Smartphone-Based Balance Assessment Using Machine Learning

Marjan Nassajpour[1], Mustafa Shuqair[1], Amie Rosenfeld, D.P.T.[2], Magdalena I. Tolea, Ph.D.[2],
James E. Galvin, M.D, M.P.H.[2], and Behnaz Ghoraani, Ph.D.[1*]
[1]Department of Electrical Engineering & Computer Science, Florida Atlantic University, Boca Raton, FL, USA
[2]Comprehensive Center for Brain Health, Department of Neurology, University of Miami, Boca Raton, FL

2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) | 979-8-3503-7149-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/EMBC53108.2024.10781862

***Abstract*—This study explores the potential of smartphones to objectively assess balance, which is crucial for the elderly and individuals recovering from various medical conditions. We propose an innovative methodology to estimate the Modified Clinical Test of Sensory Interaction on Balance (m-CTSIB) scores using the accelerometer sensor of a smartphone coupled with machine learning techniques. Our dataset consists of 28 participants, aged 21 to 88 years. Notably, the XGBOOST algorithm demonstrates a strong correlation (0.92) with the ground truth balance scores. These ground truth scores are obtained using a force plate system collected simultaneously with the smartphone data, ensuring precise and reliable comparisons. This methodology offers an objective, accessible, and convenient means for balance assessment, greatly facilitating at-home monitoring and enhancing the potential for remote health monitoring. Our findings underscore the method's reliability and potential impact on telemedicine and patient care, offering notable improvements in the quality of life.**

***Index Terms*—Smartphone-Based Balance Assessment, Machine Learning in Healthcare, Remote Health Monitoring**

I. Introduction

The widespread use of smartphones in healthcare signifies a shift to accessible, remote, and cost-effective clinical assessments. Equipped with various sensors like accelerometers, smartphones facilitate broader healthcare evaluations. This advancement is especially relevant in balance assessment, crucial for the well-being of the elderly and individuals recovering from injuries, surgeries, or neurological conditions like Parkinson's and Alzheimer's [1]–[3].

Balance, commonly associated with stability and the regulation of posture, is crucial in preventing falls, which are a prevalent threat to the elderly [4], [5]. Traditional balance assessment methods, such as the Modified Clinical Test of Sensory Interaction on Balance (m-CTSIB), rely on clinician-administered tests and specialized equipment like force plates, e.g., the Falltrak II (MedTrak VNG, Inc.), to measure balance [6]–[8]. However, these methods suffer from drawbacks like high costs, limited availability, and potential subjectivity.

In this context, the question arises: Can smartphones be utilized to objectively estimate balance, akin to the assessments conducted in a clinical office with specialized equipment and trained staff? Recent studies have begun to explore the reliability and validity of smartphones in measuring balance scores in various tests, including the Timed Up and Go (TUG) test and the Berg Balance Scale (BBS) test. These studies primarily focused on extracting changes in acceleration vectors and total time for assessment [9]–[12].

This study aims to extend this exploration by using a smartphone combined with machine learning algorithms to predict m-CTSIB test scores as measures of deviations of the center of pressure (COP) from the center of mass (COM). By securely positioning the smartphone on subjects' abdomen, we gather accelerometer data and apply machine learning to predict scores. This method could significantly enhance diagnosing and managing balance and cognitive issues in older adults, improving their quality of life and independence with remote assessments, cutting clinic visits.

