±0.021	0.85±0.017	0.75±0.128	0.68±0.173	0.69±0.107
Smartphone and Shoes	1.00±0.000	0.82±0.068	0.90±0.042	0.96±0.026	0.99±0.006	0.98±0.013	0.84±0.086	0.99±0.003	0.91±0.051
Smartwatch and Shoes	0.99±0.001	0.79±0.086	0.88±0.056	0.96±0.024	0.99±0.013	0.98±0.016	0.83±0.112	0.99±0.005	0.90±0.068
All Devices	1.00±0.000	0.77±0.091	0.87±0.059	0.96±0.029	0.99±0.002	0.98±0.015	0.84±0.100	0.99±0.014	0.91±0.059

The table reports values in the format [Mean±std].

Figure 10 reports all correlation and Bland-Altman plots related to Dataset C.

APPENDIX II: SELECTED FEATURES FOR EACH DEVICE

For the phone the 16 selected features are: the standard deviation and Shannon entropy of the accelerometer on the X-axis; the range and Shannon entropy of the accelerometer on the Y-axis; the mean, standard deviation, Shannon entropy, range, and number of peaks of the accelerometer on the Z-axis; the mean, range, and mean crossing rate of the orientation azimuth; the mean, range, and peak frequency of the orientation pitch; and Shannon entropy of the orientation roll. For the watch the 8 selected features are: the standard deviation, range, and Shannon entropy of the accelerometer on the X-axis; the standard deviation, root mean square, and Shannon entropy of the accelerometer on the Y-axis; and the standard deviation and Shannon entropy of the accelerometer on the Z-axis. For the left shoe the 13 selected features consist of: the mean and number of peaks of the accelerometer on the X-axis; the mean, standard deviation, range, and Shannon entropy of the accelerometer on the Z-axis; the mean, standard deviation, and number of peaks of pressure sensor 1; the mean and Shannon entropy of pressure sensor 2; and the mean and number of peaks of pressure sensor 3. For the right shoe the 16 selected features include: the mean and root mean square of the accelerometer on the X-axis; the mean and Shannon entropy of the accelerometer on the Y-axis; the mean, standard deviation, range, and Shannon entropy of the accelerometer on the Z-axis; the standard deviation, range, and Shannon entropy of pressure sensor 1; the mean and standard deviation of pressure sensor 2; and the standard deviation, root mean square, and Shannon entropy of pressure sensor 3.

DATA AVAILABILITY STATEMENT

The dataset is available on Zenodo (https://zenodo.org/records/16949732, accessed on 21 September 2025).

ETHICAL APPROVAL

Informed consent was obtained from all subjects. Ethical approval for acquisitions on healthy subjects was granted by the University of Pisa (N° 6/2024). Regarding the clinical study data ethical approval was issued for all sites by the Ethical Commission (EC) of the Medical Association of Schleswig-Holstein (Bad Segeberg; vote 074/23 ff), EC of the Tuscany Region - North West Area (Pisa; vote CET10/2023), and EC of Parc de Salut Mar (Barcelona; vote 2023/11230). All potential participants must sign an informed consent form before taking part in the study. The study was registered at ClinicalTrials.gov on 15/12/2023 and titled "COPD Exacerbation Modelling Using Unobtrusive Sensors - the TO-LIFE Clinical Study A (CSA)" (ClinicalTrials.gov Identifier: NCT06172712).

ACKNOWLEDGMENT

Funded by the European Union. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency (HADEA). Neither the European Union nor the granting authority can be held responsible for them.

REFERENCES

[1] V. P. T. Hornyak, J. M. VanSwearingen, and J. S. Brach, “Measurement of gait speed,” Topics in Geriatric Rehabilitation, vol. 28, no. 1, pp. 27–32, 2012.

[2] T. Buracchio, H. Dodge, D. Howieson, D. Wasserman, and J. Kaye, “The trajectory of gait speed preceding mild cognitive impairment,” Archives of Neurology, vol. 67, no. 8, pp. 980–986, 2010.

[3] S. Fritz and M. Lusardi, “White paper: "walking speed: The sixth vital sign",” Journal of Geriatric Physical Therapy, vol. 32, no. 2, pp. 46–49, 2009.

[4] M. D. Czech, D. Psaltos, H. Zhang et al., “Age and environment-related differences in gait in healthy adults using wearables,” npj Digital Medicine, vol. 3, p. 127, 2020.

