

Northumbria Research Link

Citation: Nouredanesh, Mina, Ojeda, Lauro, Alexander, Neil B., Godfrey, Alan, Schwenk, Michael, Melek, William and Tung, James (2022) Automated Detection of Older Adults' Naturally-Occurring Compensatory Balance Reactions: Translation From Laboratory to Free-Living Conditions. IEEE Journal of Translational Engineering in Health and Medicine, 10. pp. 1-13. ISSN 2168-2372

Published by: IEEE

URL: <https://doi.org/10.1109/JTEHM.2022.3163967>
<<https://doi.org/10.1109/JTEHM.2022.3163967>>

This version was downloaded from Northumbria Research Link:
<http://nrl.northumbria.ac.uk/id/eprint/49134/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)

Received 17 November 2021; revised 26 February 2022; accepted 9 March 2022.
Date of publication 22 April 2022; date of current version 5 May 2022.

Digital Object Identifier 10.1109/JTEHM.2022.3163967

Automated Detection of Older Adults' Naturally-Occurring Compensatory Balance Reactions: Translation From Laboratory to Free-Living Conditions

MINA NOUREDANESH^{1,2}, LAURO OJEDA³, NEIL B. ALEXANDER^{4,5},
ALAN GODFREY⁶, (Senior Member, IEEE), MICHAEL SCHWENK^{7,8},
WILLIAM MELEK¹, (Senior Member, IEEE), AND JAMES TUNG¹

¹Department of Mechanical and Mechatronics Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada

²School of Rehabilitation Science, McMaster University, Hamilton, ON L8S 1C7, Canada

³Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI 48109, USA

⁴Division of Geriatric and Palliative Medicine, Department of Internal Medicine, University of Michigan, Ann Arbor, MI 48109, USA

⁵Geriatric Research Education and Clinical Center, VA Ann Arbor Healthcare System, Ann Arbor, MI 48105, USA

⁶Department of Computer and Information Sciences, Northumbria University, Newcastle upon Tyne NE1 8ST, U.K.

⁷Network Aging Research (NAR), Heidelberg University, 69115 Heidelberg, Germany

⁸Human Performance Research Centre, University of Konstanz, 78464 Konstanz, Germany

CORRESPONDING AUTHOR: M. NOUREDANESH (m2noured@uwaterloo.ca)

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada under Grant RGPIN-2015-05317, in part by the University of Michigan Injury Prevention Center by the Centers for Disease Control and Prevention under Grant R49-CE-002099, in part by the National Institute of Aging under Grant AG024824, and in part by the National Center for Advancing Translational Science under Grant UL1TR002240.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Medical Faculty, Tübingen University, Germany, under Application No. 266/2016MP2, the University of Michigan Institutional Review Board under Application No. HUM00073568, and the Northumbria University Research Ethics Committee under Reference No. 17589.

This article has supplementary downloadable material available at <https://doi.org/10.1109/JTEHM.2022.3163967>, provided by the authors.

ABSTRACT *Objective:* Older adults' falls are a critical public health problem. The majority of free-living fall risk assessment methods have investigated fall predictive power of step-related digital biomarkers extracted from wearable inertial measurement unit (IMU) data. Alternatively, the examination of characteristics and frequency of naturally-occurring compensatory balance reactions (CBRs) may provide valuable information on older adults' propensity for falls. To address this, models to automatically detect naturally-occurring CBRs are needed. However, compared to steps, CBRs are rare events. Therefore, prolonged collection of criterion standard data (along with IMU data) is required to validate model's performance in free-living conditions. *Methods:* By investigating 11 fallers' and older non-fallers' free-living criterion standard data, 8 naturally-occurring CBRs, i.e., 7 trips (self-reported using a wrist-mounted voice-recorder) and 1 hit/bump (verified using egocentric vision data) were localized in the corresponding trunk-mounted IMU data. Random forest models were trained on independent/unseen datasets curated from multiple sources, including in-lab data captured using a perturbation treadmill. Subsequently, the models' translation/generalization to older adults' out-of-lab data were assessed. *Results:* A subset of models differentiated between naturally-occurring CBRs and free-living activities with high sensitivity (100%) and specificity ($\geq 99\%$). *Conclusions:* The findings suggest that accurate detection of naturally-occurring CBRs is feasible. *Clinical/Translational Impact-* As a multi-institutional validation study to detect older adults' naturally-occurring CBRs, suitability for larger-scale free-living studies to investigate falls etiology, and/or assess the effectiveness of perturbation training programs is discussed.

INDEX TERMS Compensatory balance reactions, free-living digital biomarker, falls, machine learning, fall risk assessment.

I. INTRODUCTION

Falls in older adults, which may lead to serious physical and/or psychological consequences [1], [2], are one of the most important public health problems world-wide. Fall risk assessment (FRA) is the initial step for fall prevention programs and interventions, which aims to identify different risk factors for falls including intrinsic (e.g., gait and balance impairments, cognitive status alterations) and extrinsic/environmental (e.g., low-friction surfaces, uneven terrains) [3]. Despite significant advances in FRAs, falls are still resistant to preventive interventions. The majority of FRAs, such as Timed Up and Go [4] or instrumented methods (e.g., pressure sensitive mat), are confined to clinical settings (controlled conditions), where patients may alter their performance due to awareness of being observed (i.e., Hawthorne effect [5]). Moreover, FRAs conducted under controlled conditions could miss the examination of specific environmental and behavioural risks that can lead to falls. Therefore, new free-living FRAs are needed to address the aforementioned limitations.

Wearable sensor systems (e.g., smartwatches, inertial measurement units (IMUs) data loggers), have facilitated the emergence of *free-living FRAs* to monitor older adults' daily activities in out-of-lab conditions. To address this, early studies (reviewed in [6]) have explored the relationships between IMU-based free-living digital biomarkers (FLDBs) and the frequency of prospective or retrospective falls in older adults. These FLDBs include macro (e.g., quantity of: steps [7] and turns [8]) and micro (e.g., spatiotemporal measures such as step time [7], or frequency measures [9] including index of harmonicity) measures. However, many of these FLDBs, which were mostly dependent on the detection of steps, exhibited inconsistent fall predictive powers across studies, indicating that they may not be *stable* in distinguishing fall-prone individuals. Moreover, the relationships between falls and free-living dynamic postural control measures, such as step width [10] and the frequency of naturally-occurring compensatory balance reactions (CBRs), have yet to be investigated in depth [6]. Considering balance impairment as one of the strongest risk factors for falls [11], the investigation of balance-related FLDBs may lead to more stable risk assessments and provide new insights into fall prevention in older adults. This necessitates the development of robust models to identify balance impairment in older adults under free-living conditions.

Also known as missteps or near-falls, CBRs are reactions (e.g., trips, slip-like, crossover) to recover stability following a loss of balance, characterized by rapid movements to broaden the base of support. Findings from controlled studies support the view that impaired ability to execute CBRs is associated with a higher risk of falling [12], however, there have been only a limited number of studies investigating biomarkers that are related to the naturally-occurring CBRs. These studies either used self-reports [13], [14] or wearable sensor data [15]–[17] to quantify the frequency

of CBRs. For instance, Srygley *et al.* [13] described that the quantity of self-reported missteps was positively associated with the frequency of prospective falls in older adults. In contrast, Gazibara *et al.* [14] showed that self-reported near-falls were not linked to prospective falls in people with Parkinson's disease (PD). However, these findings were limited to self-reported observations with no further verification (e.g., video evidence) and lack spatial and temporal resolution. An objective approach was used in [15], where the quantity of 'suspected' missteps detected in 3 days of IMU recordings was reported to be strongly associated with retrospective falls in people with PD. The thresholds used in this CBR detection approach were mostly determined based on trial and error [15]. The highest number of suspected missteps was 1,007 within 4,148 gait windows (window length: 5 s, \approx 5.7 hours of gait), while the lowest number of suspected missteps was 4 within 95 gait windows (or 7.8 minutes of gait). The high rate of false positives was attributed to the presence of high amplitudes in the vertical (V) acceleration signal and more inconsistent gait patterns compared to controlled conditions [15]. The term 'suspected' for this FLDB highlights the lack of criterion (gold) standard data to reliably validate the employed *threshold-based* CBR detection approach in free-living conditions.

There have been *machine learning-based* CBR detection methods, developed based on surface electromyography (sEMG) [18], [19] or IMU [19]–[21] features, where the IMU-based models presented a more satisfactory performance compared to the sEMG-based ones [19]. These models were developed (trained and tested) using healthy young participants' data collected in controlled conditions, and achieved high detection accuracies. However, their translation to detect older adults' naturally-occurring CBRs has remained uninvestigated. Considering the aforementioned findings, further research on the validity of CBR detection models needs to be undertaken to reliably examine the associations between the frequency of naturally-occurring CBRs, as a *stand-alone* FLDB, and falls in older populations.

