# Transductive Learning Models for Accurate Ambulatory Gait Analysis in Elderly Residents of Assisted Living Facilities

Huanghe Zhang, Ton T. H. Duong, Ashwini K. Rao, Pietro Mazzoni, Sunil K. Agrawal, Yi Guo, *Senior Member, IEEE*, and Damiano Zanotto, *Member, IEEE*

***Abstract*—Instrumented footwear represents a promising technology for spatiotemporal gait analysis in out-of-the-lab conditions. However, moderate accuracy impacts this technology's ability to capture subtle, but clinically meaningful, changes in gait patterns that may indicate adverse outcomes or underlying neurological conditions. This limitation hampers the use of instrumented footwear to aid functional assessments and clinical decision making. This paper introduces new transductive-learning inference models that substantially reduce measurement errors relative to conventional data processing techniques, without requiring subject-specific labelled data. The proposed models use subject-optimized input features and hyperparameters to adjust the spatiotemporal gait metrics (i.e., stride time, length, and velocity, swing time, and double support time) obtained with conventional techniques, resulting in computationally simpler models compared to end-to-end machine learning approaches. Model validity and reliability were evaluated against a gold-standard electronic walkway during a clinical gait performance test (6-minute walk test) administered to N $= 95$ senior residents of assisted living facilities with diverse levels of gait and balance impairments. Average reductions in absolute errors relative to conventional techniques were $-42.0\%$ and $-33.5\%$ for spatial and gait-phase parameters, respectively, indicating the potential of transductive learning models for improving the accuracy of instrumented footwear for ambulatory gait analysis.**

***Index Terms*—Ambulatory gait analysis, transductive learning, support vector regression, wearable technology, instrumented footwear.**

Manuscript received July 27, 2021; revised December 8, 2021 and January 4, 2022; accepted January 10, 2022. Date of publication January 13, 2022; date of current version January 28, 2022. This work was supported in part by the U.S. National Science Foundation under Grant IIS-1838799 and Grant IIS-1838725 and in part by the Columbia-Coulter Translational Research Partnership under Grant UR008231. *(Corresponding author: Damiano Zanotto.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Columbia University Institutional Review Board.

Huanghe Zhang, Ton T. H. Duong, and Damiano Zanotto are with the Department of Mechanical Engineering, Stevens Institute of Technology, Hoboken, NJ 07030 USA (e-mail: damiano.zanotto@stevens.edu).

Ashwini K. Rao is with the Department of Rehabilitation and Regenerative Medicine, Columbia University, New York, NY 10032 USA.

Pietro Mazzoni is with the Institute of Clinical and Translational Sciences, Washington University in St. Louis, St. Louis, MO 63110 USA.

Sunil K. Agrawal is with the Department of Mechanical Engineering, Columbia University, New York, NY 10027 USA.

Yi Guo is with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030 USA.

Digital Object Identifier 10.1109/TNSRE.2022.3143094

## I. Introduction

SPATIOTEMPORAL gait analysis can aid the diagnosis of neurological disorders in the elderly [1] and help identify frail individuals [2] and those at increased risk of falls [3], thereby informing more timely and effective treatment plans [4]. However, expensive laboratory equipment and the need for expert operation hamper the widespread use of gait analysis in clinical settings. Instrumented footwear such as smart shoes [5]–[8], insoles [9], [10], and on-shoe sensors [11], [12], have the potential to provide a low-cost scalable solution for gait analysis in unconstrained environments. However, these devices have so far demonstrated moderate validity and reliability in capturing spatiotemporal gait parameters in populations with gait impairments [13], [14]. Moderate accuracy makes it difficult to detect subtle changes in gait patterns that may be indicative of underlying neurological conditions. Moreover, moderate levels of agreement with gold-standard instrumentation preclude meaningful comparisons with the vast body of published normative gait data [15].

The computation of temporal and gait-phase parameters from raw time-series is typically achieved using force sensitive resistors (FSR) [16] or foot-attached inertial measurement units (IMUs) [13]. Estimating spatial gait parameters is more challenging. In principle, these metrics can be obtained by double integrating the accelerometric signals after compensating for gravity. However, this method is hardly accurate due to sensors noise and low-frequency drift. To mitigate the effect of sensor drift, a widely used approach relies on zero velocity update (ZUPT) and velocity drift compensation (VDC). ZUPT re-initializes the integration process after each gait cycle by setting the velocity of the foot to be zero every time the foot is stationary on the ground (foot-flat phase, FF [17]). VDC models the velocity drift between two consecutive FF phases using linear [13], [18] or nonlinear [11] models. The estimated drift is then subtracted from the velocity signal. The possibility to apply ZUPT and VDC represents a key advantage of foot-attached inertial sensors over their trunk- and wrist-attached conterparts [19]. However, ZUPT and VDC are not always effective with pathological gait, wherein stationary FF phases might not be easily detectable [13], [20].

To overcome the limitations of ZUPT and VDC, some researchers have investigated the application of classical machine learning (ML) regression models, such as Gaussian process regression (GPR) [21] and support vector regression (SVR) [22]. In these studies, an inference model is trained, which maps walking-related inertial features extracted from fixed-duration epochs to a target spatiotemporal gait metric (e.g., walking speed). Although these regression models are not affected by integration error accumulation, they cannot estimate walking speed with stride-by-stride granularity (i.e., the standard for spatiotemporal gait analysis) and their accuracy is heavily affected by the choice of hyperparameters and input features. Therefore, in these approaches grid search or random search with cross-validation are often required to appropriately tune the hyperparameters and select the optimal input features [23].

Other authors have proposed the use of learning-based *end-to-end* inference models. Leveraging the vast expressive power of deep-learning methods, these models directly take the raw time-series from inertial sensors as inputs to estimate a target gait parameter and therefore they are not affected by the drawbacks of ZUPT and VDC. For example, to measure stride length, Xing and co-workers used artificial neural networks (ANNs) [24], Hannink *et al.* applied deep convolutional neural networks (DCNNs) [20], while Wang and colleagues proposed the use of DCNNs and deep recurrent neural networks (DRNNs) [25]. End-to-end *subject-specific* models, in which training and testing data originate from the same individual, are typically accurate. However, due to inter-subject differences, the performances of these methods is significantly degraded when subject-specific labelled observations are not available during training [26]. Additionally, training deep-learning models typically requires larger datasets compared to classical machine learning models [26].

Collecting labelled data from each user is often unpractical and expensive in behavioral and biomedical research, since it requires tedious manual procedures or the use of specialized laboratory instrumentation as ground truth. One solution is offered by the *transfer learning* paradigm [27]. Transfer learning deals with the problem of domain adaptation, i.e., altering an inference model previously learned in a source domain to fit a new target domain for which limited or no labelled training data are available. In the context of wearable sensing devices, the target domain may represent a new user with different anthropometric attributes or different body sensor placements. Transfer learning approaches have been successfully applied to activity classification problems [27] and, to a lesser extent, also to predict continuous biomechanical variables [28]. However, the performances of a transfer learning model are greatly affected by the availability of labelled data from the target domain [29]. An alternative approach that does not suffer from this limitation is the *transductive learning* paradigm [30]. Unlike traditional (inductive) learning, transductive learning exploits knowledge about the location of the unlabelled test data with respect to the training data to provide *direct inference on the test data*, thereby trading model generalizability for improved performances [31]. This approach has great potential for use with connected wearables, given the growing availability of affordable cloud storage and computing services where previously collected labelled data can be used to train transductive models on demand. While transductive learning models have been proven to be effective in both classification [32], [33] and regression [34]–[36] problems, *none of these past studies has investigated the application of transductive learning methods to ambulatory gait analysis.*

Building upon our previous works on SVR models for ambulatory gait analysis [37]–[40], this paper introduces the following new contributions: (i) a novel transductive learning framework to improve validity and reliability of instrumented footwear in estimating spatiotemporal gait parameters, without the need for subject-specific labelled data; (ii) the validation of the proposed method against gold-standard instrumentation, in relation to conventional data processing techniques and to our previous SVR models, with a cohort of elderly residents of assisted living facilities; (iii) a study of the sensitivity of the proposed transductive models to different levels of gait impairment (as measured by a standardized clinical assessment) and to the use of mobility aids; and (iv) a feature analysis for the proposed models.

The rest of the paper is organized as follows. Section II illustrates the experimental protocol. Section III introduces the transductive learning framework for SVR models, and Section IV describes the approach for data analysis. Results are presented in Section V and discussed in Section VI. Concluding remarks are reported in Section VII.

## II. Experimental Protocol

Ninety-five seniors age 65 and older were recruited from five long-term care (LCT) facilities in the New York metropolitan area. Wearing instrumented footwear of appropriate size (Fig. 1), each participant completed the 6-minute Walk Test (6MWT) and the Timed Up And Go (TUG) test under the supervision of a physical therapist. Both assessments took place in the LCT facility where the participant lived. The instrumented footwear is a research prototype developed in the Columbia University Robotics and Rehabilitation Laboratory [41]. During the 6MWT, raw time-series acquired from an IMU and four FSRs embedded in each footwear were logged at 500 Hz. A 6-meter electronic walkway (Zeno Walkway, Protokinetics LLC, Havertown, PA, US) located in the middle of a 25-meter long course served as the gold-standard system for validating the inference models. The two systems were synchronized using a custom-engineered wireless board. Descriptive characteristics of the study participants are reported in Table I. The experimental protocol was approved by the Columbia University Institutional Review Board and all participants provided informed consent.