II. Study design

A. Dataset

The study comprised 28 participants aged between 21 and 88. Of the total participants, 9 were male, while 19 were female. The research protocol was approved by an Institutional Review Board (IRB) and adhered to the Helsinki Declaration guidelines. All participants provided informed consent through signed consent forms, ensuring ethical compliance and participant awareness. A comprehensive assessment of m-CTSIB scores was conducted to evaluate the balance control of participants. The m-CTSIB test includes four distinct conditions designed to systematically alter sensory inputs. These conditions assess the integration of visual, somatosensory, and vestibular information:

- **EOSS (Eyes Open, Stable Surface):** Challenges the somatosensory system with feedback from the feet on a stable surface.
- **ECSS (Eyes Closed, Stable Surface):** Tests the somatosensory and visual systems.
- **EOFS (Eyes Open, Foam Surface):** Evaluates visual and vestibular systems with a reduced somatosensory input on an unstable surface.
- **ECFS (Eyes Closed, Foam Surface):** Focuses on the vestibular system, minimizing somatosensory and visual inputs.

To assess m-CTSIB scores, we employed the Falltrak II, a force plate technology designed to measure participants' balance using path length (PL) and average velocity (AV). Stability is evaluated by assessing the deviation of the center of pressure (COP) from the center of mass (COM) during the test. Path length (PL) represents the extent of COP movement,

This study was supported by the National Science Foundation (IIS 1942669) and Ed and Ethel Moore Alzheimer's Disease Research Program at the Florida Department of Health (AWD-001693).

