Real-world Balance Assessment while Standing for Fall Prediction in Older Adults

This article has been accepted for publication in IEEE Transactions on Biomedical Engineering. 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/TBME.2023.3326306

Real-world Balance Assessment while Standing for Fall Prediction in Older Adults

Jose Albites-Sanabria, Graduate Student Member, IEEE, Pierpaolo Palumbo, Jorunn L. Helbostad, Stefania Bandinelli, Sabato Mellone, Luca Palmerini, Lorenzo Chiari

***Abstract*— Postural control naturally declines with age, leading to an increased risk of falling. Within clinical settings, the deployment of balance assessments has become commonplace, facilitating the identification of postural instability and the development of targeted interventions to forestall falls among older adults. A dearth of studies has ventured beyond the controlled laboratory, leaving, however, a gap in our understanding of balance in real-world scenarios. In response, this study combined previously reported algorithms to build a finite-state machine (FSM) with four states: walking, turning, sitting, and standing. The FSM was validated against video annotations (gold standard) in an independent dataset with data collected on 20 older adults. Later, the FSM was applied to data from 168 community-dwelling older people in the InCHIANTI cohort. The InCHIANTI participants were evaluated both in the laboratory and then remotely in real-world conditions for a week. In identifying fallers, mean frequency, sway path, and jerk, computed during standing, revealed significant relationships with fall risk. A 70/30 data split with recursive feature selection and resampling techniques was used to train and test four machine-learning models. Our findings revealed that the best-performing model (Lasso Regression) built on real-world balance features had a higher area under the curve (AUC, 0.76) than one built on lab-based assessments (0.57). This study shows, for the first time, that real-world balance characteristics while standing differ significantly from lab-based assessments and are more predictive than lab-based assessments in identifying older adults at higher risk of falling.**

***Index Terms*— lab-based, real-world, inertial sensors, balance assessment, fall risk**

I. Introduction

Falls are a significant health concern for older adults, as they can lead to severe injuries and loss of independence. According to the World Health Organization (WHO), falls are the second leading cause of accidental or unintentional injury deaths worldwide[1].

Identifying individuals at risk for falls is therefore critical for preventing falls and reducing associated health care costs. One important aspect of fall risk assessment is balance evaluation. Poor balance has been identified as a significant risk factor for falls in older adults [2]–[4]. Traditional balance assessment methods include clinical measures such as the Berg Functional Balance Scale [5] and the Balance Evaluation Systems Test [6], or force plate stabilometry [7]. Nonetheless, administering those tools requires trained personnel and specialized laboratory or clinical settings.

Although laboratory assessments demonstrated a high potential for identifying balance and mobility issues, it is currently unknown whether they accurately reflect the complexity and diversity of balance during daily life activities or whether they are related to responses to perturbations of balance that occur in real life [8], [9]. Indeed, several gaps remain in our understanding of the use of wearable sensors for real-world balance assessment. Considering fall risk in older adults, these gaps are particularly relevant and require urgent responses. First, there is a need to evaluate the feasibility of using real-world recordings to identify standing events among different daily activities. Second, it is still unclear whether and to what extent real-world balance assessments using wearable sensors are comparable to laboratory-based assessments. Finally, to the best of our knowledge, no research has ever investigated the prognostic ability of a wearable sensor-based, real-world balance assessment for falls.

In this study, we aimed to develop and validate a tool to assess balance while standing in real-world conditions with a single inertial sensor placed at the lower back. We carried out three tasks to achieve this objective: i) we developed a finite-state machine (FSM) algorithm and validated it against a gold standard (video annotations); ii) we compared measurements

The study has been funded in part by the European Union (Seventh Framework Program (FP7/2007-2013) under grant agreement No.: 288940 (FARSEEING project)), and by the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 820820 (Mobilise-D). This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA. The InCHIANTI follow-up study 4 (2013–2015) was supported by the Agency for Regional Health Care of Tuscany (Reg. n° 586/R, 26/10/2012). *(Corresponding author: Jose Albites-Sanabria.)*

J. Albites-Sanabria is with the Department of Electrical, Electronic, and Information Engineering "Guglielmo Marconi", University of Bologna, Bologna 40136, Italy and the Institute of Advanced Studies, University of Bologna, Bologna 40136, Italy (e-mail: jose.albitessanabri2@unibo.it).

P. Palumbo is with the Department of Electrical, Electronic, and Information Engineering "Guglielmo Marconi", University of Bologna, Bologna 40136, Italy (e-mail: pierpaolo.palumbo@unibo.it).

J. L. Helbostad is with the Department of Neuroscience, Norwegian University of Science and Technology, Trondheim 7491, Norway (e-mail: jorunn.helbostad@ntnu.no).

S. Bandinelli is with the Azienda Sanitaria Toscana Centro, Firenze, Piero Palagi Hospital, Firenze, Italy (e-mail: stefania1.bandinelli@uslcentro.toscana.it).

S. Mellone, L. Palmerini and L. Chiari are with the Department of Electrical, Electronic, and Information Engineering "Guglielmo Marconi", University of Bologna, Bologna 40136, Italy and the Health Sciences and Technologies—Interdepartmental Center for Industrial Research, University of Bologna, Bologna, 40136, Italy (e-mail: sabato.mellone@unibo.it, luca.palmerini@unibo.it , lorenzo.chiari@unibo.it).

This article has been accepted for publication in IEEE Transactions on Biomedical Engineering. 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/TBME.2023.3326306

obtained from laboratory and real-world balance assessments; and iii) we analyzed the laboratory and real-world balance assessments in terms of their prognostic ability in identifying individuals at risk of falling.