## III. Methods

The proposed inference framework is illustrated in Fig. 2. In line with the transductive inference paradigm [31], the locations of the test samples in the input feature space are exploited to improve model predictions. To this end, for each target user (i.e., for each iteration of the leave-one-out cross-validation (LOOCV) loop), the algorithm produces an individualized SVR model by leveraging the input feature vector

Fig. 1. Instrumented footwear used in this study [41].

TABLE I
SUBJECT INFORMATION

| | N = 95 |
|---|---|
| Age, years (SD) | 84.7 (8.1) |
| Sex | |
| Male, n (%) | 28 (29.5%) |
| Female, n (%) | 67 (70.5%) |
| Height, cm (SD) | 163.7 (9.1) |
| Weight, kg (SD) | 67.3 (16.8) |
| 6MWT, m (SD) | 230.4 (84.1) |
| TUG, s (SD) | 21.8 (13.4) |
| Risk of falling ($\geq$12.0 s [42]) | 80.0% |
| Gait speed, cm/s (SD) | 69.5 (25.0) |
| Normal ($\geq$ 100 cm/s [43]) | 11.6% |
| Risk for adverse events (60-80 cm/s [44]) | 32.6% |
| Impaired individuals (<60 cm/s [44]) | 34.7% |
| Walking conditions | |
| Non-assisted, n (%) | 44 (46.3%) |
| Cane, n (%) | 10 (10.5%) |
| Walker, n (%) | 41 (43.2%) |

$X^{te}$ extracted from the target user. A detailed description of the steps involved in the proposed method is reported in Sections III-A to III-D.

### A. Kernel Distance as Similarity Measure

Given a gait parameter of interest and a target user, the procedure first identifies the subset of individuals that most closely resemble the target user. These individuals will form the hold-out validation subset for the feature selection step described in Sec. III-B. To this end, the generalized distances between the training dataset and the test dataset in the feature space are computed as [36], [45]:

$$\begin{aligned} d_{i,j} &= ||\phi(X_i^{tr}(I)) - \phi(X_j^{te}(I))|| \\ &= \sqrt{K_{i,i}^{tr}(I) - 2K_{i,j} + K_{j,j}^{te}} \end{aligned} \tag{1}$$

$d_{i,j}$ is the generalized distance characterizing the degree of similarity between the $i$-th training instance and the $j$-th test instance. Vector $I$ contains the indices of the selected input features, which are measured by the instrumented footwear. $X^{tr}(I)$ and $X^{te}(I)$ are the vectors of the selected input features corresponding to the training dataset and the target user, respectively. $\phi(\cdot)$ represents the feature mapping function, which maps the selected subset of input features $X(I)$ into the high-dimensional feature space. $K_{i,i}^{tr}$ and $K_{j,j}^{te}$ are entries of the kernel matrices corresponding to the training dataset and the test dataset. $K$ is the kernel matrix defined as

$$K = \begin{bmatrix} k(X_1^{tr}(I), X_1^{te}(I)) & \cdots & k(X_1^{tr}(I), X_M^{te}(I)) \\ \vdots & \ddots & \vdots \\ k(X_L^{tr}(I), X_1^{te}(I)) & \cdots & k(X_L^{tr}(I), X_M^{te}(I)) \end{bmatrix}, \tag{2}$$

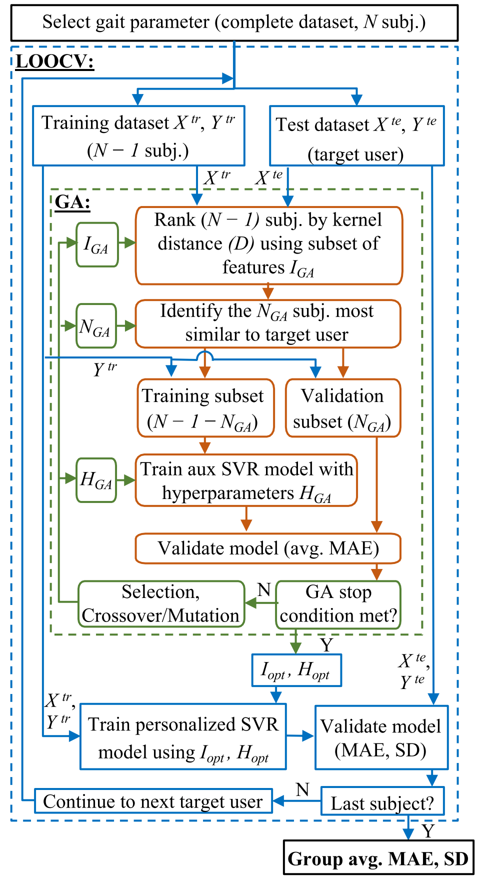


Fig. 2. Proposed transductive learning framework for a chosen gait parameter. $X^{tr}$ and $X^{te}$ represent the training dataset and the test dataset, respectively. $X^{te}$ is obtained from the selected target user (i.e., a particular iteration of the leave-one-out cross-validation (LOOCV) loop), for whom the individualized SVR model is being trained. $X^{tr}$ is extracted from the other $N-1$ subjects. The best subset of input features $I_{opt}$ and the best hyperparameters $H_{opt}$ for the individualized SVR model are estimated by a genetic algorithm (GA), using a subgroup of individuals whose input features are the most similar to the target user, according to a normalized kernel distance $D$. Accuracy (mean absolute error, MAE) and precision (standard deviation, SD) of the individualized SVR model relative to the gait measurements obtained with gold-standard instrumentation ($Y^{te}$) are computed for model validation purposes. It is worth noting that $Y^{te}$ is only used for model validation, therefore it is not available at train time.

where $k(\cdot)$ is the kernel function, and $L$, $M$ are the total number of instances in the training and the test dataset, respectively. The normalized distance $D_l$ between the $l$-th individual in the training dataset and the target user is given by

$$D_l = \frac{\sum_{i=1}^{N_{s,l}} \sum_{j=1}^{M} d_{i,j}}{N_{s,l}}, \tag{3}$$

where $N_{s,l}$ is the total number of training instances in the $l$-th subject's dataset.

### B. Feature Selection and Hyperparameters Tuning

As shown in Fig. 2, an efficient heuristic optimization method (genetic algorithm, GA [46]) determines the best subset of input features $I_{opt}$ and the best SVR hyperparameters $H_{opt}$ for the target user. The GA operates through a nested cross-validation loop, wherein the $N_{GA}$ individuals that are the most similar to the target user in terms of normalized distances (3) form the hold-out validation subset, and the remaining $(N - 1 - N_{GA})$ individuals form the training subset for an auxiliary SVR model. The number of individuals in the validation subset $N_{GA}$, the input feature subset $I_{GA}$, and the SVR hyperparameters $H_{GA}$ are the optimization variables. For each iteration of the GA, model performance (as quantified by the mean absolute error, MAE) is evaluated for each of the $N_{GA}$ individuals in the validation subset, and the average MAE is used as the cost function for the GA. This procedure is repeated until a stopping criterion is met.

In summary, because $I_{opt}$ and $H_{opt}$ maximize model accuracy on the group of individuals in the training dataset that most closely resemble the target user, they are regarded as the optimal parameters to train an individualized SVR model based on the training dataset $\left(X^{tr}(I_{opt}),\ Y^{tr}\right)$ consisting of $N - 1$ individuals, by following the SVR procedure discussed in Sec. III-D. It is worth noting that, since model training is performed on the largest available dataset (as opposed to the $N_{opt}$ most similar individuals), the resulting SVR model is less prone to overfitting.

## C. Feature Extraction

The set of candidate input features include the following variables:

*1) Spatiotemporal Gait Parameters:* The conventional procedure to estimate spatiotemporal gait parameters using instrumented footwear starts from the determination of gait events. The timing of the initial contacts (IC) events is computed based on underfoot FSRs [16]. Toe-off (TO) events are detected from IMU data using the method described in [47]. This was motivated by preliminary tests indicating that FSRs could not accurately capture TO events in this population of older adults. Stride time (ST) is defined as the time interval between two consecutive ICs of the same foot. Swing time (SW) is computed as the time interval between the TO of the current stride and the IC of the same foot's next stride, and SW% is determined as the ratio between SW and the corresponding ST. Double support time (DS) is calculated as the time interval between the IC of the contralateral foot and the TO of the ipsilateral foot. FSRs are used to determine foot-flat (FF) periods. Foot displacement over consecutive FF periods is computed by double integration of the gravity-corrected accelerometric signals, using ZUPT and VDC as detailed in [41]. The integration interval between consecutive FF periods is delimited by the time instant with the lowest acceleration magnitude within each FF period. Stride length (SL) is calculated as the L2-norm of the displacement vector and stride velocity (SV) is defined as the ratio between SL and the corresponding ST.