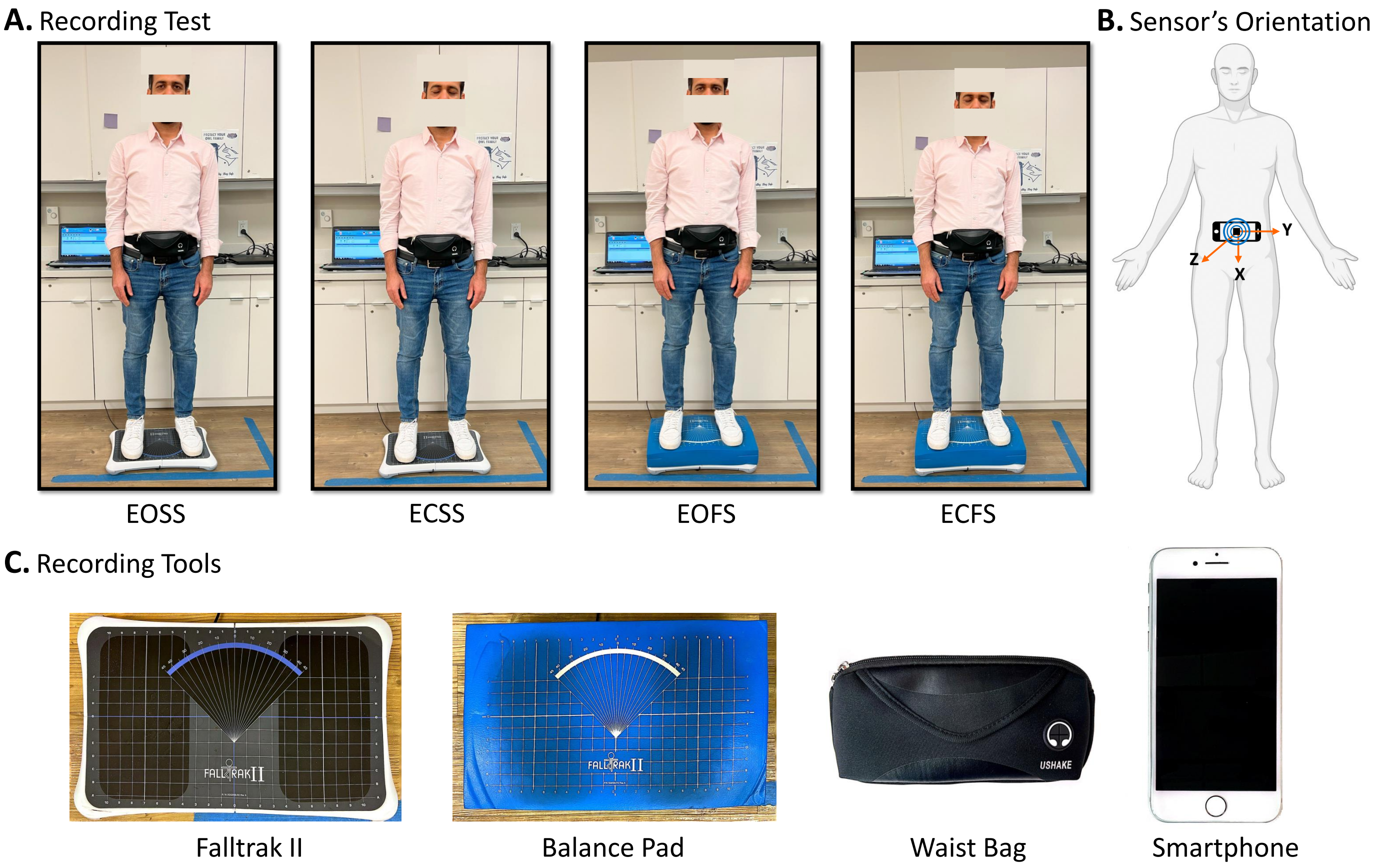

Fig. 1. Experimental setup and data collection. textbf(A) Illustrates the recording process for the four conditions of the Modified Clinical Test of Sensory Interaction on Balance (m-CTSIB). **(B)** Shows the sensor orientation of the smartphone placed on the subject's abdomen during the experiment. **(C)** Depicts the tools used for data recording.

with a shorter path length indicating improved balance performance. Average velocity (AV) measures the speed of COP movement during the test. Fig 1 illustrates the experimental setup and the four conditions.

During the test, participants stood on the Falltrak II platform. At the same time, an iPhone 7 (Model A1778) was positioned at their umbilical region to record three-dimensional accelerometer data (X, Y, and Z axes) at a 100 Hz sampling rate. The phone was securely positioned in a waist bag of matching size to minimize rotation and movement. The 'Just Record' application facilitated access to the accelerometer data from the phone. For each of the four conditions, data were recorded separately by utilizing the stop and start buttons within the app. Fig 1.B illustrates the sensor's orientation, where the X-axis corresponds to vertical (VT) movement, the Y-axis to medial-lateral (ML) movement, and the Z-axis to anterior-posterior (AP) movement.

We observed significant correlations between AV and PL scores in all test conditions, with a correlation of 0.94 for EOSS and a perfect 1.00 for ECSS, EOFS, and ECFS, suggesting that AV changes are consistently mirrored by corresponding alterations in PL and vice versa. Hence, our machine learning model was designed to estimate AV scores exclusively as AV is integral for evaluating balance and stability, where higher AV scores generally indicate greater instability.

B. Data Pre-processing and Feature Extraction

In preprocessing, we removed a 0.5-second interval from the start and end of the smartphone sensor data to reduce potential artifacts during subjects' transitions between conditions. For each subject under various conditions, 42 features were extracted from the smartphone accelerometer in the X, Y, and Z dimensions, resulting in 168 features. Table I details the extracted features. Feature selection was omitted, as the machine learning algorithms (XGBOOST, Decision tree, and Random Forest) that were employed inherently prioritize relevant features during training. This leverages the models' built-in feature selection, streamlining the analysis pipeline and ensuring a robust modeling process without needing a separate feature selection stage.

C. Machine Learning Techniques

In this study, we employed three machine learning algorithms—XGBOOST, Decision tree, and Random forest—to model the relationship between the input, consisting of extracted features, and the output represented by AV balance scores. The methodology involved leave-one-out cross-