II. Materials and Methods

A. Study Protocol

The present study is based on data from two different cohorts of community-dwelling older adults: the ADAPT [10] (A Personalized Fall Risk Assessment System for promoting independent living) and the InCHIANTI [11] ("Invecchiare nel Chianti") studies. A total of 20 older adults (76.4±5.6 years), 75% female, participated in the ADAPT study, performing various Activities of Daily Living (ADL). The ADAPT data collection protocol was divided into two sessions: an in-lab semi-structured supervised protocol and a free-living unsupervised protocol (out-of-lab). In the in-lab semi-structured protocol, subjects were asked to follow a series of tasks in the "Usability Laboratory" at the Faculty of Medicine and Health Sciences at the Norwegian University of Science and Technology, Trondheim, while being monitored by ceiling-mounted cameras. As a part of the free-living protocol, subjects were instructed to perform their usual ADLs naturally and to include predefined activities without any instruction or supervision on how to perform them. A GoPro camera was attached to the chest as a gold standard [10]. Several inertial sensing units were placed at various body locations. Our analysis used only the sensor worn on the lower back (uSense, 3D accelerometer, 3D gyroscope, 100 Hz sampling frequency).

The second part of the study is based on data from the 4th follow-up of the InCHIANTI study (clinical trial: NCT01331512). One hundred and sixty-eight community-dwelling older adults over 65 years (79.7±6.6), 50.9% female, were monitored using a smartphone embedded with a tri-axial accelerometer and gyroscope (100 Hz sampling frequency), worn on the lower back in a belt. During the laboratory assessment, participants were evaluated with the Romberg test [12]. The first quiet standing condition (eyes open on a rigid surface for 30 seconds) was used in the analyses of this study. For the real-world analysis, participants received a dedicated smartphone and information on how to wear it (using a belt) and take care of it properly. They were instructed to wear it after dressing, from morning to night, during their usual daily activities for a weekly monitoring period, and then return it to the clinical staff. A fall was defined as "an unexpected event in which the person comes to rest on the ground, floor, or lower level [13]". Prospective fall incidence was ascertained through monthly telephone interviews for 6 months and at the 12th month from the start of continuous monitoring. Participants who did not fall prospectively were defined as non-fallers (NFs). Participants who fell one or more times were defined as fallers (Fs). Table I provides demographic information for both studies. The Ethical Committee of the Italian National Institute of Research and Care of Aging [11] approved the InCHIANTI study protocol. The protocols of both studies comply with the Declaration of Helsinki. All participants received a detailed description of the study purpose and procedures and gave their written informed consent.

TABLE I
ADAPT and InCHIANTI study cohort characteristics

	ADAPT	InCHIANTI (4th follow-up)		
	Participants (N=20)	Non-Fallers [NFs] (N=140)	Fallers [Fs] (N=28)	Combined (N=168)
Gender (M/F)	5/15	72/68	11/17	83/85
Age (years)	76.4 ± 5.6	79.4 ± 6.7	81.1 ± 5.6	79.6 ± 6.6
Height (cm)	167 ± 7.2	159.8 ± 9.1	158.8 ± 9.6	159.6 ± 9.2
Weight (kg)	73.7 ± 11.4	70.9 ± 13	69.4 ± 13.9	70.7 ± 13.1
MMSE	-	27.3 ± 1.9	27.1 ± 1.8	27.3 ± 1.9

B. Data Processing

Inertial sensor data from the lower back were acquired from both datasets. The signals from the sensor were first low-pass filtered with a cutoff frequency of 5 Hz to remove high-frequency noise. Then, gait, sitting transfers, and turn identification algorithms were implemented in Python 3.8 based on previously validated studies [14]–[16].

Standing events were identified using a finite-state machine (FSM) logic based on the identified gait, sitting, and turn events. An FSM is a model abstraction for any system with a limited number of conditional states of being, and it has an excellent advantage for real-world applications, given its predictability. The proposed FSM identified standing events by first detecting all stationary periods. Stationary periods were defined as epochs in which the participant's lower back was almost entirely still ($|a| < 0.05\, m/s^2$) [15]. The algorithm then examined the surrounding activities (gait, turn, and sitting transfers) to determine whether the stationary period was an actual standing event. A standing event was identified if it occurred in between gait activities and after or before sitting transfers (Fig. 1 and Algorithm 1 in Appendix).

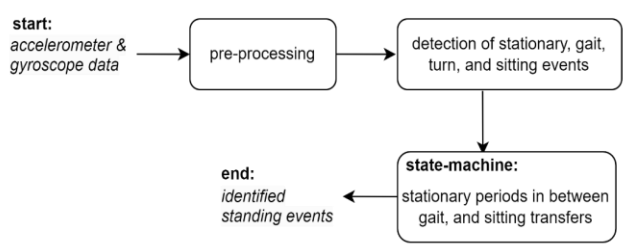


Fig. 1. Finite-state machine (FSM) algorithm development

Balance features were selected from the literature. Only features showing good test-retest reliability were considered for the study (Table II) [10]–[12]. Balance features were calculated from the sensor data during the first 30 seconds of the standing laboratory assessments. For the real-world assessment, all the identified standing events with a duration of at least 30 seconds were considered for the analysis. Still, balance features were calculated only during the first 30 seconds of each event. Each

This article has been accepted for publication in IEEE Transactions on Biomedical Engineering. 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/TBME.2023.3326306

real-world balance feature was then averaged across all the identified standing events for each subject. In addition, two additional features were extracted: the number of stands per hour (stands/h) and the duration of the standing events (actual identified duration of the event).

TABLE II
SUMMARY OF EXTRACTED BALANCE FEATURES. IF NOT SPECIFIED, FEATURES' DEFINITIONS AND FORMULAS WERE TAKEN FROM [17]

Abbreviation	Description
Time-domain	
Range	Range of acceleration $[m/s^2]$
Distance	Mean distance of acceleration center $[m/s^2]$
mean Frequency	Rounds required to cover full acceleration trajectory $[Hz]$
Path	Sway path, total length of trajectory $[m/s^2]$
Area	Sway area computed as area spanned from acceleration per unit of time $[m^2/s^5]$
Ellipse area	Ellipse that encapsulates 90% of the data points $[m/s^2]$
Frequency-domain	
Total power	Power content versus frequency $[Hz]$
95% Freq	95% power frequency (F95) $[Hz]$
Spectral centroid	Centroidal Frequency $[Hz]$
Complex	
Jerk	Log dimensionless jerk (LDLJ-A)[28]
SampleE	Sample Entropy: regularity and unpredictability of acceleration and gyroscope
LDE	Maximum finite-time Lyapunov exponent[29]

C. Validation and Statistical Analysis

The FSM was validated using the video-labeled annotations (gold standard) in the ADAPT dataset. Data from both in-lab and out-lab protocols were merged for the analysis. Sensitivity, specificity, accuracy, and F1-score were computed based on the labels from the video camera recordings (gold standard). Bland-Altman plots were used to assess the agreement between the algorithm and the gold standard concerning the start time and the end of the identified standing events.