*2) Inertial Features:* Inertial features are computed between two consecutive FF periods of the same foot. They include the L2-norm of the foot acceleration ($|a|$), as well as the root mean square ($rms$), maximum ($max$), sum, energy,[1] and sample entropy[2] of the vertical ($a_z$), anteroposterior ($a_x$) and mediolateral ($a_y$) projections of the foot acceleration.

[1]Energy of a time-series is defined as the sum over the squared values of the time series data.

[2]Sample entropy quantifies the degree of randomness in the time series data.

*3) Anthropometric Features:* Anthropometric attributes include height, weight, body mass index (BMI), shoe size, age, gender, and type of walking assistance (i.e., walker/cane, or no assistive device). Dummy coding was used to include categorical features in the SVR models.

## D. Support Vector Regression (SVR)

The goal of the SVR models is to reduce measurement errors in the gait parameters obtained with the conventional data processing techniques described in Sec. III-C.1. SVR was selected over neural networks because the former has superior generalization accuracy and global optimization properties [48]. SVR models estimate a gait parameter at the i-$th$ stride as

$$\hat{Y}_i^{tr} = f(X_i^{tr}(I), \beta) = \beta^T \phi\left(X_i^{tr}(I)\right) \tag{4}$$

$Y_i^{tr}$ is the reference value of the gait parameter at the i-$th$ stride measured with the gold-standard equipment and $\hat{Y}_i^{tr}$ is the corresponding SVR estimate. $X_i^{tr}(I)$ is the vector of input features (restricted to the subset of features $I$), measured over the same stride. The weights vector $\beta$ is determined by numerically solving the constrained convex optimization problem

$$\min \frac{||\beta||^2}{2} + C\sum_{i=1}^{L}(\xi_i + \xi_i^*) \tag{5}$$

$$\begin{aligned} \text{subject to}\quad & Y_i^{tr} - \beta^T\phi\left(X_i^{tr}(I)\right) \le \varepsilon + \xi_i^*, \\ & i = 1,\ldots,L \\ & \beta^T\phi\left(X_i^{tr}(I)\right) - Y_i^{tr} \le \varepsilon + \xi_i, \quad i = 1,\ldots,L \\ & \xi_i, \xi_i^* \ge 0, \quad i = 1,\ldots,L \end{aligned} \tag{6}$$

where $\varepsilon$ and $C$ are the SVR hyperparameters. The former is a user-defined tolerance, the latter determines the trade-off between the flatness of $f(X(I), \beta)$ and the extent to which deviations larger than $\varepsilon$ are penalized. $\xi$ and $\xi^*$ are slack variables, bounding regression errors that are tolerated. Based on our previous work [37]–[39], Gaussian radial basis function (RBF) was selected as the type of kernel function, and the sets of candidate values for the hyperparameters were restricted to the following: $C \in$ [1 2 5 10 100], $\varepsilon \in$ [0.1 0.2 0.5 0.8 1 1.2 1.5 2 2.5 3].

## E. Data Processing Methods

To highlight the advantages of the proposed transductive models in estimating stride-by-stride spatiotemporal gait metrics, relative to the conventional data processing procedure and relative to our previous work [39], we compared validity and reliability of the following methods:

*1) Conventional Method (CONV):* This is the conventional data processing procedure described in Sec. III-C.1. This method, unlike the following ones, does not rely on inference models and therefore it does not require training. The CONV method was included in the analysis to verify whether the complexity of SVR models is well-justified by improvements in accuracy and precision.

*2) Subject-Specific SVR Models (SS):* With the SS method, SVR models are trained independently for each subject, using that subject's labelled data. For a given set of input features, this method yields the best performance that can be obtained using SVR. However, since subject-specific labelled data are required for each individual, the method has limited practical applicability. Each model is evaluated using 10-fold cross-validation. Within each fold, feature selection and hyperparameters optimization is achieved using GA, via a nested 9-fold cross-validation procedure wherein the cost function is the average MAE across the 9 folds [39].

*3) Generic SVR Models (GN):* This method generates *one-size-fits-all* models using LOOCV [39]. Within each training dataset (i.e., data from *(N-1)* individuals), feature selection and hyperparameters optimization is achieved using GA, via a nested 10-fold cross-validation procedure similar to the one used for SS, in which the cost function is the average MAE across the 10 folds.

*4) Transductive SVR Models (TR):* This method represents the main contribution of this paper. It generates *individualized* models using an optimized subset of labelled data taken from other individuals, as described in Sec. III-A and Sec. III-B.

## IV. Data Analysis

### A. Statistical Analysis

For each gait parameter, we assessed the interrater reliability of each data processing method relative to the ground-truth system using intraclass correlation coefficients (ICC), by employing a single-measurement, absolute-agreement, two-way mixed-effects model. Point estimates of the ICC were rated as poor ($< 0.5$), moderate (0.5 to 0.75), good (0.75 to 0.9), and excellent (0.9 to 1) [49].

Validity of the data processing methods was evaluated by calculating the MAE and the standard deviation of the errors (SD) for each gait parameter relative to the ground-truth values. MAE and SD were regarded as metrics of accuracy and precision, respectively, since they are widely used in the validation of wearable sensors [11], [13]–[15], [20], [21], [26], [50]. We applied repeated-measures ANOVA with *processing method* (CONV, SS, GN, TR) as within-subject factor, *walking condition* (non-assisted vs. assisted walking) as between-subject factor, and MAE and SD of the five gait parameters as dependent variables. Study participants who used a walker or a cane to perform the 6MWT were included in the *assisted walking* group. Mauchly's test was applied to check sphericity, and the Huynh-Feldt correction was applied if Mauchly's test indicated that the assumption of sphericity had been violated. When significant ($\alpha < 0.05$) effects were identified, post-hoc comparisons using the Bonferroni-Holm correction were applied as appropriate.

The level of accuracy of inertial sensors for ambulatory gait analysis typically degrades when those sensors are applied to clinical populations with motor deficits. Thus, Spearman's correlation analysis was carried out separately for each gait parameter to assess the effect of gait and balance impairment on the accuracy of the four data processing methods. The TUG score was selected as a surrogate metric for gait and balance impairment [51], and percentage MAE values were regarded as representative of model accuracy, to account for inter-subject biometric differences. The strength of the correlation was interpreted as negligible ($|\rho| < 0.3$), low ($0.3 \le |\rho| < 0.5$), moderate ($0.5 \le |\rho| < 0.7$), high ($0.7 \le |\rho| < 0.9$), or very high ($0.9 \le |\rho| < 1.0$) [52]. All statistical analysis was carried out in SPSS v28 (IBM Corporation, Armonk, NY).

### B. Feature Analysis

For each gait parameter, we determined the permutation feature importance $\mathrm{PFI}_{l,i}$, which yields the importance of the $l$-th input feature as related to the measurement errors of the TR model tailored to the $i$-th target user [53], [54]:

$$\mathrm{PFI}_{l,i} = \frac{\mathrm{MAE}_{l,i}^{perm}}{\mathrm{MAE}_{i}^{orig}} \tag{7}$$

$\mathrm{MAE}_{i}^{orig}$ is the original MAE of the $i$-th TR model for a given gait parameter, while $\mathrm{MAE}_{l,i}^{perm}$ is the MAE obtained by randomly permuting the $l$-th input feature within $X^{te}$, and feeding the altered input vector to the same TR model. The $\mathrm{PFI}_{l,i}$ values are averaged across all subjects for whom the $l$-th input feature is included in $I_{opt}$, resulting in a single scalar metric per each gait parameter. Because this procedure would not be applicable to the anthropometric features (since they remain the same for any given target user), the following modified equation was used in place of (7) for all anthropometric features

$$\tilde{\mathrm{PFI}}_{l,i} = \frac{1}{N_l}\frac{\sum_{n=1}^{N_l}\mathrm{MAE}_{l,i}^{n}}{\mathrm{MAE}_{i}^{orig}}, \tag{8}$$

where $N_l$ is the number of possible values that the $l$-th feature can take (e.g., if $l$ indicates sex, then $N_l = 2$) and $\mathrm{MAE}_{l,i}^{n}$ was computed by setting the $l$-th feature in $X^{te}$ to its $n$-th value.

## V. Results

A total of 9253 strides were collected simultaneously by the instrumented footwear and by the reference system. The number of strides per individual varied from 42 to 159 (97.4±24.6 mean and SD), depending on their velocity during the 6MWT. SL varied from 11.13 to 164.20 cm (83.83±25.93 cm), SV varied from 10.29 to 151.60 cm/s (70.40±26.73 cm/s), ST varied from 0.74 to 2.41 s (1.24±0.22 s), SW varied from 0.08 to 0.89 s (0.39±0.08 s), and DS varied from 0.05 to 0.93 s (0.23±0.08 s).