TABLE I
THE EXTRACTED FEATURES FOR EACH SUBJECT

Feature Name	Signals	# Features
1- Standard deviation	X,Y,Z	12
2- Shannon Entropy	X,Y,Z	12
3- Sample Entropy	X,Y,Z	12
4- Skewness	X,Y,Z	12
5- kurtosis	X,Y,Z	12
6- Frequency-domain entropy	X,Y,Z	12
7- Energy of the main frequency	X,Y,Z	12
8- Main frequency	X,Y,Z	12
9- Energy of the secondary frequency	X,Y,Z	12
10- Secondary frequency	X,Y,Z	12
11- Sparsity	X,Y,Z	12
12- Difference sum	X,Y,Z	12
13- Average Jerk	X,Y,Z	12
14- Cross Correlation XY	X,Y	4
15- Cross Correlation XZ	X,Z	4
16- Cross Correlation YZ	Y,Z	4
Total Number of Features		168

validation, randomization, and partitioning of the remaining data into an 80% training set and a 20% validation set. Subsequently, all training, validation, and test data underwent normalization using the mean and standard deviation of the training data. The hyperparameter tuning was conducted through a grid search method tailored to each algorithm's unique parameters: for XGBOOST, the number of trees, decision tree depth, and feature sampling rate; for Random Forest, the number of trees, decision tree depth, minimum samples required to split an internal node, and minimum samples required to be at a leaf node; and for the Decision Tree, the minimum samples required to split an internal node, minimum samples required to be at a leaf node, and the number of features considered when seeking the best split. Optimal models for all algorithms were selected based on criteria encompassing correlation (r) and minimum absolute error (MAE).

D. Model Interpretation

To assess the impact of features on our model predictions, we employed SHAP (SHapley Additive exPlanations) summary plots, an advanced visualization tool in machine learning. These plots use Shapley values from cooperative game theory to attribute each feature's contribution fairly. Shapley values represent the average marginal contribution of a feature across all possible feature combinations. In the SHAP summary plot, features are ranked and visualized based on their absolute Shapley values, highlighting their contribution to pushing the model's prediction higher or lower.

This method is potent because it considers interaction effects between features, offering a nuanced view compared to traditional feature importance methods. SHAP plots provide a transparent understanding of model behavior by quantifying the exact impact of each feature on the output. This approach is crucial for validating the trustworthiness and reliability of our machine learning models, ensuring that the most significant contributors to predictions are both clinically relevant and interpretable [13].

III. RESULTS AND DISCUSSION

We trained and evaluated three machine learning models—Decision Tree, Random Forest, and XGBOOST—using features from smartphone accelerometers for AV score estimation. Model performance was optimized by maximizing correlation (r) and minimizing MAE.

Table II presents the top-performing results of these three machine learning models across all training, validation, and test datasets. Regarding training, XGBOOST achieved a high r (1) and low MAE (0.06), outperforming Decision Tree and Random Forest. However, when considering the validation set, Random Forest emerged as the most effective model, achieving a perfect r (0.96) and the lowest MAE (0.12). In the testing phase, XGBOOST exhibited strong performance with a r of 0.92 and the lowest MAE of 0.30, surpassing Decision Tree and Random Forest. Overall, XGBOOST demonstrated superior performance across all three datasets, making it the most effective model for estimating AV scores in this study.

To assess the influence of individual features on predicting AV balance scores, we conducted a comprehensive analysis using the optimal XGBOOST model and performed a SHAP value analysis. Fig. 2 showcases the top 10 features significantly influencing m-CTSIB balance score estimation according to their SHAP values. In the graph, concordant changes in the feature values and SHAP values for a given feature signify a direct relationship with the model prediction. Conversely, when the changes in feature values and SHAP values exhibit an inverse direction, we characterize the feature as having an opposite relationship with the model prediction. A key finding from this analysis is the prominence of standard deviation (STD) in all three directions: AP, ML, and VT. The data reveals a direct correlation between larger STD values and higher SHAP values. This suggests that subjects displaying more pronounced sway during the m-CTSIB balance test—indicated by greater STD values—lead to more accurate predictions by the model. This observation is consistent with the premise that increased body sway or motion, especially in these three directions, may indicate compromised postural stability, reflecting a less controlled response during the balance test. On the other hand, sample entropy, particularly in the AP and ML directions, shows an inverse relationship with model predictions. Lower values in sample entropy correspond to higher SHAP values. This indicates that subjects with lower variability or unpredictability in their movement patterns, as measured by sample entropy