In the InCHIANTI study, the Mann-Whitney U test was performed to analyze differences between fallers and non-fallers, and adjusted p-values (Benjamini-Hochberg) were computed to account for the false discovery rate. Univariate associations between balance features and prospective fallers were assessed using logistic regression analysis. The predictive performance of balance features obtained from real-world recordings and lab assessments was evaluated by fitting four machine learning classification models: Logistic Regression, Lasso Regression, Support Vector Machine (SVM), and Decision Tree. A 70-30% data split was performed, with 70% of the dataset used for training and 30% for testing the models. These models have demonstrated the ability to discern patterns and relationships within complex datasets and have been used in previous studies for fall risk assessment [4], [18]–[20]. To account for imbalanced data issues, random undersampling, Synthetic Minority Oversampling Technique (SMOTE), and near-miss techniques were analyzed for each classification technique. Recursive feature elimination was used in the training dataset to identify relevant features in each model. Finally, the area under the curve (AUC), sensitivity, specificity, accuracy, F1-score, and geometric mean were computed as performance metrics for the trained models in the test data. These analyses were performed using the SciPy [21] and Scikit-learn [22] libraries in Python 3.8.

III. RESULTS

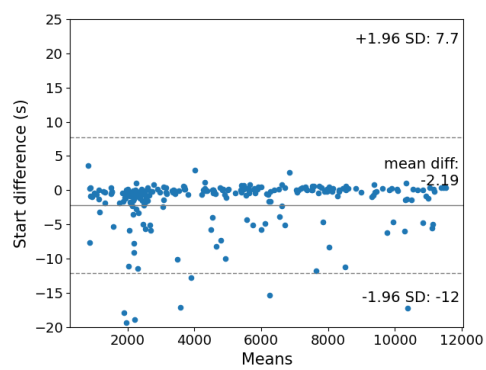
A. Standing events: validation of the algorithm

Table III reports the performance of the FSM in the ADAPT dataset. The algorithm successfully identified standing events that occurred between gait and sitting transfers and excluded sitting and lying events, resulting in a high specificity for standing events (0.99). However, the sensitivity of the algorithm was low, identifying about half of the video-labeled standing events (sensitivity of 0.48 in-lab, 0.39 out-lab).

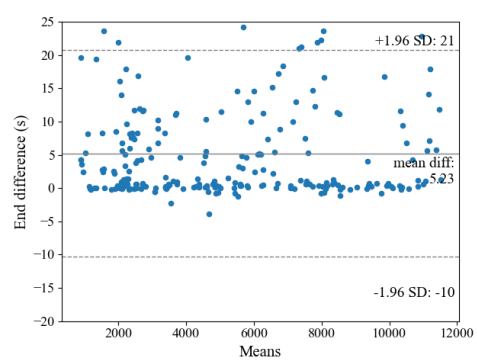
TABLE III
STANDING ALGORITHM VALIDATION IN THE ADAPT DATASET

	in-lab	out-lab
Sensitivity	0.48	0.35
Specificity	0.99	0.99
Accuracy	0.97	0.97
F1-score	0.77	0.71

From the Bland-Altman plots, one can see that the FSM detects the start time of a standing event with a mean delay of 2.04 s and marks the end of the standing period on average in advance by 5.06 s (Fig. 2).



(a)



(b)

Fig. 2. Difference between gold standard and FSM algorithm in the identification of the start time (a) and the end (b) of standing events

This article has been accepted for publication in IEEE Transactions on Biomedical Engineering. 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/TBME.2023.3326306

B. Fall risk associations with instrumented laboratory balance assessment

Table IV reports the summary statistics for the balance features computed for the participants of the InCHIANTI study during the laboratory testing. Range (max-min) was calculated for both the anteroposterior acceleration (AP Acc Range) and the mediolateral acceleration (ML Acc Range).

Univariate analysis (Fig. 3 and Table BI-Appendix) showed that Jerk (OR: 1.24, 95% CI: 1.028-1.495) and LDE (OR: 1.18, 95% CI: 1.021-1.363) were significant features associated with 12-month prospective falls ($p < 0.05$).

TABLE IV

BALANCE FEATURES (MEDIAN AND INTERQUARTILE RANGE, IQR) DURING LABORATORY ASSESSMENT AND SIGNIFICANT DIFFERENCES (U MANN-WHITNEY) IN THE INCHIANTI STUDY

Feature	Non-Faller median (IQR)	Faller median (IQR)	p-values	adjusted-*p* (Benjamini-Hochberg)
AP Acc Range	0.332 (0.145)	0.358 (0.19)	0.221	0.287
ML Acc Range	0.234 (0.093)	0.249 (0.089)	0.244	0.289
Distance	0.042 (0.013)	0.046 (0.015)	0.078	0.180
mean Frequency	0.159 (0.023)	0.16 (0.028)	0.427	0.463
Path	128.27 (40.596)	140.542 (46.714)	0.083	0.180
Area	0.02 (0.016)	0.022 (0.015)	0.180	0.880
Ellipse area	0.071 (0.041)	0.082 (0.071)	0.080	0.180
Jerk	-14.985 (0.444)	-14.762 (0.635)	**0.032**	0.180
Total power	20.042 (2.701)	18.887 (4.549)	0.107	0.199
95% Freq	44.214 (1.233)	43.799 (2.173)	0.141	0.223
Spectral centroid	0.334 (0.045)	0.313 (0.071)	0.074	0.180
SampleE	1.307 (0.237)	1.227 (0.404)	0.155	0.223
LDE	0.006 (0.002)	0.007 (0.003)	**0.035**	0.180