[5] A. Mirelman, P. Bonato, R. Camicioli, T. D. Ellis, N. Giladi, J. L. Hamilton, C. J. Hass, J. M. Hausdorff, M. Mancini, A. Nieuwboer, and L. Rochester, “Gait impairments in parkinson’s disease,” The Lancet Neurology, vol. 18, no. 7, pp. 697–708, 2019.

[6] D. Renggli, C. Graf, N. Tachatos, N. Singh, M. Meboldt, W. R. Taylor, L. Stieglitz, and M. Schmid Daners, “Wearable inertial measurement units for assessing gait in real-world environments,” Frontiers in Physiology, vol. 11, 2020. [Online]. Available: https://www.frontiersin.org/journals/physiology/articles/10.3389/fphys.2020.00090

[7] L. e. a. Rochester, “A roadmap to inform development, validation and approval of digital mobility outcomes: The mobilise-d approach,” Digital Biomarkers, vol. 4 (Suppl. 1), p. 13–27, 2020.

[8] M. Wang, L. Li, M. Feng, and Z. Liu, “Advances in artificial intelligence applications for the management of chronic obstructive pulmonary disease,” Frontiers in Medicine, vol. 12, p. 1685254, 2025, edited

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3648857

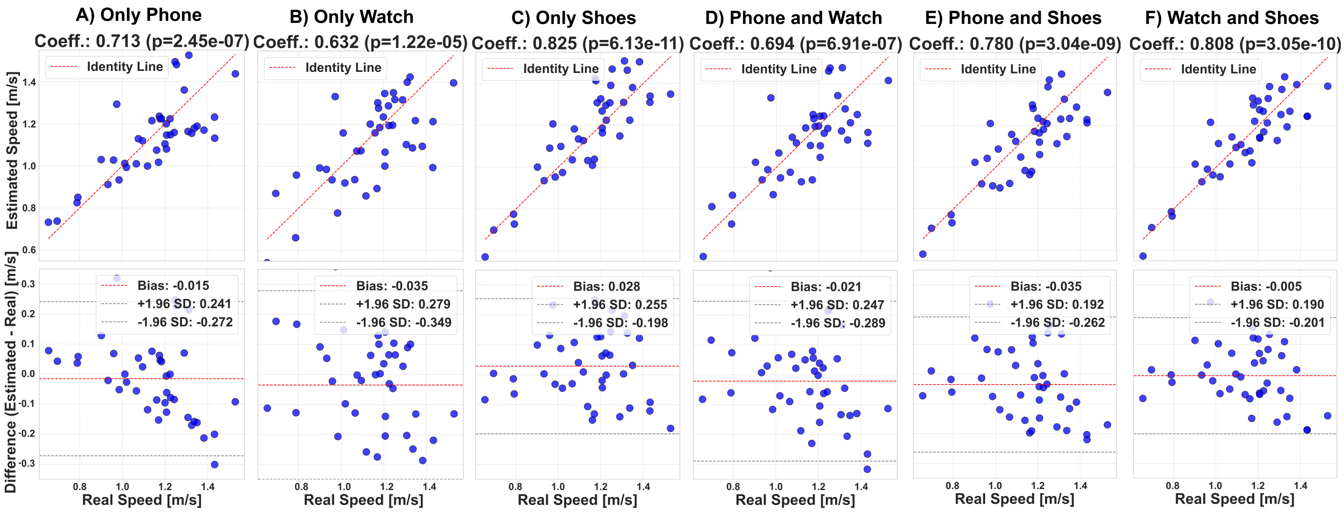


FIGURE 10: Correlation (with Spearman coefficient) and Bland-Altman plots for only phone (A), only watch (B), only shoes (C), phone and watch (D), phone and shoes (E), watch and shoes (F).

by Liang Zhao, Reviewed by Zhao Shuai and Jianting Wu. [Online]. Available: https://doi.org/10.3389/fmed.2025.1685254

[9] T. Kronborg, S. Hangaard, S. H. Laursen, L. K. E. Hæsum, J. Egmose, C. Bender, P. H. Secher, O. Hejlesen, and F. W. Udsen, "Impact of telemonitoring with exacerbation prediction algorithm versus telemonitoring alone on hospitalizations and health-related quality of life in patients with copd," Respiratory Care, vol. 70, no. 8, pp. 954–961, 2025, pMID: 40192545. [Online]. Available: https://doi.org/10.1089/respcare.12611

[10] S. Mei, X. Li, Y. Zhou et al., "Deep learning for detecting and early predicting chronic obstructive pulmonary disease from spirogram time series," npj Systems Biology and Applications, vol. 11, p. 18, 2025. [Online]. Available: https://doi.org/10.1038/s41540-025-00489-y

[11] C. Karpman and R. Benzo, "Gait speed as a measure of functional status in copd patients," International Journal of Chronic Obstructive Pulmonary Disease, pp. 1315–1320, 2014.

[12] D. Ilgin, S. Ozalevli, H. Karaali, A. H. Cimrin, and E. S. Ucan, "Gait speed as a functional capacity indicator in patients with chronic obstructive pulmonary disease," Annals of Thoracic Medicine, vol. 6, no. 3, pp. 141–146, 2011.

[13] B. Waschki, A. Kirsten, O. Holz, K. C. Mueller, M. Schaper, A.-L. Sack, T. Meyer, K. F. Rabe, H. Magnussen, and H. Watz, "Physical activity is the strongest predictor of all-cause mortality in patients with copd: A prospective cohort study," Chest, vol. 140, no. 2, pp. 331–342, 2011.

[14] L. e. a. Delgado-Ortiz, "Real-world walking cadence in people with copd," ERJ Open Research, vol. 10, no. 2, 2024. [Online]. Available: https://publications.ersnet.org//content/erjor/10/2/00673-2023

[15] A. Åkerberg, M. Lindén, and M. Folke, "How accurate are pedometer cell phone applications?" Procedia Technology, vol. 5, pp. 787–792, 2012.

[16] J. Leong and J. Wong, "Accuracy of three android-based pedometer applications in laboratory and free-living settings," Journal of Sports Sciences, vol. 35, no. 1, pp. 14–21, 2017, epub 2016 Mar 7.

[17] V. J. Beltrán-Carrillo et al., "Validity of the 'samsung health' application to measure steps: A study with two different samsung smartphones," Journal of Sports Sciences, vol. 37, no. 7, pp. 788–794, 2018.

[18] S. e. a. Del Din, "Analysis of free-living gait in older adults with and without parkinson's disease and with and without a history of falls: Identifying generic and disease-specific characteristics," The Journals of Gerontology: Series A, vol. 74, no. 4, p. 500–506, 2019.

[19] R. Romijnders, E. Warmerdam, C. Hansen, G. Schmidt, and W. Maetzler, "A deep learning approach for gait event detection from a single shank-worn imu: Validation in healthy and neurological cohorts," Sensors, vol. 22, no. 10, p. 3859, 2022.

[20] R. e. a. Romijnders, "Ecological validity of a deep learning algorithm to detect gait events from real-life walking bouts in mobility-limiting diseases," Frontiers in Neurology, vol. 14, 2023. [Online]. Available: https://www.frontiersin.org/journals/neurology/articles/10.3389/fneur.2023.1247532

[21] J. Ren, A. Wang, H. Li, X. Yue, and L. Meng, "A transformer-based neural network for gait prediction in lower limb exoskeleton robots using plantar force," Sensors, vol. 23, p. 6547, 2023.

[22] M. Endo, K. L. Poston, E. V. Sullivan, L. Fei-Fei, K. M. Pohl, and E. Adeli, "Gaitforemer: Self-supervised pre-training of transformers via human motion forecasting for few-shot gait impairment severity estimation," in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2022, ser. Lecture Notes in Computer Science, vol. 13438, 2022, pp. 130–139.

[23] F. e. a. Kluge, "Real-world gait detection using a wrist-worn inertial sensor: Validation study," JMIR Formative Research, vol. 8, p. e50035, 2024.

[24] Y. E. Brand, F. Kluge, L. Palmerini et al., "Self-supervised learning of wrist-worn daily living accelerometer data improves the automated detection of gait in older adults," Scientific Reports, vol. 14, p. 20854, 2024. [Online]. Available: https://doi.org/10.1038/s41598-024-71491-3

[25] C. Kirk, A. Küderle, M. E. Micó-Amigo et al., "Mobilise-d insights to estimate real-world walking speed in multiple conditions with a wearable device," Scientific Reports, vol. 14, p. 1754, 2024.

[26] M. Zanoletti, P. Bufano, F. Bossi, F. Di Rienzo, C. Marinai, G. Rho, C. Vallati, N. Carbonaro, A. Greco, M. Laurino et al., "Combining different wearable devices to assess gait speed in real-world settings," Sensors, vol. 24, p. 3205, 2024.

[27] tolife, "Tolife | ai and smart sensing for copd | horizon europe project," web, 2024.