TABLE II
ESTIMATION RESULTS FOR M-CTSIB SCORES

Methods	Train		Validation		Test	
	r	MAE	r	MAE	r	MAE
Decision Tree	0.96	0.10	0.92	0.16	0.76	0.39
Random Forest	0.98	0.06	0.96	0.12	0.82	0.30
XGBOOST	**1**	**0.03**	0.91	0.17	**0.92**	**0.30**

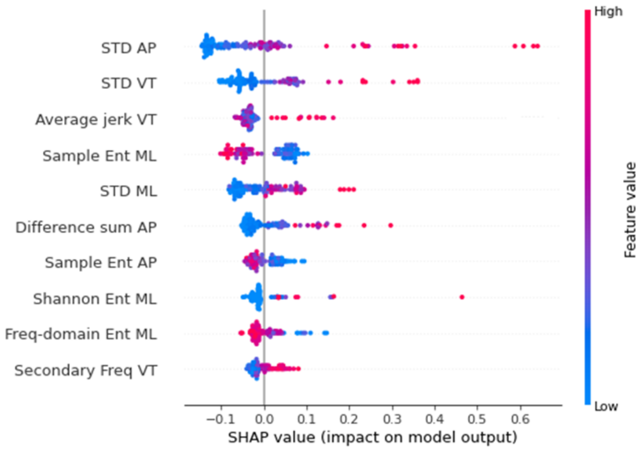

Fig. 2. Visualization of SHAP values extracted from the top-performing XGBOOST model, depicting the 10 features with the highest mean absolute SHAP values. Each point corresponds to an individual subject and its respective feature (row), with the color indicating the feature values. The position along the x-axis reflects the influence of each feature on the model's prediction for a specific subject. Positive SHAP values contribute to higher predictions, while negative values influence lower predictions. Features are arranged on the y-axis in order of importance. The abbreviations are STD (standard deviation), Ent (Entropy), and Freq (Frequency).

in these specific directions, significantly impact the model's predictive accuracy.

Comparatively, our study advances upon previous smartphone-based balance assessments. Almajid *et al.* conducted single and dual-task TUG tests with a smartphone placed on the umbilical region, relying solely on accelerometer signals for start and stop time identification. Hou *et al.* affixed a smartphone to the back of the trunk at the second sacrum spine level, focusing on a singular feature – the combined changes in the acceleration vector from the AP and ML axes as an indicator of postural control. In contrast, our methodology involved extracting 42 diverse features, and we employed machine learning methods to estimate AV balance scores, providing a more nuanced and multifaceted evaluation. Notably, we benchmarked our results against the objective measurements from Falltrak II, offering a quantitative basis for comparison. In contrast, Almajid *et al.* and Hou *et al.* studies resorted to subjective assessments by clinicians. Furthermore, our study boasted a larger participant size of 28, surpassing the sample sizes of Almajid *et al.* (20 subjects) and Hou *et al.* (18 subjects). This strengthens the robustness of our findings and contributes to a more thorough understanding of smartphone-based balance assessment.

IV. Conclusion

This study represents a pioneering effort in employing smartphones for the objective and remote estimation of balance, particularly addressing the growing concern of fall risk and neurological diseases like Alzheimer's among older adults. Smartphones, with their widespread availability and advanced sensor capabilities, offer a unique opportunity to monitor balance in a manner that is both accessible and cost-effective. In our research, we utilized a single iPhone 7, attached to the umbilical region of participants, to estimate the m-CTSIB scores. Our participant cohort comprised 28 individuals who underwent the m-CTSIB test under four sensory conditions. These tests were simultaneously monitored using the Falltrak II system, providing ground truth measurements for comparison. The XGBOOST algorithm correlated strongly with our ground truth scores (0.92). Additionally, our SHAP analysis provided insightful findings, revealing that certain features, particularly the standard deviation in various axes, had significant predictive power for balance impairment. The outcomes of this study highlight the potential of smartphones as a practical solution for remote balance monitoring, paving the way for further innovations in telemedicine and patient care.