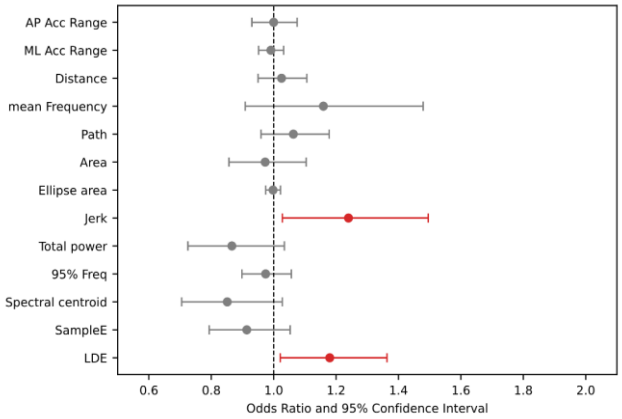
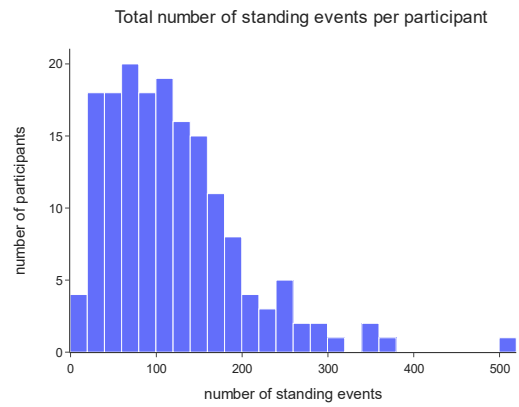


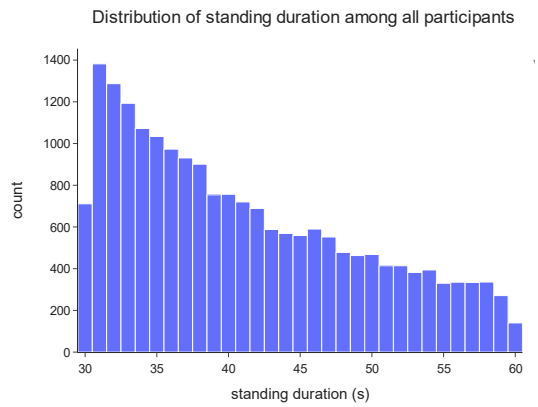
Fig. 3. Forest plot of univariate analyses (z-scored), laboratory balance features' associations with 12-month prospective falls.

C. Fall risk associations with real-world balance assessments

Participants were monitored from 5 to 9 days (6.4±1.2 days). A total of 20,021 standing events with a duration of at least 30 seconds were identified by the FSM for all the participants of the InCHIANTI study throughout the monitoring period. The duration of the identified events ranged from 30 to 60 seconds (Fig. 4).



(a)



(b)

Fig. 4. Total number of standing events per participant identified through the monitoring period (a). Distribution of standing duration across all participants (b).

Table V shows the computed balance features of the participants of the InCHIANTI study (median and interquartile range) during real-world conditions.

Univariate analysis (Fig. 5 and Table BII-Appendix) showed that standing duration (OR: 1.422, 95% CI: 1.086-1.862), distance (OR: 0.699, 95% CI: 0.517-0.945), and mean frequency (OR: 1.546, 95% CI: 1.166-2.048) were significantly associated with 12-month prospective falls ($p<0.05$), indicating their potential importance in real-world balance assessments.

This article has been accepted for publication in IEEE Transactions on Biomedical Engineering. 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/TBME.2023.3326306

TABLE V

LAB VS. REAL-WORLD BALANCE FEATURES (MEDIAN AND INTERQUARTILE RANGE, IQR), STATISTICAL DIFFERENCE (WILCOXON SIGNED-RANK TEST)

Feature	Non-Faller median (IQR)	Faller median (IQR)	*p*-values	adjusted-*p* (Benjamini-Hochberg)
Stands/h	1.302 (1.145)	1.039 (0.823)	0.2479	0.6197
Duration	40.88 (1.521)	41.581 (0.976)	**0.0049**	**0.0365**
AP Acc Range	0.666 (0.109)	0.678 (0.073)	0.7512	0.9390
ML Acc Range	0.706 (0.164)	0.676 (0.185)	0.1250	0.3749
Distance	0.069 (0.016)	0.063 (0.014)	**0.0152**	0.0759
mean Frequency	0.163 (0.015)	0.175 (0.019)	**0.0026**	**0.0365**
Path	204.888 (45.763)	203.335 (29.82)	0.3437	0.7364
Area	0.039 (0.021)	0.034 (0.018)	0.0585	0.2194
Ellipse area	0.244 (0.12)	0.243 (0.102)	0.4078	0.7646
Jerk	-14.283 (0.372)	-14.284 (0.51)	0.8565	0.9542
Total power	14.429 (1.64)	14.023 (2.379)	0.7415	0.9390
95% Freq	41.07 (1.755)	41.093 (2.364)	0.7159	0.9390
Spectral centroid	0.252 (0.03)	0.253 (0.045)	0.9542	0.9542
SampleE	0.72 (0.14)	0.716 (0.213)	0.6970	0.9390
LDE	0.008 (0.001)	0.008 (0.001)	0.9542	0.9542

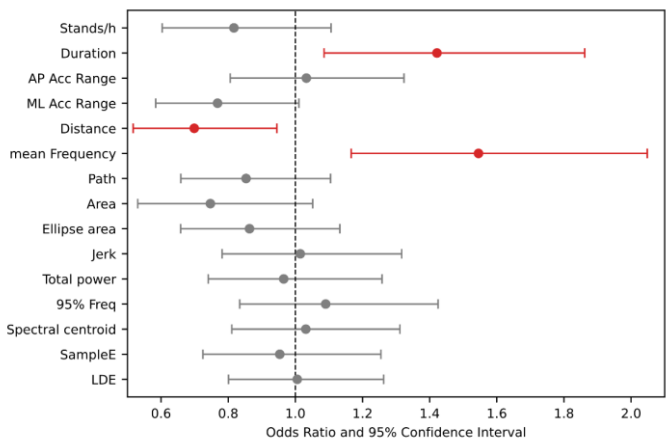


Fig. 5. Forest plot of univariate analysis (z-scored), real-world balance features' associations with 12-month prospective falls.