[28] N. Carbonaro et al., "Smart sensors for daily-life data collection toward precision and personalized medicine: The tolife project approach," in MEDICON'23 and CMBEBIH'23. MEDICON CMBEBIH 2023, ser. IFMBE Proceedings, A. Badnjević and L. Gurbeta Pokvić, Eds. Springer, Cham, 2024, vol. 93.

[29] F. D. Rienzo et al., "Using multiple devices for patient monitoring in clinical studies: The tolife experience," in 2024 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events, 2024.

[30] N. Carbonaro, F. Lorussi, and A. Tognetti, "Assessment of a smart sensing shoe for gait phase detection in level walking," Electronics, vol. 5, no. 4, p. 78, 2016.

[31] M. Avvenuti, N. Carbonaro, M. G. C. A. 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, p. 3811, 2018.

[32] M. Zanoletti, P. Bufano, F. Bossi, F. Di Rienzo, C. Marinai, G. Rho, C. Vallati, N. Carbonaro, A. Greco, M. Laurino, and A. Tognetti, "Combining different wearable devices to assess gait speed in real-world setting [data set]," 2024. [Online]. Available: https://doi.org/10.5281/zenodo.11091279

[33] X. T. B.V., "Mvn gait report white paper," 2021, https://www.xsens.com/hubfs/Downloads/Whitepapers/Xsens_MVN_Gait_report_white_paper.pdf Accessed: 2025-03-27.

[34] J. Annegarn, M. A. Spruit, H. H. Savelberg, P. J. Willems, C. van de Bool, A. M. Schols, E. F. Wouters, and K. Meijer, "Differences in walking pattern during 6-min walk test between patients with copd and healthy subjects," PLoS One, vol. 7, no. 5, p. e37329, 2012. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0037329

[35] A. C. on Proficiency Standards for Clinical Pulmonary Function Laboratories, "Ats statement: guidelines for the six-minute walk test," American Journal of Respiratory and Critical Care Medicine, vol. 166, no. 1, pp. 111–117, July 2002, erratum in: Am J Respir Crit Care Med. 2016 May 15;193(10):1185. doi: 10.1164/rccm.19310erratum.

[36] T. K. Koo and M. Y. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," Journal of Chiropractic Medicine, vol. 15, no. 2, pp. 155–163, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1556370716000158

[37] R. Vallat, "Pingouin: statistics in python," Journal of Open Source Software, vol. 3, no. 31, p. 1026, Nov 2018.

[38] A. Mannini and A. M. Sabatini, "On-line classification of human activity and estimation of walk-run speed from acceleration data using support vector machines," in 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 2011, pp. 3302–3305.

[39] A. Soltani et al., "Algorithms for walking speed estimation using a lower-back-worn inertial sensor: A cross-validation on 556 speed ranges," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 29, pp. 1955–1964, 2021.

[40] A. Shrestha and M. Won, "Deepwalking: Enabling smartphone-based walking speed estimation using deep learning," in 2018 IEEE Global Communications Conference (GLOBECOM). IEEE, December 2018, pp. 1–6.

[41] R. S. McGinnis et al., "A machine learning approach for gait speed estimation using skin-mounted wearable sensors: From healthy controls to individuals with multiple sclerosis," PLOS ONE, vol. 12, no. 6, p. e0178366, June 2017.

[42] A. Soltani, H. Dejnabadi, M. Savary, and K. Aminian, "Real-world gait speed estimation using wrist sensor: A personalized approach," IEEE Journal of Biomedical and Health Informatics, vol. PP, pp. 1–1, 2019.

[43] I. Hillel, E. Gazit, A. Nieuwboer, L. Avanzino, L. Rochester, A. Cereatti, U. D. Croce, M. O. Rikkert, B. R. Bloem, E. Pelosin, S. Del Din, P. Ginis, N. Giladi, A. Mirelman, and J. M. Hausdorff, "Is every-day walking in older adults more analogous to dual-task walking or to usual walking? elucidating the gaps between gait performance in the lab and during 24/7 monitoring," Eur Rev Aging Phys Act, vol. 16, p. 6, 2019.

[44] M. A. Brodie, M. J. Coppens, S. R. Lord, N. H. Lovell, Y. J. Gschwind, S. J. Redmond, M. B. Del Rosario, K. Wang, D. L. Sturnieks, M. Persiani, and K. Delbaere, "Wearable pendant device monitoring using new wavelet-based methods shows daily life and laboratory gaits are different," Med Biol Eng Comput, vol. 54, no. 4, pp. 663–674, 2016.