References

[1] A. R. Alashram, G. Annino, M. Raju, and E. Padua, "Effects of physical therapy interventions on balance ability in people with traumatic brain injury: A systematic review," *NeuroRehabilitation*, vol. 46, no. 4, pp. 455–466, 2020.

[2] M. F. Corrà, N. Vila-Chã, A. Sardoeira, C. Hansen, A. P. Sousa, I. Reis, *et al.*, "Peripheral neuropathy in parkinson's disease: prevalence and functional impact on gait and balance," *Brain*, vol. 146, no. 1, pp. 225–236, 2023.

[3] M. Seifallahi, A. H. Mehraban, J. E. Galvin, and B. Ghoraani, "Alzheimer's disease detection using comprehensive analysis of timed up and go test via kinect v. 2 camera and machine learning," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 1589–1600, 2022.

[4] A. S. Pollock, B. R. Durward, P. J. Rowe, and J. P. Paul, "What is balance?," *Clinical rehabilitation*, vol. 14, no. 4, pp. 402–406, 2000.

[5] 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, 2015.

[6] L. B. Horn, T. Rice, J. L. Stoskus, K. H. Lambert, E. Dannenbaum, and M. R. Scherer, "Measurement characteristics and clinical utility of the clinical test of sensory interaction on balance (ctsib) and modified ctsib in individuals with vestibular dysfunction," *Archives of physical medicine and rehabilitation*, vol. 96, no. 9, pp. 1747–1748, 2015.

[7] D. L. Freeman, G. Gera, F. B. Horak, M. T. Blackinton, M. Besch, and L. King, "The instrumented test of sensory integration for balance: A validation study," *Journal of geriatric physical therapy (2001)*, vol. 41, no. 2, p. 77, 2018.

[8] P. Suttanon, K. D. Hill, K. J. Dodd, and C. M. Said, "Retest reliability of balance and mobility measurements in people with mild to moderate alzheimer's disease," *International Psychogeriatrics*, vol. 23, no. 7, pp. 1152–1159, 2011.

[9] J. A. Moral-Munoz, B. Esteban-Moreno, E. Herrera-Viedma, M. J. Cobo, and I. J. Pérez, "Smartphone applications to perform body balance assessment: a standardized review," *Journal of Medical Systems*, vol. 42, pp. 1–8, 2018.

[10] B. D. C. Corrêa, E. G. R. Santos, A. Belgamo, G. H. L. Pinto, S. S. Xavier, C. C. Silva, *et al.*, "Smartphone-based evaluation of static balance and mobility in long-lasting covid-19 patients," *Frontiers in Neurology*, vol. 14, 2023.

[11] Y.-R. Hou, Y.-L. Chiu, S.-L. Chiang, H.-Y. Chen, and W.-H. Sung, "Development of a smartphone-based balance assessment system for subjects with stroke," *Sensors*, vol. 20, no. 1, p. 88, 2019.

[12] R. Almajid and R. Goel, "Assessment of dual-tasking during a dynamic balance task using a smartphone app: a pilot study," *Journal of physical therapy science*, vol. 34, no. 2, pp. 115–121, 2022.

[13] G. P. McFall, L. Bohn, M. Gee, S. M. Drouin, H. Fah, W. Han, *et al.*, "Identifying key multi-modal predictors of incipient dementia in parkinson's disease: a machine learning analysis and tree shap interpretation," *Frontiers in Aging Neuroscience*, vol. 15, 2023.