D. Contextual differences: lab vs. real-world assessments

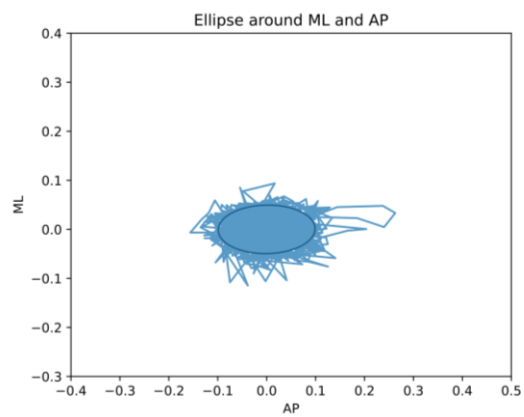
Fig. 6 illustrates representative traces (and confidence ellipse areas) for a lab assessment and a real-world measurement during a standing event.

The size and jerkiness of the accelerometer traces are larger during real-world assessments. Measurements were significantly different under the Wilcoxon signed rank test ($p<0.05$) between the two settings, with spatiotemporal features being significantly larger under real-world conditions (Table VI). Antero-posterior and mediolateral ranges and ellipse areas obtained in real-world settings showed an average increase of over 50% with respect to laboratory measurements (Fig. 7).

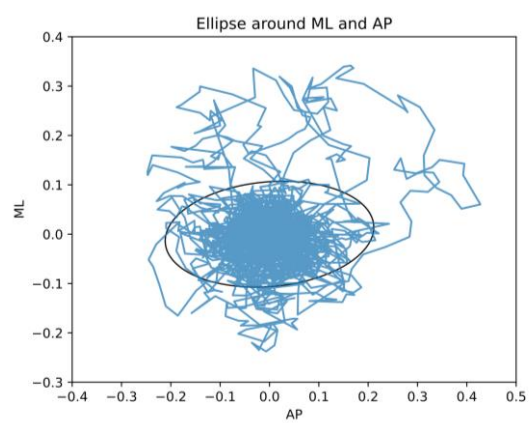
TABLE VI

LAB VS. REAL-WORLD BALANCE FEATURES (MEDIAN AND INTERQUARTILE RANGE, IQR), STATISTICAL DIFFERENCE (WILCOXON SIGNED-RANK TEST)

Features	Lab assessment median (IQR)	Real-world median (IQR)	*p*-value Wilcoxon
AP Acc Range	0.334 (0.15)	0.668 (0.103)	2.76E-24
ML Acc Range	0.236 (0.088)	0.704 (0.164)	3.82E-24
Distance	0.043 (0.014)	0.068 (0.017)	3.39E-22
mean Frequency	0.159 (0.024)	0.165 (0.017)	3.76E-07
Path	131.643 (40.604)	204.888 (42.86)	2.16E-22
Area	0.021 (0.016)	0.037 (0.021)	3.64E-26
Ellipse area	0.072 (0.042)	0.244 (0.118)	1.18E-24
Jerk	-14.959 (0.524)	-14.283 (0.389)	4.62E-22
Total power	19.861 (3.081)	14.404 (1.766)	3.90E-26
95% Freq	44.177 (1.416)	41.078 (1.778)	9.83E-22
Spectral centroid	0.333 (0.051)	0.253 (0.031)	1.04E-23
SampleE	1.301 (0.266)	0.72 (0.159)	1.31E-25
LDE	0.006 (0.002)	0.008 (0.001)	1.03E-15



(a)



(b)

Fig. 6. Ellipse area (90% confidence) representative traces for a 30 s lab (a) and a 30s real-world assessment (b)

This article has been accepted for publication in IEEE Transactions on Biomedical Engineering. 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/TBME.2023.3326306

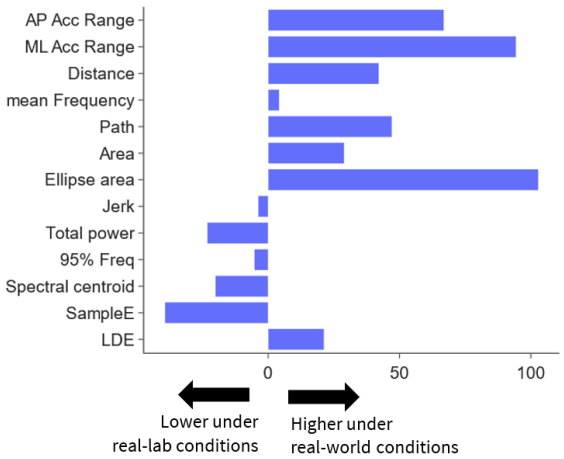


Fig. 7. Percentage change for features measured under real-world conditions compared with in-lab conditions.

We built four machine-learning models to assess the prognostic ability of laboratory and real-world balance assessments for prospective fallers. On the training dataset, recursive feature selection identified the top five most relevant features for both laboratory (distance, ellipse area, jerk, spectral centroid, LDE) and real-world (duration, distance, mean frequency, area, path) assessments, which were subsequently used on the test dataset. Prognostic models based on real-world balance features outperformed those based on laboratory assessments. The AUC of models built on lab evaluations was in the range 0.35-0.56, whereas, for real-world assessments, it was in the range 0.6-0.76 (Fig. 8). Furthermore, adopting a resampling strategy for training appeared to improve the performance of the models.

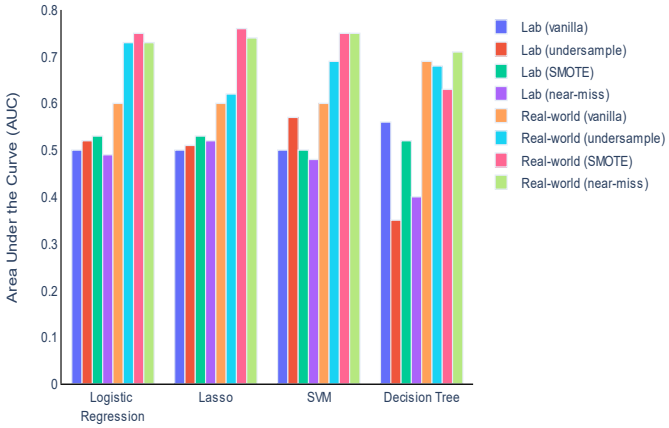


Fig. 8. AUC for each model and setting.