Performing a validation study in the context of CBR detection is logistically challenging. Compared to other gait events such as steps and turns, naturally-occurring CBRs are rare events and hard to capture. For instance, only 46 CBRs (trips) were self-reported by three older adults in 107 person-day of data [17]. Therefore, prolonged acquisition of criterion standard data (e.g., egocentric vision) along with IMU data from older adults is required to capture naturally-occurring CBRs. The integration of criterion standard data allows accurate identification/localization of CBR onsets in the corresponding IMU data and may provide information on the circumstances leading to false alarms. This information can be used to assess the performance of the IMU-based CBR detection models.

This paper presents a multi-institutional collaborative effort and proposes a machine learning-based framework for the detection of multidirectional CBRs, which has been

validated using fallers' and older non-fallers' free-living or out-of-lab data. The key considerations for model development and validation have been discussed in subsection I-A.

A. KEY CONSIDERATIONS FOR CBR DETECTION MODELS' TRAINING AND VALIDATION

Previous studies conducted in controlled conditions have suggested that a single IMU placed on participants' trunk (e.g., sternum [19], waist [20]) or pelvis [21] outperforms all other single IMU placement sites, including ankles and thighs, for the purpose of CBR detection, possibly as it performs better at approximating the linear mechanics acting through the whole-body center of mass. Moreover, the use of IMUs mounted on waist (close to pelvis) and bilaterally on ankles and thighs (5 total IMUs) resulted in slightly higher CBR detection accuracies compared to a single waist-mounted IMU (96.6% vs 94.7%) [21]. The marginal improvement in accuracy shown with multi-IMU methods coupled with the need to minimize obtrusiveness indicate the potential for a single sensor location suitable for prolonged field studies. Therefore, the data of a trunk-mounted IMU were considered in the present study.

Although CBRs happen more often during gait [22], a CBR detection model dependent on a gait detection algorithm (e.g., [15], [22]) may exhibit limited performance in some scenarios (see section IV). Thus, differentiation between CBRs from all other activities of daily living was hypothesized to be a more promising approach, and considered in this study.

Previous CBR detection studies [19]–[22] considered alternate methods of model training and performance assessment, such as k-fold and leave-one-subject-out cross-validation. Similar to our previous research works [21], [23], we hypothesize in the current study that incorporating a training dataset curated from data sources that are independently collected from the test dataset would result in the machine learning models with more realistic results in terms of generalization to unseen data (although lower accuracies are expected to be obtained compared to the cross-validation methods where training and test datasets share very similar distributions, e.g., k-fold [21]). Specifically, this study examines the use of in-lab perturbation data for model training as a viable approach to detect real-world CBRs (in the test dataset). This approach also facilitates forming a balanced training dataset (i.e., balanced set of CBR and non-CBR events, and balanced distribution over different CBR classes), which is otherwise very challenging to be achieved in free-living studies due to the rarity of naturally-occurring CBRs (discussed earlier) and the varying occurrence frequencies for different types. For instance, trips were reported to be the most common CBR type (e.g., in PD fallers [14]) and are potentially easier to be captured compared to the other CBR types during free-living data collection. While the investigation of CBRs in the sagittal plane has attracted more attention from the researchers (e.g., in [17], [22]), the ability to detect different CBR types, including those in the frontal plane, may provide a more

comprehensive insight into older adults' balance impairment. This can be addressed by training models on a comprehensive dataset that includes samples from different CBR types (e.g., crossover, sidestep, slip-like).

The findings of our previous study [21] indicated that a perturbation treadmill (PT) is a safe and reproducible option to elicit multidirectional CBRs (PT-CBRs). Additionally, it was hypothesized that incorporating PT and free-living data in the training dataset can augment the performance of CBR detection models [21].

To address the aforementioned points, two models were developed:

- 1) Model 1 was trained using an open access benchmark dataset, i.e., the Inertial Measurement Unit Fall Detection (referred to as the 'IMUFD') dataset [24], which includes young adults' simulated CBR and non-CBR events (simulated activities of daily living),
- 2) Model 2's training dataset was formed by adding an equal number of CBR and non-CBR events from (a) the PT dataset (young adults' data) [21] and (b) one older adult's out-of-lab activities' data from Multimodal Ambulatory Gait and Fall Risk Assessment in the wild (MAGFRA-W) dataset, to the IMUFD dataset.

While the incorporation of the aforementioned training datasets comes with multiple advantages, previous research showed that CBR detection models developed based on controlled data may generate high rate of false positives when applied to unseen/free-living data [15], [21]. In contrast to falls, which result in coming to rest inadvertently on the ground, CBRs are often accompanied by subtle changes in posture, and subsequently, may be confused with other activities of daily living [21]. Considering that the majority of samples in the training datasets for Models 1 and 2 were acquired from controlled data, several criteria were considered to *automatically* compensate for the prominent discrepancies between the training and validation/test datasets, when required (detailed in section II-D.2.a).

The dataset used to validate the proposed framework includes a subset of 11 fallers' and older non-fallers' multimodal data from a) Free-living IMU and Voice Recorder (FIVR) and b) MAGFRA-W datasets, which encompasses 8 naturally-occurring CBRs. The CBRs were verified using criterion standard data (e.g., visual verification in MAGFRA-W using egocentric vision data, Fig. 1). Using this independent validation/test dataset, the models' performance was further assessed by investigating their:

- generalizability to detect naturally-occurring CBRs executed by older adults with different characteristics (e.g., history of falls, with walking aids),
- robustness against false alarms generation in different indoor and outdoor contexts.

The translation results to detect naturally-occurring CBRs are discussed in section III and the framework's clinical implications are further highlighted in section IV.

II. METHODS AND PROCEDURES

A. DATASETS AND MULTI-INSTITUTIONAL STUDIES

Previous research showed excellent agreement between spatiotemporal measures estimated from L5- (of the lumbar spine) and waist-mounted (right hip) accelerometers [25]. Therefore, despite discrepancies in the exact anatomical location of the trunk-mounted IMU across the multi-institutional datasets, inertial data collected from pelvis-, lower back- (L5-), and waist-mounted IMUs, i.e., trunk-mounted, were considered comparable for the task of CBR detection model development in this study. While multiple IMUs were used to collect data in the studies discussed here, data recorded by trunk-mounted IMUs were considered to develop CBR detection models. The sensor's three orthogonal axes were checked to be aligned with the three anatomical axes in upright posture for the MAGFRA-W dataset as well as the IMUFD, PT, and FIVR datasets as detailed elsewhere [15], [16], [21].

1) IMUFD

The IMUFD dataset includes 150 CBRs and 240 non-CBR epochs simulated by 10 healthy young participants between 22 and 32 yrs [24]. Five types of CBRs (commonly observed in videos recorded in long-term care facilities) were simulated: 1) trips, 2) slips, 3) hit and bump (by another person), 4) incorrect transfer while rising from sitting to standing, and 5) misstep during gait. The simulated non-CBR epochs include the following activities: 1) walking, 2,3) ascending and descending stairs, 4) standing, 5) sitting to standing, 6) standing to sitting, 7) standing to lying, and 8) picking up an object from the ground. Only data from the waist-mounted IMU (APDM Opal, Portland, USA), were considered in the study (sampling frequency of $f_s = 128$ Hz, triaxial accelerometers range: ± 6 g; triaxial gyroscope, range: ± 1500 deg/s).

2) PT DATASET

As detailed elsewhere [21], nine healthy young participants (mean age = 26 yrs) wore five IMUs (Opal model, APDM Inc., accelerometers and gyroscopes were set to operating ranges of ± 16 g and ± 2000 deg/s, respectively) at the pelvis, and bilaterally on thighs and ankles, and walked over a perturbation treadmill (speed = 1.1 m/s). Perturbations in 4 directions (right, left, backward, forward) were induced during the right or left leg stance phases (in 2 separate 20-minute sets). This process resulted in eight different classes of PT-CBRs, such as sidestep, crossover, slip-like, trip-like (80 PT-CBRs/set for each participant, overall 160 PT-CBRs/participant). Here, the pelvis-mounted IMU data recorded from 6 participants were considered for model development. The study has received ethics clearance and was reviewed and approved by the Medical Faculty, Tübingen University, Germany (No: 266/2016MP2).