Training and testing of the inference models were conducted on a 4 GHz Intel® Core™ i7-6700K using MATLAB (The Mathworks Inc., Natick, MA). In the TR models, $N_{GA}$ ranged from 1 to 10 subjects. For each each gait parameter, it took approximately 180 minutes, 40 minutes, and 8 minutes to train GN, TR, and SS models, respectively.

### A. Interrater Reliability

Intraclass correlation coefficients are reported in Table II. In general, SVR models showed superior equivalence with the gold-standard equipment compared to the CONV method. However, differences in equivalence between CONV and SVR

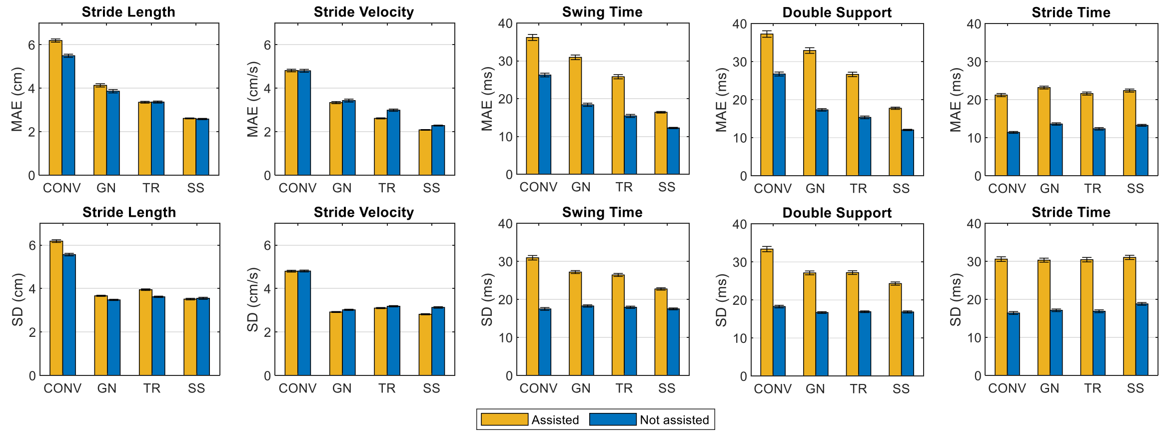


Fig. 3. Mean absolute errors (MAE) and standard deviation of the error (SD) for each gait parameter, averaged across all study participants who completed the 6MWT with (yellow) and without (blue) a mobility aid. Clusters represent the four processing methods analyzed in this study. Error bars indicate ±1SE.

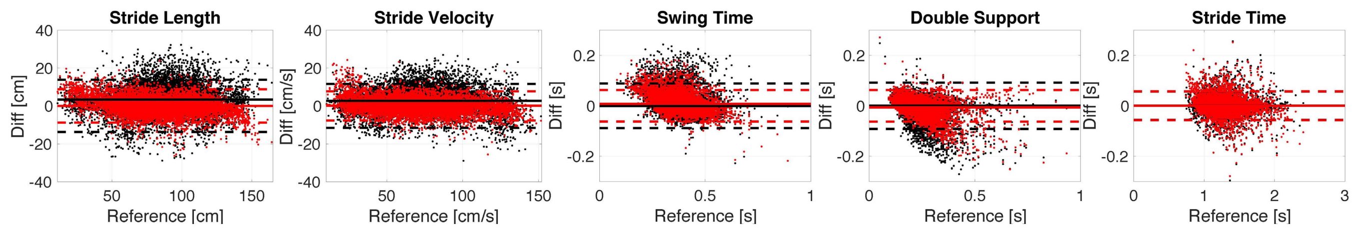


Fig. 4. Bland-Altman plots for the spatiotemporal gait parameters analyzed in this study, comparing the CONV methods (black) with the proposed TR models (red). In these plots, each dot represents a footfall, while dashed and solid horizontal lines indicate 95% limits of agreement (±1.96SD) and average errors, respectively.

TABLE II

Intraclass Correlation Coefficients for the Four Processing Methods. Values Between Parentheses Indicate 95% CI

| | CONV | GN | TR | SS |
|---|---|---|---|---|
| SL | .98 (.86-.99) | .99 (.98-.99) | 1.00 (.99-1.00) | 1.00 (1.00-1.00) |
| SV | .99 (.93-1.00) | .99 (.99-.99) | 1.00 (1.00-1.00) | 1.00 (1.00-1.00) |
| ST | 1.00 (1.00-1.00) | 1.00 (1.00-1.00) | 1.00 (1.00-1.00) | 1.00 (1.00-1.00) |
| SW | .84 (.77-.89) | .90 (.86-.93) | .94 (.90-.96) | 1.00 (1.00-1.00) |
| DS | .89 (.84-.93) | .94 (.91-.96) | .96 (.94-.98) | 1.00 (1.00-1.00) |

TABLE III

Accuracy (MAE) and Precision (SD) of the Estimated Gait Parameters (Pooled Data, Including Both Assisted and Non-Assisted Walking conditions)

| | CONV | | GN | | TR | | SS | |
|---|---|---|---|---|---|---|---|---|
| | MAE | SD | MAE | SD | MAE | SD | MAE | SD |
| SL [cm] | 5.87 | 5.90 | 4.01 | 3.58 | 3.36 | 3.80 | 2.60 | 3.53 |
| SV [cm/s] | 4.81 | 4.80 | 3.38 | 2.97 | 2.79 | 3.14 | 2.18 | 2.96 |
| ST [ms] | 16.66 | 24.04 | 18.74 | 24.22 | 17.31 | 24.17 | 18.15 | 25.39 |
| SW [ms] | 31.58 | 24.73 | 25.11 | 23.07 | 20.99 | 22.50 | 14.46 | 20.33 |
| DS [ms] | 32.60 | 26.69 | 26.00 | 22.50 | 21.63 | 22.64 | 15.21 | 21.01 |

methods were significant only in SW and DS, for the TR and SS models (as indicated by the confidence intervals (CIs) in Table II). The interrater reliability of SVR models slightly degraded when subject-specific training data were not available, and this effect was more marked for GN models than it was for TR models. Importantly, TR models demonstrated better stability (i.e., smaller CIs) than GN models.

## *B. Validity*

Fig. 3 shows the error metrics (MAE and SD) in the gait parameters for the four data processing methods, grouped by walking condition. Fig. 4 illustrates the Bland-Altman plots of all gait parameters for the CONV and the TR methods. The error metrics and the Bland-Altman plots were determined based on reference stride-by-stride data obtained with the gold-standard system. The MAE averaged across all study participants are reported in Table III, along with their SD. Results of the repeated-measures ANOVA are shown in Table IV and post-hoc analyses are reported in Table V.

TABLE IV

*p*-Values of the Two-Way Repeated-Measures ANOVA (PM = Processing Method, WC = Walking Condition)

| | MAE | | | SD | | |
|---|---|---|---|---|---|---|
| | PM | WC | PM*WC | PM | WC | PM*WC |
| SL | $<0.001$ | ns | ns | $<0.001$ | ns | ns |
| SV | $<0.001$ | ns | ns | $<0.001$ | ns | ns |
| ST | $<0.001$ | $<0.001$ | ns | $<0.05$ | $<0.001$ | ns |
| SW | $<0.001$ | $<0.001$ | $<0.05$ | $<0.001$ | $<0.001$ | $<0.001$ |
| DS | $<0.001$ | $<0.001$ | $<0.05$ | $<0.001$ | $<0.001$ | $<0.01$ |

The processing method significanlty affected the accuracy of SL and SV. MAE values in these gait metrics were significantly smaller when using all SVR models compared to the CONV method (Table V). SS models outperformed both GN and TR models in these two gait metrics, indicating the advantages of exploiting subject-specific labelled data, when they are available. On the other hand, it is worth noting that TR models yielded significantly smaller MAE than GN models, despite relying on the same set of available labelled observations (i.e., non subject-specific). The processing method also affected the MAE of ST. Interestingly, the CONV method yielded the most accurate results in ST (Tab. III), confirming the validity of conventional approaches based on FSRs to estimate this temporal parameter of gait [16]. Indeed, the CONV method outperformed both SS and GN methods (but not TR) in terms of accuracy. It is worth noting, however, that the increase in MAE between CONV and GN and SS