Given the imbalanced nature of the dataset, additional metrics were computed to provide a complete picture of the performance of the models. Table VII provides performance metrics for the overall best models (based on AUC, F1-score, and geometric mean) for laboratory and real-world settings. Applying SMOTE to Lasso Regression in real-world balance features showed the best performance compared to other models, with a sensitivity of 0.67 and a specificity of 0.79.

TABLE VII

PERFORMANCE METRICS FOR SELECTED MODELS USED IN THE TWO CONDITIONS: (L) LABORATORY, (R) REAL-WORLD

Model	AUC	Sensitivity	Specificity	Accuracy (balanced)	F1-score (macro)	Geometric mean
(L) Vanilla Decision Tree	0.56	0.25	0.88	0.57	0.57	0.52
(L) Undersampling SVM	0.57	0.13	0.86	0.49	0.49	0.42
(L) SMOTE Decision Tree	0.52	0.25	0.79	0.52	0.51	0.49
(L) Near miss Logistic Regression	0.49	0.38	0.67	0.52	0.5	0.51
(R) Vanilla Decision Tree	0.69	0.56	0.84	0.70	0.67	0.69
(R) Undersampling Decision Tree	0.69	0.66	0.7	0.68	0.61	0.68
(R) SMOTE Lasso	**0.76**	0.67	0.79	0.73	0.68	0.73
(R) Near miss SVM	0.75	0.88	0.58	0.74	0.60	0.73

IV. DISCUSSION

Our study aimed to investigate the use of inertial sensors for balance assessment in older adults, both in laboratory and real-world settings. The algorithm for identifying standing events based on the proposed finite-state machine (FSM) logic is essential to establishing the validity and reliability of real-world balance assessments. As a first objective, our FSM was validated against video annotations. While the sensitivity of the algorithm was relatively low, it is essential to highlight the context of our real-world monitoring approach. In real-world conditions, the monitoring spanned over seven days, allowing us to record many more standing episodes than standard lab-based assessments. The extensive monitoring period enabled the identification of an average of over 100 standing events per participant. This contrasts with laboratory circumstances, where the number of measured experiments is frequently limited. The low sensitivity of the FSM could be attributed to the cumulative uncertainty of the gait, turn, and sitting algorithms used to identify true standing events. These algorithms prioritize specificity over sensitivity, which impacts the performance of the FSM. It is important to note that this trade-off between specificity and sensitivity was intentional and necessary for our purposes. In real-world scenarios where our algorithm was (and will be) deployed, it is more important to avoid false positives than false negatives to avert extracting balance features from non-standing events [8], [9]. Another limitation of our study is that our sample size was relatively small and may not represent the variability in the general population of older adults. Addressing these limitations may aid in increasing the identification of real-world standing occurrences, hence enhancing the accuracy of real-world fall prediction models. Furthermore, external validation in more extensive and different cohorts is a critical task for the future clinical validation of this tool [23], [24]. Our proposed tool could also be confronted with further machine-learning and deep-learning techniques, provided enough data is available for implementation and validation.

As part of our second objective, when applying the standing algorithm to real-world recordings, we found that the balance features obtained in real-world settings significantly differed from those obtained in laboratory settings. Our results are aligned with findings in other mobility domains, such as gait [9], [25]. In laboratory assessments, the environment is typically standardized (e.g., rigid standardized surfaces, no arm

This article has been accepted for publication in IEEE Transactions on Biomedical Engineering. 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/TBME.2023.3326306

movement, no noise or other disturbing input, etc.). In contrast, in real-world situations, the environment is far more complex and somewhat unpredictable (e.g., irregular surfaces, standing while talking, arm or head movement, and different interactions with the environment). Studies examining walking through crowded hallways and a city center showed that real-world environments might cause significant heterogeneity and asymmetry in mobility patterns [25].

Later, we examined the laboratory and real-world balance features in terms of their prognostic ability to identify individuals at risk of falling In the lab, jerk and LDE were significantly associated with prospective falls; fallers had higher jerk and LDE, indicating more chaotic or unpredictable postural sway patterns. In the real world, longer duration, lower distance, and higher mean frequency became predictive of falls. Longer duration may indicate a possible increase in slowness and/or increased difficulty in switching between daily life activities. This result is consistent with what was shown in [26], where prospective fallers turned less frequently and took longer to turn. On the other hand, shorter distance and greater mean frequency may indicate a reduced ability to make larger postural adjustments when needed. This may be problematic in instances requiring sudden balance corrections, such as in real-world settings. In addition, prognostic models based on real-world balance features surpassed those based on laboratory assessments. Also, combining the SMOTE sampling technique with a Lasso Regression can manage imbalanced data and reduce irrelevant characteristics' influence on classification accuracy, resulting in a more precise and robust model.

Finally, while previous studies investigated the prognostic value of various mobility domains [4], [8], [20], [26], [27], we place a deliberate emphasis on highlighting the predictive potential of balance features. We aimed to contribute a focused methodology to the distinctive predictive power of balance-related metrics, thereby complementing the existing body of literature. By concentrating on this specific aspect, we sought to deepen the understanding of fall risk assessment and provide a comprehensive framework for enhancing fall prediction models. Future and ongoing research endeavors will delve into more comprehensive digital mobility biomarker paradigms, thereby enriching the predictive accuracy and encompassing a holistic spectrum of mobility factors in fall risk assessment.

V. Conclusion

To the best of our knowledge, this is the first study to show that real-world balance features differ considerably from laboratory balance assessments (Romberg test) and have a higher predictive capacity in identifying patients at high risk of falling. These findings highlight the need to move beyond traditional laboratory-based balance measures and develop more sensitive and accurate methods for predicting falls [8], [18], [26]. Lab-based assessments may not accurately reflect the demands of daily life and may not capture the full range of balance challenges that older adults encounter. Furthermore, unlike lab assessments, real-world assessments allow for the identification of multiple events, providing additional insights into the participants' exposure and fitness. In our study, due to the characteristics of real-world monitoring, two additional features were introduced: the number of stands per hour and the duration of the standing events. The latter was found to be significantly associated with prospective falls. Further research is needed to confirm the study's findings in an external, larger, and more diverse sample of older adults and to explore the potential of real-world balance assessments for predicting falls in other populations, such as individuals with neurological diseases or mobility impairments.