3) FIVR DATASET

In order to capture real world CBRs, participants wore a wrist-mounted voice recorder and 4 body-worn IMUs (Opal,

APDM Inc., Portland, USA; $f_s = 128$ Hz, ± 16 g acceleration, ± 2000 deg/s angular rate) during waking hours on the wrist, feet, and lower back [16]. As detailed elsewhere [16], 5 participants (4 males, 76.2 ± 5.4 yrs, with a history of ≥ 2 falls in the past 6 months) were instructed to self-report any CBR (defined as an event where balance control was lost momentarily, but recovered, including slips, trips, stumbles or missteps) using the voice-recorder immediately after the event occurrence. Here, the self-reported trips (either the participant used the word 'trip' or the explained contexts that were consistent with a trip such as 'stubbed foot' or 'caught foot on') were considered. A pose estimation algorithm was used to verify the presence of CBRs within the recorded IMU data and spot their onsets. To address this, location of the feet, as well as lower back and wrist orientation data were combined to create a three-dimensional animation representing the estimated body motion [26]. Overall, 7 CBRs (all trips), with ≈ 10 minutes before and after each event (overall 140 minutes of data) were taken into account for model validation (see FIVR D1 to D7 in Fig. 2 and 3, D: dataset). The study reviewed and approved by the University of Michigan Institutional Review Board (HUM00073568).

4) MAGFRA-W DATASET

The MAGFRA-W dataset includes data collected by multiple wearable IMUs (Axivity, Newcastle upon-Tyne, UK; acceleration range: ± 8 g, angular velocity range: ± 500 deg/s, $f_s = 100$ Hz) as well as a waist-mounted camera (GoPro Hero 5 Session or Hero 6 Black camera, 30fps, wide view) in out-of-lab environments. Data collection was performed in (a) public environments within Northumbria University, during which older adults navigated through different indoor and/or outdoor environments while walking alongside a researcher, or (b) older adults' homes (indoor) or their neighbourhood (outdoor) for $\approx 1 - 2$ hours with no researcher in attendance. Outdoor data collection was performed during daylight hours. The camera was centered at each older adults' waist by means of a belt attachment and was set up to capture top-down views of feet and the regions around them, with no calibration or a strictly reproducible placement procedure on camera's angle with respect to the frontal plane. In the present study, the L5-mounted IMU data collected from 7 participants (mean age: 73.46 yrs, 1 male, 3 fallers based on the number of self-reported falls in the previous 12 months) were processed. One older adult's data (female, 80 yrs, non-faller) were used for model training (see II-C.1.c) and 6 participants' data were considered for models' validation (see II-D.1.b, Supplementary Materials I and II). One participant's age was below 65 yrs, however, as she was a recurrent faller, her data were considered for further analysis (MAGFRA-W D5: female, 55 yrs, faller). MAGFRA-W D3 (female, 80 yrs, non-faller) and MAGFRA-W D5 include 2 adults' data collected in their homes and neighbourhoods. In MAGFRA-W D5 and D6 (male, 80 yrs, faller) participants used walking aids. The project received ethics approval (reference number: 17589, approval date: 4-Oct-2019) from Northumbria University

Research Ethics Committee, Newcastle upon Tyne, UK. All participants gave written informed consent before participating in the study.

B. SIGNAL PREPROCESSING

The AX6 data in the MAGFRA-W dataset showed inconsistency with the other 3 APDM-captured datasets (IMUFD, FIVR, and PT) in terms of the units and sampling frequency. Therefore, unit conversion as well as signal upsampling (100 to 128 Hz, using MATLAB interpolation method 'pchip') were performed to obtain comparable data within all datasets. Signal detrending (removing the DC offset) was considered in the previous CBR detection studies to address slight tilts/shifts in sensor placement [15], [21]. Therefore, for each of the simulated CBR and non-CBR trials in the IMUFD dataset (with an approximate width of 15s/trial), each of the 6 acceleration (ACC) and angular velocity (Gyro) signals was detrended separately. Moreover, due to the consistency in activity type (over-treadmill walking), which resulted in relatively consistent sensor orientation, all six ACC and Gyro signals corresponding to each set were detrended separately [21]. However, as the orientation of the trunk-mounted IMU with respect to the gravitational vector is expected to change considerably during free-living activities (impacting signals' DC values), for FIVR and MAGFRA-W datasets, instead of detrending the full-length signals for each participant, non-overlapping sliding windows (SWs) with the length of 15 s (in accordance with the IMUFD segments) were applied to each of the six inertial signals, and the overlapping data were detrended separately. All data were processed using MATLAB (R2019a, MathWorks Inc, USA).

C. MODEL TRAINING

In this section, the procedure for data preprocessing and segmentation is discussed for each dataset. Overall, to form the *training datasets*, 227, 60 and 60 non-CBR and 148, 120 and 0 CBR signal segments were extracted from the IMUFD, PT and MAGFRA-W inertial data, respectively (overall 17 individuals). The subsequent segments were further used for feature extraction (see II-C.2) and preparation of the training datasets (i.e., feature matrices X) for Models 1 and 2.

1) DATA SEGMENTATION

Based on the findings reported in [18], [19], [21], in the signal vector amplitude of acceleration signal (SVA_{ACC}) recorded by a trunk-mounted accelerometer, the peaks, i.e., $argmax(SVA_{ACC})$, can be reliable signal-based indicators of CBR onsets in response to perturbations. Moreover, based on the available evidence and criteria discussed in [21], a segment width of ≈ 4.69 ($= 601$ samples at $f_s = 128$ Hz) created by cropping ≈ 2.34 s (or 300 samples with $f_s = 128$ Hz) before and after of the corresponding $argmax(SVA_{ACC})$ in all 6 ACC and Gyro signals is sufficiently wide to encompass important transitional information attributed to the mechanical and postural adjustments evoked after a perturbation.

Here, each CBR and non-CBR segment is a 6×601 matrix (6: number of signals).

a: IMUFD SIGNAL SEGMENTATION

After calculation of the SVA signal and detection of $argmax(SVA_{ACC})$ for each trial in IMUFD (section II-A.1), 227 non-CBR and 148 CBR segments were considered for feature extraction to form the training datasets for models 1 and 2 (as discussed in II-C.3).

b: PT SIGNAL SEGMENTATION

Considering the possible adaptation happening over the course of data collection (80 CBRs/set for each participant as discussed in II-A.2) [27], only the first 10 CBRs elicited in each set were considered, resulting in 120 PT-CBRs (6 participants \times 2 sets (right and left leg stance phases) \times [2 (trip-like) + 2 (slip-like) + 2 (crossover) + 2 (sidestep)]). The i^{th} PT-CBR segment was created by cropping 300 samples before and after of the sample corresponding to $argmax(SVA_{ACC,i})$ in all of the 6 signals. Additionally, 60 non-CBR segments were extracted from the 'steady state' normal over-treadmill walking intervals between the two consecutive PT-CBRs (as discussed elsewhere [21]). These segments were further considered for feature extraction (discussed in II-C.2) to form the training dataset for Model 2.

c: MAGFRA-W SIGNAL SEGMENTATION

One participant's (female, 80 yrs) data from the MAGFRA-W dataset were used to prepare the training dataset for Model 2. This participant's data were confirmed to be free of any CBR events by manual inspection of the egocentric vision data. A non-overlapping SW with the length of 5 s, i.e., $SW_{SVA_{ACC},5s}$, was applied to the SVA_{ACC} signal attributed to this participant. In each $SW_{SVA_{ACC},5s}$, the index corresponding to the peak, i.e., $argmax(SW_{SVA_{ACC},5s,j})$, is identified, and 300 samples before and after this point in all 6 signals form the segment. Overall, 60 non-overlapping non-CBR segments were selected and considered for feature extraction (discussed in II-C.2) to form the training dataset for Model 2.

2) FEATURE EXTRACTION

Extraction of discriminative features from the IMU segments is a necessary step in the proposed machine learning-based approach for the recognition of CBR patterns. In contrast to the CBR detection models proposed in [21], for which each of the 6 ACC and Gyro axes was considered independently for feature extraction, for each of the CBR and non-CBR segments, only 2 signals: 1) SVA_{ACC} and 2) the SVA of angular velocity signals (SVA_{Gyro}), were taken into account. The following 20 features were extracted from the SVA_{ACC} and SVA_{Gyro} components of each segment: 1) maximum peak, 2) root mean square (RMS), 3) mean, 4) variance, 5) skewness, 6) kurtosis, 7) number of peaks, 8) maximum autocorrelation, 9) integral (trapezoid numeric), 10) the Shannon entropy, 11) amplitude of the dominant frequency (periodogram PSD), 12) the dominant frequency in the segment,

13) maximum of signal derivative, 14) mean of the signal derivative, 15) variance of the signal derivative, 16) skewness of the signal derivative, 17) kurtosis of the signal derivative, 18) RMS of the signal derivative, 19) integral (trapezoid numeric) of the signal derivative, and 20) the Shannon entropy of signal derivative. In addition to the aforementioned features, $\text{argmax}(SVA_{Gyro})$ in each segment was considered, resulting in 41 ($= 2 \times 20 + 1$) features for each window. These features were previously taken into account for the development of CBR detection models [19]–[21].

3) TRAINING PROCEDURE

In our previous work on a similar classification problem to detect CBRs, multiple machine learning techniques were examined [18], [19] [21], where the random forest (RF) method (bootstrap-aggregated decision trees) [28] exhibited a satisfactory performance. Considering RFs permit parallel processing and demonstrates robustness against nonlinear relationships, and considering the size of the training dataset (small for the development of deep learning models), RF models were investigated.