[45] A. Mueller, H. A. Hoefling, A. Muaremi, J. Praestgaard, L. C. Walsh, O. Bunte, R. M. Huber, J. Fürmetz, A. M. Keppler, M. Schieker, W. Böcker, R. Roubenoff, S. Brachat, D. S. Rooks, and I. Clay, "Continuous digital monitoring of walking speed in frail elderly patients: Noninterventional validation study and longitudinal clinical trial," JMIR Mhealth Uhealth, vol. 7, no. 11, p. e15191, 2019.

BIOGRAPHY

Michele Zanoletti (Student Member, IEEE) obtained his specialist degree in biomedical engineering at the University of Pavia in 2019. He worked as a research fellow at the University of Pavia from 2019 to 2022. Currently he works at the Institute of Clinical Physiology, National Research Council and he is a Ph.D. student in Information Engineering at the University of Pisa. His research interests include the processing of biomedical signals and artificial intelligence algorithms applied to medicine.

Carlo Vallati is an associate professor with the Department of Information Engineering, University of Pisa. He is also the director of the Crosslab on Cloud Computing, Big Data and Cybersecurity funded by the Italian Ministry of Education and Research (MIUR) in the framework of the "Departments of Excellence" program. He is co-author of more than 100 peer-reviewed papers in international journals and conference proceedings. He has been involved in multiple national and international research projects. He has served as a program committee member for more than 40 international conferences/workshops and as Workshop Chair for IEEE IoT-SoS and IEEE SmartSys. He has served as TPC co-chair of IEEE SMARTCOMP 2020 and general vice chair of PERCOM 2022. He is a member of the EB of Ad Hoc Networks and Journal of Reliable Intelligent Environments.

Nicola Carbonaro (Member, IEEE) is an Associate Professor at the Department of Information Engineering and the Research Center "E. Piaggio," University of Pisa. He received his degree in Electronic Engineering from the University of Pisa in 2004 and a Ph.D. in Information Engineering from the same university in 2010, focusing on the development of wearable systems for human activity classification. In 2009, he was a Visiting Researcher at the "Neural Control of Movement" Laboratory at Arizona State University. He currently teaches the courses Sensor Systems for the Bachelor's Degree in Biomedical Engineering and Bionic Senses for the Master's Degree in Bionics Engineering at the University of Pisa. Dr. Carbonaro has contributed to several national and European research projects and has authored numerous scientific papers, conference proceedings, and book chapters. His research primarily focuses on the design and development of hardware and software for wearable sensing technologies for physiological and behavioral monitoring in biomedical applications.

Alberto Greco (Senior Member, IEEE) received a PhD degree in automatics, robotics, and bioengineering from the University of Pisa in 2015. He is currently an Associate professor at the Dipartimento di Ingegneria dell'Informazione and the Research Center "E. Piaggio" of theUniversity of Pisa. His main research interests are computational physiological signal processing and modeling, artificial intelligence, and wearable monitoring systems. Applications include affective computing and the assessment of mood and consciousness disorders. He is the author of several international scientific contributions in these fields. He is co-coordinator of the European research project POTION.

Alessandro Tognetti completed his degree in Electronic Engineering and earned his PhD in Bioengineering from the University of Pisa, Italy, in 2001 and 2005, respectively. He is currently Full Professor of Bioengineering at the Department of Information Engineering, University of Pisa, where he teaches courses on Biosensors, Bioelectric Phenomena, Bionic Senses, and Modelling of Multi-physics Phenomena. He has been actively involved in numerous national and international research projects (over 20 projects) and has collab-

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3648857

orated with universities, research institutes, and companies worldwide. He currently coordinates the Horizon Europe TOLIFE project. His research primarily focuses on the development of wearable and non-invasive sensors and systems for biomedical applications, innovative sensory principles, and bioelectric signal modelling for electrophysiological twins.

Marco Laurino received his B.S. and M.S. degrees in Biomedical Engineering from the University of Pisa, Italy, in 2007 and 2009, respectively. In 2014, he obtained his Ph.D. in Neuroscience from the University of Pisa. Currently he is a researcher at the Institute of Clinical Physiology (IFC), National Research Council (CNR) of Italy. His research focuses on biomedical signal and image processing, including Hd-EEG signal, EKG, skin conductance, and actigraphy. Additionally, he specializes in applying artificial intelligence techniques to medicine and biomedical systems modeling. He coordinated and participated to numerous research projects funded by European and National public institutions.