Appendix

A. Finite-State Machine

Algorithm for identifying standing events. Periods without an identified state were labeled as “unknown” events.

Algorithm 1 **Finite state-machine (FSM) for standing detection**

```
Low pass filter acceleration with fc = 5 Hz (filt_acc)
stationary = (|filt_acc| < 0.05) // stationary period
gait = Adamowicz_algorithm [18]
sit-to-stand, stand-to-sit = Pham_algorithm [19]
turns = El-Gohary_algorithm [20]
ap = gait or turns or sit-to-stand or stand-to-sit // active periods
remove any ap from stationary periods
state = build states ap and stationary events
for each state k:
  if (stationary): // look at surrounding activity
    if (state[k-1]  is “unknown” and length(state[k-1]) < 0.5s):
      state[k-1] = state[k-2]
    end if
    if (state[k+1]  is “unknown” and length(state[k+1]) < 0.5s):
      state[k+1] = state[k+2]
    end if
    if (state[k-1] or state[k+1] is “walk”):
      state[k] = "stand"
    else if (state[k-1] is “sit-to stand” or state[k+1] is “stand-to sit”):
      state[k] = "stand"
    end if
  end if
end for
```

This article has been accepted for publication in IEEE Transactions on Biomedical Engineering. 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/TBME.2023.3326306

B. Balance and Prospective falls associations

TABLE BI
ASSOCIATIONS BETWEEN BALANCE AND PROSPECTIVE FALLS IN THE LABORATORY ASSESSMENT (InCHIANTI STUDY)

	95% Confidence Interval		Odds Ratio	p-values
	2.50%	**97.50%**	OR	
AP Acc Range	0.930	1.075	1.000	0.998
ML Acc Range	0.952	1.032	0.991	0.661
Distance	0.950	1.106	1.025	0.522
mean Frequency	0.909	1.479	1.159	0.234
Path	0.959	1.178	1.063	0.243
Area	0.857	1.104	0.973	0.669
Ellipse area	0.975	1.022	0.998	0.863
Jerk	1.028	1.495	1.240	**0.025**
Total power	0.725	1.034	0.866	0.112
95% Freq	0.898	1.057	0.974	0.529
Spectral centroid	0.705	1.028	0.851	0.094
SampleE	0.793	1.053	0.914	0.212
LDE	1.021	1.363	1.180	**0.025**

TABLE BII
ASSOCIATIONS BETWEEN BALANCE AND PROSPECTIVE FALLS IN THE REAL-WORLD ASSESSMENT (InCHIANTI STUDY)

	95% Confidence Interval		Odds Ratio	p-values
	2.50%	**97.50%**	OR	
Stands/h	0.603	1.106	0.817	0.191
Duration	1.086	1.862	1.422	**0.011**
AP Acc Range	0.806	1.324	1.033	0.799
ML Acc Range	0.584	1.011	0.768	0.059
Distance	0.517	0.945	0.699	**0.020**
mean Frequency	1.166	2.048	1.546	**0.002**
Path	0.659	1.105	0.853	0.229
Area	0.531	1.052	0.747	0.095
Ellipse area	0.658	1.133	0.864	0.289
Jerk	0.782	1.317	1.015	0.913
Total power	0.741	1.258	0.965	0.794
95% Freq	0.834	1.425	1.090	0.527
Spectral centroid	0.811	1.312	1.031	0.802
SampleE	0.725	1.255	0.954	0.735
LDE	0.801	1.263	1.006	0.961

ACKNOWLEDGMENT

We want to acknowledge the participants who volunteered to participate in the clinical studies. Their contribution has been vital to the successful completion of the present study.

CONFLICT OF INTEREST

S.M., L.P. and L.C. are co-founders and own shares of mHealth Technologies (https://mhealthtechnologies.it/). P.P. holds copyrights on codes for fall risk assessment in older people. All other authors declare no competing interest.

REFERENCES

[1] WHO, *Strategies for preventing and managing falls across the life-course*. 2021.

[2] A. Zijlstra, M. Mancini, L. Chiari, and W. Zijlstra, "Biofeedback for training balance and mobility tasks in older populations: a systematic review," *J. Neuroeng. Rehabil.*, vol. 7, no. 1, p. 58, Dec. 2010, doi: 10.1186/1743-0003-7-58.

[3] B. R. Greene, K. McManus, L. G. M. Ader, and B. Caulfield, "Unsupervised Assessment of Balance and Falls Risk Using a Smartphone and Machine Learning," *Sensors*, vol. 21, no. 14, p. 4770, Jul. 2021, doi: 10.3390/s21144770.

[4] P. Palumbo, L. Palmerini, S. Bandinelli, and L. Chiari, "Fall Risk Assessment Tools for Elderly Living in the Community: Can We Do Better?," *PLoS One*, vol. 10, no. 12, p. e0146247, Dec. 2015, doi: 10.1371/journal.pone.0146247.

[5] F. B. Horak, "Clinical assessment of balance disorders," *Gait Posture*, vol. 6, no. 1, pp. 76–84, Aug. 1997, doi: 10.1016/S0966-6362(97)00018-0.

[6] F. B. Horak, D. M. Wrisley, and J. Frank, "The Balance Evaluation Systems Test (BESTest) to Differentiate Balance Deficits," *Phys. Ther.*, vol. 89, no. 5, pp. 484–498, May 2009, doi: 10.2522/ptj.20080071.

[7] L. Chiari, "Stabilometry," in *Encyclopedia of Neuroscience*, M.D. Binder, N. Hirokawa, and U. Windhorst, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 3830–3833. doi: 10.1007/978-3-540-29678-2_5623.

[8] M. Nouredanesh, A. Godfrey, J. Howcroft, E. D. Lemaire, and J. Tung, "Fall risk assessment in the wild: A critical examination of wearable sensor use in free-living conditions," *Gait Posture*, vol. 85, no. May 2020, pp. 178–190, 2021, doi: 10.1016/j.gaitpost.2020.04.010.