The training datasets for Models 1 and 2 were formed by concatenating the feature vectors extracted from the 1) IMUFD segments (a $X_{375 \times 41}$ matrix), and 2) IMUFD, PT, and MAGFRA-W segments (a $X_{615 \times 41}$ matrix). Based on the initial tests, an RF model with 19 trees (RF_{19}) showed satisfactory results on all validation datasets, while more trees resulted in excessive sensitivity to classify a considerable proportion of local peaks as a CBR (likely due to overfitting). To indicate that the results are not impacted by the inherent model randomness, another metric, i.e. ‘confidence score’ was defined. This metric considers the predictions of 50 RF_{19} ’s models trained on the corresponding datasets for Models 1 and 2 (discussed in section II-D.2.c). MATLAB defaults were used for other parameters including the minimum number of observations per tree leaf (i.e., 1 for classification) and number of variables to select at random for each decision split (i.e., square root of the number of variables for classification).

D. MODELS VALIDATION BASED ON FREE-LIVING DATA

1) VALIDATION/TEST DATASET

a: FIVR DATASET

Data discussed in II-A.3 were considered to validate the proposed CBR detection models. In each of the 7 FIVR datasets, the confirmed CBR is located in the centre of the timeseries, i.e., $t \in 600 \pm 3$ s, in FIVR D1 to D7, as shown in Fig. 2 and Fig. 3.

b: MAGFRA-W DATASET

By visual inspection of the recorded egocentric vision data in the MAGFRA-W dataset, 1 naturally-occurring (hit/bump) CBR was identified (see Fig. 1 and Fig. 3-MAGFRA-W D1, in which the CBR event happened at $\approx t = 631$ s). The participant (female, 76 yrs, non-faller) hit a light pole and

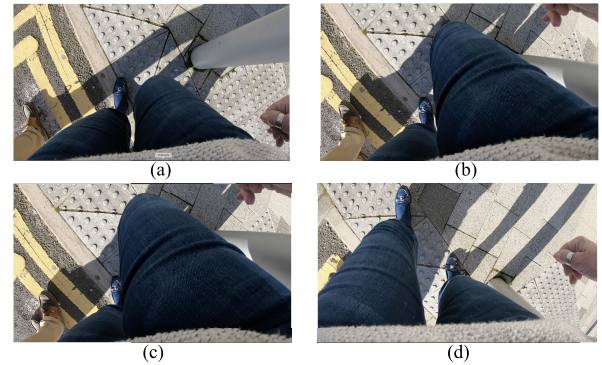


FIGURE 1. A CBR was observed in an older adult’s egocentric vision data (MAGFRA-W D1). The participant hit a light pole and lifted right leg forward.

lifted her right leg forward. The multimodal data attributed to this participant captured different movement patterns such as level walking on different surfaces, turns, the use of elevator, stair descending, and obstacle avoidance (see Supplementary Material I-Fig. 1).

Data from 5 more participants (Fig. 4 and Supplementary Material II) were examined to assess the models’ robustness across varying contexts during which the models could generate false alarms.

2) REGIONS OF INTEREST

The validation dataset was segmented similar to the method described in section II-C.1.c. To avoid confusion, data segments extracted from the validation datasets are referred to as the ‘regions of interest’ (ROIs). A $SW_{SVA_{ACC}, 5s}$ was applied to the IMU data in the FIVR and MAGFRA-W datasets, after removing ≈ 10 s from the start and end of each dataset. ROI_j (a 6×601 matrix) includes all samples $\in [ind_{ROI_j} - 300, ind_{ROI_j} + 300]$ from all 6 ACC and Gyro signals, $ind_{ROI_j} = \text{argmax}(SW_{SVA_{ACC}, 5s, j})$, where j denotes the ROI’s number in the corresponding dataset. If the distance between the peaks in the adjacent ROIs was less than 300 samples, i.e., $|ind_{ROI_{j+1}} - ind_{ROI_j}| \leq 300$, the ROI corresponding to the smaller peak was disregarded as a considerable proportion ($\geq 50\%$) of this ROI (including the peak) is being automatically included in the ROI attributed to the peak with higher amplitude. This ROI elimination approach can play an important role in large-scale free-living studies, as it reduces the overall processing time by decreasing the number of data points being examined by the CBR detection models.

a: POSSIBLY-NOISY ROIS

Preliminary results (Supplementary Material III) indicated that all CBRs were detectable either by Model 1 or 2; with the exception of 1 CBR event depicted in FIVR D2 (Fig. 2). The models’ inability to capture this CBR event was surprising as the peak corresponding to the CBR’s onset in SVA_{ACC} was higher than the other detected CBR events

(see Fig. 2 and Fig. 3). Moreover, high rate of false positives were initially observed in FIVR D2 (Fig. 2) and FIVR D6 (Fig. 3) as shown in Supplementary Material III.

As mentioned in subsection I-A, the reasons behind the aforementioned false negative and positive observations can be attributed to the differences between the training and validation/test datasets, more specifically due to the differences: 1) between young healthy and older adults' performance, and 2) between free-living and controlled data (e.g., treadmill vs. free-living walking).

Hof *et al.* [29] reported that older adults corrected perturbations with higher variability and less accuracy in foot placement, regained balance with more steps, and demonstrated higher attentional demand. However, all CBRs in the IMUFD and PT datasets were collected from young healthy adults. Therefore, age-related difference are likely to play a role in translating models trained on these datasets to detect older adults' CBR onsets. Moreover, while previous research showed that gait speed could impact compensatory stepping characteristics [30], all multidirectional PT-CBRs were elicited while participants were walking with a constant speed on the treadmill.

Moreover, previous research have highlighted differences between in-lab and free-living gait [31]–[34] as well as discrepancies between older and young adults' gait [35]. Compared to controlled gait, acceleration signals attributed to free-living gait represent lower regularity [32], more diverse range, e.g., in the anterior-posterior (AP) [34] and V [15], [34] axes, which may result in the generation of false positives. Treadmill and overground walking patterns were also reported to be different in terms of smoothness and rhythmicity [36], and some of the digital biomarkers extracted from older adults' treadmill and free-living gait data were significantly different [34]. However, a considerable proportion of the training datasets for Models 1 and 2 include young adults' data collected under controlled conditions. In contrast, the validation datasets captured walking patterns from older adults in diverse contexts (e.g., while walking on an uneven surface covered by gravels), with various gait speeds, different walking bout lengths (e.g., short, long), and contains gait events such as turns.

Considering the aforementioned points, as both CBR and non-CBR data acquired under controlled conditions typically demonstrate more regular and smoother acceleration signals compared to free-living data, we hypothesized that by detecting and filtering/smoothing 'possibly-noisy' ROIs in free-living data, we can compensate for inconsistencies between the training and validation datasets and subsequently, improve the overall performance of the CBR detection models.

As mentioned earlier, gait speed impacts compensatory stepping characteristics [30]. Based on knowledge that walking speed was strongly correlated with range in ACC_V and ACC_{AP} signals [9], and considering that both ACC_V and ACC_{AP} demonstrated significantly higher ranges during free-living gait compared to their in-lab counterparts [34]

(with little-to-no difference was reported in mediolateral direction), we hypothesized that the range in AP and V directions can be used to define the 'possibly-noisy' condition for a ROI. Therefore, while defining this condition warrants deeper investigation of controlled and free-living data, we propose a 'possibly-noisy' ROI definition if the range in ACC_{AP} or ACC_V in a window (with the length of $\phi = 2.32s$) before or after of the ROI is above a certain threshold ($\theta_{AP} = 8.55m/s^2$ and $\theta_V = 11.36 m/s^2$). These hyperparameters/thresholds were obtained based on the results reported in [9], in which free-living data were collected from more than three hundred older adults (including fallers and non-fallers) using a lower back-mounted IMU. Window size, ϕ , was obtained based on the average stride frequency in older adults' free-living data ($\phi = 2 \times \text{average stride time} = 2 \times 1/0.86Hz \approx 2.32 s$). Based on the same study, $8.55 m/s^2$ and $11.36 m/s^2$ were the average range values for ACC_{AP} and ACC_V signals during gait, respectively [9]. These parameters for the identification of possibly-noisy conditions were selected based on the assumption that CBRs are more likely to occur during gait. If the windows before and after a ROI overlap with non-gait regions (e.g., sedentary), the possibly-noisy condition is less likely to be met for the ROI.

To compensate for inter-dataset differences, we hypothesized that applying a low-pass butterworth filter with the cut-off frequency of 10 Hz and order of 1 to the possibly-noisy ROIs would make the underlying acceleration signals smoother, while it can preserve important kinematic information related to CBRs and other activities. Subsequently, for each detected possibly-noisy ROI, all 6 ACC and Gyro signals were filtered and then the SVA_{ACC} and SVA_{Gyro} signals were recalculated. The ROI remained unchanged (no filter was applied), if the possibly-noisy condition was not met.

b: FEATURE EXTRACTION FROM ROIS

For each ROI, either low-pass filtered or unchanged, the 41 features discussed in II-C.2 were extracted.

c: ROI'S CONFIDENCE SCORE

For each ROI, the average of outputs (1: CBR, 0: non-CBR) from 50 RF_{19} 's in each model was defined as the ROI's confidence score. Subsequently, a ROI encompasses a CBR if its corresponding confidence score is ≥ 0.9 (i.e., at least 45 out of 50 RF_{19} 's in Model 1 or 2 classified the ROI as a CBR).

III. RESULTS

From 8 verified free-living CBR events, 4 and 6 CBRs were correctly detected (confidence score ≥ 0.9) by Models 1 and 2, respectively (see Table 1). Model 1 was unable to detect 4 valid CBRs: 1) FIVR D1 (confidence score: 0.00), 2) FIVR D4 (confidence score: 0.24), 3) FIVR D5 (confidence score: 0.12), and 4) MAGFRA-W D1 (confidence score: 0.54), and achieved the sensitivity of 50.00%. CBRs in FIVR D1 and FIVR D5 were correctly classified by 29 and 19 out

of the 50 RF_{19} 's in Model 2, respectively, however, their corresponding confidence scores did not exceed the threshold of 0.9 (i.e., considered as false negatives). Therefore, Model 2 achieved 75.00% sensitivity. From a total of 4,047 non-CBR ROIs tested (extracted from all 12 datasets), 8 and 7 false positives were generated by Models 1 and 2, respectively. Subsequently, Models 1 and 2 achieved the overall specificities of 99.80% and 99.82%, respectively.

Among 50 trained RF_{19} 's in Model 2, 5 models detected all 8 CBRs. However, no subset of RF_{19} 's in Model 1 was able to detect all 8 CBRs (e.g., confidence score of 0.00 was achieved for the CBR in FIVR D5). To probe further, Model 2' was considered by including this subset of 5 RF_{19} 's in Model 2. Model 2', which achieved 100% sensitivity, was applied to all 12 validation datasets to examine its robustness against generating false positives (Table 1). This model generated 13 false positives yielding the overall specificity of 99.80%.

By comparing the preliminary CBR detection results in Supplementary Material III (obtained for 8 datasets: FIVR D1-D7 and MAGFRA-W D1) with the corresponding results in Table 1, it was observed that detecting and filtering the possibly-noisy ROIs improved the overall models' performance. For these 8 datasets, the overall sensitivity and specificity of Model 1 increased from 37.50% and 98.71% to 50.00% and 99.87%, respectively. Similarly, the sensitivity and specificity of Model 2 increased from 62.50% and 95.82% to 75.00% and 99.74%, respectively. While a reduction in the quantity of false positives was observed for FIVR D2 (Model 2: 27 to 0) and FIVR D6 (Model 1: 13 to 1; Model 2: 34 to 2), the results for some datasets including FIVR D1 and FIVR D5 did not change after applying the possibly-noisy condition. Before considering this condition, the CBR in FIVR D2 (Fig. 2) was not detectable by the models, however, it was successfully detected by Models 1 and 2, when its corresponding ROI, identified as possibly-noisy, was filtered. The corresponding ROI to the CBR in FIVR D6 also met the criteria for being counted as possibly-noisy, and was still detectable by the models after being filtered.

While an equal number of 236 SWs were extracted from each of the FIVR datasets, different numbers of ROIs were reported across these datasets (Table 1) due to the integration of the ROI elimination approach. This resulted in $\approx 16.3\%$ reduction in the total number of ROIs for FIVR D1 to D7, while saving the ROIs corresponding to the CBR events.

To visualize the range of contexts captured, sample multimodal data for one participant (MAGFRA-W D2: female, 76 yrs, 0 falls in the prior year), who walked on different indoor and outdoor surfaces (e.g., stairs, gravel, grass, transitions), are included in Fig. 4. The only context leading to the false positive in this dataset was a sudden change in walking direction on carpet (indoor environment). Moreover, Supplementary Material I represents an anticipatory pattern, obstacle avoidance, which could generate false positives according to previous studies' reports [15]. While this event could have been confused with sidestep and crossover CBRs, the models

did not generate false positives (see MAGFRA-W D1 in Fig. 3). There were also several peaks with higher SVA_{ACC} amplitudes than the spotted CBRs', e.g., in Fig. 2-FIVR D4, FIVR D5, as well as data points with high amplitudes (e.g., in MAGFRA-W D4 and D6, Supplementary Material II) for which Models 2 and 2' did not generate false positives, indicating the models' robustness against such signal features.

IV. DISCUSSION/CONCLUSION

This paper presents one of the first CBR detection frameworks validated using criterion standard data (including egocentric vision) captured from older adults under free-living conditions. The validation/test dataset were captured from 11 fallers and older non-fallers with different levels of mobility impairment while interacting with different indoor and outdoor environments. The mobility patterns considered in the validation dataset include various walking speeds, turns, ascending/descending stairs, transitions, and anticipatory reactions (e.g., obstacle avoidance). To represent a pragmatic picture of model's generalizability to unseen datasets (i.e., complex free-living data captured from older adults with different characteristics), rather than using cross-validation approaches in which training and test datasets share very similar distributions (e.g., k-fold), we formed training datasets by curating data from multiple sources, including young adults' controlled data that were independently collected from the test dataset. Moreover, the integration of PT-induced CBRs was hypothesized to provide satisfactory proxies for the lack of available data of multidirectional naturally-occurring CBRs in the target older adult populations to form a sufficiently large training dataset. Therefore, Model 1 was trained on an open access dataset (IMUFD), and Model 2 was trained on a curated dataset from young adults' (IMUFD, 120 PT-CBRs, and 60 non-CBR events from PT) and one older adult's (60 non-CBR events from MAGFRA-W) data (i.e., 120 PT-CBR and 120 non-CBR events were added to IMUFD). A condition was further defined to automatically detect possibly-noisy signal segments to further compensate for the prominent discrepancies between the training and validation/test datasets. Model 2 showed a higher sensitivity compared to Model 1 (75% vs 50%) and generated slightly fewer false positives (7 vs 8). From the 50 trained RF_{19} 's in Model 2, 5 models (formed Model 2') detected all 8 CBRs, indicating that an optimized subset of RF 's can be found to achieve a high sensitivity (100% here) in the detection of CBRs. This model was more prone to generating false positives (overall specificity: 99.67%) compared to Models 1 (overall specificity: 99.80%) and 2 (overall specificity: 99.82%). However, considering its lower processing time ($\approx \times 1/10$ of Model 2) and higher sensitivity, Model 2' can be considered superior to Models 1 and 2, and thus, suitable for being tested in larger-scale studies.

The higher sensitivity of Model 2 (and 2'), compared to Model 1, is due to the inclusion of PT-CBRs as well as one older adult's out-of-lab data in the training dataset.

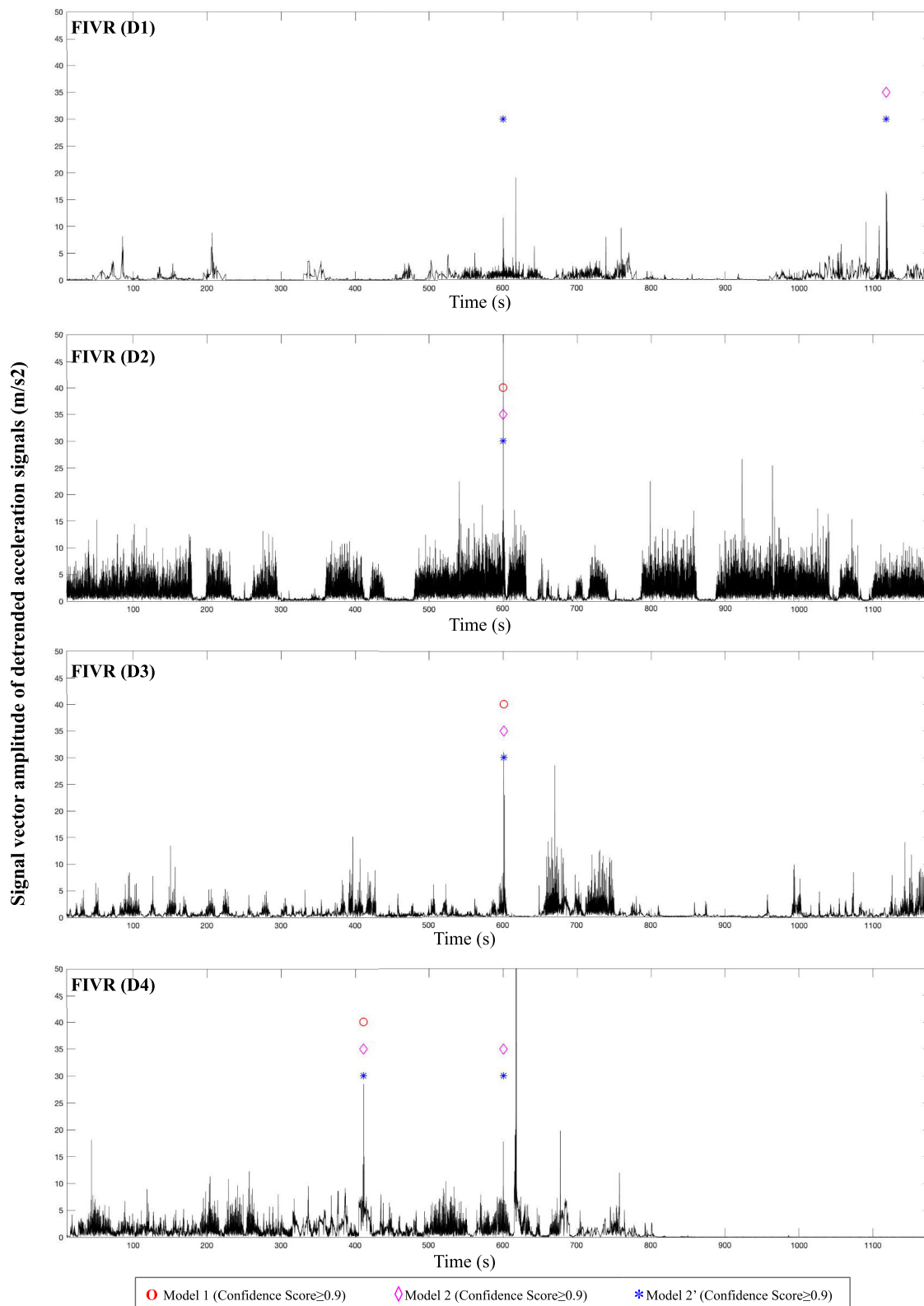


FIGURE 2. CBR detection Models 1, 2, and 2' were applied to the FIVR datasets (the CBR events are located at 600 ± 3 s).

The simulated CBRs performed by participants in the IMUFD [20] mostly include anticipatory adjustments preceding voluntary movements [21]. However, as reactive

responses to unanticipated threats to dynamic equilibrium during gait, CBRs must be rapidly executed often without anticipatory adjustments to provide stability in the face

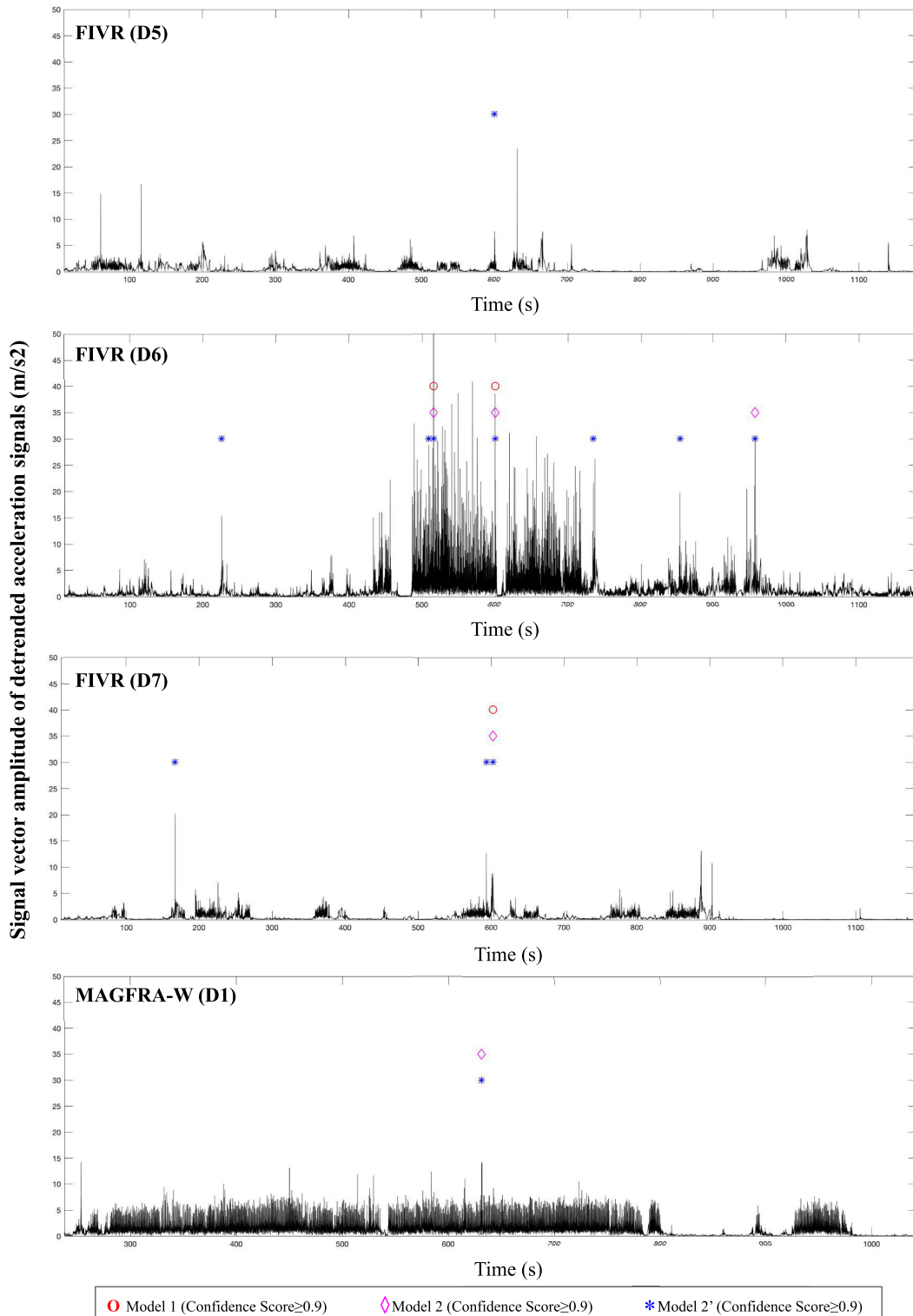


FIGURE 3. CBR detection Models 1, 2, and 2' were applied to three FIVR datasets (the CBR events are located at 600 ± 3 s) and one older adult's data from the MAGFRA-W dataset (the CBR event is located at $t = 631$ s).

of environmental challenges, and are performed automatically with no attention [21], [29]. Considering the reproducibility and safety of the PT approach, as well as the findings of the present study, this approach is suggested

to be used to collect larger-scale multidirectional PT-CBR datasets elicited in different gait speeds, to boost the generalizability of the proposed models. Collecting larger training datasets would subsequently facilitate the development of

TABLE 1. Test results obtained after applying the models on the extracted regions of interest (ROIs) from the validation datasets. FP: false positive.

	#ROIs	Model 1			Model 2			Model 2'		
		Sensitivity	Specificity	#FPs	Sensitivity	Specificity	#FPs	Sensitivity	Specificity	#FPs
FIVR D1	190	0.00	100	0	0.00	99.47	1	100	99.47	1
FIVR D2	203	100	100	0	100	100	0	100	100	0
FIVR D3	197	100	100	0	100	100	0	100	100	0
FIVR D4	198	0.00	99.49	1	100	99.49	1	100	99.49	1
FIVR D5	200	0.00	100	0	0.00	100	0	100	100	0
FIVR D6	201	100	99.50	1	100	99.00	2	100	97.00	6
FIVR D7	194	100	100	0	100	100	0	100	98.96	2
MAGFRA-W D1	173	0.00	100	0	100	100	0	100	100	0
MAGFRA-W D2	65	N/A	98.46	1	N/A	98.46	1	N/A	98.46	1
MAGFRA-W D3	1100	N/A	99.72	3	N/A	99.81	2	N/A	99.81	2
MAGFRA-W D4	205	N/A	100	0	N/A	100	0	N/A	100	0
MAGFRA-W D5	990	N/A	100	0	N/A	100	0	N/A	100	0
MAGFRA-W D6	139	N/A	98.56	2	N/A	100	0	N/A	100	0

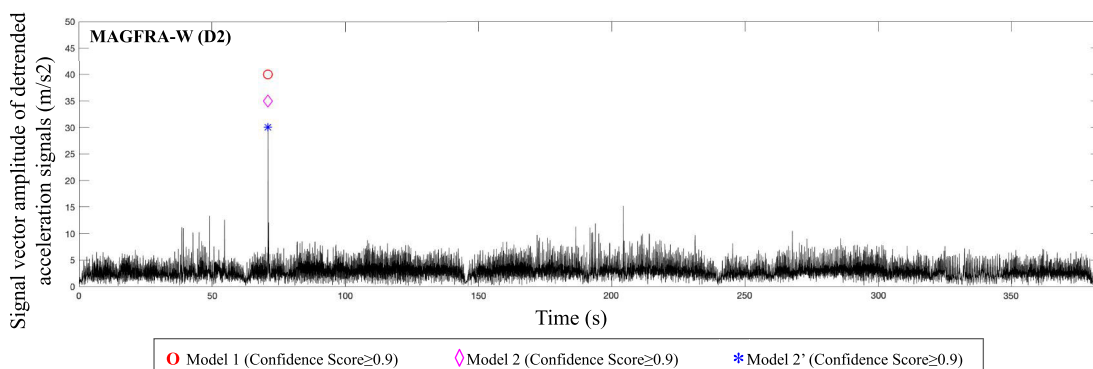


FIGURE 4. Multimodal data collected from one older adult. Upper panel: the detrended signal vector amplitude of trunk-mounted acceleration signals and the false positives generated by the models. Lower panel: sample frames captured by a waist-mounted camera showing different indoor and outdoor walking surfaces the participant walked over.

deep learning models, which may outperform the random forest models and the engineered features discussed in the present study [21].

By exploring the findings of previous research works, which examined differences between free-living and controlled digital biomarkers, hyperparameters (e.g., ϕ , V and AP range) were considered to automatically detect possibly-noisy ROIs in the validation dataset. Although confirming the suitability of these hyperparameters for the detection of highly irregular ROIs requires a deeper investigation, they led to promising results in the present study. The results obtained after applying a 10-Hz low-pass filter to the detected possibly-noisy ROIs indicated these signal segments can become smoother and potentially more comparable to the training dataset, while important kinematic information in their underlying CBRs (e.g., in FIVR D6 and D2) can be

preserved. Subsequently, an increase in the overall models' sensitivity and specificity was observed. Among all datasets, FIVR D6 generated the highest rate of false positives. Even after the consideration of the possibly-noisy condition, 6 out of the total 13 false positives generated by Model 2' were attributed to this dataset. We attribute this high rate of false positives to the significant differences between the movement task(s) in this dataset (walking in a construction site, which resulted in high amplitudes) and training dataset. By incorporating a more inclusive training dataset, the models' performance is expected to be improved. Although the collection of older adults' free-living non-CBR events is not as challenging as capturing their naturally-occurring CBRs, only 60 samples from one older adult's out-of-lab activities were considered in the training dataset for Model 2 (and 2') so that the balance between the number of CBR and non-CBR events

could be maintained. Considering a more inclusive training dataset consisting of different individuals' non-CBR and CBR events captured in various conditions may also bypass the requirement for detecting possibly-noisy ROIs. This will be investigated in our future studies.

Orientation signals (obtained from raw IMU signals) were considered to detect CBRs in previous studies [22] and used to verify the presence of CBRs in the FIVR dataset (see II-A.3). Here, although CBRs were detected with high sensitivity and specificity, incorporating the SVA signals (rather than individual ACC and Gyro signals) as well as the detrending process made the proposed models potentially less sensitive to sensor orientation misconfiguration and everyday orientation changes. The incorporation of orientation signal features may impact the performance of the proposed CBR detection framework, and may be considered in our future studies.

Due to the challenges associated with the collection of naturally-occurring CBRs, only 8 CBRs were verified and investigated in the present study. Larger scale free-living studies are required to be conducted to better understand the natural statistics of CBRs in the real world and deeper assess the performance of the incorporated: machine learning models, hyperparameters for defining the possibly-noisy condition, detrending process (e.g., optimal SW size), and ROI elimination approach. Overall, considering the large range of free-living movement patterns captured in the validation/test dataset, and considering state-of-the-art models may not generalize well to new users whose data have not been used in the training process [37], the proposed framework exhibited a satisfactory performance.

A. NO REQUIREMENT FOR GAIT DETECTION

Previous work suggested a two-step approach for CBR detection, which require gait detection as the first step [15], [22]. However, CBRs may not necessarily occur during walking (e.g., incorrect transfer while rising from sitting to standing). Moreover, poor performance of an employed gait detection approach may decrease the overall sensitivity of the subsequent CBR detection model. For instance, while short walking bouts constitute a considerable proportion of daily walking bouts in older adults [38], they can be missed/disregarded by commonly used gait detection algorithms [6]. The majority of gait detection algorithms rely on the identification of heel strike events in the acceleration signals. However, when it comes to free-living conditions, these events may not always be identified by distinctive peaks [39], due to reduced gait speed [40] and different variations of gait patterns (e.g., scuffling, dragging of the feet [41]) happening frequently during activities such as household cleaning [39]. This may lead to misidentification of gait events and potentially reduce the sensitivity of CBR detection models. As opposed to detection during gait only, distinguishing CBRs from all activities of daily living, as proposed here, may outperform models focused solely on detection of during-gait CBRs.

B. CLINICAL APPLICATIONS

Future research aims to focus on applying the described models to large-scale free-living IMU datasets collected from older fallers and non-fallers in a longitudinal manner to better understand the associations between falls and CBR-related FLDBs (e.g., direction, duration, number of steps to recover balance as well as the signal-based features discussed in section II-C.2). This would further allow the identification of stable CBR-related FLDBs in terms of detecting older fallers. Moreover, there are a number of perturbation training programs that are currently being tested/developed in clinical settings [42]–[44]. However, the transfer of balance recovery skills gained during these in-clinic programs to everyday scenarios has not been well-investigated. The models proposed here help track responsiveness to these programs by providing objective information on the timing and frequency of naturally-occurring reactive responses induced by real-life perturbations.

The egocentric vision data captured in MAGFRA-W provided rich contextual information about the factors leading to CBRs (e.g., a light pole, Fig. 1) and contexts that may lead to the generation of false alarm (e.g., a sudden change in walking direction, Fig. 4). By identifying contexts associated with verified CBRs, risky features of the environment can be detected. Thus, by taking appropriate actions such as the modification of environment (e.g., removing obstacles, securing fall areas) as well as rehabilitation interventions (e.g., training to negotiate stairs and transitions), future falls are expected to be prevented.

ACKNOWLEDGMENT

Mina Nouredanesh received the Vector Institute Postgraduate Affiliate Award (from: Vector Institute, Toronto, Canada), AGE-WELL ACCESS Award, and AGE-WELL Graduate Student Award in Technology and Aging (from: AGE-WELL NCE (Canada's technology and aging network), Canada).

REFERENCES

- [1] K. A. Hartholt *et al.*, "Societal consequences of falls in the older population: Injuries, healthcare costs, and long-term reduced quality of life," *J. Trauma, Injury, Infection Crit. Care*, vol. 71, no. 3, pp. 748–753, 2011.
- [2] M. Terroso, N. Rosa, A. T. Marques, and R. Simoes, "Physical consequences of falls in the elderly: A literature review from 1995 to 2010," *Eur. Rev. Aging Phys. Activity*, vol. 11, no. 1, pp. 51–59, Apr. 2014.
- [3] L. Z. Rubenstein, C. M. Powers, and C. H. MacLean, "Quality indicators for the management and prevention of falls and mobility problems in vulnerable elders," *Ann. Internal Med.*, vol. 135, no. 8, pp. 686–693, 2001.
- [4] A. Shumway-Cook, S. Brauer, and M. Woollacott, "Predicting the probability for falls in community-dwelling older adults using the timed up & go test," *J. Amer. Phys. Therapy Assoc.*, vol. 80, no. 9, pp. 896–903, 2000.
- [5] V. Robles-García *et al.*, "Spatiotemporal gait patterns during overt and covert evaluation in patients with Parkinson's disease and healthy subjects: Is there a Hawthorne effect?" *J. Appl. Biomech.*, vol. 31, no. 3, pp. 189–194, Jun. 2015.
- [6] 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, pp. 178–190, Mar. 2021.
- [7] S. Del Din *et al.*, "Analysis of free-living gait in older adults with and without Parkinson's disease and with and without a history of falls: Identifying generic and disease-specific characteristics," *J. Gerontol.*, A, vol. 74, no. 4, pp. 500–506, Mar. 2019.

- [8] M. Mancini et al., "Continuous monitoring of turning mobility and its association to falls and cognitive function: A pilot study," *J. Gerontol. A, Biomed. Sci. Med. Sci.*, vol. 71, no. 8, pp. 1102–1108, 2016.
- [9] K. S. van Schooten et al., "Daily-life gait quality as predictor of falls in older people: A 1-year prospective cohort study," *PLoS ONE*, vol. 11, no. 7, Jul. 2016, Art. no. e0158623.
- [10] S. Lord, B. Galna, J. Verghese, S. Coleman, D. Burn, and L. Rochester, "Independent domains of gait in older adults and associated motor and nonmotor attributes: Validation of a factor analysis approach," *J. Gerontol. A, Biol. Sci. Med. Sci.*, vol. 68, no. 7, pp. 820–827, Jul. 2013.
- [11] M. E. Tinetti and C. Kumar, "The patient who falls: It's always a trade-off," *JAMA*, vol. 303, no. 3, pp. 258–266, 2010.
- [12] B. E. Maki and W. E. McIlroy, "Control of compensatory stepping reactions: Age-related impairment and the potential for remedial intervention," *Physiotherapy Theory Pract.*, vol. 15, no. 2, pp. 69–90, 1999.
- [13] J. M. Srygley, T. Herman, N. Giladi, and J. M. Hausdorff, "Self-report of missteps in older adults: A valid proxy of fall risk?" *Arch. Phys. Med. Rehabil.*, vol. 90, no. 5, pp. 786–792, 2009.
- [14] T. Gazibara et al., "Near-falls in people with Parkinson's disease: Circumstances, contributing factors and association with falling," *Clin. Neurol. Neurosurg.*, vol. 161, pp. 51–55, Oct. 2017.
- [15] T. Iluz et al., "Automated detection of missteps during community ambulation in patients with Parkinson's disease: A new approach for quantifying fall risk in the community setting," *J. Neuroeng. Rehabil.*, vol. 11, no. 1, p. 48, Dec. 2014.
- [16] S. Handelzalts, N. B. Alexander, N. Mastruserio, L. V. Nyquist, D. M. Strasburg, and L. V. Ojeda, "Detection of real-world trips in at-fall risk community dwelling older adults using wearable sensors," *Frontiers Med.*, vol. 7, p. 514, Sep. 2020.
- [17] L. V. Ojeda, P. G. Adamczyk, J. R. Rebola, L. V. Nyquist, D. M. Strasburg, and N. B. Alexander, "Reconstruction of body motion during self-reported losses of balance in community-dwelling older adults," *Med. Eng. Phys.*, vol. 64, pp. 86–92, Feb. 2019.
- [18] M. Nouredanesh, S. L. Kukreja, and J. Tung, "Detection of compensatory balance responses using wearable electromyography sensors for fall-risk assessment," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 1680–1683.
- [19] M. Nouredanesh and J. Tung, "IMU, sEMG, or their cross-correlation and temporal similarities: Which signal features detect lateral compensatory balance reactions more accurately?" *Comput. Methods Programs Biomed.*, vol. 182, Dec. 2019, Art. no. 105003.
- [20] O. Aziz, E. J. Park, G. Mori, and S. N. Robinovitch, "Distinguishing near-falls from daily activities with wearable accelerometers and gyroscopes using support vector machines," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2012, pp. 5837–5840.
- [21] M. Nouredanesh, K. Gordt, M. Schwenk, and J. Tung, "Automated detection of multidirectional compensatory balance reactions: A step towards tracking naturally occurring near falls," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 2, pp. 478–487, Feb. 2020.
- [22] J. Hauth et al., "Automated loss-of-balance event identification in older adults at risk of falls during real-world walking using wearable inertial measurement units," *Sensors*, vol. 21, no. 14, p. 4661, Jul. 2021.
- [23] M. Nouredanesh, A. W. Li, A. Godfrey, J. Hoey, and J. Tung, "Chasing feet in the wild: A proposed egocentric motion-aware gait assessment tool," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2018, pp. 176–192.
- [24] O. Aziz, M. Musngi, E. J. Park, G. Mori, and S. N. Robinovitch, "A comparison of accuracy of fall detection algorithms (threshold-based vs. machine learning) using waist-mounted tri-axial accelerometer signals from a comprehensive set of falls and non-fall trials," *Med. Biol. Eng. Comput.*, vol. 55, no. 1, pp. 45–55, 2017.
- [25] S. Del Din, A. Hickey, N. Hurwitz, J. C. Mathers, L. Rochester, and A. Godfrey, "Measuring gait with an accelerometer-based wearable: Influence of device location, testing protocol and age," *Physiol. Meas.*, vol. 37, no. 10, p. 1785, 2016.
- [26] L. Ojeda and J. Borenstein, "Non-gps navigation for security personnel and first responders," *J. Navigat.*, vol. 60, no. 3, pp. 391–407, Sep. 2007.
- [27] F. Madehkhaksar, J. Klenk, K. Sczuka, K. Gordt, I. Melzer, and M. Schwenk, "The effects of unexpected mechanical perturbations during treadmill walking on spatiotemporal gait parameters, and the dynamic stability measures by which to quantify postural response," *PLoS ONE*, vol. 13, no. 4, Apr. 2018, Art. no. e0195902.
- [28] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [29] A. Hof, S. Vermerris, and W. Gjaltema, "Balance responses to lateral perturbations in human treadmill walking," *J. Exp. Biol.*, vol. 213, no. 15, pp. 2655–2664, 2010.
- [30] T. Bhatt, J. D. Wening, and Y.-C. Pai, "Influence of gait speed on stability: Recovery from anterior slips and compensatory stepping," *Gait Posture*, vol. 21, no. 2, pp. 146–156, Feb. 2005.
- [31] S. Del Din, A. Godfrey, B. Galna, S. Lord, and L. Rochester, "Free-living gait characteristics in ageing and Parkinson's disease: Impact of environment and ambulatory bout length," *J. Neuroeng. Rehabil.*, vol. 13, no. 1, p. 46, Dec. 2016.
- [32] I. Hillel et al., "Is every-day walking in older adults more analogous to dual-task walking or to usual walking? Elucidating the gaps between gait performance in the lab and during 24/7 monitoring," *Eur. Rev. Aging Phys. Activity*, vol. 16, no. 1, pp. 1–12, Dec. 2019.
- [33] I. Galperin et al., "Associations between daily-living physical activity and laboratory-based assessments of motor severity in patients with falls and Parkinson's disease," *Parkinsonism Rel. Disorders*, vol. 62, pp. 85–90, May 2019.
- [34] S. M. Rispens et al., "Fall-related gait characteristics on the treadmill and in daily life," *J. Neuroeng. Rehabil.*, vol. 13, no. 1, pp. 1–9, 2016.
- [35] T. Iluz et al., "Can a body-fixed sensor reduce Heisenberg's uncertainty when it comes to the evaluation of mobility? Effects of aging and fall risk on transitions in daily living," *J. Gerontol. A, Biol. Sci. Med. Sci.*, vol. 71, no. 11, pp. 1459–1465, Nov. 2016.
- [36] B. S. R. Lazzarini and T. J. Kataras, "Treadmill walking is not equivalent to overground walking for the study of walking smoothness and rhythmicity in older adults," *Gait Posture*, vol. 46, pp. 42–46, May 2016.
- [37] A. Stisen et al., "Smart devices are different: Assessing and Mitigating Mobile sensing heterogeneities for activity recognition," in *Proc. 13th ACM Conf. Embedded Netw. Sensor Syst.*, Nov. 2015, pp. 127–140.
- [38] M. A. Brodie, S. R. Lord, M. J. Coppens, J. Annegarn, and K. Delbaere, "Eight-week remote monitoring using a freely worn device reveals unstable gait patterns in older fallers," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 11, pp. 2588–2594, Nov. 2015.
- [39] A. Hickey, S. Del Din, L. Rochester, and A. Godfrey, "Detecting free-living steps and walking bouts: Validating an algorithm for macro gait analysis," *Physiol. Meas.*, vol. 38, no. 1, pp. N1–N15, Jan. 2017.
- [40] C. Soaz and K. Diepold, "Step detection and parameterization for gait assessment using a single waist-worn accelerometer," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 5, pp. 933–942, May 2016.
- [41] A. H. Snijders, B. P. van de Warrenburg, N. Giladi, and B. R. Bloem, "Neurological gait disorders in elderly people: Clinical approach and classification," *Lancet Neurol.*, vol. 6, no. 1, pp. 63–74, 2007.
- [42] M. M. Rieger, S. Papegaaij, F. Steenbrink, J. H. van Dieën, and M. Pijnappels, "Perturbation-based gait training to improve daily life gait stability in older adults at risk of falling: Protocol for the REACT randomized controlled trial," *BMC Geriatrics*, vol. 20, no. 1, pp. 1–12, Dec. 2020.
- [43] Y. Okubo, D. L. Sturnieks, M. A. Brodie, L. Duran, and S. R. Lord, "Effect of reactive balance training involving repeated slips and trips on balance recovery among older adults: A blinded randomized controlled trial," *J. Gerontol. A*, vol. 74, no. 9, pp. 1489–1496, Aug. 2019.
- [44] Y.-C. Pai, T. Bhatt, F. Yang, and E. Wang, "Perturbation training can reduce community-dwelling older adults' annual fall risk: A randomized controlled trial," *J. Gerontol. A, Biol. Sci. Med. Sci.*, vol. 69, no. 12, pp. 1586–1594, Dec. 2014.