TABLE V

ADJUSTED $p$-VALUES OF THE POST-HOC ANALYSES FOLLOWING ONE-WAY REPEATED MEASURES ANOVA. THE BONFERRONI-HOLM METHOD WAS USED TO ADJUST FOR MULTIPLE PAIRWISE COMPARISONS ACROSS THE PROCESSING METHODS

| | CONV/GN | | CONV/TR | | CONV/SS | | GN/TR | | GN/SS | | TR/SS | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MAE | SD | MAE | SD | MAE | SD | MAE | SD | MAE | SD | MAE | SD |
| Pooled Data | | | | | | | | | | | | |
| SL | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.01 | <0.001 | ns | <0.001 | ns |
| SV | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | ns | <0.001 | ns |
| ST | <0.001 | ns | ns | ns | <0.001 | ns | <0.001 | ns | ns | ns | ns | ns |
| Non-assisted | | | | | | | | | | | | |
| SW | <0.01 | ns | <0.001 | ns | <0.001 | ns | <0.01 | ns | <0.001 | ns | <0.05 | ns |
| DS | <0.001 | ns | <0.001 | ns | <0.001 | ns | ns | ns | <0.001 | ns | <0.05 | ns |
| Assisted | | | | | | | | | | | | |
| SW | ns | ns | <0.001 | <0.01 | <0.001 | <0.001 | <0.05 | ns | <0.001 | <0.001 | <0.05 | <0.01 |
| DS | ns | <0.01 | <0.001 | <0.01 | <0.001 | <0.001 | <0.05 | ns | <0.001 | <0.05 | <0.05 | <0.05 |

was very modest (on average, less than 2.1 ms, which in our sample corresponded to only 0.09∼0.3% of the stride time), thereby indicating that all methods yielded excellent validity in estimating ST. Regardless of the processing method, errors in ST were larger for individuals who walked with a mobility aid. We reason that walking with a mobility aid augments an individual's forward lean, which in turn negatively impacts the accuracy with which IC events are detected, by altering the foot loading patterns upon which IC detection algorithms are based. In terms of gait phase variables (SW and DS), the accuracy of the processing methods depended on whether individuals walked with or without a mobility aid, as indicated by the significant interactions (Table IV). Separate one-way repeated-measures ANOVA (Table V) indicated that the TR and SS methods significantly outperformed the CONV method in both SW and DS for individuals requiring a walking aid, but the improvement was only marginal for the GN method. For the non-assisted walking condition, instead, results replicated those obtained for the MAE of SL and SV (i.e., all SVR models outperformed the CONV method). Unlike ST, gait phase metrics require the detection of both IC and TO events. We explain the difference in the performance of the CONV methods between ST and gait phase metrics by noting that capturing TO events is far more challenging than detecting IC when using foot-attached sensors, especially with elderly and impaired gait.

The analysis of the error variability (SD) for SL and SV revealed a significant effect of the processing method, with all the SVR models significantly outperforming the CONV method (Table V). The SVR models for SL and SV performed similarly in terms of precision, except for the TR method, which resulted in significantly worse precision compared to the GN method. We explain this result with the type of cost function chosen for the GA optimizer, which only accounted for accuracy. Similarly to the MAE, the precision of ST was negatively affected by the use of a walking aid, irrespective of the processing method. We found a significant effect of the processing method on the precision of ST, but the post-hoc analysis did not reveal significant differences among CONV, GN, TR and SS, suggesting that this effect was marginal. Additionally, the precision of the processing methods in determining SW and DS depended on the walking condition. For individuals walking without a mobility aid, no significant differences in precision were found across the processing methods in these two gait metrics. For those who walked with a mobility aid, however, all SVR models produced an improvement in precision relative to the CONV method

TABLE VI

SPEARMAN'S $\rho$ AND CORRESPONDING $p$-VALUES BETWEEN TUG AND THE PERCENT MAE OF GAIT PARAMETERS, FOR EACH DATA PROCESSING METHOD

| MAE | TUG | | | |
|---|---|---|---|---|
| % | CONV | GN | TR | SS |
| SL | 0.478 (<0.001) | 0.687 (<0.001) | 0.583 (<0.001) | 0.640 (<0.001) |
| SV | 0.517 (<0.001) | 0.711 (<0.001) | 0.579 (<0.001) | 0.676 (<0.001) |
| ST | 0.647 (<0.001) | 0.630 (<0.001) | 0.606 (<0.001) | 0.653 (<0.001) |
| SW | 0.295 (<0.01) | 0.615 (<0.001) | 0.543 (<0.001) | 0.668 (<0.001) |
| DS | ns | 0.321 (<0.01) | 0.277 (<0.01) | 0.461 (<0.001) |

(with SS performing better than TR and GN), even though the improvement was not significant for the GN method in estimating SW.

In summary, for spatial gait parameters (SL, SV), all inference models consistently outperformed the CONV method, both in terms of accuracy and precision, regardless of the walking condition. This was also the case for the gait phase variables (SW and DS), except that we did not find significant improvements in precision when applying SVR models to highly functioning individuals who did not use a mobility aid. These results clearly indicate that when subject-specific labelled data are not available for SL, SV, SW, and DS, the TR method should be preferred to the CONV method and to one-size-fits-all (GN) models. On the other hand, we found no evidence that the application of SVR models may further improve the validity of ST beyond the high levels of accuracy afforded by the CONV method.

### C. Effects of Gait Impairment on Validity

The results of the bivariate correlation analysis are summarized in Table VI. Except for DS estimated using the CONV method, the percentage MAE of all gait parameters, regardless of the data processing method, were positively associated with the TUG scores (i.e., the longer the time required to complete the TUG, the larger the measurement errors). This suggests that the more severe the gait impairment is, the more challenging it is to extract accurate spatiotemporal gait parameters using instrumented footwear. Nonetheless, all correlation coefficients were low or moderate.

### D. Feature Analysis

Table VII shows the number of input features selected by the GA optimizer for the GN and TR models. Compared to the GN method, the number of optimal input features for the TR method was consistently smaller across all gait metrics, indicating simpler and more efficient inference models. The PFI values for the TR models are reported in Tab. VIII.

TABLE VII

MEAN AND STANDARD DEVIATION (SD) OF THE NUMBER OF FEATURES SELECTED BY THE GA OPTIMIZER FOR THE GN AND TR MODELS

| | SL | | SV | | ST | | SW | | DS | |
|---|---|---|---|---|---|---|---|---|---|---|
| | GN | TR | GN | TR | GN | TR | GN | TR | GN | TR |
| MEAN | 23.8 | 15.6 | 23.7 | 15.3 | 18.7 | 11.2 | 22.6 | 14.0 | 23.5 | 14.7 |
| SD | 1.4 | 2.9 | 1.4 | 2.9 | 1.7 | 2.5 | 1.4 | 2.6 | 1.7 | 3.0 |

As expected, the accuracy of the TR models was mainly affected by the estimations obtained with the conventional method (indicated as $p^{CONV}$). The second most relevant feature for SL and SV was $sum(a_x)$, which is an indirect measure of the amount of velocity drift error accumulated after each stride.[3] Indeed, this value is exactly what the VDC method exploits to cancel out the effects of velocity drift. We conjecture that this input feature allows the inference models to compensate for the residual effects of velocity drift after applying VDC, thereby affecting the accuracy of both SV and SL.

It is worth noting that, for ST, the PFI of $p^{CONV}$ was approximately six times larger than any other input features. This provides strong evidence that inference models using this candidate set of input features might not be as effective in correcting the already modest measurement errors in ST produced by the CONV method. For SW and DS, the second most important feature was the wearer's $weight$. Because the magnitude of the propulsive forces at TO is associated with an individual's weight, this parameter may correct the accuracy with which TO events (and therefore SW and DS) are detected.

## VI. DISCUSSION

Unlike traditional inductive learning, transductive learning exploits knowledge about the positions of the unlabelled test data with respect to the training data in the input feature space to provide direct inference on the test data, resulting in improved performances [31]. To the best of the authors' knowledge, this study is the first one applying transductive learning models to ambulatory gait analysis. Taking full advantage of the transductive learning paradigm, the proposed approach generates personalized inference models without requiring subject-specific reference data to train the models, and therefore is applicable to *out-of-the lab* and *in-clinic* gait assessments, for which laboratory equipment is often not available. The TR models were validated against gold-standard instrumentation in a cohort of elderly residents of assisted living facilities and demonstrated significantly better accuracy than CONV techniques in all spatiotemporal gait parameters except for ST. Unlike the *end-to-end* ML methods proposed by recent research [20], [24]–[26], the TR models introduced in this paper do not replace conventional data processing techniques. Instead, the outputs generated using conventional techniques are embedded into the inference models as domain knowledge and augmented with a subject-tailored subset of input features that substantially reduce measurement errors. This results in more efficient learning (i.e., less complex and more accurate models) from the same training dataset and

[3]For the instrumented footwear shown in Fig. 1, the $x$ axis is aligned with the longitudinal axis of the sandal, pointing forward, and the $z$ axis is orthogonal to the footwear's sole, pointing downward.

TABLE VIII

PERMUTATION FEATURE IMPORTANCE FOR TR MODELS. THE TWO MOST IMPORTANT FEATURES FOR EACH GAIT PARAMETER ARE MARKED IN BOLD FONT. $p^{CONV}$ INDICATES THE CORRESPONDING GAIT PARAMETER (I.E., SL, SV, ST, SW OR DS) OBTAINED WITH THE CONV METHOD

| Features | SL | SV | ST | SW | DS |
|---|---|---|---|---|---|
| $p^{CONV}$ | **2.48** | **2.77** | **6.26** | **2.41** | **1.74** |
| $\|a\|$ | 1.08 | 1.12 | 1.05 | 1.06 | 1.07 |
| ST | 1.05 | 1.03 | - | 1.04 | 1.06 |
| SW% | 1.09 | 1.15 | 1.02 | - | 1.09 |
| rms($a_x$) | 1.12 | 1.08 | 1.06 | 1.22 | 1.08 |
| rms($a_y$) | 1.07 | 1.05 | **1.12** | 1.11 | 1.11 |
| rms($a_z$) | 1.16 | 1.20 | 1.07 | 1.08 | 1.19 |
| max($a_x$) | 1.00 | 1.02 | 1.01 | 1.03 | 1.01 |
| max($a_y$) | 1.01 | 1.02 | 1.01 | 1.03 | 1.02 |
| max($a_z$) | 1.01 | 1.01 | 1.00 | 1.04 | 1.03 |
| min($a_x$) | 1.01 | 1.01 | 1.01 | 1.00 | 1.01 |
| min($a_y$) | 1.01 | 1.03 | 1.01 | 1.02 | 1.00 |
| min($a_z$) | 1.00 | 1.01 | 1.00 | 1.01 | 1.01 |
| sum($a_x$) | **1.23** | **1.23** | 1.01 | 1.02 | 1.02 |
| sum($a_y$) | 1.02 | 1.03 | 1.00 | 1.03 | 1.01 |
| sum($a_z$) | 1.21 | 1.14 | 1.01 | 1.04 | 1.00 |
| energy($a_x$) | 1.11 | 1.17 | 1.04 | 1.11 | 1.10 |
| energy($a_y$) | 1.09 | 1.08 | 1.11 | 1.10 | 1.09 |
| energy($a_z$) | 1.14 | 1.10 | 1.02 | 1.13 | 1.10 |
| entropy($a_x$) | 1.01 | 1.03 | 1.00 | 1.02 | 1.00 |
| entropy($a_y$) | 1.02 | 1.03 | 1.00 | 1.03 | 1.01 |
| entropy($a_z$) | 1.05 | 1.05 | 1.00 | 1.05 | 1.04 |
| height | 1.09 | 1.04 | 1.00 | 1.00 | 1.11 |
| weight | 1.14 | 1.11 | 1.03 | **1.36** | **1.40** |
| BMI | 1.10 | 1.06 | 1.04 | 1.19 | 1.14 |
| shoe size | 1.14 | 1.15 | 1.06 | 1.21 | 1.30 |
| age | 1.08 | 1.04 | 1.01 | 1.05 | 1.05 |
| gender | 1.13 | 1.11 | 1.02 | 1.13 | 1.21 |
| assistance | 1.12 | 1.08 | 1.05 | 1.15 | 1.19 |

sensor hardware. Compared with our previous *one-size-fits-all* models (here indicated as GN), the TR models required approximately 20% of training time and 60% of the input features, yet they further reduced measurement errors by up to 18%.

We found good or excellent interrater reliability in all the spatiotemporal metrics extracted with the CONV method, indicating that conventional metrics and ground truth data are well-correlated. This result, however, must be interpreted with caution, since interrater reliability does not imply construct validity. Indeed, in our sample the CONV method yielded an average MAE of approximately 5 cm/s in SV (Tab. III), which is the minimal clinically important difference (MCID) for community dwelling older adults [55] and individuals with Parkinson's disease [56]. This indicates that about half of the SV values obtained with the CONV method were affected by measurement errors that exceeded the MCID. The TR inference models can mitigate this problem by offering excellent reliability and enhanced validity in all spatiotemporal parameters, without the need for subject-specific labelled data. When applying the TR models, the average MAE in SV dropped to approximately 3.4 cm/s.

Among the spatiotemporal gait parameters investigated in this work, ST represented an exception, since the measures produced by the CONV method demonstrated excellent inter-rater reliability (ICC = 1.00) and validity (MAE% = 1.3%), which could not be further improved by any of the proposed SVR models. We explain this result by noting that IC events can be reliably detected by underfoot FSRs, even when the IC does not occur under the hindfoot [57]. On the other hand, a limited number of underfoot FSRs might not yield

TABLE IX

ACCURACY OF IMU-BASED WEARABLE SYSTEMS REPORTED IN RECENT WORKS (**TOP:** TRADITIONAL METHODS; **MIDDLE:** SUBJECT-SPECIFIC INFERENCE MODELS; **BOTTOM:** MODELS NOT REQUIRING SUBJECT-SPECIFIC LABELLED DATA). THE TRANSUCTIVE-LEARNING MODELS (**TR-SVR**) REPRESENT THE MAIN CONTRIBUTION OF THE PAPER

| Ref. | Subjects | Population | # Strides | Methods | MAE% | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | SL | SV | ST | SW | DS |
| Mariani *et al.* [11] | 10 elderly | healthy | 492 | CONV | $4.9\%^a$ | $4.3\%^a$ | - | - | - |
| Rampp *et al.* [13] | 116 elderly | geriatric inpatients | 1220 | CONV | $7.8\%^b$ | - | $2.4\%^b$ | $6.7\%^b$ | - |
| Trojaniello *et al.* [60] | 10 elderly | healthy | 576 | CONV | 3.0% | - | 1.0% | 6.0% | - |
| Trojaniello *et al.* [60] | 10 elderly | Parkinson's | 532 | CONV | 2.0% | - | 1.0% | 7.0% | - |
| Ferrari *et al.* [14] | 14 elderly | Parkinson's | 1314 | CONV | $4.0\%^a$ | $8.0\%^a$ | $6.4\%^a$ | - | - |
| Kluge *et al.* [61] | 4 elderly | Parkinson's | 129 | CONV | 4.2% | 4.2% | 1.3% | 8.5% | - |
| **Our work** | **95 elderly** | **assisted living facilities** | **9253** | **CONV** | **7.5%** | **7.7%** | **1.3%** | **8.6%** | **14.7%** |
| Xing *et al.* [24] | 13 young | healthy | - | ANNs (SS model) | >2.0% | - | - | - | - |
| Hannink *et al.* [20] | 116 elderly | geriatric inpatients | 1220 | DCNNs (SS model) | $4.9\%^b$ | - | - | - | - |
| Wang *et al.* [25] | 5 young | healthy | 10000 | DRNNs (SS model) | >4.7% | - | - | - | - |
| **Our work** | **95 elderly** | **assisted living facilities** | **9253** | **SS-SVR** | **3.6%** | **3.7%** | **1.4%** | **3.9%** | **6.5%** |
| Zrenner *et al.* [26] | 27 young | healthy | 2377 | DCNNs | 5.9% | 5.9% | - | - | - |
| Zihajehzadeh *et al.* [21] | 15 young | healthy | - | GPR | - | 4.9% | - | - | - |
| McGinnis *et al.* [22] | 17 young | healthy | - | SVR | - | >6.0% | - | - | - |
| Zhang *et al.* [38] | 14 young | healthy | 5085 | GN-SVR | 1.4% | 1.2% | - | - | - |
| **Our work** | **95 elderly** | **assisted living facilities** | **9253** | **GN-SVR** | **5.7%** | **6.0%** | **1.5%** | **6.9%** | **10.9%** |
| **Our work** | **95 elderly** | **assisted living facilities** | **9253** | **TR-SVR** | **4.7%** | **4.9%** | **1.4%** | **5.9%** | **8.9%** |

[a] Estimated based on the reported accuracy% ± precision%, assuming a normal distribution.
[b] Estimated as the ratio between mean absolute error and mean estimations.

accurate estimates of TO events in elderly and pathological gait [58], which motivated the IMU-based TO detection approach adopted in this work. Even so, the estimates of SW and DS obtained with the CONV method using IMU-based TO detection demonstrated only good reliability and moderate validity (9% and 15% MAE, respectively), suggesting that the effectiveness of this approach might be affected by the reduced ankle plantarflexion angle at push-off, a common feature of elderly gait [59] that may have lowered the signal-to-noise ratio in the IMU time-series used to detect TO events.

When applying the CONV method, the results we obtained were in line with previous research on ambulatory gait analysis employing similar conventional data processing techniques (Tab. IX, top panel). This confirms that the baseline level of accuracy that the proposed inference models aimed to improve upon was comparable with state-of-the-art results. We attribute the differences in MAE to the different target populations (i.e., institutionalized elderly in this study, vs. community dwelling elderly in [11] and [60]), differences in gait event detection methods (i.e., IMU+FSR in this study vs. IMU in [13]), and to the small numbers of analyzed strides [13] and study participants [14], [60], [61] in most of the previous studies, which might not yield a truthful representation of the intra- and inter-subject variability of the population. It is also important to note that about one half of our study participants completed the 6MWT using a mobility aid (Tab. I). Among the related literature shown in Tab. IX, only the study in [13] analyzed the effects of walking aids on the accuracy of a wearable system.

When using SS models to estimate SL, we found better results than what was reported in [20] and [25], and worse results than [24], which all relied on neural networks (Tab. IX, middle panel). However, for [24] and [25], the validation was limited to a small sample of healthy young adults, thereby making it difficult to compare performances with this study. On the other hand, the percentage MAEs of SL and SV obtained with the GN models were in line with those presented in [26] for DCNN models (Tab. IX, bottom panel), which did not rely on subject-specific training data. The accuracy in SV for GN models was better than what was reported in [22] using SVR models, but worse than [21] using GPR models. However, these two studies estimated the average walking speed during a 5-second window, which is related, but not equivalent, to the definition of stride velocity commonly used in spatiotemporal gait analysis. Additionally, such an average metric cannot be used to estimate stride-to-stride variability, which is important in elderly and pathological gait.

The most interesting results in our study came from the TR models. In terms of accuracy, *TR models outperformed GN models in all the analyzed gait parameters, despite relying on the same available set of labelled data*, thereby providing strong support for the use of personalized inference models in ambulatory gait analysis. However, TR models did not reduce error variability compared to GN. We explain this result with the type of cost function used in the GA optimizer, which was blind to SD. Notably, compared to past research (Tab. IX), TR models yielded more accurate SL estimations than the DCNNs in [20], even though the latter work relied on subject-specific training data. Moreover, the MAE in SV obtained with TR models was similar to the results presented in [21], which collected data from healthy individuals. Despite these encouraging results, we found a significant gap in validity between SS models and TR models, suggesting that subject-specific training data should be used whenever possible, if validity is a priority.

The results from the correlation analyses suggested that the level of gait impairment affects the validity of the inference models and the CONV methods in a similar fashion. Although the strength of the correlation was low or moderate, this result emphasizes the importance of validating new technologies and computational methods for ambulatory gait analysis on the clinical population of interest, since results obtained with highly functional individuals might not transfer to those with gait impairments. Further research is warranted to better elucidate the sensitivity of wearable systems for ambulatory gait analysis to the use of walking aids, given that our results evidenced a general trend towards degraded accuracy and precision in all temporal and gait-phase parameters in the individuals who used a cane or a walker.

This study has several limitations. First, the experimental protocol did not allow us to quantify the test-retest reliability of the TR models. While the GN models are not affected by this problem and the SS models showed excellent test-retest reliability [38], rigorously evaluating the extent to which the accuracy of TR models decreases when a previously trained TR model is applied to subsequent walking tests will inform us about the need to train session-specific TR models, which has important practical implications given the computational time required to train each model. For example, using previously-trained TR models would enable real-time gait analysis (e.g., for use with wearable biofeedback systems [62]). Conversely, should the TR models demonstrate poor test-retest reliability, gait data could be generated in real-time using previously trained GN models, while higher-accuracy data could be made available off-line, after training session-specific TR models. Second, the benefits of the TR approach rely on the availability of labelled observations from individuals who have similar gait patterns to the test subject. Consequently, different labelled datasets will likely be required when studying different populations (e.g., pediatric vs. adult populations). Third, although most standardized clinical gait performance tests include bouts of straight-line walking on even surfaces similar to those analyzed in this study, the results we presented might not apply to real-life walking, which often consists of short walking bouts on even and uneven surfaces, turns, episodes of gait initiation/termination, and other motor tasks such as stairs negotiation [63]. Therefore, we plan to direct future studies toward validating the TR models in motor tasks that more closely represent real-life walking, to investigate the range of applicability of these methods to extended-time gait monitoring in real-life environments. Furthermore, a future study will focus on extending the use of transductive learning methods for weights optimization in the SVR models, and on improving the GA-based feature selection procedure to increase the precision of the models (e.g., through a multi-objective optimization). We also plan to develop similar ML inference models to extract kinetic gait parameters (such as center-of-pressure trajectories [64]), which were not addressed in this study. Lastly, it is worth noting that, even though the proposed inference models were validated on a labelled dataset collected using the instrumented footwear shown in Fig. 1, these methods can be readily applied to any IMU-based footwear system, such as instrumented insoles and on-shoe sensors.

## VII. Conclusion

This paper introduced novel transductive learning models to improve validity and reliability of instrumented footwear for spatiotemporal gait analysis. The performance of the proposed models were compared with those of conventional data processing techniques using overground walking data collected from a cohort of older adults. Conventional methods demonstrated high validity and reliability only in measuring stride time, however they were outperformed by the transductive models in terms of stride length, velocity, double support time, and swing time. Additionally, the transductive models showed better accuracy than previously developed *one-size-fits-all* inference methods (despite relying on the same available set of labelled data), while reducing model complexity and training time. By shifting the focus from traditional inductive inference to direct inference on a target dataset, transductive models may represent a promising method to improve the validity and range of applicability of instrumented footwear for ambulatory gait analysis.

## Acknowledgment

The authors would like to thank Dr. Enrique Jinete and the staff at *Healing Therapeutics, LLC* for their help with recruiting, scheduling, and administering the functional assessment tests to the study participants.

## References

[1] Y. Moon, J. Sung, R. An, M. E. Hernandez, and J. J. Sosnoff, "Gait variability in people with neurological disorders: A systematic review and meta-analysis," *Hum. Movement Sci.*, vol. 47, pp. 197–208, Jun. 2016.

[2] 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.

[3] J. Verghese, R. Holtzer, R. B. Lipton, and C. Wang, "Quantitative gait markers and incident fall risk in older adults," *J. Gerontol., A*, vol. 64, no. 8, pp. 896–901, 2009.

[4] T. A. L. Wren, G. E. Gorton, III, S. Õunpuu, and C. A. Tucker, "Efficacy of clinical gait analysis: A systematic review," *Gait Posture*, vol. 34, no. 2, pp. 149–153, Jun. 2011.

[5] N. Hegde *et al.*, "The pediatric SmartShoe: Wearable sensor system for ambulatory monitoring of physical activity and gait," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 2, pp. 477–486, Feb. 2018.

[6] J. Montes, D. Zanotto, S. D. Young, R. Salazar, D. C. De Vivo, and S. Agrawal, "Gait assessment with solesound instrumented footwear in spinal muscular atrophy," *Muscle Nerve*, vol. 56, no. 2, pp. 230–236, Aug. 2017.

[7] D. Zanotto *et al.*, "Dizziness handicap inventory score is highly correlated with markers of gait disturbance," *Otology Neurotol.*, vol. 38, no. 10, pp. 1490–1499, 2017.

[8] B. Szeto *et al.*, "Hearing loss is associated with increased variability in double support period in the elderly," *Sensors*, vol. 21, no. 1, p. 278, Jan. 2021.

[9] M. S. Oerbekke, M. J. Stukstette, K. Schütte, R. A. de Bie, M. F. Pisters, and B. Vanwanseele, "Concurrent validity and reliability of wireless instrumented insoles measuring postural balance and temporal gait parameters," *Gait Posture*, vol. 51, pp. 116–124, Jan. 2017.

[10] H. Zhang, D. Zanotto, and S. K. Agrawal, "Estimating CoP trajectories and kinematic gait parameters in walking and running using instrumented insoles," *IEEE Robot. Autom. Lett.*, vol. 2, no. 4, pp. 2159–2165, Oct. 2017.

[11] B. Mariani, C. Hoskovec, S. Rochat, C. Büla, J. Penders, and K. Aminian, "3D gait assessment in young and elderly subjects using foot-worn inertial sensors," *J. Biomech.*, vol. 43, no. 15, pp. 2999–3006, Nov. 2010.

[12] J. Rudisch *et al.*, "Agreement and consistency of five different clinical gait analysis systems in the assessment of spatiotemporal gait parameters," *Gait Posture*, vol. 85, pp. 55–64, Mar. 2021.

[13] A. Rampp, J. Barth, S. Schülein, K.-G. Gaßmann, J. Klucken, and B. M. Eskofier, "Inertial sensor-based stride parameter calculation from gait sequences in geriatric patients," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 4, pp. 1089–1097, Apr. 2015.

[14] A. Ferrari, P. Ginis, M. Hardegger, F. Casamassima, L. Rocchi, and L. Chiari, "A mobile Kalman-filter based solution for the real-time estimation of spatio-temporal gait parameters," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 7, pp. 764–773, Jul. 2016.

[15] S. Chen, J. Lach, B. Lo, and G.-Z. Yang, "Toward pervasive gait analysis with wearable sensors: A systematic review," *IEEE J. Biomed. Health Inform.*, vol. 20, no. 6, pp. 1521–1537, Nov. 2016.

[16] J. M. Hausdorff, Z. Ladin, and J. Y. Wei, "Footswitch system for measurement of the temporal parameters of gait," *J. Biomechanics*, vol. 28, no. 3, pp. 347–351, Mar. 1995.

[17] A. Peruzzi, U. Della Croce, and A. Cereatti, "Estimation of stride length in level walking using an inertial measurement unit attached to the foot: A validation of the zero velocity assumption during stance," *J. Biomech.*, vol. 44, no. 10, pp. 1991–1994, Jul. 2011.

[18] J. R. Rebula, L. V. Ojeda, P. G. Adamczyk, and A. D. Kuo, "Measurement of foot placement and its variability with inertial sensors," *Gait Posture*, vol. 38, no. 4, pp. 974–980, 2013.

[19] A. Atrsaei, F. Dadashi, B. Mariani, R. Gonzenbach, and K. Aminian, "Toward a remote assessment of walking bout and speed: Application in patients with multiple sclerosis," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 11, pp. 4217–4228, Nov. 2021.

[20] J. Hannink *et al.*, "Mobile stride length estimation with deep convolutional neural networks," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 2, pp. 354–362, Mar. 2018.

[21] S. Zihajehzadeh and E. J. Park, "Experimental evaluation of regression model-based walking speed estimation using lower body-mounted IMU," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 243–246.

[22] 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, Jun. 2017, Art. no. e0178366.

[23] A. Géron, *Hands-on Machine Learning With Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. Newton, MA, USA: O'Reilly Media, 2019.

[24] H. Xing, J. Li, B. Hou, Y. Zhang, and M. Guo, "Pedestrian stride length estimation from IMU measurements and ANN based algorithm," *J. Sensors*, vol. 2017, pp. 1–10, Feb. 2017.

[25] Q. Wang, L. Ye, H. Luo, A. Men, F. Zhao, and Y. Huang, "Pedestrian stride-length estimation based on LSTM and denoising autoencoders," *Sensors*, vol. 19, no. 4, p. 840, Feb. 2019.

[26] M. Zrenner, S. Gradl, U. Jensen, M. Ullrich, and B. Eskofier, "Comparison of different algorithms for calculating velocity and stride length in running using inertial measurement units," *Sensors*, vol. 18, no. 12, p. 4194, Nov. 2018.

[27] A. Akbari *et al.*, *Transfer Learning or Wearable Computers*. Amsterdam, The Netherlands: Elsevier, 2021, pp. 435–459.

[28] T. T. Dao, "From deep learning to transfer learning for the prediction of skeletal muscle forces," *Med. Biol. Eng. Comput.*, vol. 57, no. 5, pp. 1049–1058, 2019.

[29] A. Arnold, R. Nallapati, and W. W. Cohen, "A comparative study of methods for transductive transfer learning," in *Proc. 7th IEEE Int. Conf. Data Mining Workshops (ICDMW)*. IEEE, 2007, pp. 77–82.

[30] V. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer, 1995.

[31] V. Vapnik, *Estimation of Dependences Based on Empirical Data*. New York, NY, USA: Springer, 2006.

[32] T. Joachims, "Transductive inference for text classification using support vector machines," in *Proc. 16th Int. Conf. Mach. Learn.*, vol. 99, 1999, pp. 200–209.

[33] M. Ji *et al.*, "Graph regularized transductive classification on heterogeneous information networks," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*. Berlin, Germany: Springer, 2010, pp. 570–586.

[34] A. Schwaighofer and V. Tresp, "Transductive and inductive methods for approximate Gaussian process regression," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 15, 2002, pp. 977–984.

[35] D. Tang, T.-H. Yu, and T.-K. Kim, "Real-time articulated hand pose estimation using semi-supervised transductive regression forests," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 3224–3231.

[36] M. Li, L. Xie, and Z. Wang, "A transductive model-based stress recognition method using peripheral physiological signals," *Sensors*, vol. 19, no. 2, p. 429, 2019.

[37] H. Zhang, M. O. Tay, Z. Suar, M. Kurt, and D. Zanotto, "Regression models for estimating kinematic gait parameters with instrumented footwear," in *Proc. 7th IEEE Int. Conf. Biomed. Robot. Biomechatronics (Biorob)*, Aug. 2018, pp. 1169–1174.

[38] H. Zhang, Y. Guo, and D. Zanotto, "Accurate ambulatory gait analysis in walking and running using machine learning models," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 1, pp. 191–202, Jan. 2020.

[39] H. Zhang, Z. Chen, D. Zanotto, and Y. Guo, "Robot-assisted and wearable sensor-mediated autonomous gait analysis§," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2020, pp. 6795–6802.

[40] T. T. H. Duong *et al.*, "Validation of insole-based gait analysis system in young children with a neurodevelopmental disorder and autism traits," in *Proc. 8th IEEE RAS/EMBS Int. Conf. Biomed. Robot. Biomechatronics (BioRob)*. IEEE, 2020, pp. 715–720.

[41] S. Minto, D. Zanotto, E. M. Boggs, G. Rosati, and S. K. Agrawal, "Validation of a footwear-based gait analysis system with action-related feedback," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 9, pp. 971–980, Sep. 2016.

[42] (2020). *Timed Up & Go (TUG)*. Online Report. Centers for Disease Control and Prevention. [Online]. Available: https://www.cdc.gov/steadi/pdf/TUG_Test-print.pdf

[43] M. Montero-Odasso *et al.*, "Gait velocity as a single predictor of adverse events in healthy seniors aged 75 years and older," *J. Gerontol. A, Biol. Sci. Med. Sci.*, vol. 60, no. 10, pp. 1304–1309, 2005.

[44] N. M. Peel, S. S. Kuys, and K. Klein, "Gait speed as a measure in geriatric assessment in clinical settings: A systematic review," *J. Gerontol. A, Biol. Sci. Med. Sci.*, vol. 68, no. 1, pp. 39–46, 2012.

[45] B. Schölkopf, "The kernel trick for distances," in *Proc. Adv. Neural Inf. Process. Syst.*, 2001, pp. 301–307.

[46] J. Huang, Y. Cai, and X. Xu, "A hybrid genetic algorithm for feature selection wrapper based on mutual information," *Pattern Recognit. Lett.*, vol. 28, no. 13, pp. 1825–1844, May 2007.

[47] A. M. Sabatini, C. Martelloni, S. Scapellato, and F. Cavallo, "Assessment of walking features from foot inertial sensing," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 3, pp. 486–494, Mar. 2005.

[48] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY, USA: Springer, 2009.

[49] T. K. Koo and M. Y. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," *J. Chiropractic Med.*, vol. 15, no. 2, pp. 155–163, Jun. 2016.

[50] M. Bertoli *et al.*, "Estimation of spatio-temporal parameters of gait from magneto-inertial measurement units: Multicenter validation among parkinson, mildly cognitively impaired and healthy older adults," *Biomed. Eng. OnLine*, vol. 17, no. 1, p. 58, Dec. 2018.

[51] A. Christopher, E. Kraft, H. Olenick, R. Kiesling, and A. Doty, "The reliability and validity of the timed up and go as a clinical tool in individuals with and without disabilities across a lifespan: A systematic review," *Disability Rehabil.*, vol. 43, pp. 1–15, Oct. 2019.

[52] M. M. Mukaka, "A guide to appropriate use of correlation coefficient in medical research," *Malawi Med. J.*, vol. 24, no. 3, pp. 69–71, 2012.

[53] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[54] A. Fisher, C. Rudin, and F. Dominici, "All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously," *J. Mach. Learn. Res.*, vol. 20, no. 177, pp. 1–81, 2019.

[55] G. Pulignano *et al.*, "Incremental value of gait speed in predicting prognosis of older adults with heart failure," *JACC, Heart Failure*, vol. 4, no. 4, pp. 289–298, Apr. 2016.

[56] C. J. Hass *et al.*, "Defining the clinically meaningful difference in gait speed in persons with Parkinson disease," *J. Neurol. Phys. Therapy*, vol. 38, no. 4, pp. 233–238, 2014.

[57] J. Rueterbories, E. G. Spaich, B. Larsen, and O. K. Andersen, "Methods for gait event detection and analysis in ambulatory systems," *Med. Eng. Phys.*, vol. 32, no. 6, pp. 545–552, Jul. 2010.

[58] J. A. Ramirez-Bautista, J. A. Huerta-Ruelas, S. L. Chaparro-Cárdenas, and A. Hernández-Zavala, "A review in detection and monitoring gait disorders using in-shoe plantar measurement systems," *IEEE Rev. Biomed. Eng.*, vol. 10, pp. 299–309, 2017.

[59] Y. Barak, R. C. Wagenaar, and K. G. Holt, "Gait characteristics of elderly people with a history of falls: A dynamic approach," *Phys. Therapy*, vol. 86, no. 11, pp. 1501–1510, 2006.

[60] D. Trojaniello *et al.*, "Estimation of step-by-step spatio-temporal parameters of normal and impaired gait using shank-mounted magneto-inertial sensors: Application to elderly, hemiparetic, Parkinsonian and choreic gait," *J. NeuroEng. Rehabil.*, vol. 11, p. 152, 2014.

[61] F. Kluge, H. Gaßner, J. Hannink, C. Pasluosta, J. Klucken, and B. M. Eskofier, "Towards mobile gait analysis: Concurrent validity and test-retest reliability of an inertial measurement system for the assessment of spatio-temporal gait parameters," *Sensors*, vol. 17, no. 7, p. 1522, Jun. 2017.

[62] H. Zhang *et al.*, "Wearable biofeedback system to induce desired walking speed in overground gait training," *Sensors*, vol. 20, no. 14, p. 4002, 2020.

[63] M. S. Orendurff, "How humans walk: Bout duration, steps per bout, and rest duration," *J. Rehabil. Res. Develop.*, vol. 45, no. 7, pp. 1077–1090, Dec. 2008.

[64] T. T. H. Duong *et al.*, "Gaussian process regression for COP trajectory estimation in healthy and pathological gait using instrumented insoles," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2021, pp. 9548–9553.