[9] E. Warmerdam *et al.*, "Personal View Long-term unsupervised mobility assessment in movement disorders," *Lancet Neurol.*, vol. 19, no. 5, pp. 462–470, 2020, doi: 10.1016/S1474-4422(19)30397-7.

[10] A. K. Bourke, E. A. F. Ihlen, R. Bergquist, P. B. Wik, B. Vereijken, and J. L. Helbostad, "A physical activity reference data-set recorded from older adults using body-worn inertial sensors and video technology—The ADAPT study data-set," *Sensors (Switzerland)*, vol. 17, no. 3, 2017, doi: 10.3390/s17030559.

[11] L. Ferrucci *et al.*, "Subsystems Contributing to the Decline in Ability to Walk: Bridging the Gap Between Epidemiology and Geriatric Practice in the InCHIANTI Study," *J. Am. Geriatr. Soc.*, vol. 48, no. 12, pp. 1618–1625, Dec. 2000, doi: 10.1111/j.1532-

5415.2000.tb03873.x.

[12] Y. Agrawal, J. P. Carey, H. J. Hoffman, D. A. Sklare, and M. C. Schubert, "The Modified Romberg Balance Test," *Otol. Neurotol.*, vol. 32, no. 8, pp. 1309–1311, Oct. 2011, doi: 10.1097/MAO.0b013e31822e5bee.

[13] J. Klenk *et al.*, "The FARSEEING real-world fall repository: a large-scale collaborative database to collect and share sensor signals from real-world falls," *Eur. Rev. Aging Phys. Act.*, vol. 13, no. 1, pp. 1–7, 2016, doi: 10.1186/s11556-016-0168-9.

[14] L. Adamowicz, Y. Christakis, M. D. Czech, and T. Adamusiak, "SciKit Digital Health: Python Package for Streamlined Wearable Inertial Sensor Data Processing," *JMIR mHealth uHealth*, vol. 10, no. 4, p. e36762, Apr. 2022, doi: 10.2196/36762.

[15] M. H. Pham, "Validation of a Lower Back ' Wearable ' -Based Sit-to-Stand and Stand-to-Sit Algorithm for Patients With Parkinson ' s Disease and Older Adults in a Home-Like Environment," vol. 9, no. August, pp. 1–11, 2018, doi: 10.3389/fneur.2018.00652.

[16] M. El-Gohary *et al.*, "Continuous Monitoring of Turning in Patients with Movement Disability," *Sensors*, vol. 14, no. 1, pp. 356–369, Dec. 2013, doi: 10.3390/s140100356.

[17] R. A. W. Felius, M. Geerars, S. M. Bruijn, N. C. Wouda, J. H. Van Dieën, and M. Punt, "Reliability of IMU-based balance assessment in clinical stroke rehabilitation," *Gait Posture*, vol. 98, pp. 62–68, Oct. 2022, doi: 10.1016/j.gaitpost.2022.08.005.

[18] K. McManus, B. R. Greene, L. G. M. Ader, and B. Caulfield, "Development of Data-Driven Metrics for Balance Impairment and Fall Risk Assessment in Older Adults," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 7, pp. 2324–2332, Jul. 2022, doi: 10.1109/TBME.2022.3142617.

[19] K. S. van Schooten, M. Pijnappels, S. M. Rispens, P. J. M. Elders, P. Lips, and J. H. van Dieën, "Ambulatory Fall-Risk Assessment: Amount and Quality of Daily-Life Gait Predict Falls in Older Adults," *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.*, vol. 70, no. 5, pp. 608–615, May 2015, doi: 10.1093/gerona/glu225.

[20] L. Montesinos, R. Castaldo, and L. Pecchia, "Wearable inertial sensors for fall risk assessment and prediction in older adults: A systematic review and meta-analysis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 3, pp. 573–582, 2018, doi: 10.1109/TNSRE.2017.2771383.

[21] P. Virtanen *et al.*, "SciPy 1.0: fundamental algorithms for scientific computing in Python," *Nat. Methods*, vol. 17, no. 3, pp. 261–272, Mar. 2020, doi: 10.1038/s41592-019-0686-2.

[22] O. Kramer, "Scikit-Learn," 2016, pp. 45–53. doi: 10.1007/978-3-319-33383-0_5.

[23] F. K. Id *et al.*, "PLOS ONE Consensus based framework for digital mobility monitoring," pp. 1–14, 2021, doi: 10.1371/journal.pone.0256541.

[24] C. Mazzà *et al.*, "Technical validation of real- - world monitoring of gait : a multicentric observational study," no. 12246987, pp. 1–14, 2021, doi: 10.1136/bmjopen-2021-050785.

[25] P. Tamburini, F. Storm, C. Buckley, M. C. Bisi, R. Stagni, and C. Mazzà, "Moving from laboratory to real life conditions: Influence on the assessment of variability and stability of gait," *Gait Posture*, vol. 59, pp. 248–252, Jan. 2018, doi: 10.1016/j.gaitpost.2017.10.024.

[26] J. M. Leach, S. Mellone, P. Palumbo, S. Bandinelli, and L. Chiari, "Natural turn measures predict recurrent falls in community-dwelling older adults: a longitudinal cohort study," *Sci. Rep.*, vol. 8, no. 1, p. 4316, Dec. 2018, doi: 10.1038/s41598-018-22492-6.

[27] A. Weiss, T. Herman, N. Giladi, and J. M. Hausdorff, "Objective Assessment of Fall Risk in Parkinson's Disease Using a Body-Fixed Sensor Worn for 3 Days," *PLoS One*, vol. 9, no. 5, p. e96675, May 2014, doi: 10.1371/journal.pone.0096675.

[28] A. Melendez-Calderon, C. Shirota, and S. Balasubramanian, "Estimating Movement Smoothness From Inertial Measurement Units," Front. Bioeng. Biotechnol., vol. 8, Jan. 2021, doi: 10.3389/fbioe.2020.558771.

[29] S. M. Bruijn, D. J. J. Bregman, O. G. Meijer, P. J. Beek, and J. H. van Dieën, "Maximum Lyapunov exponents as predictors of global gait stability: A modelling approach," Med. Eng. Phys., vol. 34, no. 4, pp. 428–436, May 2012, doi: